Estimating C Under Restrictions

Anders Warne

February 27, 2008

Abstract: This note discusses a simple switching algorithm for estimating a VAR model subject to cointegration restrictions under linear restriction on the moving average impact (C) matrix. The algorithm is based on the Gaussian likelihood function, but the estimated parameters are generally not maximum likelihood. Only in the rare event that either the cointegration space is known or we are unwilling to directly influence this space through restrictions on the C matrix can the estimates of the free parameters be considered maximum likelihood.

Keywords: C matrix, Cointegration, Maximum Likelihood, Vector Autoregression.

JEL Classification Numbers: C32.

1. Setup

The model is written as

$$y = \Psi X + \epsilon, \tag{1}$$

where y is $p \times T$ with typical column Δx_t, X is $d + p(k-1) + r \times T$ with typical column $X_t = [D_t, \Delta x_{t-1}, \ldots, \Delta x_{t-k+1}, \beta' x_{t-1}]$ where D_t is a vector with deterministic variables (constant, linear trend, seasonal dummies, etc.), while ϵ is $p \times T$, with $\epsilon_t \sim N(0, \Omega)$. We may, without loss of generality, also let any constant term or linear trend in an error correction model be restricted by a minor change in the definition of X_t. Regarding Ψ we may express it as $\Psi = [\mu \quad \Gamma_1 \cdots \Gamma_{k-1} \quad \alpha]$.

The restrictions of interest can be written as:

$$f(\Psi, \beta) = R\text{vec}(C) - r, \tag{2}$$

where $C = \beta_\perp (\alpha'_\perp \Gamma \beta_\perp)^{-1} \alpha'_\perp$, with $\Gamma = I_p - \sum_{i=1}^{k-1} \Gamma_i$.

The idea is to estimate Ψ and β under the restrictions in (2) using a switching algorithm, where Ψ (and Ω) is estimated conditional on β and β is then re-estimated conditional on Ψ (and Ω). We continue with this until we have obtained convergence.

2. Estimation of Ψ and Ω Conditional on β

By setting up the usual log-likelihood function we have that

$$\Lambda_T = -\frac{1}{2} pT \log(2\pi) - \frac{T}{2} \log\det(\Omega) - \frac{1}{2} \text{tr}(\Omega^{-1} \epsilon\epsilon'). \tag{3}$$

We shall maximize this function with respect to Ψ, Ω subject to (2). Letting $\psi = \text{vec}(\Psi)$ it can be shown that the following must be satisfied by a solution to the optimization problem:

$$\bar{\psi} = \hat{\psi} - [(XX')^{-1} \otimes \bar{\Omega}] g'_\psi \{g_\psi [(XX')^{-1} \otimes \bar{\Omega}] g'_\psi\}^{-1} f(\hat{\Psi}, \beta), \tag{4}$$

while $\bar{\Omega} = T^{-1} \bar{\epsilon}\bar{\epsilon}'$, where $\hat{\psi}$ is the unrestricted estimate of ψ. Moreover, g_ψ are the partial derivatives of f with respect to ψ', i.e.,

$$g_\psi = \frac{\partial f}{\partial \psi'} = R[\xi' \otimes C], \tag{5}$$

where $\xi' = C'(\iota_{k-1}' \otimes I_p)J_\Gamma + (C'\Gamma' - I_p)\alpha(\alpha'\alpha)^{-1}J_\alpha$, while $\Psi J_\Gamma' = [\Gamma_1 \cdots \Gamma_{k-1}]$ and $\Psi J_\alpha' = \alpha$. A similar expression for the partial derivatives of vec(C) with respect to ψ is given in Johansen's (1996) book and in Paruolo's (1997) article on the C matrix.

Notice first that I have used a first order Taylor approximation of the restrictions to obtain (4), namely

$$f(\bar{\psi}, \beta) = f(\hat{\psi}, \beta) + g_\psi(\bar{\psi} - \hat{\psi}). \tag{6}$$

Second, the equations do not provide us with a solution for $\bar{\psi}$ and $\bar{\Omega}$. They merely represent a form of expressing conditions which the restricted parameters must satisfy. Furthermore, they rely on $f(\bar{\psi}, \beta) = 0$. Finally, one need to keep in mind that multiple restrictions on C may turn out to be linearly dependent. In that case, the row rank of g_ψ is lower than the number of rows. Any such dependencies must be removed before estimation is attempted. This can typically be done by using random numbers for all the free parameters and checking the row rank of g_ψ. Should it be less than the number of rows, then remove all rows which are linearly dependent from g_ψ and remove the corresponding restrictions from $f(\Psi, \beta)$.

To use this in practise one approach is to proceed as follows:

(1) For the first iteration, replace $\bar{\Omega}$ with $\hat{\Omega}$ while g_ψ is evaluated at $\hat{\psi}$ and β (where the latter is just sloppy notation for the original estimate of β) in equation (4). This allows us to calculate $\bar{\psi}_1$ and $\bar{\Omega}_1$;
(2) For iteration i ($i = 2, 3, \ldots$) we replace $\hat{\psi}$ with $\bar{\psi}_{i-1}$, $\bar{\Omega}$ with $\bar{\Omega}_{i-1}$, g_ψ and f are evaluated at $\bar{\psi}_{i-1}$ and β in equation (4). This gives us $\bar{\psi}_i$ and $\bar{\Omega}_i$;
(3) We continue with the previous step until $f(\bar{\psi}_i, \beta)$ is sufficiently close to zero.

3. Estimation of β Conditional on Ψ and Ω

If we want to re-estimate β, which we probably want to since it can increase the value of the log-likelihood, then β is estimated conditional on $\bar{\psi}$. For that case, it is not necessary to involve the restrictions on C since they are already satisfied by the initial β.[1] Instead, one can estimate β conditional on $\bar{\psi}$ and $\bar{\Omega}$ subject to whatever restrictions we have on β. Once we have an updated β we merely return to the estimation of ψ conditional on the new β.

While there are several ways we can proceed, one natural approach is to return directly to the log-likelihood function, condition it on Ψ and Ω and maximize it with respect to β. Let us therefore consider the following representation of equation (1)

$$Z_0 - \Phi Z_2 - \alpha\beta' Z_1 = \epsilon, \tag{7}$$

where $\Phi = [\mu \quad \Gamma_1 \cdots \Gamma_{k-1}]$, $Z_0 = y$, etc. Substituting for ϵ in equation (3), evaluating (Ψ, Ω) at $(\bar{\Psi}, \bar{\Omega})$, and differentiating the log-likelihood function with respect to β we obtain:

$$\frac{\partial \Lambda_T}{\partial \mathrm{vec}(\beta)'} = \mathrm{vec}(Z_1\epsilon')'[\Omega^{-1}\alpha \otimes I_p]. \tag{8}$$

Next, we can express the restrictions on β in the following form for the columns of β:

$$\beta_i = h_i + H_i\varphi_i, \qquad i = 1, \ldots, r, \tag{9}$$

or

$$\mathrm{vec}(\beta) = h + H\varphi, \tag{10}$$

where φ gives all the free parameters of β. Technically, we may also consider further generalizations for restrictions on β, but for the purposes here the above format will do.

Taking the restrictions on β into account for the maximization of the conditional log-likelihood function means that the ML estimator of φ has to satisfy

$$\frac{\partial \Lambda_T}{\partial \varphi'} = \mathrm{vec}(Z_1\epsilon')'[\Omega^{-1}\alpha \otimes I_p]H = 0. \tag{11}$$

[1] They could, of course, still be included.

This implies that the (conditional) ML estimator of φ is given by:

$$\bar{\varphi} = \big[H'[\alpha'\Omega^{-1}\alpha \otimes M_{11}]H\big]^{-1} \\ H'[\alpha'\Omega^{-1} \otimes I_p]\big(\mathrm{vec}(M_{10} - M_{12}\Phi') - [\alpha \otimes M_{11}]h\big), \tag{12}$$

where $M_{ij} = (1/T)Z_iZ_j'$ while Φ, α, and Ω are evaluated at $\bar{\Phi}$, $\bar{\alpha}$, and $\bar{\Omega}$ from the previous step in the algorithm.

Notice that the β estimation part of the algorithm directly gives us a new value without resorting to any iterations. Had we used a general form for our restrictions on β this need not have been possible. Still, as long as the restrictions on β we wish to use are linear, then this part of the algorithm does not require iterations. For example, linear cross-equation (or, more accurately, cross-column) restrictions can be handled and would be more general than the simple restrictions in (9). Linear cross-equation restrictions can be expressed in the form of (10) and the expression for the (conditional) ML estimator of φ is therefore unchanged.

Given $\bar{\varphi}$ we have a new value for β and we may now return to the estimation of ψ conditional on this value for β. To begin with we evaluate the restrictions to check if they are sufficiently close to zero (a new value for β may affect the value for these restrictions). If the answer is no we repeat the steps for estimating ψ, Ω conditional on β and then β conditional on ψ.

4. Comments

This algorithm does not produce ML estimates unless β is known, in which case it doesn't have to be estimated, or we wish to exclude restrictions on C from appearing directly on β. A simple example of why the algorithm generally does not produce ML estimates is the following bivariate case with $k = 1$ and $r = 1$. Letting $\beta = [1 \quad \beta_1]'$ and $\alpha = [\alpha_1 \quad \alpha_2]'$ we find that C is given by:

$$C = \frac{1}{\alpha_1 - \alpha_2\beta_1}\begin{bmatrix} \alpha_2\beta_1 & -\alpha_1\beta_1 \\ -\alpha_2 & \alpha_1 \end{bmatrix}.$$

Suppose we want to estimate such a model under the restriction that $C_{12} = 0$. This holds if either $\alpha_1 = 0$ or $\beta_1 = 0$. Both of these restrictions cannot be satisfied as that would violate the condition that C has rank 1 (in fact, that would make the process for x_t I(2) unless $\alpha_2 = 0$). The algorithm will impose the $\alpha_1 = 0$ restriction, while $\beta_1 = 0$ would not be considered. Since the latter case may produce a higher value for the likelihood function than the former, the algorithm generally won't produce ML estimates. However, if β_1 is known or we wish to exclude the possibility that the first variable in the x_t vector is stationary, then the outcome of the algorithm is indeed ML. In practise, such circumstances will be rare.

Estimating Ψ, β and Ω under ML may require that the partial derivatives of C with respect to β are somehow taken into account. A switching algorithm based on the ideas above and this feature could potentially produce ML estimates under linear restrictions on C.

The asymptotic distribution of Ψ under C restrictions can easily be derived from equation (4). With the exception of parameters on the deterministic variables the limiting distribution of $\sqrt{T}(\bar{\psi} - \psi)$ is Gaussian with zero mean. Specifically, the asymptotic behavior of $\sqrt{T}(\bar{\psi} - \psi)$ is linearly related to the asymptotic behavior of $\sqrt{T}(\hat{\psi} - \psi)$ through:

$$\sqrt{T}(\bar{\psi} - \psi) = \big[I - [(XX')^{-1} \otimes \Omega]g_\psi'\{g_\psi[(XX')^{-1} \otimes \Omega]g_\psi'\}^{-1}g_\psi\big]\sqrt{T}(\hat{\psi} - \psi) + o_P(1).$$

The algorithm can also be used when additional restrictions on Ψ need to be considered. Such restrictions may be either linear or non-linear. For such cases we simply replace equations (2) and (5) with the corresponding form of the restrictions and their partial derivatives. Part of these may be linear restrictions on C. Whatever these restrictions may be it is important to keep in mind that g_ψ need not have full row rank, i.e., some of the restrictions may be linearly dependent. Should this occur, then the "redundant" restrictions need to be removed from f prior to estimation.

References

Johansen, S. (1996), *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, 2nd edition, Advanced Texts in Econometrics, Oxford: Oxford University Press.

Paruolo, P. (1997), "Asymptotic Inference on the Moving Average Impact Matrix in Cointegrated I(1) Systems", *Econometric Theory*, 13, 79–118.