Estimation and Testing for Common Cycles

Anders Warne

February 27, 2008

Abstract: This note discusses estimation and testing for the presence of common cycles in cointegrated vector autoregressions. A simple two-stage estimator is considered where the cointegration vectors are estimated in the first stage and the remaining parameters, subject to common cycles, in the second stage. The latter stage is equivalent to using reduced rank regression in a stationary framework. Simple procedures for iterating on these two stages are discussed with emphasis on estimating the cointegration space conditional on the common cycles restriction. It is shown that the two-stage estimator of the parameters describing the dynamics is asymptotically Gaussian and efficient. Furthermore, an estimator of the co-feature matrix is examined and its asymptotic properties are derived. Finally, two asymptotically equivalent methods for computing the likelihood ratio test for the null of s versus $s+g$ common cycles are presented along with the limiting behavior of the tests.

Keywords: Asymptotics, Cointegration, Common Cycles, Likelihood Ratio Test.

JEL Classification Numbers: C32.

1. Setup

The model we shall consider is a standard cointegrated VAR model written on error correction form as:

$$\Delta x_t = \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + \alpha\beta' x_{t-1} + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where x_t is p dimensional, the parameters α and β are $p \times r$ with full column rank $r \in \{0, 1, \ldots, p-1\}$, $x_0, \ldots, x_{1-k}$ are fixed, while ε_t is i.i.d. Gaussian with zero mean and positive definite covariance matrix Ω. For simplicity, but without loss of generality, I have excluded all deterministic variables from the model since the common cycles hypothesis does not impose any restrictions on the parameters of such variables.

Following Vahid and Engle (1993) we say that x_t has s *common cycles* if (and only if)

$$\Theta = \begin{bmatrix} \Gamma_1 & \ldots & \Gamma_{k-1} & \alpha \end{bmatrix} = \xi\eta', \tag{2}$$

where ξ is a full rank $p \times s$ matrix and η is a full rank $p(k-1)+r \times s$ matrix for $s \in \{r, \ldots, p-1\}$. Given that x_t has s common cycles it follows that there exists a $p \times (p-s)$ matrix $\xi_\perp$ such that $\xi_\perp' \Delta x_t$ is serially uncorrelated (since $\xi_\perp' \xi = 0$ by construction). The matrix $\xi_\perp$ is called the *co-feature* matrix.[1]

In the event that $k = 1$, s is always equal to r, the cointegration rank, and $\xi_\perp = \alpha_\perp$. This means that if we wish to test for the presence of s common cycles under $k = 1$ it is equivalent to testing for the rank of the $p \times p$ matrix $\Pi = \alpha\beta'$. Once we have established a proper rank for this matrix, there is no need to pursue any further estimation or testing regarding common cycles. When the lag order is greater than 1, however, then s can be any integer between r and $p-1$. It is clear that s cannot be smaller than r since that would violate the rank assumption for α. Furthermore, the case $s = p$ is uninteresting since this implies that Θ is a full rank matrix,

[1] Hecq, Palm and Urbain (2000) suggest two forms of reduced rank for Θ. The strong form is equivalent to equation (2) being satisfied, while their weak form only requires that the Γ_i matrices can be written as functions of ξ, i.e., that $\xi_\perp'(\Delta x_t - \alpha\beta' x_{t-1}) = \xi_\perp' \varepsilon_t$. The weak form is not discussed in this note since, e.g., it is not invariant to the specification of the error correction form, i.e., if x appears in levels on the right hand side of (1) for $t-1$ or for $t-k$.

i.e., common cycles would not impose any restrictions on Θ and the co-feature matrix would be empty. Rather, the case $s = p$ may be used as one alternative hypothesis when we are interested in testing the null hypothesis that Θ has reduced rank $s < p$.

Assuming that x_t is subject to s common cycles, it follows that Θ has less than $p(p(k-1)+r)$ unique parameters. Since $\xi\eta' = \xi MM^{-1}\eta' = \xi^*\eta^{*\prime}$ for any full rank $s \times s$ matrix M we know that, e.g., only the space spanned by the columns of η can identified. To uniquely determine η (and ξ) let us therefore select a basis for the space spanned by its columns, expressed as:

$$\eta = h + h_\perp \psi, \tag{3}$$

where h is a known $p(k-1) + r \times s$ matrix such that, for simplicity, $h'h = I_s$, $h_\perp$ is $p(k-1) + r \times p(k-1) + r - s$ of full rank such that $h'_\perp h = 0$ and $h'_\perp h_\perp = I_{p(k-1)+r-s}$, while ψ is $p(k-1) + r - s \times s$ contains the free parameters of η.

From this discussion it directly follows that if x_t has s common cycles, this imposes exactly $q = (p-s)(p(k-1)+r-s)$ restrictions on the Γ_i and α parameters of the cointegrated VAR model relative to the case when Θ has full rank p. Generally, suppose we wish to test the hypothesis that Θ has rank s versus the alternative that it has rank $s + g$, the number of restrictions on Γ_i and α is $q = g(pk + r - 2s - g)$. Against all $g \in \{1, \ldots p - s\}$, the number of restrictions is positive when $k \geq 2$. Again, for $k = 1$ we cannot impose any additional restrictions on α based on the common cycles hypothesis since $s = r$ by assumption.

2. Estimation: A Two-Stage Algorithm

As discussed above, the assumption of s common cycles imposes a second reduced rank condition on the cointegrated VAR model whenever the lag order of the VAR, k, is greater than or equal to 2. Since $k = 1$ is trivial, we shall assume that $k \geq 2$.

First of all, suppose that $s = r$. In this case the two reduced rank conditions can be merged into one reduced rank condition since $\xi = \alpha M$ for some full rank $r \times r$ matrix M. This means that

$$\begin{aligned} \Delta x_t &= \alpha \begin{bmatrix} \tilde{\eta}'_2 & \beta' \end{bmatrix} \begin{bmatrix} \Delta x_{t-1} \\ \vdots \\ \Delta x_{t-k+1} \\ x_{t-1} \end{bmatrix} + \varepsilon_t \\ &= \alpha\beta^{*\prime} Z^*_{1t} + \varepsilon_t, \end{aligned}$$

where $[\Gamma_1 \ \cdots \ \Gamma_{k-1}] = \alpha\tilde{\eta}'_2$. This means that $\Theta = \alpha[\tilde{\eta}'_2 \ \ I_r]$ so that $h = [0 \ \ I_r]'$. Hence, $\xi = \alpha$ and $\psi = \tilde{\eta}_2$, while the co-feature matrix $\xi_\perp$ spans the space spanned by $\alpha_\perp$, as in the case when $k = 1$. The parameters α and β^* can be estimated by maximum likelihood through reduced rank regression. In case β should satisfy some restrictions, e.g., exact identification, procedures analysed by Johansen (1996) can be used directly.

Second, suppose $s \geq r$ so that the two reduced rank conditions cannot necessarily be merged as above. One approach to estimation of the cointegrated VAR with s common cycles, is to use a procedure similar to the one considered for the I(2) model by Johansen (1995a); see also Paruolo (2000).[2] The first stage is to determine the cointegration rank, r, and obtain an estimate of β using, e.g., reduced rank regression, as in Johansen (1996). This stage may also involve estimation of β subject to linear restrictions.

[2] Maximum likelihood estimation may be considered using an algorithm similar to the one studied by Johansen (1997) when $s > r$. Alternatively, one may consider the iterative Gaussian reduced rank estimator suggested by Ahn (1997), where (β, ξ, ψ) (and Ω) are computed through an approximate Newton-Raphson algorithm; see also Ahn and Reinsel (1988). Furthermore, if β is known then the estimator of (ξ, ψ, Ω) discussed in this section is the maximum likelihood estimator.

In stage two, we condition on the estimated β from stage one and estimate ξ and η using reduced rank regression; see Hecq, Palm and Urbain (2001) for an alternative two-stage approach. That is, let $z_t = [\Delta x'_{t-1} \;\cdots\; \Delta x'_{t-k+1} \;\; (\beta' x_{t-1})']'$ so that

$$\Delta x_t = \xi \eta' z_t + \varepsilon_t. \tag{4}$$

From the reduced rank regression procedure discussed in, e.g., Johansen (1996, Chap. 6) we obtain estimates of ξ^* and η^* such that $(1/T)\sum_{t=1}^{T} \eta^{*\prime} z_t z_t' \eta^* = I_s$, while $\xi^* \eta^{*\prime} = \xi \eta'$. Since $h'h = I_s$ it follows that the estimate of η, which satisfies (3), can also be computed from $\eta = \eta^*(h'\eta^*)^{-1}$ while $\psi = h'_\perp \eta$.[3] Accordingly, the estimate of ξ must satisfy $\xi = \xi^* \eta^{*\prime} h$. Given $\hat{\Theta} = \hat{\xi}\hat{\eta}'$ we compute $\hat{\Omega} = (1/T)\sum_{i=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_t'$, where $\hat{\varepsilon}_t = \Delta x_t - \hat{\Theta} z_t$.[4]

Since the estimated Θ is asymptotically independent of the estimated β, and the latter converges weakly at a faster rate than the former, we may stop here. In practise, however, it may be advantagenous to consider iteration over the two stages, but where the first stage is slighly modified. In case β is generically identified in the sense of Johansen (1995b) we can write it as a function of its unique parameters. Let β_i be the i:th column of β with

$$\beta_i = m_i + M_i \varphi_i, \quad i \in \{1, \ldots, r\}, \tag{5}$$

where m_i and M_i are known and φ_i contains the free parameters for β_i. Letting vec be the column stacking operator, we can express the identified β as

$$\begin{aligned} \mathrm{vec}(\beta) &= \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_r \end{bmatrix} + \begin{bmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & M_r \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_r \end{bmatrix} \\ &= m + M\varphi. \end{aligned} \tag{6}$$

The gradient vector for the log-likelihood function in the direction of φ can be expressed as:

$$\frac{\partial \ln L}{\partial \varphi'} = T\big(\mathrm{vec}(\alpha)\big)' \Big[I_r \otimes \Omega^{-1}\big(M_{01} - \alpha\beta' M_{11} - \Gamma M_{21}\big)\Big] M, \tag{7}$$

where $\otimes$ is the Kronecker product, $\Gamma = [\Gamma_1 \;\cdots\; \Gamma_{k-1}]$, $M_{ij} = (1/T)\sum_{t=1}^{T} Z_{it} Z'_{jt}$, $Z_{0t} = \Delta x_t$, $Z_{1t} = x_{t-1}$, and $Z_{2t} = [\Delta x'_{t-1} \;\cdots\; \Delta x'_{t-k+1}]'$. Plugging the estimated parameters and product moment matrices into (7) we first check if the gradient is sufficiently close to zero. If it is, we are done. If not, we can solve for φ conditional on Θ and Ω by setting the gradient vector to zero. This provides us with

$$\varphi = \Big[M'\big(\alpha'\Omega^{-1}\alpha \otimes M_{11}\big) M \Big]^{-1} M' \Big[\mathrm{vec}\big([M_{10} - M_{12}\Gamma']\Omega^{-1}\alpha\big) - \big[\alpha'\Omega^{-1}\alpha \otimes M_{11}\big] m \Big]. \tag{8}$$

Conditional on the new β we can then re-estimate ξ, η, and Ω as in stage two. Given the outcome of this stage, we can compute the gradient vector in (7) and check if it is sufficiently close to zero. If it is the iterations are finished, if not we re-estimate β using (8), and so on until we have convergence.[5]

---

[3] To show this, notice that for $\eta = \eta^*(h'\eta^*)^{-1}$ the following $p(k-1) + r$ equations hold:

$$\begin{bmatrix} h' \\ h'_\perp \end{bmatrix} \eta = \begin{bmatrix} I_s \\ h'_\perp \eta^* (h'\eta^*)^{-1} \end{bmatrix}.$$

Since the inverse of the matrix which is multiplies by η is given by $[h \;\; h_\perp]$, premultiplication of both sides by this inverse gives us $\eta = h + h_\perp h'_\perp \eta^*(h'\eta^*)^{-1} = h + h_\perp \psi$. If h were such that $h'h \neq I_s$, the above relations would have to be changed to account for this.

[4] The vector z_t is actually an estimate of the unknown vector z_t whenever an estimate of β is used.

[5] It may be noted that the estimator of β could also be based on, e.g., cross-equation restrictions; see, e.g., Pesaran and Shin (2002). What is important here is that β is identified and that $\mathrm{vec}(\beta) = m + M\varphi$.

Suppose instead that β has not been generically identified as in (6) but has been obtained through reduced rank regression as in Johansen (1996, Chap. 6). Let us decompose η such that $\Gamma = \xi\eta_2'$ and $\alpha = \xi\eta_1'$. This means that (1) can be expressed as:

$$Z_{0t} = \xi\eta_2' Z_{2t} + \xi\eta_1'\beta' Z_{1t} + \varepsilon_t. \tag{9}$$

To determine if re-estimation of β is useful in the first place we can use the following expression (which we may think of as a "gradient matrix"), with β evaluated at its previous value:

$$\alpha'\Omega^{-1}\big(M_{01} - \Gamma M_{21} - \alpha\beta' M_{11}\big) = \eta_1\xi'\Omega^{-1}\big(M_{01} - \xi\eta_2' M_{21} - \xi\eta_1'\beta' M_{11}\big). \tag{10}$$

If this matrix is sufficiently close to zero we do not need to re-estimate β. If it is not close to zero, we may proceed by setting (10) to zero and solve for β conditional on ξ, η, and Ω. This gives us:

$$\beta = M_{11}^{-1}\big(M_{10} - M_{12}\eta_2\xi'\big)\Omega^{-1}\xi\eta_1'\big(\eta_1\xi'\Omega^{-1}\xi\eta_1'\big)^{-1}. \tag{11}$$

This represents the modified stage one for the iterations when β is not generically identified. As above, we can re-estimate ξ, η, and Ω as in stage two, but conditional on the new β. Given the outcome of this stage we can compute the right hand side of (10) and check if it is sufficiently close to zero. If it is, we are done, if not we continue until the algorithm has converged.

3. Asymptotic Properties of the Estimators

Suppose first that β is known. This means that the estimator of (ξ, η, Ω) is the maximum likelihood estimator and is equivalent to reduced rank regression using, e.g., the algorithm presented by Johansen (1996, Chap. 6); see, also, Anderson (1951). Accordingly, the estimated parameters solve the first order conditions for the log-likelihood function.

Let $\theta = \mathrm{vec}(\Theta)$ and $\gamma = [\mathrm{vec}(\xi)' \;\; \mathrm{vec}(\psi)']'$. This means that the partial derivatives of θ with respect to the free parameters γ is:

$$G(\gamma) \equiv \frac{\partial\theta}{\partial\gamma'} = \Big[\big[\eta \otimes I_p\big] \quad \big[h_\perp \otimes \xi\big] K_{p(k-1)+r-s,s}\Big], \tag{12}$$

where $K_{m,n}$ is the $mn \times mn$ commutation matrix; see Magnus and Neudecker (1988). That is, $\mathrm{vec}(\psi') = K_{p(k-1)+r-s,s}\mathrm{vec}(\psi)$. If there are exactly s common cycles, then $G(\gamma)$ has full column rank. The first order condition for γ is equal to the partial derivatives of the log-likelihood function with respect to θ times the matrix $G(\gamma)$. Let $M_{zz} = (1/T)\sum_{t=1}^{T} z_t z_t'$ while $M_{xz} = (1/T)\sum_{t=1}^{T} \Delta x_t z_t'$.[6] We then have that the gradient vector is given by

$$\frac{\partial \ln L}{\partial\gamma'} = T\big(\mathrm{vec}(M_{xz} - \Theta M_{zz})\big)'\big[I_{p(k-1)+r} \otimes \Omega^{-1}\big]G(\gamma). \tag{13}$$

Furthermore, since our estimator of γ is a maximum likelihood estimator, z_t is stationary, and the log-likelihood function satisfies standard regularity conditions (being the Gaussian log-likelihood), we know that $\hat\gamma$ is a consistent estimator of γ and, furthermore, that $\sqrt{T}(\hat\gamma - \gamma)$ converges in distribution to a Gaussian with zero mean and asymptotic covariance matrix given by the inverse of the information matrix for γ. Furthermore, $\hat\Omega$ is a consistent estimator of Ω, it is asymptotically independent of $\hat\gamma$ (which is the reason why the asymptotic covariance matrix for $\hat\gamma$ is given by the inverse of the information matrix for γ), and converges in distribution at the rate $\sqrt{T}$ to a Gaussian.

Specifically, by a Taylor expansion of the gradient in (13) and some standard rearrangement we have that:

$$\sqrt{T}\big(\hat\gamma - \gamma\big) = \Big[G(\gamma)'\big[M_{zz} \otimes \Omega^{-1}\big]G(\gamma)\Big]^{-1} G(\gamma)'\big[I \otimes \Omega^{-1}\big]\mathrm{vec}\bigg(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t z_t'\bigg) + o_p(1), \tag{14}$$

---

[6] Notive that $M_{xz} = [M_{02} \;\; M_{01}\beta]$, while

$$M_{zz} = \begin{bmatrix} M_{22} & M_{21}\beta \\ \beta' M_{12} & \beta' M_{11}\beta \end{bmatrix}.$$

where

$$\frac{\partial^2 \ln L}{\partial\gamma\partial\gamma'} = -TG(\gamma)'\left[M_{zz}\otimes\Omega^{-1}\right]G(\gamma), \tag{15}$$

Since $M_{zz} \xrightarrow{\text{p}} \mu$, a positive definite matrix, we find from, e.g., Lütkepohl (1991, Chap. 3) that

$$\text{vec}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t z_t'\right) \xrightarrow{\text{d}} N\big(0,\mu\otimes\Omega\big),$$

where $\xrightarrow{\text{p}}$ denotes convergence in probability and $\xrightarrow{\text{d}}$ convergence in distribution. It then follows that

$$\sqrt{T}\big(\hat{\gamma}-\gamma\big) \xrightarrow{\text{d}} N(0,\Sigma_\gamma), \tag{16}$$

where

$$\Sigma_\gamma = \Big[G(\gamma)'\big[\mu\otimes\Omega^{-1}\big]G(\gamma)\Big]^{-1},$$

is the inverse of the information matrix. Hence, the estimator of γ is asymptotically efficient since its asymptotic covariance matrix is equal to the Cramér-Rao lower bound. Furthermore, the asymptotic covariance matrix is equal to the one presented in Ahn (1997, Theorem 1) for the approximate Newton-Raphson algorithm for estimating all parameters jointly; see, also, Section 4.

Now, as noted above, θ is a nonlinear and continuously differentiable function of γ.[7] Based on well known results for such functions (see, e.g., Serfling, 1980, Chap. 3) we therefore have that

$$\begin{aligned}\sqrt{T}\big(\hat{\theta}-\theta\big) &= G(\gamma)\sqrt{T}\big(\hat{\gamma}-\gamma\big) + o_p(1)\\ &\xrightarrow{\text{d}} N\big(0,G(\gamma)\Sigma_\gamma G(\gamma)'\big).\end{aligned} \tag{17}$$

The asymptotic covariance matrix for θ therefore has reduced rank, and its rank is equal to the rank of Σ_γ, i.e., $s(pk+r-s)$.

Furthermore, and following the results in, e.g., Magnus and Neudecker (1988), the asymptotic behavior of the maximum likelihood estimator of Ω can directly be derived. Let $\omega = \text{vech}(\Omega)$, where vech is the column stacking operator which takes only the values on and below the diagonal. It can easily be shown (see, e.g., Magnus and Neudecker, 1988, or Lütkepohl, 1991) that $\hat{\omega}$ and $\hat{\gamma}$ are asymptotically independent by checking the matrix with second order cross partials. Furthermore, let D_p be the $p^2\times(p+1)p/2$ duplication matrix which equates $\text{vec}(\Omega) = D_p\omega$, while $D_p^+ = (D_p'D_p)^{-1}D_p'$ is its Moore-Penrose inverse. We then find that

$$\sqrt{T}\big(\hat{\omega}-\omega\big) \xrightarrow{\text{d}} N\big(0,2D_p^+[\Omega\otimes\Omega]D_p^{+\prime}\big). \tag{18}$$

Aagain, the asymptotic covariance matrix for $\hat{\omega}$ is equal to the inverse of the information matrix for ω and the estimator is therefore asymptotically efficient.

It may also be of interest to estimate the co-feature matrix $\xi_\perp$ and determine its asymptotic properties. Given a value of ξ we know the space spanned by the columns of $\xi_\perp$. However, we need to select a basis for this space to uniquely determine the elements of $\xi_\perp$ and to derive its asymptotic behavior.

To this end, let

$$\xi_\perp = H - H_\perp\big(\xi'H_\perp\big)^{-1}\xi'H, \tag{19}$$

where H is a known $p\times(p-s)$ matrix with full rank, and $H_\perp$ is $p\times s$ such that $\xi'H_\perp$ has rank s and $H'H_\perp = 0$. It is easily seen that $\xi'\xi_\perp = 0$ for this choice of basis for $\xi_\perp$. We then find that the partial derivatives of the elements of $\xi_\perp$ with respect to the elements of ξ is a $p(p-s)\times ps$

[7] The relationship between these parameters can actually be written as:

$$\theta = \left[\left[h + \frac{1}{2}h_\perp\psi\otimes I_p\right] \quad \left[\frac{1}{2}h_\perp\otimes\xi\right]K_{p(k-1)+r-s,s}\right]\gamma = QG(\gamma)\gamma,$$

where Q is a diagonal matrix with some elements equal to 1 and others equal to $1/2$.

matrix given by

$$\frac{\partial \text{vec}(\xi_\perp)}{\partial \text{vec}(\xi)'} = -\Big[\xi_\perp' \otimes H_\perp\big(\xi' H_\perp\big)^{-1}\Big] K_{p,s}, \tag{20}$$

since $H' - H'\xi(H_\perp'\xi)^{-1}H_\perp' = \xi_\perp'$. Since $\xi_\perp$ does not depend on ψ it follows that:

$$F(\gamma) \equiv \frac{\partial \text{vec}(\xi_\perp)}{\partial \gamma'} = \left[\frac{\partial \text{vec}(\xi_\perp)}{\partial \text{vec}(\xi)'} \quad 0\right]. \tag{21}$$

Let the estimator of the co-feature matrix $\xi_\perp$ be denoted as $\hat{\xi}_\perp$ and suppose that its obtained by inserting the estimator of ξ into equation (19). We then find that

$$\begin{aligned} \sqrt{T}\big(\hat{\xi}_\perp - \xi_\perp\big) &= F(\gamma)\sqrt{T}\big(\hat{\gamma} - \gamma\big) + o_p(1) \\ &\xrightarrow{\text{d}} N\big(0, F(\gamma)\Sigma_\gamma F(\gamma)'\big). \end{aligned} \tag{22}$$

Since $F(\gamma)$ has rank $(p-s)s$ it follows that the asymptotic covariance matrix for the estimator of the unique elements of $\xi_\perp$ has full rank (where the latter can be represented by the $s \times (p-s)$ matrix $H_\perp'\xi_\perp$). Moreover, since $\hat{\gamma}$ is asymptotically efficient $\hat{\xi}_\perp$ is efficient in this sense as well.

The above results are derived for the assumption that β is known. However, all results also hold when β is estimated and the estimator satisfies

$$\sqrt{T}\big(\hat{\beta} - \beta\big) \xrightarrow{\text{p}} 0.$$

Since, e.g., the Gaussian maximum likelihood estimator of β satisfies this super-consistency requirement, the above asymptotic results are valid for the two-stage approach to estimating a cointegrated VAR with common cycles restrictions. To see this, notice first that the $\hat{z}_t$ regressors, based on the estimated β, is related to the z_t regressors, using the true value of β, through

$$\hat{z}_t = \begin{bmatrix} \Delta x_{t-1} \\ \vdots \\ \Delta x_{t-k+1} \\ \hat{\beta}' x_{t-1} \end{bmatrix} = z_t + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ (\hat{\beta} - \beta)' x_{t-1} \end{bmatrix}.$$

It then follows that M_{zz} based on $\hat{z}_t$ converges in probability to μ, just like M_{zz} based on knowing β. Furthermore, and importantly for convergence in distribution, let $\hat{\varepsilon}_t = \Delta x_t - \Theta \hat{z}_t$. Then,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{z}_t' = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t z_t' + o_p(1). \tag{23}$$

Hence, the limiting behavior of $(1/\sqrt{T})\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{z}_t'$ is equivalent to the limiting behavior of $(1/\sqrt{T})\sum_{t=1}^{T} \varepsilon_t z_t'$. At the same time, using an estimator of β instead of a known matrix introduces a bias, which although asymptotically negligable, may still be relevant in small samples.

4. Estimation: A One-Stage Algorithm

Ahn (1997) has suggested that we can estimate φ and γ simultaneously through an iterative approximate Newton-Raphson algorithm. Hence, the algorithm is based on β being (generically) identified.[8] Let $\phi = [\varphi' \;\; \gamma']'$, the Ahn estimator is obtained from iteration on the following

[8] The algorithm in Ahn (1997) is actually based on $\beta = [I_r \;\; \beta_0']'$, where β_0 is an $(p-r) \times r$ matrix with the free parameters of the exactly identified β.

system:

$$
\begin{aligned}
\hat{\phi}^{(i+1)} &= \hat{\phi}^{(i)} - \left\{ \left[\frac{\partial^2 \ln L}{\partial\phi\partial\phi'} \right]^{-1} \frac{\partial \ln L}{\partial\phi} \right\}_{\phi=\hat{\phi}^{(i)}} \\
&= \hat{\phi}^{(i)} + \left\{ \begin{bmatrix} M'\left[\alpha'\Omega^{-1}\alpha \otimes M_{11}\right]M & M'K_{r,p}\left[\alpha'\Omega^{-1}\otimes M_{1z}\right]G(\gamma) \\ G(\gamma)'\left[\Omega^{-1}\alpha\otimes M_{z1}\right]K_{p,r}M & G(\gamma)'\left[M_{zz}\otimes\Omega^{-1}\right]G(\gamma) \end{bmatrix}^{-1} \times \right. \\
&\qquad \left. \times \begin{bmatrix} M'\left[I_r \otimes (M_{10} - M_{11}\beta\alpha' - M_{12}\Gamma')\Omega^{-1}\right]\mathrm{vec}(\alpha) \\ G(\gamma)'\left[I_{p(k-1)+r}\otimes\Omega^{-1}\right]\mathrm{vec}(M_{xz} - \Theta M_{zz}) \end{bmatrix} \right\}_{\phi=\hat{\phi}^{(i)}},
\end{aligned}
\tag{24}
$$

where $M_{z1} = (1/T)\sum_{t=1}^{T} z_t Z_{1t}' = [M_{12} \;\; M_{11}\beta]'$ and $M_{1z} = M_{z1}'$. The iterations have converged when the gradient vector is sufficiently close to zero and as starting values we may, e.g., choose the estimate of φ for unrestricted Θ, and the estimator of γ from the two-stage algorithm which is conditioned on the this initial estimate of φ. Asymptotically, the two-stage and the approximate Newton-Raphson estimators of φ and γ are equivalent. Furthermore, the two-stage estimator of φ and the maximum likelihood estimator of φ for unrestricted Θ are also asymptotically equivalent. Hence, the limiting behavior of all these estimators of φ does not depend on the reduced rank structure of Θ. Ahn (1997) notes, however, that for finite samples the efficiency in the sense of smaller mean squared error can be gained by exploring the structure among the parameters on stationary processes.

5. Testing

A natural approach to testing for s common cycles versus the alternative of $s+g$ (where $g \in \{1,\dots p-s\}$) within the current framework is to use a likelihood ratio test. Estimation under the null can be achieved as discussed above, while estimation under the alternative is as discussed above for all $g < p-s$ and using the unrestricted estimate of Θ when $g = p-s$.

Let $\hat{\Omega}_{s_j}$ be the estimator of Ω subject to s_j common cycles, with $s_j = s$ under the null and $s_j = s+g$ under the alternative. Asymptotically, it does not matter whether the two-stage or the one-stage algorithm has been used here. The likelihood ratio statistic can be written as:

$$
LR = T \ln\left(\frac{\det(\hat{\Omega}_s)}{\det(\hat{\Omega}_{s+g})} \right). \tag{25}
$$

Given the asymptotic normal behavior of $\hat{\theta}$ under the null hypothesis, the fact that all the common cycles restrictions are expressed as restrictions on parameters on stationary processes only, i.e., the restrictions do not depend on β (see, e.g., Sims, Stock and Watson, 1990), and the fact that the estimator of β converges to its true value in probability at the rate $\sqrt{T}$ (i.e., it is super-consistent), it follows that LR is asymptotically χ^2 with $q = g(pk + r - 2s - g)$ degrees of freedom. In the event that $g = p - s$, the number of degrees of freedom $q = (p-s)(p(k-1)+r-s)$.

An asymptotically equivalent form of the likelihood ratio test may also be derived. The estimator of η^*, i.e., the estimator obtained from the reduced rank regression before employing the normalization in (3), is obtained (just like in Johansen, 1996) by solving the eigenvalue problem

$$
\det\bigl(\lambda M_{zz} - M_{zx}M_{xx}^{-1}M_{xz}\bigr) = 0,
$$

for eigenvalues $1 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_p > \hat{\lambda}_{p+1} = \cdots = \hat{\lambda}_{p(k-1)+r} = 0$ and eigenvectors $\hat{V} = [\hat{V}_1 \; \cdots \; \hat{V}_{p(k-1)+r}]$, normalized such that $\hat{V}'M_{zz}\hat{V} = I$. Here, $M_{xx} = (1/T)\sum_{t=1}^{T}\Delta x_t \Delta x_t' = M_{00}$, while $M_{zx} = M_{xz}'$. The estimator of η^* is given by $\hat{\eta}^* = [\hat{V}_1 \; \cdots \; \hat{V}_s]$ under the null, and by $\hat{\eta}^* = [\hat{V}_1 \; \cdots \; \hat{V}_{s+g}]$ under the alternative. Since the maximized likelihood function is given by

$$
L_{\max}^{-2/T}(s_j) = \det\bigl(M_{xx}\bigr)\prod_{i=1}^{s_j}\bigl(1 - \hat{\lambda}_i\bigr),
$$

for any choice of $s_j \in \{r, \ldots, p\}$, the likelihood ratio test of s common cycles versus $s + g$ common cycles can then be formulated as:

$$LR_{\text{alt}} = -T \sum_{i=s+1}^{s+g} \ln\left(1 - \hat{\lambda}_i\right). \tag{26}$$

The two test statistics in (25) and (26) are numerically equivalent if (and only if) the estimator of β is not updated under the two-stage approach. In that case, $\hat{\Omega}_s$ and $\hat{\Omega}_{s+g}$ in (25) are computed using the same β matrix. The statistic in (26) always uses the same β matrix under the null and the alternative hypotheses.

The test statistic in (26) is similar to the statistic proposed by Vahid and Engle (1993) when $g = p - s$. In fact, the only difference is that (26) uses the number of observations, while the statistic in Vahid and Engle employs the number of observations corrected for the number of lags.

References

Ahn, S. K. (1997), "Inference of Vector Autoregressive Models with Cointegration and Scalar Components", *Journal of the American Statistical Association*, 92, 350–356.

Ahn, S. K. and Reinsel, G. C. (1988), "Nested Reduced-Rank Autoregressive Models for Multiple Time Series", *Journal of the American Statistical Association*, 83, 849–856.

Anderson, T. W. (1951), "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions", *Annals of Mathematical Statistics*, 22, 327–351.

Hecq, A., Palm, F. C. and Urbain, J.-P. (2000), "Permanent-Transitory Decomposition in VAR Models with Cointegration and Common Cycles", *Oxford Bulletin of Economics and Statistics*, 62, 511-532.

Hecq, A., Palm, F. C. and Urbain, J.-P. (2001), *Testing for Common Cyclical Features in VAR Models with Cointegration*, CESifo Working Paper No. 451.

Johansen, S. (1995a) "A Statistical Analysis of Cointegration for I(2) Variables", *Econometric Theory*, 11, 25–59.

Johansen, S. (1995b), "Identifying Restrictions of Linear Equations with Applications to Simultaneous Equations and Cointegration", *Journal of Econometrics*, 69, 111–132.

Johansen, S. (1996), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed., Oxford: Oxford University Press.

Johansen, S. (1997), "Likelihood Analysis of the I(2) Model", *Scandinavian Journal of Statistics*, 24, 433–462.

Lütkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Heidelberg: Springer-Verlag.

Magnus, J. R. and Neudecker, H. (1988), *Matrix Differential Calculus: With Applications in Statistics and Econometrics*, New York: John Wiley & Sons.

Paruolo, P. (2000), "Asymptotic Efficiency of the Two Stage Estimator in I(2) Systems", *Econometric Theory*, 16, 524–550.

Pesaran, M. H. and Shin, Y. (2002), "Long-Run Structural Modelling", *Econometric Reviews*, 21, 49–87.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, John Wiley: New York.

Sims, C. A., Stock, J. H. and Watson, M. W. (1990), "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 58, 113–144.

Vahid, F. and Engle, R. F. (1993), "Common Trends and Common Cycles", *Journal of Applied Econometrics*, 8, 341–360.