CONFIDENCE SETS IN REGRESSIONS WITH HIGHLY SERIALLY CORRELATED REGRESSORS

James H. Stock

Kennedy School of Government, Harvard University
and the NBER

and

Mark W. Watson*

Woodrow Wilson School, Princeton University
and the NBER

December 1996

*We thank Don Andrews, Graham Elliott, Bruce Hansen, Peter Phillips, and Jonathan Wright for helpful comments and discussions. An earlier version of this paper was circulated under the title, "Confidence Sets in Cointegrating Regressions with Nearly Integrated Regressors." This research was supported in part by National Science Foundation grant no. SBR-9409629.

ABSTRACT

Small deviations from exact unit roots can produce large coverage rate distortions for conventional confidence sets for cointegrating coefficients (Elliott [1994]). We therefore propose new methods for constructing confidence sets for long-run coefficients with highly serially correlated regressors which do not necessarily have a unit root. Although the standard bootstrap is shown to be asymptotically invalid, a modified, valid bootstrap is developed. Invariant confidence sets that are optimal (highest average accuracy) are obtained but are difficult to implement in practice. An approximately optimal invariant method is proposed; this works almost as well as the optimal method, at least for a single persistent regressor.

Key Words: cointegration, local to unit roots, money demand

1. Introduction

Asymptotically efficient estimators of cointegrating coefficients are widely used in empirical work to perform inference about long-run relations. These estimators are appropriate when some of the roots of the system are exactly equal to one and the others are well within the range of stationarity. However, Elliott (1994) showed that inferences based on these procedures depend delicately on the largest roots being exactly rather than nearly one. Indeed, in large samples, values of the largest autoregressive roots that can be distinguished from one by unit root tests only with small asymptotic probability can result in tests of hypotheses on cointegrating coefficients having sizes far in excess of their nominal levels. Similarly, confidence sets for the coefficients of long-run relations can have coverage rates far from their nominal coverage rate if the regressors are highly serially correlated but do not have an exact unit root.

This problem is of empirical relevance. For example, numerous authors have considered the problem of estimating long-run income and interest elasticities of money demand, and the recent literature has focused overwhelmingly on the use of cointegrating techniques (cf. Baba, Hendry, and Starr [1992], Hafer and Jansen (1991), Hoffman and Rasche [1991], Miller (1991), and Stock and Watson [1993]). However, the debate about whether real GDP has an exact unit root remains unresolved, and a 90% confidence interval for the largest root root in quarterly U.S. real GDP from 1970 to 1994 is (0.744, 1.029) (Stock [1996]). Additionally, the exact unit root model must be an imperfect characterization of interest rate dynamics, because the exact unit root model implies that the nominal interest rate can (and will) go below zero and grow without bound. At best, the exact unit root assumption, upon which this long run money demand literature literature is predicated, is questionable.

In this paper we consider a variety of possible methods for constructing confidence regions for the coefficients in a single cointegrating equation when there are two asymptotically important

vectors of nuisance parameters: the largest autoregressive roots and the drift parameters of the regressor process. The presence of a small deterministic trend adds an additional layer of complication to this problem: with no deterministic trend, unit-root type distribution theory applies; with a large deterministic trend, some coefficients can have asymptotic normal distributions (West [1988]); and with a small deterministic trend, the distributions fall somewhere in between (Stock and West [1988]). To focus on the problematic values of these parameters, we model them as falling in, respectively, a T^{-1} region of unity (Elliott's [1994] approach, cf. Phillips [1987], Chan and Wei [1987], and Cavanagh, Elliott and Stock [1995]) and a $T^{-\frac{1}{2}}$ region of zero. These neighborhoods are chosen so that tests of a unit root, or of zero time trend, are not consistent against these alternatives. These nestings are designed capture the circumstance in which the applied researcher is faced with data that appear to have large roots and/or a small trend component, but the empirical evidence on these assumptions is ambiguous.

The bootstrap provides one way to construct confidence regions when the performance of confidence sets based on first order asymptotics is poor. We show, however, that the bootstrap (more precisely, the parametric bootstrap percentile-t interval in Hall's (1992) terminology) is invalid in this application in the sense that its asymptotic coverage rate differs from its nominal coverage rate.

We therefore consider alternative confidence regions that have controlled asymptotic coverage rates. One of these is a modification of the bootstrap, where the nominal coverage rate is adjusted so that the actual coverage rate is controlled. A second is a sup-bound confidence region. Although these alternative procedures achieve the desired asymptotic coverage rate, both are found to produce wide confidence sets (and tests with low power) for some values of the nuisance parameters.

We then turn to the problem of producing optimal confidence sets, where optimality is defined in terms of best average accuracy among invariant tests. There is no evident analytic solution to

this problem, but it can be solved numerically (with some approximations) and we do so for the case of a single right hand variable. Because these optimal tests are difficult to implement, we develop a simpler version based on the same principle. This simpler, approximately optimal test is found to perform almost as well as as the optimal test.

There is a small literature on this problem, which is reviewed by Stock (1996). If the joint model has a finite-order autoregressive representation of order p, then one approach is to augment the model by additional lags which are known to have zero coefficients; inference on the coefficients with unknown values is now χ^2. 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 (1995) extend it to a general VAR(p). This approach can produce wide confidence sets in Monte Carlo simulations, however, and indeed the confidence sets have width that is $O_p(T^{-1/2})$ rather than the optimal rate T^{-1}. Campbell and Dufour (1995) propose a nonparametric approach to this problem in the case of a known intercept and a scalar regressor, but the extension to an unknown intercept and especially to multiple regressors is unclear. The only general approach that achieves the optimal T^{-1} rate is the method proposed by Wright (1996a). All the procedures considered in this paper achieve the optimal rate.

The paper is organized as follows. The model and preliminary asymptotic results are presented in section 2. Bootstrap and related confidence sets are examined in section 3. Optimal and approximately optimal invariant confidence sets are developed in section 4. Numerical results are presented in section 5, and section 6 contains an application to long run money demand in the U.S. Section 7 concludes.

2. The Model and Asymptotic Framework

2.1. The Model

The model considered is the n-dimensional cointegrated VAR, written in triangular form:

$$X_t = \gamma_0 + \gamma_1 t + V_{X,t} \tag{2.1}$$

$$y_t = \alpha_y + \theta' X_t + u_{y,t} \tag{2.2}$$

where y_t is a scalar and X_t is $k\times 1$ (so $n=k+1$). The error process $V_{X,t}$ evolves as:

$$V_{X,t} = AV_{X,t-1} + u_{X,t} \tag{2.3}$$

Let $U_t = (u_{X,t}', u_{y,t})'$ denote the n dimensional error vector. U_t is assumed to be a second order stationary stochastic process. We work with the two-sided triangular representation, $U_t = H(L)\epsilon_t$, where ϵ_t is a white noise sequence with $E(\epsilon_t\epsilon_t')=I$ (cf. Watson [1994, sec. 3.2]). Partition H(L) and ϵ_t conformably with $(u_{X,t}', u_{y,t})'$; then, in the obvious notation, $\epsilon_t = (\epsilon_{1,t}', \epsilon_{2,t})'$. Without loss of generality, let $H_{11}(L)$ and $H_{22}(L)$ be one-sided, $H_{12}(L)=0$, $H_{21}(L)$ be two sided in general, $H_{11}(0)$ be lower triangular, and the diagonal elements of H(0) be non-negative. It is further assumed that $H_{11}(L)$ and $H_{22}(L)$ are invertible. These assumptions uniquely determine H(L) from the autocovariances of U_t.

It will be convenient to use two alternative representations of this model. The first takes advantage of the lower triangular representation of H(L) to reexpress the cointegrating equation (2.2) with an error that is uncorrelated with the right hand side variables at all leads and lags. Let $d(L)=H_{21}(L)H_{11}(L)^{-1}$ and $\eta_t=H_{22}(L)\epsilon_{2,t}$, and substitute (2.3) into (2.1). The model can then be expressed as,

(2.4) $$X_t = \alpha_X + \rho_X t + AX_{t-1} + u_{X,t}$$

(2.5) $$y_t = \tilde{\alpha}_y + \theta' X_t + d(L)(X_t - \rho_X t - AX_{t-1}) + \eta_t$$

where $\alpha_X = (I-A)\gamma_0 + A\gamma_1$, $\rho_X = (I-A)\gamma_1$, and $\tilde{\alpha}_y = \alpha_y - d(1)\alpha_X$. We will assume that the error $u_{X,t}$ follows a VAR(q), and it will be convenient to reexpress (2.4) explicitly as a VAR. Obtained by multiplying (2.4) by $H_{11}(L)^{-1}$ and rearranging, the VAR representation of X_t is,

(2.6) $$\Delta X_t = \Psi_1 X_{t-1} + \Psi_2 t + \Psi_3 + \Psi_4(L)\Delta X_{t-1} + \tilde{\epsilon}_{1t}$$

where $\tilde{\epsilon}_{1t} = H_{11}(0)\epsilon_{1t}$, $\Psi_1 = H_{11}(0)H_{11}(1)^{-1}(A-I)$, $\Psi_2 = H_{11}(0)H_{11}(1)^{-1}\rho_X$, $\Psi_3 = H_{11}(0)[H_{11}(1)^{-1}\alpha_X + H^{11}*(1)\rho_X]$, and $\Psi_4 = H_{11}(0)[L^{-1}(H_{11}(0)^{-1} - H_{11}(L)^{-1}) + H^{11}*(L)(A-I)]$, where $H^{11}*(L) = \sum_{i=0}^{q} H_i^{11} L^i$ where $H_i^{11}* = -\sum_{j=i+1}^{q} H_j^{11}$, and where H_j^{11} is the jth lag coefficient matrix in $H_{11}(L)^{-1}$.

A second, equivalent representation of the model is in a coordinate system chosen so that the key parameters are unitless. Let θ_0 denote a particular value of θ; in the subsequent discussion of hypothesis tests, θ_0 will denote the value of θ under the null hypothesis. Define $Z_t = H_{11}(1)^{-1}X_t$ and $w_t = H_{22}(1)^{-1}(y_t - \theta_0' X_t)$. Then the system can be rewritten,

(2.7) $$Z_t = \delta_0 + \delta_1 t + V_{Z,t}, \quad V_{Z,t} = BV_{Z,t} + \nu_{Z,t},$$

where $\delta_0 = H_{11}(1)^{-1}\gamma_0$, $\delta_1 = H_{11}(1)^{-1}\gamma_1$, $B = H_{11}(1)^{-1}AH_{11}(1)$, $V_{Z,t} = H_{11}(1)^{-1}V_{X,t}$, and $\nu_{Z,t} = H_{11}(1)^{-1}u_{X,t}$. Similarly, in these transformed coordinates (2.4) and (2.2) become,

(2.8) $$Z_t = \alpha_Z + \rho_Z t + BZ_{t-1} + \nu_{Z,t}$$

(2.9) $$w_t = \alpha_w + (\beta-\beta_0)'Z_t + \nu_{w,t}$$

where $\alpha_Z=(I-B)\delta_0+B\delta_1$, $\rho_Z=(I-B)\delta_1$, $\alpha_w=H_{22}(1)^{-1}\alpha_y$, $\beta=H_{11}(1)'\theta H_{22}(1)^{-1}$, $\beta_0=H_{11}(1)'\theta_0 H_{22}(1)^{-1}$, and $\nu_{w,t}=H_{22}(1)^{-1}u_{y,t}$. The errors $\nu_t = (\nu_{Z,t}' \ \nu_{w,t})'$ in the transformed model obey $F(L)\nu_t = \epsilon_t$, where $F(L)=H(L)^{-1}\text{diag}(H_{11}(1),H_{22}(1))$ so $F_{11}(L)=H_{11}(L)^{-1}H_{11}(1)$, $F_{12}(L)=0$, $F_{21}(L)=-H_{22}(L)^{-1}H_{21}(L)H_{11}(L)^{-1}H_{11}(1)$, and $F_{22}(L)=H_{22}(L)^{-1}H_{22}(1)$, where F(L) is partitioned conformably with ν_t.

2.2. Asymptotic Framework and Assumptions

This paper considers the possibilities that the matrix of largest autoregressive roots of X_t is nearly but not necessarily the identity matrix, and that X_t has a small, possibly nonzero time trend component. The assumptions are made in the transformed coordinates (2.7)-(2.9). Specifically, B is modeled as being sufficiently close to I_k that a test of the hypothesis $B=I_k$ would fail to reject the null with positive probability, even asymptotically. Similarly, δ_1 is modeled as being sufficiently small that a test of $\delta_1=0$ would fail to reject the null with positive probability asymptotically. Accordingly, a local reparameterization is adopted for the asymptotic analysis:

> *Assumption A.* $B = I+C/T$, $\delta_1 = \omega/T^{1/2}$, and $\beta = \beta_0 + b/T$, where C is a fixed $k\times k$ matrix and ω and b are fixed $k\times 1$ vectors. Let $r = -C\omega$.

The nesting $B=I+C/T$ is the familiar multivariate local-to-unity model for large autoregressive roots, cf. Phillips (1987).

Several additional technical assumptions are made for the asymptotic analysis. These are collected as assumption B.

Assumption B.

(i) ϵ_t is a martingale difference sequence with $\max_i \sup_t E\epsilon_{it}^4 < \infty$ and $E\epsilon_t\epsilon_t' = I$.

(ii) $T^{-1/2}V_{X,0} \xrightarrow{p} 0$

(iii) Sample first and second moments of I(0) variables are consistent for their expectations.

(iv) $H_{11}(L)^{-1}$ and $H_{22}(L)^{-1}$ have maximum order q and fixed roots that are outside the unit circle, and $d(L) = \sum_{j=-p}^{p} d_j L^j$, where p and q are known.

The asymptotic results involve representations in terms of functionals of Brownian motion. Under assumption B,

$$T^{-1/2}\sum_{t=1}^{[Ts]}\epsilon_t \Rightarrow W(s) \tag{2.10}$$

where W is a n-dimensional standard Brownian motion, $\Rightarrow$ denotes weak convergence on $D[0,1]^n$, and $[\bullet]$ denotes the greatest lesser integer. Partition W as $W=(W_1' \; W_2)'$ conformably with ϵ_t. Let $J_{C,\omega}(s)$ be the continuous time sum of a deterministic trend and the diffusion process $J_C(s)$:

$$J_{C,\omega}(s) = \omega s + J_c(s), \tag{2.11}$$

$$dJ_C(s) = CJ_C(s)ds + dW_1(s). \tag{2.12}$$

Under assumptions A and B,

$$T^{-1/2}Z_{[Ts]} \Rightarrow J_{C,\omega}(s) \tag{2.13}$$

where this limit is joint with (2.10).

Most of the statistics in sections 3 and 4 involve estimators of functions of H(L), and for some of the results these estimators must be $T^{1/2}$-consistent. It is assumed that a $T^{1/2}$-consistent estimator of $H(L)^{-1}$, $\hat{H}(L)^{-1}$, is available:

Assumption C. Under assumptions A and B, $\hat{H}(z)^{-1}-H(z)^{-1} = O_p(T^{-1/2})$ for fixed z, $\|z\| \leq 1$.

Because $H(L)^{-1}$ is finitely parameterized (assumption B(iv)), there are several such estimators. One which is readily computed is obtained by imposing the restrictions C=0 and ω=0. Under these restrictions, the VAR lag coefficients in (2.6) simplify to $\Psi_4(L) = L^{-1}(I-H_{11}(0)H_{11}(L)^{-1})$. Because $E(\tilde{\epsilon}_{1t}\tilde{\epsilon}_{1t}') = H_{11}(0)H_{11}(0)'$ and $H_{11}(0)$ is assumed to be lower triangular, let $\hat{H}_{11}(0)$ be the Cholesky factor of the residual covariance matrix from (2.6), and set $\hat{H}_{11}(L)^{-1} = \hat{H}_{11}(0)^{-1}[I-L\hat{\Psi}_4(L)]$, where $\hat{\Psi}_4(L)$ is the OLS estimator of $\Psi_4(L)$ from (2.6). Similarly, imposing ρ_X=0 and A=I and estimating (2.5) by OLS, where d(L) has q leads and lags, produces an estimator $\hat{d}(L)$, and $\hat{H}_{21}(L)=\hat{d}(L)\hat{H}_{11}(L)$ serves an estimator of $H_{21}(L)$. Estimation of an AR(p) using the residuals from that regression produces the estimator $\hat{H}_{22}(L)^{-1}$. Estimators of other functions of H(L) can in turn be constructed from these estimators; for example, an estimator of F(L) is $\hat{F}(L) = \hat{H}(L)^{-1}\text{diag}(\hat{H}_{11}(1), \hat{H}_{22}(1))$. Under assumption A these estimators are based on mispecified models. Nonetheless, calculations like those in Elliott (1994) show that they satisfy assumption C. In practice, other estimators of H(L) might be preferred, particularly if C and ω are far from zero. We therefore do not specify a particular set of such estimators for the theoretical development but rather require only that they satisfisfy assumption C.

Finally, it is useful to adopt some additional notation. Let $J^{\mu}_{C,\omega}(s)=J_{C,\omega}(s)-\int J_{C,\omega}$, $s^{\mu}=s-1/2$, and $J^{\dagger}_{C,\omega} = [J^{\mu}_{C,\omega}\ s^{\mu}]$. Unless explicit limits are given, summations of sample variables are over $p+1 \leq t \leq T-p$, so for example $\sum_{s=p+1}^{T-p} Z_s$ is denoted $\sum Z_s$. Let the superscript μ on sample variables denote demeaning, for example $Z^{\mu}_t=Z_t-(T-2p)^{-1}\sum Z_s$.

3. Bootstrap and Related Confidence Intervals

3.1. Test Statistics and Preliminary Asymptotics

This section examines the performance of confidence intervals constructed using the bootstrap, a size-adjusted bootstrap, and a sup-bound approach. The versions of these intervals considered here are based on statistics obtained from the system (2.5) and (2.6), and these statistics and their asymptotic distributions are presented before turning to the confidence intervals.

The parameters of (2.5) cannot be estimated in a single equation because A and ρ_X are unknown. However, a family of feasible estimators is obtained by using hypothesized or estimated values of A and ρ_X. Specifically, let C_T and r_T denote hypothesized or estimated sequences that relate to C and r. The sequences C_T and r_T can be either stochastic or nonstocastic (examples are discussed below). Let $\tilde{A}=I+T^{-1}H_{11}(1)C_TH_{11}(1)^{-1}$, $\tilde{\rho}_X = T^{-3/2}H_{11}(1)r_T$, and $\tilde{u}_{X,t}\equiv X_t-\tilde{\rho}_X t-\tilde{A}X_{t-1}$. Then (2.5) can be written,

$$y_t = \tilde{\alpha}_y + \theta' X_t + d(L)\tilde{u}_{X,t} + \tilde{\eta}_t \tag{3.1}$$

where $\tilde{\eta}_t = \eta_t + d(L)(u_{X,t}-\tilde{u}_{X,t})$.

Let $\hat{\theta}(C_T,r_T)$ denote the OLS estimator of θ from the regression (3.1). The autocorrelation-consistent Wald statistic testing the null hypothesis $R\theta=R_0$ from this regression is,

$$S(C_T,r_T) = [R\hat{\theta}(C_T,r_T)-R_0]'[R\hat{V}(C_T,r_T)R']^{-1}[R\hat{\theta}(C_T,r_T)-R_0] \tag{3.2}$$

where $\hat{V}(C_T,\omega_T) = M^{\theta\theta}\hat{H}_{22}(1)^2$, where M is the moment matrix of the regressors, and $M^{\theta\theta}$ denotes the elements of M^{-1} corresponding to θ.

One special case is when $C_T=0$ and $r_T=0$ for all T. The estimator $\hat{\theta}(0,0)$ is the dynamic OLS (DOLS) estimator of Stock and Watson (1993) (also see Phillips and Loretan [1991], Saikkonen [1991]). This is an efficient estimator of the cointegrating coefficients under the standard definition of cointegration, $C=0$ and $\omega=0$ (so $r=0$), and in this case the Wald statistic S(0,0) has a limiting χ^2 distribution. However, in the case $\omega=0$ and $C\neq 0$ Elliot (1994) showed that the DOLS Wald statistic no longer has a χ^2 distribution and that critical values for S(0,0) depend on C.

Another special case is when C_T and r_T are the OLS estimators of these parameters obtained from estimation of the VAR (2.6). Let $\hat{\Psi}_1$ and $\hat{\Psi}_2$ denote the estimators of Ψ_1 and Ψ_2 from OLS estimation of (2.6). The estimators $\hat{C}$ and $\hat{r}$ based on $\hat{\Psi}_1$ and $\hat{\Psi}_2$ are,

$$\hat{C} = T\hat{H}_{11}(0)^{-1}\hat{\Psi}_1\hat{H}_{11}(1) \tag{3.3a}$$

$$\hat{r} = T^{3/2}\hat{H}_{11}(0)^{-1}\hat{\Psi}_2. \tag{3.3b}$$

The resulting estimator of θ, $\hat{\theta}(\hat{C},\hat{r})$, is asymptotically the Gaussian MLE when no a-priori information is available regarding C and ω.

The asymptotic behavior of S is given in the following theorem.

Theorem 1. Suppose that (2.4), (2.5), and assumptions A-C hold. Then:

(a) Let (C_T, r_T) equal the fixed values $(\bar{C}, \bar{r})$. Then

$$\begin{aligned} S(C_T,r_T) \Rightarrow\ & [(\int J^{\mu}_{C,\omega}J^{\mu}_{C,\omega}{}')^{-1}\int J^{\mu}_{C,\omega}dW_2+g]'K' \\ & \times[K(\int J^{\mu}_{C,\omega}J^{\mu}_{C,\omega}{}')^{-1}K']^{-1}K[(\int J^{\mu}_{C,\omega}J^{\mu}_{C,\omega}{}')^{-1}\int J^{\mu}_{C,\omega}dW_2+g] \\ \equiv\ & S^*(\bar{C},\bar{r}) \end{aligned}$$

where $g = b-[(\int J^{\mu}_{C,\omega}J^{\mu}_{C,\omega}{}')^{-1} \int J^{\mu}_{C,\omega}s^{\mu}(\bar{r}-r)'+(\bar{C}-C)']F_{21}(1)'$ and $K=RH_{11}(1)^{-1}{}'$. Conditional on $\{J_{C,\omega}\}$, this is distributed as a noncentral $\chi^2_{rank(R)}$, with noncentrality parameter $g'K'[K(\int J^{\mu}_{C,\omega}J^{\mu}_{C,\omega}{}')^{-1})K']^{-1}Kg$.

(b) Let (C_T,r_T) equal the VAR estimates $(\hat{C},\hat{r})$ given in (3.3). Then

$$[\hat{C}\text{-}C \;\; \hat{r}\text{-}r] \Rightarrow (\int J^{\dagger}_{C,\omega}dW_1')'(\int J^{\dagger}_{C,\omega}J^{\dagger}_{C,\omega}{}')^{-1} \equiv [C^* \;\; r^*]$$

and $S(\hat{C},\hat{r}) \Rightarrow S^*(C^*+C,r^*+r)$.

All proofs are contained in the appendix.

The essential idea of this theorem is due to Elliott (1994), who obtained the limit $S(0,0) \Rightarrow S^*(0,0)$ when $\omega=0$; this theorem extends his result to tests that use nonzero (C_T,r_T) and to nonzero ω. Elliott (1994) showed that if $F_{21}(1)\neq 0$, that is, if X_t is not long-run exogenous, then the distribution of $S^*(0,0)$ depends on C. Theorem 1 shows that this remains true for general $S(C_T,r_T)$. The exception is when (C_T,r_T) equals (C,r), in which case the expression in part (a) simplifies because $g=b$, and S(C,r) has a χ^2 null distribution, just as the DOLS Wald statistic S(0,0) has a χ^2 null distribution when C and r are exactly zero. Of course, a-priori knowledge of the precise values of (C,r) is rarely available in practice.

The dependence of the limiting distribution of S on the nuisance parameters (C,r) is the essential feature that complicates the problem of inference on θ.

3.2. Bootstrap regions

Let $\kappa(\alpha;C,\omega)$ denote the $100\alpha\%$ critical values of the distribution of $S^*(C^*+C,r^*+r)$ for population parameters (C,ω), and let

(3.4) $$\hat{\omega}_2 = T^{1/2}\hat{H}_{11}(1)^{-1}\hat{\gamma}_1$$

where $\hat{\gamma}_1$ is the OLS estimator of γ_1 in (2.1). Consider the following four algorithms:

Algorithm B1 (parametric percentile-t bootstrap, version 1).

(1) Estimate (2.6) by OLS, construct $\hat{C}$ and $\hat{r}$ using (3.3), and let $\hat{\omega}_1 = -\hat{C}^{-1}\hat{r}$. Let $\hat{\Sigma}_{11}$ be the OLS estimator of the error variance covariance matrix in (2.6).

(2) Using $(C_T, r_T) = (\hat{C}, \hat{r})$, estimate (3.1) by OLS to obtain $\hat{\theta}(\hat{C}, \hat{r})$. Estimate an AR(p) model for the residual from this regression, with estimated innovation variance $\hat{\Sigma}_{22}$ and estimated long run variance $\hat{H}_{22}(1)^2$.

(3) Compute $S(\hat{C}, \hat{r})$.

(4) Use the estimated models from steps (1) and (2) to construct artificial samples (x_t, y_t), where ϵ_t are pseudorandom i.i.d. $N(0, \text{diag}(\hat{\Sigma}_{11}, \hat{\Sigma}_{22}))$ draws. For each realization, let $(\hat{\hat{C}}, \hat{\hat{r}})$ be the OLS estimates of (C,r), and compute $S(\hat{\hat{C}}, \hat{\hat{r}})$. Compute the $100\alpha\%$ critical value of the bootstrap distribution of $S(\hat{\hat{C}}, \hat{\hat{r}})$; denote this $\kappa_{B1}(\alpha; \hat{C}, \hat{\omega}_1)$.

Algorithm B2 (asymptotic percentile-t distribution, version 1).

(1) - (3) Same as in B1, except also compute $\hat{H}_{11}(1)$ and $\hat{F}_{21}(1)$.

(4) Evaluate the asymptotic distribution in theorem 1(b) substituting $(\hat{C}, \hat{\omega}_1, \hat{F}_{21}(1), \hat{H}_{11}(1))$ for $(C, \omega, F_{21}(1), H_{11}(1))$, and compute the $100\alpha\%$ critical value of this asymptotic distribution; denote this $\kappa_{B2}(\alpha; \hat{C}, \hat{\omega}_1)$.

Algorithm B3 (parametric percentile-t bootstrap, version 2).

(1)-(3) Same as B1, except also compute $\hat{\omega}_2$.

(4) Same as B1 except that the pseudo realizations of x_t are constructed using the estimated version of (2.1) with $V_{x,t}$ generated from the estimated VAR (2.6) with the constant and

time trend term suppresed. Compute the $100\alpha\%$ critical value of the bootstrap distribution of $S(\hat{\hat{C}},\hat{\hat{r}})$; denote this $\kappa_{B3}(\alpha;\hat{\hat{C}},\hat{\omega}_2)$.

Algorithm B2 (asymptotic percentile-t distribution, version 2).

(1)-(4) Same as B2 except that $\hat{\omega}_2$ is used in place of $\hat{\omega}_1$. Denote the critical value $\kappa_{B4}(\alpha;\hat{\hat{C}},\hat{\omega}_2)$.

Algorithm B1 is a multivariate version of what Hall (1992) calls the parametric percentile-t method for computing confidence regions. The standard nonparametric bootstrap used in the VAR literature is due to Runkle (1987). Its implementation here differs from algorithm B1 in two ways: in step (4), the bootstrap draws are constructed using resampled estimated residuals rather than pseudorandom normal variates, and the other percentile method (in Hall's [1992] terminology) based on bootstrap realizations of $\hat{\hat{\theta}}(\hat{\hat{C}},\hat{\hat{r}})$ is used to compute the confidence interval, rather than the percentile-t method (cf. Kilian [1996], Sims and Zha [1995], and Wright [1996b]).

Algorithm B2 typically would be implemented using a Monte Carlo simulation of the asymptotic distribution, evaluated at the estimated parameter values. In this sense this algorithm might not normally be considered a bootstrap.

Algorithms B3 and B4 differ from B1 and B2 in the estimated value of ω used to generate pseudo realizations of X_t and to evaluate the asymptotic critical value, respectively. They are introduced because $\hat{\omega}_1$ is a poor estimator of ω when C is close to 0, which may lead to poor performance of the bootstrap. Canjels and Watson (1997) consider the case $k=1$ and find that $\hat{\omega}_2$ is better behaved, and the numerical results in section 5 are based on this estimator.

The percentile-t bootstrap confidence region for $R\theta$ region is computed by inverting the quadratic form of the test statistic. For simplicity, consider the case $R=I_k$. Then for algorithm B1, the parametric percentile-t region with nominal coverage rate $100(1-\alpha)\%$ is,

(3.5) $\quad I_{B1} = \{\theta_0\colon\ [\hat\theta(\hat C,\hat r)-\theta_0]'\hat V(\hat C,\hat r)^{-1}[\hat\theta(\hat C,\hat r)-\theta_0] \le \kappa_{B1}(\alpha;\hat C,\hat\omega_1)\}.$

The confidence region based on algorithm B2 is computed as in (3.5), except that $\kappa_{B1}(\alpha;\hat C,\hat\omega_1)$ is replaced by $\kappa_{B2}(\alpha;\hat C,\hat\omega_1)$, and the confidence intervals for algorithms B3 and B4 are computed analogously with $\hat\omega_2$ replacing $\hat\omega_1$.

The next theorem characterizes the critical values from these algorithms in large samples.

Theorem 2. Under the conditions of theorem 1

(i) $\kappa_{B1}(\alpha;\hat C,\hat\omega_1) - \kappa_{B2}(\alpha;\hat C,\hat\omega_1) \xrightarrow{p} 0$ and $\kappa_{B2}(\alpha;\hat C,\hat\omega_1) - \kappa(\alpha;\hat C,\hat\omega_1) \xrightarrow{p} 0.$

(ii) $\kappa_{B3}(\alpha;\hat C,\hat\omega_2) - \kappa_{B4}(\alpha;\hat C,\hat\omega_2) \xrightarrow{p} 0$ and $\kappa_{B3}(\alpha;\hat C,\hat\omega_2) - \kappa(\alpha;\hat C,\hat\omega_2) \xrightarrow{p} 0.$

It is inessential for these results that pseudonormal errors be used in step (4) of B1 or B3; theorem 2 obtains under the weaker assumption that these errors are a homoskedastic martingale difference sequence with at least four moments.

A consequence of theorems 1 and 2 is that the bootstrap critical values are random variables, even asymptotically. For concreteness consider algorithm B3. Although the resulting critical values approach those based on the asymptotic distribution in theorem 1(b), evaluated at $(\hat C,\hat\omega_2)$, $\hat C$ and $\hat\omega_2$ are themselves $O_p(1)$ random variables. From theorem 1, $\hat C \Rightarrow C^*+C$ and, by a straightforward calculation, $\hat\omega_2 \Rightarrow \omega_2^*$, where $\omega_2^* = 12\int s^\mu J^\mu_{C,\omega}$, where the limits are joint. Thus the critical values have the limiting representation, $\kappa_{B3}(\alpha;\hat C,\hat\omega_2) \Rightarrow \kappa(\alpha;C^*+C,\omega_2^*)$. The asymptotic rejection rate under the null of the test based on B3 thus has the limit,

$$\begin{aligned}(3.6)\quad \text{Prob}[S(\hat C,\hat r) > \kappa_{B3}(\alpha;\hat C,\hat\omega_2)\,|\,b=0,C,\omega] &= \text{Prob}[(\kappa(\alpha;C,\omega)/\kappa_{B3}(\alpha;\hat C,\hat\omega_2))S(\hat C,\hat r) > \kappa(\alpha;C,\omega)\,|\,b=0,C,\omega] \\ &\to \text{Prob}[(\kappa(\alpha;C,\omega)/\kappa(\alpha;C^*+C,\omega_2^*))S^*(C^*+C,r^*+r) > \kappa(\alpha;C,\omega))\,|\,b=0,C,\omega]\end{aligned}$$

Because C* and ω_2^* are $O_p(1)$ random variables, the distribution of $(\kappa(\alpha;C,\omega)/\kappa(\alpha;C^*+C,\omega_2^*))S^*(C^*+C,r^*+r)$ differs from the distribution of $S^*(C^*+C,r^*+r)$, and in general $\kappa(\alpha;C,\omega)$ is not its $100\alpha\%$ critical value. Thus, for any C,ω, the limiting rejection rate in (3.6) in general is not α. The asymptotic size of the test is the supremum over (C,ω) of the limiting expression in (3.6). The magnitude of the size distortion depends on the distribution of (C^*,ω_2^*) and on the curvature of κ. In general, unless X_t is long-run exogenous, the asymptotic size of the test based on B3, or equivalently on B4, does not equal its nominal level.

Analogous arguments apply to algorithms B1 and B2, with the modification that $\hat{\omega}_1 \Rightarrow \omega_1^*$, where $\omega_1^* = -(C^*+C)^{-1}(r^*+r)$.

The failure of the bootstrap here accords with and extends Basawa et. al.'s (1991) finding that the bootstrap is invalid in the univariate unit root model.

These results were developed for the OLS estimator of (3.1). Because η_t is serially correlated, an alternative is to estimate (3.1) by GLS. It is readily shown that one-step GLS estimation of θ is asymptotically equivalent to OLS because the regressor X_t is growing stochastically (this is shown in Stock and Watson (1993) for the case $C=0$, $\omega=0$). The GLS version of the Wald statistic S also is asymptotically equivalent to the autocorrelation-consistent OLS Wald statistic considered here. Thus theorem 2 and the discussion of this section also applies to bootstrap tests and confidence regions based on the GLS Wald statistic.

3.3. Size-adjusted bootstrap regions

Although the bootstrap tests outlined above have size distortions, it is possible to compute a size adjusted bootstrap test which is asymptotically valid. To be concrete, consider the bootstrap test based on algorithm B3 with nominal level α', that is, with critical value $\kappa_{B3}(\alpha';\hat{C},\hat{\omega}_2)$. Then a size-adjusted parametric bootstrap percentile-t region based on the OLS statistic can be computed by finding the value of α' that solves,

(3.7) $$\sup_{C,\omega}\text{Prob}[S(C^*+C,r^*+r) > \kappa(\alpha';C^*+C,\omega_2^*) \mid b=0,C,\omega] = \alpha$$

The value of α' that solves (3.7) will be no greater than α and will depend on $F_{21}(1)$ and, if rank(R)$\neq$k, on $RH_{11}(1)^{-1\prime}$; it equals α in the special case that X_t is long run exogenous (so that $F_{21}(1)=0$). This produces a feasible test, which in general has rejection rate less than α for most parameter values. Equation (3.7) can be solved numerically, and values of α' for various values of the estimable and/or known nuisance parameters are tabulated in section 5.

3.4. Sup-bound regions

An alternative to the bootstrap is to use confidence regions based on an asymptotically conservative test. This was explored in the k=1, ω=0 case in Cavanagh, Elliott, and Stock (1995). The sup-bound test rejects when $S(\hat{C},\hat{r}) > \kappa_{SB}(\alpha)$, where $\kappa_{SB}(\alpha)$ solves,

(3.8) $$\sup_{C,\omega}\text{Prob}[S(C^*+C,r^*+r) > \kappa_{SB}(\alpha) \mid b=0,C,\omega] = \alpha.$$

By construction the size of the sup-bound test equals its nominal level.

An advantage of the sup-bound test is that the critical values $\kappa_{SB}(\alpha)$ can be summarized in a response surface that depends on $F_{21}(1)$ and, if rank(R)<k, on $RH_{11}(1)^{-1\prime}$. A disadvantage of this test is that it is asymptotically conservative in the sense that, for most values of C and ω, the rejection rate will be less than α.

3.5. Other Methods

Methods other than those discussed so far are available for constructing asymptotically valid confidence sets; these include Bonferroni and Scheffe methods. These are discussed by

Cavanagh, Elliott and Stock (1995) in the related problem of Granger causality tests with nearly unit roots. Looking ahead, however, we find that the relatively simple approximately optimal regions proposed in section 4 work nearly as well as the optimal regions, so these other approaches are not pursued here.

4. Efficient and Nearly Efficient Tests and Confidence Regions

This section investigates efficient tests and confidence regions for b. The first step is to obtain a set $\hat{Q}$ of maximal invariant asymptotically minimal sufficient statistics for (C,ω,b), which are derived under the assumption of i.i.d. Gaussian errors. Under Gaussianity, asymptotically efficient invariant tests of $(C,\omega,b) = (0,0,0)$ will be functions of only these statistics. The dimension of $\hat{Q}$ exceeds the dimension of (C,ω,b), and there is no uniformly most powerful invariant test of $(C,\omega,b) = (0,0,0)$. Neither is there an asymptotic pivot for testing $b=0$ by which the dependence of the distribution of $\hat{Q}$ on C and ω can be eliminated.

In the absence of analytical results on the construction of optimal tests in this context, two numerical solutions are proposed. The first is to construct tests and confidence regions that are efficient in the sense of maximizing weighted average power (or weighted average accuracy) using all statistics in $\hat{Q}$. The second approach is to construct approximately efficient confidence intervals based on a subset of the statistics in $\hat{Q}$.

4.1 Maximal Invariant Asymptotically Minimal Sufficient Statistics for (C,ω,b)

Consider the null and alternative hypotheses,

(4.1) $$H_0: B=I, \delta_1=0, \text{ and } \beta=\beta_0 \text{ vs. } H_1: B\neq I, \delta_1\neq 0, \text{ or } \beta\neq\beta_0.$$

This testing problem is invariant to the following tranformations:

(4.2) G_1: $Z_t \to Z_t + a_Z$, $w_t \to w_t + a_w$, $t \to t + a_t$

(4.3) G_2: $Z_t \to DZ_t$, $DD' = I$

(4.4) G_3: $w_t \to -w_t$

where a_Z is a $k \times 1$ vector of constants, a_w and a_t are fixed scalars, and D is a $k \times k$ matrix.

The tests commonly used for empirical cointegration analysis, for example tests of restrictions on θ based on Johansen's (1988) method, are invariant to G_1, G_2 and G_3. The transformation G_1 is a location shift; for example, invariance to G_1 means that inference does not depend on whether the data are in logarithms of thousands or millions of dollars. The transformation G_2 is designed to resolve the ambiguity associated with writing the model in the (Z_t, w_t) coordinate system. Recall that (2.8) and (2.9) were obtained by defining $Z_t = H_{11}(1)^{-1}X_t$, where $H_{11}(1)H_{11}(1)'$ is 2π times the spectral density of $u_{X,t}$ at frequency zero. Because $H_{11}(1)H_{11}(1)' = H_{11}(1)DD'H_{11}(1)'$ for D satisfying (4.3), the definition of $H_{11}(1)$ is unique only up to this orthonormal rotation, and this nonuniqueness was resolved arbitrarily by taking $H_{11}(0)$ to be lower triangular with non-negative diagonal elements. To be more concrete, invariance to G_2 means for example that in a case with two regressors, inference about β should not depend on which regressor is labeled "Z_{1t}" and which is labeled "Z_{2t}". The transformation G_3 says that inference should not depend on the sign of w_t; for example, if w_t is the (scaled) log of a real bilateral exchange rate, then inference should not depend on which country is chosen as the basis for measuring the log spread.

The invariant statistics are formed by constructing canonical versions of Z_t and w_t that are invariant to G_1, G_2 and G_3. The general strategy is to subject $\{X_t, y_t\}$ to a series of

transformations that leave them invariant under G_1, G_2 and G_3. Invariance with respect to G_1 is achieved by subtracting sample means. Achieving (maximal) invariance with respect to G_2 is somewhat more involved, and this is accomplished by first premultiplying X_{t-1} by $\hat{H}_{11}(1)^{-1}$ and $y_t-\theta_0'X_t$ by $\hat{H}_{22}(1)^{-1}$; this produces feasible counterparts of Z_t and w_t. This feasible counterpart of Z_t is premultiplied by a particular orthonormal matrix P that leaves the result invariant under G_2. Invariance with respect to G_3 is achieved by normalizing the sign of w_t so that the regression coefficient on the first canonical regressor is positive.

Let $\hat{Z}_t = \hat{P}\hat{H}_{11}(1)^{-1}X_t$, $\hat{w}_t = \hat{H}_{22}(1)^{-1}(y_t-\theta_0'X_t)$, and $\hat{\zeta}_t =$ $\text{diag}(\hat{P},1)\hat{H}(L)^{-1}[\Delta X_t' \;\; y_t-\theta_0'X_t]'$, where $\hat{P}=\hat{P}_2\hat{P}_1$, where

$$\hat{P}_1 = \text{evec}[\hat{H}_{11}(1)^{-1}T^{-2}\sum X_t^{\mu}X_t^{\mu\prime}\hat{H}_{11}(1)^{-1\prime}] \tag{4.5}$$

$$\hat{P}_2 = \text{diag}[\{\text{sgn}[\hat{P}_1\hat{H}_{11}(1)^{-1}(X_{T-p}-X_{p+1})]_i\},i=1,\ldots,k], \tag{4.6}$$

where evec(M) denotes the eigenvectors of the square matrix M. The matrix P_1 diagonalizes the regressor cross-product matrix, while P_2 normalizes the sign of these orthogonalized regressors.

The invariant statistics can be computed from the following regressions:

$$\hat{\zeta}_{1t} = \alpha_{\zeta_1} + r(t/T^{3/2}) + C(\hat{Z}_{t-1}/T) + e_{1,t} \tag{4.7}$$

$$\hat{w}_t = \alpha_{w_1} + b'(\hat{Z}_{t-1}/T) + \sum_{i=-p}^{p} d_i(\Delta\hat{Z}_t-\tilde{C}\hat{Z}_{t-1}/T-\tilde{r}t/T^{3/2}) + e_{2,t} \tag{4.8}$$

$$\hat{Z}_t = \delta_0 + \omega(t/T^{1/2}) + e_{3,t} \tag{4.9}$$

$$\hat{w}_t = \alpha_{w_2} + \xi(t/T^{3/2}) + e_{4,t} \tag{4.10}$$

where $e_{1,t},\ldots, e_{4,t}$ are error terms, and $\tilde{r}$ and $\tilde{C}$ denote the OLS estimators of r and C from (4.7). Let $\tilde{b}$ denote the OLS estimator of b in (4.8), let $\tilde{t}_b$ be the k-vector of t statistics testing $b_i=0$ in (4.8), i=1,...,k, with each element divided by $\text{sgn}(\tilde{b}_1)$, and let $S\tilde{E}_b$ denote the vector of

standard errors of $\tilde{b}$. Let $\tilde{\omega}$ be the OLS estimator of ω in (4.9). Let $\tilde{\xi}$ be the OLS estimator of ξ in (4.10), divided by $\text{sgn}(\tilde{b}_1)$. Let

$$\hat{Q} = \{\tilde{C}, \tilde{r}, \tilde{t}_b, S\tilde{E}_b, \tilde{\omega}, \tilde{\xi}\}. \tag{4.11}$$

Because $\hat{Z}_t$ has been orthogonalized, the F-statistic testing $b=0$ in (4.8) is $\tilde{F}_b = \tilde{t}_b'\tilde{t}_b$.

Theorem 3. Suppose that assumptions A-C hold and that $\{\epsilon_t\}$ is i.i.d. N(0,I). Then $\hat{Q}$ is a set of maximal invariant asymptotic minimum sufficient statistics for (C,ω,b) under G_1, G_2 and G_3.

Theorem 4. Under assumptions A-C,

(a) $[\tilde{C} \;\; \tilde{r}] \Rightarrow P^*[C^*+C \;\; r^*+r]\text{diag}(P^{*\prime},1)$,

(b) $S\tilde{E}_b \Rightarrow SE^*_b$, where $SE^*_{b,i} = (P^* \int J^{\mu}_{C,\omega} J^{\mu}_{C,\omega}' P^{*\prime})^{-\frac{1}{2}}_{i\,i}$, $i=1,\ldots,k$,

(c) $\tilde{t}_b \Rightarrow t^*_b$, where $t^*_{b,i} = b^*_i/[SE^*_{b,i}\text{sgn}(b^*_1)]$, $i=1,\ldots,k$,

(d) $\tilde{\omega} \Rightarrow 12P^* \int s^{\mu} J^{\mu}_{C,\omega}$,

(e) $\tilde{\xi} \Rightarrow 12\{\int s^{\mu} J^{\mu}_{C,\omega}'b + [-F_{21}(1) \;\; 1] \int s^{\mu} dW\}/\text{sgn}(b^*_1)$,

where C^* and r^* are defined in theorem 1, $P^*=P^*_2P^*_1$, where $P^*_1 = \text{evec}(\int J^{\mu}_{C,\omega} J^{\mu}_{C,\omega}')$, $P^*_2 = \text{diag}[\text{sgn}(P^*_1 J^{\mu}_{C,\omega}(1))]$, and $b^* = P^*\{b + (\int J^{\mu}_{C,\omega} J^{\mu}_{C,\omega}')^{-1}[\int J^{\mu}_{C,\omega} dW_2 - (\int J^{\mu}_{C,\omega} J^{\mu}_{C,\omega}')C^{*\prime}F_{21}(1)' - \int s^{\mu} J^{\mu}_{C,\omega} r^{*\prime} F_{21}(1)']\}$. These limits are joint with those in theorems 1 and 2.

Although the construction outlined here is different, $\tilde{F}_b$ is asymptotically equivalent to the Wald statistic $S(\hat{C},\hat{r})$ in section 3 when $R=I$.

4.2. Efficient Tests and Confidence Regions

The dimension of the statistics in $\hat{Q}$ exceeds the dimension of (C,ω,b), and there is no uniformly most powerful test of the null hypothesis $\beta_0=0$. Moreover, the distribution of $\hat{Q}$ depends on (C,ω,b) (as well as consistently estimable nuisance parameters). Under this local nesting, there appears to be no pivot for constructing tests of $\beta=\beta_0$, even asymptotically. This complicates significantly the development of efficient tests and confidence sets.

We therefore set out the general problem of constructing efficient tests that maximize power against a weighted average of alternatives. Let ϕ_T denote a test of $\beta=\beta_0$ against the alternative $\beta\neq\beta_0$, so that $\phi_T=1$ if and only if the null is rejected. Let π be the asymptotic rejection rate of that test:

$$(4.12) \qquad \pi(b;C,\omega,\phi) = \lim_{T\to\infty}\text{Prob}[\phi_T=1 \,|\, B=1+C/T,\delta_1=T^{-1/2}\omega,\beta=\beta_0+b/T].$$

The test of asymptotic level α that has the highest weighted average power, as determined by the weighting function $dG(b,C,\omega)$, solves:

$$(4.13) \qquad \max_{\phi} \int \pi(b;C,\omega,\phi)dG(b,C,\omega) \text{ subject to } \sup_{C,\omega}\pi(0;C,\omega,\phi)\leq\alpha.$$

Among all tests that are invariant to G_1, G_2 and G_3, the solution to (4.13) will depend only on the maximal invariant asymptotically minimal sufficient statistic $\hat{Q}$, that is, the rejection region of this test can be written,

$$(4.14) \qquad f(\hat{Q}) > 0.$$

Numerical solutions to the problem of finding the function $f(\hat{Q})$ that solves (4.13), that is, that yields the efficient invariant test, are examined in the next section.

A confidence region for β can be constructed by inverting the acceptance region of this efficient test. The confidence regions maximize the average accuracy in the sense that the test on which they are based maximize the average power in (4.13).

4.3. Approximately Efficient Tests and Confidence Regions

Because of the high dimension of $\hat{Q}$, the approach described in the previous subsection is numerically cumbersome. We therefore consider tests that are approximately efficient, in the sense that they solve (4.13) among functions ϕ that depend on a subset of $\hat{Q}$. Because the key nuisance parameters entering the limiting distribution of $\tilde{t}_b$ and $\tilde{F}_b$ are C and ω, a natural choice for the subset of Q is $\{\tilde{F}_b, \tilde{C}, \tilde{\omega}\}$. Specifically, we consider functions such that the test rejects if

$$f(\tilde{F}_b,\tilde{C},\tilde{\omega}) > 0. \tag{4.15}$$

5. Numerical Results

5.1. Model Used for Numerical Analysis

All calculations are carried out using the model:

$$Z_t = (\omega/T^{1/2})t + V_{Z,t} \tag{5.1}$$

$$\Delta V_{Z,t} = (C/T)V_{Z,t-1} + \nu_{Z,t} \tag{5.2}$$

$$w_t = (b/T)'Z_t + \nu_{w,t} \tag{5.3}$$

where $(\nu_{Z,t}' \ \nu_{w,t})'$ is NIID with mean zero, $E(\nu_{Z,t}\nu_{Z,t}')=I$, $E(\nu_{w,t}\nu_{Z,t}')=F_{21}(1)$, $E(\nu_{w,t}^2)=1+F_{21}(1)F_{21}(1)'$, and $V_{Z,0}=0$. The intercepts are set to zero in (5.1) and (5.3) without

loss of generality because all the statistics considered are invariant to the location shift G_1. Because of computational constraints, the discussion focuses primarily on the model with k=1, although at the end of this section size adjustments for the bootstrap are given for k=2. Because the Wald statistic for testing b=0 in (5.3) is invariant to the sign of the regressor, size adjustments for the bootstrap and sup-bound critical values depends only on $|F_{21}(1)|$, and we parameterize this dependence using the parameter $\lambda = F_{21}(1)^2/(1+F_{21}(1)^2)$, which is the squared correlation between $\nu_{Z,t}$ and $\nu_{w,t}$.

5.2. Size Adjustments for the Bootstrap and Sup-bound Critical Values

The asymptotic rejection rates under the null for algorithm B3/B4 bootstrap tests with a nominal size of 10% are plotted in Figure 1 for $\lambda=0.5$ as a function of C and ω. (The results in this section are approximations based on 4000 draws of data from (5.1)-(5.3) with T=400.) Evidently there is a large size distortion, and the actual size of the test (the peak of this function) is 24%. Panel A of Table 1 shows the sizes for other value of values of λ, for bootstrap tests with nominal sizes of 5% and 10%. The size distortions are large, even for values of λ as low as 0.25.

Panel B of Table 1 presents the size adjustments -- the values of the nominal size for the bootstrap -- that yield tests with actual 5% and 10% sizes, for five different values of λ. For example, when $\lambda=.50$ a nominal size of 3% will yield an actual size of 10%, and to achieve a size of 5%, the bootstrap must be carried out with a nominal size of 1.3%.

Figure 2 shows 10% critical values for $S^*(C^*+C,r^*+r)$ when $\lambda=0.50$ as a function of C and ω. For extreme values of C and ω these approach the χ^2_1 critical value of 2.7. However, the critical value is much larger for values of (C,ω) near (0,0). The sup-bound critical value is the peak of this function, 5.46 in this case, and occurs near C=1 and $\omega=0$. Table 2 shows the 5% and 10% sup-bound critical values for a range of values of λ.

5.3. Efficient and Nearly Efficient Tests

As discussed in section 4, the efficient invariant tests that solve (4.13) have critical regions determined by $f(\hat{Q})>0$. We approximated $f(\hat{Q})$ by a polynomial expansion, modified so that standard χ^2 critical regions obtain when $|\tilde{C}|$ or $|\tilde{\omega}|$ are large. Let $\tilde{F}_b = \tilde{t}_b'\tilde{t}_b$. The specific approximation used is,

$$f^a(\hat{Q}) = \tilde{F}_b - \chi^2_{1,\alpha} - [f_o + \hat{Q}'f_1 + \hat{Q}'f_2\hat{Q}]\times\exp(m_c\tilde{C}^2 + m_\omega\tilde{\omega}^2) \tag{5.4}$$

where $\chi^2_{1,\alpha}$ denotes the 1-α'th quantile of the χ^2_1 distribution, and $(f_0, f_1, f_2, m_c, m_\omega)$ are parameters. The scale of the statistic $f(\hat{Q})$ is fixed by setting $f_{2,11}=0$. When $m_c=m_\omega=0$, $f^a(\hat{Q})$ is just a quadratic approximation of $f(\hat{Q})$. When $m_c<0$ and $m_\omega<0$, the critical regions approach $\tilde{F}_b>\chi^2_{1,\alpha}$ as $|\tilde{C}|$ or $|\tilde{\omega}|$ approach ∞. The parameters were chosen to solve,

$$\max \int \pi(b;C,\omega)dG(b,C,\omega), \text{ subject to } \sup_{C,\omega}\pi(0;C,\omega)\le\alpha \tag{5.5}$$

where $\pi(b;C,\omega)=\text{Prob}[f^a(\hat{Q})>0 \mid C,\omega,b]$.

The weighting function $dG(b,C,\omega)$ was chosen as uniform on $-30\le b\le 30$, $-25\le C\le 5$, and $0\le\omega\le 10$, where the restriction to positive values of ω is without of generality because the test is invariant the sign of the regressors and the weighting function is symmetric in b. The solution to (5.5) was approximated as follows. First $G(b,C,\omega)$ was approximated by a step function using 20 values of b, 16 values of C and 7 values of ω. Then for each of the 2240 grid points of (b,C,ω), independent draws of $\hat{Q}$ were constructed from (5.1)-(5.3) using $T=400$. For each value of the parameter vector, f_0 was chosen to satisfy the sample analogue of the size constraint constructed from 1000 draws of $\hat{Q}$ for each value of (C,ω). The sample analogue of $\int \pi(b;C,\omega)dG(b,C,\omega)$ was then constructed using 100 draws of $\hat{Q}$ for each value of (b,C,ω). (Fewer draws were necessary

to evaluate the average power function because of the additional averaging.) This function evaluation algorithm together with a simulated annealing algorithm was used to determine an approximate solution to (5.5). At the resulting optimized parameter values, a more accurate size adjustment was then calculated by choosing f_0 to satisfy the sample analogue of the size constraint constructed from 4000 independent draws of $\hat{Q}$ for each value of (C,ω) in the descrete grid. The test functions $f^a(\hat{Q})$ were calculated for two different sizes (5% and 10%) and four values of λ (0.25, 0.50, 0.75 and 1.0). The resulting power function depends only on $|\lambda|$ (equivalently, $|F_{21}(1)|$) because of the invariance of the statistics with respect to sign changes in Z_t and the symmetry of $G(b,C,\omega)$ in b.

Analogous calculations were performed using a restricted set of statistics $\hat{Q}^R=(\tilde{F}_b, \tilde{C}, \tilde{\omega})$. These will be denoted as the restricted optimal tests.

5.4. Comparison of the procedures when k=1

Figure 3 compares the power of the various tests with size of 5% (equivalently, the accuracy of confidence intervals with confidence level 95%) when $\lambda=0.5$. Panels (a) and (b) show power as a function of b for two specific values of (C,ω). Panel (c) shows power as a function of b averaged over C and ω using the uniform weighting function used to solve (5.5). Panel (d) shows averages of the values in panel (c) over both positive and negative values of b. Panels (a)-(c) were calculated under the assumption that $F_{21}(1)$ was positive. Changing the sign of $F_{21}(1)$ results in a change in the sign of b, so that results for negative values of $F_{21}(1)$ are mirror images of the results shown.

Panel (a) shows that when $C=0$ and $\omega=0$ (as in the benchmark cointegrated regression) the sup-bound and bootstrap tests are badly biased. This bias arises because $\hat{b}$ is negatively biased when $F_{21}(1)>0$ (as is the estimated regression coefficient in the unit-root AR(1) regression). The optimal tests essentially eliminates this bias and have power functions more symmetric around 0.

Similar results obtain in panel (b), when C=-10 and $\omega=1$, and in panel (c), which shows power averaged over all values of C and ω. Evidently there is little loss in power in moving from the fully optimal tests to the restricted optimal tests.

Panels (a) and (b) show that the optimal tests do not uniformly dominate the other procedures over all values of b. However, from panel (c), this is nearly the case for the power averaged over C and ω: there are small power differences across the procedures for positive values of b and there are large power gains for the optimal procedures when b is negative.

The use of invariant testing procedures suggests indifference about the sign of b, and panel (d) compares the power of the procedures averaged over C, ω and positive and negative values of b. (These are averages of the results in panel (c) over $\pm$b.) One way to compare the powers of the different procedures is to use panel (d) to compare the values of b where each procedure achieves a given power, say 50%. The ratio of these values of b corresponds to the ratio of sample sizes necessary to achieve 50% power (again, averaged over both positive and negative values of β-β_0 and all C and ω), which is the asymptotic relative efficiency of the test at 50% power.

Values of these ratios are shown in Table 3, where efficiencies are shown relative to the optimal test. Results are shown for 5% and 10% tests, for different values of λ, and for powers of 50% and 80%. In many cases, the efficiency losses for the bootstrap and sup-bound procedures are large. In most cases, the restricted optimal test performs nearly as well as the optimal test. In all cases, the restricted optimal test is more efficient asymptotically than the bootstrap or sup-bound tests, often substantially so.

5.5. Extensions to k=2

As in the case with k=1, confidence regions in higher dimensions are calculated by inverting tests of the hypothesis Rb=0 using the model (5.1)-(5.3). The computational problems associated with solving the optimal problem (5.5) when k>1 are demanding, and our discussion is limited to the bootstrap and sup-bound procedures.

To begin, it is useful to simplify the dependence of the confidence region on $F_{21}(1)$, which is now a 1×2 vector. When rank(R)=2, the Wald statistic used for both the bootstrap and sup-bound procedures is invariant to transformations of the form $Z_t \rightarrow DZ_t$ with D non-singular (group G_2). Thus we can choose D without loss of generality so that $DD'=I$ (so that the errors in (5.1) still have an identity covariance matrix), but $F_{21}(1)=[F_{21,1}\ 0]$. As in the model with k=1, the dependence of the confidence region on $F_{21}(1)$ can then be parameterized as $\lambda=F_{21,1}^2/[1+F_{21}(1)'F_{21}(1)]$, the R^2 from the regression of $\nu_{w,t}$ onto $\nu_{Z,t}$. When rank(R)=1, the problem is more complicated. Begin by transforming $Z_t \rightarrow DZ_t$, with $DD'=I$ so that the resulting $R=[1\ 0]$. The Wald statistic is now invariant to multiplying each of the columns of Z_t by ±1, and so we can, without of generality, assume $F_{21}(1)$ has non-negative elements. The dependence of the confidence region on $F_{21}(1)$ can now be characterized by $\lambda_1=F_{21,1}^2/[1+F_{21}(1)'F_{21}(1)]$ and $\lambda_2=F_{21,2}^2/[1+F_{21}(1)'F_{21}(1)]$. Note, λ_1 is the squared correlation between $\nu_{w,t}$ and the first element of $\nu_{Z,t}$ (in the transformed model with $R=[1\ 0]$), λ_2 is the squared correlation between $\nu_{w,t}$ and the second element of $\nu_{Z,t}$ in this model, and $\lambda_1+\lambda_2$ is the R^2 from the regression of $\nu_{w,t}$ onto $\nu_{Z,t}$. With the dependence of the distribution on $F_{21}(1)$ parameterized by λ, λ_1 and λ_2 we now turn to a discussion of the sup-bound and bootstrap procedures.

Sup-bound critical values are very large for this model with k=2. This arises because the critical value surface for the Wald statistics, as a function of the elements of C, has a steep upward sloping ridge along $C_{11}=C_{22}$, for $C_{12}=C_{21}=0$ with $\omega=0$. This ridge has a peak near $C_{11}=C_{22}=40$ where the 10% critical value is approximately 70. (This can be compared to the χ_2^2 critical value of 4.61). The case $C_{11}=C_{22}=40$ is extremely explosive and is of little or no practical interest for economic time series. Thus, inference based on sup-bound critical values will be extremely conservative for values of C of practical interest, and this will lead to poor test power and poor accuracy of confidence intervals. For this reason, detailed values of the sup-bound critical values are not tabulated.

A modification of the bootstrap procedure protects it from this drawback. Essentially, larger critical values need to be used when it is likely that the data were generated by a model with C diagonal and C_{11} and C_{22} large. There are many ways to do this; the procedure used here was to calculate the maximum eigenvalue of $\hat{C}$, and to use the standard bootstrap critical value if maxeval($\hat{C}$)$\leq$2, and to use the critical value from the C=diag[maxeval($\hat{C}$)] distribution otherwise. Since maxeval($\hat{C}$) is only rarely greater than 2 for models with unit or stationary roots, this has little effect on the properties of the bootstrap for these models, which are the cases of primary practical interest.

Table 4 summarizes the size adjustments necessary for this modified bootstrap (algorithms B3/B4). Panel A shows the required adjustments when rank(R)=2, so that adjustments depend only of λ. Panel B shows the results when rank(R)=1, so that the adjustments depend on λ_1 and λ_2. These adjustments can be quite large. For example, when rank(R)=1 and $\lambda_1=\lambda_2=.25$, the bootstrap must be carried out using a nominal size of 0.5% to achieve an actual size of 5%.

6. Long-Run Money Demand

There is a large literature on long-run money demand functions in OECD countries. A typical specification relates real balances to log income and an interest rate, and the coefficients to be estimated are the long run income elastiticy β_y and interest semielasticity β_R:

$$(6.1) \qquad m_t\text{-}p_t = \alpha + \beta_y y_t + \beta_R R_t + u_t$$

where $m_t\text{-}p_t$ denotes the logarithm of real balances, y_t denotes the logarithm of aggregate output and R_t is the nominal short term interest rate. In this section we examine estimates of this system

based on 90 years of annual data for the United States. The data are M1, real net national product, the net national product price deflator, and the commercial paper rate, 1900-1989; see Stock and Watson (1993) for a definitions and data sources.

Unit root tests fail to reject the null hypothesis of two unit roots in the joint (y_t, R_t) process and in the joint (m_t-p_t, y_t, R_t) process. These results suggest that (6.1) might be viewed as a cointegration relation, so that β_y and β_R can be estimated using an efficient cointegration estimator. Results for this exercise using $\hat{\theta}(0,0)$ and S(0,0), that is, the DOLS estimator and confidence regions, are shown in the first row of Table 5, panel A. The estimated income elasticity is close to unity and the interest rate semi-elasticity is -.10. The corresponding confidence intervals suggest that the coefficients are estimated reasonably precisely.

However, these results are of dubious validity because they are predicated on two exact unit roots in the joint income and interest rate process. The VAR(1) for y_t and R_t ((2.6) with one lag estimated by OLS) is shown in Panel B of the figure. Also shown are the derived estimates $\hat{A}$ $(=I+\hat{\Psi}_1)$, $\hat{C}$ (eq. (3.3a)), $\hat{r}$ (eq. (3.3b)), and $\hat{\omega}_2$ (eq. (3.4)). $\hat{A}$ has two roots of 0.86, and the elements of $\hat{C}$ matrix are large and negative. While unit root pre-tests do not reject the null hypothesis that both of the roots are unity (equivalently that C=0), neither do they reject many other values for the roots that imply large negative values of C. This reinforces the a-priori concern that the confidence intervals constructed from the efficient cointegrating estimates and their standard errors are unreliable.

The second row of Panel A shows the results from estimating the coefficients without imposing unit roots in the system. The estimated income elasticity is changed slightly, but the estimated interest rate semielasticity is the same to three digits. Also shown are confidence intervals constructed using the bootstrap, with size adjustments interpolated from Table 4. Not surprisingly, these confidence intervals are somewhat wider than the DOLS intervals, because they do not use the extra (possibly incorrect) information that C=0. However, for these data, there is

only a small increase in the width of the intervals. One reason for this is that the estimated value of λ is small (.03) so that there little loss of power from using the size adjusted bootstrap.

Confidence ellipses constructed using both DOLS estimators are presented in figure 4. The modified, size-adjusted bootstrap confidence elipse is larger than the DOLS elipse and is centered at a slightly different point. From a substantive standpoint, however, the size-adjusted bootstrap elipse still produces fairly precise inference. Some have argued, for example, that an income elasticity of one-half is consistent with these or similar data (cf. Baba, Hendry and Starr (1992)), but this value is clearly inconsistent with the evidence reported here.

7. Conclusions

With the exception of the bootstrap, this paper has not dwelled on demonstrating the invalidity of standard methods for construction of confidence intervals in this model. Still, it is worth noting that other methods which might initially appear valid for this problem are not. An example is confidence intervals based on a two step procedure, where the initial step is a pretest for an exact unit root and, based on the pretest, either an I(0) or I(1) (cointegration) strategy is adopted. If the size of the pretest is fixed and there is a root local to unity, then this procedure will randomly choose between these two incorrect models, with the probability of choosing I(0) equal to the local asymptotic power of the pretest. If the size tends to zero suitably as the number of observations increases, so that the pretest is consistent, then this method will asymptotically choose the incorrect I(1) model. In either case, the coverage rates of the resulting confidence sets will in general be distorted. Similar remarks apply to procedures based on consistent model selection algorithms that choose between I(0) and I(1) or cointegration specifications.

In contrast, the size-adjusted bootstrap, sup-bound, approximately optimal and optimal methods for construction of confidence sets developed here are all asymptotically valid when there

are roots nearly but not necessarily exactly unity, and when there might be a time trend in the regressor process. All these procedures achieve the optimal rate: the resulting confidence sets are $O_p(T^{-1})$. Considerable differences were found in the asymptotic accuracy of the various confidence sets. Although the bootstrap and sup-bound methods can produce considerable losses in asymptotic accuracy, the approximately optimal procedure was found to work nearly as well as the optimal procedure.

Appendix

Proof of theorem 1

(a) By standard rate arguments and the algebra of OLS,

$$T[\hat{\theta}(C_T,r_T)-\theta] = (T^{-2}\sum X_t^{\mu}X_t^{\mu\prime})^{-1}T^{-1}\sum X_t^{\mu}\tilde{\eta}_t^{\mu} + o_p(1)$$

where $\tilde{\eta}_t^{\mu} = \eta_t^{\mu} + [d(L)(\tilde{\rho}_X-\rho_X)t + d(L)(\tilde{A}-A)X_{t-1}]^{\mu}$. Neglecting terms that are asymptotically negligible and changing to the coordinates of the transformed model using $X_t^{\mu} = H_{11}(1)Z_t^{\mu}$, $C_T-C = H_{11}(1)^{-1}T(\tilde{A}-A)H_{11}(1)$, $r_T-r=T^{3/2}H_{11}(1)^{-1}(\tilde{\rho}_X-\rho_X)$, and $d(1)=H_{21}(1)H_{11}(1)^{-1}$, we have,

$$\begin{aligned}
T[\hat{\theta}(C_T,r_T)-\theta_0] &= H_{11}(1)^{-1\prime}(T^{-2}\sum Z_t^{\mu}Z_t^{\mu\prime})^{-1} \\
&\quad\times\{T^{-1}\sum Z_t^{\mu}\epsilon_{2t}H_{22}(1) + T^{-3/2}\sum Z_t^{\mu}(t/T)^{\mu}(r_T-r)H_{21}(1)' \\
&\quad + T^{-2}\sum Z_t^{\mu}Z_t^{\mu\prime}(C_T-C)'H_{21}(1)'\} + H_{11}(1)^{-1\prime}bH_{22}(1) + o_p(1) \\
&\Rightarrow H_{11}(1)^{-1\prime}(\int J_{C,\omega}^{\mu}J_{C,\omega}^{\mu\prime})^{-1}\{\int J_{C,\omega}^{\mu}dW_2 \\
&\quad - [\int J_{C,\omega}^{\mu}s^{\mu}(\bar{r}-r)' + \int J_{C,\omega}^{\mu}J_{C,\omega}^{\mu\prime}(\bar{C}-C)']F_{21}(1)\}H_{22}(1) + H_{11}(1)^{-1\prime}bH_{22}(1)
\end{aligned}$$

where $F_{21}(1) = -H_{21}(1)/H_{22}(1)$. Now, $T^2M^{\theta\theta} = (T^{-2}\sum X_t^{\mu}X_t^{\mu\prime})^{-1} + o_p(1) = H_{11}(1)^{-1\prime}(T^{-2}\sum Z_t^{\mu}Z_t^{\mu\prime})^{-1}H_{11}(1)^{-1} + o_p(1) \Rightarrow H_{11}(1)^{-1\prime}(\int J_{C,\omega}^{\mu}J_{C,\omega}^{\mu\prime})^{-1}H_{11}(1)^{-1}$. Also $\hat{H}_{22}(1)^2 \xrightarrow{p} H_{22}(1)^2$ by assumption C. Define $K=RH_{11}(1)^{-1\prime}$. Then $S(C_T,r_T) \Rightarrow D_1'D_2^{-1}D_1$, where $D_1 = K[\int J_{C,\omega}^{\mu}J_{C,\omega}^{\mu\prime})^{-1}\int J_{C,\omega}^{\mu}dW_2 + g]$, where g is defined in the statement of the theorem, and $D_2 = K(\int J_{C,\omega}^{\mu}J_{C,\omega}^{\mu\prime})^{-1}K'$. The conditional noncentral χ^2 limiting distribution follows from noting that $J_{C,\omega}^{\mu}$ is independent of W_2 (because W_1 and W_2 are independent).

(b) Define $X_t^{\dagger} = [T^{-1/2}X_t^{\mu\prime}\ (t/T)^{\mu}]'$ and $Z_t^{\dagger} = [T^{-1/2}Z_t^{\mu\prime}\ (t/T)^{\mu}]'$, and note that $X_t^{\dagger} = \mathrm{diag}(H_{11}(1),1)Z_t^{\dagger}$. By standard rate arguments and the algebra of OLS,

$$[T(\hat{\Psi}_1-\Psi_1) \quad T^{3/2}(\hat{\Psi}_2-\Psi_2)] = (T^{-1/2}\sum H_{11}(0)\epsilon_{1t}X_{t-1}^{\dagger\prime})(T^{-1}\sum X_{t-1}^{\dagger}X_{t-1}^{\dagger\prime})^{-1} + o_p(1).$$

Also note that $\hat{H}_{11}(0)^{-1}T\Psi_1\hat{H}_{11}(1) = C + o_p(1)$ and $T^{3/2}\hat{H}_{11}(0)^{-1}\Psi_2 = r + o_p(1)$. Thus, using the definitions in (3.3), we have,

$$[\hat{C}\text{-}C \;\; \hat{r}\text{-}r] = (T^{-1/2}\sum \epsilon_{1t}Z_{t-1}^{\dagger})(T^{-1}\sum Z_{t-1}^{\dagger}Z_{t-1}^{\dagger\prime})^{-1} + o_p(1)$$
$$\Rightarrow (\int J_{C,\omega}^{\dagger}dW_1')'(\int J_{C,\omega}^{\dagger}J_{C,\omega}^{\dagger\prime})^{-1}.$$

The limit $S(\hat{C},\hat{r}) \Rightarrow S^*(C^*+C,r^*+r)$ follows because the previous limits in this theorem are all joint. □

Proof of theorem 2

By construction the pseudorandom data from algorithm B1 satisfy the conditions of theorem 1, where the population values of C and ω are the empirical estimates of $\hat{C}$ and $\hat{\omega}_1$. It follows that, conditional on $(\hat{C},\hat{\omega}_1)$, $S(\hat{\hat{C}},\hat{\hat{r}}) \Rightarrow S^*(C^*+\hat{C},r^*+\hat{r})$, with critical values $\kappa(\alpha;\hat{C},\hat{\omega}_1)$ (where $\hat{C}$ and $\hat{\omega}_1$ are treated as fixed). Thus $\kappa_{B1}(\alpha;\hat{C},\hat{\omega}_1)\text{-}\kappa(\alpha;\hat{C},\hat{\omega}_1) \xrightarrow{p} 0$. The same argument applies to $\kappa_{B2}(\alpha;\hat{C},\hat{\omega}_1)$. The proof for $\kappa_{B3}(\alpha;\hat{C},\hat{\omega}_2)$ and $\kappa_{B4}(\alpha;\hat{C},\hat{\omega}_2)$ follows the same argument.

Proof of Theorem 3.

The maximal invariant asymptotically minimum sufficient statistics $\hat{Q}$ are derived in six steps. Let $\Gamma = \{\gamma_0, \alpha_y, H(L)\}$.

(a) A set of statistics Q_1 is obtained which is asymptotically minimum sufficient for (C,ω,b) in the case that Γ is known.

(b) The invariance group G_1 is applied to reduce Q_1 to the maximal invariant Q_2.

(c) The invariance group G_2 is applied to reduce Q_2 to the maximal invariant Q_3.

(d) When Γ is unknown, Q_3 is infeasible. A feasible set of statistics $\hat{Q}_3$ is proposed and it is shown that $\hat{Q}_3 - Q_3 \xrightarrow{p} 0$.

(e) $\hat{Q}_3$ is reexpressed in a more convenient form, $\hat{Q}_4$.

(f) It is shown that $\hat{Q}$ is a maximal invariant of $\hat{Q}_4$ under G_3. It follows that $\hat{Q}$ is a set of feasible statistics that is maximal invariant (under G_1, G_2 and G_3) asymptotically minimum sufficient for (C,ω,b).

(a) Let $\mathcal{L}$ denote the Gaussian log-likelihood, parameterized by $(B,\delta_1,\beta,\Gamma)$. Now

$$\begin{aligned}\mathcal{L}(Z_1,\ldots,Z_T,w_1,\ldots,w_T;B,\delta_1,\beta,\Gamma) &= \mathcal{L}(Z_1,\ldots,Z_p) + \mathcal{L}(Z_{p+1},\ldots,Z_{T-p}|Z_1,\ldots,Z_p)\\ &+ \mathcal{L}(Z_{T-p+1},\ldots,Z_T|Z_1,\ldots,Z_{T-p}) + \mathcal{L}(w_1,\ldots,w_p|Z_1,\ldots,Z_T)\\ &+ \mathcal{L}(w_{p+1},\ldots,w_{T-p}|w_1,\ldots,w_p,w_{T-p+1},\ldots,w_T,Z_1,\ldots,Z_T)\\ &+ \mathcal{L}(w_{T-p+1},\ldots,w_T|w_1,\ldots,w_p,Z_1,\ldots,Z_T)\end{aligned}$$

and the log-likelihood ratio statistic of the test of the null $B=I$, $\delta_1=0$, $\beta=\beta_0$ is

$$LR(B,\delta_1,\beta;\Gamma) = 2[\mathcal{L}(Z_1,\ldots,Z_T,w_1,\ldots,w_T;B,\delta_1,\beta,\Gamma) - \mathcal{L}(Z_1,\ldots,Z_T,w_1,\ldots,w_T;I,0,\beta_0,\Gamma)].$$

This can be written in terms of the difference of the six components in the expression for the log likelihood under the null and alternative. Of these, all but the second and fifth term involve p observations on terms with $O_p(1)$ stochastic components. Because of the local parameterization, it can be shown that the differences between the relevant quadratic forms for these four terms are $o_p(1)$. In addition, when Γ is known, the remaining two terms in the likelihood ratio have simple forms. Let $\zeta_t = H(L)^{-1}[\Delta X_t' \ (y_t-\theta_0'X_t-\alpha_y)]' = F(L)[\Delta Z_t' \ (w_t-\alpha_w)]'$, so that under (2.8) and (2.9), $\zeta_t = \mu_t+\epsilon_t$, where $\mu_t = F(L)[(\alpha_Z+\rho_Z t+(B-I)Z_{t-1})' \ (\beta-\beta_0)'Z_t]'$. Thus

(A.1) $LR(B,\delta_1,\beta;\Gamma) = tr\{\sum[(\zeta_{1t}-\mu_{1t})(\zeta_{1t}-\mu_{1t})'-\zeta_{1t}\zeta_{1t}']\} + \sum[(\zeta_{2t}-\mu_{2t})(\zeta_{2t}-\mu_{2t})'-\zeta_{2t}\zeta_{2t}'] + o_p(1).$

Recall that $F_{11}(1)=I$ and $F_{22}(1)=1$, and define $F^*(L) = (I-L)^{-1}[F(L)-F(1)]$. Let $F^*(L)$ be partitioned conformably with $F(L)$. Now

(A.2a) $\mu_{1t} = T^{-1/2}\omega + T^{-1/2}r(t/T) + T^{-1/2}C(Z_{t-1}/T^{1/2}) + T^{-1}F^*_{11}(L)C\Delta Z_{t-1} + O(T^{-1}),$

(A.2b) $\mu_{2t} = T^{-1/2}F_{21}(1)\omega + T^{-1/2}F_{21}(1)r(t/T) + T^{-1/2}[F_{21}(1)C+b'](Z_{t-1}/T^{1/2})$
$+ T^{-1}[F^*_{21}(L)LC+(1+F^*_{22}(L))b']\Delta Z_t + O(T^{-1}),$

where the $O(T^{-1})$ terms are uniform in t. Tedious calculations now show that

$$\sum[(\zeta_{1t}-\mu_{1t})(\zeta_{1t}-\mu_{1t})'-\zeta_{1t}\zeta_{1t}'] = -[T^{-1/2}\sum\zeta_{1t}]\omega' - [T^{-1/2}(t/T)\zeta_{1t}]r' - [T^{-1}\sum\zeta_{1t}Z_{t-1}']C'$$
$$+ r[T^{-3/2}\sum(t/T)Z_{t-1}]'C' + C[T^{-2}\sum Z_{t-1}Z_{t-1}']C' + \omega[T^{-3/2}\sum Z_{t-1}']C'$$
$$+ \text{transposes of some of these terms} + \text{constant} + o_p(1).$$

Similar terms obtain for $\sum[(\zeta_{2t}-\mu_{2t})(\zeta_{2t}-\mu_{2t})'-\zeta_{2t}\zeta_{2t}']$, except that the coefficients are different and ζ_{1t} is replaced by ζ_{2t}. Collecting the $O_p(1)$ statistics, this yields the asymptotically minimum sufficient statistic,

$$\bar{Q}_1 = \{T^{-2}\sum Z_{t-1}Z_{t-1}', T^{-1}\sum Z_{t-1}\zeta_{1t}', T^{-1}\sum Z_{t-1}\zeta_{2t}', T^{-3/2}\sum(t/T)Z_{t-1},$$
$$T^{-1/2}\sum(t/T)\zeta_{1t}, T^{-1/2}\sum(t/T)\zeta_{2t}, T^{-1/2}\sum\zeta_{1t}, T^{-1/2}\sum\zeta_{2t}, T^{-3/2}\sum Z_{t-1}\}.$$

This can be reduced by recalling that $\zeta_{1t} = F_{11}(L)\Delta Z_t$ so that $T^{-1/2}\sum\zeta_{1t} = T^{-1/2}(Z_{T-p}-Z_{p+1}) + o_p(1)$. However, $(T^{-1}\sum Z_{t-1}\zeta_{1t}')_{ii} = \frac{1}{2}[T^{-1/2}(Z_{T-p}-Z_{p+1})_i]^2 + \text{constant} + o_p(1)$. Thus $T^{-1/2}\sum\zeta_{1t}$ is redundant up

to its sign, so it can be replaced by $\text{sgn}(Z_{T-p}-Z_{p+1})$. This yields the asymptotically minimum sufficient statistic,

$$Q_1 = \{T^{-2}\sum Z_{t-1}Z_{t-1}',\ T^{-1}\sum Z_{t-1}\zeta_{1t}',\ T^{-1}\sum Z_{t-1}\zeta_{2t}',\ T^{-3/2}\sum (t/T)Z_{t-1},$$
$$T^{-1/2}\sum (t/T)\zeta_{1t},\ T^{-1/2}\sum (t/T)\zeta_{2t},\ \text{sgn}(Z_{T-p}-Z_{p+1}),\ T^{-1/2}\sum \zeta_{2t},\ T^{-3/2}\sum Z_{t-1}\}.$$

(b) Under a location shift in Z_t, $Z_t \to Z_t+a_Z$, the maximal invariant of $\{Z_t\}$ is $\{Z_t-\hat{\gamma}_{0,GLS}\}$, where $\hat{\gamma}_{0,GLS}$ is the GLS estimator of γ_0. However, $\{Z_t-\hat{\gamma}_{0,GLS},\ w_t\}$ is not invariant to $w_t \to w_t+a_w$, and a maximal invariant under $w_t \to w_t+a_w$ is, by standard arguments, $\{Z_t^{\mu},\ w_t^{\mu}\}$. This is also invariant to $Z_t \to Z_t+a_Z$. Thus, under G_1 a maximal invariant of $\{Z_t,\ w_t,\ t\}$, $t = p+1,\ldots,T-p$ is $\{Z_t^{\mu},\ w_t^{\mu},\ t^{\mu}\}$. It follows that Q_2 is maximal invariant asymptotically minimum sufficient for (C,ω,b) under G_1, where

$$Q_2 = \{T^{-2}\sum Z_{t-1}^{\mu}Z_{t-1}^{\mu}{}',\ T^{-1}\sum Z_{t-1}^{\mu}\zeta_{1t}^{\mu}{}',\ T^{-1}\sum Z_{t-1}^{\mu}\zeta_{2t}^{\mu},\ T^{-3/2}\sum (t/T)^{\mu}Z_{t-1}^{\mu},$$
$$T^{-1/2}\sum (t/T)^{\mu}\zeta_{1t}^{\mu},\ T^{-1/2}\sum (t/T)^{\mu}\zeta_{2t}^{\mu},\ \text{sgn}(Z_{T-p}-Z_{p+1})\}$$

(c) Let the eigenvector-eigenvalue decomposition of $T^{-2}\sum Z_{t-1}^{\mu}Z_{t-1}^{\mu}{}'$ be $T^{-2}\sum Z_{t-1}^{\mu}Z_{t-1}^{\mu}{}' = P_1'\Lambda_1P_1$, where $P_1'P_1=I$ and Λ_1 is the $k\times k$ matrix with the ordered eigenvalues on the diagonal. From Lehmann [1959, pp. 298-9], Λ_1 is a maximal invariant of $T^{-2}\sum Z_{t-1}^{\mu}Z_{t-1}^{\mu}{}'$ under G_2. Let $Z_{t-1}^{(1)} = P_1Z_{t-1}^{\mu}$.

The invariant version of the statistics in Q_2 are now computed using $\{Z_{t-1}^{(1)},\ w_t\}$. The system for $\{Z_t^{(1)},\ w_t\}$ is given by (2.7), with every term premultiplied by P_1 and with $F_{11}(L)$ replaced by $P_1F_{11}(L)P_1'$, and by (2.9). Thus ζ_{1t}^{μ} becomes $P_1F_{11}(L)P_1'(\Delta P_1Z_t)^{\mu} = P_1F_{11}(L)(\Delta Z_t)^{\mu}$. Note that ζ_{2t} is unaffected by G_2.

Further reduction is possible because $Z_t^{(1)}$ is unique only up to its sign:

$\Upsilon T^{-2}\sum Z_{t-1}^{(1)}Z_{t-1}^{(1)\prime}\Upsilon = \Lambda_1$ for any k×k diagonal matrix Υ with ± 1 on the diagonal, but $\Upsilon T^{-3/2}\sum (t/T)^{\mu}Z_{t-1}^{(1)} \neq T^{-3/2}\sum (t/T)^{\mu}Z_{t-1}^{(1)}$ unless $\Upsilon = I$. Therefore let $P_2 = \text{diag}[\{\text{sgn}(Z_{T+p}^{(1)} - Z_{p+1}^{(1)})_i\}, i=1,\ldots k]$, and let $Z_{t-1}^{(2)} = P_2P_1Z_{t-1}^{\mu}$ and $\zeta_t^{(2)} = \text{diag}(P_2P_1,1)\zeta_t^{\mu}$. Now $\{Z_{t-1}^{(2)}, \zeta_t^{(2)}\}$ are maximal invariant under G_1 and G_2.

Because $\text{sgn}(Z_{T-p}^{(2)}-Z_{p+1}^{(2)})_i = P_{2,ii}\text{sgn}(Z_{T-p}^{(1)}-Z_{p+1}^{(1)})_i = 1$, when Q_2 is reexpressed in terms of $\{Z_{t-1}^{(2)}, \zeta_t^{(2)}\}$, the sign statistic is nonrandom and may be omitted. It follows that Q_3 is the maximal invariant asymptotic minimum sufficient statistic under G_1 and G_2, where

$$Q_3 = \{T^{-2}\sum Z_{t-1}^{(2)}Z_{t-1}^{(2)\prime}, T^{-1}\sum Z_{t-1}^{(2)}\zeta_{1t}^{(2)\prime}, T^{-1}\sum Z_{t-1}^{(2)}\zeta_{2t}^{(2)\prime}, T^{-3/2}\sum (t/T)^{\mu}Z_{t-1}^{(2)},$$
$$T^{-1/2}\sum (t/T)^{\mu}\zeta_{1t}^{(2)}, T^{-1/2}\sum (t/T)^{\mu}\zeta_{2t}^{(2)}\}.$$

For future reference, observe that P_1 is a continuous functions of Q_2. Let $Q_{2,i}$ refer to the i-th statistic listed in Q_2. Then $P_1 = \text{evec}(Q_{2,1})$.

(d) The next step is to obtain an asymptotically equivalent version of Q_3 that is feasible when Γ is unknown. This is constructed in three steps.

(i) First obtain a feasible version of Q_2. Let $\hat{\hat{Z}}_t = \hat{H}_{11}(1)^{-1}X_t$, $\hat{\hat{w}}_t = \hat{H}_{22}(1)^{-1}(y_t - \theta_0'X_t)$, and $\hat{\hat{\zeta}}_t = \hat{F}(L)[\Delta\hat{\hat{Z}}_t' \ \hat{\hat{w}}_t]'$. Now $\hat{\hat{Z}}_t = \hat{H}_{11}(1)^{-1}H_{11}(1)Z_t$, $\hat{\hat{w}}_t = \hat{H}_{22}(1)^{-1}H_{22}(1)w_t$, and $\hat{\hat{\zeta}}_t^{\mu} - \zeta_t^{\mu} = \hat{\Pi}(L)[\Delta Z_t^{\mu\prime} \ w_t^{\mu}]' - \bar{\Delta}$, where $\bar{\Delta}$ depends on the data but not on t, and $\hat{\Pi}(L) = \hat{F}(L)\text{diag}[\hat{H}_{11}(1)^{-1}H_{11}(1), \hat{H}_{22}(1)^{-1}H_{22}(1)] - F(L)$. Note that, because $\hat{F}(L)$ and $\hat{H}(L)$ are $T^{1/2}$-consistent, $T^{1/2}\hat{\Pi}_i = O_p(1)$, i = -p,…,p.

Define $\hat{Q}_2$ as Q_2 with Z_t and ζ_t replaced by $\hat{\hat{Z}}_t$ and $\hat{\hat{\zeta}}_t$. To show the asymptotic equivalence of $\hat{Q}_2$ and Q_2, we show that $\hat{Q}_{2,i} - Q_{2,i} \xrightarrow{p} 0$, i=1,…,7.

$\hat{Q}_{2,1}$: $\hat{Q}_{2,1}-Q_{2,1} \xrightarrow{p} 0$ because $\hat{H}_{11}(1)^{-1}H_{11}(1)-I \xrightarrow{p} 0$.

$\hat{Q}_{2,2}$: $\hat{Q}_{2,2}-Q_{2,2} = T^{-1}\sum[\hat{Z}^{\mu}_{t-1}\hat{\zeta}^{\mu}_{1t}{}' - Z^{\mu}_{t-1}\zeta^{\mu}_{1t}{}']$

$= T^{-1}\sum(\hat{Z}^{\mu}_{t-1}-Z^{\mu}_{t-1})\zeta^{\mu}_{1t}{}' + T^{-1}\sum\hat{Z}^{\mu}_{t-1}(\hat{\zeta}^{\mu}_{1t}-\zeta^{\mu}_{1t})'$.

Now $T^{-1}\sum(\hat{Z}^{\mu}_{t-1}-Z^{\mu}_{t-1})\zeta^{\mu}_{1t}{}' = (\hat{H}_{11}(1)^{-1}H_{11}(1)-I)T^{-1}\sum Z^{\mu}_{t-1}\zeta^{\mu}_{1t}{}' \xrightarrow{p} 0$ because $\hat{H}_{11}(1)^{-1}H_{11}(1)-I \xrightarrow{p} 0$, and $T^{-1}\sum Z^{\mu}_{t-1}\zeta^{\mu}_{1t}{}' = O_p(1)$. Also,

$$T^{-1}\sum\hat{Z}^{\mu}_{t-1}(\hat{\zeta}^{\mu}_{t}-\zeta^{\mu}_{t})' = \hat{H}_{11}(1)^{-1}H_{11}(1)\sum_{i=-p}^{p}[T^{-1}\sum Z^{\mu}_{t-1}(\Delta Z^{\mu}_{t-i}{}' \;\; w^{\mu}_{t-i})]\hat{\Pi}_i.$$

Consider the i=0 term. Now

$$T^{-1}\sum Z^{\mu}_{t-1}\Delta Z^{\mu}_{t} = T^{-1/2}rT^{-1}\sum Z^{\mu}_{t-1}(t/T)^{\mu} + T^{-1/2}CT^{-3/2}\sum Z^{\mu}_{t-1}Z^{\mu}_{t-1} + T^{-1}\sum Z^{\mu}_{t-1}\nu^{\mu}_{Z,t}$$

which is $O_p(1)$ by standard arguments. Similarly, $T^{-1}\sum Z^{\mu}_{t-1}w^{\mu}_{t} = T^{-2}Z^{\mu}_{t-1}Z^{\mu}_{t-1}{}'b + T^{-1}Z^{\mu}_{t-1}\nu^{\mu}_{w,t} = O_p(1)$. Because $\hat{\Pi}_i \xrightarrow{p} 0$, i =-p,…,p, it follows that $T^{-1}\sum\hat{Z}^{\mu}_{t-1}(\hat{\zeta}^{\mu}_{1t}-\zeta^{\mu}_{1t}) \xrightarrow{p} 0$.

$\hat{Q}_{2,3}$: The argument parallels $\hat{Q}_{2,2}$.

$\hat{Q}_{2,4}$: $\hat{Q}_{2,4}-Q_{2,4} = [\hat{H}_{11}(1)^{-1}H_{11}(1)-I]T^{-3/2}\sum(t/T)^{\mu}Z^{\mu}_{t-1} \xrightarrow{p} 0$.

$\hat{Q}_{2,5}$: $\hat{Q}_{2,5}-Q_{2,5} = T^{-1/2}\sum(t/T)^{\mu}(\hat{\zeta}^{\mu}_{1t}-\zeta^{\mu}_{1t}) = \sum_{i=-p}^{p}T^{1/2}\hat{\Pi}_i\{T^{-1}\sum(t/T)^{\mu}[\Delta Z^{\mu}_{t-i}{}' \;\; w^{\mu}_{t-i}]'\}$.
Consider the i=0 term. Now, $T^{-1}\sum(t/T)^{\mu}\Delta Z^{\mu}_{t} = T^{-1/2}rT^{-1}\sum[(t/T)^{\mu}]^2 + T^{-1/2}CT^{-3/2}\sum(t/T)^{\mu}Z^{\mu}_{t-1} + T^{-1}\sum(t/T)^{\mu}\nu^{\mu}_{Z,t} \xrightarrow{p} 0$, where the limit follows by standard arguments. Similarly, $T^{-1}\sum(t/T)^{\mu}w^{\mu}_{t} \xrightarrow{p} 0$. Because $T^{1/2}\hat{\Pi}_i = O_p(1)$, $\hat{Q}_{2,5}-Q_{2,5} \xrightarrow{p} 0$.

$\hat{Q}_{2,6}$: The argument parallels $\hat{Q}_{2,5}$.

$\hat{Q}_{2,7}$: Let $\hat{\hat{\upsilon}} = \hat{\hat{Z}}_{T-p} - \hat{\hat{Z}}_{p+1}$ and $\upsilon = Z_{T-p} - Z_{p+1}$, so $\hat{Q}_{2,7} = \text{sgn}(\hat{\hat{\upsilon}})$ and $Q_{2,7} = \text{sgn}(\upsilon)$. Let $\Delta_T = T^{1/2}[\hat{H}_{11}(1)^{-1}H_{11}(1) - I]$ so $\hat{\hat{\upsilon}} - \upsilon = \Delta_T \upsilon / T^{1/2}$. The strategy is to show that $\hat{\hat{\upsilon}}_i$ and υ_i have the same sign with asymptotic probability one if $|\upsilon_i| > T^{1/4}$ and that $\text{Prob}[|\upsilon_i| \le T^{1/4}] \to 0$.

$$\begin{aligned}
\text{Prob}[\hat{Q}_{2,7i} \neq Q_{2,7i}] &= \text{Prob}[\text{sgn}(\hat{\hat{\upsilon}}_i) \neq \text{sgn}(\upsilon_i)] \\
&\le \text{Prob}[|\hat{\hat{\upsilon}}_i - \upsilon_i| > T^{1/4} \,|\, |\upsilon_i| > T^{1/4}]\text{Prob}[|\upsilon_i| > T^{1/4}] + \text{Prob}[|\upsilon_i| \le T^{1/4}] \\
&\le \text{Prob}[|\hat{\hat{\upsilon}}_i - \upsilon_i| > T^{1/4}] + \text{Prob}[|\upsilon_i| \le T^{1/4}] \\
&\le \text{Prob}[\textstyle\sum_{j=1}^{k} |\Delta_{T,ij}| \, |T^{-1/2}\upsilon_j| > T^{1/4}] + \text{Prob}[|\upsilon_i| \le T^{1/4}].
\end{aligned}$$

Now $T^{-1/2}\upsilon \Rightarrow J_{C,\omega}(1) = O_p(1)$ and $\Delta_T = O_p(1)$ by assumption C. Thus $\sum_{j=1}^{k} |\Delta_{T,ij}| \, |T^{-1/2}\upsilon_j| = O_p(1)$ and the probability this exceeds $T^{1/4}$ converges to zero. Similarly, because $J_{C,\omega}(1)$ has no point mass at zero, $\text{Prob}[|T^{-1/2}\upsilon_i| \le T^{-1/4}] \to 0$. Thus $\text{Prob}[\hat{Q}_{2,7i} \neq Q_{2,7i}] \to 0$, $i = 1,\ldots,k$.

(ii) Next construct $\hat{P}_1$ and $\hat{P}_2$ as in (4.5) and (4.6). It needs to be shown that $\hat{P}_i - P_i \xrightarrow{p} 0$.

$\hat{P}_1$: For the treatment of P_2, the stronger result that $\hat{P}_1 = P_1 + O_p(T^{-1/2})$ is used, and this is now shown. Define Δ_T as in the treatment of $\hat{Q}_{2,7}$ above. Now $\hat{Q}_{2,1} = (I + T^{-1/2}\Delta_T)Q_{2,1}(I + T^{-1/2}\Delta_T') = Q_{2,1} + O_p(T^{-1/2})$, because $\Delta_T = O_p(1)$ and $Q_{2,1} = O_p(1)$. By the continuity of the eigenvector function, it follows that $\hat{P}_1 = \text{evec}(\hat{Q}_{2,1}) = \text{evec}(Q_{2,1}) + O_p(T^{-1/2}) = P_1 + O_p(T^{-1/2})$, as was to be shown.

$\hat{P}_2$: Let $\upsilon^{(1)} = Z^{(1)}_{T-p} - Z^{(1)}_{p+1} = P_1(Z_{T-p} - Z_{p+1})$ and $\hat{\hat{\upsilon}}^{(1)} = \hat{P}_1(\hat{\hat{Z}}_{T-p} - \hat{\hat{Z}}_{p+1})$, so $P_2 = \text{diag}(\{\text{sgn}(\upsilon_i^{(1)})\})$, $\hat{P}_2 = \text{diag}(\{\text{sgn}(\hat{\hat{\upsilon}}_i^{(1)})\})$, and $\hat{\hat{\upsilon}}^{(1)} - \upsilon^{(1)} = \Delta_T^{(1)}\upsilon^{(1)}/T^{1/2}$, where

$\Delta_T^{(1)} = T^{1/2}[\hat{P}_1\hat{H}_{11}(1)^{-1}H_{11}(1)P_1'-I]$. The argument that $\hat{P}_2-P_2 \xrightarrow{p} 0$ parallels the argument that $\hat{Q}_{2,7}-Q_{2,7} \xrightarrow{p} 0$ in part (d)(i), with v replaced by $v^{(1)}$, etc. This argument requires that $\Delta_T^{(1)} = O_p(1)$, which in turn follows from Assumption C and from $\hat{P}_1=P_1+O_p(T^{-1/2})$, which was shown above.

(iii) Let $\hat{Q}_3$ be defined as Q_3, constructed with $\{Z_{t-1}^{(2)}, \xi_t^{(2)}\}$ replaced by $\{\hat{Z}_t^{\mu}, \hat{\xi}_t^{\mu}\} = \{\hat{P}\hat{Z}_t^{\mu}, \text{diag}(\hat{P},1)\hat{\xi}_t^{\mu}\}$. Note that $\hat{Q}_3$ is a continuous function of $\hat{Q}_2$ and $\hat{P}$, for example $\hat{Q}_{3,1} = \hat{P}\hat{Q}_{2,1}\hat{P}'$. Because $\hat{Q}_2-Q_2 \xrightarrow{p} 0$ and $\hat{P}-P \xrightarrow{p} 0$, $\hat{Q}_3-Q_3 \xrightarrow{p} 0$. Thus $\hat{Q}_3$ is feasible and is asymptotically equivalent to Q_3.

(e) Let

$$\hat{Q}_4 = \{\tilde{C}, \tilde{r}, \tilde{\tilde{t}}_b, S\tilde{E}_b, \tilde{\omega}, \tilde{\tilde{\xi}}\},$$

where $\tilde{\tilde{t}}_b$ is the k-vector with i-th element $\tilde{\tilde{t}}_{b,i} = \tilde{b}_i/S\tilde{E}_{b,i}$, where $\tilde{\tilde{\xi}}$ is the OLS estimator of ξ in (4.10), and where the rest of the terms are defined preceding (4.11). We now show that $\hat{Q}_4$ and $\hat{Q}_3$ are equivalent.

Let $\hat{Q}_{3,1}^{\dagger}$ be the $(k+1)\times(k+1)$ matrix with blocks $(\hat{Q}_{3,1}^{\dagger})_{11} = \hat{Q}_{3,1}$, $(\hat{Q}_{3,1}^{\dagger})_{12} = (\hat{Q}_{3,1}^{\dagger})_{21}' = \hat{Q}_{3,4}$, and $(\hat{Q}_{3,1}^{\dagger})_{22} = = 1/12$. Algebraic manipulations show that,

$$\begin{aligned}
&[\tilde{C}\ \ \tilde{r}] = \hat{Q}_{3,1}^{\dagger-1}[\hat{Q}_{3,2}\ \ \hat{Q}_{3,5}],\\
&\tilde{b} = \hat{Q}_{3,1}^{-1}\{\hat{Q}_{3,3} - \hat{Q}_{3,1}\hat{C}'d(1)' - \hat{Q}_{3,4}\hat{r}'d(1)'\} + o_p(1),\\
&\tilde{\tilde{t}}_{b,i} = \tilde{b}_i/(\hat{Q}_{3,1})_{i\,i}^{1/2},\\
&S\tilde{E}_{b,i} = (\hat{Q}_{3,1})_{i\,i}^{1/2},\\
&\tilde{\omega} = 12\hat{Q}_{3,4},\\
&\tilde{\tilde{\xi}} = 12[\hat{Q}_{3,6} - \hat{Q}_{3,5}F_{11}(1)^{-1}F_{21}(1)'] + o_p(1).
\end{aligned}$$

Evidently $\hat{Q}_4$ is a continuous, one to one function of $\hat{Q}_3$ (up to $o_p(1)$ terms; recall that $\hat{Q}_{3,1}$ is diagonal), so $\hat{Q}_4$ is asymptotically equivalent to $\hat{Q}_3$.

(f) It remains to impose invariance to G_3. Under G_3, $\tilde{\tilde{t}}_b \to -\tilde{\tilde{t}}_b$ and $\tilde{\tilde{\xi}} \to -\tilde{\tilde{\xi}}$; the remaining statistics in $\hat{Q}_4$ are already invariant to G_3. We impose the sign convention that $\tilde{t}_{b,1} \geq 0$ (equality occurs with probability zero). Let $\tilde{t}_b = \tilde{\tilde{t}}_b/\text{sgn}(\tilde{b}_1)$ and $\tilde{\xi} = \tilde{\tilde{\xi}}/\text{sgn}(\tilde{b}_1)$. Then $(\tilde{t}_b, \tilde{\xi})$ are maximal invariants of $(\tilde{\tilde{t}}_b, \tilde{\tilde{\xi}})$, and it follows that $\hat{Q}$ is a maximal invariant asymptotic minimal sufficient statistic under G_1, G_2, and G_3. □

Proof of theorem 4

Let $\bar{Q} \equiv \{\bar{C}, \bar{r}, \bar{t}_b, S\bar{E}_b, \bar{\omega}, \bar{\xi}\}$ denote the corresponding elements of $\hat{Q}$, constructed using $\{Z^{(2)}_{t-1}, \zeta^{(2)}_t\}$ instead of $\{\hat{Z}^{\mu}_{t-1}, \hat{\zeta}^{\mu}_{t-1}\}$. By the arguments of the proof of theorem 3, $\bar{Q}$-$\hat{Q} \xrightarrow{p} 0$, so it suffices to find the limiting distribution of $\bar{Q}$. Recall that $Z^{(2)}_{t-1} = PZ^{\mu}_{t-1}$ and $\zeta^{(2)}_t = \text{diag}(P,1)\zeta^{\mu}_t$.

First we show that $P \Rightarrow P^*$. Now $P_1 = \text{evec}(T^{-2}\sum Z^{\mu}_{t-1}Z^{\mu}_{t-1}{}') \Rightarrow \text{evec}(\int J^{\mu}_{C,\omega}J^{\mu}_{C,\omega}{}') \equiv P^*_1$. Also, $\text{Prob}[P_{2,i}=1] = \text{Prob}[\{P_1T^{-1/2}(Z_{T-p}-Z_{p+1})\}_i>0] \Rightarrow \text{Prob}[\{P^*_1J_{C,\omega}(1)\}_i>0]$, so $P_2 \Rightarrow \text{diag}[\text{sgn}(P^*_1J_{C,\omega}(1))] \equiv P^*_2$. Let $P^* = P^*_2P^*_1$. Now turn to the elements of $\bar{Q}$.

$[\bar{C}\ \bar{r}]$: Let $\bar{Z}^{\dagger}_{t-1} = [T^{-1/2}Z^{(2)}_{t-1}{}'\ \ (t/T)^{\mu}]' = \text{diag}(P,1)Z^{\dagger}_{t-1}$. Now,

$$
\begin{aligned}
[\bar{C}\ \bar{r}] &= P(T^{-1/2}\textstyle\sum \zeta^{\mu}_{1t}\bar{Z}^{\dagger}_{t-1}{}')(T^{-1}\sum \bar{Z}^{\dagger}_{t-1}\bar{Z}^{\dagger}_{t-1}{}')^{-1} \\
&= P(T^{-1/2}\textstyle\sum \zeta^{\mu}_{1t}Z^{\dagger}_{t-1}{}')(T^{-1}\sum Z^{\dagger}_{t-1}Z^{\dagger}_{t-1}{}')^{-1}\text{diag}(P',1) \\
&= P[CP'\ r] + P[T^{-1/2}\textstyle\sum \epsilon_{1t}Z^{\dagger}_{t-1}{}' + T^{-3/2}\sum (\Gamma^*_{11}(L)C\Delta Z_{t-1} + O(T^{-1}))Z^{\dagger}_{t-1}{}'] \\
&\qquad \times (T^{-1}\textstyle\sum Z^{\dagger}_{t-1}Z^{\dagger}_{t-1}{}')^{-1}\text{diag}(P',1) \\
&\Rightarrow P^*[CP^{*\prime}\ r] + P^*(\textstyle\int dW_1J^{\dagger}_{C,\omega}{}')\int J^{\dagger}_{C,\omega}J^{\dagger}_{C,\omega}{}')^{-1}\text{diag}(P^{*\prime},1)
\end{aligned}
$$

$$= P^*[C^* + C\ \ r^* + r]\mathrm{diag}(P^{*\prime}, 1).$$

$S\bar{E}_b$: $(T^{-2}\sum Z^{(2)}_{t-1}Z^{(2)\prime}_{t-1})^{-1} = P(T^{-2}\sum Z^{\mu}_{t-1}Z^{\mu\prime}_{t-1})^{-1}P' \Rightarrow P^*(\int J^{\mu}_{C,\omega}J^{\mu\prime}_{C,\omega})^{-1}P^{*\prime}$, which is diagonal. Thus $S\bar{E}_b \Rightarrow SE^*_b$, where $SE^*_{b,i} = (P^* \int J^{\mu}_{C,\omega}J^{\mu\prime}_{C,\omega}P^{*\prime})^{-1/2}_{i\,i}$.

$\bar{t}_b$. First consider $\bar{b}$:

$$\bar{b} = Pb + P(T^{-2}\sum Z^{\mu}_{t-1}Z^{\mu\prime}_{t-1})^{-1}[T^{-1}\sum Z^{\mu}_{t-1}\epsilon_{2t} + T^{-2}\sum Z^{\mu}_{t-1}Z^{\mu\prime}_{t-1}(C-\bar{C})'d(1)'$$
$$+ T^{-3/2}\sum Z^{\mu}_{t-1}(t/T)^{\mu}(r-\bar{r})'d(1)'] + o_p(1) \Rightarrow b^*,$$

where b^* is given in the theorem. Also, $\mathrm{sgn}(\bar{b}) \Rightarrow \mathrm{sgn}(b^*)$ jointly with the other limits. Thus $\bar{t}_b \Rightarrow t^*_b$, where t^*_b is given in the theorem.

$\bar{\omega}$: $\bar{\omega} = T^{-3/2}\sum (t/T)^{\mu}Z^{(3)}_{t-1}/T^{-1}[(t/T)^{\mu}]^2 \Rightarrow 12P^* \int s^{\mu}J^{\mu}_{C,\omega}$.

$\bar{\xi}$: Let $\tilde{\bar{\xi}} = (T^{-1}\sum (t/T)^{\mu 2})^{-1}T^{-1/2}\sum (t/T)^{\mu}w_t = 12\{T^{-3/2}\sum (t/T)^{\mu}Z^{\mu}_{t-1}'b + T^{-1/2}(t/T)^{\mu}\nu_{w,t}\} \Rightarrow$ $12\{\int s^{\mu}J^{\mu}_{C,\omega}'b + [-F_{21}(1)\ \ 1]\int s^{\mu}dW\}$. It follows that $\bar{\xi} = \tilde{\bar{\xi}}/\mathrm{sgn}(\tilde{b}_1)$ has the limiting representation given in the theorem. □

References

Baba, Y., Hendry, D.F., and Starr, R.M. 1992. "The Demand for M1 in the U.S.A., 1960-1988," *Review of Economic Studies* 59: 25-61.

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: 1098-1101.

Campbell, B. and Dufour, J.M. 1995. "Exact Nonparametric Orthogonality and Random Walk Tests," *The Review of Economics and Statistics* 78: 1-16.

Canjels, E. and Watson, M.W. 1997. "Estimating Deterministic Trends in the Presence of Serially Correlated Errors," forthcoming *Review of Economics and Statistics*.

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. and Wei, C.Z. 1987. "Asymptotic Inference For Nearly Nonstationary AR(1) Processes," *Annals of Statistics* 15: 1050-1063.

Choi, I. 1993. "Asymptotic Normality of the Least-Squares Estimates for Higher Order Autoregressive Integrated Processes with Some Applications," *Econometric Theory* 9: 263-282.

Dolado, J.J. and Lutkepohl, H. 1995. "Making Wald Tests Work for Cointegrated VAR Systems," forthcoming, *Econometric Reviews*.

Elliott, G. 1994. *Application of Local to Unity Asymptotic Theory to Time Series Regression*, unpublished Ph.D. Dissertation, Department of Economics, Harvard University.

Hafer, R.W. and Jansen, D.W. 1991. "The Demand for Money in the United States: Evidence from Cointegration Tests," *Journal of Money, Credit and Banking* 23: 155-168.

Hall, P. 1992. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

Hoffman, D.L. and Rasche, R.H. 1991. "Long Run Income and Interest Elasticities of Money Demand in the United States," *Review of Economics and Statistics* 73: 665-674.

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

Kilian, L. 1996. "Small Sample Confidence Intervals for Impulse Response Functions," forthcoming, *The Review of Economics and Statistics*.

Lehmann, E.L. 1959. *Testing Statistical Hypotheses*. New York: Wiley.

Miller, S. 1991. "Monetary Dynamics: An Application of Cointegration and Error Correction," *Journal of Money, Credit and Banking* 23: 139-154.

Phillips, P.C.B. 1987. "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika* 74: 535-547.

Phillips, P.C.B. and Loretan, M. 1991. "Estimating Long-Run Economic Equilibria," *Review of Economic Studies* 58: 407-436.

Runkle, D.E. 1987. "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics* 5: 437-42.

Saikkonen, P. 1991. "Asymptotically Efficient Estimation of Cointegrating Regressions," *Econometric Theory* 7: 1-21.

Sims, C.A. and Zha, T. 1995. "Error Bands for Impulse Responses," manuscript, Yale University.

Stock, J.H. 1996. "Cointegration, Long-Run Comovements, and Long-Horizon Forecasting," forthcoming in D. Kreps and K.F. Wallis (eds), *Advances in Econometrics: Proceedings of the Seventh World Congress of the Econometric Society*. Cambridge: Cambridge University Press.

Stock, J.H. and Watson, M.W. 1993. "A Simple Estimator of Cointegrating Vectors in Higher-Order Integrated Systems," *Econometrica* 61: 783-820.

Stock, J.H., and West, K.D. 1988. "Integrated Regressors and Tests of the Permanent Income Hypothesis," *Journal of Monetary Economics 21*, no. 1, 85-96.

Toda, H.Y. and Yamamoto, T. 1995. "Statistical Inference in Vector Autoregressions with Possibly Integrated Processes," *Journal of Econometrics* 66: 225-250.

Watson, M.W. 1994. "Vector Autoregressions and Cointegration," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics, Vol. IV*, ch. 47, 2843-2915.

West, K.D. 1988. "Asymptotic Normality when Regressors have a Unit Root," *Econometrica* 56: 1397-1417.

Wright, J.H. 1996a. "Confidence Intervals for Cointegrating Vectors based on Stationarity Tests," manuscript, Harvard University.

Wright, J.H. 1996b. "Inference for Impulse Response Functions," manuscript, Harvard University.

Table 1
Results for Bootstrap Test (Algorithms B3/B4), k=1

A. Actual Asymptotic Size of Nominal 5% and 10% Bootstrap Tests

	λ				
Nominal Size	0.00	0.25	0.50	0.75	1.0
10%	0.10	0.17	0.24	0.30	0.37
5%	0.05	0.10	0.14	0.19	0.23

B. Bootstrap Size Adjustments

	λ				
Size	0.0	0.25	0.50	0.75	1.0
10%	0.10	0.045	0.030	0.020	0.015
5%	0.05	0.018	0.013	0.008	0.005

Notes: Panel A gives the actual size of the bootstrap procedures (algorithms B3 and B4) with nominal sizes given in the first column. Panel B gives the required nominal sizes for the bootstrap procedures that lead to actual sizes given in the first column. For example, to achieve a size of 5% when λ=0.50, Panel B shows that a bootstrap test with a nominal size of 1.3% should be used.

Table 2
Wald Statistic Sup-Bound Critical Values, k=1

	λ			
Size	0.25	0.50	0.75	1.0
10%	4.37	5.46	6.35	7.08
5%	5.95	7.12	8.04	8.66
1%	9.89	11.40	12.12	12.80

Notes: The entries in the table are the supremum of critical values for the Wald statistics $S^*(C^*+C, r^*+r)$ over all values of C and ω, for the sizes given in the first column of the table.

Table 3
Asymptotic Relative Efficiencies of Tests Relative to Optimal Test

A. 50% Power

	5% Size				10% Size			
	λ				λ			
Test	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00
Restricted Optimal	0.95	0.84	0.98	0.97	0.97	0.84	0.84	0.91
Bootstrap	0.55	0.59	0.53	0.45	0.56	0.62	0.55	0.44
Sup-Bound	0.53	0.56	0.51	0.43	0.54	0.57	0.51	0.41

B. 80% Power

	5% Size				10% Size			
	λ				λ			
Test	0.25	0.50	0.75	1.00	0.25	0.50	0.75	1.00
Restricted Optimal	0.99	0.95	0.96	0.98	0.97	0.93	0.91	0.94
Bootstrap	0.62	0.59	0.57	0.51	0.61	0.62	0.59	0.53
Sup-Bound	0.60	0.57	0.54	0.49	0.60	0.60	0.56	0.50

Notes: Relative Efficiencies are calculated as the limiting ratio of sample sizes necessary to achieve an average power of 50% (Panel A) and 80% (Panel B), when averaged over all C, ω and positive and negative values of β. Efficiencies are relative to the fitted optimal test from Section 5, and relative efficiencies less than 1 imply that the test requires a larger sample size than the fitted optimal test.

Table 4

Size Adjustments for the Bootstrap (Algorithms B3/B4), k=2

A. Rank(R)=2

	λ				
Size	0.0	0.25	0.50	0.75	1.0
10%	0.10	0.043	0.025	0.015	0.008
5%	0.05	0.018	0.010	0.005	0.003

B. Rank(R)=1

		λ_1				
Size	λ_2	0.00	0.25	0.50	0.75	1.00
10%	0.00	0.100	0.0325	0.0150	0.0075	0.0050
5%		0.050	0.0125	0.0050	0.0025	<.0025
10%	0.25	0.0350	0.0125	0.0075	0.0050	----
5%		0.0150	0.0050	0.0025	<.0025	
10%	0.50	0.0200	0.0075	0.0050	----	----
5%		0.0050	0.0025	<.0025		
10%	0.75	0.0125	0.0050	----	----	----
5%		0.0025	<.0025			
10%	1.0	0.0075	----	----	----	----
5%		0.0025				

Notes: The entries are size adjustments for the (algorithms B3/B4) bootstrap for Wald tests of the null hypothesis $H_0: R\beta = R_0$, where β is 2×1. The entries in the table are the required nominal sizes for the bootstrap procedures that lead to actual sizes given in the first column. See Section 5.5 for the definition of λ, λ_1 and λ_2.

Table 5
Long-run Money Demand

$$m_t = \beta_0 + \beta_y y_t + \beta_r r_t + u_t$$

A. Point Estimates and 95% Confidence Intervals

Estimator	$\hat{\beta}_y$	(Conf. Int.)	$\hat{\beta}_R$	(Conf. Int.)
$\hat{\theta}(0,0)$ (DOLS)	0.970	(0.880, 1.060)	-0.101	(-0.126, -0.076)
$\hat{\theta}(\hat{C},\hat{r})$	0.983	(0.864, 1.102)	-0.101	(-0.139, -0.063)

B. Estimated VAR for (y_t, R_t)

$$\begin{bmatrix}\Delta y_t \\ \Delta R_t\end{bmatrix} = \begin{bmatrix} 0.161 \\ -2.239\end{bmatrix} + \begin{bmatrix} 0.004 \\ -0.079\end{bmatrix} t + \begin{bmatrix}-0.141 & -0.003 \\ 2.934 & -0.144\end{bmatrix}\begin{bmatrix}y_{t-1} \\ R_{t-1}\end{bmatrix} + u_{x,t}$$

$$\hat{A} = \begin{bmatrix} 0.859 & -0.003 \\ 2.934 & 0.856\end{bmatrix}, \quad \hat{\omega}_2 = \begin{bmatrix} 4.89 \\ -0.94\end{bmatrix}, \quad \hat{C} = \begin{bmatrix}-14.3 & 14.2 \\ -5.9 & -11.4\end{bmatrix}, \quad \hat{r} = \begin{bmatrix} 63.88 \\ -77.95\end{bmatrix}$$

Notes: Results are based on Annual U.S. Data from 1900-1989 as described in the text. *Panel A:* $\hat{\theta}(0,0)$ is the DOLS estimator, and the results are taken from Stock and Watson (1993), table III. $\hat{\theta}(\hat{C},\hat{r})$ is the OLS estimator from equation (3.1) constructed using estimates of $\hat{C}$, $\hat{r}$ from equation (2.6) and (3.3). Confidence intervals based on $\hat{\theta}(\hat{C},\hat{r})$ were constructed using the size adjusted bootstrap and algorithm B4. Both $\hat{\theta}$ estimators were constructed using 2 leads and lags of $u_{X,t}$ (see equation (3.1). *Panel B:* The first equation shows the estimated values of Ψ_1, Ψ_2 and Ψ_3 from equation (2.6). $\hat{A}=I+\hat{\Psi}_1$, and $\hat{\omega}_2$ was constructed using equation (3.4).

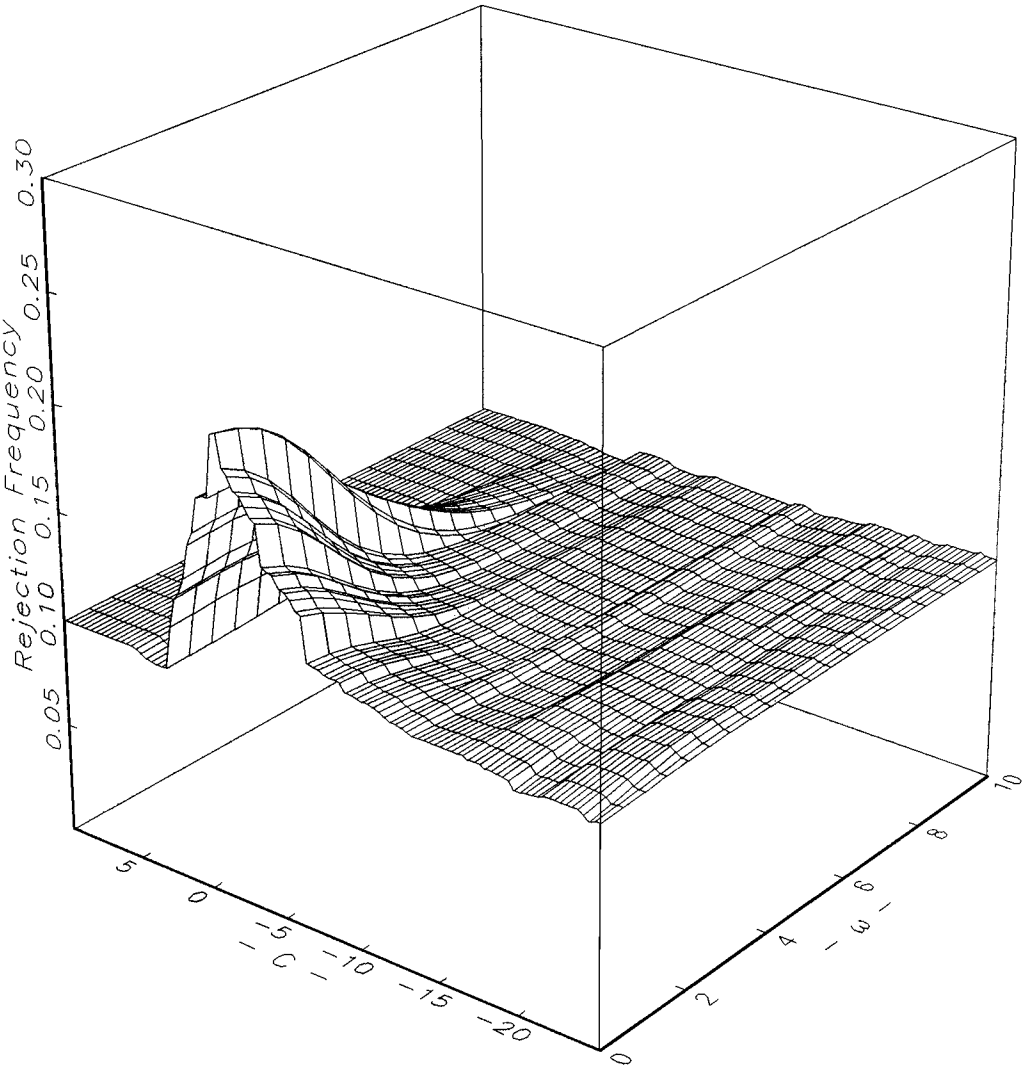

Figure 1. Bootstrap Rejection Frequency
Nominal Size = 10%, k=1, λ=0.5

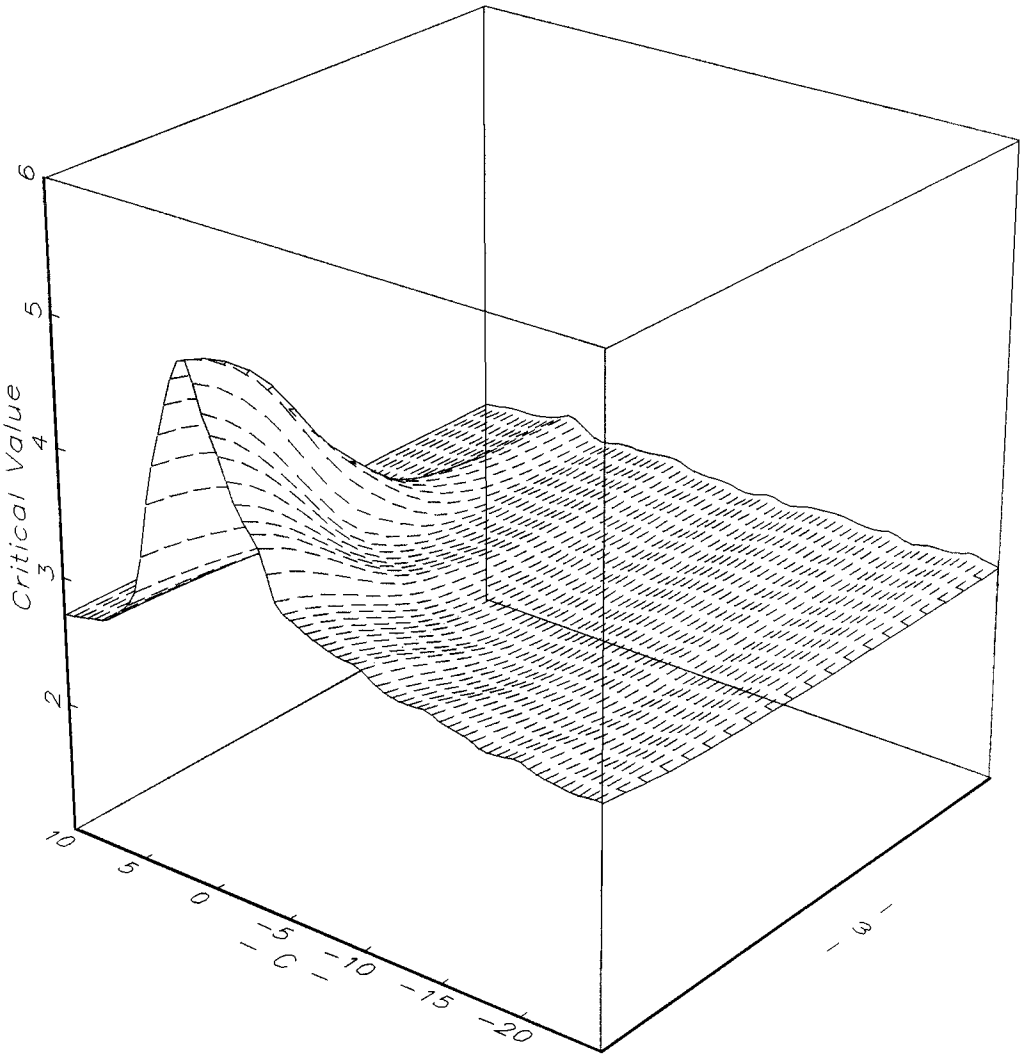

Figure 2. Wald Statistic 10% Critical Value

$k=1$, $\lambda=0.5$

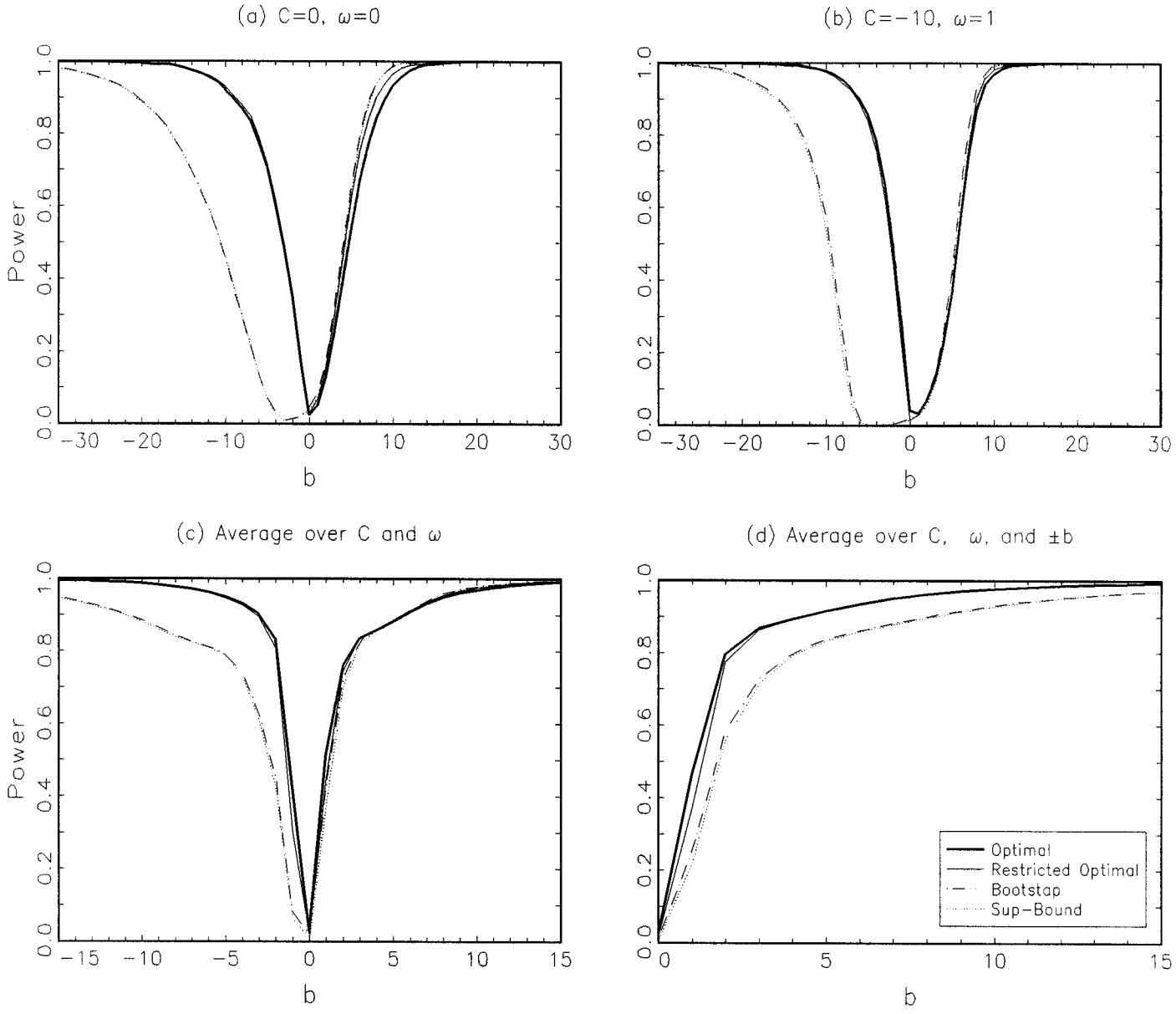

Figure 3. Power of Alternative Tests

Size = 5%, k=1, λ=0.5

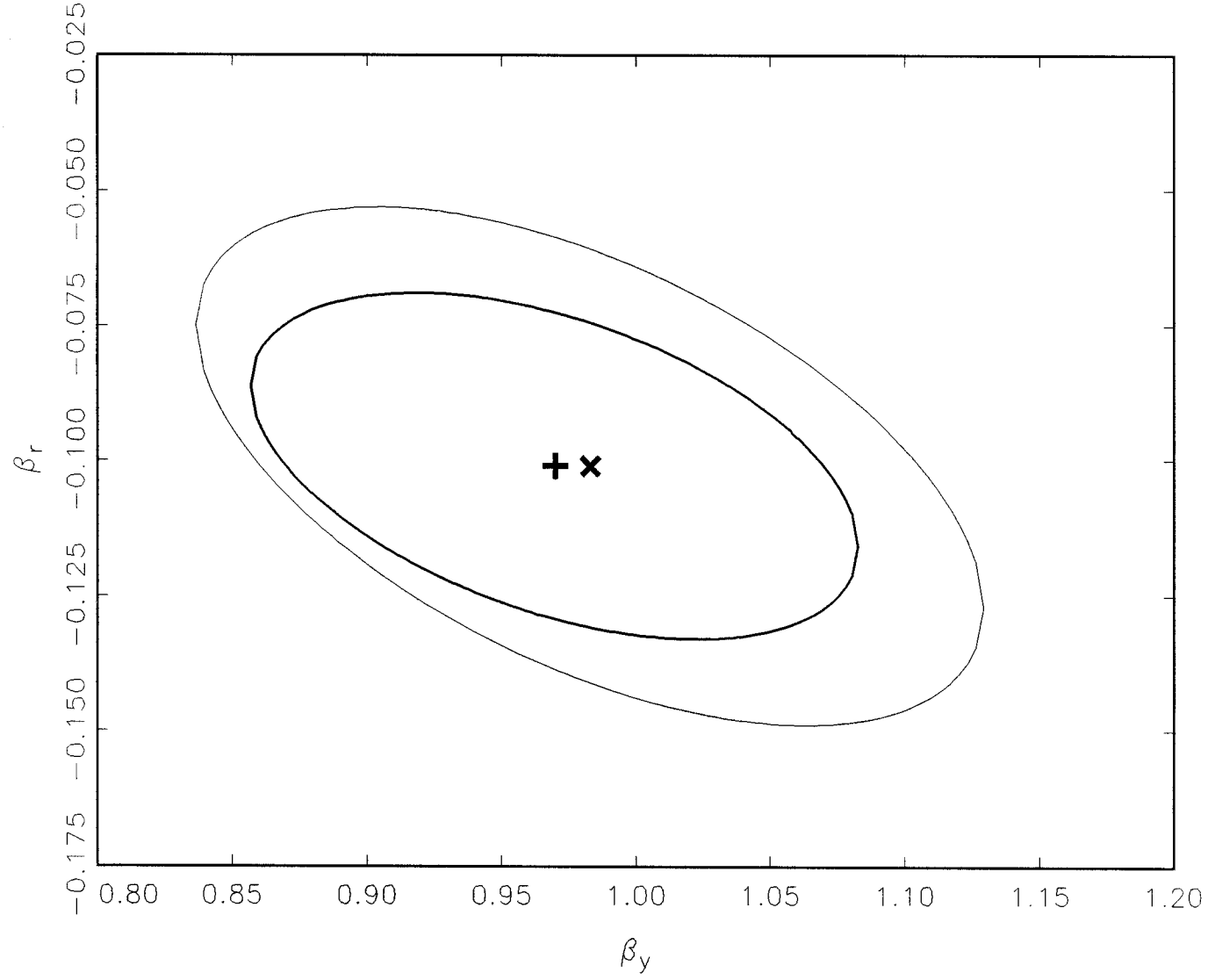

Figure 4. Long-Run Money Demand Coefficients

95% Confidence Ellipses

MLE Assuming C=0 (Thick Line and +)

MLE with C unknown (Thin Line and ×)