Author
Elliott, Graham

Publication Date
1999-10-01

99-22

<u>UNIVERSITY OF CALIFORNIA, SAN DIEGO</u>

DEPARTMENT OF ECONOMICS

ESTIMATING RESTRICTED COINTEGRATING VECTORS

BY

GRAHAM ELLIOTT

DISCUSSION PAPER 99-22
OCTOBER 1999

Estimating Restricted Cointegrating Vectors

Graham Elliott
University of California, San Diego
9500 Gilman Drive,
LA JOLLA, CA, 92093-0508

First Draft: March 1997
This Draft: May 1999

Abstract:
This paper suggests the use of simple minimum distance methods to estimate restricted cointegrating vectors. The method directly employs minimum distance methods on unrestricted cointegrating matrices estimated in the usual way to estimate restricted parameters which are linearly or nonlinearly related to the unrestricted cointegrating vector coefficients. The limiting distribution of the estimates as well as the usual test for the restrictions are derived. A Monte Carlo experiment is undertaken to examine the effectiveness of these methods for cointegrating vectors.

JEL classification : C13, C32

Keywords: Cointegration, Minimum Distance Estimation, Nonlinear Restrictions.

1. INTRODUCTION

Theory often posits the existence of long run relationships between economic quantities. If the data contain a unit root, then cointegrating methods are appropriate. Quite a large number of methods for the estimation of cointegrating models are available (the most often employed estimator is due to Johansen (1988), see Watson (1994) for a review). Methods have been developed to estimate and/or test various linear restrictions on these cointegrating vectors, however many practical problems arise. Likelihood methods can be used in many cases (Johansen 1995) however algorithms for estimation are often quite complex.

This paper suggests a very simple asymptotically efficient two step method for undertaking the estimation of restricted cointegrating models. First, obtain unrestricted cointegrating vector estimates from one of the available methods. Second, use minimum distance methods to estimate and test the restricted cointegrating vector. The methods presented are valid for nonlinear as well as linear restrictions (with some regularity conditions on the functional relationship between the restricted and unrestricted estimates). The estimates have the property that if the original estimates have an asymptotic conditional mixed normal distribution, then so do the restricted estimates. The tests for overidentfying restrictions will have χ^2 limit distributions in this case.

The paper is set up as follows. The next section introduces the general theory showing consistency of the restricted parameter estimates and derives the asymptotic distributions of these estimates and the tests for overidentifying restrictions. The following section examines the properties of the suggested method for a variety of linear and nonlinear restrictions. The methods are then applied to the model of King, Plosser, Stock and Watson (1991). The final section concludes. Proofs are contained in Appendix A.

2. MINIMUM DISTANCE AND COINTEGRATION.

2.1. Model

The data y_t can be modeled as a vector autoregression

$$y_t = m + \Psi(L) y_{t-1} + e_t^*$$

where y_t is a nx1 vector, $\Psi(L)$ is a polynomial in the lag operator (potentially infinite order), m are deterministic terms and e_t^* is a martingale difference sequence residual. This can be rewritten as

$$\Delta y_t = m + \Psi y_{t-1} + \Pi(L)\Delta y_{t-1} + e_t^* \tag{1}$$

where $\Psi = -I_n + \sum_{i=1}\Psi_i , \Pi_i = -\sum_{j=i+1}\Psi_j$ and the sums run to the order of the polynomial $\Psi(L)$.

We assume that the data y_t individually have unit roots (are I(1)) so cointegration analysis is appropriate (Engle and Granger 1987). If there is a long run relationship in the data such that βy_t is stationary when y_t is nonstationary (where β is $n_2$xn and n_2<n), then this implies that in the above model that Ψ has reduced rank. As examined in Johansen (1988) amongst others, there are many ways to normalize the cointegrating vectors β, i.e. we can partition β into $[\beta_1\ \beta_2]$ where the partition is after the n_1th column, so that we can construct new cointegrating vectors by premultiplying by the inverse if the square matrix β_2 so long as this inverse exists, and obtain the cointegrating vectors $b_2^{-1}b = [b_2^{-1}b_1, I_{n2}] = [-\Gamma, I_{n2}]$. Note that the partition is arbitrary so long as the inverse exists.

Using this normalization and following Watson (1994), we can typically write this model in triangular form (see Phillips (1991)) as

$$\begin{aligned} y_{1t} &= d_{1t} + y_{1t-1} + u_{1t} \\ y_{2t} &= d_{2t} + \Gamma y_{1t} + u_{2t} \end{aligned} \tag{2}$$

where y_{1t} is n_1x1, y_{2t} is n_2x1, (i.e. conformably partitioned with β) and u_t = [u_{1t}' u_{2t}']' is an nx1 vector of potentially serially correlated residuals. Here d_{1t} and d_{2t} are deterministic terms where $d_{it}=\delta_i' z_{it}$, δ_i are unknown parameters and we consider three cases, $z_{1t}=z_{2t}=0$, $z_{1t}=0$ and $z_{2t}=1$, and finally $z_{1t}=1$ or 0 and $z_{2t}=[1,t]$. We assume that in the first case no deterministics are included in (1), in the second a constant is included, and in the third a constant and a time trend are included in the actual regression and that the restrictions implicit on m are not tested (the extension is straightforward). As the partition of the cointegrating vector was arbitrary, so is the partition of y_t. Define Ω as the spectral density matrix of u_t at frequency zero divided by 2π, and assume that u_t satisfies $T^{-1/2}\sum_{t=1}^{[Tl]} u_t \Rightarrow B(l) = \Omega^{1/2} W(l)$ for $l \in [0,1]$ where W(λ) is an n dimensional standard Brownian motion and ⇒ denotes weak convergence.

In this paper we are interested in restrictions of the form $vec(\Gamma) = g(q)$ where θ is a qx1 vector of unknown parameters with true values θ_0 and $q<n_1n_2$ (i.e. contains less parameters than does Γ). The mapping may be linear or nonlinear. We are interested in obtaining estimates and limit distributions for θ and testing the over-identifying restriction $vec(\Gamma) = g(q)$.

With various assumptions on the error terms and the lag polynomial in (1) or their equivalents, many papers have derived estimators for the cointegrating vector Γ (Johansen 1988, Ahn and Reinsel 1990, Phillips and Hansen 1990, Stock and Watson 1993, Saikkonen 1991,1992, see Watson 1994 for a review). In this paper we will derive minimum distance methods applicable to the unrestricted estimates of $\hat{\Gamma}$ for all of these methods.

2.2. Minimum Distance

We will obtain estimates for θ using standard minimum distance methods applied to the unrestricted estimates of Γ . Thus our estimate for θ, denoted $\hat{q}$, is the θ that minimizes

$$\left\{vec(\hat{\Gamma}) - g(q)\right\}'\hat{V}^{-1}\left\{vec(\hat{\Gamma}) - g(q)\right\} \tag{3}$$

where $\hat{V}$ is an estimate of the variance covariance matrix of the vec($\hat{\Gamma}$) estimates.

For linear relationships between Γ and θ a closed form solution will be available for $\hat{q}$. In this case we have restrictions of the form $vec(\Gamma) = Gq + r$, where G is (n_1n_2 x q) with all elements known and r is (n_1n_2 x 1) again with elements known. Here the $\hat{q}$ minimizing (3) is

$$\hat{q} = (G'\hat{V}^{-1}G)^{-1}\left(G'\hat{V}^{-1}\left[vec(\hat{\Gamma}) - r\right]\right) \tag{4}$$

All linear restrictions (within and cross equation) yield a closed form solution. For restrictions that cannot be written in this form search procedures will be required to estimate θ.

2.3. Limiting Results

Theorem 1. (Consistency of $\hat{q}$)

If (a) plim $\hat{\Gamma}$ *= G , (b)* $\hat{V}$ *=> V_n for some V_n positive definite with probability one and full rank, (c) g(q) is continuous in q, (d) Q is a compact subset of* $\hat{A}^q$ *containing q_0 (the true q) (e) g(q) = g(q_0) for q Î Q implies q = q_0., then* $\hat{q} \xrightarrow{p} q_0$.

This result is not very restrictive, allowing estimation of $\hat{q}$ from any consistent estimates of $\hat{\Gamma}$. This includes those that have asymptotically mixed normal asymptotic distributions or even simple OLS estimates which are known to be inefficient. The weighting matrix used may be constant or converge to a matrix which is potentially random so long as it is positive definite (i.e. we need not use $\hat{V}$ here but could employ any positive definite weight matrix).

The main restriction placed on the problem in Theorem 1 is the limiting of the types of restrictions that can be handled (i.e. the g(.) functions), however this is just the standard restriction in the minimum distance literature. Restrictions that can be handled include all forms of linear restrictions, either within or across equations. The limitation is on the forms the nonlinear restrictions can take, they must be such that g(θ) is continuous. Thus consistent estimates of a wide range of restricted cointegrating vector parameters are available from this simple procedure.

Whilst the results of Theorem 1 show that we can turn any consistent unrestricted estimates into consistent restricted estimates, we are more likely to prefer to use estimators that efficiently (in the case of normal errors) use the information in the simultaneity of the residuals. These are the so called optimal set of estimators (see Saikkonen 1991 or Phillips 1991). Estimators in this class all have the mixed normal limit distribution conditional on y_{1t}

$$\hat{V}^{-1/2}T[vec(\hat{\Gamma}) - g(q_0)] \Rightarrow vec\left\{(\int dW_{2.1}W_1^d{}')(\int W_1^d W_1^d{}')^{-1/2}\right\} \sim MN(0, I_{n1n2})$$

when suitably scaled where $W(l) = [W_1(l)', W_2(l)']'$ and the partition is after the n_1th row of W, $W_{2.1}(l) = W_2(l) - \Omega_{21}\Omega_{11}^{-1}W_1(l)$, $\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$ where the partition is after the n_1th row and column and $W^d(l)$ is a detrended Brownian Motion where

(i) $W^d(l) = W(l)$ if no deterministic terms are included,

(ii) $W^d(l) = W(l) - \int_0^1 W(s)ds$ if a constant is included,

(iii) $W^d(l) = W(l) - (4 - 6l)\int_0^1 W(s)ds - (12l - 6)\int_0^1 sW(s)ds$ if a constant and time trend are included.

The scaling for this class of estimators (and weighting matrix used) is of the form $\hat{V} = \hat{V}_1 \otimes \hat{\Omega}_{2.1}$ where $\Omega_{2.1} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}$, $\hat{\Omega}_{2.1}$ is a consistent estimator of $\Omega_{2.1}$ and $\hat{V}_1^{-1}$ is such that it converges to $\int B_1^d B_1^d{}'$ where $B^d(l) = \Omega^{1/2}W^d(l)$ and $B_1^d(l)$ are the first n_1 rows of $B^d(l)$. Different estimators of the unrestricted cointegrating vectors suggest different estimators of $\hat{V}$. They will be examined in the next section. The remainder of the paper deals with unrestricted estimates of estimators that are in this class.

Theorem 2 (Asymptotic distribution of $\hat{q}$)

If (a) $\hat{V}^{-1/2}T[vec(\hat{\Gamma}) - g(q_0)] \Rightarrow vec\left\{(\int dW_{2.1}W_1^d{}')(\int W_1^d W_1^d{}')^{-1/2}\right\} \sim MN(0, I_{n1n2})$ (b) $\hat{V}$ => $V_{n,}$ (c) Q has a neighborhood of q_0 in which g(.) has continuous second partial derivatives and (d) rank $G(q_0)$ =q where $G(q_0) = \left.\frac{¶g(q)}{¶q'}\right|_{q_0}$ then $T(\hat{q} - q_0)$ has a mixed normal limiting distribution conditional on y_{1t} with mean zero and limiting variance covariance matrix
$(G(q_0)'V_n^{-1}G(q_0))^{-1}$.

Theorem 2 limits the results to the estimates for $\hat{\Gamma}$ that have asymptotic mixed normal distributions, obtaining asymptotic distributions for $\hat{q}$ that also have asymptotic mixed normal distributions. These are the standard estimators used in practice. An analog to Theorem 2 for

OLS estimates of $\hat{\Gamma}$ is available by the same method as the proof of Theorem 2 but results in a nonstandard asymptotic distribution for $\hat{q}$.

In practice we need to estimate the variance covariance matrix of the estimates, and G is often unknown when there are nonlinear restrictions as it is a function of the unknown restricted parameters θ_0 . This approximation can be done by using $\hat{q}$ in place of θ_0 and noting that by the consistency results of Theorem 1 the estimated $G(\hat{q})$ converges to the true $G(q_0)$.

The variance covariance matrix, as in the unrestricted case, is stochastic however this presents no problems in practice as in the unrestricted case. Hypothesis tests and confidence intervals can be set up in the usual way.

We can test the above restrictions by constructing the test

$$J_T = T^2 \left\{ vec(\hat{\Gamma}) - g(\hat{q}) \right\}' \hat{V}^{-1} \left\{ vec(\hat{\Gamma}) - g(\hat{q}) \right\}$$

where we multiply the function used to estimate θ by T^2 and replace θ with $\hat{q}$.

Theorem 3 (Asymptotic distribution of the test for overidentifying restrictions)

Under the conditions of Theorem 2 the test for overidentifying conditions J_T has a c^2 distribution conditional on y_{1t} with n_1n_2-q degrees of freedom.

This theorem shows the usual result that the standard overidentifying test has a chi squared distribution with degrees of freedom equal to the number of restrictions (n_1n_2 - q) being tested. The test itself is simple to implement as it only requires objects that have already been calculated to estimate $\hat{q}$.

Tests of a subset of the over-identifying restrictions available through these methods are available directly in a number of ways. For a subset of the possible linear restrictions, Johansen (1995) details likelihood ratio tests for restrictions. Warne (1997) considers non linear Wald tests in a Sims, Stock and Watson (1990) framework which includes cointegration. In general, all of the methods can be used to construct Wald statistics of linear restrictions on the cointegrating vectors after they have been normalized. The J_T test when restrictions are linear will result in numerically equivalent tests to these Wald tests using the same unrestricted estimates of the cointegrating vector and variance covariance matrix.

2.4. Optimality.

Saikkonen (1991) and Phillips (1991) detail optimality properties for the unrestricted cointegrating vector estimators when residuals are normally distributed. This optimality occurs within the locally asymptotically mixed normal (LAMN) family. The methods developed above retain optimality within this family by choosing the variance covariance matrix as the weighting matrix. Of course in small samples there may be a loss of efficiency in small samples as $\sqrt{T}$ convergent parameters are treated as nuisance parameters rather than estimated simultaneously. As in the regular application of minimum distance (Rothenberg 1973) any other choice of weighting matrix than the one suggested here results in a wider asymptotic distribution of the restricted estimates.

3. EXAMPLES AND A MONTE CARLO EVALUATION

This section examines a number of examples of restrictions and also examines the performance of the procedures for different initial estimators of the cointegrating vector by way of Monte Carlo experiments. In each case we examine the performance in large samples (showing the above limit theory to be applicable) and in small samples (to examine the applicability of the large sample theory for samples encountered in practice).

There are available a reasonably large number of estimators that satisfy the restrictions of Theorems 1 and 2. These were classed by Saikkonen (1991) as the efficient class of estimators and include those of Johansen (1988), Phillips and Hansen (1990), Ahn and Reinsel (1990), Phillips and Loretan (1991), Saikkonen (1991,1992), Stock and Watson (1993) etc., see Watson (1994) for a review. Each of these provide asymptotically similar estimators of the cointegrating vector and have associated with them asymptotically similar estimators of the variance covariance matrix of these estimates. Here we will evaluate the above methods for three different estimators; the Saikkonen (1992) VAR method, the Johansen (1988) full information maximum likelihood estimator, and the DOLS (Stock and Watson 1993) method. These methods are detailed in an appendix. As each differ in their small sample results, Monte Carlo results for rejection rates under the null and alternative will be presented for each of these estimators.

3.1. Setting a cointegrating vector element to a known value.

Consider a model where $n_1 = 2$ (there are two distinct unit roots in the model), $n_2=1$ (there is only one cointegrating vector) so $\Gamma = [\Gamma_1, \Gamma_2]$. Further, we will estimate Γ_2 given that we know the true value for Γ_1. In this case the restriction can be written as vec(Γ) = $G\theta + r$ where $G = [0, 1]'$ and $r = [\Gamma_1, 0]'$. The estimator for $\hat{q} = \hat{\Gamma}_2 = (G'\hat{V}^{-1}G)^{-1}(G'\hat{V}^{-1}[\hat{\Gamma}'-r])$. As results are numerically equivalent for all Γ_1 we set $\Gamma_1 = 0$.

Results when the null hypothesis is true are given for each of the three sets of unrestricted estimates for T=50, 100 and 250 and for three different values of the variance covariance matrix in Table 1A. The results presented in the first six columns are features of the original estimator - average bias and standard error of the estimated bias as well as size for each of the two coefficients in the unrestricted model. The final four columns give the same statistics for the restricted estimate as well as size for a test of the (true) restriction using the J_T statistic.

A number of results are apparent. First, the performance of the asymptotic approximation in small samples does not appear to be qualitatively any different to those for the unrestricted estimates. This can be seen by comparing the tests of the unrestricted estimates to their true values and also examining the test of the restricted estimate for its true value and the test for overidentification. The size distortion on the restricted coefficient is nearly always less than or equal to the size distortion on the unrestricted estimates. As the sample size increases the differences are less (size distortions are all disappearing). Second, the restricted estimates vary less than the unrestricted estimates. This is a reflection of their smaller sampling variance.

Third, the test for overidentifying restrictions appear also to have similar small sample properties to the unrestricted estimates. In general the size performance of these statistics is no worse than for the usual cointegrating estimates.

Table 1: Size Results for Model 3.1.

	$E(\hat{\Gamma}_1)$	$st.dev(\hat{\Gamma}_1)$	$test\ (\hat{\Gamma}_1 = \Gamma_1)$	$E(\hat{\Gamma}_2)$	$st.dev(\hat{\Gamma}_2)$	$test\ (\hat{\Gamma}_2 = \Gamma_2)$	$E(\hat{q}_1)$	$st.dev(\hat{q}_1)$	$test(\hat{\Gamma}_2 = \Gamma_2)$	J_T rej.
T=50	Ω_{12}=0.5	Ω_{13}=0.3	Ω_{23}=0.5							
Saik	0	0.089	0.121	0.006	0.091	0.124	0.002	0.064	0.113	0.124
Johan	0	0.101	0.144	-0.002	0.104	0.139	-0.001	0.07	0.123	0.139
DOLS	-0.001	0.087	0.071	0.001	0.087	0.072	0	0.063	0.069	0.072
T=100										
Saik	0	0.043	0.095	0.002	0.042	0.09	0	0.031	0.086	0.09
Johan	0	0.044	0.097	0	0.043	0.09	-0.001	0.032	0.085	0.09
DOLS	0	0.042	0.065	0	0.041	0.063	0	0.031	0.063	0.064
T=250										
Saik	0	0.016	0.063	0	0.016	0.061	0	0.012	0.06	0.061
Johan	0	0.016	0.062	0	0.016	0.061	0	0.012	0.059	0.061
DOLS	0	0.016	0.054	0	0.016	0.053	0	0.012	0.052	0.053
T=50	Ω_{12}=0.5	Ω_{13}=0.5	Ω_{23}=0.5							
Saik	0.003	0.085	0.122	0.004	0.085	0.124	0.004	0.061	0.114	0.124
Johan	-0.001	0.094	0.138	-0.001	0.096	0.133	-0.001	0.066	0.12	0.134
DOLS	-0.001	0.082	0.071	0.001	0.082	0.072	0	0.06	0.069	0.072
T=100										
Saik	0.001	0.04	0.095	0.001	0.039	0.09	0.001	0.029	0.084	0.09
Johan	0	0.041	0.096	0	0.041	0.09	0	0.03	0.083	0.09
DOLS	0	0.04	0.065	0	0.039	0.063	0	0.029	0.063	0.064
T=250										
Saik	0	0.015	0.062	0	0.015	0.061	0	0.011	0.059	0.061
Johan	0	0.015	0.062	0	0.015	0.06	0	0.011	0.059	0.06
DOLS	0	0.015	0.054	0	0.015	0.053	0	0.011	0.052	0.053
T=50	Ω_{12}=0.5	Ω_{13}=0.7	Ω_{23}=0.5							
Saik	0.005	0.072	0.122	0.002	0.072	0.118	0.004	0.052	0.111	0.118
Johan	-0.001	0.079	0.127	0	0.079	0.124	0	0.056	0.114	0.124
DOLS	-0.001	0.07	0.071	0	0.07	0.072	0	0.051	0.069	0.072
T=100										
Saik	0.001	0.034	0.091	0	0.033	0.087	0.001	0.025	0.083	0.087
Johan	0	0.035	0.091	0	0.034	0.084	0	0.025	0.08	0.084
DOLS	0	0.034	0.065	0	0.033	0.063	0	0.024	0.063	0.064
T=250										
Saik	0	0.013	0.062	0	0.012	0.059	0	0.009	0.059	0.059
Johan	0	0.013	0.06	0	0.013	0.059	0	0.01	0.058	0.059
DOLS	0	0.013	0.054	0	0.012	0.053	0	0.009	0.052	0.053

Note: In all cases unknown parameters in Γ are set to one. The data is generated according to equation (2) with no deterministics but estimated with a constant in the regressions. We set Φ(L)=I in the model and this is known in the simulations. The variance covariance matrix for $u_t=\varepsilon_t$ has ones on the diagonal and off diagonal elements as indicated. The variance covariance matrices for the unrestricted cointegrating vector estimates are calculated as in Appendix B. All the reported tests on the unrestricted estimates are Wald tests (t tests) based on these VCV matrices. Results are calculated from 5000 simulations.

Fourth, whilst the individual size performance and biases of the estimates vary slightly with the variance covariance matrix, the relationship between the performance of tests in the unrestricted model and the restricted model detailed above holds for each of the models.

Rather than compare directly to the unrestricted estimates, we can compute the (pseudo) maximum likelihood estimates of θ. In Table 2 we examine such estimates from the model

$$\Delta y_t = m + \Phi(y_{2t-1} - \Gamma y_{1t-1}) + h_t$$

where Γ is restricted as above, m and Φ are nx3 vectors of unknown parameters, and we assume that $h_t \sim N(0,\Sigma)$. This is the implied error correction form for the model investigated above, so we do not use the Johansen method as it assumes less information and hence would be less efficient.

Table 2: Restricted Estimates

Ω_{12}	Ω_{13}	Ω_{23}	Bias (mle)	Bias (md)	MSE(mle)	MSE (md)
Panel 1						
0.5	0.3	0.5	0.001	0.015	0.089	0.087
0.5	0.5	0.5	-0.001	0.014	0.078	0.078
0.5	0.7	0.5	-0.002	0.012	0.056	0.056
Panel 2						
0.5	0.3	0.5	-0.019	-0.030	0.071	0.067
0.5	0.5	0.5	-0.016	-0.023	0.060	0.057
0.5	0.7	0.5	-0.010	-0.013	0.045	0.043
Panel 3						
0.5	0.3	0.5	-0.084	0.051	0.307	0.013
0.5	0.5	0.5	0.035	0.048	0.011	0.011
0.5	0.7	0.5	-0.273	0.049	0.962	0.008
Panel 4						
0.5	0.3	0.5	0.049	0.010	0.028	0.028
0.5	0.5	0.5	0	-0.030	0.022	0.022
0.5	0.7	0.5	0.01	-0.010	0.014	0.014

Note: Reported are averages over 2000 replications with 100 observations in each pseudo-sample. The model is as in equation (2) with $u_t \sim N(0,\Omega)$, $d_{1t}=d_{2t}=0$ but a constant is estimated.

The results presented are for the same three models (parameterizations of the variance covariance matrix) and T=100. Each panel contains the average estimated bias (times 100) and mean square error (times 100) of the restricted estimates, with Panel 1 according to the model and Γ of this subsection (other panels are for models in subsections 3.2 to 3.4). For this model there are no differences between the MSE's for either estimation method, with the average bias being a little smaller for the MLE. Thus we see here that the simpler minimum distance method involves very little if any loss over the correctly specified MLE.

3.2. Cross equation linear restrictions

In this case we have n_1 =1 and n_2=2 so there are two cointegrating vectors, and $\Gamma = [\ \Gamma_1\ \Gamma_2]'$. The restriction we will test is that $\theta = \Gamma_1 = \Gamma_2$, i.e. that the two cointegrating vectors are equivalent. In this case G = [1, 1]'. In this case again the closed form solution $\hat{q} = \hat{\Gamma}_2 = (G'\hat{V}^{-1}G)^{-1}(G'\hat{V}^{-1}\hat{\Gamma})$ is available. The results here are similar in nature to those in the model in Section 3.1. For this reason only one panel of the results (comparable to Table 1) is reported in the first panel of Table 3, that for when $\Omega_{12} = \Omega_{13} = \Omega_{23} = 0.5$. The restricted estimates are more precise, and the hypothesis tests both on the restricted estimates and the overidentifying restrictions have similar size properties to the tests on the unrestricted estimates.

Panel 2 of Table 2 gives the restricted estimates for this model using minimum distance and the MLE as above. Here the correctly specified MLE's have mean square errors slightly above than the minimum distance estimator, however they are fairly close.

Table 3 : Size for Models 2.2,2.3 and 2.4 for Ω_{12}=0.5, Ω_{13}=0.5, Ω_{23}=0.5

	$E(\hat{\Gamma}_1-\Gamma_1)$	$st.dev(\hat{\Gamma}_1)$	$test\ (\hat{\Gamma}_1=\Gamma_1)$	$E(\hat{\Gamma}_2-\Gamma_2)$	$st.dev(\hat{\Gamma}_2)$	$test\ (\hat{\Gamma}_2=\Gamma_2)$	$E(\hat{q}_1-q_1)$	$st.dev(\hat{q}_1)$	$test\ (\hat{q}_1=q_1)$	J_T rej.
T=50						Model 2.2				
Saik	0.001	0.063	0.092	0.002	0.064	0.102	0.002	0.053	0.117	0.11
Johan	-0.002	0.067	0.098	-0.001	0.069	0.108	-0.001	0.056	0.122	0.118
DOLS	-0.001	0.062	0.059	-0.001	0.063	0.065	-0.001	0.051	0.072	0.061
T=100										
Saik	0.001	0.031	0.07	0	0.031	0.083	0.001	0.026	0.081	0.074
Johan	0.001	0.031	0.07	0	0.032	0.079	0	0.026	0.082	0.074
DOLS	0.001	0.03	0.057	0	0.031	0.061	0	0.025	0.063	0.059
T=250										
Saik	0	0.012	0.066	0	0.012	0.058	0	0.01	0.065	0.062
Johan	0	0.012	0.069	0	0.012	0.059	0	0.01	0.064	0.061
DOLS	0	0.012	0.062	0	0.012	0.051	0	0.01	0.059	0.053
T=50						Model 2.3				
Saik	0.003	0.085	0.122	0.004	0.085	0.124	0.002	0.023	0.115	0.127
Johan	-0.001	0.094	0.138	-0.001	0.096	0.133	0	0.026	0.123	0.14
DOLS	-0.001	0.082	0.071	0.001	0.082	0.072	0	0.023	0.07	0.076
T=100										
Saik	0.001	0.04	0.095	0.001	0.039	0.09	0	0.011	0.078	0.091
Johan	0	0.041	0.096	0	0.041	0.09	0	0.011	0.081	0.095
DOLS	0	0.04	0.065	0	0.039	0.063	0	0.011	0.058	0.069
T=250										
Saik	0	0.015	0.062	0	0.015	0.061	0	0.004	0.06	0.062
Johan	0	0.015	0.062	0	0.015	0.06	0	0.004	0.057	0.062
DOLS	0	0.015	0.054	0	0.015	0.053	0	0.004	0.051	0.052
T=50						Model 2.4				
Saik	0.001	0.063	0.092	0.002	0.064	0.102	0.001	0.032	0.12	0.098
Johan	-0.002	0.067	0.098	-0.001	0.069	0.108	-0.001	0.035	0.124	0.108
DOLS	-0.001	0.062	0.059	-0.001	0.063	0.065	-0.001	0.031	0.072	0.061
T=100										
Saik	0.001	0.031	0.07	0	0.031	0.083	0	0.016	0.085	0.073
Johan	0.001	0.031	0.07	0	0.032	0.079	0	0.016	0.084	0.074
DOLS	0.001	0.03	0.057	0	0.031	0.061	0	0.015	0.06	0.057
T=250										
Saik	0	0.012	0.065	0	0.012	0.059	0	0.006	0.059	0.062
Johan	0	0.012	0.067	0	0.012	0.059	0	0.006	0.059	0.065
DOLS	0	0.012	0.06	0	0.012	0.052	0	0.006	0.052	0.057

Note: As per Table 1.

3.3. Non- linear Restriction

This model is similar to that in Section 3.1 except that now the restriction is that $\theta = \Gamma_1 = \Gamma_2^{1/2}$. In this case there is no closed form solution so numerical optimization was used. Here the restrictions can be written as $vec(\Gamma) = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix} = \begin{bmatrix} q \\ q^2 \end{bmatrix} = g(q)$ so the matrix G is a function of θ. In constructing the variance covariance matrix for $\hat{q}$ we use the consistent estimate $\hat{q}$. In this case $G(\theta) = [1, 2\theta]'$.

Given the many approximations used to derive the limit distributions in this case we pay particular attention to the size results for the test of overidentifying conditions. Again, as the results are similar over different models only the results for the same variance covariance matrix as in Section 3.2 are reported in the second panel of Table 3. We see that there are no major problems apparent here, indeed qualitatively there is no real difference in the results from those in Section 3.1. The size performance for the tests on the restricted estimates is better in each case than the unrestricted estimates. The test of overidentifying restrictions has similar size performance to the tests on the unrestricted estimates.

Comparing the correctly specified restricted MLE's to the minimum distance estimates (Panel 3 of Table 2), in each model here the minimum distance method has either the same or slightly better MSE. The difference in the third row of the panel is due to a small number of outlier MLE's, which presumably would disappear as the sample size increases.

3.4. Cross Equation Non-Linear Restrictions.

This case is the same as in Section 3.2 except now the restriction is that $\theta = \Gamma_1 = \Gamma_2^{1/2}$. Again there is no closed form solution so optimization methods were used. The matrix G(θ) in this case is G(θ) = [1 2θ]'.

Results for the model with $\Omega_{12} = \Omega_{13} = \Omega_{23} = 0.5$ are reported in the last panel of Table 3. Again, estimates are considerably more precise that the unrestricted estimates and the size properties of hypothesis tests on unrestricted estimates are similar to tests on restricted estimates.

Comparing again the restricted MLE's to the minimum distance estimates for this model, Panel 4 of Table 2 shows similar results to that of the model in the previous subsections. Overall, the relationships between the estimators are that the results are model dependent, but fairly close for each of the models and hence there does not appear to be too much if any loss from using minimum distance over the correctly specified maximum likelihood estimator.

4. EXAMPLE

King et al (1991) identify a six variable VAR of the macro economy with cointegration amongst the variables. The variables in the system are the M2 measure of money supply (m_t), prices (p_t), real consumption (c_t), real investment (i_t), real private income (y_t^p) and the 3 month Treasury Bill (R_t). All variables except the interest rate are in logs. They have three restricted cointegrating equations in their system; these are

Consumption $\quad c_t - y_t^p = m_{21} + f_1(R_t - \Delta p_t) + u_{21t}$

Investment $\quad i_t - y_t^p = m_{22} + f_2(R_t - \Delta p_t) + u_{22t}$

and Money Demand $\quad m_t - p_t = m_{23} + b_Y y_t^p + b_R R_t + u_{23t}$.

Ordering the variables [y_t^p, R_t, Δp_t, c_t, i_t, $(m_t$-$p_t)$] where the first three variables make up y_{1t} of the notation in the theoretical part of the paper and the remaining three variables make up y_{2t} in the theoretical part of the paper we have a set of cointegrating relationships given by

$b = \begin{bmatrix} -1 & -f_1 & f_1 & 1 & 0 & 0 \\ -1 & -f_2 & f_2 & 0 & 1 & 0 \\ -b_Y & -b_R & 0 & 0 & 0 & 1 \end{bmatrix}$ thus we have $\Gamma = \begin{bmatrix} 1 & f_1 & -f_1 \\ 1 & f_2 & -f_2 \\ b_Y & b_R & 0 \end{bmatrix}$. If we were to estimate the unrestricted cointegrating vector, as reported in Table 3 Panel A of King et al (1991), we would obtain nine unrestricted estimates. The restrictions show that there are four free parameters in the cointegrating vector - we have two within equation restrictions and three known values so have 5 overidentifying restrictions.

To estimate the restricted cointegrating vectors, there is no natural system reparameterization that allows restricted estimation. King et al. (1991) estimate the model equation by equation, imposing the restrictions for each equation. The methods above allow the restrictions to be directly employed to obtain the efficient system estimates of the restricted cointegrating vector coefficients. As the restrictions are linear here a closed form solution is available. The restrictions can be written as vec(Γ) = Gθ + r where

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\ r = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ and } q = \begin{bmatrix} f_1 \\ f_2 \\ b_Y \\ b_R \end{bmatrix}.$$

The restricted estimates and their standard estimates from King et al (1991) are reproduced in the first two columns of Table 4 below (All regressions here correspond to those in King et al (1991) including sample length (quarterly data 1954:1 to 1988:4, earlier data used for lags) and number of lags in DOLS estimation procedure (5 leads and lags, 4 periods used to estimate autoregressive robust errors, we thank Jim Stock for supplying data for this analysis and programs to reproduce these results). The remaining two columns give the minimum distance system restricted estimates and their standard errors.

Table 4: Restricted Cointegrating Vector Estimates

	King et al.(1991)		*Minimum Distance*	
	Est.	s.e	Est.	s.e
ϕ_1	0.0033	0.0022	0.0030	0.0015
ϕ_2	-0.0028	0.0050	0.0041	0.0029
β_Y	1.197	0.062	1.242	0.041
β_R	-0.013	0.004	-0.015	0.003

Note: Est. refers to the point estimate, s.e. refers to the estimate of the standard error.

Several points are noticeable. First, the minimum distance estimates all have smaller estimated standard errors. The standard error on the income elasticity in the demand for money equation for the equation by equation approach is 51% larger than the minimum distance system approach standard error. The standard error on the effect of the real interest rate on investment for the equation by equation approach is 72% larger than the minimum distance system approach standard error.

Second, the system point estimates differ from the equation by equation point estimates by up to a (equation by equation) standard error in most cases. The income elasticity coefficient in the system estimation is almost one standard deviation larger than that in the equation by equation estimation. The effect of the real interest rate, negative and significantly different from zero in the equation by equation estimates, is in the minimum distance estimation more negative with a smaller standard error. Of course, with a single sample we cannot be sure that the reduction of the standard errors is due to greater efficiency, however the change in the standard errors is in the direction we expect given the likely high degree of simultaneity in the model and the effect that using this information would have on the standard errors.

The test for overidentifying restrictions here is equal to 7.59, numerically equivalent to the Wald test for these restrictions reported in King et al (1991) Table 3 Panel B. This corresponds to a p-value of 0.18 (the asymptotic distribution is c_5^2) so the over-identifying restrictions are not rejected.

5. CONCLUSION

The minimum distance methods give an extremely simple way to estimate and test quite a large range of models with restrictions on the cointegrating vectors. Restrictions on the cointegrating vectors may be linear or nonlinear. For linear restrictions, a closed form solution for the restricted estimates exists; for both cross equation and within equation restrictions. For nonlinear restrictions, search procedures are required but only over the dimension of the cointegrating parameters rather than the whole model.

These methods thus make available simple procedures for estimating restricted cointegrating vectors. An example of an application is in King et al (1991), where in estimating restricted cointegrating vectors in a macroeconomic model they had to undertake substantial rearrangement of the model and then estimated each equation individually. Such an approach, whilst yielding consistent estimates, provides inefficient estimates as the methods do not jointly use all of the restrictions implied by the economic model. For their problem in the minimum distance framework, a closed form solution for the restrictions and a joint test of the restrictions is trivial.

Thanks to S. Johansen, P. Schmidt, H. White, J. Wooldridge, seminar participants at MSU and UC. Berkeley, a referee and the editor for comments. Financial assistance from the National Science Foundation grant SBR-9720675 is gratefully acknowledged.

References:

Ahn, S.K and Reinsel, G.C. (1990), "Estimation for Partially Nonstationary Autoregressive Models", *Journal of the American Statistical Association*, 85, 813-823.

Amemiya, T. (1985) Advanced Econometrics, Harvard University Press: Cambridge.

Chan, N.H. and Wei, C.Z. (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes", *Annals of Statistics*, 16, 367-401.

Engle, R.F. and Granger, C.W. (1987), "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 55, 251-76.

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

Johansen, S. (1995), Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models, Oxford University Press, Oxford.

King, R.G, Plosser, C.I., Stock, J.H. and Watson, M.W. (1991), “Stochastic Trends and Economic Fluctuations”, *American Economic Review*, 81, 819-40.

Phillips P.C.B. (1991), "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, 283-306.

Phillips, P.C.B. and Hansen, B.E. (1990), "Statistical Inference in Instrumental Variables Regression with I(1) processes", *Review of Economic Studies*, 57, 99-125.

Phillips, P.C.B. and Loretan, M (1991), "Estimating Long-Run Economic Equilibria", *Review of Economic Studies*, 59, 407-36.

Rothenberg, T.J. (1973), Efficient Estimation with A Priori Information, Cowles Foundation for Research in Economics Monograph 23.

Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regressions", *Econometric Theory*, 7, 1-21.

Saikkonen, P. (1992), "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation", *Econometric Theory*, 8, 1-27.

Sims, C.A., Stock, J.H. and Watson, M.W. (1990), "Inference in Linear Time Series Models with Some Unit Roots", *Econometrica*, 58, 113-44.

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

Warne, A. (1997), “Inference in Cointegrated VAR Systems”, *Review of Economics and Statistics*, LXXIX, 508-511.

Watson, M.W. (1994), "Vector Autoregressions and Cointegration", in McFadden and R.F.Engle eds. Handbook of Econometrics V4, North Holland, 2843-915.

White, H. (1984) Asymptotic Theory for Econometricians, Academic Press: San Diego.

White, H, (1994) Estimation, Inference and Specification Analysis, Cambridge University Press: Cambridge, UK.

Appendix A: Proofs

Theorem 1.

The proof of consistency is similar the usual proof except that the weight matrix is stochastic in the limit.

Define $d_n(\theta) = [vec(\hat{\Gamma}) - g(\theta)]'\hat{V}^{-1}[vec(\hat{\Gamma}) - g(\theta)]$. Rewrite this as

$$d_n(\theta) = [vec(\hat{\Gamma}) - g(\theta_0)]'V_n^{-1}[vec(\hat{\Gamma}) - g(\theta_0)] + 2[g(\theta_0) - g(\theta)]'V_n^{-1}[vec(\hat{\Gamma}) - g(\theta_0)]$$
$$+[vec(\hat{\Gamma}) - g(\theta)]'[\hat{V}^{-1} - V_n^{-1}][vec(\hat{\Gamma}) - g(\theta)] + [g(\theta_0) - g(\theta)]'V_n^{-1}[g(\theta_0) - g(\theta)]$$

First, we need to show that $d_n(\theta) \xrightarrow{p} d_0(\theta)$ where $d_0(\theta) = [g(\theta_0) - g(\theta)]'V_n^{-1}[g(\theta_0) - g(\theta)]$ for all θ. Each of the first three pieces converge in probability to zero regardless of the choice of θ. First, as V_n is positive definite with probability one then the first term converges to zero in probability by condition (a) of the theorem. Noting that $[g(\theta_0) - g(\theta)]$ is constant for all θ the second term also converges to zero by condition (a). Finally, as $[vec(\hat{\Gamma}) - g(\theta)]$ is $0_p(1)$ for all $\theta \neq \theta_0$ and converges to zero otherwise then by condition (b) the third term converges to zero. The remaining term is $d_0(\theta)$. Thus $d_n(\theta) \xrightarrow{p} d_0(\theta)$ for all θ. This implies that $\hat{\theta} \xrightarrow{p} \theta_0$ following arguments analogous to Amemiya (1985) Theorem 4.1.1.

Theorem 2.

This result follows directly through application of the Mean Value theorem and application of the continuous mapping theorem and functional central limit theorem.

Define $S(\tilde{\theta}) = \left.\frac{\partial d_n(\theta)}{\partial \theta}\right|_{\tilde{\theta}} = -2\left[\left.\frac{\partial g(\theta)}{\partial \theta'}\right|_{\tilde{\theta}}\right]'\hat{V}^{-1}[vec(\hat{\Gamma}) - g(\tilde{\theta})]$ for any $\tilde{\theta} \in \Theta$.

By the mean value theorem, where θ* lies between $\hat{\theta}$ and θ_0, we have

$$S(\hat{\theta}) = 0 = S(\theta_0) + \left[\left. \frac{\partial S(\theta)}{\partial \theta'} \right|_{\theta^*} \right] (\hat{\theta} - \theta_0) .$$

This can be rearranged to obtain

$$T(\hat{\theta} - \theta_0) = C^{-1} G(\theta^*)' \hat{V}^{-1} T[vec(\hat{\Gamma}) - g(\theta_0)]$$

where $C = \frac{1}{2} \left[\left. \frac{\partial S(\theta)}{\partial \theta'} \right|_{\theta^*} \right]$ and we define $G(\theta^*) = \left[\left. \frac{\partial g(\theta)}{\partial \theta'} \right|_{\theta^*} \right]$. The limit results for $T[vec(\hat{\Gamma}) - g(\theta_0)]$ and $\hat{V}^{-1}$ come from conditions (a) and (b) of the theorem.

For the stated result we require that $C \Rightarrow (G(\theta_0)' V_n^{-1} G(\theta_0))$. By differentiation, we have that $C = \frac{\partial}{\partial \theta'} \left\{ - \left[\left. \frac{\partial g(\theta)}{\partial \theta'} \right|_{\theta^*} \right]' \hat{V}^{-1} [vec(\hat{\Gamma}) - g(\theta^*)] \right\} = G(\theta^*)' \hat{V}^{-1} G(\theta^*) + R_2$ where R_2 has typical element $- \left[\left. \frac{\partial^2 g_j(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta^*} \right] \hat{V}^{-1} [vec(\hat{\Gamma}) - g(\theta^*)]$. First, note that R_2 is $o_p(1)$ as from the consistency of $\hat{\theta}$, that θ* is between $\hat{\theta}$ and θ_0, and assumption c) then $\left[\left. \frac{\partial^2 g_j(\theta)}{\partial \theta_i \partial \theta_j} \right|_{\theta^*} \right]$ is $O_p(1)$ in the limit, $\hat{V}^{-1}$ is $0_p(1)$ in the limit and $[vec(\hat{\Gamma}) - g(\theta^*)]$ is $o_p(1)$ by assumption a).

Second, note that from White (1984) that from the consistency of $\hat{\theta}$, that θ* is between $\hat{\theta}$ and θ_0, and continuity of the derivative that plim $G(\theta^*) = G(\theta_0)$. Thus from condition b) and the continuous mapping theorem we have $G(\theta^*)' \hat{V}^{-1} G(\theta^*) \Rightarrow G(\theta_0)' V_n^{-1} G(\theta_0)$ as required.

Putting these results together using the FCLT and continuous mapping theorem we have that

$$T(\hat{\theta} - \theta_0) \Rightarrow (G(\theta_0)' V_n^{-1} G(\theta_0))^{-1} G(\theta_0)' V_n^{-1/2} vec\left\{ (\int dW_{2.1} W_1^d)(\int W_1^d W_1^d{}')^{-1/2} \right\}$$

This implies that $(G(\hat{q})'V_n^{-1}G(\hat{q}))^{1/2}T(\hat{q}-q_0)$ has an asymptotic mixed normal distribution conditional on y_{1t} with mean zero and variance covariance matrix I_q.

Theorem 3.

Write $T(vec(\hat{\Gamma})-g(\hat{q}))=T(vec(\hat{\Gamma})-g(q_0))-T(g(\hat{q})-g(q_0))$. From the Mean Value Theorem $g(\hat{q})=g(q_0)+G(q^*)(\hat{q}-q_0)$ so with the results in Theorem 2 we have

$$\begin{aligned}T(vec(\hat{\Gamma})-g(\hat{q})) &= \left\{I_{n1n2}-(G(q^*)C^{-1}G(q^*)'\hat{V}^{-1})\right\}T(vec(\hat{\Gamma})-g(q_0))\\ &= \hat{V}^{1/2}P\hat{V}^{-1/2}T(vec(\hat{\Gamma})-g(q_0))\end{aligned}$$

where $P=I_{n1n2}-\hat{V}^{-1/2}G(q^*)C^{-1}G(q^*)'\hat{V}^{-1/2}{}'$.

We now have

$$\begin{aligned}J_T &= \left\{T(vec(\hat{\Gamma})-g(\hat{q}))\right\}'\hat{V}^{-1}\left\{T(vec(\hat{\Gamma})-g(\hat{q}))\right\}\\ &= \left\{T(vec(\hat{\Gamma})-g(q_0))\right\}'\hat{V}^{-1/2}{}'\left\{P'\hat{V}^{1/2}{}'\hat{V}^{-1}\hat{V}^{1/2}P\right\}\hat{V}^{-1/2}\left\{T(vec(\hat{\Gamma})-g(q_0))\right\}\\ &= \left\{\hat{V}^{-1/2}T(vec(\hat{\Gamma})-g(q_0))\right\}'P\left\{\hat{V}^{-1/2}T(vec(\hat{\Gamma})-g(q_0))\right\}\end{aligned}$$

as P is idempotent (it is a projection matrix). By the results of Theorem 1 and Theorem 2 we have that $P\Rightarrow I_{n1n2}-V_n^{-1/2}G(q_0)(G(q_0)'V_n^{-1}G(q_0))^{-1}G(q_0)'V_n^{-1/2}{}'$ and by condition (a) $\left\{\hat{V}^{-1/2}T(vec(\hat{\Gamma})-g(q_0))\right\}$ has an asymptotic mixed standard normal distribution. By the results of Chan and Wei (1988) the two pieces are asymptotically independent. Thus the statistic is of the form z'Kz where z is normally distributed and asymptotically independent of K (the limit of P). Further, the limit for P is of rank n_1n_2-q and has unit eigen values so by lemma 8.2 of White (1994) this quadratic has a χ^2 limit distribution with n_1n_2-q degrees of freedom.

Appendix B: Constructing $\hat{\Gamma}$ and $\hat{V}$ for each of the three methods examined.

Saikkonen (1992).

Run the VAR in equation (1) and construct $\hat{\Sigma} = \frac{1}{T}\sum \hat{e}_t * \hat{e}_t *'$ and $\hat{S} = \hat{\Psi}'\hat{\Sigma}^{-1}\hat{\Psi} = \begin{bmatrix} \hat{S}_{11} & \hat{S}_{12} \\ \hat{S}_{21} & \hat{S}_{22} \end{bmatrix}$.

Then the unrestricted estimates of the cointegrating vector are $\hat{\Gamma} = -\hat{S}_{22}^{-1}\hat{S}_{21}$, $\hat{\Omega}_{2.1} = \hat{S}_{22}^{-1}$, and $\hat{V}_1$ is the upper $n_1 x n_1$ block of $\Lambda^{-1}\sum x_t * x_t *'\Lambda^{-1}$ where $x_t* = [y_{1t-1}' \; z_t']'$ and Λ is a square scaling matrix with upper left hand block TI_{n1} , second block with zeros in off diagonals, (1,1) element $\sqrt{T}$ and (2,2) element $T^{3/2}$ (if time is included as a regressor) with all other elements zero. The variance covariance matrix is then $\hat{V} = \hat{V}_1 \otimes \hat{\Omega}_{2.1}$

Johansen (1988).

Solve the equation $|\hat{l}S_{kk} - S_{k0}S_{00}^{-1}S_{0k}| = 0$, $\hat{l}$ being a diagonal matrix of eigen values $\hat{l}_i$, $S_{ij}=\Sigma R_{it}R_{jt}$, i,j=0,k where R_{0t} (resp. R_{kt}) are the residuals of a regression of Δy_t (y_{t-1}) on lags of Δy_t. Normalize the $n_2 x n$ eigen vectors ($\hat{b}$) associated with the smallest n_2 eigen values by premultiplying by $\hat{b}_2^{-1}$ where we have partitioned β into $[\beta_1 \; \beta_2]$ where the partition is after the n_1th column. Now $\hat{b}_2^{-1}\hat{b} = [\hat{b}_2^{-1}\hat{b}_1, I_{n2}] = [-\hat{\Gamma}, I_{n2}] = \hat{\Gamma}*$. A natural estimator for $\hat{\Omega}_{2.1} = (\hat{a}'\hat{\Sigma}^{-1}\hat{a})^{-1}$ where $\hat{a}$ is the nxn_2 matrix of error correction coefficients. The MLE for these estimates depends on the normalization used and is $\hat{a} = -(\hat{\Gamma}*S_{kk}\hat{\Gamma}*')^{-1}\hat{\Gamma}*S_{k0}$. We can use the upper $n_1 x n_1$ block of TS_{kk} if y_{1t} is ordered first as an estimate of $\hat{V}_1$.

Phillips and Loretan (1991), Saikkonen (1991), Stock and Watson (1993),

This method involves the running of the set of regressions $y_{2t} = m_2 + \Gamma y_{1t} + d(L)\Delta y_{1t} + h_t$ where d(L) includes leads and lags (two sided polynomial lag operator). The unrestricted estimates for $\hat{\Gamma}$ are simply the equation by equation OLS estimates of this model. An estimate for $\hat{\Omega}_{2.1}$ can be obtained by any method of estimating the spectral density at frequency zero of $\hat{h}_t$, the OLS residuals from the above equation. The estimator for $\hat{V}_1$ comes from the $n_1 x n_1$ block of the denominator of the OLS estimates of the coefficient vector corresponding to the y_{1t} coefficients in the DOLS regression above.