

ON THE ROBUSTNESS OF HYPOTHESIS TESTING BASED ON FULLY MODIFIED VECTOR AUTOREGRESSION WHEN SOME ROOTS ARE ALMOST ONE

HEIKKI KAUPPI
University of Helsinki

This paper proves that the fully modified vector autoregression (FM-VAR) estimator has second-order bias effects when some roots are local to unity. These bias effects are shown to result in potentially severe size distortions in FM-VAR testing when the hypothesis involves near unit root variables. In addition, the paper reveals that with the FM-VAR method near unit roots become estimated as exact unit roots with convergence speed faster than the order of the sample size. Also this result implies problems for FM-VAR testing, as such “hyper-consistent” estimates give rise to degenerate limit distributions under the null hypothesis.

1. INTRODUCTION

Fully modified vector autoregression (FM-VAR) was designed by Phillips (1995) to allow for robust statistical inference on an unrestricted vector autoregression (VAR) without any need to examine whether the data are stationary or possibly contain unit roots and cointegration. To allow one to ignore the number of unit roots and their location (i.e., the cointegrating relations), the FM-VAR estimator treats all variables as potential unit root processes and accordingly corrects the ordinary least squares (OLS) estimator for any harmful correlation effects and for the endogeneity of the regressors that may arise from cointegrating relations between these variables. All these corrections are performed by using nonparametric kernel estimators in the manner originally developed by Phillips and Hansen (1990). Phillips (1995)

I am grateful to Pentti Saikkonen, Jim Stock, Markku Lanne, Jukka Nyblom, and three referees for very helpful comments on earlier drafts. This paper is a part of the research program of the Research Unit on Economic Structures and Growth (RUEG) at the Department of Economics at the University of Helsinki. Financial support from the ASLA Fulbright, the Yrjö Jahnsson Foundation, and the Finnish Cultural Foundation is gratefully acknowledged. The usual disclaimer applies. Address correspondence to: Heikki Kauppi, University of Helsinki, Department of Economics, P.O. Box 17 (Arkadiankatu 7), FIN-00014 University of Helsinki, Finland; e-mail: heikki.kauppi@helsinki.fi.

shows that whether the data are stationary or contain unit roots and cointegration these corrections yield convenient asymptotic estimation theory and related Wald tests on levels VARs have limiting distributions that are bounded above by the χ^2 distribution with degrees of freedom equal to the number of restrictions. Therefore, conventional critical values can be applied to obtain valid (but conservative) asymptotic tests of hypothesis on the coefficients of a VAR.¹

On the other hand, Elliott (1998) shows that a set of commonly used procedures that are designed for testing hypotheses on cointegrating parameters and that require pretesting for cointegrating rank tend to suffer from size distortions when some roots are large but not exactly equal to one. His message is that these techniques can easily fail to produce valid inference when some individual variables are inferred to be exact unit root processes although they are in fact generated by highly autocorrelated processes with roots slightly less than unity. Although Elliott (1998) shows that the problem arises in testing procedures for normalized cointegrating vectors based on the full information maximum likelihood estimator or an asymptotically equivalent estimator, his analysis does not deal with the FM-VAR approach. Given the promises of the FM-VAR method, especially that it does not require knowledge of the location and number of unit roots, one could easily believe that it could overcome the problem introduced by Elliott (1998).

However, this paper proves that although the FM-VAR testing procedure requires neither pretesting for cointegrating rank nor explicitly imposing unit root and cointegration restrictions in its formulations it is basically faced with the same problem as the testing procedures studied by Elliott (1998). With local to unit root parametrization and related asymptotic theory, we demonstrate that the FM-VAR estimator has second-order bias effects if some roots are nearly one. The bias effects are present in coefficient estimates of local to unit root variables, and therefore Wald tests based on the FM-VAR estimator have the potential to be severely size distorted in the same way as the tests based on the methods studied by Elliott (1998). This is shown analytically, and a comparison to the result of Elliott (1998) is provided. A simulation study is reported showing that although the size distortions of the FM-VAR testing procedure can be smaller in magnitude than those of the procedures covered by Elliott (1998), they can often be unsatisfactorily high. In addition, we show that with the FM-VAR estimator near unit roots become estimated as exact unit roots with a convergence speed faster than the order of the sample size. Tests based on estimates of this kind are not generally valid if the investigated hypothesis happens to identify the corresponding direction of the parameter space. The only part of the parameter space where the FM-VAR estimator works properly involves coefficients related to those variables or directions of the process where the standard stationary asymptotics provide a good approximation.

2. SETUP AND RESULTS

The analysis is carried out in terms of the first-order n -vector autoregression

$$y_t = Ay_{t-1} + \varepsilon_t, \quad t = 1, \dots, T, \tag{1}$$

where ε_t is $iid(0, \Sigma_{\varepsilon\varepsilon})$ with $\Sigma_{\varepsilon\varepsilon}$ positive definite and with finite fourth-order cumulants. The initial values in y_0 can be any random variables, including constants, whose distribution is independent of T . Although model (1) is a special case of the one studied by Phillips (1995), it suffices for our purposes and allows us to make our general point on the FM-VAR method that can be extended to any higher order VARs that possibly include a constant and a linear time trend.

Suppose our primary interest is in testing an economic hypothesis that can be expressed as linear restrictions on A such that

$$R \text{vec}(A) = r, \tag{2}$$

where R and r are known $(q \times n^2)$ matrix of rank q and q -dimensional vector, respectively. If the system contains unit roots, then standard test statistics for this hypothesis such as a Wald statistic based on OLS estimation of (1) generally do not have standard asymptotic distributions, such as a χ^2 distribution. This result arises from the fact that the sample covariance of the nonstationary linear combinations of the components of y_{t-1} and the error of the system does not converge to zero, but rather, it converges weakly to a nonstandard distribution consisting of functionals of components of a vector Brownian motion. The associated distribution is mislocated or shifted away from the true parameter value, and this fact generally distorts hypothesis tests based on the OLS estimates (for a detailed discussion, see, e.g., Phillips, 1995).

There are different ways to try to overcome the inferential problems caused by the possible presence of unit roots in a VAR. One alternative is to employ an error correction representation of VAR (see, e.g., Johansen, 1991), on which equivalent restrictions to those on the original VAR model can be formulated. However, this approach requires pretesting for cointegrating rank, which is known to induce size distortions and pretest bias in many cases (cf. Elliott, 1998). In contrast, the FM-VAR procedure of Phillips (1995) attempts to obtain robust statistical inference on a levels VAR without the need to pretest the data concerning unit roots and cointegration.

To set up the formulas of the FM-VAR estimator and test statistic, respectively, we define the following generic notation. For any pair of covariance stationary series $\{a_t\}$ and $\{b_t\}$ the long-run covariance matrix $\Omega_{ab} = \sum_{k=-\infty}^{\infty} E(a_{t+k} b_t')$, and the one-sided long-run covariance matrix $\Delta_{ab} = \sum_{k=0}^{\infty} E(a_{t+k} b_t')$. Correspondingly, kernel estimators of these matrices are defined by

$$\hat{\Omega}_{ab} = \sum_{j=-T+1}^{T-1} w(j/K) \hat{\Gamma}_{ab}(j), \tag{3}$$

$$\hat{\Delta}_{ab} = \sum_{j=0}^{T-1} w(j/K) \hat{\Gamma}_{ab}(j), \tag{4}$$

where $w(\cdot)$ is a kernel function with a lag truncation or bandwidth parameter K and

$$\hat{\Gamma}_{ab}(j) = T^{-1} \sum_{1 \leq t, t+j \leq T} a_{t+j} b'_t, \quad \hat{\Gamma}_{ab}(-j) = \hat{\Gamma}'_{ba}(j). \tag{5}$$

Often, the series a_t and b_t in (5) have to be replaced by appropriate estimators, in which case the subscripts are modified accordingly. The following assumptions are from Phillips (1995).

Assumption KL (Kernel Condition). The kernel function $w(\cdot): R \rightarrow [-1, 1]$ is a twice continuously differentiable even function with

- (a) $w(0) = 1, w'(0) = 0, w''(0) \neq 0$, and either
- (b) $w(x) = 0, |x| \geq 1$, with $\lim_{|x| \rightarrow 1} w(x)/(1 - |x|)^2 = \text{constant}$, or
- (b') $w(x) = O(x^{-2})$, as $|x| \rightarrow \infty$.

Assumption BW (Bandwidth Expansion Rate). The bandwidth parameter K in the kernel estimates (3) and (4) has an expansion rate of the form $K = O_e(T^k)$ for some $k \in (\frac{1}{4}, \frac{2}{3})$, where the expansion rate order symbol O_e is defined in Phillips (1995, p. 1032).

Now, applying formula (43) of Phillips (1995) the FM-VAR estimator of A in (1) is given by

$$\hat{A}^+ = \left[\sum_{t=1}^T y_t y'_{t-1} - \hat{\Omega}_{\hat{\varepsilon}\Delta y} \hat{\Omega}_{\Delta y \Delta y}^{-1} \left(\sum_{t=1}^T \Delta y_{t-1} y'_{t-1} - T \hat{\Delta}_{\Delta y \Delta y} \right) \right] \left[\sum_{t=1}^T y_{t-1} y'_{t-1} \right]^{-1}, \tag{6}$$

where the subscripts $\hat{\varepsilon}$ and Δy in the estimated long-run covariance matrices refer to the residual series $\hat{\varepsilon}_t = y_t - \hat{A}y_{t-1}$ from an OLS estimation of (1) and the series Δy_{t-1} , respectively. The corrections associated with the kernel estimators in (6) are designed to remove any harmful correlations between the nonstationary directions of the regressors and the errors of the model while preserving the standard asymptotic theory for the stationary part of the parameter space (for details of the argument, see Phillips, 1995). The FM-VAR-based Wald test statistic to test for the restrictions given in (2) is

$$W = (R \text{vec}(\hat{A}^+) - r)' \left\{ R \left[\hat{\Sigma}_{\varepsilon\varepsilon} \otimes \left(\sum_{t=1}^T y_{t-1} y'_{t-1} \right)^{-1} \right] R' \right\}^{-1} (R \text{vec}(\hat{A}^+) - r), \tag{7}$$

where $\hat{\Sigma}_{\varepsilon\varepsilon} = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}'_t$ is the OLS estimator of $\Sigma_{\varepsilon\varepsilon}$ from model (1).

The asymptotic theory for the estimator and the test statistic in (6) and (7), respectively, can be found from Theorems 5.7 and 6.1 of Phillips (1995). To get an idea how these theories should be modified when some of the roots in the model are close to but not exactly equal to one, suppose that A has the simple form

$$A = \begin{pmatrix} 0 & G \\ 0 & F \end{pmatrix}, \tag{8}$$

where G is an $r \times (n - r)$ matrix and F is an $(n - r) \times (n - r)$ matrix. Partitioning $y_t = (y'_{1t}, y'_{2t})'$ conformably with A the model (1) may be written as

$$y_{1t} = Gy_{2t-1} + \varepsilon_{1t}, \tag{9}$$

$$y_{2t} = Fy_{2t-1} + \varepsilon_{2t}. \tag{10}$$

In particular, we now assume that $F = I + T^{-1}C$, where C is a fixed diagonal matrix. If all diagonal elements in C are zeros, y_{2t} is a vector random walk and the model (1) has $n - r$ (exact) unit roots. In this case the model reduces to the leading example used by Phillips (1995) to illustrate and motivate the FM-VAR approach. However, if a diagonal element is negative, say, then the corresponding variable in y_{2t} is mean reverting and the system has a root that is only local to one. Using this parametrization we obtain asymptotic results that provide more accurate approximations than those obtained assuming a fixed parameter (F) when the underlying process for y_{2t} is slowly mean reverting and the sample size is moderate (cf. Elliott, 1998). The following theorem establishes the limiting behavior of the FM-VAR estimator when the diagonal elements in C may be nonzero. Note that the error covariance matrix assumes the partition $\Sigma_{\varepsilon\varepsilon} = [\Sigma_{ij}]$, $(i, j = 1, 2)$ conformably with that of y_t (or ε_t).

THEOREM 1. *Let \hat{A}^+ be an FM-VAR estimator for model (1) where A is given by (8) with $F = I + T^{-1}C$; and define $e_1 = [I_r \ 0_{r \times (n-r)}]$, $e_2 = [0_{(n-r) \times r} \ I_{n-r}]$, $\beta' = [I_r \ -G]$, and $\beta'_\perp = [G' \ I_{n-r}]$. Then, under Assumptions KL and BW, as $T \rightarrow \infty$,*

- (a) $T^{1/2}(\hat{A}^+ - A)\beta \Rightarrow N(0, \Sigma_{\varepsilon\varepsilon} \otimes (I_r + G'G)\Sigma_{\beta\beta}^{-1}(I_r + G'G))$,
- (b1) $Te_1(\hat{A}^+ - A)\beta_\perp \Rightarrow \Sigma_{1,2}^{1/2} \int_0^1 dW_{1,2}(s) J_C(s)' (\int_0^1 J_C(s) J_C(s)' ds \Sigma_{22}^{1/2'})^{-1} - \Sigma_{12} \Sigma_{22}^{-1} C$,
- (b2) $Te_2(\hat{A}^+ - A)\beta_\perp + C \xrightarrow{p} 0$,

where $\Sigma_{\beta\beta} = \beta' \Sigma_{\varepsilon\varepsilon} \beta$, $\Sigma_{1,2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}$, and $J_C(s)$ is an Ornstein–Uhlenbeck process generated by the multivariate stochastic differential equation $dJ_C(s) = CJ_C(s) ds + dW_2(s)$, $J_C(0) = 0$, where $W_2(s)$ denotes an $(n - r)$ -vector standard Brownian motion defined on $[0, 1]$ that is given by the weak limit of the partial sum $\Sigma_{22}^{-1/2} T^{-1/2} \sum_{t=1}^{[Ts]} \varepsilon_{2t}$. Furthermore, $W_{1,2}(s)$ is an r -vector standard Brownian motion independent of $W_2(s)$.

Part (a) of Theorem 1 gives the asymptotic behavior of the FM-VAR estimator to the stationary directions in the same way as part (a') of Theorem 5.7 of Phillips (1995). We notice that the value of C makes no difference to these directions, and thus, the coefficients of the “clearly” stationary variables in the model are estimated with the same limiting theory whether the parameter F is close to or equal to a unit root (identity) matrix.

Part (b1) of Theorem 1 gives the asymptotic properties of the FM-VAR estimator of the parameter G in (9). It shows that the Brownian motion that is present in the distribution when y_{2t} has only exact unit roots is replaced by an Ornstein–Uhlenbeck process when some of the roots are just local to unity. This, of course, reflects the asymptotic properties of the local to unit root parameterization. More important, the result of part (b1) of Theorem 1 shows that a small deviation from an exact unit root can result in a second-order bias term, $-\Sigma_{12}\Sigma_{22}^{-1}C$, in the limiting distribution of the FM-VAR estimator. In general, this term disappears only when there is no simultaneity in the model, i.e., when $\Sigma_{12} = 0$. Note also that the bias effect is especially present in the estimator of the cointegrating coefficient for which the FM-VAR estimator is efficient in the same way as conventional cointegrating parameter estimators (cf. Phillips, 1995). This observation is closely related to Theorem 1 of Elliott (1998) and indicates that near unit roots distort the FM-VAR estimator especially in that part of the parameter space where the estimator behaves optimally when these roots are exactly one. Furthermore, it can be seen that the bias effects appear only for those coefficients that are on variables with near unit roots, whereas parameters on variables with exact unit roots are unaffected by the presence of near unit roots in the system—this same observation holds for the conventional cointegration estimators also (see Elliott, 1998).

Part (b2) of Theorem 1 shows that with the FM-VAR method near unit roots become estimated as exact unit roots with convergence speed that is faster than the order of the sample size.² This result is in contrast with the analysis of Phillips (1995), which indicates that such “hyperconsistent” rates of convergence can only occur when the system has a full set of unit roots (this would be n unit roots in the present model). However, Theorem 1 shows that this can happen even more generally and that estimates converge to exact unit roots even when the true roots are just nearly ones. Note that the result holds whether the system errors are contemporaneously correlated or not.

The result of part (b2) of Theorem 1 has severe implications for FM-VAR testing. First, it is clear that FM-VAR hypothesis tests involving estimates of this kind are not valid in the usual sense of a test, because their limit distributions are degenerate under the null hypothesis $F = I$, say.³ This problem would be relevant if the FM-VAR method were used to test for the cointegrating rank, which is effectively a test for the number of unit roots in the system. Second, such tests would have no asymptotic power against alternatives within the T^{-1} neighborhood of unity. This would be a weakness in the case of a cointegration test, because conventional tests for the cointegrating rank have power against

such local alternatives (see, e.g., Saikkonen and Lütkepohl, 1999). These remarks indicate that the endogeneity corrections tend to invalidate any statistical inference on (near) unit root coefficients in a VAR.

The focus of the rest of this section is on showing how FM-VAR hypothesis testing may be distorted by the local to unit root bias effects of part (b1) of Theorem 1. Suppose the Wald test statistic in (7) is applied to test a linear restriction on the coefficients of y_{2t} in the first r equations of the model (1) when A has the structure given in (8). We then have the following result.

COROLLARY 1. *Suppose A has the form given in (8) and let W_G be an FM-VAR-based Wald test statistic from (1) for the hypothesis $R \text{vec}(A) = r$, where r is a q -dimensional vector and R is a $q \times n^2$ matrix of rank q ($q \leq r(n - r)$) imposing restrictions on G only. Define a matrix R_G such that $R = R_G(e_1 \otimes e_2)$. Then, as $T \rightarrow \infty$,*

$$W_G \Rightarrow Z' \Pi Z + \mathcal{B}, \tag{11}$$

with

$$\Pi = [R_G V R'_G]^{1/2} [R_G S R'_G]^{-1} [R_G V R'_G]^{1/2}$$

and

$$\mathcal{B} = [R_G \text{vec}(b)]' [R_G S R'_G]^{-1} R_G \text{vec}(b) + 2[R_G \text{vec}(b)]' \Pi^{1/2} Z, \tag{12}$$

where $b = -\Sigma_{12} \Sigma_{22}^{-1} C$, $S = \Sigma_{11} \otimes (\Sigma_{22}^{1/2} \int_0^1 J_C(s) J_C(s) ds \Sigma_{22}^{1/2})^{-1}$, $V = (\Sigma_{1.2} \Sigma_{11}^{-1} \otimes I) S$, and Z is a q -vector of independent normal variables with mean zero and variance unity. Furthermore, Z is independent of $\int_0^1 J_C(s) J_C(s)' ds$ in (12).

If $C = 0$ in Corollary 1 we have $\mathcal{B} = 0$ in (12) and $W_G \Rightarrow \sum_{i=1}^q \pi_i \chi_i^2$, where χ_i^2 are independent χ^2 variates and the weights π_i are the eigenvalues of the matrix $[R_G V R'_G] [R_G S R'_G]^{-1}$. This result would be identical to one implied by Theorem 6.1 of Phillips (1995) when $F = I$. Note that $\Sigma_{1.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12} \leq \Sigma_{11}$ implies $\Pi \leq I$ and, thus, the weights π_i satisfy $0 < \pi_i \leq 1$. Therefore, the FM-VAR test is bounded above by the usual χ_q^2 distribution when $F = I$. As this same distribution is used to obtain critical values for the test, it is clear that the FM-VAR test is asymptotically conservative. Note that in an “exact stationarity” case, i.e., if F were fixed with the corresponding characteristic roots outside the unit root circle, the test statistic W_G would be asymptotically χ_q^2 distributed (cf. Phillips, 1995, Theorem 6.1).

The most interesting result of Corollary 1 is that when $F = I + T^{-1}C$ with nonzero diagonal elements in C , then if there is simultaneity in the model, i.e., $\Sigma_{12} \neq 0$, the FM-VAR Wald test has a bias term given by \mathcal{B} in (12). The first and second terms in \mathcal{B} , respectively, characterize the mean and variance of the bias, which both depend upon Σ and C . The mean term is always nonnegative,

and, thus, a deviation from an exact unit root tends to increase the size of the test. When Σ_{12} is nonzero, this effect is absent only from hypothesis tests that do not impose restrictions on coefficients of near unit root variables (cases where $R_G \text{vec}(b) = 0$). This latter observation is similar to one obtained for the conventional cointegrating parameter estimators and indicates that tests of hypotheses imposing restrictions only on coefficients of variables with exact unit roots are unaffected by the presence of near unit roots in the model (the “partially misspecified” case in Elliott, 1998).

To further illustrate the result of Corollary 1 and to see how it relates to that of Elliott (1998) we consider a simple example. We assume that y_{1t} and y_{2t} are scalars (i.e., $n = r = 1$) and that we are testing for the hypothesis that the variable y_{2t} has no Granger-causal effect on y_{1t} (i.e., $G = 0$). Then from Corollary 1 we see that the corresponding FM-VAR-based Wald test statistic, W_g , say, has the limit theory

$$W_g \Rightarrow (1 - \rho^2)Z^2 + \mathcal{B}_g, \tag{13}$$

with

$$\mathcal{B}_g = C^2 \rho^2 \int_0^1 J_C(s)^2 ds + 2C \sqrt{\rho^2(1 - \rho^2)} \sqrt{\int_0^1 J_C(s)^2 ds} Z, \tag{14}$$

where $\rho = \Sigma_{12} / \sqrt{\Sigma_{11} \Sigma_{22}}$ and Z is a normal variate with mean zero and variance unity. Now, let W_g^* denote a Wald test statistic for this same restriction that has been computed by applying the full information maximum likelihood estimator of G assuming y_{2t} is generated by a unit root process. From the corollary of Elliott (1998), W_g^* would then have the limiting distribution

$$W_g^* \Rightarrow Z^2 + \mathcal{B}_g^*, \tag{15}$$

where

$$\mathcal{B}_g^* = \frac{C^2 \rho^2}{1 - \rho^2} \int_0^1 J_C(s)^2 ds + 2C \sqrt{\frac{\rho^2}{1 - \rho^2}} \sqrt{\int_0^1 J_C(s)^2 ds} Z, \tag{16}$$

with ρ and Z just as in (14). We notice that the bias term \mathcal{B}_g^* in (16) is larger than \mathcal{B}_g in (14) and tends to infinity as ρ goes to one. However, this difference between the two distributions is just a reflection of the fact that the FM-VAR test is asymptotically conservative in our data generating process. The magnitude of this difference can be evaluated through simulations.

Table 1 gives simulated approximations to the rejection probabilities $\Pr(W_g > 3.84)$ and $\Pr(W_g^* > 3.84)$, where W_g and W_g^* are given in (13) and (15), respectively, and 3.84 is the 5% critical value based on the χ_1^2 -distribution. The results are in line with our theoretical findings, and we observe that the asymptotic size of the FM-VAR test is clearly unsatisfactory with moderate values of ρ although it is not as high as the size of the testing procedures

TABLE 1. Asymptotic rejection rates^a

ρ	$C = 0$		$C = -10$		$C = -20$	
	W_g	W_g^*	W_g	W_g^*	W_g	W_g^*
0	5.07	5.07	5.02	5.02	5.15	5.15
0.1	4.83	4.94	5.52	5.67	5.95	6.09
0.2	4.59	5.06	6.74	7.32	9.04	9.78
0.3	4.00	5.02	8.81	10.49	14.59	16.89
0.4	3.17	5.02	11.90	15.77	22.26	27.76
0.5	2.32	5.05	16.14	24.50	32.54	43.55
0.6	1.49	5.04	21.12	36.88	45.05	63.29
0.7	0.62	5.03	27.35	54.71	60.31	83.33
0.8	0.12	5.13	35.33	77.65	76.36	96.82
0.9	0.00	5.12	45.92	96.87	91.63	99.97

^aThe first column gives different values for $\rho = \Sigma_{12} / \sqrt{\Sigma_{11}\Sigma_{22}}$, ($\Sigma_{11} = \Sigma_{22} = 1$). Entries in the rest of the columns give simulated approximations to $\Pr(W_g > 3.84)$ and $\Pr(W_g^* > 3.84)$, where 50,000 replications and 2,000 observations for each combination of ρ and C were used.

covered by Elliott (1998). Although this example indicates that the conservative nature of the FM-VAR test can help to reduce size distortions from a near unit root in some situations, it must be kept in mind that this property comes with the price of lower power to reject false null hypotheses. It can be seen theoretically that if W_g in (13) were size adjusted by dividing with $(1 - \rho^2)$, then the FM-VAR test would have size distortions as large as in the methods studied by Elliott (1998).

Table 2 examines small sample properties of the actual FM-VAR test for the hypothesis $G = 0$ in (1) when the sample size T is 200, and the data are generated by equations (9) and (10) with $G = 0$, $\Sigma_{11} = \Sigma_{22} = 1$, $y_0 = 0$, and with various values of $\rho = \Sigma_{12}$ and C , respectively. In these simulations the FM-VAR estimator is computed by using the Parzen kernel function, and in accordance with Assumption BW, we have chosen to experiment values of the bandwidth parameter K that are the closest integers to $T^{0.26}$, $T^{0.36}$, $T^{0.46}$, $T^{0.56}$, and $T^{0.66}$, respectively. The actual values of K are indicated in the table. The results of Table 2 are clearly in line with those of Table 1, although the choice of the bandwidth parameter seems to play a role in these results. However, this is not much of a surprise because Yamada and Toda (1997) have recently demonstrated by using another simulation setup that the size performance of the FM-VAR test can be highly dependent upon the choice of the value of the bandwidth parameter even with relatively large sample sizes. Therefore, our simulation results confirm our theoretical findings and indicate that FM-VAR

TABLE 2. Rejection rates of the FM-VAR test with sample size $T = 200^a$

ρ	C	$K = 4$	$K = 7$	$K = 11$	$K = 19$	$K = 33$
0	0	2.04	3.49	4.53	5.73	6.40
0.3	0	1.41	2.84	3.79	4.35	5.12
0.6	0	0.72	1.10	1.54	2.28	2.72
0.9	0	0.01	0.03	0.03	0.16	0.46
0	-10	3.13	4.87	6.23	7.68	8.14
0.3	-10	3.60	6.75	8.15	9.24	8.83
0.6	-10	6.12	11.52	13.57	13.45	12.38
0.9	-10	12.53	21.82	25.19	24.10	18.18
0	-20	3.43	6.04	7.23	7.82	8.08
0.3	-20	5.45	8.45	9.94	9.90	8.83
0.6	-20	10.59	17.69	18.27	16.55	11.62
0.9	-20	24.34	37.11	37.06	28.74	19.52

^aThe first two columns give different values for $\rho = \Sigma_{12} / \sqrt{\Sigma_{11} \Sigma_{22}}$, ($\Sigma_{11} = \Sigma_{22} = 1$), and C , respectively. The rest of the columns report rejection rates of the FM-VAR test W_g with different combinations of values of ρ and C . The Parzen kernel function was applied with the bandwidth parameter values K that are indicated in the column headings. Ten thousand replications were used, and the data were generated by using equation (1), where A is as in (8) with $G = 0$, $F = 1 + C/T$, and $T = 200$. An initial value of $y_0 = 0$ was imposed, and the first 500 observations were dropped. The 5% χ^2_1 critical value was used, and the actual nominal size for each ρ (when $C = 0$) can be seen from Table 1 (columns under heading W_g).

hypothesis testing can suffer from potentially severe size distortions when some roots are almost but not exactly equal to one.

3. CONCLUSION

This paper examined the robustness of hypothesis testing based on the FM-VAR estimator when some of the roots of a VAR are large but less than one in absolute value. It was shown that the FM-VAR test can produce severe size distortions when the hypothesis imposes restrictions on variables with near unit roots. As this problem occurs in the particular part of the parameter space where the FM-VAR estimator is efficient in the same way as conventional cointegrating parameter estimators, this finding corresponds with that of Elliott (1998) and confirms that estimation and testing procedures that are somehow based on an assumption about exact unit roots and that are trying to treat these optimally tend to fail when there is just a slight deviation from this assumption. In addition, we showed that with the FM-VAR method near unit roots become estimated as exact unit roots with convergence speed that is faster than the order of the sample size. Also this result implies problems for FM-VAR hypothesis testing, as such “hyperconsistent” estimates tend to give rise to degenerate limit distributions under the null hypothesis. The only part of the parameter space

where the FM-VAR estimator is robust involves coefficients related to those variables or directions of the process where the standard stationary asymptotics provide a good approximation. Therefore, overall, our analysis indicates that the FM-VAR method can easily produce invalid inference if some variables in the system are considerably highly autocorrelated.

As far as possible solutions to the problem studied here are concerned, there is currently only one alternative for FM-VAR that does not require prior knowledge of the number of (near) unit roots and their location in a VAR. This is the lag augmentation procedure of Toda and Yamamoto (1995). Although this method can provide valid inference on coefficients of a levels VAR with near unit roots it has the disadvantage of being inefficient because of its ultimate intention to overfit the model. It also suffers from difficulties arising from uncertainties in finding sufficiently high lag order.

NOTES

1. The usefulness of the FM-VAR approach has been recently discussed and promoted by Freeman, Houser, Kellstedt, and Williams (1998) and Quintos (1998).
2. Yamada and Toda (1998, pp. 62–63) note that this holds for exact unit roots.
3. Here we refer to the Wald test statistic defined in the usual way such as the one in (7).

REFERENCES

- Elliott, G. (1998) On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica* 66, 149–158.
- Freeman, J., D. Houser, P.M. Kellstedt, & J.T. Williams (1998) Long-memoried processes, unit roots, and causal inference in political science. *American Journal of Political Science* 42, 1289–1327.
- Johansen, S. (1991) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated processes, part 1. *Econometric Theory* 4, 468–497.
- Phillips, P.C.B. (1991) Spectral regression for cointegrated time series. In W. Barnett, J. Powell, & G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 413–435. Cambridge University Press.
- Phillips, P.C.B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.
- Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies* 57, 99–125.
- Quintos, C.E. (1998) Fully modified vector autoregressive inference in partially nonstationary models. *Journal of the American Statistical Association* 93, 783–795.
- Saikkonen, P. & H. Lütkepohl (1999) Local power of likelihood ratio tests for the cointegrating rank of a VAR process. *Econometric Theory* 15, 50–78.
- Toda, H.Y. & T. Yamamoto (1995) Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 66, 225–250.
- Yamada, H. & H.Y. Toda (1997) A note on hypothesis testing based on the fully modified vector autoregression. *Economics Letters* 56, 27–39.
- Yamada, H. & H.Y. Toda (1998) Inference in possibly integrated vector autoregressive models: Some finite sample evidence. *Journal of Econometrics* 86, 55–95.

APPENDIX

Proof of Theorem 1. Let $F = [\beta; \beta_\perp]$ and write

$$(\hat{A}^+ - A)F = [A_T - B_T C_T] D_T^{-1} F' F, \tag{A.1}$$

where

$$A_T = \sum_{t=1}^T \varepsilon_t y'_{t-1} F, \quad B_T = \hat{\Omega}_{\varepsilon \Delta y} F [F' \hat{\Omega}_{\Delta y \Delta y} F]^{-1},$$

$$C_T = F' \left[\sum_{t=1}^T \Delta y_{t-1} y'_{t-1} - T \hat{\Delta}_{\Delta y \Delta y} \right] F, \quad D_T = \sum_{t=1}^T F' y_{t-1} y'_{t-1} F.$$

We first derive asymptotic results for $A_T, B_T, C_T,$ and D_T and then subsequently apply these to prove the main theorem. From (9) and (10) we can derive the following notation for later use:

$$\beta' y_t = U_{1t} = \underbrace{\beta' \varepsilon_t}_{u_{1t}} + \underbrace{(-G(F-I)y_{2t-1})}_{\tilde{u}_{1t}}; \tag{A.2}$$

$$\beta' \Delta y_t = \Delta U_{1t} = \underbrace{\beta' \Delta \varepsilon_t}_{\Delta u_{1t}} + \underbrace{(-G(F-I)^2 y_{2t-2} - G(F-I)\varepsilon_{2t-1})}_{\Delta \tilde{u}_{1t}}; \tag{A.3}$$

$$\beta'_\perp y_t = Y_{2t} = (I + G'G)y_{2t-1} + \underbrace{\beta'_\perp \varepsilon_t}_{e_t} + (F-I)y_{2t-1}; \tag{A.4}$$

$$\beta'_\perp \Delta y_t = U_{2t} = \underbrace{(I + G'G)\varepsilon_{2t}}_{u_{2t}} + G'\beta' \Delta \varepsilon_t + \underbrace{(F-I)y_{2t-1} + G'G(F-I)y_{2t-2}}_{y_{2t}}. \tag{A.5}$$

Consider A_T . Using $F - I = T^{-1}C$ and well-known limiting results we obtain

$$T^{-1/2} A_T e'_1 = T^{-1/2} \sum_{t=1}^T \varepsilon_t u'_{1t-1} + O_p(T^{-1/2}) \Rightarrow N(0, \Sigma_{\varepsilon \varepsilon} \otimes \Sigma_{\beta \beta}), \tag{A.6}$$

where $\Sigma_{\beta \beta} = \beta' \Sigma_{\varepsilon \varepsilon} \beta$, and

$$T^{-1} A_T e'_2 = T^{-1} \sum_{t=1}^T \varepsilon_t y'_{2t-1} (I + G'G) + O_p(T^{-1/2}), \tag{A.7}$$

with

$$T^{-1} \sum_{t=1}^T \varepsilon_t y'_{2t-1} (I + G'G) \Rightarrow \left(\begin{array}{c} \Sigma_{11}^{-1/2} \int_0^1 dW_1 J'_C \Sigma_{22}^{-1/2'} \\ \Sigma_{22}^{-1/2} \int_0^1 dW_2 J'_C \Sigma_{22}^{-1/2'} \end{array} \right) (I + G'G), \tag{A.8}$$

where $W_1 \equiv W_1(s)$ and $W_2 \equiv W_2(s)$ are standard vector Brownian motions on $[0,1]$ that are given by the weak limits of the partial sums $\Sigma_{11}^{-1/2} T^{-1/2} \sum_{t=1}^{[Ts]} \varepsilon_{1t}$ and $\Sigma_{22}^{-1/2} T^{-1/2} \sum_{t=1}^{[Ts]} \varepsilon_{2t}$, respectively, whereas $J_C \equiv J_C(s)$ is an Ornstein–Uhlenbeck process defined by the multivariate stochastic differential equation $dJ_C(s) = CJ_C(s) ds + dW_2(s)$, $J_C(0) = 0$.

Consider B_T and C_T and use (A.2)–(A.5) to write

$$\hat{\Omega}_{\hat{\varepsilon}\Delta y} F = [\hat{\Omega}_{\hat{\varepsilon}\Delta U_1} \quad \hat{\Omega}_{\hat{\varepsilon}U_2}] \tag{A.9}$$

$$F' \hat{\Omega}_{\Delta y \Delta y} F = \begin{bmatrix} \hat{\Omega}_{\Delta U_1 \Delta U_1} & \hat{\Omega}_{\Delta U_1 U_2} \\ \hat{\Omega}_{\Delta U_1 U_2} & \hat{\Omega}_{U_2 U_2} \end{bmatrix}, \tag{A.10}$$

where the subscripts ΔU_1 and U_2 refer to the series ΔU_{1t-1} and U_{2t-1} , respectively. Similarly, write

$$T^{-1} C_T = \begin{bmatrix} T^{-1} \sum_{t=1}^T \Delta U_{1t-1} U'_{1t-1} - \hat{\Delta}_{\Delta U_1 \Delta U_1} & T^{-1} \sum_{t=1}^T \Delta U_{1t-1} Y'_{2t-1} - \hat{\Delta}_{\Delta U_1 U_2} \\ T^{-1} \sum_{t=1}^T U_{2t-1} U'_{1t-1} - \hat{\Delta}_{U_2 \Delta U_1} & T^{-1} \sum_{t=1}^T U_{2t-1} Y'_{2t-1} - \hat{\Delta}_{U_2 U_2} \end{bmatrix}. \tag{A.11}$$

Lemma 1, which follows, summarizes the asymptotic properties of each of the elements in (A.9)–(A.11) similarly to Lemma 8.1 of Phillips (1995) for the $C = 0$ case. In the proof of Lemma 1 we work as in the proofs of Phillips (1995) only with long-run covariance matrix estimates that satisfy part (b) of Assumption KL (for additional comments that apply here also, see Phillips, 1995, pp. 1057–1058).

LEMMA 1. *Under Assumptions KL and BW, the following hold:*

- (a) $\hat{\Omega}_{\Delta U_1 \Delta U_1} = K^{-2} M_1 + o_p(K^{-2})$;
- (b) $\hat{\Omega}_{\Delta U_1 U_2} = K^{-2} M_2 + O_p((KT)^{-1/2}) + o_p(K^{-2}) + O_p(T^{-1})$;
- (c) $\hat{\Omega}_{U_2 U_2} = (I + G'G) \Sigma_{22} (I + G'G) + o_p(1)$;
- (d) $\hat{\Omega}_{\hat{\varepsilon} \Delta U_1} = K^{-2} M_3 + O_p((KT)^{-1/2}) + o_p(K^{-2}) + O_p(T^{-1})$;
- (e) $\hat{\Omega}_{\hat{\varepsilon} U_2} = \Sigma_{\varepsilon 2} (I + G'G) + o_p(1)$, where $\Sigma_{\varepsilon 2} = (\Sigma'_{12} \quad \Sigma'_{22})'$;
- (f) $T^{-1} \sum_{t=1}^T \Delta U_{1t-1} U'_{1t-1} - \hat{\Delta}_{\Delta U_1 \Delta U_1} = O_p(K^{-2}) + O_p(T^{-1})$;
- (g) $T^{-1} \sum_{t=1}^T U_{2t-1} U'_{1t-1} - \hat{\Delta}_{U_2 \Delta U_1} = K^{-2} Q_1 + O_p((KT)^{-1/2}) + o_p(K^{-2}) + O_p(T^{-1})$;
- (h) $T^{-1} \sum_{t=1}^T \Delta U_{1t-1} Y'_{2t-1} - \hat{\Delta}_{\Delta U_1 U_2} = T^{-1} u_{1,T-1} y'_{2,T-1} + K^{-2} Q_2 + O_p((KT)^{-1/2}) + o_p(K^{-2})$;
- (i) $T^{-1} \sum_{t=1}^T U_{2t-1} Y'_{2t-1} - \hat{\Delta}_{U_2 U_2} = (I + G'G) [T^{-1} \sum_{t=1}^T \varepsilon_{2t-1} y'_{2t-2} + T^{-2} C \times \sum_{t=1}^T y_{2t-2} y'_{2t-2}] (I + G'G) + o_p(1)$.

The error terms of $O_p((KT)^{-1/2})$ that appear in parts (b), (d), (g), and (h) are sharp.

Proof. Part (a): Use (A.3) to write

$$\hat{\Omega}_{\Delta U_1 \Delta U_1} = \hat{\Omega}_{\Delta u_1 \Delta u_1} + 2\hat{\Omega}_{\Delta u_1 \Delta \bar{u}_1} + \hat{\Omega}_{\Delta \bar{u}_1 \Delta \bar{u}_1}, \tag{A.12}$$

where the first term corresponds to $\hat{\Omega}_{\Delta u_1, \Delta u_1}$ in part (a) of Lemma 8.1 in Phillips (1995, p. 1058). Thus part (a) follows from Phillips (1995) provided that the two last terms in (A.12) are at most $o_p(K^{-2})$. First, use (A.3) to write

$$\hat{\Omega}_{\Delta u_1, \Delta \tilde{u}_1} = -T^{-2} \hat{\Omega}_{\Delta u_1, y_2} C^2 G' - T^{-1} \hat{\Omega}_{\Delta u_1, \varepsilon_2} C G', \tag{A.13}$$

where $\hat{\Omega}_{\Delta u_1, \varepsilon_2} = O_p(K^{-2})$ by part (b) of Lemma 8.1 in Phillips (1995). Second, the first term on the right-hand side of (A.13) is $o_p(T^{-2})$, because

$$\hat{\Omega}_{\Delta u_1, y_2} = O_p(1), \tag{A.14}$$

which can be proved by using similar lines to those in equation (P14) in the proof of part (c) of Lemma 8.1 in Phillips (1995, p. 1064). An important part of these derivations is the following intermediate result:

$$K^{-1} \sum_{j=-K+2}^{K-1} w'(j^*/K) \hat{\Gamma}_{y_2, u_1}(j) = O_p(1), \tag{A.15}$$

where $j^* \in [j - 1, j]$ is defined for each j . This result follows from arguments given in the proof of Theorem 3.1 of Phillips (1991, pp. 432–433) in the same way as the respective result in Phillips (1995, p. 1065). (Note that in the proof of Theorem 3.1 of Phillips, 1991, similar [intermediate] convergence results [as, e.g., Phillips, 1991, p. 433, the first result] can be proved when an exact unit root process is replaced by a local to unit root one.)

Next, use the definition of $\Delta \tilde{u}_1$ in (A.3) to write

$$\hat{\Omega}_{\Delta \tilde{u}_1, \Delta \tilde{u}_1} = T^{-4} G C^2 \hat{\Omega}_{y_2, y_2} C^2 G' + 2T^{-3} G C \hat{\Omega}_{\varepsilon_2, y_2} C^2 G' + T^{-2} G C \hat{\Omega}_{\varepsilon_2, \varepsilon_2} C G', \tag{A.16}$$

where the first term is $o_p(T^{-2})$, because

$$\begin{aligned} \|T^{-2} \hat{\Omega}_{y_2, y_2}\| &= \left\| T^{-2} \sum_{j=-K+1}^{K-1} w(j/K) \hat{\Gamma}_{y_2, y_2}(j) \right\| \\ &= \left\| T^{-3} \sum_{j=-K+1}^{K-1} w(j/K) \left(\sum_{1 \leq t-3+j \leq T} y_{2, t-3+j} y'_{2, t-3} \right) \right\| \\ &\leq T^{-1} \sum_{t=1}^T \frac{K}{T} K^{-1} \sum_{j=-K+1}^{K-1} T^{-1/2} \|y_{2, t-3+j}\| T^{-1/2} \|y_{2, t-3}\| \\ &= \frac{K}{T} T^{-1} \sum_{t=1}^T K^{-1} \sum_{j=-K+1}^{K-1} O_p(1) = \frac{K}{T} O_p(1) = o_p(1), \end{aligned} \tag{A.17}$$

where the last equality follows from Assumption BW. Using the same argument as that in (A.15) we can show that the second term in (A.16) is $o_p(T^{-2})$, whereas the last term in (A.16) is $O_p(T^{-2})$, because $\hat{\Omega}_{\varepsilon_2, \varepsilon_2} \xrightarrow{p} \Omega_{\varepsilon_2, \varepsilon_2}$. Thus, $\hat{\Omega}_{\Delta \tilde{u}_1, \Delta \tilde{u}_1} = O_p(T^{-2})$, and part (a) follows.

Part (b): Use (A.3) and (A.5) to write

$$\hat{\Omega}_{\Delta U_1 U_2} = \hat{\Omega}_{\Delta u_1 v} + \hat{\Omega}_{\Delta u_1 y_2} + \hat{\Omega}_{\Delta \tilde{u}_1 v} + \hat{\Omega}_{\Delta \tilde{u}_1 y_2}, \tag{A.18}$$

where the first term is an analog to the term $\hat{\Omega}_{u_2 \Delta u_1}$ in the second result of part (b) of Lemma 8.1 in Phillips (1995). Now, the result of part (b) follows from Phillips (1995) provided that the three last terms in (A.18) are at most $O_p(T^{-1})$. This is easily seen by applying similar arguments to those in the proof of part (a).

Part (c): The proof is similar to those of parts (a) and (b) and thus is omitted here.

Part (d): Use (A.3) to decompose

$$\hat{\Omega}_{\hat{\varepsilon} \Delta U_1} = \hat{\Omega}_{\hat{\varepsilon} \Delta u_1} + \hat{\Omega}_{\hat{\varepsilon} \Delta \tilde{u}_1} \tag{A.19}$$

and use similar algebra to that in (P13) in Phillips (1995, p. 1064) to obtain

$$\begin{aligned} \hat{\Omega}_{\hat{\varepsilon} \Delta u_1} &= \hat{\Omega}_{\varepsilon \Delta u_1} - w((K-1)/K)(\hat{A} - A)\hat{\Gamma}_{y_{u_1}}(K) + w((-K+1)/K)(\hat{A} - A)\hat{\Gamma}_{y_{u_1}}(-K+1) \\ &\quad + \sum_{j=-K+2}^{K-1} [w(j/K) - w((j-1)/K)](\hat{A} - A)\hat{\Gamma}_{y_{u_1}}(j), \end{aligned} \tag{A.20}$$

where the subscript y denotes the series y_{t-1} . The first term in (A.20) is analogous to $\hat{\Omega}_{u_0 \Delta u_1}$ in the first result of part (b) of Lemma 8.1 in Phillips (1995) and delivers the first three terms on the right-hand side of part (d). Thus, it remains to be shown that the rest of the terms in (A.20) and $\hat{\Omega}_{\hat{\varepsilon} \Delta \tilde{u}_1}$ in (A.19) are at most $O_p(T^{-1})$. These results can be derived with similar arguments to those in the proofs of parts (a) and (b) and by applying similar lines to those in equations (P13) and (P14) in Phillips (1995, p. 1064); details are omitted here.

Part (e): The proof is similar to that of part (d) and hence is omitted here.

Part (f): Use the definition of $\Delta \tilde{u}_1$ in (A.3) to write

$$\hat{\Delta}_{\Delta u_1 \Delta \tilde{u}_1} = -T^{-2} \hat{\Delta}_{\Delta u_1 y_2} C^2 G' - T^{-1} \hat{\Delta}_{\Delta u_1 \varepsilon_2} C G',$$

where $\hat{\Delta}_{\Delta u_1 \varepsilon_2} = O_p(1)$, by similar arguments to those used right after (A.20) and the ones used in (A.15), whereas $\hat{\Delta}_{\Delta u_1 \varepsilon_2} = o_p(1)$ by arguments given in the proof of Lemma 8.1 in Phillips (1995). Therefore, $\hat{\Delta}_{\Delta u_1 \Delta \tilde{u}_1} = o_p(T^{-1})$. Furthermore, similar derivations to those for (A.16) imply $\hat{\Delta}_{\Delta \tilde{u}_1 \Delta \tilde{u}_1} = O_p(T^{-2})$. These lines together with well-known limiting arguments yield

$$T^{-1} \sum_{t=1}^T \Delta U_{1t-1} U'_{1t-1} - \hat{\Delta}_{\Delta U_1 \Delta U_1} = \left[T^{-1} \sum_{t=1}^T \Delta u_{1t-1} u'_{1t-1} - \hat{\Delta}_{\Delta u_1 \Delta u_1} \right] + O_p(T^{-1}), \tag{A.21}$$

where the term in the square brackets is $O_p(K^2)$ (see Phillips, 1995, Lemma 8.1, part (e)). Thus, part (f) follows.

Part (g): In the same way as in the proof of part (f) we can deduce that

$$T^{-1} \sum_{t=1}^T U_{2t-1} U'_{1t-1} - \hat{\Delta}_{U_2 \Delta U_1} = \left[T^{-1} \sum_{t=1}^T u_{2t-1} u'_{1t-1} - \hat{\Delta}_{u_2 \Delta u_1} \right] + O_p(T^{-1}),$$

where the term in the square brackets has similar limit theory as in part (f) of Lemma 8.1 of Phillips (1995), and thus, part (g) follows.

Part (h): As in the proofs of parts (f) and (g) we have

$$\begin{aligned} T^{-1} \sum_{t=1}^T \Delta U_{1t-1} Y'_{2t-1} - \hat{\Delta}_{\Delta U_1 U_2} &= T^{-1} \sum_{t=1}^T \Delta u_{1t-1} y'_{2t-1} (I + G'G) - \hat{\Delta}_{\Delta u_1 u_2} + O_p(T^{-1}) \\ &= T^{-1} u_{1T-1} y'_{2T-1} - T^{-1} \sum_{t=1}^T u_{1t-2} \varepsilon'_{2t-2} (I + G'G) + O_p(T^{-1}), \end{aligned} \tag{A.22}$$

where the last line corresponds to equation (P18) in the proof of part (g) of Lemma 8.1 of Phillips (1995), and thus, part (h) follows from Phillips (1995, p. 1067).

Part (i): Use (A.4) and (A.5) to write

$$\begin{aligned} T^{-1} \sum_{t=1}^T U_{2t-1} Y'_{2t-1} - \hat{\Delta}_{U_2 U_2} &= T^{-1} \sum_{t=1}^T v_{t-1} Y'_{2t-1} + T^{-1} \sum_{t=1}^T y_{2t-1} Y'_{2t-1} \\ &\quad - \hat{\Delta}_{vv} - \hat{\Delta}_{vy_2} - \hat{\Delta}_{y_2 v} - \hat{\Delta}_{y_2 y_2}, \end{aligned}$$

where

$$T^{-1} \sum_{t=1}^T y_{2t-1} Y'_{2t-1} = (I + G'G)CT^{-2} \sum_{t=1}^T y_{2t-2} y'_{2t-2} (I + G'G) + O_p(T^{-1}).$$

Furthermore, applying well-known limiting arguments we get

$$T^{-1} \sum_{t=1}^T v_{t-1} Y'_{2t-1} = (I + G'G)T^{-1} \sum_{t=1}^T \varepsilon_{2t-1} y'_{2t-2} (I + G'G) + V_T + O_p(T^{-1})$$

with

$$\begin{aligned} V_T &= T^{-1} \sum_{t=1}^T G' \beta' \Delta \varepsilon_{t-1} y'_{2t-2} (I + G'G) + T^{-1} \sum_{t=1}^T v_{t-1} \varepsilon'_{t-1} \beta_{\perp} \\ &= -T^{-1} \sum_{t=1}^T G' \beta' \varepsilon_{t-2} \varepsilon'_{2t-2} (I + G'G) \\ &\quad + T^{-1} \sum_{t=1}^T [(I + G'G) \varepsilon_{2t-1} + G' \beta' \Delta \varepsilon_{t-1}] \varepsilon'_{t-1} \beta_{\perp} + o_p(1) \\ &= G' \beta' \Sigma_{22} (I + G'G) + (I + G'G) \Sigma_{2\varepsilon} \beta_{\perp} + G' \beta' \Sigma_{\varepsilon\varepsilon} \beta_{\perp} + o_p(1) = \Delta_{vv} + o_p(1), \end{aligned}$$

where the last equality follows from the definition of Δ_{vv} . Now, as $\hat{\Delta}_{vv} \xrightarrow{p} \Delta_{vv}$, and $\hat{\Delta}_{v_2 y_2}, \hat{\Delta}_{y_2 v_2}$ and $\hat{\Delta}_{y_2 y_2}$ are $o_p(1)$ by similar arguments to those used in the proof of part (f), we obtain part (i). ■

It follows from Lemma 1 that we can apply analogous lines to those in Phillips (1995, p. 1065) to obtain

$$B_T = [N + O_p((K^3 T^{-1})^{1/2}) + o_p((K^3 T^{-1})^{1/2}) \quad ; \quad \Sigma_{\varepsilon_2} \Sigma_{22}^{-1} (I + G'G)^{-1} + o_p(1)], \tag{A.23}$$

where N is a fixed nonzero matrix and $\Sigma_{\varepsilon_2} = (\Sigma'_{12} \quad \Sigma_{22})'$. Combining (A.23) and parts (f)–(i) of Lemma 1 yields

$$T^{-1} B_T C_T e'_1 = O_p(K^{-2}) + O_p((KT)^{-1/2}) \tag{A.24}$$

and

$$\begin{aligned} T^{-1} B_T C_T e'_2 &= \Sigma_{\varepsilon_2} \Sigma_{22}^{-1} \left[T^{-1} \sum_{t=1}^T \varepsilon_{2t-1} y'_{2t-2} + T^{-2} C \sum_{t=1}^T y_{2t-2} y'_{2t-2} \right] (I + G'G) \\ &\quad + O_p(T^{-1/2}) + O_p(K^{3/2} T^{-1}) + o_p(1). \end{aligned} \tag{A.25}$$

From Assumption BW we have $O_p(K^{-2} T^{1/2}) + O_p(K^{-1/2}) = o_p(1)$ (cf. Phillips, 1995, Discussion 8.5.(i)), and thus, (A.24) and (A.25) combined with well-known limiting arguments imply

$$T^{-1/2} B_T C_T = [o_p(1) \quad ; \quad O_p(T^{1/2})]. \tag{A.26}$$

Next applying the formula of the partitioned inverse we get

$$D_T^{-1} = \begin{bmatrix} [a_T - b_T c_T^{-1} b'_T]^{-1} & -[a_T - b_T c_T^{-1} b'_T]^{-1} b_T c_T^{-1} \\ -c_T^{-1} b'_T [a_T - b_T c_T^{-1} b'_T]^{-1} & [c_T - b'_T a_T^{-1} b_T]^{-1} \end{bmatrix},$$

where, by well-known limiting arguments,

$$T^{-1} a_T = T^{-1} \sum_{t=1}^T U_{1t-1} U'_{1t-1} + O_p(T^{-1}) \xrightarrow{p} \Sigma_{\beta\beta},$$

$$T^{-1} b_T = T^{-1} \sum_{t=1}^T U_{1t-1} Y'_{2t-1} = O_p(1),$$

$$T^{-2} c_T = T^{-2} \sum_{t=1}^T Y_{2t-1} Y'_{2t-1} = (I + G'G) T^{-2} \sum_{t=1}^T y_{2t-2} y'_{2t-2} (I + G'G) + O_p(T^{-1}).$$

Using these results it is seen that

$$TD_T^{-1}F'\beta = \begin{bmatrix} \Sigma_{\beta\beta}^{-1}(I + GG') + o_p(1) \\ O_p(T^{-1}) \end{bmatrix} \tag{A.27}$$

and

$$T^2D_T^{-1}F'\beta_{\perp} = \left[(I + G'G)^{-1} \left[T^{-2} \sum_{t=1}^T y_{2t-2}y'_{2t-2} \right]^{-1} + o_p(1) \right]. \tag{A.28}$$

We are ready to complete the proof of Theorem 1. For part (a) normalize the first column in (A.1) by \sqrt{T} to get

$$\begin{aligned} \sqrt{T}(\hat{A}^+ - A)\beta &= T^{-1/2}[A_T - B_T C_T]TD_T^{-1}F'\beta \\ &= T^{-1/2}A_TTD_T^{-1}F'\beta + o_p(1), \end{aligned}$$

where the last equation follows from (A.26) and (A.27). Thus, for part (a) it remains to combine the results in (A.6)–(A.8) with the one in (A.27).

For part (b1) use (A.6)–(A.8) and (A.24)–(A.28) to see that

$$e_1T(\hat{A}^+ - A)\beta_{\perp} = e_1[T^{-1}A_T - T^{-1}B_T C_T]T^2D_T^{-1}F'\beta_{\perp} = P_T - \Sigma_{12}\Sigma_{22}^{-1}C + o_p(1),$$

where

$$P_T = \left[T^{-1} \sum_{t=1}^T \varepsilon_{1t}y'_{2t-1} - T^{-1}\Sigma_{12}\Sigma_{22}^{-1} \sum_{t=1}^T \varepsilon_{2t-1}y'_{2t-2} \right] \left[T^{-2} \sum_{t=1}^T y_{2t-2}y'_{2t-2} \right]^{-1}$$

converges weakly to

$$\begin{aligned} &\left[\Sigma_{11}^{1/2} \int_0^1 dW_1 J_C' \Sigma_{22}^{1/2'} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{22}^{1/2} \int_0^1 dW_2 J_C' \Sigma_{22}^{1/2'} \right] \left[\Sigma_{22}^{1/2} \int_0^1 J_C J_C' \Sigma_{22}^{1/2'} \right]^{-1} \\ &\equiv \Sigma_{1,2}^{1/2} \int_0^1 dW_{1,2} J_C' \Sigma_{22}^{1/2'} \left(\Sigma_{22}^{1/2} \int_0^1 J_C J_C' \Sigma_{22}^{1/2'} \right)^{-1}, \end{aligned}$$

where \equiv denotes equality in distribution and $W_{1,2} = W_1 - \Sigma_{12}\Sigma_{22}^{-1}W_2$ is a standard Brownian motion independent of W_2 .

For part (b2) note that

$$e_2T(\hat{A}^+ - A)\beta_{\perp} = T(\hat{F}^+ - F) = e_2[T^{-1}A_T - T^{-1}B_T C_T]T^2D_T^{-1}F'\beta_{\perp},$$

where $\hat{F}^+ = e_1 \hat{A}^+ e_2'$. It is then straightforward to apply (A.6)–(A.8) and (A.24)–(A.28) to show that

$$T(\hat{F}^+ - F) = \left(T^{-1} \sum_{t=1}^T \varepsilon_{2t} y'_{2t-1} - \left[T^{-1} \sum_{t=1}^T \varepsilon_{2t-1} y'_{2t-2} + T^{-2} C \sum_{t=1}^T y_{2t-2} y'_{2t-2} \right] \right) \times \left(T^{-2} \sum_{t=1}^T y_{2t-2} y'_{2t-2} \right)^{-1} + o_p(1). \tag{A.29}$$

Given $F = I + T^{-1}C$ we see that C cancels out from both sides of (A.29) and thus it can be written as

$$T(\hat{F}^+ - I) = \left[T^{-1} \sum_{t=1}^T \varepsilon_{2t} y'_{2t-1} - T^{-1} \sum_{t=1}^T \varepsilon_{2t-1} y'_{2t-2} \right] \left(T^{-2} \sum_{t=1}^T y_{2t-2} y'_{2t-2} \right)^{-1} + o_p(1). \tag{A.30}$$

This shows why \hat{F}^+ converges to an identity matrix and why we have added C on the left-hand side of part (b2) of Theorem 1. It remains to observe that the difference between the two sums in the square brackets in (A.30) is just $T^{-1} \varepsilon_{2T} y'_{2,T-1} - T^{-1} \varepsilon_{20} y'_{2,-1} = o_p(1)$, whereas the inverse in (A.30) is $O_p(1)$. ■

Proof of Corollary 1. Because the hypothesis imposes restrictions on G only, there indeed exists a $q \times (n(n-r))$ matrix R_G such that $R \text{vec}(A) = R_G(e_1 \otimes e_2) \text{vec}(A) = R_G \text{vec}(G) = r$. Therefore, letting

$$\hat{S} = [e_1 \hat{\Sigma}_{\varepsilon\varepsilon} e_1'] \otimes \left[e_2 \left(T^{-2} \sum_{t=1}^T y_{t-1} y'_{t-1} \right)^{-1} e_2' \right]$$

and using the fact that $R_G \text{vec}(G) = r$ under the null hypothesis we can write

$$W_G = (R_G \text{vec}[T(\hat{G}^+ - G)])' [R_G \hat{S} R_G']^{-1} (R_G \text{vec}[T(\hat{G}^+ - G)]), \tag{A.31}$$

where $\hat{S} \Rightarrow S$, by similar arguments to those used in the proof of Theorem 1, and because $e_1 \hat{\Sigma}_{\varepsilon\varepsilon} e_1' = \hat{\Sigma}_{11} = \Sigma_{11} + o_p(1)$.

Next, because $W_{1.2}(s)$ and $W_2(s)$ are vectors of independent standard Brownian motions, $\text{vec}(\int_0^1 dW_{1.2} J_C' (\int_0^1 J_C J_C')^{-1/2})$ is equal in distribution to $N(0, I_{r(n-r)})$ (e.g., Park and Phillips, 1988). Thus it follows directly from part (b1) of Theorem 1 that

$$[R_G \hat{V} R_G']^{-1/2} \{R_G \text{vec}[T(\hat{G}^+ - G)] - R_G \text{vec}(b)\} \Rightarrow Z \equiv N(0, I_q), \tag{A.32}$$

where \hat{V} is \hat{S} when $e_1 \hat{\Sigma}_{\varepsilon\varepsilon} e_1'$ is replaced by $\Sigma_{1.2}$. Taken together (A.31) and (A.32) show the result in (11), whereas the independence statement holds because Z in (A.32) and J_C are jointly normal and uncorrelated. ■