Seminar Paper No. 556

INFERENCE IN COINTEGRATED VAR SYSTEMS

by

Anders Warne

INSTITUTE FOR INTERNATIONAL ECONOMIC STUDIES

University of Stockholm

ISSN 0347-8769

Seminar Paper No. 556

INFERENCE IN COINTEGRATED VAR SYSTEMS

by

Anders Warne

Seminar Papers are preliminary material circulated to stimulate discussion and critical comment.

November 1993

Institute for International Economic Studies
S-106 91 Stockholm
Sweden

Inference in Cointegrated VAR Systems

Anders Warne*

First version: February, 1992

This version: November 1993

Abstract

This paper examines the asymptotic properties of the Wald statistic in vector autoregressions that may have unit roots. Within this framework, we extend the distribution theory of Sims, Stock and Watson (1990) to nonlinear restrictions. It is shown that the limiting distribution of the Wald statistic is nonstandard when the restrictions concern the constant term or constrain the cointegration or common trends spaces. To exemplify the procedure, we study constraints derived from linear(ized) rational expectations models, i.e. nonlinear cross equation restrictions. A Monte Carlo study is performed and we find that the test statistic is somewhat oversized in small samples.

KEYWORDS: Cointegration, cross equation restrictions, unit roots, vector autoregression, Wald statistic.

JEL CLASSIFICATION NUMBERS: C12, C15, C32.

---

*Institute for International Economic Studies, Stockholm University, 10691 Stockholm, Sweden.

I began writing this paper and its companion "Inference in Vector Autoregressions with Integrated Time Series" while visiting the Institute of Economics at the University of Aarhus, Denmark. The hospitality of the members of the institute is gratefully acknowledged. In particular, many thanks to Niels Haldrup and Svend Hylleberg. For comments and suggestions I thank Peter Boswijk, Søren Johansen, Jim Stock and seminar participants at IIES, Institute of Mathematical Statistics at the University of Copenhagen, and the ESEM92 in Brussels. Tor Jacobson has checked some of the simulations (with more sophisticated software) and for that I owe him. Errors and obscurities are, of course, solely my responsibility. Financial support from *Humanistisk-Samhällsvetenskapliga Forskningsrådet* is gratefully acknowledged.

1 Introduction

It is well known that vector autoregressions (VAR's) are convenient and powerful tools to describe the behavior of a large number of economic time series. For some time it has been recognised that many economic time series appear to be nonstationary and in the past first differences of logs were frequently used in practise. As a response to the paper by Engle and Granger (1987) on the connection between VAR's and cointegration (i.e. Granger's representation theorem), econometricians have recently consecrated a great deal of effort to study long run relations for macroeconomic and financial variables within such a framework. In particular, much research has been devoted to estimating and testing restrictions on the cointegration and common trends spaces using various techniques (see e.g. Johansen and Juselius (1992), King, Plosser, Stock and Watson (1991), and Mellander, Vredin and Warne (1992)). Moreover, VAR's have frequently been used to test different neutrality concepts (as in Bergman and Warne (1993), Christiano and Ljungqvist (1988), and so on) as well as restrictions derived from dynamic economic models that are based on rational expectations (see, for instance, Campbell (1987), Campbell and Shiller (1987), and Ito (1988)). For tests of such hypotheses it is common knowledge that inference in statistical models with nonstationary time series is sometimes based on the wrong asymptotic distribution when the standard χ^2 is used. To cope with this difficulty, researchers have generally adjusted data by taking first differences and by using cointegration restrictions (see Campbell and Perron (1991) for a review on integration and cointegration).

The general distribution theory for testing linear restrictions in linear time series models that may have unit roots is given in a paper by Sims, Stock and Watson (1990). They show that the Wald statistic has a limiting representation which is constructed from functionals of a multivariate Brownian motion. Generally, its distribution is not the χ^2 but nonstandard so that application of the theory can be computationally demanding. Warne (1992) uses the Sims, Stock and Watson framework in a VAR setting (when data is jointly integrated at most of order one) to study the question how the asymptotic distribution of the Wald test depends on the restrictions. There it is shown that for all linear restrictions which constrain either the cointegration or the common trends space

or the constant term, the Wald test generally has a nonstandard limiting distribution. Moreover, for all other restrictions, the distribution is the usual χ^2 .

In this paper the objective is to extend the Sims, Stock and Watson (1990) theory to nonlinear restrictions using the model considered in Warne (1992). Although a VAR system (which is cointegrated of order (1,1) in the sense of Johansen (1991)) only is a special case of the general model studied in Sims, Stock and Watson, it is still rich enough to cover the most interesting cases. Furthermore, analysis within the general model is inconvenient (and sometimes even intractable) when we focus on the interpretation of the results. Specifically, the limiting distribution of the Wald test in such a model is far more complicated than the one we obtain in a VAR, and a study based on the former model would merely provide few (if any) additional insights. Moreover, the cases we consider here include a substantial number of empirical data models.

The paper is structured as follows. Section 2 presents the basic VAR that may have unit roots. The properties of the regressors are discussed in terms of an observationally equivalent representation of the conditional mean. For that model the regressors are labeled canonical as they are “cleverly” derived from a separation of the different forms of behavior which may be present in the original regressors. The asymptotic properties of the ordinary least squares (OLS) estimator are stated and discussed in section 3 (see also Warne (1992)). In section 4 we examine general restrictions on the coefficients of a VAR and an asymptotic representation of the Wald statistic is given. Section 5 deals with economic examples of nonlinear restrictions. In particular, we consider the type of nonlinear cross equation restrictions that are typically implied by exact linear rational expectations models (see Hansen and Sargent (1981)). It is shown that the limiting distribution of the Wald statistic is generally nonstandard since such restrictions constrain the common trends space when the VAR is subject to unit roots. In section 6 we study the small sample behavior of the test by Monte Carlo simulation. For a bivariate VAR with one lag we find that the test statistic is somewhat oversized. In addition, we make a comparison to the Toda and Yamamoto (1993) approach, where the test statistic has an asymptotic χ^2 distribution. The main conclusions are summarized in section 7. Finally, proofs of the propositions in the paper are given in the Appendix.

2 The Model

In this paper we shall analyse a standard VAR subject to some unit roots. The model is a special case of the class of linear time series models considered by Sims, Stock and Watson (1990). I shall make use of the Sims, Stock and Watson (SSW) notation extensively to facilitate the comparison between this paper and their general setting. Since the model is also discussed by Warne (1992) here we shall focus on its main properties.

Let $\{x_t\}$ denote an n dimensional real valued vector discrete time series generated by the VAR of finite order p :

$$x_t = \begin{bmatrix} A_1 & \cdots & A_p & | & \rho \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \vdots \\ x_{t-p} \\ 1 \end{bmatrix} + \varepsilon_t = \begin{bmatrix} b_1 & | & b_2 \end{bmatrix} Y_{t-1} + \varepsilon_t, \quad (2.1)$$

for $t = 1, \dots, T$. The A_k matrices are $n \times n$ and the vector ρ is n dimensional so that $b = [b_1 \ b_2]$ is an $n \times (np + 1)$ matrix with $b_2 = \rho$. The $n \times 1$ vector of disturbances $\{\varepsilon_t\}$ is assumed to be a sequence of martingale differences with $E[\varepsilon_t | \varepsilon_1, \dots, \varepsilon_{t-1}] = 0$ and $E[\varepsilon_t \varepsilon_t' | \varepsilon_1, \dots, \varepsilon_{t-1}] = \Sigma$, a nonsingular matrix. Furthermore, the $(np + 1) \times 1$ vector Y_0 is fixed (see e.g. Gallant (1987) or Helland (1982) for definitions).

Let $A(z) = I_n - \sum_{k=1}^p A_k z^k$, with I_n the $n \times n$ identity matrix and $\det[A(z)] = 0$ if and only if $|z| > 1$ or $z = 1$. This implies that $\{x_t\}$ is integrated of order d , with d being a nonnegative integer. Furthermore, let $A(1) = \gamma \alpha'$, where γ and α are $n \times r$ matrices with $r \in \{0, 1, \dots, n\}$ and $\text{rank}[A(1)] = r$. Let the $n \times (n - r)$ matrices $\gamma_\perp$ and $\alpha_\perp$ satisfy $\gamma'_\perp \gamma = 0$, $\gamma'_\perp \gamma_\perp = I_{(n-r)}$, $\alpha'_\perp \alpha = 0$ and $\alpha'_\perp \alpha_\perp = I_{(n-r)}$. Johansen (1991) shows that if the $(n - r) \times (n - r)$ matrix

$$\gamma'_\perp \left(\frac{dA(z)}{dz} \Big|_{z=1} \right) \alpha_\perp,$$

is nonsingular, then $\{\alpha' x_t\}$ and $\{\Delta x_t\}$ are integrated of order zero, while $\{x_t\}$ is integrated of order one. That is, Johansen's condition (together with the root assumptions above) implies that $A(z)$ has exactly $(n - r)$ unit roots and $(n(p - 1) + r)$ roots which lie outside the unit disc and, hence, $\{x_t\}$ is cointegrated of order (1,1) with r cointegration vectors, the columns of α .

Furthermore, from Granger's representation theorem (see Johansen (1991)) we find that x_t has the following two moving average representations:

$$\Delta x_t = \mu + C(L)\varepsilon_t, \quad (2.2)$$

and

$$x_t = \mu t + C(1)\xi_t + \tilde{C}(L)\varepsilon_t. \quad (2.3)$$

Here, I have assumed that $x_0 = 0$ and $\varepsilon_s = 0$ for $s = 0$. This is notationally convenient but not necessary to prove the results stated below. Concerning notation, L is the lag operator and Δ is the difference operator. Also, $C(z) = C(1) + (1 - z)\tilde{C}(z)$, where $C(z)$ is an infinite order matrix polynomial with $C(0) = I_n$ while $\tilde{C}(z)$ is absolutely summable. The random vector $\xi_t = \sum_{i=1}^t \varepsilon_i$ and $\text{rank}[C(1)] = (n - r)$. In fact, Johansen shows that

$$C(1) = \alpha_{\perp} \left(\gamma'_{\perp} \left[\sum_{k=1}^p k A_k \right] \alpha_{\perp} \right)^{-1} \gamma'_{\perp},$$

where $(dA(z)/dz)|_{z=1} = -\sum_{k=1}^p k A_k$. Hence, x_t is influenced by $(n - r)$ stochastic trends and these can be represented by $\gamma'_{\perp} \xi_t$. In addition, the $n \times 1$ vector $\mu = C(1)\rho$ so that x_t may be driven by linear deterministic trends as well as linear stochastic trends in those cases when it is nonstationary.

When there are unit roots in x_t , the generating process (as given in (2.3)) can be said to include both stationary and nonstationary elements. To derive the asymptotic distributions of the OLS estimator of b and of the Wald statistic for testing hypotheses about b , the SSW approach is to transform the original regressors, Y_{t-1} , so that these properties show up in the transformed or canonical regressors, denoted by Z_{t-1} .

To formalize the idea of canonical regressors we begin by observing that when $\{x_t\}$ is cointegrated of order (1,1) with r cointegration vectors, there exists exactly $(n(p - 1) + r)$ stationary canonical regressors and $(n - r)$ nonstationary canonical regressors (corresponding to the numbers of stable and unit roots). The first group includes the r cointegration relationships and $n(p - 1)$ lagged first differences. The second group contains the $(n - r)$ stochastic trends with drift. These latter regressors can be further divided into one group where each element is dominated by a stochastic trend and another group where the element is dominated by the linear deterministic trend.

In Warne (1992) it is shown how appropriate canonical regressors are obtained for the VAR in (2.1). Let $\mu = \alpha_{\perp} \mu^*$ and $\nu = \alpha_{\perp} \mu_{\perp}^* \nu^*$, where $\mu^{*'} \mu_{\perp}^* = 0$, $\mu_{\perp}^{*'} \mu_{\perp}^* = I_{(n-r-1)}$, and ν^* is an $(n-r-1) \times (n-r-1)$ nonsingular matrix. It can be shown that if $\mu \neq 0$ then the canonical regressor $Z_t^4 = \mu' x_t$ is dominated by a linear deterministic trend and that the canonical regressors $Z_t^3 = \nu' x_t$ are dominated by stochastic trends. In contrast, when $\mu = 0$ there does not exist a Z_t^4 regressor but instead $(n-r)$ Z_t^3 regressors with $\nu = \alpha_{\perp}$. By dominance it should be understood that it reflects the component which determines the long run or limiting behavior of the time series. Hence, although the time series $\{Z_t^4\}$ includes a random walk component it behaves (in the limit) as if it were a linear deterministic trend.

It is straightforward to rewrite the VAR in (2.1) by adding and subtracting terms such that

$$x_t = \rho + \sum_{j=1}^{p-1} A_j^* \Delta x_{t-j} + (I_n - A(1)) x_{t-1} + \varepsilon_t, \quad (2.4)$$

with $A_j^* = -\sum_{k=j+1}^p A_k$. From this expression it can be seen that there are exactly $n(p-1)$ regressors in first differences. We can therefore let the $(n(p-1)+r)$ mean zero stationary canonical regressors be given by $Z_t^1 = [(\Delta x_t - \mu)' \cdots (\Delta x_{t-p+2} - \mu)' (\alpha' x_t)']'$.

We thus have that the total number of stochastic canonical regressors is equal to np . Since Y_t contains $np+1$ elements where np are stochastic and 1 is a constant the final canonical regressor should also be a constant. If we let

$$U = \begin{bmatrix} \alpha & \nu & \mu \end{bmatrix}',$$

we can rewrite the VAR in (2.4) by using $(I_n - A(1))x_{t-1} = (I_n - A(1))U^{-1}Ux_{t-1}$ such that

$$\begin{aligned} x_t = & \sum_{j=1}^{p-1} A_j^* (\Delta x_{t-j} - \mu) + (\alpha(\alpha'\alpha)^{-1} - \gamma)\alpha' x_{t-1} + K\rho \\ & + \nu(\nu'\nu)^{-1}\nu' x_{t-1} + \mu(\mu'\mu)^{-1}\mu' x_{t-1} + \varepsilon_t, \end{aligned} \quad (2.5)$$

with $K = I_n + (\sum_{j=1}^{p-1} A_j^*)C(1)$. Unless $\rho = 0$, we find that $K\rho \neq 0$ and the transformed VAR therefore includes a constant as well as the stochastic canonical regressors discussed above. This final canonical regressor is denoted by $Z_t^2 = 1$. Furthermore, from (2.5) it can be seen that all the canonical regressors we discussed above appear as regressors on the right hand side. In fact, they are the only regressors in the transformed model.

To summarize, let k_i be the number of Z_t^i regressors. We then know that: (i) $k_1 = n(p-1) + r$, (ii) $k_2 = 1$, (iii) $k_3 = n - r - k_4$, whereas (iv) $k_4 = 1$ if $r \leq n-1$ and $\mu \neq 0$ while $k_4 = 0$ otherwise. It should be emphasized that the transformed model in (2.5) is observationally equivalent with respect to the conditional mean in equation (2.1). We shall use this equivalence when we study the asymptotic properties of the OLS estimator of b and of the Wald test of restrictions on b .

A more convenient form of writing the transformed model in equation (2.5) is often called for. To accomplish this, let us stack the canonical regressors such that $Z_t = [Z_t^{1'} \ Z_t^2 \ Z_t^{3'} \ Z_t^4]'$ with $Z_t = DY_t$. The $(np+1) \times (np+1)$ matrix D is nonsingular and given by

$$D = \left[\begin{array}{cccccc|c} I_n & -I_n & 0 & \cdots & 0 & 0 & -\mu \\ 0 & I_n & -I_n & & 0 & 0 & -\mu \\ \vdots & & \ddots & \ddots & & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & -I_n & 0 & -\mu \\ 0 & 0 & 0 & & I_n & -I_n & -\mu \\ \alpha' & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ \nu' & 0 & 0 & \cdots & 0 & 0 & 0 \\ \mu' & 0 & 0 & \cdots & 0 & 0 & 0 \end{array} \right] = \left[\begin{array}{c|c} D_1 & D_2 \\ 0 & 1 \\ D_3 & 0 \\ D_4 & 0 \end{array} \right]. \quad (2.6)$$

The matrix D_1 is thus $(n(p-1) + r) \times np$, D_2 is $(n(p-1) + r) \times 1$, D_3 is $k_3 \times np$, and the matrix D_4 is $k_4 \times np$. Accordingly, $Z_t^1 = [D_1 \ D_2]Y_t$, $Z_t^2 = [0 \ 1]Y_t$, $Z_t^3 = [D_3 \ 0]Y_t$, whereas $Z_t^4 = [D_4 \ 0]Y_t$. We then have that

$$x_t = cZ_{t-1} + \varepsilon_t, \quad (2.7)$$

where $c = bD^{-1}$. It is now straightforward to show that

$$\begin{aligned} c &= \left[\begin{array}{cccc|c|c|c|c} A_1^* & \cdots & A_{p-1}^* & (\alpha(\alpha'\alpha)^{-1} - \gamma) & K\rho & \nu(\nu'\nu)^{-1} & \mu(\mu'\mu)^{-1} \end{array} \right] \\ &= \left[\begin{array}{c|c|c|c} c_1 & c_2 & c_3 & c_4 \end{array} \right]. \end{aligned} \quad (2.8)$$

Here, the matrix c_i represents the coefficients on the canonical regressor Z_t^i for $i = 1, 2, 3, 4$. As these regressors have very different behavior we can expect that least squares estimates of the c_i 's will have fundamentally different asymptotic properties. Furthermore, we have

that $b_1 = c_1 D_1 + c_3 D_3 + c_4 D_4$ and $b_2 = c_1 D_2 + c_2$. Hence, the asymptotic behavior of the coefficients on the original regressors may, in turn, be expected to depend crucially on the asymptotic behavior of the coefficients on the canonical regressors.

3 Asymptotic Distribution of the OLS Estimator

In stacked form, the VAR model in equation (2.1) may be written as

$$X = bY + \epsilon = cZ + \Sigma^{1/2}f.$$

Here, $X = [x_1 \cdots x_T]$ is an $n \times T$ matrix, $Y = [Y_0 \cdots Y_{T-1}]$ is $(np + 1) \times T$, and $\epsilon = [\epsilon_1 \cdots \epsilon_T] = \Sigma^{1/2}[\varphi_1 \cdots \varphi_T] = \Sigma^{1/2}f$, where $\Sigma^{1/2}\Sigma^{1/2'} = \Sigma$ and $\Sigma^{1/2}$ is an $n \times n$ matrix. Let vec denote the column stacking operator, $\otimes$ the Kronecker product, $\xrightarrow{P}$ convergence in probability, $\xrightarrow{d}$ convergence in distribution, $\Rightarrow$ weak convergence of the associated probability measures in the sense of Billingsley (1968), $\mathcal{N}$ the (multivariate) normal distribution, and $B(u)$ an n dimensional standard Brownian motion (Wiener process) defined on the n dimensional unit interval. Below, the integral $\int_0^1 B(u)B(u)'du$ is an $n \times n$ matrix of stochastic variables found as ordinary Riemann integrals of the continuous functions $B_i(u)B_j(u)$, while the integral $\int_0^1 dB(u)B(u)'$ is an $n \times n$ matrix where the (i, j) :th element is given by a stochastic integral of the process B_j with respect to the process B_i . Naturally, $\int_0^1 dB(u) = B(1) \sim \mathcal{N}(0, I_n)$.

The stacked “single equations form” of the VAR model is given by

$$\begin{aligned} x &= (Y' \otimes I_n)\beta + \epsilon \\ &= (Z' \otimes I_n)\delta + (I_T \otimes \Sigma^{1/2})\varphi. \end{aligned} \tag{3.1}$$

We here have that the $n(np + 1) \times 1$ vector $\beta = \text{vec}(b)$, $\delta = \text{vec}(c)$, $x = \text{vec}(X)$, $\epsilon = \text{vec}(\epsilon)$, and $\varphi = \text{vec}(f)$. Also, $\beta_i = \text{vec}(b_i)$ and $\delta_j = \text{vec}(c_j)$. The OLS estimator of β is now given by

$$\begin{aligned} \hat{\beta} &= [(YY')^{-1} \otimes I_n][Y \otimes I_n]x \\ &= [D' \otimes I_n][(ZZ')^{-1} \otimes I_n][Z \otimes I_n]x \\ &= [D' \otimes I_n]\hat{\delta}. \end{aligned} \tag{3.2}$$

Hence, the stochastic behavior of $\hat{\beta}$ is fully described by $\hat{\delta}$. It therefore suffices to examine the asymptotic distribution of the OLS estimator of δ to obtain the asymptotic distribution of $\hat{\beta}$.

From equations (3.1) and (3.2) we now have that:

$$\hat{\delta} - \delta = [(ZZ')^{-1} \otimes I_n] [Z \otimes \Sigma^{1/2}] \varphi. \quad (3.3)$$

Given the choice of a finite order VAR process, the following special case of Condition 1 in SSW is sufficient to assure weak convergence of the relevant random matrices in (3.3).

Assumption 3.1 *There exists some $\kappa_4 < \infty$ such that $E[\varphi_{i,t}^4] < \kappa_4$ for all $i \in \{1, \dots, n\}$.*

Here, $\varphi_{i,t}$ is the i :th element of φ_t where $\varepsilon_t = \Sigma^{1/2}\varphi_t$. Hence, $E[\varphi_t\varphi_t'|\varphi_1, \dots, \varphi_{t-1}] = I_n$.

Finally, to handle the different orders of convergence of the canonical regressors, let us define the $(np + 1) \times (np + 1)$ scaling matrix Υ_T by

$$\Upsilon_T = \begin{bmatrix} T^{1/2}I_{k_1} & 0 & 0 & 0 \\ 0 & T^{1/2} & 0 & 0 \\ 0 & 0 & TI_{k_3} & 0 \\ 0 & 0 & 0 & T^{3/2}I_{k_4} \end{bmatrix}.$$

Thus, we expect that sums of squares of the stationary and constant canonical regressors will converge at the standard rate T^{-1} , that sums of squares of the canonical regressors which are dominated by stochastic trends will converge at the rate T^{-2} , and that the sum of squares of the canonical regressor which is dominated by a linear deterministic trend will converge at the rate T^{-3} .

We can now state the following version of Theorem 1 in SSW (see also Warne (1992)):

Proposition 3.1 *Under Assumption 3.1,*

$$[\Upsilon_T \otimes I_n] (\hat{\delta} - \delta) = \begin{bmatrix} T^{1/2}(\hat{\delta}_1 - \delta_1) \\ T^{1/2}(\hat{\delta}_2 - \delta_2) \\ T(\hat{\delta}_3 - \delta_3) \\ T^{3/2}(\hat{\delta}_4 - \delta_4) \end{bmatrix} \Rightarrow [V^{-1} \otimes I_n] \phi, \quad (3.4)$$

with $\phi = [\phi_1' \phi_2' \phi_3' \phi_4']'$, where $\phi_1 \sim \mathcal{N}(0, V_{11} \otimes \Sigma)$ is $n(n(p-1) + r) \times 1$, $\phi_2 \sim \mathcal{N}(0, \Sigma)$ is $n \times 1$, the $nk_3 \times 1$ random vector $\phi_3 = [\nu' C(1) \Sigma^{1/2} \otimes \Sigma^{1/2}] \text{vec}[\int_0^1 dB(u) B(u)']$, and the

$nk_4 \times 1$ vector $\phi_4 \sim \mathcal{N}(0, ((\mu'\mu)^2/3)\Sigma)$, and where ϕ_1 is independent of ϕ_j for $j = 2, 3, 4$. Also, $E[\phi_2\phi_4'] = (\mu'\mu/2)\Sigma$ so ϕ_2 and ϕ_4 are not independent. Moreover,

$$V^{-1} = \begin{bmatrix} V_{11}^{-1} & 0 & 0 & 0 \\ 0 & V_{22}^+ & V_{23}^+ & V_{24}^+ \\ 0 & V_{23}^{+'} & V_{33}^+ & V_{34}^+ \\ 0 & V_{24}^{+'} & V_{34}^{+'} & V_{44}^+ \end{bmatrix},$$

where $T^{-1} \sum_{t=1}^T Z_{t-1}^1 Z_{t-1}^{1'} \xrightarrow{P} V_{11}$, a deterministic and nonsingular matrix. Furthermore, for (a) $r < n - 1$ and $\mu \neq 0$,

$$\begin{aligned} V_{22}^+ &= \left[\Xi + \frac{6}{\mu'\mu} \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right] \right]' C(1)' \nu V_{33}^+ \nu' C(1) \left[\Xi + \frac{6}{\mu'\mu} \left[\frac{\mu'\mu}{2} \Xi + \Lambda \right] \right] + 4 \\ V_{23}^+ &= - \left[\Xi + \frac{6}{\mu'\mu} \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right] \right]' C(1)' \nu V_{33}^+ \\ V_{24}^+ &= - \frac{12}{(\mu'\mu)^2} \left[\Xi + \frac{6}{\mu'\mu} \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right] \right]' C(1)' \nu V_{33}^+ \nu' C(1) \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right] - \frac{6}{\mu'\mu} \\ V_{33}^+ &= \left\{ \nu' C(1) \left[\Psi - \Xi \Xi' - \frac{12}{(\mu'\mu)^2} \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right] \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right]' \right] C(1)' \nu \right\}^{-1} \\ V_{34}^+ &= \frac{12}{(\mu'\mu)^2} V_{33}^+ \nu' C(1) \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right] \\ V_{44}^+ &= \frac{12}{(\mu'\mu)^2} + \frac{144}{(\mu'\mu)^4} \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right]' C(1)' \nu V_{33}^+ \nu' C(1) \left[\frac{\mu'\mu}{2} \Xi - \Lambda \right]. \end{aligned}$$

(b) For $r = n - 1$ and $\mu \neq 0$ we have that $V_{22}^+ = 4$, $V_{24}^+ = -[6/(\mu'\mu)]$, $V_{44}^+ = [12/(\mu'\mu)^2]$, while V_{i3}^+ are empty matrices. (c) For $r \leq n - 1$ and $\mu = 0$

$$\begin{aligned} V_{22}^+ &= 1 + \Xi' C(1)' \nu V_{33}^+ \nu' C(1) \Xi, \\ V_{23}^+ &= -\Xi' C(1)' \nu V_{33}^+, \\ V_{33}^+ &= \{ \nu' C(1) [\Psi - \Xi \Xi'] C(1)' \nu \}^{-1}, \end{aligned}$$

where V_{i4}^+ are empty matrices. (d) For $r = n$, $V_{22}^+ = 1$ while V_{ij}^+ are empty matrices for $j = 3, 4$. Finally, the random vector $\Xi = \Sigma^{1/2} [\int_0^1 B(u) du]$, while the random matrix $\Psi = \Sigma^{1/2} [\int_0^1 B(u) B(u)' du] \Sigma^{1/2'}$, and the random vector $\Lambda = \Sigma^{1/2} [\int_0^1 u B(u) du]$.

Accordingly, the estimated coefficients on the stationary canonical regressors converge in distribution to a Gaussian with mean δ_1 and covariance matrix $(V_{11}^{-1} \otimes \Sigma)$ at the rate $T^{1/2}$ whereas the other estimated coefficients generally converge weakly to random variables which have nonstandard distributions. However, if there are no regressors which

are dominated by stochastic trends and $\{x_t\}$ is nonstationary (part (b)), the limiting distribution of the OLS estimator of δ is normal. This is, of course, in full correspondence with the results obtained by for example West (1988). It thus seems fair to claim that the root to the wicked expressions in Proposition 3.1 lies in Z_t^3 , the canonical regressors which are dominated by stochastic trends.

Let us now turn to the asymptotic distribution of the OLS estimator of β . We then find that

Proposition 3.2 *Under Assumption 3.1 and $k_1 > 0$,*

$$T^{1/2} \begin{bmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{bmatrix} \Rightarrow \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}, \quad (3.5)$$

where the random vector $\zeta_1 \sim \mathcal{N}(0, D_1' V_{11}^{-1} D_1 \otimes \Sigma)$ is $n^2 p$ dimensional, and $\zeta_2 = (D_2' V_{11}^{-1} \otimes I_n) \phi_1 + V_{22}^+ \phi_2 + (V_{23}^+ \otimes I_n) \phi_3 + V_{24}^+ \phi_4$ is n dimensional. Moreover, if $k_1 = 0$ and $k_3 > 0$, then

$$T (\hat{\beta}_1 - \beta_1) \Rightarrow [D_3' \otimes I_n] \{ [V_{23}^{+'} \otimes I_n] \phi_2 + [V_{33}^+ \otimes I_n] \phi_3 + [V_{34}^+ \otimes I_n] \phi_4 \}, \quad (3.6)$$

while $\zeta_2 = V_{22}^+ \phi_2 + (V_{23}^+ \otimes I_n) \phi_3 + V_{24}^+ \phi_4$. Finally, if $k_1 = k_3 = 0$, then

$$T^{3/2} (\hat{\beta}_1 - \beta_1) \Rightarrow \zeta_1^* \sim \mathcal{N}(0, (48/\mu^2) \Sigma), \quad (3.7)$$

and $\zeta_2 \sim \mathcal{N}(0, 4\Sigma)$ while $E[\zeta_1^* \zeta_2] = -(6/\mu) \Sigma$.

An interesting observation we can make from Proposition 3.2 is that the OLS estimator of the autoregressive coefficients always has a limiting normal distribution unless $p = 1$, $r = 0$ and $n \geq 2$ or $\mu = 0$. In that case

$$x_t = \rho + x_{t-1} + \varepsilon_t,$$

where $\mu = \rho$. This is clearly a very unusual case and we shall not discuss it further.

The asymptotic covariance matrix of $\hat{\beta}_1$ is generally singular. In particular,

$$\text{rank}[(D_1' V_{11}^{-1} D_1 \otimes \Sigma)] = n(n(p-1) + r) \leq n^2 p,$$

with strict inequality whenever $r < n$. Furthermore, the OLS estimator of β_2 has a nonstandard limiting distribution when $k_3 > 0$. The reason for this stems from the observations that (i) the estimated coefficients on the canonical constant are asymptotically

correlated with the estimated coefficients on the nonstationary canonical regressors, Z_t^3 , while (ii) the stationary canonical regressors have mean zero so that no component of the estimated coefficients on these regressors can be asymptotically correlated with the estimated coefficients on the constant term. In fact, the (stochastic) matrix $[V \otimes \Sigma]$ may be regarded as the “covariance” matrix of ϕ . From this covariance matrix we find that ϕ_1 is uncorrelated with ϕ_j , $j \geq 2$, and that $E[\phi_2\phi_4'] = [\mu'\mu/2]\Sigma$. The result that ϕ_1 is also independent of ϕ_j is based on Theorem 2.2 of Chan and Wei (1988). However, ζ_1 and ζ_2 are not independent since both vectors are functions of ϕ_1 . As long as we limit our attention to the autoregressive coefficients in (2.1) the asymptotic correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$ will not lead to inferential difficulties. Finally, with $\beta_1 = (D_1' \otimes I_n)\delta_1 + (D_3' \otimes I_n)\delta_3 + (D_4' \otimes I_n)\delta_4$, the superconsistency of the estimated coefficients on the nonstationary canonical regressors implies that the limiting distribution of these coefficients does not influence the asymptotic distribution of the estimated β_1 coefficients (unless $k_1 = 0$).

4 Asymptotic Distribution of the Wald Statistic

Some of the hypotheses that attract interest from empirical researchers concern nonlinear restrictions on the coefficients of a VAR model. In section 5 we shall examine one such case. Here we shall analyse a larger family of restrictions and focus on the asymptotics while the economic motivation is given in the next section.

The linear restrictions on β which are examined in a companion paper by Warne (1992) are special cases of the general restrictions $F(\beta) = 0$. The vector function $F(\beta)$ is assumed to be real valued and of dimension q . Furthermore, we shall limit our attention to functions which satisfy the following:

Assumption 4.1 (i) $F(\beta)$ is continuously differentiable, (ii) the first order Taylor expansion

$$F(\hat{\beta}) - F(\beta) = \left. \frac{\partial F(\beta)}{\partial \beta'} \right|_{\beta=\hat{\beta}} (\hat{\beta} - \beta) + \varrho_{(\hat{\beta},\beta)}, \quad (4.1)$$

satisfies

$$\text{rank} \left[\left. \frac{\partial F(\beta)}{\partial \beta'} \right|_{\beta=\hat{\beta}} \right] = q, \quad (4.2)$$

with probability 1, and (iii) the q dimensional remainder term $\varrho_{(\hat{\beta},\beta)}$ is $o_p(T^{-1/2})$.

For linear functions of β that are well defined Assumption 4.1(ii) is always satisfied. However, it is possible that for other functions the $q \times n(np + 1)$ matrix with partial derivatives, $R(\beta) := (\partial F(\beta)/\partial \beta')$, has a reduced row rank without implying that some of the original restrictions are “generically” dependent. For example, if $R(\beta)$ is associated with a saddle point of some elements of $F(\beta)$, then $R(\hat{\beta})$ has rank less than q with probability 1. In other words, the linearized restrictions may be linearly dependent for certain β but not for other. Here we abstract from such F functions (see e.g. Andrews (1987) and Lütkepohl (1989)). Finally, part (iii) in Assumption 4.1 can be justified on the basis of being easy to check for economic specifications of $F(\beta)$. Although this assumption is rarely made directly, standard regularity conditions in nonlinear statistical models imply that the remainder term is $o_p(T^{-1/2})$ (see e.g. Gallant (1987, Theorem 7.12)).

The Wald statistic applied to test the null hypothesis, $F(\beta) = 0$, vs. the alternative, $F(\beta) \neq 0$, is given by:

$$W = F(\hat{\beta})' [R(\hat{\beta}) [(YY')^{-1} \otimes \hat{\Sigma}] R(\hat{\beta})']^{-1} F(\hat{\beta}). \quad (4.3)$$

The estimator of the covariance matrix Σ is equal to $\hat{\Sigma} = T^{-1}\hat{\epsilon}\hat{\epsilon}'$, with $\hat{\epsilon} = X - \hat{b}Y = X - \hat{c}Z$, and Warne (1992) shows that $\hat{\Sigma}$ is consistent. In this section we shall present an asymptotic representation of the Wald statistic in equation (4.3) and sufficient conditions for that random variable to be distributed as χ_q^2 .

By the well known theorem of Mann and Wald (see e.g. Theorem 1.7 in Serfling (1980)) we have that if $\hat{\beta} \xrightarrow{P} \beta$, then $F(\hat{\beta}) \xrightarrow{P} F(\beta)$ and $R(\hat{\beta}) \xrightarrow{P} R(\beta)$. In other words, when the OLS estimator of β is consistent we know that the OLS estimator of the vector function F and the matrix with partial derivatives of F are consistent as well. As an immediate consequence of Proposition 3.2 and Assumption 4.1 we find that (for $k_1 > 0$)

$$T^{1/2} (F(\hat{\beta}) - F(\beta)) \Rightarrow R(\beta)\zeta, \quad (4.4)$$

where $\zeta = [\zeta_1' \ \zeta_2']'$. If ζ is normally distributed it follows that the conclusion in Theorem 3.3.A in Serfling (1980) holds true. Moreover, if the restrictions only concern the autoregressive coefficients in (2.1), i.e. $F(\beta) = F(\beta_1)$, we have that $R(\beta)\zeta$ is normally distributed.

It is convenient to partition the matrix of partial derivatives with respect to β as

follows

$$R(\beta) = \begin{bmatrix} R_1(\beta) & R_2(\beta) \end{bmatrix}. \quad (4.5)$$

The matrix $R_1(\beta) = (\partial F(\beta)/\partial \beta'_1)$ is $q \times n^2 p$ and $R_2(\beta) = (\partial F(\beta)/\partial \beta'_2)$ is $q \times n$. Since $\beta = (D' \otimes I_n)\delta$ we have that $R(\beta)\beta = P(\delta)\delta$ for:

$$P(\delta) := \frac{\partial F(\delta)}{\partial \delta'} = R([D' \otimes I_n]\delta)[D' \otimes I_n].$$

By equation (2.6) we then find that

$$\begin{aligned} P(\delta) &= \begin{bmatrix} R_1(D'_1 \otimes I_n) + R_2(D'_2 \otimes I_n) & R_2 & R_1(D'_3 \otimes I_n) & R_1(D'_4 \otimes I_n) \end{bmatrix} \\ &= \begin{bmatrix} P_1(\delta) & P_2(\delta) & P_3(\delta) & P_4(\delta) \end{bmatrix}, \end{aligned} \quad (4.6)$$

where $R_i \equiv (\partial F(\beta)/\partial \beta'_i)$. Accordingly, the $P_j(\delta)$ matrices contain the partial derivatives of F with respect to δ_j and $P_j(\delta)$ is therefore a $q \times nk_j$ matrix. As discussed in SSW and in Warne (1992) concerning linear restrictions, these $P_j(\delta)$ matrices are generally not convenient to use when we derive the limiting properties of the Wald statistic. Instead, let $H(\delta)$ be a $q \times q$ nonsingular matrix and let $\tilde{P}(\delta) = H(\delta)P(\delta)$. The matrix $H(\delta)$ is chosen such that $\tilde{P}(\delta)$ is upper block triangular, i.e.

$$\tilde{P}(\delta) = \begin{bmatrix} \tilde{P}_{11}(\delta) & \tilde{P}_{12}(\delta) & \tilde{P}_{13}(\delta) & \tilde{P}_{14}(\delta) \\ 0 & \tilde{P}_{22}(\delta) & \tilde{P}_{23}(\delta) & \tilde{P}_{24}(\delta) \\ 0 & 0 & \tilde{P}_{33}(\delta) & \tilde{P}_{34}(\delta) \\ 0 & 0 & 0 & \tilde{P}_{44}(\delta) \end{bmatrix}. \quad (4.7)$$

Here we have that the rank of the $q_j \times nk_j$ matrix $\tilde{P}_{jj}(\delta)$ is equal to q_j , and $q_j = \min\{\text{rank}[P_j(\delta)], q - \sum_{i=1}^{j-1} q_i\}$ for all $j \geq 3$, whereas the matrix $\tilde{P}_{11}(\delta)$ is $q_1 \times nk_1$ and $q_1 = \text{rank}[P_1(\delta)]$ while $q_2 = (\text{rank}[P_1(\delta) \ P_2(\delta)] - \text{rank}[P_1(\delta)])$. Accordingly, $q = \sum_{j=1}^4 q_j$ and the interpretation of q_j is that it represents the number of linearized restrictions which are well defined on δ_j but not on δ_{j-i} for $1 \leq i \leq j-1$.

In the limit some of the $\tilde{P}_{ij}(\delta)$ matrices with $i < j$ do not influence the limiting representation of W . The reason is, of course, that the OLS estimator of δ_3 and of δ_4 is superconsistent. Therefore, let

$$\tilde{P}(\delta) (\Upsilon_T^{-1} \otimes I_n) = \Upsilon_T^{*-1} P_T^*(\delta),$$

where the $q \times q$ matrix Υ_T^* is given by

$$\Upsilon_T^* = \begin{bmatrix} T^{1/2}I_{q_1} & 0 & 0 & 0 \\ 0 & T^{1/2}I_{q_2} & 0 & 0 \\ 0 & 0 & TI_{q_3} & 0 \\ 0 & 0 & 0 & T^{3/2}I_{q_4} \end{bmatrix}.$$

As in Warne (1992) and SSW the matrix Υ_T^* takes into account that well defined restrictions in $\tilde{P}(\delta)$ are associated with estimated coefficients which converge weakly at different rates. Let $P^*(\delta) = \lim_{T \rightarrow \infty} P_T^*(\delta)$. It now follows from the discussion above that if $\hat{\delta}$ is a consistent estimator of δ , then $P_T^*(\hat{\delta})$ is a consistent estimator of $P^*(\delta)$. Furthermore, by Assumption 4.1(ii) we have that

$$\text{rank} [P_T^*(\hat{\delta})] = \text{rank} [P^*(\delta)] = q,$$

with probability 1. It should be emphasized that if $F(\beta)$ does not satisfy the rank condition in Assumption 4.1, then the limiting behavior of the Wald statistic is generally unknown. However, if the estimated “covariance matrix” of the restrictions in (4.3) has rank q^* with probability 1 a Wald statistic based on its generalized inverse may have a limiting $\chi_{q^*}^2$ distribution (see Andrews (1987)). When $q^* < q$ and unknown such information is generally of little interest for a practitioner.

We can now state the main result concerning the asymptotic behavior of the Wald statistic in (4.3):

Proposition 4.1 *Under Assumptions 3.1, 4.1, and $F(\beta) = 0$,*

$$W \Rightarrow \left(P^*(\delta) [V^{-1} \otimes I_n] \phi \right)' \left[P^*(\delta) [V^{-1} \otimes \Sigma] P^*(\delta)' \right]^{-1} \left(P^*(\delta) [V^{-1} \otimes I_n] \phi \right), \quad (4.8)$$

where V^{-1} and ϕ are given in Proposition 3.1.

In relation to Theorem 2 in SSW we here find that under the additional conditions in Assumption 4.1 the Wald statistic still converges to a nondegenerate random variable. Moreover, the distribution of this variable depends on the stochastic matrix V and on the random vector ϕ . It then follows that the structure of the deterministic matrix $P^*(\delta)$ establishes the limiting distribution of the Wald statistic.

The following corollary provides sufficient conditions for the Wald statistic in (4.3) to have an asymptotic χ_q^2 distribution.

Corollary 4.1 *Under Assumptions 3.1, 4.1, $F(\beta) = 0$, $R_2(\beta) = 0$, and $\text{rank}[P_1(\delta)] = q$,*

$$W \Rightarrow \omega_1 \sim \chi_q^2, \quad (4.9)$$

where

$$\omega_1 = \left(P_1(\delta) \left[V_{11}^{-1} \otimes I_n \right] \phi_1 \right)' \left[P_1(\delta) \left[V_{11}^{-1} \otimes \Sigma \right] P_1(\delta)' \right]^{-1} \left(P_1(\delta) \left[V_{11}^{-1} \otimes I_n \right] \phi_1 \right).$$

First, the condition that $R_2(\beta) = 0$ is generally easy to check in practise. Since this matrix contains the partial derivatives of F with respect to β_2 , i.e. the constant vector ρ in (2.1), a sufficient (and for practical purposes necessary) condition is that the restrictions do not involve the constants. Second, a sufficient condition for $\text{rank}[P_1(\delta)] = q$ is that the restrictions do *not* constrain the column or row spaces of $A(1)$ and such restrictions are generally, as we shall see below, easy to spot. In addition, we shall show that this condition is by no means necessary. The only condition in Corollary 4.1 which may be troublesome from the practitioners' point of view is Assumption 4.1(ii). However, the difficulties that may occur due to such an assumption are not unique to the current setting but always arise in nonlinear models when the analysis is based on first order Taylor expansions.¹

Restrictions on the row or column spaces of $A(1)$ serve as an important reference point for understanding the limiting behavior of the Wald statistic. In particular, apart from restrictions on the constant term, they are the only restrictions that may lead to a nonstandard asymptotic distribution. Simple examples of such restrictions are that one of the columns or one of the rows of $A(1)$ is equal to zero. Let e_i be the i :th column of I_n . The hypothesis $A(1)e_j = 0$ gives us a restriction on the column space of $A(1)$ while $e_i'A(1) = 0$ is a restriction on the row space. Under both these two hypotheses, the rank of $A(1)$ is clearly less than n . Furthermore, with $A(1) = \gamma\alpha'$ and $r \geq 1$, the first hypothesis implies that $\alpha'e_j = 0$, i.e. $e_j \in \text{sp}(\alpha_\perp)$, while the second implies that $e_i'\gamma = 0$, i.e. $e_i \in \text{sp}(\gamma_\perp)$. In other words, the former restrictions concern the cointegration space and the latter deal with the common trends space. Below, we shall first discuss the restrictions on the column space of $A(1)$.

Without loss of generality, suppose $p = 1$. With $A(1) = I_n - A_1$ the restriction on α

---

¹See Gregory and Veall (1985) for a discussion on small sample problems when using the Wald statistic under linearizations of nonlinear restrictions.

can be written

$$\begin{aligned} F(\beta) &= \begin{bmatrix} A_1 & \rho \end{bmatrix} \begin{bmatrix} e_j \\ 0 \end{bmatrix} - e_j \\ &= \begin{bmatrix} (e'_j \otimes I_n) & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} - e_j = 0. \end{aligned}$$

Accordingly, $R_1 = (e'_j \otimes I_n)$ has rank $q = n$ while $R_2 = 0$. Since $P = R[D' \otimes I_n]$ and $P_1 = R_1[D'_1 \otimes I_n]$ with $D'_1 = \alpha$ it follows that

$$P_1 = [e'_j \alpha \otimes I_n] = 0,$$

under the hull hypothesis. Furthermore, if $k_3 > 0$ then

$$P_3 = [e'_j \nu \otimes I_n],$$

while if $k_4 = 1$ then $P_4 = e'_j \mu I_n$. Since $\nu_i \in \text{sp}(\alpha_\perp)$ for all $i = 1, \dots, k_3$, it follows that $e'_j \nu \neq 0$ and $\text{rank}[P_3] = q = n$. Accordingly, we may let $H(\delta) = I_n$ and by Proposition 4.1 it follows that the limiting distribution of the Wald statistic is nonstandard as long as $k_3 > 0$.

Next, for the n restrictions on the row space of $A(1)$, i.e. $e'_i A(1) = 0$, we find that the null hypothesis can be written

$$F(\beta) = \begin{bmatrix} (I_n \otimes e'_i) & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} - e_i = 0.$$

We thus have that $R_1 = [I_n \otimes e'_i]$ while $R_2 = 0$. Furthermore, $\text{rank}[R_1] = q = n$.

With $P_j = R_1[D'_j \otimes I_n]$ for $j = 1, 3, 4$ and $P_2 = R_2 = 0$ we have that

$$P = \begin{bmatrix} (\alpha \otimes e'_i) & 0 & (\nu \otimes e'_i) & \mu e'_i \end{bmatrix}.$$

From this it can be seen that $\text{rank}[P_1] = r$, whereas $\text{rank}[P_3] = (n - r - k_4)$ and $\text{rank}[P_4] = k_4$. Suppose that $\mu \neq 0$ and $r < n - 1$, i.e. $k_4 = 1$ and $k_3 > 0$. We may then choose $H(\delta) = (U')^{-1}$ and

$$\tilde{P} = \begin{bmatrix} I_r \otimes e'_i & 0 & 0 & 0 \\ 0 & 0 & I_{(n-r-1)} \otimes e'_i & 0 \\ 0 & 0 & 0 & e'_i \end{bmatrix} = P^*.$$

From Proposition 4.1 it follows that the Wald statistic has a nonstandard limiting distribution which, among other things, depends on the number of cointegration vectors.

The conclusions from these two examples are also valid for more general restrictions on the row or column space of $A(1)$. In particular, if $k_3 > 0$ then the Wald statistic “always” has a nonstandard limiting distribution if the restrictions concern the cointegration or common trends space. (The word always is here given in quotation marks for reasons that will be explained below.) Apart from restrictions on ρ , for all other cases the asymptotic distribution of the test statistic is χ_q^2 . In section 5 we shall study nonlinear cross equation restrictions. Based on our analysis thus far a major concern will be if such constraints restrict the cointegration or common trends space. Furthermore, if they do then we would like to know what the dimension of the relevant space is and, preferably, its “coefficients”. This information will be vital in order to simplify inference based on nonstandard distributions.

Let us now consider why restrictions on the column or row spaces of $A(1)$ do not necessarily lead to nonstandard asymptotic inference. Toda and Yamamoto (1993) show that if we know the maximum order of integration, $d_{\max}$, and the maximum lag length, $p_{\max}$, it is always possible to use χ^2 based inference as long as the null hypothesis does not involve ρ . Their general idea is that by increasing the lag length from $p_{\max}$ to $(p_{\max} + d_{\max})$, the additional regressors should fully capture the nonstationarity properties. Since we know that the true value of the coefficients on these additional regressors is zero, we only need to test the null on the coefficients of the first $p_{\max}$ lags.

To formalize this idea and show that Theorem 1 in Toda and Yamamoto is, in fact, a special case of Corollary 4.1, we here have that $d_{\max} = 1$. In addition, let $p_{\max} = p - 1$, i.e. we know that $A_p = 0$. Let $\beta_{11} = \text{vec}([A_1 \cdots A_{p-1}])$ and $\beta_{1p} = \text{vec}(A_p)$ so that $\beta'_1 = [\beta'_{11} \ \beta'_{1p}]$. Suppose the restrictions only concern the autoregressive coefficients of the VAR. Since $\beta_{1p} = 0$ it immediately follows that $F(\beta) = F(\beta_{11})$. Accordingly,

$$R_1(\beta) = \begin{bmatrix} \frac{\partial F(\beta)}{\partial \beta'_{11}} & 0 \end{bmatrix} = \begin{bmatrix} R_{11}(\beta) & 0 \end{bmatrix}.$$

Under Assumption 4.1 $R_{11}(\beta)$ is a $q \times n^2(p-1)$ matrix with full row rank. Since $R_2(\beta) = 0$ we have that

$$R(\beta) = \begin{bmatrix} R_{11}(\beta) & 0 & 0 \end{bmatrix}.$$

Now, for $j = 1, 3, 4$, let us define the matrices $D_{j,i}$, $i = 1, 2$, from D_j such that $D_{j,1}$ contains the first $n(p-1)$ columns of D_j and $D_{j,2}$ the remaining n columns. From equation (2.6) we find that the D matrix can be written

$$D = \begin{bmatrix} D_{1,1} & D_{1,2} & D_2 \\ 0 & 0 & 1 \\ D_{3,1} & 0 & 0 \\ D_{4,1} & 0 & 0 \end{bmatrix},$$

with $D_{1,1}$ being an $(n(p-1) + r) \times n(p-1)$ matrix while $D_{1,2}$ is $(n(p-1) + r) \times n$. Note that $D_{1,1}$ has full column rank while D_1 has full row rank. As we shall see below, this turns out to be crucial.

From the definition of $P(\delta)$ we now find that

$$P_1(\delta) = R_{11}(\beta) [D'_{1,1} \otimes I_n].$$

Since $D'_{1,1}$ and $R_{11}(\beta)$ both have full row rank, it follows that $\text{rank}[P_1(\delta)] = q$. Hence, the conditions in Corollary 4.1 are satisfied and the Wald statistic therefore has an asymptotic χ^2 distribution. Moreover, this shows that restrictions which constrain the row or column spaces of $A(1)$ need not result in a nonstandard limiting distribution.

More generally, we can deduce from this analysis that for all hypotheses such that the restrictions neither involve the constant vector, ρ , nor *all* A_k matrices, the Wald statistic has a limiting χ^2 distribution. One immediate implication is, of course, that for a standard lag length hypothesis, e.g. $A_p = 0$, the usual χ^2 based inference applies.

From these results one may be inclined, like Toda and Yamamoto (1993), to conclude that as long as we know the maximum order of integration, a good strategy for testing the hypothesis of interest is to (i) determine the maximum lag length (e.g. by standard lag length tests), and (ii) estimate a VAR with $p = (p_{\max} + d_{\max})$ lags and compute the test statistic based on the estimated coefficients on the first $p_{\max}$ lags. While the theory tells us that the Wald statistic has an asymptotic χ^2 distribution, we also know that variable addition increases the asymptotic standard errors of the estimated autoregressive coefficients. Hence, the Toda and Yamamoto approach may suffer from low power in small samples. In addition, the limiting distribution may still be χ^2 when the test is based on estimates from a VAR with $p_{\max}$ lags.

Let us now briefly examine what happens in case the null hypothesis is not consistent with the coefficients of the VAR. In the Appendix, the following result is proven:

Proposition 4.2 *Under Assumptions 3.1, 4.1, and $F(\beta) \neq 0$, the probability that the Wald statistic in (4.3) is less than or equal to any finite real number W^* converges to zero at all rates greater than T^{-2d} , where $d \in \{1/2, 1, 3/2\}$ and is determined from the T -factor in $\Upsilon_T^* H(\delta) F(\beta)$ with nonzero coefficient for which the exponent is the largest.*

In other words, the probability of accepting a false null hypothesis often converges to zero at a faster rate when the VAR model is subject to unit roots than when $r = n$. This suggests that the power of the test in small samples may be greater in nonstationary than in stationary VAR systems. On the other hand, it is difficult to compute the (nonstandard) asymptotic distribution with accuracy for nonstationary VAR's (see Jacobson and Larsson (1991)) and, hence, the uncertainty about the estimated asymptotic critical values may be a cause for concern. We shall return to size and power issues in section 6.

5 Nonlinear Cross Equation Restrictions

The rational expectations hypothesis has become widely used in the theoretical and empirical macro and finance literature over the past 20 years. It is well known that (linear) economic models which include such an assumption usually imply highly nonlinear cross equation (NCE) restrictions on the coefficients of a VAR (see Hansen and Sargent (1981) and the review by Baillie (1989)). For example, the foreign exchange market efficiency hypothesis and the expectations hypothesis of the term structure of interest rates are generally associated with such constraints (see e.g. Campbell and Shiller (1987), Ito (1988), Sargent (1979), and Warne (1990)). In Warne (1992) the case of linear cross equation (LCE) constraints is studied in some detail and the conclusion there is that the Wald test is typically associated with a nonstandard limiting distribution when such restrictions are tested in VAR's subject to unit roots. Moreover, the intuition behind this result is that CE restrictions, as the terminology suggests, reduce the forms of nonstationarity which are possible under the null by constraining the row space of b (and thus of $A(1)$).

A purpose here is to examine whether this conclusion is extended to the nonlinear case. Given what we know about the LCE restrictions, we certainly expect to find that NCE

constraints are also (generally) associated with a nonstandard asymptotic distribution of the Wald test. Furthermore, I shall discuss the importance of Assumption 4.1 in more detail.

In order to construct a useful representation of the VAR system for computing conditional expectations, let the $n \times np$ matrix $J = [I_n \ 0 \ \cdots \ 0]$ while the $np \times np$ matrix A is defined by

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_n & 0 & \cdots & 0 & 0 \\ & \ddots & & & \vdots \\ & & \ddots & & \vdots \\ 0 & 0 & & I_n & 0 \end{bmatrix}.$$

Using this new notation we can rewrite equation (2.1) as

$$x_{t+i} = \sum_{j=0}^{i-1} JA^j J' \rho + JA^i Y_t^* + \sum_{j=0}^{i-1} JA^j J' \varepsilon_{t+i-j}, \quad (5.1)$$

where $Y_t = [Y_t^{*'} \ 1']'$, $t \geq 0$ is fixed, and $i \geq 1$. With $\mathcal{E}_t := \{x_s\}_{s=1-p}^t$, the history of x at t , it can be seen from (5.1) that

$$E[x_{t+i} | \mathcal{E}_t] = \sum_{j=0}^{i-1} JA^j J' \rho + JA^i Y_t^* = E[x_{t+i} | Y_t^*],$$

for all $i \geq 1$ and t . Consequently, there is no additional information in x_{t-s} for any $s \geq p$ (given the information in Y_t^*) which would improve our forecast of x_{t+i} . It then follows that any hypothesis about the prediction properties of the VAR model concerns the information set Y_t^* .

Exact linear rational expectations models, such as those considered by Hansen and Sargent (1981), always generate hypotheses which can be described by the following system of equations:

$$\begin{aligned} \sum_{i=0}^s N_i E[x_{t+i} | \mathcal{A}_t] &= \lambda, \\ \Gamma \lambda &= \lambda_0. \end{aligned} \quad (5.2)$$

The matrices N_i are $m \times n$ with $m \leq n$ and $\text{rank}[N_0 \ \cdots \ N_s] = m$, while the prediction horizon $s \geq 1$ and possibly infinite. Furthermore, the vectors λ and λ_0 are m and l dimensional, respectively, whereas Γ is $l \times m$ with $\text{rank}[\Gamma] = l$. Naturally, we assume that all these parameters are known to the econometrician. Note that if $l = 0$ then the

constant vector λ is unconstrained whereas $l = m$ implies that all elements of λ are given by economic theory. Also, the cases $0 < l < m$ occur when λ is only partially restricted. Finally, the information set which economic agents use to predict x_{t+i} at t , denoted by $\mathcal{A}_t$, is assumed to contain the history of x up to and including t . That is, $\mathcal{E}_t \subset \mathcal{A}_t$ for all t . Given the assumption that agents have rational expectations, as in equation (5.2), the additional assumption that their information set is at least as large as that of the econometrician seems quite harmless.²

To derive the NCE restrictions on the coefficients of the VAR system in (2.1) from the hypothesis in (5.2), note first that the law of iterated expectations implies that

$$\begin{aligned}\sum_{i=0}^s N_i E[x_{t+i}|Y_t^*] &= \lambda, \\ \Gamma \lambda &= \lambda_0.\end{aligned}\tag{5.3}$$

In contrast to equation (5.2), the hypothesis in (5.3) can be tested since the information set Y_t^* can be observed whereas $\mathcal{A}_t$ is partially unknown. Second, substituting the conditional expectation of x_{t+i} in equation (5.1) into (5.3) we obtain

$$\sum_{i=0}^s N_i J A^i Y_t^* = \lambda - \sum_{i=0}^s N_i \left(\sum_{j=0}^{i-1} J A^j J' \right) \rho.\tag{5.4}$$

For this to hold true for all t it is clear that

$$(i) \quad \sum_{i=0}^s N_i \left(\sum_{j=0}^{i-1} J A^j J' \right) \rho = \lambda,$$

and

$$(ii) \quad \sum_{i=0}^s N_i J A^i = 0.$$

It is important to note that (i) does not impose any restriction on ρ or A since λ may be free. If $l \geq 1$, however, then $\Gamma \lambda = \lambda_0$ provides us with

$$(iii) \quad \Gamma \sum_{i=0}^s N_i \left(\sum_{j=0}^{i-1} J A^j J' \right) \rho = \lambda_0.$$

---

²If $N_0 = 0$ then the top section of equation (5.2) may be interpreted as being derived from a rational expectations model with $\sum_{i=1}^s N_i E[x_{t+i}|\mathcal{A}_{t+1}] = \lambda + \eta_{t+1}$. In the terminology of Hansen and Sargent (1981), this hypothesis does not originate from an exact rational expectations model. However, if we assume that $\mathcal{A}_t \subset \mathcal{A}_{t+1}$ for all t and $E[\eta_{t+1}|\mathcal{A}_t] = 0$, the hypothesis can be transformed into (5.2) by conditioning on $\mathcal{A}_t$ instead of $\mathcal{A}_{t+1}$ (see Baillie (1989)).

This constrains the values which A and ρ can take on. In fact, (ii) and (iii) describe all the CE restrictions on the VAR which are implied by the rational expectations model. Accordingly, the number of restrictions is $q = (mnp + l)$. Employing the notation in section 4 we find that

$$F(\beta) = \begin{bmatrix} \text{vec}(\sum_{i=0}^s N_i J A^i) \\ \Gamma \sum_{i=0}^s N_i \left(\sum_{j=0}^{i-1} J A^j J' \right) \rho - \lambda_0 \end{bmatrix}. \quad (5.5)$$

The $(mnp + l) \times n(np + 1)$ matrix with partial derivatives can now be shown to be given by

$$R(\beta) = \sum_{i=0}^s \sum_{j=0}^{i-1} \begin{bmatrix} [(A')^{(i-1-j)} \otimes N_i J A^j J'] & | & 0 \\ \sum_{h=0}^{j-1} [\rho' J (A')^{(j-1-h)} \otimes \Gamma N_i J A^h J'] & | & \Gamma N_i J A^j J' \end{bmatrix}. \quad (5.6)$$

In general, this matrix has rank $q = (mnp + l)$ under the null. For example, suppose $\text{rank}[A] < np$ so that the VAR system in (2.1) is overparameterized. As an immediate consequence we find that the matrix $[(A')^{(i-1-j)} \otimes N_i J A^j J']$ has rank less than mnp for some i and j , whereas for other combinations of exponents the rank may be mnp .³ The double summation of these matrices, however, provides us with a matrix which has full row rank under the null.⁴ In fact, the only obvious situation when the rank of $R(\beta)$ is less than q occurs at $\rho = 0$ and when $\Gamma \sum_{i=0}^s N_i (\sum_{j=0}^{i-1} J A^j J')$ has reduced row rank. Whether this may at all be consistent with the null hypothesis is not known, but this is probably not the case.

Let us now show that part (iii) of Assumption 4.1 is satisfied. Let $F_1(\beta)$ denote the first mnp cross equation restrictions and $F_2(\beta)$ the remaining l restrictions. By application of

---

³For example, if N_1 has rank m , then for $i = 1$ and $j = 0$ the rank of this matrix is always mnp .

⁴To rigorously verify this claim for the general case is virtually impossible. However, at an intuitive level it may be realized from the observations that (i) the maximum prediction horizon, s , is independent of the lag order of the VAR, (ii) the N_i matrices form a system of m independent relationships, and that (iii) under the null these matrices also impose a rank condition on the cointegration space. In particular, the number of cointegration vectors can never be less than $\text{rank}[\sum_{i=0}^s N_i] \leq m$ (this is shown below). The first observations is directly related to the fact that too many lags in the VAR system does not imply that $R(\beta)$ has reduced row rank. The second and third observation mean that the rank of critical functions of the A_i matrices can never be less than m . I suggest that readers who are still in doubt try out simple examples to convince themselves or try to prove me wrong.

Lemma 1 in Schmidt (1973), and some tedious algebra, it follows that

$$\begin{aligned} F_1(\hat{\beta}) - F_1(\beta) &= \sum_{i=0}^s \sum_{j=0}^{i-1} \left[(A')^{i-1-j} \otimes N_i J \hat{A}^j J' \right] (\hat{\beta}_1 - \beta_1) \\ &= R_{1.1}(\beta, \hat{\beta}) (\hat{\beta}_1 - \beta_1). \end{aligned}$$

Let ϱ_1 denote the first mnp elements of the remainder term och ϱ_2 the final l elements. From the above expression and (5.6) we find that

$$\varrho_1 = \left[R_{1.1}(\beta, \hat{\beta}) - R_{1.1}(\hat{\beta}) \right] (\hat{\beta}_1 - \beta_1).$$

By Proposition 3.2 it follows that this vector converges to zero in probability at the rate $T^{1/2}$.

Next, let the final l rows of $R(\beta)$ be represented by the $l \times n^2 p$ matrix $R_{1.2}(\beta)$ and the $l \times n$ matrix $R_{2.2}(\beta)$. Some algebra later we end up with:

$$\begin{aligned} F_2(\hat{\beta}) - F_2(\beta) &= R_{2.2}(\hat{\beta}) (\hat{\beta}_2 - \beta_2) + \\ &\quad \sum_{i=0}^s \sum_{j=0}^{i-1} \sum_{h=0}^{j-1} \left[\rho' J (A')^{(j-1-h)} \otimes \Gamma N_i J \hat{A}^h J' \right] (\hat{\beta}_1 - \beta_1) \\ &= R_{2.2}(\hat{\beta}) (\hat{\beta}_2 - \beta_2) + R_{1.2}(\beta, \hat{\beta}) (\hat{\beta}_1 - \beta_1). \end{aligned}$$

Hence, the final l elements of the remainder term can be expressed as

$$\varrho_2 = \left[R_{1.2}(\beta, \hat{\beta}) - R_{1.2}(\hat{\beta}) \right] (\hat{\beta}_1 - \beta_1).$$

Again, by Proposition 3.2 we find that ϱ_2 converges to zero in probability at the rate $T^{1/2}$.

Since part (i) and (iii) and, in all likelihood, part (ii) of Assumption 4.1 are satisfied, we can now proceed and examine the $P(\delta)$ matrix. By equation (2.6) we find that the first nk_1 columns of this matrix are given by:

$$P_1(\delta) = \sum_{i=0}^s \sum_{j=0}^{i-1} \left[\begin{array}{c} [(A')^{(i-1-j)} D'_1 \otimes N_i J A^j J'] \\ \sum_{h=0}^{j-1} [\rho' J (A')^{(j-1-h)} D'_1 \otimes \Gamma N_i J A^h J'] + \Gamma N_i J A^j J' [D'_2 \otimes I_n] \end{array} \right].$$

Notice how the first mnp rows of this matrix resemble the P_1 matrix we obtained for the row space restrictions in section 4. In fact, if $s = 1$, $N_1 = e'_i$ (and $p = 1$), the above matrix is identical to the one we found earlier. Clearly, these mnp rows of $P_1(\delta)$ form a matrix with reduced row rank and, hence, we know that the Wald test has a

nonstandard asymptotic distribution whenever $k_3 > 0$.⁵ For example, if the constant vector is unconstrained (that is, $l = 0$), the fact that $P_1(\delta)$ has rank q_1 , with $q - q_1 = m(n - r)$, implies that the final condition in Corollary 4.1 is violated for all $r < n$. By Proposition 4.1 it then follows that the Wald statistic has a nonstandard asymptotic distribution unless $r = n - 1$ and $\mu \neq 0$ (i.e., when there is no canonical regressor which is dominated by stochastic trends). If that turns out to be the case, it can be shown that (even for $l \geq 0$) the Wald statistic has a limiting $\chi^2_{(mnp+l)}$ distribution under the null.

To shed some light on how to determine the parameters of the limiting representation of the Wald test of the NCE restrictions, note that the hypothesis in (5.3) can be written as:

$$\sum_{i=0}^s N_i E[x_{t+i}|Y_t^*] = \sum_{j=1}^s N_j^* E[\Delta x_{t+j}|Y_t^*] + N_0^* x_t, \quad (5.7)$$

with $N_j^* = \sum_{i=j}^s N_i$ for all $j \in \{0, 1, \dots, s\}$. The first term on the right hand side solely consists of linear relationships between stationary variables. The second term, however, can be rewritten as

$$N_0^* x_t = N_0^* \left[\alpha(\alpha'\alpha)^{-1} (\alpha'x_t) + \nu(\nu'\nu)^{-1} (\nu'x_t) + \mu(\mu'\mu)^{-1} (\mu'x_t) \right]. \quad (5.8)$$

If $\text{rank}[N_0^*] = n$, then the null hypothesis implies that $\{x_t\}$ is integrated of order zero. This follows from the observation that the first term on the right hand side in (5.7) is stationary. Hence, the second term must be stationary as well since the sum is equal to the constant vector λ . In fact, this line of reasoning asserts that $r \geq \text{rank}[N_0^*]$ is a *necessary* condition for the null hypothesis to be consistent with the VAR system in (2.1). Moreover, if the number of cointegration vectors is equal to the rank of N_0^* , then the cointegration space is fully described by the hypothesis in (5.7), i.e. $N_0^* = \alpha'$. Such a hypothesis can easily be tested with the statistic in (4.3) and the corresponding limiting distribution of W in Proposition 4.1 is nonstandard unless $k_3 = 0$. Furthermore, when $r > \text{rank}[N_0^*]$, it is still true that $N_0^{*\prime} \in \text{sp}(\alpha)$. Accordingly, $N_0^* \alpha_\perp = 0$ so that $N_0^* x_t = N_0^* \alpha(\alpha'\alpha)^{-1} (\alpha'x_t)$ by equation (5.8).

---

⁵In fact, the limiting representation of the Wald statistic (under the null) can be decomposed into one term which is distributed as $\chi^2_{q_1}$ and a second term which has a nonstandard distribution when $k_3 > 0$ and a $\chi^2_{(q-q_1)}$ distribution for $k_3 = 0$. Furthermore, these two terms are independent when there are no restrictions on ρ so the sum (generally) has a nonstandard distribution.

6 Size and Power in Finite Samples

In this section we shall examine small sample properties of the Wald statistic when the VAR is subject to some unit roots. We begin by setting up the simplest example of NCE restrictions and discuss what the constraints imply for the VAR coefficients. Next, we turn to the Monte Carlo results and compare size and power when inference is based on nonstandard distributions to the approach suggested by Toda and Yamamoto (1993).

Let us consider an example of NCE restrictions for what is commonly known as the unbiasedness or the foreign exchange market efficiency hypothesis. Let S_t and $F_t(2)$ denote the log of the spot and the 2-period forward rates between two currencies. The unbiasedness hypothesis then states that

$$E[S_{t+2}|\mathcal{A}_t] = F_t(2).$$

Let us assume that $x_t = [S_t \ F_t(2)]'$ and, for simplicity, we exclude the constant from the VAR. In terms of the notation in section 5 we now have that $m = 1$, $N_0 = [0 \ -1]$, $N_2 = [1 \ 0]$, while $N_1 = 0$. Furthermore, suppose that the true lag order is 1.

Below we shall study two cases. First, that a VAR(1) model is fitted, and, second, that a VAR(2) model is estimated. In the latter case we shall pursue the Toda and Yamamoto (1993) suggestion and only test the restrictions on the first lag. Both cases imply that $q = 2$ and that the NCE restrictions are described by

$$F(\beta) = \begin{bmatrix} a_{11}^2 + a_{12}a_{21} \\ a_{12}(a_{11} + a_{22}) - 1 \end{bmatrix} = 0.$$

Here, a_{ij} denotes the element in row i and column j of A_1 . For these constraints the 2×4 matrix of partial derivatives with respect to β' for the VAR(1) is given by

$$R(\beta) = \begin{bmatrix} 2a_{11} & a_{12} & a_{21} & 0 \\ a_{12} & 0 & a_{11} + a_{22} & a_{12} \end{bmatrix}.$$

For the Toda and Yamamoto approach, the above matrix is equal to $R_{11}(\beta)$ while $R(\beta)$ is 2×8 and the final 4 columns are zero. Since $a_{12} \neq 0$ is a *necessary* condition for the null to be true, the rank of $R(\beta)$ is clearly 2 when the NCE restrictions are satisfied. Hence, the restrictions $F(\beta)$ satisfy Assumption 4.1(ii) under the null as was claimed

above. Under alternatives, $R(\beta)$ has rank less than 2 if (i) $a_{11} = a_{12} = 0$, or if (ii) $a_{12} = 0$ and $a_{11} = -a_{22}$. Case (i) implies that the first restriction is satisfied but not the second, while case (ii) implies that the second restriction is violated and the first may be zero or nonzero.

If $\{x_t\}$ is integrated of order zero (i.e. $r = 2$), the Wald test obviously has an asymptotic χ^2_2 distribution under the null. However, if $r = 1$ we know that this need not be true. Our N_i matrices give rise to $N_0^* = [1 \ -1]$. When $\{x_t\}$ is subject to unit roots and the unbiasedness hypothesis is satisfied by the VAR coefficients, it follows that $N_0^* = \alpha'$. Furthermore, $N_1^* = N_2^* = [1 \ 0]$ and the stationary transformation in equations (5.7) and (5.8) implies that $\alpha'_\perp = [1 \ 1]$, while

$$E[\Delta S_{t+2}|Y_t^*] + E[\Delta S_{t+1}|Y_t^*] + (S_t - F_t(2)) = 0.$$

Accordingly, when $\{x_t\}$ is integrated of order one, the spot–forward spread must be stationary under the null. Moreover, letting γ_i denote the i :th element of the 2×1 vector γ , we have that:

$$A = A_1 = \begin{bmatrix} 1 - \gamma_1 & \gamma_1 \\ -\gamma_2 & 1 + \gamma_2 \end{bmatrix}.$$

A necessary and sufficient condition for this matrix to have exactly one unit eigenvalue and one eigenvalue inside the unit circle (i.e. for $A(z)$ not to have double unit roots or one unit root and one explosive root) is that $2 > \gamma_1 - \gamma_2 > 0$. Also, the null hypothesis implies that $\gamma_1 \neq 0$. Substituting for the cointegration consistent a_{ij} coefficients into $F(\beta)$ we obtain

$$F(\beta) = \begin{bmatrix} (1 - \gamma_1)^2 - \gamma_1\gamma_2 \\ \gamma_1(2 + \gamma_2 - \gamma_1) - 1 \end{bmatrix}.$$

These two functions are equal to zero when $\gamma_2 = (1 - \gamma_1)^2\gamma_1^{-1}$. In addition, the assumption that $A(z)$ does not contain explosive or additional unit roots implies that $\gamma_1 > 1/2$.

When $r = 1$ and $\alpha' = N_0^*$ we find that $R(\beta)$ has rank 2 under the null and the alternatives. The models we shall study below will therefore all be based on one cointegration vector which implies that the spot–forward spread is integrated of order zero. The matrix with partial derivatives of F with respect to δ' for the VAR(1) is given by:

$$P(\delta) = \begin{bmatrix} 2(1 - \gamma_1) + \gamma_2 & \gamma_1 & 2(1 - \gamma_1) - \gamma_2 & \gamma_1 \\ -2(1 - \gamma_1) - \gamma_2 & -\gamma_1 & 2 + \gamma_2 & \gamma_1 \end{bmatrix}.$$

Note that for any values of γ_1 and γ_2 , this matrix has rank 2. Furthermore, the first two columns of $P(\delta)$ give us $P_1(\delta)$ while the remaining two columns form $P_3(\delta)$. Thus, $\text{rank}[P_1(\delta)] = 1$ so that only one linearized restriction is well defined on the coefficients on the stationary canonical regressor. The remaining restriction is therefore well defined on the canonical regressor which is dominated by a stochastic trend. Let the 2×2 matrix $H(\delta)$ be lower triangular with unit elements on and below the diagonal. It now follows that

$$P^*(\delta) = \begin{bmatrix} 2(1 - \gamma_1) + \gamma_2 & \gamma_1 & 0 & 0 \\ 0 & 0 & 2(2 - \gamma_1) & 2\gamma_1 \end{bmatrix}.$$

According to Proposition 4.1 the limiting representation of the Wald statistic for the VAR(1) model is given by $\omega := \omega_1 + \omega_3$. The random variable $\omega_1 \sim \chi_1^2$ and is independent of the random variable ω_3 , which is represented by

$$\omega_3 := \frac{\phi_3' \tilde{P}_{33}(\delta)' \tilde{P}_{33}(\delta) \phi_3}{\left(\tilde{P}_{33}(\delta) \Sigma \tilde{P}_{33}(\delta)' \right) (\alpha_\perp' C(1) \Psi C(1)' \alpha_\perp)}. \quad (6.1)$$

The parameters are given by $\tilde{P}_{33}(\delta) = [2(2 - \gamma_1) \ 2\gamma_1]$, $\alpha_\perp' = [1 \ 1]$, while

$$C(1) = \frac{1}{\gamma_1 - \gamma_2} \begin{bmatrix} -\gamma_2 & \gamma_1 \\ -\gamma_2 & \gamma_1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix},$$

and $\Sigma^{1/2}$ is the lower triangular Choleski matrix. In the simulations below we assume that ε_t is i.i.d. $\mathcal{N}(0, \Sigma)$. Also, to estimate the limiting distribution of ω we replace ϕ_3 and Ψ with

$$\phi_{3,S} := \left[\alpha_\perp' C(1) \Sigma^{1/2} \otimes \Sigma^{1/2} \right] \text{vec} \left(S^{-1} \sum_{s=1}^S \varphi_s \Phi_{s-1}' \right),$$

and

$$\Psi_S := \Sigma^{1/2} \left[S^{-2} \sum_{s=1}^S \Phi_{s-1} \Phi_{s-1}' \right] \Sigma^{1/2'}.$$

Here, $\Phi_0 = 0$, $\Phi_s = \sum_{j=1}^s \varphi_j$, and S is the step length. In the simulations we let $S = 800$ and use the random number generator in RATS386, version 3.11.⁶

In the Toda and Yamamoto case, the 2×6 matrix $P_1(\delta)$ is equal to

$$P_1(\delta) = \begin{bmatrix} 2(1 - \gamma_1) & \gamma_1 & -\gamma_2 & 0 & 2(1 - \gamma_1) + \gamma_2 & \gamma_1 \\ \gamma_1 & 0 & (2 + \gamma_2 - \gamma_1) & \gamma_1 & -2(1 - \gamma_1) - \gamma_2 & -\gamma_1 \end{bmatrix}.$$

---

⁶Tor Jacobson, Department of Statistics, University of Uppsala, has also simulated the limiting distribution of ω using FORTRAN 77 and the standard normal variates were generated from a ratio-of-uniforms routine. His results are fully in line with those presented below.

Also, the 2×2 matrix $P_3(\delta)$ is equal to the $P_3(\delta)$ matrix in the VAR(1). Clearly, we now find that $\text{rank}[P_1(\delta)] = 2$ under the null (and under the alternatives) so that the Wald statistic has a limiting χ_2^2 distribution.

The basic models we shall examine are denoted by M_i , $i = 1, 2, 3, 4$. In addition, for each basic model we consider one null model and 3 alternative models. We use the notation $M_{i,j}$, where $j = 0, 1, 2, 3$. Specifically,

$$\begin{aligned} M_{1,j} &= \left\{ \gamma_1 = 3/4; \gamma_2 = (1/12) - \tau_j; \sigma_1^2 = .001; \sigma_2^2 = .001; \sigma_{12} = 0 \right\}, \\ M_{2,j} &= \left\{ \gamma_1 = .51; \gamma_2 = \frac{(.49)^2 - .75\tau_j}{.51}; \sigma_1^2 = .001; \sigma_2^2 = .001; \sigma_{12} = 0 \right\}, \\ M_{3,j} &= \left\{ \gamma_1 = 3/4; \gamma_2 = (1/12) - \tau_j; \sigma_1^2 = .001; \sigma_2^2 = .001; \sigma_{12} = .0009 \right\}, \\ M_{4,j} &= \left\{ \gamma_1 = 3/4; \gamma_2 = (1/12) - \tau_j; \sigma_1^2 = .010; \sigma_2^2 = .001; \sigma_{12} = 0 \right\}. \end{aligned} \tag{6.2}$$

Furthermore,

$$\tau_0 = 0, \quad \tau_1 = .05, \quad \tau_2 = .10, \quad \tau_3 = .25.$$

For τ_0 the null hypothesis is true, while for $\tau_j \neq 0$, the alternative is true. It can be shown that for all these models

$$F(\beta) = \begin{bmatrix} (3/4)\tau_j \\ -(3/4)\tau_j \end{bmatrix}, \quad j = 0, 1, 2, 3.$$

Hence, deviation from the null are the same for each basic model. In addition, we have that $\Upsilon_T^* H(\delta) F(\beta) = [T^{1/2}(3/4)\tau_j \ 0]'$ for the VAR(1). When $\tau_j \neq 0$ it thus follows by Proposition 4.2 that the Wald statistic explodes at the usual rate T .

For each given τ_j , the four models have been selected so that specific properties of the $\{x_t\}$ process are highlighted. The autocorrelation function for the spot–forward spread is given by $(1 + \gamma_2 - \gamma_1)^h$, $h = 0, 1, \dots$. The difference between $M_{1,j}$ and $M_{2,j}$ is that the latter model results in substantially larger autocorrelations. In particular, for $M_{1,0}$ the autocorrelation function is $(1/3)^h$ and for $M_{2,0}$ it is approximately $(.96)^h$. Using $M_{1,j}$ as the benchmark model, $M_{3,j}$ differs with respect to the correlation between the disturbances. In $M_{1,j}$ the correlation is zero and in $M_{3,j}$ it is .9. Finally, $M_{4,j}$ results in a 10 times higher variance for $\varepsilon_{1,t}$ than for $\varepsilon_{2,t}$, while in $M_{1,j}$ the variances are equal. Below we shall be concerned with how important these features are for the size and power of the Wald statistic in finite samples.

In Table 1 we report critical values from simulations of ω . Comparing these to the corresponding critical values from the χ^2_2 distribution (see top of Table 2), it is striking how similar they are. Still, for all models (except $M_{4,0}$) the nonstandard distribution is slightly skewed to the right of the χ^2_2 .⁷ Moreover, for the three sample sizes, $T \in \{50, 100, 500\}$, it can be seen that the asymptotic distribution is a rather good approximation of the empirical small sample distribution. The worst case is found for $M_{2,0}$, where the cointegration relation has a very high degree of persistence. Here, the empirical size for $T = 50$ is roughly 2.5 percent higher than the level in the upper tail of the asymptotic distribution. It is worth pointing out that in these experiments we have used the true values of the parameters when simulating the asymptotic distributions and, hence, since we in practice would have to rely on estimates the possible size problem may be underestimated here.⁸

The empirical sizes for $M_{1,0}$ – $M_{4,0}$, when the Toda and Yamamoto test has been carried out, are given in Table 2. Here we find that for $T = 500$, the limiting χ^2_2 distribution approximates the empirical distributions well. For sample sizes that are common in practice ($T \in \{50, 100\}$), however, the test statistic is clearly oversized. In terms of the 95 percent quantile, the empirical size is about 7.5 percent for $T = 100$ and 10 percent for $T = 50$. Comparing these numbers to those in Table 1, it seems like the Toda and Yamamoto approach suffers from substantial size problems.

In Table 3 critical values from the empirical distributions of the Wald statistic for different sample sizes and choices of parameters are given along with size adjusted estimates of power. For example, the empirical critical value from the 95 percent quantile of $(M_{1,0}, T = 100)$ when the VAR(1) has been estimated is equal to 6.377. If data is instead generated according to $M_{1,1}$, the estimated probability of obtaining a larger value

---

⁷As k_3 increases we expect the difference between the nonstandard and χ^2 distributions to become larger.

⁸In principle, for a fixed α the OLS estimates of β and Σ may be used (since they are consistent), but it is probably wise to use estimates that are conditioned on the restrictions. Also, the parameters k_i need to be specified. One possibility is to choose the smallest r which is consistent with the null hypothesis and $k_4 = 1$ if at least one of the time series has a “clear” linear deterministic trend. Inference based on the resulting limiting distribution is likely to be conservative and may thus suffer from lack of power. We shall not discuss this issue further here, but it is clearly a very important one when it comes to applying the nonstandard distributions given in this paper.

for the test statistic is 9.3 percent. As we move to the data model $M_{1,3}$, the probability is estimated to equal 82.7 percent. For the test based on fitting a VAR(1) model to data, power is typically the best in $M_{1,j}$ and the worst in $M_{3,j}$, while $M_{2,j}$ and $M_{4,j}$ are more difficult to rank. The reason why the test has more power in $M_{1,j}$ than in $M_{3,j}$ is easily explained by the fact that the asymptotic standard errors for the restrictions (they are, of course, equal) are much larger in $M_{3,j}$ than in $M_{1,j}$. For $M_{3,2}$ the standard errors are both equal to 2.617 and for $M_{1,2}$ they are .614 (cf. Proposition 3.2 and equation (4.4)).

Next, if we compare the size adjusted power of the Wald test in the VAR(1) to that obtained for the test based on the Toda and Yamamoto approach, the picture is quite clear. When power is low for the VAR(1) (like in $M_{3,j}$), i.e. close to the empirical level, it is also low for the Toda and Yamamoto case. In all other circumstances, power is better in the VAR(1). For example, if $M_{2,3}$ is the true model and the sample size is 100, then the probability of getting a value for W in excess of 6.513 (the critical value from the 95 percent quantile of the empirical distribution for $M_{2,0}$) is roughly 68 percent in the VAR(1). Similarly, the probability of W being larger than 7.146 is about 41 percent when the Toda and Yamamoto approach is used.⁹

To summarize the conclusions from this simulation exercise, we have found that for common sample sizes, $T \in \{50, 100\}$, the Wald test is somewhat oversized in the VAR(1) models and substantially oversized in the Toda and Yamamoto cases. Furthermore, in terms of size adjusted power, the results are also highly in favor of “the VAR(1)” to the Toda and Yamamoto approach. Hence, while the Toda and Yamamoto testing approach is easier to conduct when we rely on limiting distributions, the addition of variables is not something to be taken lightly when inference is carried out in small samples.

---

⁹Note that in some cases the estimated size adjusted power is lower than the empirical level of the test. There may be two reasons for this. First, I have used different $\{\varphi_t\}$ sequences in, e.g., $(M_{4,0}, T = 50)$ and $(M_{4,1}, T = 50)$. Second, the estimates of β and Σ will be different in the two settings even if the same disturbance sequences are used. Hence, even if we did control for the first case, there is still no guarantee in small samples that the empirical distribution of the Wald test, when $M_{4,1}$ is generating the data, is skewed to the right of the empirical distribution of the test for $M_{4,0}$. Moreover, it should be remembered that the empirical small sample distributions are estimates and, hence, subject to sampling uncertainty. Since we have not taken this uncertainty into account we cannot say that a size adjusted power of, say, 4 percent is lower than the 5 percent empirical level.

7 Concluding Comments

Nonlinear restrictions on the coefficients of a VAR system typically occur when we consider testing hypotheses about prediction properties of the data. Allowing for unit roots, we have derived the limiting distribution of the Wald statistic used to test such constraints. It is shown that this distribution is generally nonstandard when the restrictions constrain the cointegration or common trends space. On the other hand, for restrictions which do not reduce the number of free parameters in the cointegration, common trends and in the constant vectors, the asymptotic distribution is the usual χ^2 . In addition, as suggested and proven in Toda and Yamamoto (1993), by adding a sufficient number of lags to the VAR and only test restrictions on the original coefficients, the corresponding Wald statistic has a limiting χ^2 distribution given that the constant in the VAR is not involved. Here, we have given a simple proof of this claim.

An example of nonlinear restrictions in VAR's are those derived from hypotheses associated with linear(ized) rational expectations models. These constraints are known as cross equation restrictions since they impose relationships between coefficients across equations in the VAR system. We have shown that the Wald test of such restrictions generally has a nonstandard limiting distribution. Furthermore, we have tried to give a formal and intuitive explanation why this turns out to be the case. Specifically, the cross equation restrictions operate on the row space of the total impact matrix of the VAR and thereby constrain the common trends space. In addition, we describe how to derive theoretical cointegration vectors from these constraints and we point out that a necessary condition for the cross equation restrictions to be consistent with the data is that the theoretical cointegration vectors span the cointegration space of the VAR system. In other words, cross equation restrictions impose a lower bound for the dimension of the cointegration space and a subspace for the cointegration vectors. If the number of cointegration vectors is equal to the lower bound, then the cross equation restrictions specify the whole cointegration space. This is very important in nonstationary VAR's since it greatly facilitates inference from a nonstandard limiting distribution.

Finally, we have performed a small Monte Carlo study, comparing inference based on a nonstandard distribution to inference based on the standard χ^2 , using the Toda and

Yamamoto (1993) approach of variable addition. The results we have obtained here are fully in favor of using the nonstandard distribution and, thus, of estimating VAR models in levels with as few lags as data allows. While the Wald statistic is somewhat oversized in this case, it is substantially oversized in the Toda and Yamamoto case. Moreover, in terms of size adjusted power, the Toda and Yamamoto approach is outperformed by the approach based on nonstandard limiting distributions. It should be pointed out, though, that we have not examined all difficulties that may arise when inference relies on nonstandard distributions. In particular, the questions of how to select the number of cointegration vectors and whether or not to allow for a linear trend in the series need further attention. When nonlinear cross equation restrictions are tested within a VAR model, the underlying economic hypothesis typically, as mentioned above, gives us a good idea of how to address these questions.

Table 1: Critical values from simulated asymptotic distributions and the size of the Wald statistic in 2 dimensional VAR(1) models with 1 cointegration vector.

Model	T	The VAR(1)							
		.10	.50	.80	.85	.90	.95	.975	.99
$M_{1.0}$	∞	.218	1.410	3.211	3.813	4.607	6.107	7.515	9.457
	Empirical quantiles								
	500	.101	.498	.796	.849	.899	.950	.976	.990
	100	.101	.488	.788	.839	.884	.944	.970	.987
	50	.100	.496	.779	.831	.881	.936	.965	.984
$M_{2.0}$	∞	.212	1.410	3.263	3.880	4.684	6.014	7.519	9.278
	Empirical quantiles								
	500	.094	.488	.800	.849	.899	.945	.973	.988
	100	.101	.486	.780	.834	.880	.938	.969	.985
	50	.091	.467	.759	.813	.868	.923	.959	.978
$M_{3.0}$	∞	.231	1.453	3.364	3.961	4.739	6.115	7.519	9.197
	Empirical quantiles								
	500	.098	.487	.790	.840	.889	.943	.973	.987
	100	.109	.509	.797	.847	.891	.944	.970	.986
	50	.102	.488	.784	.835	.884	.935	.964	.983
$M_{4.0}$	∞	.216	1.421	3.261	3.830	4.589	5.978	7.238	9.108
	Empirical quantiles								
	500	.101	.504	.798	.845	.893	.942	.969	.986
	100	.105	.496	.781	.827	.872	.927	.956	.976
	50	.104	.495	.776	.824	.868	.924	.949	.971

Notes: The limiting distributions are based on a step length $S = 800$ and 10,000 replications. For the empirical small sample distributions, we have also performed 10,000 replications. The random number generator in RATS386, version 3.11, has been used and the ranking function in GAUSS-386i, version 3.1.4.

Table 2: Critical values from the χ^2_2 distribution and the size of the Wald statistic based on the Toda and Yamamoto approach.

Model	T	Toda and Yamamoto							
		.10	.50	.80	.85	.90	.95	.975	.99
	∞	.211	1.386	3.219	3.794	4.605	5.991	7.378	9.210
Empirical quantiles									
$M_{1.0}$	500	.098	.492	.791	.842	.894	.945	.973	.988
	100	.096	.489	.777	.824	.872	.927	.958	.979
	50	.095	.466	.758	.806	.856	.904	.935	.961
$M_{2.0}$	500	.097	.505	.798	.850	.900	.947	.972	.986
	100	.100	.485	.769	.815	.869	.924	.954	.974
	50	.092	.462	.741	.790	.842	.900	.934	.960
$M_{3.0}$	500	.102	.495	.789	.839	.894	.942	.969	.985
	100	.098	.483	.772	.823	.875	.925	.955	.976
	50	.090	.457	.750	.803	.853	.910	.942	.965
$M_{4.0}$	500	.103	.500	.794	.842	.890	.940	.966	.983
	100	.095	.488	.773	.819	.869	.915	.942	.961
	50	.092	.478	.760	.803	.851	.897	.922	.942

Table 3: Critical values from empirical distributions and size adjusted estimates of power for the Wald statistic in 2 dimensional VAR(1) models with 1 cointegration vector and based on the Toda and Yamamoto approach.

Model	T	The VAR(1)						Toda and Yamamoto					
		.80	.85	.90	.95	.975	.99	.80	.85	.90	.95	.975	.99
$M_{1.0}$	500	3.244	3.825	4.625	6.097	7.442	9.531	3.303	3.898	4.739	6.190	7.533	9.796
	100	3.350	3.973	4.908	6.377	7.885	10.208	3.501	4.207	5.218	6.924	8.745	11.590
	50	3.439	4.095	5.027	6.773	8.451	10.743	3.708	4.502	5.856	8.242	10.959	16.001
Probabilities of larger values													
$M_{1.1}$	500	.485	.423	.344	.229	.153	.078	.423	.360	.283	.185	.127	.065
	100	.293	.233	.164	.093	.052	.032	.272	.214	.152	.091	.052	.024
	50	.266	.211	.148	.078	.043	.020	.253	.196	.141	.069	.037	.013
$M_{1.2}$	500	.885	.850	.796	.689	.592	.457	.805	.757	.690	.576	.474	.338
	100	.466	.394	.311	.213	.140	.070	.395	.333	.255	.165	.107	.055
	50	.362	.294	.221	.136	.082	.042	.328	.266	.191	.111	.063	.026
$M_{1.3}$	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	.950	.929	.894	.827	.748	.626	.895	.864	.811	.712	.611	.468
	50	.783	.730	.636	.521	.407	.287	.706	.647	.547	.402	.281	.153
Model	T	.80	.85	.90	.95	.975	.99	.80	.85	.90	.95	.975	.99
$M_{2.0}$	500	3.262	3.892	4.711	6.214	7.701	9.677	3.238	3.790	4.610	6.127	7.626	10.220
	100	3.463	4.115	5.090	6.513	7.977	10.288	3.579	4.304	5.263	7.146	9.339	13.139
	50	3.709	4.370	5.353	6.987	8.949	11.323	3.937	4.749	6.001	8.445	11.346	16.621
Probabilities of larger values													
$M_{2.1}$	500	.577	.506	.425	.306	.215	.131	.282	.228	.163	.089	.050	.020
	100	.270	.210	.141	.082	.049	.021	.206	.153	.104	.048	.024	.008
	50	.225	.174	.119	.065	.032	.013	.189	.140	.092	.045	.023	.009
$M_{2.2}$	500	.900	.866	.819	.721	.614	.444	.552	.491	.400	.275	.184	.086
	100	.420	.348	.265	.174	.114	.057	.255	.192	.130	.066	.032	.008
	50	.303	.237	.170	.097	.048	.023	.203	.152	.096	.046	.022	.007
$M_{2.3}$	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.998	.995	.989	.967
	100	.883	.845	.782	.680	.573	.419	.720	.649	.559	.406	.264	.116
	50	.652	.585	.494	.358	.233	.131	.480	.400	.298	.170	.086	.026

Table 3: (Continued) Critical values from empirical distributions and size adjusted power estimates.

Model	T	The VAR(1)						Toda and Yamamoto					
		.80	.85	.90	.95	.975	.99	.80	.85	.90	.95	.975	.99
$M_{3.0}$	500	3.487	4.086	4.935	6.353	7.752	9.818	3.329	3.930	4.746	6.305	7.926	10.116
	100	3.395	4.009	4.895	6.359	7.922	9.967	3.517	4.163	5.214	7.106	9.083	11.851
	50	3.538	4.161	5.144	6.762	8.263	10.319	3.731	4.559	5.666	7.919	10.510	14.674
Probabilities of larger values													
$M_{3.1}$	500	.213	.158	.106	.053	.028	.010	.207	.158	.110	.054	.027	.010
	100	.205	.152	.100	.051	.026	.011	.197	.150	.100	.051	.027	.010
	50	.195	.149	.095	.048	.026	.011	.208	.154	.105	.054	.027	.011
$M_{3.2}$	500	.266	.209	.148	.083	.045	.020	.260	.203	.146	.077	.039	.016
	100	.223	.173	.115	.059	.032	.013	.207	.160	.106	.051	.026	.011
	50	.204	.154	.102	.050	.025	.012	.211	.161	.110	.056	.027	.011
$M_{3.3}$	500	.602	.536	.451	.334	.245	.146	.602	.538	.454	.330	.221	.128
	100	.304	.245	.182	.111	.064	.033	.300	.239	.171	.092	.052	.024
	50	.266	.208	.147	.080	.047	.022	.250	.197	.140	.075	.041	.018
Model	T	The VAR(1)						Toda and Yamamoto					
		.80	.85	.90	.95	.975	.99	.80	.85	.90	.95	.975	.99
$M_{4.0}$	500	3.277	3.901	4.755	6.231	7.670	9.992	3.288	3.904	4.834	6.456	8.224	10.604
	100	3.485	4.188	5.143	6.886	8.999	12.001	3.535	4.243	5.494	8.090	11.547	18.655
	50	3.525	4.224	5.336	7.284	9.642	12.176	3.750	4.574	6.104	10.584	19.073	39.827
Probabilities of larger values													
$M_{4.1}$	500	.532	.448	.351	.219	.132	.047	.479	.405	.302	.170	.088	.031
	100	.224	.159	.096	.039	.014	.003	.194	.137	.071	.023	.008	.002
	50	.200	.141	.080	.036	.015	.006	.171	.123	.073	.026	.011	.005
$M_{4.2}$	500	.989	.980	.964	.920	.854	.706	.981	.972	.948	.878	.768	.584
	100	.521	.418	.298	.146	.056	.014	.451	.352	.220	.079	.025	.005
	50	.348	.262	.165	.070	.029	.012	.282	.203	.113	.031	.012	.004
$M_{4.3}$	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	1.000	.980	1.000	1.000	.999	.995	.955	.695
	50	.992	.984	.962	.884	.737	.566	.979	.957	.895	.590	.261	.090

Mathematical Appendix

Proof of Proposition 3.1 From equation (3.3) we have that

$$[\Upsilon_T \otimes I_n] (\hat{\delta} - \delta) = \left[(\Upsilon_T^{-1} Z Z' \Upsilon_T^{-1})^{-1} \otimes I_n \right] [\Upsilon_T^{-1} Z \otimes \Sigma^{1/2}] \varphi.$$

From Lemma 1 and 2 in SSW we can derive the limiting representation of the 10 unique blocks of the matrix $V_T = (\Upsilon_T^{-1} Z Z' \Upsilon_T^{-1})$ and of the four vectors in $[\Upsilon_T^{-1} Z \otimes \Sigma^{1/2}] \varphi$. Moreover, by Theorem 2.4 in Chan and Wei (1988) the joint convergence of these matrices and vectors is assured.

Let $\Phi_t := \sum_{s=1}^t \varphi_s$, with $\Phi_0 = 0$. Letting $Z_t^1 = \tilde{F}_{11}(L) \Sigma^{1/2} \varphi_t$, where $\tilde{F}_{11}(z) = \sum_{j=0}^{\infty} \tilde{F}_{11,j} z^j$ is a $k_1 \times n$ matrix polynomial, we have that

$$T^{-1} \sum_{t=1}^T Z_{t-1}^1 Z_{t-1}^{1'} \xrightarrow{P} V_{11} := \sum_{j=0}^{\infty} \tilde{F}_{11,j} \Sigma \tilde{F}_{11,j}', \quad (\text{A.1})$$

by part (f) of Lemma 1 in SSW. Similarly,

$$T^{-1} \sum_{t=1}^T Z_{t-1}^1 Z_{t-1}^2 \xrightarrow{P} 0, \quad (\text{A.2})$$

since $E[Z_{t-1}^1] = 0$ and $Z_{t-1}^2 = 1$. Next,

$$T^{-3/2} \sum_{t=1}^T Z_{t-1}^1 Z_{t-1}^{3'} \xrightarrow{P} 0, \quad (\text{A.3})$$

by Lemma 2 in SSW, while the same Lemma gives us that

$$T^{-2} \sum_{t=1}^T Z_{t-1}^1 Z_{t-1}^4 \xrightarrow{P} 0. \quad (\text{A.4})$$

Turning to the next column block of V_T ,

$$T^{-1} \sum_{t=1}^T (Z_{t-1}^2)^2 = 1, \quad (\text{A.5})$$

while

$$\begin{aligned} T^{-3/2} \sum_{t=1}^T Z_{t-1}^3 Z_{t-1}^2 &= T^{-3/2} \sum_{t=1}^T Z_{t-1}^3 \\ &= \nu' C(1) \Sigma^{1/2} T^{-3/2} \sum_{t=1}^T \Phi_{t-1} + o_p(1) \\ &\Rightarrow \nu' C(1) \Sigma^{1/2} \int_0^1 B(u) du, \end{aligned} \quad (\text{A.6})$$

by part (a) of Lemma 1 in SSW. Also,

$$\begin{aligned}
T^{-2} \sum_{t=1}^T Z_{t-1}^4 Z_{t-1}^2 &= T^{-2} \sum_{t=1}^T Z_{t-1}^4 \\
&= \mu' \mu [T(T-1)/2T^2] + o_p(1) \\
&\xrightarrow{P} \mu' \mu / 2.
\end{aligned} \tag{A.7}$$

Moving to the third column block of V_T we first have that

$$\begin{aligned}
T^{-2} \sum_{t=1}^T Z_{t-1}^3 Z_{t-1}^{3'} &= \nu' C(1) \Sigma^{1/2} \left[T^{-2} \sum_{t=1}^T \Phi_{t-1} \Phi_{t-1}' \right] \Sigma^{1/2'} C(1)' \nu + o_p(1) \\
&\Rightarrow \nu' C(1) \Sigma^{1/2} \left[\int_0^1 B(u) B(u)' du \right] \Sigma^{1/2'} C(1)' \nu,
\end{aligned} \tag{A.8}$$

by part (b) of Lemma 1 in SSW. Moreover,

$$\begin{aligned}
T^{-5/2} \sum_{t=1}^T Z_{t-1}^3 Z_{t-1}^4 &= \mu' \mu \nu' C(1) \Sigma^{1/2} \left[T^{-5/2} \sum_{t=1}^T (t-1) \Phi_{t-1} \right] + o_p(1) \\
&\Rightarrow \mu' \mu \nu' C(1) \Sigma^{1/2} \left[\int_0^1 u B(u) du \right],
\end{aligned} \tag{A.9}$$

by part (a) of Lemma 1. Finally,

$$\begin{aligned}
T^{-3} \sum_{t=1}^T (Z_{t-1}^4)^2 &= (\mu' \mu)^2 \frac{T(T-1)(2T-1)}{6T^3} + o_p(1) \\
&\xrightarrow{P} (\mu' \mu)^2 / 3.
\end{aligned} \tag{A.10}$$

Combining the expressions in (A.1)–(A.10) we find that

$$V_T \Rightarrow \begin{bmatrix} V_{11} & 0 & 0 & 0 \\ 0 & 1 & \Xi' C(1)' \nu & \mu' \mu / 2 \\ 0 & \nu' C(1) \Xi & \nu' C(1) \Psi C(1)' \nu & \mu' \mu \nu' C(1) \Lambda \\ 0 & \mu' \mu / 2 & \mu' \mu \Lambda' C(1)' \nu & (\mu' \mu)^2 / 3 \end{bmatrix}, \tag{A.11}$$

where Ξ , Λ , and Ψ are defined in Proposition 3.1.

Turning to the vector $[\Upsilon_T^{-1} Z \otimes \Sigma^{1/2}] \varphi = \text{vec}(\Sigma^{1/2} f Z' \Upsilon_T^{-1})$ we first observe that

$$f Z' \Upsilon_T^{-1} = \begin{bmatrix} T^{-1/2} \sum_{t=1}^T \varphi_t Z_{t-1}^{1'} & T^{-1/2} \sum_{t=1}^T \varphi_t & T^{-1} \sum_{t=1}^T \varphi_t Z_{t-1}^{3'} & T^{-3/2} \sum_{t=1}^T \varphi_t Z_{t-1}^4 \end{bmatrix}.$$

It follows that

$$\text{vec}(f Z' \Upsilon_T^{-1}) = \begin{bmatrix} \text{vec}(T^{-1/2} \sum_{t=1}^T \varphi_t Z_{t-1}^{1'}) \\ T^{-1/2} \sum_{t=1}^T \varphi_t \\ \text{vec}(T^{-1} \sum_{t=1}^T \varphi_t Z_{t-1}^{3'}) \\ T^{-3/2} \sum_{t=1}^T \varphi_t Z_{t-1}^4 \end{bmatrix}. \tag{A.12}$$

By a central limit theorem for stationary processes with finite fourth moments or a functional limit theorem (see e.g. Helland (1982)), we get

$$\left[I_{(n(p-1)+r)} \otimes \Sigma^{1/2} \right] \text{vec} \left(T^{-1/2} \sum_{t=1}^T \varphi_t Z_{t-1}' \right) \Rightarrow \phi_1 \sim \mathcal{N}(0, V_{11} \otimes \Sigma). \quad (\text{A.13})$$

By the same theorem

$$\Sigma^{1/2} T^{-1/2} \sum_{t=1}^T \varphi_t \Rightarrow \phi_2 := \Sigma^{1/2} \int_0^1 dB(u) \sim \mathcal{N}(0, \Sigma). \quad (\text{A.14})$$

By part (e) of Lemma 1 in SSW

$$T^{-1} \sum_{t=1}^T \varphi_t \Phi_{t-1}' \Rightarrow \int_0^1 dB(u) B(u)'$$

Accordingly, if $k_3 > 0$ we have that

$$\begin{aligned} \left[I_{k_3} \otimes \Sigma^{1/2} \right] \text{vec} \left(T^{-1} \sum_{t=1}^T \varphi_t Z_{t-1}' \right) &= \left[\nu' C(1) \Sigma^{1/2} \otimes \Sigma^{1/2} \right] \times \\ &\quad \text{vec} \left(T^{-1} \sum_{t=1}^T \varphi_t \Phi_{t-1}' \right) + o_p(1) \\ &\Rightarrow \phi_3 := \left[\nu' C(1) \Sigma^{1/2} \otimes \Sigma^{1/2} \right] \times \\ &\quad \text{vec} \left(\int_0^1 dB(u) B(u)' \right). \end{aligned} \quad (\text{A.15})$$

Also,

$$\begin{aligned} \Sigma^{1/2} T^{-3/2} \sum_{t=1}^T \varphi_t Z_{t-1}' &= \mu' \mu \Sigma^{1/2} T^{-1/2} \sum_{t=1}^T [(t-1)/T] \varphi_t + o_p(1) \\ &\Rightarrow \phi_4 := \mu' \mu \Sigma^{1/2} \int_0^1 u dB(u) \sim \mathcal{N}(0, [(\mu' \mu)^2 / 3] \Sigma), \end{aligned} \quad (\text{A.16})$$

as given in part (d) of Lemma 1 in SSW. Finally, ϕ_1 is independent of ϕ_j , $j = 2, 3, 4$, by Lemma 2 in SSW (which relies on Theorem 2.2 in Chan and Wei (1988)). The intuition behind this result is given in equation (A.11). $\square$

Proof of Proposition 3.2 We shall limit the proof to the case when $k_1 > 0$. We know from equations (2.6) and (3.2) that

$$\begin{aligned} T^{1/2} (\hat{\beta}_1 - \beta_1) &= T^{1/2} [D_1' \otimes I_n] (\hat{\delta}_1 - \delta_1) + T^{1/2} [D_3' \otimes I_n] (\hat{\delta}_3 - \delta_3) + \\ &\quad T^{1/2} [D_4' \otimes I_n] (\hat{\delta}_4 - \delta_4) \\ &= [D_1' \otimes I_n] T^{1/2} (\hat{\delta}_1 - \delta_1) + o_p(1), \end{aligned} \quad (\text{A.17})$$

by Proposition 3.1. In addition, we know that $T^{1/2}(\hat{\delta}_1 - \delta_1) \xrightarrow{d} \mathcal{N}(0, V_{11}^{-1} \otimes \Sigma)$. Next, by equations (2.6) and (3.2) we also find that

$$\begin{aligned} T^{1/2}(\hat{\beta}_2 - \beta_2) &= T^{1/2}[D'_2 \otimes I_n](\hat{\delta}_1 - \delta_1) + T^{1/2}(\hat{\delta}_2 - \delta_2) \\ &\Rightarrow [D'_2 V_{11}^{-1} \otimes I_n] \phi_1 + V_{22}^+ \phi_2 + [V_{23}^+ \otimes I_n] \phi_3 + V_{24}^+ \phi_4, \end{aligned} \quad (\text{A.18})$$

by Proposition 3.1. $\square$

Proof of Proposition 4.1 Suppose $F(\beta) = 0$. Substituting for $F(\hat{\beta})$ in equation (4.1) into equation (4.3), and using the fact that $\beta = (D' \otimes I_n)\delta$, $Z = DY$, $\tilde{P}(\delta) = H(\delta)R(\beta)[D' \otimes I_n]$, we obtain

$$\begin{aligned} W &= (\tilde{P}(\hat{\delta})(\hat{\delta} - \delta))' [\tilde{P}(\hat{\delta}) [(ZZ')^{-1} \otimes \hat{\Sigma}] \tilde{P}(\hat{\delta})']^{-1} (\tilde{P}(\hat{\delta})(\hat{\delta} - \delta)) \\ &\quad + 2\varrho'_{(\hat{\beta}, \beta)} H(\delta)' [\tilde{P}(\hat{\delta}) [(ZZ')^{-1} \otimes \hat{\Sigma}] \tilde{P}(\hat{\delta})']^{-1} (\tilde{P}(\hat{\delta})(\hat{\delta} - \delta)) \\ &\quad + \varrho'_{(\hat{\beta}, \beta)} H(\delta)' [\tilde{P}(\hat{\delta}) [(ZZ')^{-1} \otimes \hat{\Sigma}] \tilde{P}(\hat{\delta})']^{-1} H(\delta) \varrho_{(\hat{\beta}, \beta)}. \end{aligned} \quad (\text{A.19})$$

Using the relationship between the $\tilde{P}(\delta)$ and the $P_T^*(\delta)$ matrices, the Wald statistic in (A.19) is equal to

$$\begin{aligned} W &= (P_T^*(\hat{\delta})[\Upsilon_T \otimes I_n](\hat{\delta} - \delta))' [P_T^*(\hat{\delta})[V_T^{-1} \otimes \hat{\Sigma}] P_T^*(\hat{\delta})']^{-1} \\ &\quad \times (P_T^*(\hat{\delta})[\Upsilon_T \otimes I_n](\hat{\delta} - \delta)) \\ &\quad + 2\varrho'_{(\hat{\beta}, \beta)} H(\delta)' \Upsilon_T^* [P_T^*(\hat{\delta})[V_T^{-1} \otimes \hat{\Sigma}] P_T^*(\hat{\delta})']^{-1} (P_T^*(\hat{\delta})[\Upsilon_T \otimes I_n](\hat{\delta} - \delta)) \\ &\quad + \varrho'_{(\hat{\beta}, \beta)} H(\delta)' \Upsilon_T^* [P_T^*(\hat{\delta})[V_T^{-1} \otimes \hat{\Sigma}] P_T^*(\hat{\delta})']^{-1} \Upsilon_T^* H(\delta) \varrho_{(\hat{\beta}, \beta)} \\ &= W_T^1 + W_T^2 + W_T^3, \end{aligned} \quad (\text{A.20})$$

where $V_T := \Upsilon_T^{-1} Z Z' \Upsilon_T^{-1}$. By Assumption 3.1 and part (i) and (ii) of Assumption 4.1 we find that $V_T \Rightarrow V$, $[\Upsilon_T \otimes I_n](\hat{\delta} - \delta) \Rightarrow [V^{-1} \otimes I_n]\phi$, $\hat{\Sigma} \xrightarrow{p} \Sigma$, and $P_T^*(\hat{\delta}) \xrightarrow{p} P^*(\delta)$, which has full row rank. Hence, W_T^1 in (A.20) converges weakly to the random variable in equation (4.8). To complete the proof it remains to show that $\Upsilon_T^* H(\delta) \varrho_{(\hat{\beta}, \beta)}$ is $o_p(1)$, i.e. that $W_T^2, W_T^3 \in o_p(1)$.

The remainder term in equation (4.1) can always be represented as the product of two terms. Specifically,

$$\varrho_{(\hat{\beta}, \beta)} = \varrho_{(\hat{\beta}, \beta)}^* (\hat{\beta} - \beta), \quad (\text{A.21})$$

where $\varrho_{(\hat{\beta}, \beta)}^*$ is a $q \times n(np + 1)$ matrix. Assumption 4.1(iii) implies that $\varrho_{(\hat{\beta}, \beta)}^* \in o_p(1)$. Defining $\bar{\varrho}_{T,(\hat{\delta}, \delta)}$ from

$$\Upsilon_T^{*-1} \bar{\varrho}_{T,(\hat{\delta}, \delta)} = \varrho_{(\hat{\beta}, \beta)}^* \left[D' \Upsilon_T^{-1} \otimes I_n \right],$$

it follows that $\bar{\varrho}_{T,(\hat{\delta}, \delta)} \in o_p(1)$. Making use of these properties we obtain

$$\Upsilon_T^* H(\delta) \varrho_{(\hat{\beta}, \beta)} = H(\delta) \bar{\varrho}_{T,(\hat{\delta}, \delta)} [\Upsilon_T \otimes I_n] (\hat{\delta} - \delta). \quad (\text{A.22})$$

The right hand side of this expression converges to zero in probability by Assumptions 3.1 and 4.1. $\square$

Proof of Proposition 4.2 Suppose $F(\beta) \neq 0$. By equation (A.20) and Assumption 4.1 we have that

$$\begin{aligned} W &= (\Upsilon_T^* H(\delta) F(\beta))' \left[P_T^*(\hat{\delta}) \left[V_T^{-1} \otimes \hat{\Sigma} \right] P_T^*(\hat{\delta})' \right]^{-1} (\Upsilon_T^* H(\delta) F(\beta)) \\ &\quad + 2(\Upsilon_T^* H(\delta) F(\beta))' \left[P_T^*(\hat{\delta}) \left[V_T^{-1} \otimes \hat{\Sigma} \right] P_T^*(\hat{\delta})' \right]^{-1} \\ &\quad \times \left(P_T^*(\hat{\delta}) + H(\delta) \bar{\varrho}_{T,(\hat{\delta}, \delta)} \right) [\Upsilon_T \otimes I_n] (\hat{\delta} - \delta) + W_T^1 + o_p(1). \end{aligned} \quad (\text{A.23})$$

Under the alternative, the first and the second term on the right hand side are nonzero. Furthermore, as the sample size increases both terms become larger (at least for all $T \geq T^*$) at the rates given by the nonzero coefficient in $\Upsilon_T^* H(\delta) F(\beta)$ whose T -factor is the largest. This factor can always be found in the first term since it is based on the square of $\Upsilon_T^* H(\delta) F(\beta)$ and can only take on the values T , T^2 , or T^3 . $\square$

References

- [1] Andrews, D.W.K. (1987), “Asymptotic Results for Generalized Wald Tests”, *Econometric Theory*, 3:348–358.
- [2] Baillie, R.T. (1989), “Econometric Tests of Rationality and Market Efficiency”, *Econometric Reviews*, 8:151–186.
- [3] Bergman, M. and A. Warne (1993), “Money–Income Causality and the Neutrality of Money”, Manuscript, Institute for International Economic Studies, Stockholm University.
- [4] Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley, New York.
- [5] Campbell, J.Y. (1987), “Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis”, *Econometrica*, 55:1249–1273.
- [6] Campbell, J.Y. and P. Perron (1991), “Pitfalls and Opportunities: What Macroeconomists should know about Unit Roots”, in O.J. Blanchard and S. Fischer (eds), *NBER Macroeconomics Annual*, MIT Press, 141–201.
- [7] Campbell, J. Y. and R.J. Shiller (1987), “Cointegration and Tests of Present Value Models”, *Journal of Political Economy*, 95:1062–1088.
- [8] Chan, N. H. and C. Z. Wei (1988), “Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes”, *Annals of Statistics*, 16:367–401.
- [9] Christiano, L.J. and L. Ljungqvist (1988), “Money does Granger–Cause Output in the Bivariate Money–Output Relation”, *Journal of Monetary Economics*, 22:217–235.
- [10] Engle, R.F. and C.W.J. Granger (1987), “Co–Integration and Error Correction: Representation, Estimation and Testing”, *Econometrica*, 55:251–276.
- [11] Gallant, A.R. (1987), *Nonlinear Statistical Models*, John Wiley, New York.
- [12] Gregory, A.W. and M.R. Veall (1985), “Formulating Wald Tests of Nonlinear Restrictions”, *Econometrica*, 53:1465–1468.

- [13] Hansen, L.P. and T.J. Sargent (1981), “Exact Linear Rational Expectations Models: Specification and Estimation”, Federal Reserve Bank of Minneapolis, Research Department, Staff Report # 71.
- [14] Helland, I.S. (1982), “Central Limit Theorems for Martingales with Discrete or Continuous Time”, *Scandinavian Journal of Statistics*, 9:79–94.
- [15] Ito, T. (1988), “Use of (Time–Domain) Vector Autoregressions to Test Uncovered Interest Parity”, *Review of Economics and Statistics*, 70:296–305.
- [16] Jacobson, T. and R. Larsson (1991), “Numerical Aspects of a Likelihood Ratio Statistic for Cointegration Rank”, Manuscript, Department of Statistics, University of Uppsala.
- [17] Johansen, S. (1991) “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”, *Econometrica*, 59:1551–1580.
- [18] Johansen, S. and K. Juselius (1992), “Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for U.K.”, *Journal of Econometrics*, 53:211–244.
- [19] King, R.G., C.I. Plosser, J.H. Stock, and M.W. Watson (1991), “Stochastic Trends and Economic Fluctuations”, *American Economic Review*, 81:819–840.
- [20] Lütkepohl, H. (1989), “Testing for Causation Between Two Variables in Higher Dimensional VAR Models”, Manuscript, Institut für Statistik und Ökonometrie der Christian–Albrechts–Universität, Kiel, Germany.
- [21] Mellander, E., A. Vredin, and A. Warne (1992), “Stochastic Trends and Economic Fluctuations in a Small Open Economy”, *Journal of Applied Econometrics*, 7:369–394.
- [22] Sargent, T.J. (1979), “A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure”, *Journal of Monetary Economics*, 5:133–143.

- [23] Schmidt, P. (1973), “The Asymptotic Distribution of Dynamic Multipliers”, *Econometrica*, 41:161–164.
- [24] Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley, New York.
- [25] Sims, C.A., J.H. Stock, and M.W. Watson (1990), “Inference in Linear Time Series Models with Some Unit Roots”, *Econometrica*, 58:113–144.
- [26] Toda, H.Y. and T. Yamamoto (1993), “Statistical Inference in Vector Autoregressions with Possibly Integrated Processes”, Manuscript, Institute of Socio–Economic Planning, University of Tsukuba, Japan.
- [27] Warne, A. (1992), “Inference in Vector Autoregressions with Integrated Time Series”, Manuscript, Institute for International Economic Studies, Stockholm University.
- [28] Warne, A. (1990), *Vector Autoregressions and Common Trends in Macro and Financial Economics*, PhD thesis, Stockholm School of Economics.
- [29] West, K.D. (1988), “Asymptotic Normality, when Regressors have a Unit Root”, *Econometrica*, 56:1397–1417.