OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 58, 4 (1996)
0305-9049

VAR, ERROR CORRECTION AND PRETEST FORECASTS AT LONG HORIZONS

*James H. Stock**

I. INTRODUCTION

The production of economic forecasts is an important way that economists contribute to day-to-day policymaking. This paper focuses on a difficult but important aspect of economic forecasting, the construction of forecasts over long horizons. A typical long horizon forecast of economic activity might span four years using 20 to 40 years of data; a longer horizon might span one or two decades based on a comparable data set. In the extreme, some policy questions involve very long horizons indeed: forecasts which enter the debate about global warming entail predictions over horizons of 100 years.[1] A theme of this paper is that the presence of persistence in the form of large, possibly unit autoregressive roots, presents particular difficulties for long horizon forecasting. The technical problem is that both the true long horizon conditional mean and the standard formula for a valid prediction interval depend on the largest autoregressive roots of the system, and that these largest roots cannot be measured with sufficient precision to obtain asymptotically unbiased forecasts or prediction intervals with asymptotically correct coverage rates, uniformly over the possible values of the largest roots.

Recently, related observations have been made by Phillips (1995) and Stock (1995) using asymptotic arguments. Following those papers, we adopt a definition of a long horizon which depends on the length of the data set at hand. From a statistical perspective, a horizon of four years is arguably 'shorter' with 100 years of data than with 20, in the sense that the longer data set has more information about behaviour over four-year spans. This idea is formalized by taking the horizon h to be a fixed

*The author thanks Mark Watson and the participants at the Conference on the Econometrics of Economic Policy, European Economic Institute, September 1995, for helpful suggestions This research was supported in part by National Science Foundation grant no SBR-9409629

[1]For example, Nordhaus and Yang (1995) consider global per capita income scenarios for the years 2100 and 2200, and provide model solutions (consumption, income, optimal carbon taxes, carbon emissions) through 2090, a horizon of 100 years (also see Manne and Richels, 1992).

fraction of the sample size: a long horizon is $h=[T\lambda]$, where $[\bullet]$ denotes the greatest lesser integer, where $\lambda>0$.

This paper makes two new contributions relative to the recent literature. First, one focus here is on interval forecasts. It is shown that interval forecasting at long horizons faces an additional difficulty beyond those confronted by long horizon point forecasting. Second, we examine pretest forecasts, in which the decision about modelling the number and location of unit roots is based on pretesting for unit roots and/or cointegration. It is argued that, although pretest (or model selection) forecasts can work well when there truly is a unit root, in theory pretest forecasts can work poorly for deviations from an exact unit root which are sufficiently small that they often will not be detected even by efficient pretests.

Specifically, we consider three strategies for long horizon forecasting in multivariate systems. In the first, forecasts are based on a vector autoregression (VAR) in levels, which is estimated by ordinary least squares (OLS). In the second, forecasts are made using a vector error correction model (VECM), in which the researcher specifies the number of unit roots and cointegrating vectors, and the space in which the cointegrating vectors fall, based on *a priori* knowledge. The VECM is then estimated using efficient cointegration estimation techniques. In the third strategy, unit root and cointegration pretests are used to ascertain the number and location of unit roots in the system. Based on the results of this analysis, either a VAR or a VECM is specified, and estimation proceeds as above for each model. Forecasts and forecast confidence intervals are then computed conditional on the number and location of the unit roots as determined by the pretest. This pretest procedure more closely resembles current practice than either the pure VAR or pure VECM strategies.

These three strategies are analyzed using asymptotic techniques. Because efficient unit root tests have nondegenerate asymptotic power functions against alternatives in a T^{-1} neighbourhood of one, these asymptotics adopt the local to unity model in which the largest autoregressive roots are $1+O(T^{-1})$ (cf. Bobkoski, 1983; Cavanagh, 1985; Chan and Wei, 1987; Phillips, 1987). For these roots, an applied researcher might consider the conventional unit roots/cointegration model to be a satisfactory approximation to the true data generating process. In this nesting, the estimated or imposed roots in these models are within $O_p(T^{-1})$ of the true roots, and the imprecision resulting from their estimation is asymptotically negligible (to first-order) for short horizon point or interval forecasts. Over long horizons, however, the error from estimation of the largest roots is of the same order as the effect of future disturbances, and this can result in substantially biased point forecasts and prediction intervals with actual coverage rates far different than the nominal coverage rate.

When there is a single large unit root in the system, the multivariate forecasting problem effectively reduces to the problem of long-horizon forecating in a univariate AR(1) with a root local to unity. This case is examined numerically for the three estimation strategies. In this simple model, the pretest forecast based on a nearly efficient unit root pretest proposed by Elliott, Rothenberg and Stock (1996) is found to produce prediction intervals with fairly stable coverage rates as a function of the forecast horizon and the largest root, if the largest root is no greater than unity; forecasts based on a Dickey–Fuller (1979) pretest do less well by this criterion. It must be emphasized, however, that these numerical results do not pertain to the multivariate, multiple large root case more relevant to practical forecasting. Thus these results do not provide an endorsement of pretest forecasts or prediction intervals for long horizons based on a multivariate system which has more than one large root. Moreover, even in the univariate setting, pretest forecasts are discontinuous functions of the data, a property which is arguably undesirable in practice.

The rest of the paper is organized as follows. The model, assumptions and theoretical results for VARs and VECMs for n variables with k large, possibly unit roots are presented in Section II. Asymptotic results for the VAR, VECM and pretest forecasts are given in Section III, and theoretical results for prediction intervals are given in Section IV. Numerical results are presented in Section V, and Section VI concludes.

II. THE MODEL AND PRELIMINARY ASYMPTOTIC RESULTS

Suppose that the n-dimensional time series X_t has k large autoregressive roots. The sample consists of $(X_1, \ldots, X_T)$. It is convenient to order X_t so that the first k elements of X_t, X_{1t}, are $I(1)$ or nearly $I(1)$ and not cointegrated, while the remaining elements of X_t, X_{2t} are cointegrated with X_{1t}. The model, written in triangular form, is,

$$X_{1t} = \beta_1 + V_t, \quad (I - AL)V_t = U_{1t} \tag{1a}$$

$$X_{2t} = \beta_2 + \theta X_{1t} + U_{2t}. \tag{1b}$$

Note that in this specification no element of X_t has a drift (a deterministic time trend component).

The disturbances $U_t = (U'_{1t}, U'_{2t})'$ are assumed to follow a linear process $U_t = B(L)\varepsilon_t$, where ε_t is an n-dimensional martingale difference sequence with $2+\delta$ moments, $\delta > 0$ and with $E\varepsilon_t\varepsilon'_t = \Sigma$. It is also assumed that $B(L)$ is invertible and one-summable, that is, $\Sigma_{j=0}^{\infty} j|B_j| < \infty$, where $|M| = \max_{i,j}|M_{i,j}|$ for any matrix M. It is assumed that V_1 is $O_p(1)$.

In this representation, the k large roots of the system are contained in A, and the disturbances in (1) are assumed to be integrated of order zero ($I(0)$). Specifically, for the asymptotic analysis, we let $A = I + C/T \cong \exp(C/T)$ (where $\exp(\bullet)$ is the matrix exponential), where the roots of C are finite (cf. Phillips, 1987). If $A = I$, this reduces to the standard cointegration model in triangular form. This model generalizes the concept of cointegration to the case that the elements of X_t are individually highly persistent, with roots nearly but not necessarily exactly one, but some linear combinations are $I(0)$. Elliott (1994) studied the properties of efficient estimators of cointegrating parameters, in which $A=I$ is imposed, in this model for general C (with $k=1$).

The system (1) can be written as a restricted VAR,

$$X_t=\mu+\Gamma[A\ 0]X_{t-1}+D(L)\tilde{X}_{t-1}+e_t \tag{2}$$

where $\Gamma=[I_k, \theta']'$ (where I_k is the $k\times k$ identity matrix), $\mu=RB(1)^{-1}[\beta_1'(I-A)', \beta_2']'$, $D(L)=RL^{-1}(I-B(L)^{-1})$, $\tilde{X}_t=[((I-AL)X_{1t})'(X_{2t}-\theta X_{1t})']'$ and $e_t=R\varepsilon_t$, where R is the $n\times n$ matrix with blocks $R_{11}=I_k$, $R_{12}=0$, $R_{21}=\theta$, $R_{22}=I_{n-k}$. (The derivation of (2) is sketched in the appendix.) Note that in this representation $\tilde{X}_t$ is $I(0)$ so that the only nearly nonstationary regressor in (2) is X_{t-1}. When referring to the VAR form, it is assumed henceforth that $D(L)$ has finite order $p-1$ so that in levels X_t follows a VAR(p).

The next theorem provides an asymptotic approximation to the distribution of X_{T+h} conditional on X_T. Because A is nearly one, X_T and X_{T+h} are $O_p(T^{1/2})$ and the results are presented after scaling by $T^{-1/2}$. Let $\bar{\lambda}<\infty$ denote the maximum horizon λ as a fraction of the sample size under consideration. Throughout, we adopt the notation $X_{t|s}=E(X_t\,|\,\{X_r\}, r\leq s)$.

Theorem 1

Suppose that $|C|<\infty$ and $B(L)$ is one-summable. Then:

(a) $T^{-1/2}X_{T+[T\lambda]}-\Gamma A^{[T\lambda]}T^{-1/2}X_{1T}\Rightarrow\Gamma\phi(\lambda)$ and $T^{-1/2}X_{T+[T\lambda]|T}-\Gamma A^{[T\lambda]}T^{-1/2}X_{1T}\xrightarrow{p}0$ uniformly in λ, $0\leq\lambda\leq\bar{\lambda}$, where $\phi(\lambda)=J_c(1+\lambda)-\exp(C\lambda)J_c(1)$, where $dJ_c=CJ_c+[I, 0]B(1)\,dW$, where W is a n-dimensional Brownian motion with covariance matrix Σ.

(b) For λ fixed, the distribution of $T^{-1/2}X_{T+[T\lambda]}$ conditional on $T^{-1/2}X_{1T}=x$ is asymptotically $N(\Gamma\exp(C\lambda)x, V(\lambda))$, where $V(\lambda)=\Gamma\{\int_{s=0}^{\lambda}\exp(Cs)\Omega_{11}\exp(Cs)'\,ds\}\Gamma'$, where $\Omega=B(1)\Sigma B(1)'$.

The proof is sketched in the Appendix.

Because X_{2t} is cointegrated with X_{1t}, the long-horizon conditional expectations of X_{1t} and X_{2t} are asymptotically equivalent, up to the matrix of cointegrating coefficients θ. Thus the joint conditional distribution of $T^{-1/2}X_{T+h}$ is degenerate, with the limiting covariance matrix $V(\lambda)$ having rank k.

III VAR, VECM, AND PRETEST FORECASTS

1. *VAR and Cointegration Forecasts*

Two conventional estimators of VAR coefficients in systems such as (2) are by OLS in levels and using an asymptotically efficient cointegration estimator such as Johansen's (1988) method. With Gaussian disturbances, the two estimators have maximum likelihood interpretations, in the first case with no restrictions on the magnitude or location of the roots, in the second case imposing k unit roots and the existence of $n-k$ cointegrating vectors.

Although both methods yield consistent parameter estimators in the local-to-unity model (Elliott, 1994), their long-term forecasts differ by $O_p(T^{1/2})$. First consider the unrestricted levels VAR estimator. The device of rearranging the VAR regressors used in Sims, Stock and Watson (1990) for the case $C=0$, extended to the local to unity case as in Stock (1991) and Elliott (1994), can be used to obtain a limiting representation for $T(\hat{A}^{\text{VAR}}-I)\equiv\hat{C}^{\text{VAR}}$, where $\hat{A}^{\text{VAR}}$ is the VAR estimator of A. The expression for this representation is inconsequential for the development here so we simply denote this limit as $\hat{C}^{\text{VAR}}\Rightarrow\tilde{C}^{\text{VAR}}$. Let $\hat{X}^{\text{VAR}}_{T+h|T}$ denote the VAR forecast of X_{T+h} and let $e^{\text{VAR}}_{T+h}(h)$ denote the VAR forecast error, so $e^{\text{VAR}}_{T+h}(h)=X_{T+h}-\hat{X}^{\text{VAR}}_{T+h|T}$. Then $T^{-1/2}X_{1T}\Rightarrow J_c(1)$ and,

$$T^{-1/2}\hat{X}^{\text{VAR}}_{T+h|T}\Rightarrow\Gamma\exp(\tilde{C}^{\text{VAR}}\lambda)J_c(1) \tag{3}$$

$$T^{-1/2}e^{\text{VAR}}_{T+h}(h)\Rightarrow\Gamma\{\exp(C\lambda)-\exp(\tilde{C}^{\text{VAR}}\lambda)\}J_c(1)+\Gamma\phi(\lambda) \tag{4}$$

for $h=[T\lambda]$, where $\phi(\lambda)$ and $\tilde{C}^{\text{VAR}}$ are independent.

Next consider forecasts made using an efficient cointegration estimator in which A is set to the identity matrix. For example, this would be obtained by applying Johansen's (1988) algorithm (with a constant term) after imposing n–k cointegrating vectors. An asymptotically equivalent estimator is obtained in the triangular form (1a,b) imposing $A=I$; cf. Stock and Watson (1993). These estimators of θ are T-consistent even if $A=I+C/T$ (see Elliott (1994) for the $k=1$ case). Accordingly, for $h=[T\lambda]$ the long horizon forecast and forecast error based on an efficient cointegrating estimator have the limits:

$$T^{-1/2}\hat{X}^{CI}_{T+h|T}\Rightarrow\Gamma J_c(1) \tag{5}$$

$$T^{-1/2}e^{CI}_{T+h}(h)\Rightarrow\Gamma\{\exp(C\lambda)-I\}J_c(1)+\Gamma\phi(\lambda). \tag{6}$$

An important implication of the expressions (3)–(6) is that the treatment of A results in different asymptotic behaviour of the long-horizon forecasts. To see this, consider the case $C=0$, so the first term in (6) drops out and $T^{-1/2}e^{CI}_{T+h}(h)\Rightarrow\Gamma\phi(\lambda)$. Similarly, $T^{-1/2}e^{\text{VAR}}_{T+h}(h)\Rightarrow\Gamma\{I-\exp(\tilde{C}^{\text{VAR}}\lambda)\}J_c(1)+\Gamma\phi(\lambda)$. Because $\phi(\lambda)$ is independent of $(\tilde{C}^{\text{VAR}},J_c(1))$, the levels VAR forecast error has a larger variance than the cointegration forecast error. Because the cointegration forecast is unbiased when $C=0$, in this case the VAR forecast has a larger root mean square error (RMSE) as well. Evidently, the first order asymptotic performance of the long-horizon forecast depends on the estimation method, a feature that distinguishes the long-horizon/nearly $I(1)$ forecasting problem from the usual short-horizon problem.

Because the VAR forecast is an odd function of the data, if the disturbances are distributed symmetrically around zero then the VAR forecast is unconditionally unbiased (cf. Magnus and Pesaran (1991)). However, the roots of the VAR estimator of A will be biased towards zero so the forecasts are pulled towards zero. Thus a more useful concept for forecasting is bias conditional on $X_{1T} \geq 0$ or, more strongly, bias conditional on X_{1T} (cf. Phillips, 1979). Throughout, when we refer to a forecast being biased, it is meant that it is biased conditional on $X_{1T} \geq 0$. In this sense, when $C=0$ the VAR forecast is biased towards zero, although the cointegration forecast is not.

If $C \neq 0$, then both the levels VAR and cointegration forecasts are biased. Numerical results on their root mean squared errors are presented in the next section. Because C is rarely known to be precisely zero on *a priori* grounds, these results suggest investigating alternative forecasting procedures with reduced bias.

2. *Pretest Forecasts when k = 1*

In practice a cointegration estimator is used after pretesting for unit roots and/or cointegration. When $k>1$, the sequence of unit root and cointegration pretests is rather involved and many options are available. To sharpen the results, we therefore focus on the case $k=1$. Specifically, suppose that the hypothesis of a unit root in X_{1t} is tested, and let $\delta_T = 1$ if the pretest rejects the unit root null and $=0$ otherwise. In this strategy, if $\delta_T = 0$ then $C=0$ is imposed, but if $\delta_T = 1$ then the VAR estimator of C is used. This yields the pretest estimator,

$$\hat{C}^{PRE} = \delta_T \hat{C}^{\text{VAR}}. \tag{7}$$

Under the local-to-unity model, if the size of the test is fixed, then $(\delta_T, \hat{C}^{\text{VAR}}) \Rightarrow (\tilde{\delta}, \tilde{C}^{\text{VAR}})$ (say), where $\tilde{\delta}$ is a limiting Bernoulli random variable which equals one with unconditional probability equal to the power of the pretest. The random variables $\tilde{\delta}$ and $\tilde{C}^{\text{VAR}}$ are in general dependent. The limiting representation of $\tilde{\delta}$ depends on precisely which unit root pretest is used (for examples see Stock, 1994).

The forecast and forecast error based on the pretest estimator have the limits,

$$T^{-1/2}\hat{X}^{\text{PRE}}_{T+h|T} \Rightarrow \Gamma \exp(\tilde{\delta}\tilde{C}^{\text{VAR}}\lambda) J_c(1) \tag{8}$$

$$T^{-1/2}e^{\text{PRE}}_{T+h}(h) \Rightarrow \Gamma\{\exp(C\lambda) - \exp(\tilde{\delta}\tilde{C}^{\text{VAR}}\lambda)\} J_c(1) + \Gamma\phi(\lambda) \tag{9}$$

for $h = [T\lambda]$, where $\phi(\lambda)$ is independent of $(\tilde{\delta}, \tilde{C}^{\text{VAR}})$.

IV. ASYMPTOTIC COVERAGE RATES OF PREDICTION INTERVALS

The conditional distribution of X_{T+h} given X_T, presented in Theorem 1, is approximately normal for h large but depends on C both in its mean and its variance. Were C consistently estimable, then prediction intervals could be formed using consistent estimates of the mean and variance, and these would be valid to first-order asymptotically. However, C is not consistently estimable. Moreover there appears to be no pivot for X_{T+h} given X_T which is valid asymptotically. Thus conventional methods for the construction of prediction intervals are not applicable in the long-horizon forecasting problem when A is local to unity.

Expressions for true coverage rates of standard prediction intervals can be obtained from the previous results. Consider the case of forecasting a linear combination of X_{T+h}, say $b'X_{T+h}$ for $h=[T\lambda]$. For λ fixed, the conventional 68 percent prediction interval for $b'X_{T+h}$ is $b'\hat{X}_{T+h|T} \pm T^{1/2}[b'\hat{V}(h/T)b]^{1/2}+o_p(1)$, where $\hat{V}(\lambda)=\hat{\Gamma}\{\int_{s=1}^{1+\lambda}\exp(\hat{C}(\lambda-s))\hat{\Omega}_{11}\exp(\hat{C}(\lambda-s))'\,ds\}\hat{\Gamma}'$, where $(\hat{C}, \hat{\Gamma}, \hat{\Omega}_{11})$ are estimators of (C, Γ, Ω_{11}). As discussed in Section II, MLEs of $\hat{\Gamma}$ and $\hat{\Omega}_{11}$ are consistent (for finite C) if $A=I_k$ is imposed. A consequence of Theorem 1 is that the coverage rate of a '$\pm$one standard deviation' prediction interval based on some estimator $\hat{C}$, when $\hat{C}\Rightarrow\tilde{C}$ (say), is,

$$P[b'X_{T+h}\in(b'\hat{X}_{T+h|T}-T^{1/2}[b'\hat{V}(h/T)b]^{1/2}, b'\hat{X}_{T+h|T}+T^{1/2}[b'\hat{V}(h/T)b]^{1/2})]$$
$$\Rightarrow E\Phi[v(\lambda)+\tau(\lambda)]-E\Phi[v(\lambda)-\tau(\lambda)] \tag{10}$$

where $\Phi(\bullet)$ is the cumulative normal cdf, $v(\lambda)=b'\Gamma[\exp(\tilde{C}\lambda)-\exp(C\lambda)]J_c(1)/[b'V(\lambda)b]^{1/2}$, and $\tau(\lambda)^2=[b'\Gamma\{\int_{s=0}^{\lambda}\exp(\tilde{C}s)\Omega_{11}\exp(\tilde{C}s)'\,ds\}\Gamma'b]/b'V(\lambda)b$, where the expectation is taken over the joint distribution of $(J_c(1), \tilde{C})$.

Evidently, the true asymptotic coverage rate in (10) will not in general be the nominal rate (here, 68%). The distortions arise from misestimation of both the conditional mean and the variance.

An implication of these results is that prediction intervals for X_{2t} will in general not have the desired coverage rate, even if X_{1t} is strictly exogenous. This contrasts with the problem of constructing confidence intervals for the cointegrating coefficients θ based on the cointegrating MLE, which have the desired coverage rates even if A is local-to-unity as long as X_{1T} is exogenous.

V. NUMERICAL RESULTS

This section presents numerical results on the performance of the alternative point and interval forecasts discussed in the preceding sections. All results are for the univariate AR(1) version of (1a) with $\beta_1=0$, so that X_t obeys,

$$X_t=\alpha X_{t-1}+\varepsilon_t \tag{11}$$

where ε_t is $N(0, 1)$ and $\alpha=1+C/T$. Because the short-run dynamics do not affect long-horizon forecasts and because the forecasts of X_{1T+h} and X_{2T+h}

are equivalent to order $O_p(1)$, the asymptotic results suggest that qualitatively similar findings would obtain for forecasts in a VAR of general order with a single large root.

Six alternative long-horizon forecasts were considered. All apply to the case that a constant (but no time trend) is included when the largest root is estimated; however, consistent with (11) and with the treatment in Sections II–IV, this constant is set to zero when forecasts are made. Thus for example the OLS forecast is $\hat{\alpha}^{\mu[T\lambda]}X_T$, where $\hat{\alpha}^{\mu}$ is OLS estimate of α obtained by regressing X_t on $(1, X_{t-1})$. Similarly, the unit root pretests are computed in the case that a constant enters the regression (the 'demeaned' case); if the statistic rejects, the OLS forecast $\hat{\alpha}^{\mu[T\lambda]}X_T$ is used, while if it fails to reject, the random walk forecast X_T is used.

The six forecasts examined are: (1) the OLS levels forecast; (2) the random walk forecast, which imposes the unit root in X_t; (3) the forecast based on the pretest estimator with a 5 percent Dickey–Fuller (1979) t-test for a unit root; (4) the pretest forecast based on a 10 percent Dickey–Fuller t-test; (5) the pretest forecast based on the 5 percent DF–GLS t-test proposed by Elliott, Rothenberg and Stock (1996); and (6) the pretest forecast based on a 10 percent DF–GLS t-test. The motivation for considering the DF–GLS test is that it has higher asymptotic power than the standard DF test in the 'demeaned' case and effectively achieves the asymptotic power envelope; it is worth considering the possibility that a more efficient pretest might improve the performance of the pretest forecast. Note that the random walk forecast is the univariate counterpart of the cointegration forecast discussed in Section II. As a basis of comparison, the infeasible forecast based on the true value of α, $\alpha^h X_T$, was also computed.

Simulations were performed for various values of α and various horizons using $T=100$ observations, with 20,000 Monte Carlo replications for each value of (α, λ). For each draw, prediction intervals were computed as a conventional '$\pm$one standard deviation' interval as discussed in Section IV; that is, once α was estimated, prediction intervals were computed using the estimated value of α.

Table 1 presents the RMSE of each of the four forecasts for various horizons and various values of α, relative to the RMSE of the infeasible forecast which would be made were α known. When $\alpha=1$, the random walk estimator imposes the correct value of α, so in this case the ratio of RMSE's is exactly one. In all other cases, the ratio exceeds one, although not necessarily to the three digit accuracy of the table. Most of the cases in Table 1 pertain to large roots and fairly long horizons. However, to facilitate comparison to the more familiar stationary/short horizon case, results are also reported for $C=-20$ and $\lambda=0.02$, which correspond to $\alpha=0.8$ and $h=2$ for $T=100$.

The results in Table 1 accord with the predictions of the theory. For the smallest values of α, the OLS forecast has lower RMSE than the unit

TABLE 1

Relative Root Mean Square Error of $\hat{X}_{T+h|T}-X_{T+h}$ for Various Forecasts, $T=100$

Entries are ratio of forecast error RMSE based on estimated α to forecast error RMSE based on true α

	Forecast					
			DF Pretest		DF-GLS Pretest	
α	OLS	Random Walk	5%	10%	5%	10%
			A Horizon $h=2$ ($\lambda=0.02$)			
0.80	1.005	1.102	1.027	1.015	1.007	1.005
0.90	1.010	1.054	1.049	1.043	1.029	1.019
0.93	1.015	1.033	1.038	1.039	1.036	1.027
0.95	1.016	1.026	1.029	1.033	1.033	1.028
0.97	1.034	1.017	1.030	1.035	1.028	1.032
0.98	1.047	1.010	1.025	1.033	1.020	1.024
0.99	1.073	1.004	1.022	1.032	1.013	1.019
1.00	1.117	1.000	1.028	1.042	1.007	1.012
1.01	1.193	1.033	1.058	1.072	1.038	1.043
1.02	1.219	1.419	1.426	1.431	1.422	1.424
			B Horizon $h=10$ ($\lambda=0.10$)			
0.80	1.003	1.348	1.091	1.043	1.011	1.004
0.90	1.014	1.226	1.190	1.160	1.104	1.058
0.93	1.021	1.158	1.149	1.136	1.115	1.083
0.95	1.037	1.121	1.121	1.123	1.116	1.099
0.97	1.066	1.072	1.085	1.093	1.085	1.089
0.98	1.104	1.048	1.067	1.080	1.062	1.072
0.99	1.181	1.019	1.051	1.073	1.033	1.048
1.00	1.331	1.000	1.056	1.091	1.013	1.025
1.01	1.579	1.151	1.197	1.228	1.160	1.169
1.02	1.777	2.469	2.476	2.482	2.471	2.474
			C Horizon $h=20$ ($\lambda=0.20$)			
0.80	1.000	1.406	1.102	1.049	1.014	1.002
0.90	1.008	1.340	1.282	1.241	1.159	1.079
0.93	1.018	1.276	1.252	1.232	1.192	1.136
0.95	1.033	1.212	1.205	1.197	1.183	1.149
0.97	1.057	1.139	1.143	1.144	1.142	1.136
0.98	1.101	1.090	1.100	1.110	1.098	1.103
0.99	1.186	1.040	1.064	1.081	1.049	1.060
1.00	1.405	1.000	1.051	1.088	1.011	1.021
1.01	1.796	1.285	1.323	1.353	1.291	1.298
1.02	2.244	3.343	3.347	3.352	3.345	3.347
			D Horizon $h=50$ ($\lambda=0.50$)			
0.80	1.000	1.422	1.108	1.049	1.012	1.002
0.90	1.001	1.405	1.339	1.289	1.186	1.091
0.93	1.007	1.385	1.351	1.318	1.264	1.177
0.95	1.031	1.347	1.331	1.315	1.294	1.237
0.97	1.043	1.271	1.262	1.253	1.254	1.232
0.98	1.086	1.211	1.206	1.202	1.206	1.199
0.99	1.173	1.101	1.107	1.113	1.102	1.105
1.00	1.488	1.000	1.028	1.051	1.004	1.010
1.01	2.131	1.610	1.628	1.645	1.612	1.616
1.02	3.269	4.979	4.980	4.982	4.979	4.980

Notes Computed by Monte Carlo simulation with $T=100$ observations for the Gaussian AR(1) model (20,000 replications) The forecasting methods are discussed in the text root forecast, and the OLS forecast is almost as efficient as if α were known. Also, for the shortest horizon ($\lambda=0.02$), for values of $\alpha \leq 1$, the RMSEs of all six forecasts are similar. However, as λ increases the features predicted by the theory emerge. For example, for α nearly one, the random walk forecast has lower RMSE because of the bias in the OLS estimator. The values of $C=T(\alpha-1)$ for which the random walk forecast has lower RMSE are approximately $-2 \leq C \leq 1$, depending on the horizon.

The RMSEs of the pretest forecasts reflect their tendency to use the random walk forecast when C is nearly zero and the OLS forecast when $C \ll 0$. The performance of the DF pretest forecasts is worst when $C \cong -10$, in which case the OLS forecast has substantially lower RMSE than the random walk forecast but the unit root null is rejected with relatively low probability. The 10 percent DF pretest forecast outperforms the 5 percent pretest forecast for all C except $-3 \leq C \leq 1$ (approximately), and the performance of the two forecasts is fairly close even in this range.

For almost all values of C, the DF–GLS pretest forecasts have lower RMSEs than the DF pretest forecasts. Our interpretation of this finding is that the improved performance of the DF–GLS pretest forecast stems from the better power of the DF–GLS unit root test. The OLS forecast has a lower RMSE than the unit root forecast for $C \leq -3$; the greater is the power of the pretest, the more likely is the pretest procedure to use this better-performing OLS forecast. On the other hand, because size is controlled, both the DF–GLS and DF pretest forecasts tend to choose the random walk forecast for C nearly zero, so the two procedures perform similarly for $-2 < C \leq 0$.

The differences among the forecasts becomes stronger as the horizon increases. It is noteworthy, however, that even at the relatively short horizon of 10 percent of the sample the RMSEs can differ by as much as 20 percent. At the horizon of 50 percent of the sample, the 10 percent DF–GLS pretest forecast is the only one to have a relative RMSE uniformly under 1.25 for $C \leq 0$, and its RMSE is less than those of the DF pretest forecasts for all C considered.

Coverage rates for prediction intervals are reported in Table 2. As in Table 1, the results accord with the theory. For small values of α, the OLS forecast interval coverage rates approach the desired 68 percent, but for α nearly one or greater than one, the OLS intervals have very low coverage rates indeed, particularly at long horizons. Similarly, the random walk intervals are far too wide when α is in fact much less than one, but perform relatively well when α is nearly one. Perhaps the most striking result in Table 2 is that the DF–GLS prediction intervals have coverage rates which are approximately constant and are only slightly below the nominal coverage rate for a wide range of $C \leq 0$. In contrast, the DF pretest coverage rates, while better than the OLS or random walk coverage rates, still depend fairly strongly on C. In particular, the DF

TABLE 2

Asymptotic Coverage Rates of Nominal 68% Prediction Intervals, $T=100$

Entries are the probability that the h-step ahead prediction interval computed as $\hat{X}_{T+h|T}\pm$one standard deviation, using the estimated value of α, contains X_{T+h}

	Forecast					
			DF Pretest		DF–GLS Pretest	
α	OLS	Random Walk	5%	10%	5%	10%
		A Horizon $h=2$ ($\lambda=0.02$)				
0.80	0.620	0.703	0.620	0.619	0.620	0.620
0.90	0.635	0.692	0.656	0.644	0.636	0.633
0.93	0.639	0.696	0.670	0.658	0.648	0.639
0.95	0.638	0.681	0.667	0.658	0.649	0.641
0.97	0.636	0.687	0.674	0.666	0.664	0.655
0.98	0.628	0.684	0.670	0.660	0.666	0.656
0.99	0.618	0.681	0.669	0.661	0.668	0.661
1.00	0.598	0.676	0.661	0.652	0.668	0.661
1.01	0.572	0.661	0.650	0.643	0.655	0.651
1.02	0.572	0.512	0.508	0.504	0.509	0.507
		B Horizon $h=10$ ($\lambda=0.10$)				
0.80	0.629	0.865	0.633	0.627	0.629	0.628
0.90	0.613	0.781	0.686	0.651	0.619	0.609
0.93	0.598	0.756	0.698	0.665	0.625	0.599
0.95	0.580	0.732	0.693	0.664	0.631	0.595
0.97	0.556	0.712	0.681	0.659	0.646	0.609
0.98	0.544	0.705	0.676	0.655	0.657	0.627
0.99	0.502	0.693	0.667	0.647	0.657	0.630
1.00	0.455	0.680	0.654	0.634	0.658	0.640
1.01	0.386	0.591	0.577	0.565	0.578	0.565
1.02	0.350	0.286	0.280	0.275	0.278	0.273
		C Horizon $h=20$ ($\lambda=0.20$)				
0.80	0.626	0.952	0.645	0.629	0.627	0.626
0.90	0.600	0.865	0.733	0.676	0.616	0.598
0.93	0.585	0.822	0.741	0.692	0.633	0.590
0.95	0.556	0.786	0.733	0.694	0.642	0.587
0.97	0.528	0.746	0.708	0.680	0.657	0.605
0.98	0.491	0.731	0.696	0.668	0.662	0.615
0.99	0.447	0.704	0.673	0.649	0.657	0.622
1.00	0.382	0.684	0.656	0.634	0.654	0.630
1.01	0.290	0.516	0.501	0.487	0.501	0.485
1.02	0.235	0.189	0.184	0.178	0.180	0.171
		D Horizon $h=50$ ($\lambda=0.50$)				
0.80	0.629	0.998	0.658	0.636	0.630	0.629
0.90	0.604	0.974	0.811	0.739	0.651	0.613
0.93	0.585	0.944	0.843	0.777	0.686	0.615
0.95	0.548	0.898	0.829	0.777	0.700	0.613
0.97	0.503	0.837	0.788	0.751	0.712	0.632
0.98	0.450	0.786	0.743	0.710	0.693	0.625
0.99	0.386	0.743	0.707	0.677	0.681	0.631
1.00	0.285	0.681	0.650	0.626	0.644	0.611
1.01	0.168	0.364	0.350	0.338	0.345	0.327
1.02	0.108	0.087	0.083	0.080	0.079	0.073

See the notes to Table 1

pretest coverage rates are typically too tight (and are centered in the wrong place) for α exceeding one and are too wide for C approximately -10, a value of C for which the pretest forecast often incorrectly picks the random walk model.

The asymptotic theory predicts that the distortions in Tables 1 and 2 will persist as the sample size increases, when the local to unity parameter C is held constant. This is verified in Table 3 (relative RMSEs) and Table 4 (coverage rates) for selected values of C for 500 observations. A comparison of the corresponding rows of Tables 1 and 3 (e.g. $\alpha=0.9$ in Table 1, $C=-10$ in Table 3), and of Tables 2 and 4, demonstrates close agreement between the results for $T=100$ and $T=500$, confirming this aspect of the asymptotic theory.

TABLE 3

Relative Root Mean Square Error of $\hat{X}_{T+h|T}-X_{T+h}$ for Various Forecasts, $T=500$

Entries are ratio of forecast error RMSE based on estimated α to forecast error RMSE based on true α

	Forecast					
			DF Pretest		DF–GLS Pretest	
C	OLS	Random Walk	5%	10%	5%	10%
			A Horizon $\lambda=0$ 02			
−10	1 013	1 048	1 046	1 043	1 031	1 022
−5	1.022	1 023	1 031	1 035	1 029	1 031
−3	1 031	1 015	1 023	1 028	1 021	1 022
−2	1 046	1 008	1 018	1 026	1 014	1 018
0	1 129	1 000	1 030	1 046	1 003	1 006
			B Horizon $\lambda=0$ 10			
−10	1 014	1 206	1 177	1 153	1 103	1 059
−5	1 039	1.111	1 113	1 114	1 110	1 102
−3	1 066	1 074	1 083	1 088	1 081	1 082
−2	1 108	1 044	1 061	1 074	1 053	1 059
0	1 335	1 000	1 055	1 093	1 006	1 012
			C Horizon $\lambda=0$ 20			
−10	1 009	1 322	1 273	1 233	1 165	1.090
−5	1.031	1 206	1 197	1 191	1 188	1 167
−3	1 066	1 136	1 137	1 136	1 139	1 135
−2	1 103	1 093	1 100	1 107	1 096	1 100
0	1 402	1 000	1 047	1 084	1 006	1 012
			D Horizon $\lambda=0$ 50			
−10	1 001	1.421	1 356	1 304	1 210	1 109
−5	1 027	1 352	1 335	1 320	1 311	1 266
−3	1 050	1 272	1 265	1 256	1 264	1 250
−2	1 102	1 205	1 201	1 197	1 202	1 199
0	1 441	1 000	1 027	1 053	1 003	1 006

See notes to Table 1

VI. DISCUSSION AND CONCLUSIONS

The numerical results indicate that the magnitude of bias of VAR and VECM point forecasts, and of the distortions of their associated coverage rates, can be substantial, even for horizons of only 10 percent or 20 percent of the sample size. These findings can be interpreted either from an asymptotic or finite sample perspective. From an asymptotic perspective, under the local-to-unity nesting these distortions persist even with arbitrarily large sample sizes. From a finite-sample perspective, the results were computed for an AR(1) with 100 observations and thus they can alternatively be viewed as indicative of problems that arise with roots of,

TABLE 4

Asymptotic Coverage Rates of Nominal 68% Prediction Intervals, T = 100

Entries are the probability that the h-step ahead prediction interval computed as $\hat{X}_{T+h|T}$ ± one standard deviation, using the estimated value of α, contains X_{T+h}

	Forecast					
			DF Pretest		DF–GLS Pretest	
α	OLS	Random Walk	5%	10%	5%	10%
			A Horizon λ = 0.02			
−10	0.653	0.705	0.675	0.663	0.655	0.651
−5	0.644	0.692	0.677	0.667	0.663	0.652
−3	0.644	0.686	0.675	0.669	0.672	0.664
−2	0.631	0.685	0.674	0.667	0.674	0.667
0	0.608	0.682	0.668	0.660	0.677	0.674
			B Horizon λ = 0.10			
−10	0.614	0.789	0.694	0.655	0.623	0.608
−5	0.588	0.739	0.701	0.673	0.654	0.617
−3	0.565	0.718	0.688	0.668	0.669	0.638
−2	0.540	0.702	0.676	0.658	0.668	0.642
0	0.461	0.682	0.659	0.640	0.668	0.656
			C Horizon λ = 0.20			
−10	0.607	0.870	0.743	0.687	0.628	0.605
−5	0.556	0.791	0.739	0.698	0.664	0.604
−3	0.529	0.755	0.718	0.688	0.685	0.637
−2	0.495	0.731	0.699	0.675	0.682	0.643
0	0.380	0.678	0.652	0.632	0.660	0.643
			D Horizon λ = 0.50			
−10	0.617	0.975	0.827	0.753	0.670	0.628
−5	0.554	0.898	0.832	0.779	0.733	0.648
−3	0.499	0.838	0.791	0.751	0.742	0.672
−2	0.453	0.795	0.755	0.723	0.726	0.668
0	0.289	0.684	0.656	0.631	0.660	0.637

See notes to Table 1

for example, 0.9 with that sample size. These conclusions are similar in spirit to those drawn from recent work on testing values of long-run parameters, specifically that the lack of knowledge of whether the largest autoregressive root is exactly one introduces substantial difficulties for inference. Moreover, if the exact unit root model is imposed, hypothesis tests on long-run parameters can have coverage rates differing substantially from those in finite samples; cf. Elliott (1994) and Cavanagh, Elliott and Stock (1995).

One practical implication of this work is that pretest forecasts and intervals are arguably preferable to the OLS/VAR estimator if there is reason to believe that the largest roots are close to one. The DF–GLS pretest forecasts generally outperform the DF pretest forecasts. In addition, coverage rates of DF–GLS pretest prediction intervals were found to be fairly constant across different values of $C \leq 0$; for $C \leq 0$, the nominal 68 percent interval based on the 10 percent DF–GLS pretest has an asymptotic coverage rate of approximately 60 percent for forecast horizons up to 50 percent.

These results therefore provide a cautious endorsement of pretest forecasts and intervals if they are based on an efficient pretest and if the researcher has reason to believe that the largest root is at most unity. It should be emphasized, however, that these results strictly apply only to the univariate AR(1) model, and the performance of pretest forecasts could deteriorate substantially in multivariate models and/or with lag structures typically found in empirical work.

Although the pretest procedures considered here focus on tests with fixed size, the results are informative about forecasts based on consistent model selection strategies, which can loosely be thought of as pretests with critical values that depend on the sample size. For a model selection strategy to be consistent, the size of the pretest must go to zero as $T \to \infty$. It follows that, asymptotically, a consistent model selection procedure will classify a root local to unity as exactly unity with probability tending to one; that is, it will pick the VECM model asymptotically for all finite C. Thus the random walk columns of Tables 1–4 can be reinterpreted as numerical approximations to the asymptotic performance of a consistent model selection algorithm. Evidently, long horizon forecasting with consistent model selection produces point forecasts which can have large asymptotic RMSEs and prediction intervals which are far wider than desirable, depending on the true value of C. At issue is the subtle point that, although consistent model selection procedures asymptotically choose the correct model for any *fixed* value of A, they do not select the correct model *uniformly* in A. This pitfall of long horizon forecasting using a consistent model selection algorithm parallels the failure of consistent model selection procedures to produce tests of long-run parameters which control size uniformly in A (cf. Elliott and Stock (1994)).

Kennedy School of Government, Harvard University

Date of Receipt of Final Manuscript, August 1996

REFERENCES

Bobkoski, M. J. (1983). *Hypothesis Testing in Nonstationary Time Series*, unpublished PhD dissertation, Department of Statistics, University of Wisconsin.

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*, Vol. 11, pp. 1131–47.

Chan, N. H. and Wei, C. Z. (1987). 'Asymptotic Inference for Nearly Nonstationary AR(1) Processes', *Annals of Statistics*, Vol. 15, pp. 1050–63.

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*, Vol. 74, pp. 427–31.

Elliott, G. (1994). *Application of Local to Unity Asymptotic Theory to Time Series Regression*, unpublished PhD dissertation, Department of Economics, Harvard University.

Elliot, G., Rothenberg, T. J. and Stock, J. H. (1996). 'Efficient Tests for an Autoregressive Unit Root', *Econometrica*, Vol. 64, pp. 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*, Vol. 10, pp. 672–700.

Johnansen, S. (1988). 'Statistical Analysis of Cointegrating Vectors', *Journal of Economic Dynamics and Control*, Vol. 12, pp. 231–54.

Magnus, J. R. and Pesaran, B. (1991). 'The Bias of Forecasts from a First-Order Autoregression', *Econometric Theory*, Vol. 7, pp. 222–35.

Manne, A. S. and Richels, R. (1992). *Buying Greenhouse Insurance: The Economic Costs of CO_2 Emission Limits*, MIT Press, Cambridge, Mass.

Nordhaus, W. D. and Yang, Z. (1995). 'RICE: A Regional Dynamic General Equilibrium Model of Optimal Climate-Change Policy', manuscript, Department of Economics, Yale University.

Phillips, P. C. B. (1979). 'The Sampling Distribution of Forecasts from a First-Order Autoregression', *Journal of Econometrics*, Vol. 9, pp. 241–61.

Phillips, P. C. B. (1987). 'Towards a Unified Asymptotic Theory for Autoregression', *Biometrika*, Vol. 74, pp. 535–47.

Phillips, P. C. B. (1995). 'Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VAR's', manuscript, Cowles Foundation for Research in Economics, Yale University.

Sims, C. A., Stock, J. H. and Watson, M. W. (1990). 'Inference in Linear Time Series Models with Some Unit Roots', *Econometrica*, Vol. 58, pp. 113–44.

Stock, J. H. (1991). 'Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time Series', *Journal of Monetary Economics*, Vol. 28, pp. 435–60.

Stock, J. H. (1994). 'Unit Roots, Structural Breaks, and Trends', ch. 46 in Engle,

R. and McFadden, D. (eds.), *Handbook of Econometrics, Volume IV*, Elsevier, Amsterdam.

Stock, J. H. (1995). 'Cointegration, Long-Run Comovements, and Long-Horizon Forecasting', forthcoming in Kreps, D. and Wallis, K. F. (eds.), *Advances in Econometrics: Proceedings of the Seventh World Congress of the Econometric Society*, Cambridge University Press, Cambridge.

Stock, J. H. and Watson, J. W. (1993). 'A Simple Estimator of Cointegrating Vectors in Higher-Order Integrated Systems', *Econometrica*, Vol. 61, pp. 783–820.

APPENDIX

Derivation of (2)

Premultiply (1a) by $(I-AL)$ and use the definition of $\tilde{X}_t$ to write,

$$\tilde{X}_t=\tilde{\beta}+U_t \tag{A1}$$

where $\tilde{\beta}=[\beta_1'(I-A)'\ \beta_2']'$. Next premultiply (A1) by $B(L)^{-1}$ and rearrange terms to obtain,

$$X_{1t}=\tilde{\mu}_1+AX_{1t-1}+[I_k\ \ 0]\tilde{D}(L)\tilde{X}_{t-1}+\varepsilon_{1t} \tag{A2a}$$

$$X_{2t}=\mu_2+\theta X_{1t}+[0\ \ I_{n-k}]\tilde{D}(L)\tilde{X}_{t-1}+\varepsilon_{2t} \tag{A2b}$$

where $\tilde{\mu}=B(1)^{-1}\tilde{\beta}$ and $\tilde{D}(L)=L^{-1}(I-B(L)^{-1})$. The expression (2) obtains by substituting (A2a) into (A2b) and collecting terms.

Proof of Theorem 1

(a) Using (1a) and $(I-AL)V_t=U_{1t}$, we have,

$$T^{-1/2}X_{1T+h}=A^hT^{-1/2}X_{1T}+\phi_T(h)+m_T(h) \tag{A3a}$$

$$T^{-1/2}X_{2T+h}=\theta T^{-1/2}X_{1T+h}+T^{-1/2}U_{2T+h} \tag{A3b}$$

where $m_T(h)=T^{-1/2}\Sigma_{i=0}^{h-1}A^i(I-A)\beta_1$ and $\phi_T(h)=T^{-1/2}\Sigma_{i=0}^{h-1}A^iU_{1T+h-i}$. Substitute $h=[T\lambda]$ into (A3) and consider the various terms. Note that $A^{[T\lambda]}=(I+C/T)^{[T\lambda]}\rightarrow\exp(C\lambda)$.

First, $|m_T([T\lambda])|\leq k^2T^{-1/2}[T\lambda]\,(\max_{0\leq i\leq[T\lambda]}|A^i|)|I-A||\beta_1|\leq KT^{1/2}|I-(I+C/T)|$, where $K=k^2\bar{\lambda}\sup_{0\leq\lambda\leq\bar{\lambda}}|\exp(C\lambda)|\ \beta_1|$. The assumptions of the theorem imply $K<\infty$; because $T^{1/2}|C/T|\rightarrow 0$, $m_T([T\lambda])\rightarrow 0$ uniformly in λ, $0\leq\lambda\leq\bar{\lambda}$.

Second, for $B(L)$ one-summable and $|C|<\infty$, standard local to unity asymptotic results imply that $J_{c,T}\Rightarrow J_c$, where $J_{c,T}(s)=T^{-1/2}\Sigma_{t=1}^{[Ts]}A^{[Ts]-t}U_{1t}$. Now $\phi_T([T\lambda])=J_{c,T}(1+\lambda)-A^{[T\lambda]}J_{c,T}(1)\Rightarrow J_c(1+\lambda)-\exp(C\lambda)J_c(1)\equiv\phi(\lambda)$. Also note that $T^{-1/2}U_{2T+[T\lambda]}\xrightarrow{p}0$ for $B(L)$ one-summable. All these limits are uniform in λ, $0\leq\lambda\leq\bar{\lambda}$.

Combining these results, we have,

$$T^{-1/2}X_{1T+h}-A^{[T\lambda]}T^{-1/2}X_{1T}\Rightarrow\phi(\lambda) \tag{A4a}$$

$$T^{-1/2}X_{2T+h}-\theta T^{-1/2}X_{1T+h}\xrightarrow{p}0 \tag{A4b}$$

for $h=[T\lambda]$, uniformly in λ, so that $T^{-1/2}X_{T+h}-\Gamma A^{[T\lambda]}T^{-1/2}X_{1T}\Rightarrow\Gamma\phi(\lambda)$ uniformly in λ.

(b) This follows from part (a) and from writing $\phi(\lambda)=\int_{s=1}^{1+\lambda}\exp(C(1+\lambda-s))Q\,\mathrm{d}W(s)$, where $Q=[I_k,\ 0]B(1)$. Evidently, for fixed λ, $\phi(\lambda)$ is distributed $N(0,\int_0^\lambda\exp(C(\lambda-s)Q\Sigma Q'\exp(C(\lambda-s))'\,\mathrm{d}s)$, and the expression for the limiting conditional distribution of $T^{1/2}X_{T+h}$ follows.

Also note that $\phi(\lambda)$ is independent of $\{W(s),\ 0\leq s\leq 1\}$ and thus of $\{J_c(s),\ 0\leq s\leq 1\}$. It follows in particular that statistics which have a limiting representation solely in terms of $W(s)$, $J_c(s)$, $0\leq s\leq 1$, and nuisance parameters are asymptotically independent of $\phi(\lambda)$. Such statistics include $\hat{C}^{\mathrm{VAR}}$ and the unit root pretests discussed in Section III.