2-4-1998

Submodel Estimation of a Structural Vector Error Correction Model

William C. Horrace
University of Arizona

Economics Letters 59 (1998) 23–29

economics letters

Submodel estimation of a structural vector error correction model under cointegration

William C. Horrace*

Department of Economics, University of Arizona. PO Box 210108, Tucson, Arizona 85721-0108, USA

Received 29 May 1997; accepted 4 February 1998

Abstract

In this paper we derive the concentrated likelihood function of a mutually independent subsystem (submodel) of equations from a p-dimensional vector error component model under cointegration. The structural estimates of the subsystem parameters are identified by exclusion restrictions. The maximum likelihood estimates may be useful for counterfactual policy analysis. © 1998 Elsevier Science S.A.

Keywords: Vector error correction model; Cointegration; Limited information maximum likelihood; Counterfactual policy analysis

JEL classification: C32

1. Introduction

This paper is concerned with structural estimation of a subsystem (submodel) of equations from a vector error components model under cointegration. Our purpose is twofold. First, we wish to transform the entire system to ensure mutual independence of the submodel of interest and the remaining equations. Second, we wish to estimate the submodel of interest using subset Limited Information Maximum Likelihood (LIML) techniques incorporating restrictions on the submodel to identify its structural parameters. The resulting likelihood function is the product of two least variance ratios in the style of Johansen (Johansen, 1988, 1991) and can, in fact, be thought of as a generalization of the Johansen result.

This submodel analysis has implications for ‘‘counterfactual policy analysis’’ or ‘‘VAR transplantation’’, in which one is interested in estimating the behavior of an economy (a submodel) under alternative sets of policy rules. See for example McCallum (McCallum, 1988, 1990, 1993), and Judd and Motley (Judd and Motley, 1992, 1993). Rasche (1995) points out problems in these types of analyses and in a subsequent unpublished note identifies LIML as a potential solution. Also, submodel analysis provides computational economy for any system of equations for which full information maximum likelihood is cumbersome. See Dhrymes (Dhrymes, 1970, p. 329). However, this parsimony comes at a cost; it is well known that LIML estimates are informationally inefficient.

*Tel.: (520) 621 6230; fax: (520) 621 8450; e-mail: whorrace@u.arizona.edu

PII S0165-1765(98)00033-0

The paper is organized as follows. Section 2 details the model and derives the likelihood function. Section 3 discusses numerical estimation and a two-step estimation technique. Section 4 concludes.

2. Estimation

2.1. Specification

Consider the following error-correction representation of a structural VAR model in p dimensions.

$$\Delta Y_t A + \sum_{i=1}^{q} \Delta Y_{t-i}\Gamma_i + Y_{t-1}\beta\alpha' + X_t\Theta = \epsilon_t\,, \quad t=1,\ldots,T. \tag{1}$$

Assumptions:

A1. The ϵ_t $(t=1,\ldots,T)$ are independent p-dimensional Gaussian variables with mean zero and positive definite variance matrix Σ.
A2. The Y_t $(1\times p)$ are integrated of order 1 with β, the matrix of cointegrating vectors, and α, the matrix of error correction coefficients, being $p\times r$ matrices of rank r.
A3. The first q data points $Y_0,\ldots,Y_{q-1}$ are fixed.
A4. The parameters A, $\Gamma_i,\ldots,\Gamma_q$ and Σ (all $p\times p$ matrices), vary without restriction.
A5. X_t $(1\times s)$ are stationary s-dimensional instruments appearing only in the last $n<p$ equations. This is, $\Theta=[0, C]$; $\Theta(s\times p)$, C $(s\times n)$, where C varies without restriction.

Assumptions *A1–A4* are standard. Assumption *A5* is necessary for identification of the structural parameters. The first $m=p-n$ equations, in which the X_t do not appear, are the submodel of interest, while the last n equations, in which the X_t do appear, are not. What is typically done is to estimate the reduced form of Eq. (1) and discard the last n equations. Here, we transform the system to ensure mutual independence of the two subsystems then use limited information maximum likelihood techniques to estimate the first m equations in structural form. This structural estimate is identified by the restrictions on the parameter Θ.

2.2. Transformation

Transformation is performed per the subset LIML technique described in Dhrymes (Dhrymes, 1970, Section 7.3). Subset LIML was first considered by Rubin (1948) and Hood and Koopmans (1953). Partition Σ such that

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}; \quad \Sigma_{11}(m\times m) \text{ positive definite, } \Sigma_{12}(m\times n),\ \Sigma_{21}(n\times m),\ \Sigma_{22}(n\times n).$$

Define the transformation matrix $H(p\times p)$ such that

$$H=\begin{bmatrix} I_m & -\Sigma_{11}^{-1}\Sigma_{12}H_{22} \\ 0 & H_{22} \end{bmatrix},$$

where H_{22} is chosen to satisfy

$$H_{22}'(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{21})H_{22}=I_n.$$

Post multiplying Eq. (1) by the matrix H leaves the first m equations of the system undisturbed and yields independently distributed error terms $(\epsilon_t H)$, $t=1,\ldots,T$, a multivariate Gaussian random variable with mean 0 and block-diagonal variance matrix of the form

$$V(\epsilon_t H)=\begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Omega \end{bmatrix} \text{ positive definite,}$$

where the specific form of Ω is unimportant to the analysis since it corresponds to the n equations in which we are not interested. This block-diagonalization of the variance matrix of the transformed errors ensures the stochastic independence of the two subsystems without disturbing the variance, Σ_{11}, of the equations of interest.

2.3. The likelihood function

Let $Y_{*t}=[\Delta Y_{t-1},\ldots,\Delta Y_{t-q}]$, $\Gamma_*'=[\Gamma_i',\ldots,\Gamma_q']$. Collecting t vertically and post multiplying by H, Eq. (1) becomes,

$$\Delta YAH+Y_*\Gamma_*H+Y_{-1}\beta\alpha'H+X\Theta H=\epsilon H. \tag{2}$$

We are interested in estimating the coefficients of the first $m=p-n$ equations. To this end we partition the remaining coefficient matrices as follows,

$$A=[A_1 \quad A_2];\quad A_1(p\times m), A_2(p\times n),$$

$$\Gamma_*=[\Gamma_{*1} \quad \Gamma_{*2}];\quad \Gamma_{*1}(q\times m), \Gamma_{*2}(q\times n)$$

$$\alpha'=[\alpha_1' \quad \alpha_2'];\quad \alpha_1'(r\times m), \alpha_2'(r\times n)$$

Given the transformed and partitioned system, we would like to write down the likelihood function and concentrate out the matrices: A_2, Γ_*, α, Θ and Σ, leaving only β and A_1, the structural parameter of interest. To this end we introduce the following theorem.

Theorem 1. *Let* $W=[Y_* \quad X]$, $F=[Y_* \quad X \quad Y_{-1}\beta]$ *and* $J=[Y_* \quad Y_{-1}\beta]$. *Then the concentrated likelihood function of Eq. (2) is given by*:

$$L^{\frac{-2}{T}}(A_1,\beta)$$
$$=\left|S_{\Delta Y\Delta Y}^{W}\right|\frac{\left|A_1'\{S_{\Delta Y\Delta Y}^{F}+S_{\Delta YX}^{J}[S_{XX}^{J}]^{-1}S_{X\Delta Y}^{J}\}A_1\right|}{\left|A_1'S_{\Delta Y\Delta Y}^{F}A_1\right|}\,\frac{\left|\beta'\{S_{Y_{-1}Y_{-1}}^{W}-S_{Y_{-1}\Delta Y}^{W}[S_{\Delta Y\Delta Y}^{W}]^{-1}S_{\Delta YY_{-1}}^{W}\}\beta\right|}{\left|\beta'S_{Y_{-1}Y_{-1}}^{W}\beta\right|},$$

where the S_{ij}^{r} *are* $(p\times p)$ *residual product matrices*

$$S_{ij}^{r}=\tfrac{1}{T}i'[I_T-r(r'r)^{-1}r']j,\quad r=W,J,F,\quad i,j=\Delta Y,Y_{-1},X.$$

Proof. See Appendix A.

The inverse-likelihood function of Theorem 1, provides a convenient formulation for a numerical minimization algorithm. For an iterative search an excellent candidate for a starting value for β would be the Johansen (1991) estimate. This inverse-likelihood function is a generalization of the Johansen (1991) result which produces super consistent estimates of the model's parameters. The last ratio on the right hand side is exactly the Johansen result except for the inclusion of the stationary instruments, X. In fact, when X is excluded from the model, the likelihood reduces to the Johansen result. While the asymptotic properties of the likelihood estimates are at this point unknown, it seems reasonable to suspect that the addition of stationary instruments to the system will not adversely effect the estimates, and they should remain super consistent. We now discuss an alternative to numerical optimization of the inverse-likelihood.

3. A two step estimation procedure

Of course numerical minimization can be computationally cumbersome, and there always exists the possibility that the procedure will return a local extremum and not a global one. Unfortunately, as derived this likelihood function is not amenable to the usual partial canonical correlation analysis of Anderson (1951) and Tso (1981), because the matrices $S_{\Delta Y\Delta Y}^{F}$ and $S_{\Delta YX}^{J}[S_{XX}^{J}]^{-1}S_{X\Delta Y}^{J}$ are not independent of the parameters of the model. Specifically they are functions of β through F and J. One alternative to numerical optimization or canonical analysis is a two-step procedure that exploits the least variance ratio form of the inverse-likelihood.

In the first step, the space spanned by β is estimated using the reduced form procedure of Johansen (1991). This involves performing a least-squares regression of Y_{-1} on Y_* and X, and a least-squares regression of ΔY on Y_* and X. These yield residual product matrices, $S_{\Delta Y\Delta Y}^{W}$, $S_{Y_{-1}Y_{-1}}^{W}$ and $S_{Y_{-1}\Delta Y}^{W}$. Then $\hat{\beta}$ is constructed from the r eigenvectors associated with the r smallest eigenvalues of $S_{Y_{-1}Y_{-1}}^{W}-S_{Y_{-1}\Delta Y}^{W}[S_{\Delta Y\Delta Y}^{W}]^{-1}S_{\Delta YY_{-1}}^{W}$ in the metric of $S_{Y_{-1}Y_{-1}}^{W}$. Using the Johansen technique in the first stage seems like a reasonable approach given that the inverse-likelihood actually contains the least variance ratio derived by Johansen.

In the second step $\hat{\beta}$ is used in the three regressions of ΔY on Y_*, X and $Y_{-1}\hat{\beta}$; ΔY on Y_* and $Y_{-1}\hat{\beta}$; and X on Y_* and $Y_{-1}\hat{\beta}$. These yield residual product matrices $S_{\Delta Y\Delta Y}^{F}$, $S_{\Delta YX}^{J}$ and S_{XX}^{J}. Then $\hat{A}_1$ is similarly constructed from the m eigenvectors associated with the m smallest eigenvalues of $S_{\Delta Y\Delta Y}^{F}+S_{\Delta YX}^{J}[S_{XX}^{J}]^{-1}S_{X\Delta Y}^{J}$ in the metric of $S_{\Delta Y\Delta Y}^{F}$. So the two-step procedure amounts to two eigenvalue problems, both in the spirit of partial canonical correlations. The notion that this two-step procedure produces a truly minimized inverse likelihood may seem dubious, however the Johansen technique has gained in popularity, and there now exists software to easily perform the first stage of the estimation procedure. Therefore, the two-step approach may provide much in the way of computational economy. Additionally, this two-step procedure provides a useful algorithm for the aforementioned numerical estimation procedure.

4. Conclusions

In this paper we have derived the inverse-likelihood function of a submodel of cointegrated equations. Perhaps remarkably, the resulting function is a generalization of the Johansen (1991) reduced form inverse-likelihood function for the cointegration parameter β. The innovation here is that the likelihood includes a structural parameter of the equations of interest, A_1, which is identified by exclusion restrictions on Θ. This likelihood may be useful for counterfactual policy analysis and for its computational parsimony.

As derived, the likelihood function does not lend itself to canonical correlation analysis because the least variance ratio in the structural parameter, A_1, is a function of β. The function is, however, amenable to numerical optimization or a two-step estimation procedure outlined herein. Whether restrictions can be found that permit the use of canonical techniques remains to be seen, however initial research on the rank of the residual product matrices, S^r_{ij}, indicates that such a restriction may exist.

Acknowledgements

I would like to thank Peter Schmidt and Robert Rasche for comments on an earlier version of the paper. Portions of this research were supported by the National Science Foundation and the Ford Foundation.

Appendix A

Proof of Theorem 1. The log-likelihood function for Eq. (2) is:

$$\begin{aligned}\ln L(\beta, A, \Gamma_*, \alpha, \Theta, \Sigma) = & -(pT/2)\ln(2\pi) - (T/2)\ln|H'\Sigma H| + T\ln|AH| \\ & - (1/2)\mathrm{tr}\{(\Delta YAH + Y_*\Gamma_*H + Y_{-1}\beta\alpha'H + X\Theta H)(H'\Sigma H)^{-1}(\Delta YAH \\ & + Y_*\Gamma_*H + Y_{-1}\beta\alpha'H + X\Theta H)'\},\end{aligned}$$

Concentration of A_2, Γ_{*2}, α_2, Θ, Σ_{12}, Σ_{21}, and Σ_{22} proceeds per Dhrymes (1970, p. 332) and will not be detailed here. Define $N=[\Delta Y \quad Y_* \quad X \quad Y_{-1}\beta]$ and $M'=[A_1' \quad \Gamma_{*1}' \quad 0' \quad \alpha_1]$. Then the resulting log-likelihood function is

$$\begin{aligned}\ln L(\beta, A_1, \Gamma_{*1}, \alpha_1, \Sigma_{11}) = & -(pT/2)[\ln(2\pi) + 1] + mT/2 - (T/2)\ln|S^F_{\Delta Y\Delta Y}| - (T/2)\ln|\Sigma_{11}| \\ & + (T/2)\ln|A_1'S^F_{\Delta Y\Delta Y}A_1| - (T/2)tr(\Sigma_{11}^{-1}M'(N'N/T)M).\end{aligned} \tag{A1}$$

All that remains is concentration of Γ_{*1}, α_1, Σ_{11}. Let $\delta'=[\Gamma_{*1}' \quad \alpha_1]$, then

$$M'(N'N/T)M = A_1'(\Delta Y'\Delta Y/T)A_1 + A_1'(\Delta Y'J/T)\delta + \delta'(J'\Delta Y/T)A_1 + \delta'(J'J/T)\delta.$$

Substituting this into Eq. (A1) and taking derivatives:

$$\frac{\partial \ln L}{\partial \delta} = -(T/2)[2(J'\Delta Y/T)A_1 + 2(J'J/T)\delta]\Sigma_{11}^{-1} = 0,$$

$$\frac{\partial \ln L}{\partial \Sigma_{11}} = -(T/2)[-\Sigma_{11}^{-1} + \Sigma_{11}^{-1}A_1' S_{\Delta Y \Delta Y}^{J} A_1 \Sigma_{11}^{-1}] = 0,$$

implying $\delta = -(J'J/T)^{-1}(J'\Delta Y/T)A_1$ and $\Sigma_{11} = A_1' S_{\Delta Y\Delta Y}^{J} A_1$. Substituting δ and Σ_{11} into the log-likelihood function gives

$$\begin{aligned}\ln L(A_1, \beta) = & -(pT/2)[\ln(2\pi) + 1] + mT/2 - (T/2)\ln|S_{\Delta Y\Delta Y}^{F}| - (T/2)\ln|A_1' S_{\Delta Y\Delta Y}^{J} A_1| \\ & + (T/2)\ln|A_1' S_{\Delta Y\Delta Y}^{F} A_1|.\end{aligned}$$

Suppressing the constant terms and taking the anti-log, the likelihood function becomes:

$$L^{\frac{-2}{T}}(A_1, \beta) = \frac{|A_1' S_{\Delta Y\Delta Y}^{J} A_1|}{|A_1' S_{\Delta Y\Delta Y}^{F} A_1|}\, |S_{\Delta Y\Delta Y}^{F}|. \tag{A2}$$

Write F as $F = [W \quad Y_{-1}\beta]$, then using the rules of a partitioned inverse,

$$\begin{aligned}F(F'F/T)^{-1}F' = T^{-1}\{ & W(W'W)^{-1}W' \\ & + [I_T - W(W'W)^{-1}W']Y_{-1}\beta[\beta' S_{Y_{-1}Y_{-1}}^{W}\beta]\beta' Y_{-1}'[I_T - W(W'W)^{-1}W']\}.\end{aligned}$$

Then $S_{\Delta Y\Delta Y}^{F} = \frac{1}{T}\Delta Y'[I_T - F'(F'F)^{-1}F]\Delta Y = S_{\Delta Y\Delta Y}^{W} - S_{Y_{-1}Y_{-1}}^{W}\beta[\beta' S_{Y_{-1}Y_{-1}}^{W}\beta]^{-1}\beta' S_{Y_{-1}Y_{-1}}^{W}$. Using the results for the determinant of a partitioned matrix,

$$|S_{\Delta Y\Delta Y}^{F}| = |S_{\Delta Y\Delta Y}^{W}| \frac{\left|\beta'\{S_{Y_{-1}Y_{-1}}^{W} - S_{Y_{-1}\Delta Y}^{W}[S_{\Delta Y\Delta Y}^{W}]^{-1}S_{\Delta Y Y_{-1}}^{W}\}\beta\right|}{\left|\beta' S_{Y_{-1}Y_{-1}}^{W}\beta\right|}. \tag{A3}$$

It is also easily shown that $S_{\Delta Y\Delta Y}^{J} = S_{\Delta Y\Delta Y}^{F} + S_{\Delta YX}^{J}[S_{XX}^{J}]^{-1}S_{X\Delta Y}^{J}$. Substituting this result and Eq. (A3) into Eq. (A2), the main result follows. *QED*

References

Anderson, T.W., 1951. Estimating linear restrictions on regression coefficients for multivariate normal distributions. Annals of Mathematical Statistics 22, 327–351.

Dhrymes, P.J., 1970. Econometrics, statistical foundations and applications. Springer-Verlag, New York.

Hood, W.C., Koopmans, T.C. (Eds.), 1953. Estimation of linear relationships. In: Studies in Econometric Methods. Cowles Foundation for Research in Economics. Monograph No. 14. Wiley, New York.

Johansen, S., 1988. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231–254.

Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.

Judd, J.P., Motley, B., 1992. Controlling inflation with an interest rate instrument. Federal Reserve Bank of San Francisco Economic Review 3, 3–22.

Judd, J.P., Motley, B., 1993. Using a nominal GDP rule to guide discretionary monetary policy. Federal Reserve Bank of San Francisco Economic Review 3, 3–11.

McCallum, B.T., 1988. Robustness properties of a rule for monetary policy. Carnegie–Rochester Conference Series on Public Policy 29, 173–204.

McCallum, B.T., 1990. Targets, instruments and indicators of monetary policy. In: Haraf, W.S., Cagan, P. (Eds.), Monetary Policy for a Changing Financial Environment. AEI Press, Washington, D.C.

McCallum, B.T., 1993. Specification and analysis of a monetary policy rule for Japan. Bank of Japan Monetary and Economic Studies 11, 1–45.

Rasche, R.H., 1995. Pitfalls in counterfactual analyses of policy rules. Open Economies Review 6, 199–202.

Rubin, H., 1948. Systems of linear stochastic equations. Unpublished Ph.D. dissertation. University of Chicago, Chicago.

Tso, M.K.-S., 1981. Reduced-rank regression and canonical analysis. Journal of the Royal Statistical Society, Series B 42, 183–189.