CHAPTER 2

Cointegration, long-run comovements, and long-horizon forecasting

James H. Stock

1 INTRODUCTION

Over the past decade, methodological advances in modeling and analyzing long-run relations have produced fundamental changes in the way that econometricians approach economic time-series data. Prior to these advances, econometric analysis of time-series data either ignored the problems which arise when regressors are highly persistent, or used a preliminary transformation to induce stationarity followed by analysis of the transformed variables. Now, the concept of cointegration, developed by Granger (1981), Granger and Weiss (1983), and Engle and Granger (1987), which builds on work on error correction models (Sargan (1964) and Davidson, Hendry, Srba, and Yeo (1978)), provides a powerful and widely adopted framework for studying long-run as well as short-run relations.

This chapter undertakes a selective survey of some recent theoretical work and focuses on some problems which remain in this area.[1] The thesis of this chapter is that, although the modern methodology of cointegration/unit root analysis provides a compelling framework for some problems, such as short-term forecasting using long-run information, for some applications this methodology has important limitations. The limitations of concern here arise when some of the regressors have an autoregressive root which is large but, in contrast to the basic assumption of cointegration analysis, not exactly one.

In empirical work it is common that an econometrician has evidence that the largest autoregressive root α in a series x_t is nearly one. For example, confidence intervals for the largest autoregressive root are given in table 2.1 for three economic time series (the confidence intervals are computed by inverting the Dickey–Fuller (1979) t-statistic using the method in Stock (1991)). While the Dickey–Fuller test fails to reject a unit root for all three

Table 2.1. *Largest autoregressive roots of three economic time series*

Series	#Obs.	DF t-stat	OLS estimate of sum of lags	Median-unbiased estimate of α	90% Confidence interval
US real GDP, 70:I–94:IV (quarterly)	100	−2.96	0.887	0.872	(0.744, 1.029)
Australian wool price 74:3–94:12 (monthly)	250	−2.28	0.959	0.989	(0.940, 1.015)
90-day US Treasury bill rate, 61:9–94:12 (monthly)	400	−2.12	0.971	0.985	(0.962, 1.006)

Notes: The "DF t-stat" is the Dickey–Fuller (1979) t-statistic testing the unit root hypothesis. The OLS estimate of the sum of the lag coefficients is one plus the coefficient on the lagged level term in the Dickey–Fuller regression. The median-unbiased estimate and the 90 percent confidence interval were computed by inverting the Dickey–Fuller test statistic using the method of Stock (1991). The number of lags in the Dickey–Fuller regression was selected by the Bayes information criterion (BIC). For the 90-day US Treasury bill rate, a constant was included and the BIC chose 12 lags; for US real GDP and Australian wool prices, a constant and a time trend were included and BIC chose 1 and 0 lags, respectively.

series, neither do we reject roots close to one. Absent economic theory that specifies a particular value of α, the confidence interval in table 2.1 provides no objective basis for inferring, for example, that for the Australian wool price series $\alpha = 1.00$ rather than 0.98.

This chapter is organized around three of the practical problems which motivated much of the original theoretical work on unit roots and cointegration:

A Tests of the hypothesis that x_{t-1} does not predict y_t when x_t is serially correlated and possibly has a unit autoregressive root. A leading example is tests in linear rational expectations models with highly autocorrelated regressors, for example, tests of whether stock returns are predictable by lagged variables such as the dividend yield (cf. Mankiw and Shapiro (1985), Fama (1991)).

B Inference about the parameters of long-run relationships. A classic example is construction of a confidence interval for the income elasticity in a levels consumption function relating log consumption and log income.

C Long-run forecasting and the construction of forecast intervals for y_{T+k} or x_{T+k}, where the forecast horizon k is long in the sense that it is a non-trivial fraction λ (e.g., $\lambda \geq 0.1$) of the sample size T.

In each of these problems, if x_t is I(1) ($\alpha = 1$) then the methods of stationary time-series econometrics will generally be inappropriate but these three problems are well handled (asymptotically) using the technology of integration and cointegration. However, if α is large but not exactly one, then both the I(0) approach, which treats x_t as stationary, and the I(1) approach, which treats x_t as having an exact unit root, can produce systematic errors in inference.

In each of these three problems, the relevant distributions depend on the nuisance parameter α. A common method for handling this nuisance parameter is first to pretest for unit roots and/or cointegration and, when indicated, to impose a unit root and/or cointegration. Assuming the pretest is ideal in the sense that type I errors tend to zero in large samples, this yields asymptotically valid inference when $\alpha = 1$.[2] If, however, α is large but not necessarily one, this pretest strategy will not deliver reliable inference in problems A, B, or C, even with an ideal pretest. For example, in problem A, controlling size asymptotically means that, under the null hypothesis, the limit of the supremum (taken over α) of the rejection rate should be at most 5 percent, say. Although the supremum of the limit of the rejection rates under this pretest procedure is controlled when α is fixed, the limit of the supremum is not, and this results in size distortions which persist asymptotically. A Monte Carlo experiment in section 3 demonstrates that these size distortions can be large, particularly in problem B.

Section 2 gives precise definitions of concepts related to I(1) regression and briefly reviews the modern approach to problems A, B, and C; this will be referred to as the "I(1)" methodology. Section 3 examines the performance of the I(1) approach to problems A, B, and C when α is large but not exactly one. While perhaps this could be done using finite sample techniques, such a treatment would be cumbersome at best. Because the key source of difficulties is the single nuisance parameter α when it is large, the analysis of section 3 uses an asymptotic approach which provides great simplifications over a finite-sample treatment, yet retains the essential dependence on α. This is done by modeling the largest autoregressive root as local to unity, specifically, in a T^{-1} neighborhood of one. It has been documented elsewhere (see Stock (1994a) for references) that this device provides good approximations to finite sample distributions when α is large and there are 100 or more observations. Thus, these asymptotics provide a magnifying glass which focuses on the problematic dependence of the finite-sample distributions on α.

Section 5 contains a review of several alternative, currently unconventional approaches to problems A–C. One general conclusion, presented in section 6, is that, despite the significant advances in this literature, state-of-the-art techniques fail to provide acceptable inference in these important applications. This should not be taken as a sweeping indictment of research on cointegration and unit roots, for indeed much of value has been learned. However, the conclusion suggests that current techniques typically do not address adequately some of the problems which originally motivated this line of research, and this in turns suggests directions for future work.[3]

2 THE I(1) APPROACH TO PROBLEMS A, B, AND C

In this chapter, a univariate time series u_t, $t = 1, \ldots, T$, is said to be I(0) if the sequence of its rescaled partial sums, $T^{-\frac{1}{2}}\Sigma_{t=1}^{[Ts]}u_t$ (where $[\cdot]$ denotes the greatest lesser integer function), obeys a functional central limit theorem and converges weakly to a stochastic process which is proportional to a Brownian motion. This is implied by various primitive assumptions including mixing conditions (Herrndorf (1984)) or by u_t being a one-summable linear moving average of martingale difference sequences (Hall and Heyde (1980), Phillips and Solo (1992)). In this section, attention is limited to the case that the stochastic component of x_t is I(1).[4]

2.1 Problem A: regression when x_t is I(1)

When x_t is I(1), problem A is a special case of the more general regression problem

$$y_t = \delta_1' z_{1t} + \delta_2 + \delta_3' z_{3t} + \xi_t, \tag{1}$$

where ξ_t is a martingale difference sequence with respect to its lags and to $(z_{1t}', z_{3t}')'$ and its lags, $E\xi_t^2 = \sigma_\xi^2$, z_{1t} consists of I(0) variables (taken without loss of generality to have mean zero), and z_{3t} consists of I(1) variables which are not themselves cointegrated (the spectral density of Δz_{3t} at frequency zero has full rank). Problem A is nested in (1) by letting $z_{1t} = x_{t-1}$ if x_t is I(0) and omitting z_{3t}, or by letting $z_{3t} = x_{t-1}$ if x_t is I(1) and omitting z_{1t}.

Another example of a regression nested in (1) is the augmented Dickey–Fuller (1979) regression, which is obtained by setting $z_{1t} = (\Delta y_{t-1}, \ldots, \Delta y_{t-p})$ and $z_{3t} = y_{t-1}$.

Asymptotic distribution theory for ordinary least squares (OLS) estimation of (1) has been developed by Chan and Wei (1988), Park and Phillips (1988), and Sims, Stock, and Watson (1990). Asymptotically, the moment matrix of the regressors is block diagonal; $\{T^{\frac{1}{2}}(\hat{\delta}_1 - \delta_1), T^{\frac{1}{2}}(\hat{\delta}_2 - \delta_2), T(\hat{\delta}_3 - \delta_3)\}$ have a joint limiting distribution where $T^{\frac{1}{2}}(\hat{\delta}_1 - \delta_1)$ is independent of $\{T^{\frac{1}{2}}(\hat{\delta}_2 - \delta_2), T(\hat{\delta}_3 - \delta_3)\}$; and $T^{\frac{1}{2}}(\hat{\delta}_1 - \delta_1) \xrightarrow{d} N(0, (Ez_{1t}z'_{1t})^{-1}\sigma_\xi^2)$. Although tests involving only δ_1 have conventional χ^2 distributions, in general, the limiting distribution of the Wald tests involving δ_3 is non-standard (an exception is when the cross spectral density of $(\Delta z_{3t}, \xi_t)$ is zero at frequency zero, in which case the distribution is χ^2). In general, the Wald test statistic has a representation as a functional of Brownian motion. For example, in the augmented Dickey–Fuller example above, the t-statistic testing $\delta_3 = 1$ has the Dickey–Fuller (1979) "$\hat{\tau}_\mu$" distribution. When the distribution depends on nuisance parameters, such as a spectral density at frequency zero, it typically can be computed by simulation using consistent estimators of the nuisance parameters.

The implication of these results for application A is that the distribution of the t-statistic testing whether x_{t-1} predicts y_t will in general have either a $N(0, 1)$ distribution if $|\alpha| < 1$, or a non-standard distribution if $\alpha = 1$. The non-standard distribution is readily computed by Monte Carlo simulation.

2.2 Problem B: cointegrating regressions

Let Y_t be an $n \times 1$ vector time series. Engle and Granger (1987) defined Y_t to be cointegrated if each element of Y_t is I(1) but there exists an $n \times r$ matrix β of full column rank with $1 \le r < n$ such that $\beta' Y_t$ is I(0). Various forms are available for representing cointegrated variables. Three useful, equivalent forms are the levels vector autoregression (VAR) representation, the vector error correction model (VECM) form (Engle and Granger (1987)), and the triangular form (Campbell and Shiller (1987), Phillips (1991)); for discussion see Engle and Yoo (1991) and Watson (1994). Here, we work with the triangular form, in which Y_t is partitioned into an $r \times 1$ vector y_t and a $k \times 1$ vector x_t where $k = n - r$, and the cointegrating matrix is conformably normalized as $\beta = (I_r, -\theta)'$, where θ is $r \times k$. Including a constant term, the triangular form is

$$x_t = \mu_x + v_t, \Delta v_t = u_{1t}, \tag{2a}$$

$$y_t - \theta x_t = \mu_y + u_{2t}, \tag{2b}$$

where $u_t = (u'_{1t}\, u'_{2t})'$ is a I(0) vector process, the spectral density of which at frequency zero, Ω, has full rank. It is assumed throughout that v_0 is $O_p(1)$.

Various estimators are available for estimation of the cointegrating parameter θ. Although the OLS estimator of the regression of y_t on x_t is consistent, in general it has bias of order $O(T^{-1})$ and in finite samples this bias can be large (Stock (1987)). This $O(T^{-1})$ bias can be avoided by using an asymptotically efficient estimator. Many such estimators have been proposed (see Hargreaves (1994) for a partial list). Some are readily motivated as Gaussian maximum likelihood estimators (MLEs), or approximate MLEs, in one of the representations above. One of the most common MLEs is the Johansen (1988)/Ahn and Reinsel (1990) estimator, which is the Gaussian MLE in a finite order VECM; see Watson (1994) for further discussion.

The triangular form provides a convenient alternative starting point for developing efficient estimators. As a motivation, suppose that u_t is Gaussian, and orthogonalize the system by projecting u_{2t} on to u_{1t}, that is, write $u_{2t} = E(u_{2t} \mid \{u_{1t}\}) + \tilde{u}_{2t} = d(L)u_{1t} + \tilde{u}_{2t}$, where $d(L)$ is two sided. By construction, $\{\tilde{u}_{2t}\}$ and u_{1t} are independent. Thus (2b) becomes, $y_t = \theta x_t + d(L)\Delta x_t + \tilde{u}_{2t}$. Thus, if $d(L)$ has finite order, by the theory of seemingly unrelated regressions, θ can be estimated efficiently by generalized least squares (GLS) regression of y_t onto x_t and leads and lags of Δx_t. However, because x_t is I(1), the GLS and OLS estimators of θ in this regression are asymptotically equivalent, so OLS estimation of this augmented regression asymptotically yields the MLE. This approach can be thought of as OLS with additional variables and variants of it have been studied by Phillips and Loretan (1991), Saikkonen (1991), and Stock and Watson (1993); the particular estimator just described is the "dynamic OLS" (DOLS) estimator of Stock and Watson (1993).

The various efficient estimators of cointegrating vectors have the same asymptotic distribution, which is a random mixture of normals (Johansen (1988), Phillips (1991)). Importantly, standard Wald statistics testing q restrictions on θ, based on an efficient estimator $\hat{\theta}$, have asymptotic χ_q^2 distributions. Thus, despite the non-standard nature of the problem and the non-standard distribution of the OLS estimator and the MLE, hypothesis tests on θ and confidence intervals for θ can be constructed using standard techniques. These powerful yet simple methods for inference on θ are arguably the greatest single accomplishment of cointegration theory. These results are the foundation of the now-conventional cointegration approach to problem B, in which an efficient estimator is used to estimate the cointegration coefficients and confidence intervals are constructed based on this estimate and its standard error.

2.3 Problem C: long-run forecasting

Depending on the orders of integration and cointegration, standard models for generating forecasts are the levels VAR (stationary series), the VECM (cointegrated series), and the differences VAR (I(1), no cointegration). If the series being forecast do not contain deterministic trends, then computation of asymptotic prediction intervals conditional on the chosen model is straightforward. Parameter uncertainty becomes more important when the series contain deterministic trend components (Sampson (1991)).

3 DISTRIBUTIONS WHEN α IS LOCAL TO UNITY

This section analyzes the behavior of the I(1) procedures of the previous section when α is large but not necessarily one. As discussed in the introduction, in theory this could be done using finite sample techniques, but a simpler and more incisive treatment is possible by focusing on the problematic large root, in particular setting $\alpha = 1 + c/T$, where c is a constant, and using asymptotic approximations. In this case, the stochastic part of $T^{-\frac{1}{2}}x_{[Ts]}$ converges to an Ornstein–Uhlenbeck process (Bobkoski (1983), Cavanagh (1985), Chan (1988), Chan and Wei (1987), and Phillips (1987)). Distributions of functionals of the Ornstein–Uhlenbeck process are described by the single parameter c; however, c is not consistently estimable, although confidence intervals for c can be constructed (Stock (1991), Andrews (1993)).

The asymptotic power of efficient tests of an autoregressive unit root with fixed size against the alternative $\bar{\alpha} = 1 + \bar{c}/T$ is strictly less than one (Elliott, Rothenberg and Stock (1996), Rothenberg and Stock (1995)). It follows that if the critical value tends to infinity with the sample size, as needed if the test is to distinguish I(1) from I(0) processes consistently, then the asymptotic power of the test against the local-to-unity alternative is zero.[5]

3.1 Problem A with local-to-unity regressors[6]

Consider the bivariate model

$$x_t = \mu_x + v_t, (1 - \alpha L)v_t = u_{1t}, u_{1t} = a(L)^{-1}\varepsilon_{1t} \tag{3a}$$

$$y_t = \mu_y + \gamma x_{t-1} + u_{2t}, u_{2t} = \varepsilon_{2t}, \tag{3b}$$

where $a(L) = \Sigma_{i=0}^{k} a_i L^i, a_0 = 1$. Suppose interest is in testing the hypothesis $\gamma = \gamma_0$; in the linear rational expectations application discussed in the introduction, $\gamma_0 = 0$. Further suppose that $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ is a martingale difference sequence with $E(\varepsilon_t \varepsilon_t' \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) = \Sigma = [\sigma_{ij}], \sup_t E\varepsilon_{it}^4 < \infty, i = 1, 2, Ev_0^2 < \infty$, and the roots of $a(L)$ are fixed and greater than one.

Let $\Omega = [\omega_{ij}]$ be 2π times the spectral density of $u_t = (u_{1t}, u_{2t})'$ at frequency zero, so $\omega_{11} = \sigma_{11}/a(1)^2$ and $\omega_{22} = \sigma_{22}$. Assume that the functional central limit theorem $T^{-\frac{1}{2}}\Sigma_{t=1}^{[T\lambda]}(u_{1t}/\omega_{11}^{\frac{1}{2}}, u_{2t}/\omega_{22}^{\frac{1}{2}})' \Rightarrow W$ is satisfied, where "$\Rightarrow$" denotes weak convergence on $D[0, 1]$, where $W = (W_1, W_2)'$ is a two-dimensional Brownian motion with covariance matrix $\bar{\Omega} = [\bar{\omega}_{ij}]$, where $\bar{\omega}_{11} = \bar{\omega}_{22} = 1$ and $\bar{\omega}_{12} = \bar{\omega}_{21} = \delta = \omega_{12}/(\omega_{11}\omega_{22})^{\frac{1}{2}}$.

Let t_γ denote the t-statistic testing the $\gamma = \gamma_0$ in (3b). Then, under the null

$$t_\gamma \Rightarrow \tau_{2,c} = \delta\tau_{1,c} + (1 - \delta^2)^{\frac{1}{2}} z \tag{4}$$

where $\tau_{1,c} = (\int J_c^{\mu 2})^{-\frac{1}{2}} \int J_c^{\mu} dW_1, \tau_{2,c} = (\int J_c^{\mu 2})^{-\frac{1}{2}} \int J_c^{\mu} dW_2, J_c^{\mu}(s) = J_c(s) - \int_0^1 J_c(r)dr$, J_c is the Ornstein–Uhlenbeck process which obeys $dJ_c(s) = cJ_c(s) + dW_1(s)$, and z is a standard normal random variable distributed independently of (W_1, J_c).

Evidently the limiting distribution of t_γ depends on both c and δ. Because δ is consistently estimable (in general, by the sample cross-spectral correlation of u_t at frequency zero, here by the sample correlation between the OLS residuals $\hat{\varepsilon}_{1t}$ and $\hat{\varepsilon}_{2t}$), δ will be treated as known. However, c is not consistently estimable, and, if $\delta \neq 0, t_\gamma$ has a non-standard asymptotic distribution. Thus asymptotic inference cannot in general rely on simply substituting a suitable estimator $\hat{c}$ for c when selecting critical values for tests of γ.

Suppose the test based on t_γ is performed following a consistent pretest for a unit root in x_t: if the unit root test rejects, the $N(0, 1)$ distribution is used, if not, (4) is used. Evidently, the size of this procedure is $\sup_c P[\delta\tau_{1c} + (1 - \delta^2)^{\frac{1}{2}} z \notin (d_0, d_1)]$, where (d_0, d_1) are the critical values of $\tau_{2,0}$. Numerical evaluation indicates that the size can be large. For example, for tests with nominal level 5 percent, if $\delta = 0.9$ the size is 0.33 if a constant is included in the regression (3b), and is 0.64 if a constant and time trend are included.

3.2 Problem B with local-to-unity regressors

Consider the triangular form in the bivariate case with intercepts and where α is the largest root of x_t

$$x_t = \mu_x + v_t, (1 - \alpha L)v_t = u_{1t} \tag{5a}$$

$$y_t = \mu_y + \theta x_t + u_{2t} \tag{5b}$$

where the value of θ, θ_0, is taken to be non-zero. As in (3), let u_t have spectral density Ω and assume $T^{-\frac{1}{2}}\Sigma_{t=1}^{[T\lambda]}(u_{1t}/\omega_{11}^{\frac{1}{2}}, u_{2t}/\omega_{22}^{\frac{1}{2}}) \Rightarrow W$, where W is defined following (3). Note that, in contrast to (3b), in (5b) u_{2t} is not restricted to be a martingale difference sequence but rather is a general I(0) process.

If α is local to unity, then consistent pretests for unit roots and cointegration in (x_t, y_t) would lead one to conclude that x_t and y_t are I(1) and cointegrated, so that θ should be estimated using an efficient cointegration estimator $\hat{\theta}$. However, Elliott (1994) showed that, even if u_t has a parametric structure of known order, if $\alpha = 1 + c/T$ and θ is a scalar, $\hat{\theta}$ and its t-statistic t_θ testing $\theta = \theta_0$ have the limits

$$\{T(\hat{\theta} - \theta_0), t_\theta\} \Rightarrow \{ -\omega_{11}^{-1/2}\omega_{2.1}^{1/2}\int J_c^\mu dW_{2.1}(\int J_c^{\mu 2})^{-1} - c\delta\omega_{11}^{-1/2}\omega_{22}^{1/2}, z - c\delta(1-\delta^2)^{-1/2}(\int J_c^{\mu 2})^{1/2}\}, \quad (6)$$

where $W_{2.1}$ is a standard Brownian motion, $w_{2.1} = w_{22} - w_{21}^2/w_{11}$, and z is a standard normal random variable, where $(z, W_{2.1})$ are independent of (W_1, J_c). As in problem A, the limiting random variables in (6) have distributions which depend on c.

The implications parallel those for problem A. Tests of $\theta = \theta_0$ typically exhibit size distortions, even asymptotically, and confidence intervals for θ can have coverage rates less than their nominal asymptotic value. In fact, the asymptotic size of the Wald test of $\theta = \theta_0$ based on t_θ with standard normal critical values is arbitrarily close to one, depending on δ.

3.3 Problem C with local-to-unity regressors

The presence of a large but not necessarily unit root substantially complicates the problem of making long-run point and interval forecasts.[7] To be concrete, consider the bivariate cointegrating model (5), where it is further supposed that (u_{1t}, u_{2t}) obeys a VAR(p) with all roots strictly outside the unit circle and where $Ev_0^2 < \infty$. Then in levels, (x_t, y_t) obeys a VAR($p + 1$) with one root local to the unit circle.

Consider two alternative forecasting strategies. In the first, the econometrician computes forecasts from a levels VAR. In the second, the econometrician pretests for the number of unit roots in the system, after which forecasts are constructed using the levels VAR (no unit roots detected), a VECM with $\alpha = 1$ imposed (one unit root), or a differences VAR (two unit roots). To simplify the analysis, suppose that a consistent I(0)/I(1) classification procedure is used, so that in sufficiently large samples the pretest strategy, applied to (5), delivers a VECM with probability approaching one. Finally, suppose that p is known.

At short horizons, the levels VAR and the pretest procedures yield asymptotically equivalent forecasts and asymptotically valid prediction intervals when $\alpha = 1 + c/T$.[8] However, the problem of long-term forecasting is more complicated. Specifically, model the forecast horizon k as being a constant fraction λ of the sample size, so $k = [T\lambda]$. Let $x_{s|t}$ denote the conditional mean of x_s given $(x_t, y_t, x_{t-1}, y_{t-1}, \ldots)$. Then $T^{-1/2}(x_{T+k|T}, y_{T+k|T}) = T^{-1/2}e^{c\lambda}(x_T, \theta x_T) + o_p(1)$. Moreover, $T^{-1/2}x_{T+k} \Rightarrow e^{c\lambda}J_c(1) + \phi_c(\lambda)$ and $T^{-1/2}(y_{T+k} - \theta x_{T+k}) \xrightarrow{p} 0$, where J_c is as defined in section 3.1, $\phi_c(\lambda)$ is $N(0, \omega_{11}(1 - e^{2c\lambda})/(1 - e^{2c}))$, and ϕ_c and J_c are independent.

First consider the levels VAR procedure, with forecasts $(\hat{x}_{T+k|T}, \hat{y}_{T+k|T})$. Then $T^{-1/2}(x_{T+k} - \hat{x}_{T+k|T}) \Rightarrow [1 - e^{(c^* - c)\lambda}]e^{c\lambda}J_c(1) + \phi_c(\lambda)$, where c^* is a $O_p(1)$ random variable which depends on J_c; thus the long-run levels VAR forecast is conditionally biased of order $O(T^{1/2})$, with conditional bias $E[T^{-1/2}(x_{t+k} - \hat{x}_{T+k|T}) | x_T] \to E[(1 - e^{(c^* - c)\lambda}) | J_c(1)]e^{c\lambda}J_c(1)$. The magnitude of this bias depends on the distribution of c^* which depends on nuisance parameters. Typically, however, c^* will be biased toward zero, so the forecast will be biased toward zero as well. Moreover, because the conditional distribution of x_{T+k} depends on c but c is not consistently estimated, prediction intervals computed using standard (stationary) first-order asymptotics will have incorrect tolerance levels.

The pretest estimator is also conditionally biased and produces invalid long-run prediction intervals. For the pretest forecast, $T^{-1/2}(x_{T+k} - \tilde{x}_{T+k|T}) \Rightarrow (1 - e^{c\lambda})J_c(1) + \phi_c(\lambda)$. The limiting conditional bias is $E[T^{-1/2}(x_{T+k} - \tilde{x}_{T+k|T}) | x_T] \to (1 - e^{c\lambda})J_c(1)$, which for $c < 0$ is biased away from zero. Imposition of the unit root produces incorrect inference on $e^{c\lambda}$ and $\phi_c(\lambda)$ which results in prediction intervals with an incorrect tolerance level.

As in problems A and B, the central difficulty is the dependence of a key distribution (here, the conditional distribution of (x_{T+k}, y_{T+k})) on c, which cannot be consistently estimated.

3.4 Finite sample simulation

A small Monte Carlo experiment was performed to assess whether these issues are important in sample sizes typically encountered in econometric applications. The design consists of equations for income (y_t), consumption (c_t), and the unforecastable excess returns (r_t) on a portfolio of equities. The parameters of the income autoregression were chosen to be similar to those estimated for US real GDP, 70:I–94:IV (the first series in table 2.1). Accordingly

$$\tilde{\Delta} y_t = 0.3\tilde{\Delta} y_{t-1} + \zeta_{1t}, \quad (7a)$$

$$c_t = \theta y_t + \zeta_{2t}, \quad (7b)$$

$$r_t = \zeta_{3t} \quad (7c)$$

Table 2.2. *Monte Carlo results*

			I(0) methodology			I(1) methodology		
T	α	c	(A)	(B)	(C)	(A)	(B)	(C)
100	1.00	0	0.25	0.83	0.35	0.05	0.94	0.67
	0.975	−2.5	0.15	0.90	0.50	0.07	0.83	0.72
	0.95	−5	0.12	0.93	0.56	0.11	0.66	0.78
	0.90	−10	0.08	0.95	0.62	0.16	0.38	0.86
400	1.00	0	0.25	0.80	0.36	0.05	0.94	0.68
	0.99375	−2.5	0.15	0.88	0.50	0.07	0.84	0.75
	0.9875	−5	0.11	0.91	0.55	0.11	0.63	0.80
	0.975	−10	0.09	0.92	0.62	0.16	0.32	0.90
Nominal			0.05	0.95	0.68	0.05	0.95	0.68

Notes: Columns (A) contain Monte Carlo rejection rates of tests with nominal level 5 percent. Columns (B) contain Monte Carlo coverage rates of confidence intervals with nominal coverage rate 95 percent. Columns (C) contain Monte Carlo coverage rates of prediction intervals with nominal coverage rate 68 percent. The design, statistics and methodologies are described in the text. All regressions are run including a constant, except for the first-differences regression used to generate the I(1) forecasts, in which the constant was supressed. Results are based on 10,000 Monte Carlo replications.

where $\tilde{\Delta} = 1 - \alpha L$, ζ_t is i.i.d. $N(0, \Sigma)$ with $\Sigma_{ii} = 1$, $i = 1, \ldots, 3$, $\Sigma_{12} = 0.8$, $\Sigma_{13} = 0.9$, $\Sigma_{23} = 0.5$, and $\theta = 1$. For $T = 100$, the chosen values of α (1, ·975, ·95, ·90) fall within the 90 percent confidence interval for α for real GDP given in table 2.1; for $T = 400$, the chosen values (1, ·99375, ·9875, ·975) fall within the 90 percent confidence interval for α for the 90-day Treasury bill rate. For both sample sizes, $c = T(\alpha - 1)$ is 0, −2.5, −5 and −10.

Stylized versions of problems A, B, and C are examined by: (A) testing whether y_{t-1} forecasts r_t at the 5 percent significance level in a regression of r_t onto y_{t-1}; (B) constructing a 95 percent confidence interval for θ; and (C) constructing a 68 percent prediction interval for y_{T+k} based on a univariate forecast, where $k = 0.25T$. These problems were analyzed using two methodologies. In the "I(0)" methodology, unit root issues are ignored: in problem A, $N(0, 1)$ critical values were employed to test $\gamma = 0$ using t_γ; in problem B, the confidence intervals for θ were constructed using two stage least squares (TSLS) with y_{t-1} as an instrument (this is asymptotically full information maximum likelihood in (7) when α is unknown); and in problem C the 68 percent prediction interval for y_{T+k} is computed using standard methods after estimating an AR(2) in levels.

In the "I(1)" methodology, the econometrician assumes that $\alpha = 1$ and that (c_t, y_t) are cointegrated.[9] Accordingly, in problem A, the representation (4) is used to obtain critical values for t_γ; in problem B, the hypothesis $\theta = 1$ is tested using the Wald test statistic from the MLE for the (c_t, y_t) system (the MLE is the DOLS estimator with contemporaneous Δy_t only); and in problem C, 68 percent prediction intervals for y_{T+k} are computed from OLS estimation of an AR(1) in first differences.

The results are summarized in table 2.2. In confirmation of findings by numerous researchers, when $\alpha = 1$ the I(0) methodology does poorly, with large size distortions in problems A and B and prediction intervals which have low coverage, while the I(1) methodology does well. However, when the root is slightly less than one, the I(1) methodology fails. Confidence intervals for θ based on the efficient cointegrating estimator have low probabilities of containing the true value, and prediction intervals based on $c = 0$ are too wide (they do not reflect long-run mean reversion when $c < 0$). It should be stressed that these problems occur for values of α within the 90 percent confidence intervals for US real GDP and the 90-day Treasury bill rate in table 2.1. Under the local-to-unity nesting, the distortions are relatively stable as the sample size increases for fixed c, which accords with results elsewhere that the local-to-unity asymptotics provide a good guide to the finite distributions.

4 POTENTIAL SOLUTIONS: PROBLEMS A AND B

This section reviews alternative approaches to problems A and B which solve the size distortion problems in certain cases. This discussion focuses on the case of no deterministic terms in either the data generation process or the various statistics, although some remarks are made about extension to the case of deterministic terms. The four approaches reviewed here are conservative tests based on least squares statistics, non-parametric tests based on sign and/or rank statistics, tests of whether the regression errors are I(0), and variable augmentation schemes.[10] A detailed analysis of finite-sample and asymptotic power of these approaches is beyond the scope of this survey (see however Campbell and Dufour (1994, 1995) for problem A).

4.1 Asymptotically conservative intervals and tests

An alternative approach to inference is to use the limiting distributions of the relevant t-statistics, which depend on c, and then to evaluate these for a range of c. Cavanagh, Elliott, and Stock (1995) considered bounds tests in problem A, and their results are briefly reviewed here (also see Dufour (1990), who derived exact bounds tests for regression parameters with Gaussian AR(1) errors and strictly exogenous regressors). The extension of this approach to problem B is then briefly discussed.

First consider problem A, so that (3) holds with $\mu_x = \mu_y = 0$. Critical values for a sup-bound test are chosen so that the test based on t_y has size no more than the prescribed level, uniformly in c. For example, 5 percent critical values can be obtained as the extrema (over c) of the 2.5 percent quantiles of the limiting random variable in (4). Alternatively, valid and less-conservative critical values can be obtained by choosing the upper and lower critical values so that the supremum (over c) of the rejection rate equals the desired level of the test. This test will be conservative in the sense that, for some c, the rejection rate under the null will be less than the level of the test. Tests which are potentially less conservative can be constructed by first computing a $100(1-\eta_1)$ percent confidence set for c (using Stock's (1991) or Andrews' (1993) method), and then rejecting if t_y rejects using a $100\eta_2$ percent two-sided test based on critical values from (4) for every c in the first-stage confidence set. By Bonferroni's inequality, the combined procedure has size no more than $\eta_1 + \eta_2$. As with the sup-bound tests, η_1 and η_2 can be chosen so that the asymptotic size equals the desired level. Extension of this approach to additional deterministic terms is straightforward.

In theory, the Bonferroni approach can be extended to problem B with some modifications. To be concrete, suppose that θ is estimated using an efficient cointegration estimator which incorrectly imposes $\alpha = 1$ when in fact $\alpha = 1 + c/T$, so that the associated t-statistic testing $\theta = \theta_0$ has the limiting representation in (6). As in (4), the only nuisance parameters entering the second expression in (6) are δ, which is consistently estimable, and c. This permits construction of asymptotic confidence sets for θ given c. A first-stage confidence set for c can be constructed from univariate methods as discussed above. Unlike (4), however, the quantiles of (6) are not bounded in c, so the sup-bound test for the cointegration problem has infinite critical values. It also seems likely that Bonferroni confidence regions based on an efficient cointegration estimator will be quite wide, making this approach less appealing in problem B.

4.2 Non-parametric tests

Campbell and Dufour (1995) consider problem A (with no intercept term) and show that non-parametric tests achieve the desired size under weak conditions on x_t. For example, consider the sign test statistic

$$S(\gamma) = T^{-\frac{1}{2}}\Sigma_{t=2}^{T}\{\mathbf{1}[(y_t - \gamma x_{t-1})x_{t-1} > 0] - \tfrac{1}{2}\}. \tag{8}$$

If the conditional distribution of ε_{2t} given x_{t-1} has median zero, $\{\mathbf{1}(\varepsilon_{2t}x_{t-1} \geq 0)\}$ are i.i.d. Bernoulli random variables under general conditions on x_t, including but not limited to x_t being stationary, I(1), or local to I(1). Thus, under the null $\gamma = \gamma_0$, $S(\gamma_0)$ is a binomial random variable with an asymptotic $N(0, 1/4)$ distribution.

The local asymptotic power function of the sign test can be derived assuming that α is local to one. Specifically, suppose that (3) holds. If $\delta = 0$, the asymptotic power of the sign test of level η (with standard normal critical value d_η) against the alternative $\gamma = b/T$ is

$$\begin{aligned} P[|2S| > d_\eta] \to E[&\Phi(-d_\eta - 2bf(0)\omega_{11}^{\frac{1}{2}}\textstyle\int|J_c|) \\ &+ \Phi(-d_\eta + 2bf(0)\omega_{11}^{\frac{1}{2}}\textstyle\int|J_c|)], \end{aligned} \tag{9}$$

where f is the p.d.f. of ε_{2t} conditional on x_{t-1}.

This approach can be extended to problem B by considering an instrumental variables version of the sign test statistic (8). In the cointegration context, in general x_t is endogenous and u_{2t} is serially correlated, so $\{\mathbf{1}(u_{2t}x_t > 0\}$ are not independent Bernoullis. Suppose, however, that u_{2t} is a $(q-1)$th order moving average process, specifically, that the distribution (with p.d.f. f) of u_{2t} conditional on x_{t-q} has median zero. Accordingly, let

$$R_q(\theta) = T^{-\frac{1}{2}}\Sigma_{t-q}^{T}\{\mathbf{1}[(y_t - \theta x_t)x_{t=q} > 0] - \tfrac{1}{2}\}. \tag{10}$$

Under the null $\theta = \theta_0$, if $\alpha = 1 + c/T$ then $R_q(\theta_0) \xrightarrow{d} N(0, V)$, where $V = \Sigma_{i=-q}^{q}\{\frac{1}{2}P[\text{sgn}(u_{2t}) = \text{sgn}(u_{2t-i})] - 1/4\}$, where sgn($z$) is the sign of z. This distribution does not depend on c and thus holds if x_t is I(1) or has a local-to-unit root. A natural estimator for V is $\hat{V}(\theta) = \Sigma_{i=-q}^{q}\text{cov}(\mathbf{1}[(y_t - \theta x_t)x_{t-q} > 0], \mathbf{1}[(y_{t-i} - \theta x_{t-i})x_{t-q-i} > 0])$; then $\hat{V}(\theta_0) \xrightarrow{p} V$ uniformly in c. In (10), x_{t-q} serves the role of an instrument for x_t.

The extension of this approach to a non-zero intercept μ_y in (5b) is not automatic and appears to require joint or preliminary inference about μ_y; cf. Campbell and Dufour (1994).

4.3 Tests for residuals being I(0)

In problem A, y_t has two components, ε_{2t} which is serially uncorrelated and γx_{t-1} which is local to I(1). This suggests testing the hypothesis $\gamma = 0$ by testing whether y_t is a martingale difference sequence against the alternative that it is a random walk plus noise. Results in King (1980) and King and

Hillier (1985) can be used to derive the family of efficient tests of the null that y_t is i.i.d. Gaussian against the alternative that it is the sum of an i.i.d. Gaussian process and an independent (unobserved) Gaussian random walk (Shively (1988)). In the case that the intercept is possibly non-zero, Nyblom and Mäkeläinen (1983) showed that the locally most powerful invariant (LMPI) test of this hypothesis rejects for large values of $L^{\mu} = T^{-2}\Sigma_{t=1}^{T}(\Sigma_{s=1}^{t} y_s^{\mu})^2/\hat{\sigma}_y^2$, where $y_s^{\mu} = y_s - \bar{y}$ and $\hat{\sigma}_y^2 = T^{-1}\Sigma_{t=1}^{T} y_s^{\mu 2}$, where $\bar{y}$ is the sample mean of y_t (also see Nyblom (1986), Nabeya and Tanaka (1988), Tanaka (1990), and Saikkonen and Luukkonen (1993a)). Under the null, the test has a limiting distribution which is free of nuisance parameters and is the mean square of a scalar Brownian bridge (Anderson and Darling (1952), MacNeill (1978); critical values are tabulated by Nyblom and Mäkeläinen (1983, table 1)).

The L^{μ} test is valid when μ_y is non-zero and unknown. For comparability to the previous discussion, it is useful to consider a test which imposes $\mu_y = 0$. Saikkonen and Luukkonen (1993a) show that, in the closely related problem of testing for a unit moving average root when the intercept is known to be zero, the LMPI test rejects for large values of

$$L = T\bar{y}^2/(T^{-1}\Sigma_{t=1}^{T} y_t^2), \tag{11}$$

which has a χ_1^2 null asymptotic distribution.

Although L and L^{μ} are derived assuming that x_t is unobserved and is I(1), these tests have power against (3) with $\alpha = 1 + c/T$. The asymptotic representation of L against the alternative $\gamma = b/T$ when (x_t, y_t) satisfy (3) with $\mu_x = \mu_y = 0$ is (Wright (1996))

$$L \Rightarrow [W_2(1) + b(\omega_{11}/\omega_{22})^{1/2}\textstyle\int J_c]^2. \tag{12}$$

Note that this limiting distribution depends on δ through the dependence of W_2 and J_c.

The extension of this approach to problem B is straightforward. Here the approach has the intuitive interpretation of testing directly the premise of cointegration that $y_t - \theta x_t$ is I(0) against it being I(1). A technical complication is that, under problem B, the null is that the error correction term is a general I(0) process, whereas in problem A the corresponding error is a martingale difference sequence under the null. However, this complication can be handled either by estimating the spectral density at frequency zero (cf. Park and Choi (1988), Tanaka (1990), and Kwiatkowski *et al.* (1992)) or by suitable prefiltering (cf. Saikkonen and Luukkonen (1993a, 1993b)). Here, we focus on the first of these approaches.

Specifically, consider (5) with $\mu_y = 0$ and let $u_{2t}(\theta) = y_t - \theta x_t$ and $\tilde{\omega}_{22} = \Sigma_{m=-l_T}^{l_T} k(m/l_T)\hat{\gamma}_{\hat{u}_2}(m)$, where $\hat{\gamma}_x(m) = (T-m)^{-1}\Sigma_{t=|m|+1}^{T}(x_t - \bar{x})(x_{t-m} - \bar{x})$, k is a kernel weighting function (see Andrews (1991) for a discussion of kernel choice), and $\hat{u}_{2t} = y_t - \hat{\theta}x_t$, where $\hat{\theta}$ is a T-consistent estimator of θ uniformly in c, for example the static OLS estimator or one of the many efficient estimators which impose $\alpha = 1$. Then the suitably modified L statistic is

$$L(\theta) = T\bar{u}(\theta)^2/\tilde{\omega}_{22}. \tag{13}$$

Under suitable conditions on u_{2t} and the truncation parameter l_T (cf. Kwiatkowski *et al.* (1992), Stock (1994b)), if x_t is local to unity then $\tilde{\omega}_{22} \xrightarrow{p} \omega_{22}$ uniformly in c. Thus, under the null hypothesis, $L(\theta_0) \Rightarrow \chi_1^2$, and, under the alternative, $\theta = \theta_0 + b/T$, $L(\theta)$ has the limiting representation (12). It is straightforward to extend these results to the case of an intercept in (5b) using L^{μ}.

Evidently, when $\mu_y = 0$, $L(\theta)$ can be used to test the null that $\theta = \theta_0$ is a cointegrating vector, and confidence regions for θ can be constructed as the acceptance region of this test. Note that this test has a somewhat different interpretation than the other tests discussed for problem B because the null is the joint null that $\theta = \theta_0$ *and* that the relation is cointegrating, whereas the previous tests maintain the hypothesis that a cointegrating relation exists for $\theta = \theta_0$. In theory, confidence regions for θ_0 constructed using the modified Nyblom–Mäkeläinen statistic can be infinite or empty, either because of a type I error or because no such cointegrating relation exists.[11]

4.4 Variable augmentation schemes

Another approach to problem A is to augment the equation of interest (3b) with additional lags of x_t and then test the significance of the coefficients on all but the final lag. Choi (1993) proposed this approach in the univariate AR(1) model (a test for a unit root), and Toda and Yamamoto (1995) and Dolado and Lutkepohl (1996) proposed it for a general VAR(p). The theoretical justification of this approach is the result summarized in section 2.1 that standard t- and F-tests can be used to test restrictions on coefficients which can be rewritten as coefficients on mean-zero stationary regressors, which can be done for scalar x_t if there are at most $p - d$ linear restrictions, where d is the maximum (integer) order of integration of the regresser. Thus, Wald tests of all but one coefficient have the usual χ^2 asymptotic distribution. Although the results in Toda and Yamamoto (1995) and Dolado and Lutkepohl (1996) are for the exact $\alpha = 1$ case, calculations similar to those in Elliott (1994) can be used to extend this asymptotic χ^2 result to the case that variables are either I(1) or have a local-to-unit root.

Although these variable augmentation tests have correct size for problem A, they have poor power in the direction of a "levels effect," specifically

$\gamma \neq 0$ in (3b). Dolado and Lutkepohl (1996) show that the test has power less than one against alternatives in a $T^{-\frac{1}{2}}$ neighborhood of the null. The procedures in the previous three sections all had non-trivial power against $1/T$ alternatives, so the asymptotic relative efficiency (ARE) of the variable-augmentation test is zero. Of course, if deviations from the null are not in the direction of a levels effect, variable augmentation tests could perform satisfactorily. Currently there appears to be no finite sample study that compares this approach to the previous three approaches.[12]

5 POTENTIAL SOLUTIONS: PROBLEM C

The problem of forecasting in systems with possibly large roots has received less attention than have problems A and B. The treatment here is correspondingly brief and draws on Stock (1995). The discussion focuses on the point forecasting problem, in which the objective is to produce point forecasts which have desirable properties for a range of values of c. Precisely what constitutes a desirable property of a point forecast arguably depends on the application and, potentially, the ultimate user of the forecast. If the forecast is to be the basis of a decision by a private party, then a desirable forecast would be tailored to that individual's loss function and would incorporate the individual's prior beliefs, in accordance with decision theory. Alternatively, if the forecasts are produced on an ongoing basis and are used in general discussions of business plans and public policy, as is typically the case with government forecasts, then it is arguably desirable that the forecasts be unbiased and perform well under some standard loss function, for example squared error loss. This discussion will focus on this second case and concentrate on unbiased forecasting. In the cointegrated VAR considered in section 3.4, the conditional distributions of x_{T+k} and y_{T+k} are $O_p(T^{\frac{1}{2}})$ but $y_{T+k} - \theta x_{T+k}$ is $O_p(1)$. Moreover, the only parameter entering the asymptotic distribution of $T^{-\frac{1}{2}}(x_{T+k} - x_{T+k|T})$ which is not consistently estimable is c. Therefore, little is lost by restricting attention to a univariate AR(1) with a single large root, and this is done for the rest of this section.

Consider the problem of unbiased long-run forecasting. If $\{x_t\}$ is symmetrically distributed around zero, then any odd function of the data is unconditionally unbiased (cf. Magnus and Pesaran (1991)). Because the forecaster knows x_T, however, it seems of little solace that a forecast of, say, $0.5x_T$ is unconditionally unbiased because the forecaster could equally well have observed $-x_T$ as x_T. We therefore adopt the view, enunciated in Phillips (1979), that unconditional unbiasedness is of limited practical interest in forecasting applications, and instead we focus on forecasts which are unbiased conditional on $x_T \geq 0$. In particular, we consider the construction of median unbiased forecasts conditional on $x_T \geq 0$.

In principle, many median unbiased forecasts exist. Let τ_T be a statistic which has an asymptotic distribution that depends only on the true value c. Most if not all asymptotically similar unit root test statistics have this property; cf. Stock (1994a). Let $m(c)$ denote the median of this distribution as a function of c, and let m^{-1} denote its inverse function. If m is monotone then by definition $\hat{c}_m = m^{-1}(\tau_T)$ will exceed c one-half of the time and will be less than c one-half of the time; that is, $\hat{c}_m$ is a median unbiased estimator of c. Suppose τ_T is an even function of the data. Because $e^{c\lambda}$ is non-negative and is monotone in c, $P[e^{\hat{c}_m\lambda}x_T > e^{c\lambda}x_T \mid x_T \geq 0] = P[\hat{c}_m > c] \to \frac{1}{2}$, so $e^{\hat{c}_m\lambda}x_T$ is an asymptotically median unbiased long-run forecast of $x_{T+k|T}$ conditional on $x_T \geq 0$ for $k/T = \lambda$.

In practice this requires constructing the median function, a task which can be computationally intensive. This has been done for the Dickey–Fuller (1979) t-statistic and the Sargan–Bhargava (1983)/Bhargava (1986) statistics by Stock (1991) using local-to-unity asymptotics; for the OLS root estimator in an AR(1) by Andrews (1993) using exact sampling results for the Gaussian AR(1); and for Elliott, Rothenberg, and Stock's (1996) family of efficient unit root tests by Stock (1995). Here, we consider the performance of the median unbiased forecast based on one of these test statistics, the Dickey–Fuller t-statistic. For comparison purposes we also consider a levels OLS forecast and a pretest forecast (a 5 percent Dickey–Fuller test is used to decide whether $\{x_t\}$ has a unit root; if it rejects, the levels OLS forecast is used, otherwise $\alpha = 1$ is imposed).

Numerical results are summarized in table 2.3 for a forecast horizon equal to 25 percent of the sample size. The first three columns of results report the fraction of forecasts which fall below the true conditional mean, both as a function of $J_c(1)$ and conditional only on $J_c(1)$ being positive; the final three columns report the root mean squared error (RMSE) around the true conditional mean. Computations were performed using pseudo-random realizations of J_c with 500 observations per draw. The results are presented both conditional on $J_c(1) = j$ and on $J_c(1) > 0$. Results for distributions conditional on $J_c(1)$ were computed by writing the relevant statistics in terms of $J_c(1)$ and (independent) functionals of V_c, where $V_c(s) = J_c(s) - b(s)J_c(1)$, where $b(s) = e^{c(1-s)}(e^{2cs} - 1)/(e^{2c} - 1)$ for c non-zero and $b(s) = s$ for $c = 0$, where V_c and $J_c(1)$ are independent. Results for fixed $J_c(1)$ are based on 5,000 replications; results for $J_c(1) > 0$ are based on 20,000 replications.

These results reveal several features of these forecasts. The OLS forecast is increasingly shifted toward zero as $J_c(1)$ increases. While it is most often biased for c near zero, even for $c = -10$ it is biased toward zero nearly three-quarters of the time. The bias of the pretest procedure is not

Table 2.3. *Performance of various long-term forecasts, univariate AR(1)*

		% forecasts $< x_{T+k\|T}$			RMSE of forecast minus conditional mean		
c	$J_c(1)$	OLS	PRE	MUF	OLS	PRE	MUF
0	0.2	1.000	0.538	0.709	0.160	0.055	0.123
	0.4	1.000	0.535	0.631	0.308	0.105	0.238
	0.6	0.997	0.530	0.542	0.432	0.143	0.344
	1.5	0.865	0.511	0.281	0.711	0.195	0.812
	>0	0.96	0.52	0.50	0.50	0.14	0.54
−2.5	0.2	0.977	0.087	0.613	0.077	0.094	0.088
	0.4	0.935	0.061	0.485	0.142	0.187	0.202
	0.6	0.844	0.036	0.338	0.186	0.280	0.359
	>0	0.90	0.07	0.50	0.14	0.21	0.30
−5	0.2	0.888	0.137	0.570	0.040	0.134	0.078
	0.4	0.783	0.074	0.379	0.072	0.276	0.219
	>0	0.83	0.12	0.50	0.07	0.22	0.22
−10	0.2	0.770	0.294	0.503	0.013	0.155	0.041
	0.4	0.570	0.099	0.222	0.033	0.349	0.168
	>0	0.74	0.29	0.50	0.03	0.19	0.11

Notes: The forecasting procedures are: OLS = levels OLS; PRE = pretest using 5 percent Dickey–Fuller (1979) demeaned t-statistic; MUF = forecast using median-unbiased estimator $\tilde{c}$. The procedures are described in the text. Entries are for the distribution conditional on $J_c(1)$ taking on the value in the second column; >0 indicates the distribution is conditional on $J_c(1) > 0$.

monotone in c. For c near zero, the unit root is rarely rejected and the pretest procedure is biased up. As c becomes more negative, the percent of unit root forecasts decreases. The MUF is biased away from zero as $J_c(1)$ increases, a consequence of the positive shift in the conditional distribution of the Dickey–Fuller t-statistic as $J_c(1)$ increases (the series appears more non-stationary). Conditional only on $J_c(1) > 0$, these forecasts are of course median unbiased (within simulation error), although they can have strong conditional bias.

No single estimator has uniformly lowest RMSE. The OLS forecast works well for $c \ll 0$. The MUF is preferable to OLS for $c = 0$ and small terminal values, although it is worse than OLS unconditional on J_c for all c considered. The pretest forecast works well for c nearly zero, but its RMSE increases substantially as c becomes more negative.

Additional research is needed. Although the MUF forecast is median unbiased on average, for certain terminal values it can be badly biased, and it fails to have the smallest RMSE of the three procedures for all c considered. Also, these results apply to point forecasts and do not address prediction intervals.

6 DISCUSSION AND CONCLUSIONS

This chapter has two main points. First, despite important advances over the past decade in our ability to model long-run economic relations, commonly used techniques fail to provide satisfactory solutions to some specific problems of long-run inference, in particular the construction of confidence intervals for and tests of coefficients on regressors which have large autoregressive roots, and to long-run point and interval forecasting. In these applications, the satisfactory performance of existing techniques, which are optimal for the unit root case, hinges critically on the untestable assumption of an exact unit root. This is not to imply that the current unit root and cointegration technology has no valid applications; indeed, consistent point estimation and first-order short-run forecasting are immune to the criticisms made here. However, departures from the unit root assumption which are so small as to be detectable with only low probability can nonetheless produce substantial distortions in inference and forecasts, at least when a regressor is endogenous.

The second point of this chapter is that some new techniques are emerging which are robust to whether the largest root is exactly one. This research is, however, incomplete in several ways. The techniques are disparate and in some cases *ad hoc*, and as yet there is not a unifying theory. Some of the approaches are closely tied to the linear autoregressive model with large roots, and it might be desirable to consider procedures valid under broader concepts of persistence. This suggests that these and related problems could prove fruitful areas for future research.

A natural question is what lessons for empirical practice arise from this ongoing research. Although any answer is necessarily speculative, this survey provides some clues. At a minimum, researchers using current unit root/cointegration methodology should be cognizant of the sensitivity of long-run forecasts, and of tests and confidence intervals for long-run coefficients, to the empirically untestable assumption of an exact unit root when the regressor in question is endogenous. Unless there are compeling a priori reasons for believing an exact unit root is present, certain empirical conclusions, such as confidence intervals for cointegrating parameters, can be delicate and in the end should be unconvincing to an appropriately skeptical audience. The broader promise of this research is less reliance on unit root and cointegration econometrics as currently conceived, and more credible statements of the uncertainty associated with long-run forecasts and with the parameters in long-run relations. In some cases, this uncertainty might actually be less than is implied by current techniques, such as long-run forecasting using a VECM when roots are in fact (just) stationary, although one suspects it typically will exceed that estimated by current procedures. More precise measures of uncertainty in turn should facilitate achieving the ultimate goal of more reliable inference about economic relations and policy choices.

Notes

This chapter was prepared for presentation at the Seventh World Congress of the Econometric Society, Tokyo, Japan, August 23, 1995. The author thanks Ronald Bewley, Juan Dolado, Graham Elliott, Max King, Thomas Rothenberg, Victor Solo, Herman van Dijk, Mark Watson, Kenneth Wallis, and workshop participants at the University of New South Wales, Sydney, Australia for helpful suggestions and/or comments on an earlier draft. Jeffrey Amato and Jonathan Wright provided excellent research assistance. This research was supported in part by National Science Foundation grant no. SBR-9409629.

1 A complete review of the literature is not attempted here; rather, the interested reader is referred to the survey and/or textbook treatments of including Banerjee *et al.* (1993), Hamilton (1994, chapters 18 and 19), Johansen (1995), and Watson (1994). The theory of inference about cointegrating relations is developed in (*inter alia*) Engle and Granger (1987), Phillips and Durlauf (1986), Stock (1987), Johansen (1988, 1991), Ahn and Reinsel (1990), Stock and Watson (1988, 1993), Engle and Yoo (1991), Phillips and Ouliaris (1990), Phillips (1991, 1995a), and Saikkonen (1991, 1992). The theory of testing for autoregressive unit roots in univariate time series is well developed and is reviewed in Banerjee *et al.* (1993) and Stock (1994a). Tests for cointegration are reviewed in Banerjee *et al.* (1993) and Watson (1994). For recent Monte Carlo evidence on tests for cointegration, see Gregory (1994), Haug (1993, 1996), and Ho and Sørenson (1996).

2 Deviations of the pretest from this ideal can result in finite sample distortions even if $\alpha = 1$; cf. Elliott and Stock (1994) and Toda and Yamamoto (1995).

3 This survey focuses solely on classical (frequentist) methods. From a Bayesian perspective, these problems are conceptually straightforward and simply require integration over α with respect to a suitable prior, although in this model the results can be highly sensitive to the choice of prior. Relevant readings can be found in the special issues of the *Journal of Applied Econometrics* (October–December 1991) and *Econometric Theory* (August–October 1994); also see Sims and Zha (1994).

4 Watson (1994) provides references for the case of higher orders of integration.

5 The remarks in this paragraph extend to the more general category of processes for which $T^{-\frac{1}{2}}x_{[Ts]}$ converges weakly to $Z(s)$, an $O_p(1)$ stochastic process which is not proportional to a Brownian motion. One example of a "local-to-$I(1)$" model is an autoregression with $\alpha = 1 + c/T$. Another example is the error components model, $\Delta x_t = \zeta_{1t} + hT^{-1}\zeta_{2t}$, where $(\zeta_{1t}, \Delta\zeta_{2t})$ are jointly stationary processes, var$(\Delta\zeta_{2t})$ is fixed, and $Ex_0^2 < \infty$; cf. Nyblom and Mäkeläinen (1983).

6 The discussion here follows Cavanagh, Elliott, and Stock (1995).

7 The results in this section are taken from Stock (1995). The literature on long-run forecasting with large roots is small. Phillips (1995b) considers forecast errors from VARs specified in levels and estimated by OLS when there are multiple unit or local-to-units roots, and generalizations of the levels of VAR results in this subsection can be found there. Remarks similar in spirit on the topic of VAR impulse response functions can be found in Sims and Zha (1994).

8 There can be, however, substantial finite problems for OLS forecasts from stable autoregressions; see for example Breidt, Davis, and Dunsmuir (1995), Kemp (1991, 1992), Maekawa (1987), Magnus and Pesaran (1991), Phillips (1979), and Stine (1987).

9 One could alternatively include unit root and cointegration pretests and use the I(1) or I(0) methodology depending on the outcome. If however $\alpha = 1 + c/T$ and the pretest is a consistent decision rule, then for reasons which parallel the discussion in section 2, the pretest will result in the I(1) methodology with probability tending to one.

10 One might conjecture that another alternative would be to compute the distributions in question using the bootstrap. However, in the AR(1) model with $\alpha = 1$ (a special case of problems A and C), Basawa *et al.* (1991) showed that the parametric bootstrap, based on the OLS estimate of α, is not consistent for the distribution of the OLS estimator of α. The bootstrap fails because it requires consistent initial estimation of c, which is not possible.

11 Tests related to these which test only for the existence of cointegration and do not involve a hypothesized value of θ have been developed by Shin (1994) and Harris and Inder (1994). The approach in section 4.3 is due to Wright (1996).

12 Extension of this approach to the case of deterministic terms is straightforward; cf. Toda and Yamamoto (1995) and Dolado and Lutkepohl (1995).

References

Ahn, S.K. and Reinsel, G.C. (1990). "Estimation for partially nonstationary multivariate autoregressive models." *Journal of the American Statistical Association*, 85: 813–23.

Anderson, T. W. and Darling, D. (1952). "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes." *Annals of Mathematical Statistics*, 23: 193–212.

Andrews, D.W.K. (1991). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation." *Econometrica*, 59: 817–58.

(1993). "Exactly median-unbiased estimation of first order autoregressive/unit root model." *Econometrica*, 61: 139–66.

Banerjee, A., Dolado, J., Galbraith, J. W., and Hendry, D. F. (1993). *Co-Integration*, *Error Correction, and the Econometric Analysis of Non-Stationary Data.* Oxford: Oxford University Press.

Basawa, I. V., Mallik, A. K., McCormick, W. P., Reeves, J. H., and Taylor, R. L. (1991). "Bootstrapping unstable first-order autoregressive processes." *Annals of Statistics*, 19: 1098–101.

Bhargava, A. (1986). "On the theory of testing for unit roots in observed time series." *Review of Economic Studies*, 53: 369–84.

Bobkoski, M. J. (1983). "Hypothesis Testing in Nonstationary Time Series." Unpublished Ph.D Thesis, Department of Statistics, University of Wisconsin.

Breidt, F. J., Davis, R. A., and Dunsmuir, W. T. M. (1995). "Improved bootstrap prediction intervals for autoregression." *Journal of Time Series Analysis*, 16: 177–200.

Campbell, B. and Dufour, J.-M. (1994). "Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter." Manuscript, CRDE, University of Montreal.

(1995). "Exact nonparametric orthogonality and random walk tests." *The Review of Economics and Statistics*, 78: 1–16.

Campbell, J. Y. and Shiller, R. J. (1987). "Cointegration tests of present value models." *Journal of Political Economy*, 95: 1062–88.

(1988). "Stock prices, earnings and expected dividends." *Journal of Finance*, 43: 661–76.

Cavanagh, C. L. (1985). "Roots local to unity." Manuscript, Department of Economics, Harvard University.

Cavanagh, C. L., Elliott, G. and Stock, J. H. (1995). "Inference in models with nearly integrated regressors." *Econometric Theory*, 11: 1131–47.

Chan, N. H. (1988). "On the parameter inference for nearly nonstationary time series." *Journal of the American Statistical Association*, 83: 857–62.

Chan, N. H. and Wei, C. Z. (1987). "Asymptotic inference for nearly nonstationary AR(1) processes." *Annals of Statistics*, 15: 1050–63.

(1988). "Limiting distributions of least squares estimates of unstable autoregressive processes." *Annals of Statistics*, 16: 367–401.

Choi, I. (1993). "Asymptotic normality of the least-squares estimates for higher order autoregressive integrated processes with some applications." *Econometric Theory*, 9: 263–82.

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

Dickey, D. A. and Fuller, W. A. (1979). "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American Statistical Association*, 74: 427–31.

Dolado, Juan J. and Lutkepohl, Helmut (1996). "Making Wald tests work for cointegrated VAR systems." *Econometric Reviews*, 15.

Dufour, J. M. (1990). "Exact tests and confidence sets in linear regressions with autocorrelated errors." *Econometrica*, 58: 475–94.

Elliott, G. (1994). "Application of Local to Unity Asymptotic Theory to Time Series Regression." Unpublished Ph.D. Dissertation, Department of Economics, Harvard University.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). "Efficient tests for an autoregressive unit root." *Econometrica*, 64: 813–36.

Elliott, G. and Stock, J. H. (1994). "Inference in time series regression when the order of integration of a regressor is unknown." *Econometric Theory*, 10: 672–700.

Engle, R. F. and Granger, C. W. J. (1987). "Cointegration and error correction: representation, estimation, and testing." *Econometrica*, 55: 251–76.

Engle, R. F. and Yoo, B. S. (1987). "Forecasting and testing in co-integrated systems." *Journal of Econometrics*, 35: 143–59.

(1991). "Cointegrated economic time series: an overview with new results." In Engle, R. F. and Granger, C. W. J. (eds.), *Long-Run Economic Relationships.* Oxford: Oxford University Press, pp. 237–66.

Fama, E. F. (1991). "Efficient capital markets II." *Journal of Finance*, 46: 1575–617.

Granger, C. W. J. (1981). "Some properties of time series data and their use in econometric model specification." *Journal of Econometrics*, 16: 121–30.

Granger, C. W. J. and Weiss, A. A. (1983). "Time series analysis of error correcting models." In Karlin, S., Amemiya, T., and Goodman, L. A. (eds.), *Studies in Econometrics, Time Series and Multivariate Statistics.* New York: Academic Press.

Gregory, A. W. (1994). "Testing for cointegration in linear quadratic models." *Journal of Business and Economic Statistics*, 12: 347–60.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Applications.* New York: Academic Press.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton: Princeton University Press.

Hargreaves, C. P. (1994). "A review of methods of estimating cointegrating relationships." In Hargreaves, C. P. (ed.), *Nonstationary Time Series Analysis and Cointegration.* Oxford: Oxford University Press, pp. 87–132.

Harris, D. and Inder, B. (1994). "A test of the null hypothesis of cointegration." In Hargreaves, C. P. (ed.), *Nonstationary Time Series Analysis and Cointegration.* Oxford: Oxford University Press, pp. 133–52.

Haug, A. A. (1993). "Residual based tests for cointegration." *Economics Letters*, 41: 345–51.

(1996). "Tests for cointegration: a Monte Carlo comparison." *Journal of Econometrics*, 71.

Herrndorf, N. A. (1984). "A functional central limit theorem for weakly dependent sequences of random variables." *Annals of Probability*, 12: 141–53.

Ho, M. S. and Sørenson, B. (1996). "Finding cointegration rank in high dimensional systems using the Johansen test." *The Review of Economics and Statistics.*

Johansen, S. (1988). "Statistical analysis of cointegrating vectors." *Journal of Economic Dynamics and Control*, 12: 231–54.

(1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models." *Econometrica*, 59: 1551–80.

(1995). *Likelihood Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Kemp, G. C. R. (1991). "The joint distribution of forecast errors in the AR(1) model." *Econometric Theory*, 7: 497–518.
(1992). "The distribution of forecast errors from an estimated random walk." Manuscript, University of Essex.
King, M. L. (1980). "Robust tests for spherical symmetry and their application to least squares regression." *Annals of Statistics*, 8: 1265–71.
King, M. L. and Hillier, H. (1985). "Locally best invariant tests of the error covariance matrix of the linear regression model." *Journal of the Royal Statistical Society, Series B*, 47: 98–102.
Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). "Testing the null hypothesis of stationarity against the alternatives of a unit root: how sure are we that economic time series have a unit root?" *Journal of Econometrics*, 54: 159–78.
MacNeill, I. B. (1978). "Properties of sequences of partial sums of polyomial regression residuals with applications to tests for change of regression at unknown times." *Annals of Statistics*, 6: 422–33.
Maekawa, K. (1987). "Finite sample properties of several predictors from an autoregressive model." *Econometric Theory*, 3: 359–70.
Magnus, J. R. and Pesaran, B. (1991). "The bias of forecasts from a first-order autoregression." *Econometric Theory*, 7: 222–35.
Mankiw, N. G. and Shapiro, M. D. (1985). "Trends, random walks and tests of the permanent income hypothesis." *Journal of Monetary Economics*, 16: 165–74.
Nabeya, S. and Tanaka, K. (1988). "Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative." *Annals of Statistics*, 16: 218–35.
Nyblom, J. (1986). "Testing for deterministic linear trend in time series." *Journal of the American Statistical Association*, 81: 545–9.
Nyblom, J. and Mäkeläinen, T. (1983). "Comparisons of tests for the presence of random walk coefficients in a simple linear model." *Journal of the American Statistical Association*, 78: 856–64.
Park, J. and Choi, C. (1988). "A new approach to testing for a unit root." Working Paper No. 88-23, Center for Analytical Economics, Cornell University.
Park, J. Y. and Phillips, P. C. B. (1988). "Statistical inference in regressions with integrated processes: Part I," *Econometric Theory*, 4: 468–97.
Phillips, P. C. B. (1979). "The sampling distribution of forecasts from a first-order autoregression." *Journal of Econometrics*, 9: 241–61.
(1987). "Toward a unified asymptotic theory for autoregression." *Biometrika*, 74: 535–47.
(1991). "Optimal inference in co-integrated systems." *Econometrica*, 59: 282–306.
(1995a). "Fully modified least squares and vector autoregression." *Econometrica*, 63: 1023–79.
(1995b). "Impulse response and forecast error variance asymptotics in nonstationary VARs." Manuscript, Cowles Foundation, Yale University.
Phillips, P. C. B. and Durlauf, S. N. (1986). "Multiple time series regression with integrated processes." *Review of Economic Studies*, 53: 473–96.

Phillips, P. C. B. and Loretan, M. (1991). "Estimating long-run economic equilibria." *Review of Economic Studies*, 58: 407–36.
Phillips, P. C. B. and Ouliaris, S. (1990). "Asymptotic properties of residual based tests for cointegration." *Econometrica*, 58: 165–94.
Phillips, P. C. B. and Solo, V. (1992). "Asymptotic for linear processes." *Annals of Statistics*, 20(2): 971–1001.
Rothenberg, T. J. and Stock, J. H. (1995). "Inference in a nearly integrated autoregressive model with nonnormal innovations." Manuscript, University of California, Berkeley.
Saikkonen, P. (1991). "Asymptotically efficient estimation of cointegrating regressions." *Econometric Theory*, 7: 1–21.
(1992). "Estimation and testing of cointegrated systems by an autoregressive approximation." *Econometric Theory*, 8: 1–27.
Saikkonen, P. and Luukkonen, R. (1993a). "Testing for moving average unit root in autoregressive integrated moving average models." *Journal of the American Statistical Association*, 88: 596–601.
(1993b). "Point optimal tests for the moving average unit root hypothesis." *Econometric Theory*, 9: 343–62.
Sampson, M. (1991). "The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models." *Journal of Applied Econometrics*, 6: 67–76.
Sargan, J. D. (1964). "Wages and prices in the United Kingdom: a study in econometric methodology." In Hart, P. E., Mills, G., and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*. London: Butterworth. Reprinted in Hendry, D. F. and Wallis, K. F. (eds.), *Econometrics and Quantitative Economics*. Oxford: Basil Blackwell, 1984.
Sargan, J. D. and Bhargava, A. (1983). "Testing for residuals from least squares regression for being generated by the Gaussian random walk." *Econometrica*, 51: 153–74.
Shin, Y. (1994). "A residual-based test of the null of cointegration against the alternative of no cointegration." *Econometric Theory*, 10: 91–116.
Shively, T. S. (1988). "An exact test for a stochastic coefficient in a time series regression model." *Journal of Time Series Analysis*, 9: 81–8.
Sims, C. A., Stock, J. H., and Watson, M. W. (1990). "Inference in linear time series models with some unit roots." *Econometrica*, 58: 113–44.
Sims, Christopher A. and Zha, Tao (1994). "Error Bands for Impulse Responses." Manuscript, Cowles Foundation, Yale University.
Stine, R. A. (1987). "Estimating properties of autoregressive forecasts." *Journal of the American Statistical Association*, 82: 1072–8.
Stock, J. H. (1987). "Asymptotic properties of least squares estimators of cointegrating vectors." *Econometrica*, 55: 1035–56.
(1991). "Confidence intervals for the largest autoregressive root in U.S. economic time series." *Journal of Monetary Economics*, 28: 435–60.
(1994a). "Unit roots, structural breaks and trends." In. Engle, R. F. and McFadden, D. (eds.), *Handbook of Econometrics, Vol. IV*. Amsterdam: Elsevier,

chapter 46.

(1994b). "Deciding between I(1) and I(0)." *Journal of Econometrics*, 63: 105–31.

(1995). "Long Run Forecasting." Manuscript, Kennedy School of Government, Harvard University.

Stock, J. H. and Watson, M. W. (1988). "Testing for common trends." *Journal of the American Statistical Association*, 83: 1097–107.

(1993). "A simple estimator of cointegrating vectors in higher-order integrated systems." *Econometrica*, 61: 783–820.

Tanaka, K. (1990). "Testing for a moving average unit root." *Econometric Theory*, 6: 433–44.

Toda, Hiro Y. and Yamamoto, Taku (1995). "Statistical inference in vector autoregressions with possibly integrated processes." *Journal of Econometrics*, 66: 225–50.

Watson, M. W. (1994). "Vector autoregressions and cointegration." In Engle, R. F. and McFadden, D. (eds.), *Handbook of Econometrics, Vol. IV*. Amsterdam: Elsevier, chapter 47.

Wright, J. H. (1996). "Confidence intervals in a cointegrative regression based on stationarity tests." Manuscript, Harvard University.