Discussion Paper 19

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

November 1989

INSTRUMENTAL VARIABLE ESTIMATORS FOR STATE SPACE MODELS

Masanao Aoki
Institute for Empirical Macroeconomics
and University of California, Los Angeles

ABSTRACT

The state vector in the innovation representation is asymptotically the most efficient instrumental variable estimator for the observation matrix C. The paper compares small sample properties of IV estimators for C, the dynamic matrix A and other matrices with the system theoretic estimators described in Aoki (1987) by a small scale Monte Carlo simulations. The IV estimators appear to be about the same as the system theoretic ones as far as their small sample properties are concerned. The covariance matrix of the state vector calculated from the IV point of view are also compared with the solutions of the Riccati equations. The simulation results show that they have quite similar sample means and standard deviations. This method of calculating the state vector covariance matrices may be computationally faster than solving the Riccati equation by the Schur decomposition algorithm.

Keywords: Instrumental Variable Estimator, State Space, Innovation, Monte Carlo Simulation, Small Sample Properties

This material is based on work supported by the National Science Foundation under Grant No. SES-8722451. The Government has certain rights to this material.

Introduction

Earlier, Havenner and Aoki (1988) provided an instrumental variable estimator interpretation for the state space model parameter estimation algorithm described in Aoki (1987), which was derived by extending an algorithm of the deterministic and stochastic realization theory to state space innovation models. See Aoki (1987) for references on the realization theory. More specifically, they pointed out that the estimator of the matrix C in the observation equation of the innovation model

$$y_t = Cz_t + e_t,$$

is identical with the instrumental variable (IV) estimator in which the stacked data vector $y_{t-1}^- = (y_{t-1}', y_{t-2}', \ldots, y_{t-k}')'$ for some $k > 0$ is used as the instrument vector. They also suggested that the state vector z_t is asymptotically more efficient than y_{t-1}^-. To construct a (balanced) innovation representation of state space models, the matrices A and B in the dynamic equation, called the state transition equation, $z_{t+1} = Az_t + Be_t$, and the matrix C in the observation equation, $y_t = Cz_t + e_t$, must be estimated together with the noise covariance matrix Δ (and the covariance matrix of the state vector Π). This paper discusses IV estimators of these system matrices in the innovative state space model. In the Aoki algorithm the matrices A, C and $M = E(z_{t+1}y_t')$ $= A\Pi C' + B\Delta$ are directly estimated from the singular value decomposition of the covariance matrix between the stacked future and stacked past observations. Then the matrices B, Δ, and Π are calculated based on the minimal solution of a certain Riccati equation. In the IV approach, we need not solve the Riccati equation. However, we lose the nestedness properties[1] of the estimates of the matrices A, C, and M in the IV approach.

An asymptotically most efficient estimator for the matrix C is obtained by using the state vector z_t as the instruments which is a certain linear combination of y^-_{t-1}. This can be shown as in Sargan (1988, p. 45). This paper uses this and other instrumental variable vectors to estimate the system matrices in the balanced representation of the innovation models, and improve them to have the same asymptotic efficiency as their maximum likelihood estimates.

We also establish that appropriate choices of the weight matrices in the generalized method of movement estimator produce the system estimators and the IV estimators. Since the best weight matrix corresponding to the maximization of the concentrated log likelihood function is complicated, approximating the best weight matrix with computationally easier ones are of interest. Other choices of weight matrices are also suggested.

In the econometric approach for modeling time series the state vector is not used as instruments because the components of the state vectors are not generally available to be used as instruments. One of the contributions of the system theory in modeling time series makes them available by establishing the fact that the vectors are related by

$$z_t = Sy^-_{t-1}$$

where the matrix S is directly estimated. This matrix is equal to

$$S = \Omega R_-^{-1}$$

with

$$\text{cov } y_{t-1}^- = R_-$$

and the matrix S is known since Ω is calculated from the singular value decomposition $U\Sigma V'$ of the Hankel matrix which is the cross covariance matrix between the stacked future data $y_t^+ = (y_t', y_{t+1}', y_{t+2}', \ldots, y_{t+J}')'$ for some $J > 0$ and y_{t-1}^-. The Hankel matrix is factored in two ways; $\Theta\Omega$ and $U\Sigma V'$, where Θ is the matrix which related y_t^+ to z_t, i.e., the regression coefficient matrix when y_t^+ is regressed on z_t. See Appendix 1, and (6) below. In the the balanced model, the first half of the singular value decomposition of the Hankel matrix $U\Sigma^{1/2}$ is taken to be the matrix Θ and the second half $S^{1/2}V'$ as Ω. See Aoki (1987) for example. Since this fact is basic and is easy to demonstrate, we collect some facts regarding it in Appendix 1. This estimator of Θ is later shown to be the same as the IV estimate with z_t as instrument.

IV Estimates of System Matrices

Matrix C

The sample version of the state vector is

$$(1) \qquad z_t = \hat{\Omega}\hat{R}_-^{-1} y_{t-1}^-$$

where "^" denotes the sampled value,

$$\hat{R}_- = \text{cov } y_{t-1}^- = T^{-1}\Sigma y_{t-1}^- y_{t-1}^{-'}$$

and where $\hat{\Omega}$ is the second half of the singular value decomposition of the sample Hankel matrix

$$\hat{H} = T^{-1}\sum y_t^+ y_{t-1}^{-\prime} = \hat{U}\ \hat{\Sigma}\ \hat{V}'$$

i.e.,

$$\hat{\Omega} = \hat{\Sigma}^{\frac{1}{2}}\ \hat{V}'. \tag{2}$$

We immediately obtain an estimate for the state vector covariance matrix as

$$\begin{aligned} \hat{\Pi} &= T^{-1}\Sigma z_t z_t' \\ &= \hat{\Omega}\hat{R}_-^{-1}\hat{\Omega}'. \end{aligned} \tag{3}$$

The IV estimator of matrix C, using z_t as the instrumental variable vector is obtained from the observation equation

$$y_t = Cz_t + e_t$$

as

(4)[2] $$\hat{C} = (T^{-1}\sum y_t z_t')\hat{\Pi}^{-1}$$

where

$$\begin{aligned} T^{-1}\sum y_t z_t' &= [\hat{\Lambda}_1 \hat{\Lambda}_2 \ldots]\hat{R}_-^{-1}\hat{\Omega}' \\ &= \hat{H}_{1.}\hat{R}_-^{-1}\hat{\Omega}' \end{aligned}$$

where $\hat{H}_{1.}$ denotes the first submatrix row (pxkp) of the Hankel matrix $\hat{H}$.

Substitute y_t out from (4) to see

$$\delta C\hat{\Pi} = T^{-1}\Sigma e_t z_t'$$

where δC is the difference between the estimate and the true matrix C. Its vectorized version is

$$(\hat{\Pi}\otimes I)\mathrm{vec}\,\delta\hat{C} = T^{-1}\Sigma z_t \otimes e_t.$$

Since $T^{-1/2}\Sigma_t z_t \otimes e_t$ converges to a mean zero normal random vector with the covariance matrix $\Pi\otimes\Delta$, we have

$$T(\Pi\otimes I)\overline{(\mathrm{vec}\,\delta C)(\mathrm{vec}\,\delta C)'}(\Pi\otimes I)$$

$$\to \Pi\otimes\Delta,$$

i.e.,

$$T\overline{(\mathrm{vec}\,\delta C)(\mathrm{vec}\,\delta C)'} \to \Pi^{-1}\otimes\Delta$$

as T goes to infinity. Therefore, the expression $\hat{\Pi}^{-1}\otimes\hat{\Delta}$ is used to calculate an estimate of the standard error of $\hat{C}$.

From (4) $T^{-1}\Sigma y_t z_\tau'$ is equal to $\hat{C}\hat{\Pi}$. Hence, $\hat{\Delta} = T^{-1}\Sigma\hat{e}_t\hat{e}_t'$ where $\hat{e}_t = y_t - \hat{C}z_t$ is equal to

$$\hat{\Delta} = \hat{\Lambda}_0 - \hat{C}\hat{\Pi}\hat{C}',$$

i.e., the noise covariance matrix is estimated from the relation

$$\Lambda_0 = C\Pi C'+\Delta \tag{5}$$

as

$$\hat{\Delta} = \hat{\Lambda}_0 - \hat{C}\hat{\Pi}\hat{C}'.$$

<u>The Observability Matrix Θ</u>

The matrices CA^i, $i = 0, 1, \ldots$ are estimated by using the stacked observation equation. The IV estimator of CA^i, $i = 0, 1, \ldots$ can be read off from $\hat{O}$, the IV estimate of the observability matrix which appears in (6)

(6) $$y_t^+ = \Theta z_t + Ge_t^+.$$

The structure of the matrices Θ and G are described in Appendix 1. The IV estimator is defined from (6) as the solution of

(7) $$T^{-1}\sum y_t^+ z_t' = \hat{\Theta}\hat{\Pi}$$

where the left-hand side is $T^{-1}\sum y_t^+ y_{t-1}^{-\prime} \hat{R}_t^{-1}\hat{\Omega}' = \hat{H}\hat{R}_t^{-1}\hat{\Omega}'$, e.g., $\hat{\Theta} = U\Sigma^{1/2}$.

From (6) and (7), the estimation error in Θ is given by

(8) $$\delta\Theta\hat{\Pi} = GT^{-1}\Sigma e_t^+ z_t',$$

from which we obtain

$$T\overline{(\mathrm{vec}\delta\Theta)(\mathrm{vec}\delta\Theta)'} \to \Pi^{-1}\otimes G(I\otimes\Delta)G'.$$

Matrix A

To estimate the dynamic matrix in the state transition equation, advance t by one unit in (6), replace z_{t+1} by $Az_t + Be_t$, and multiply the stacked future observation vector

$$\begin{aligned} y_{t+1}^+ &= \Theta z_{t+1} + Ge_{t+1}^+ \\ &= \Theta Az_t + \Theta Be_t + Ge_{t+1}^+ \end{aligned}$$

from the right by the transpose of the state vector to define $\hat{A}$ as the solution of $T^{-1}\Sigma y_{t+1}^+ z_t' = \hat{\Theta}\hat{A}\hat{\Pi}$.

The left hand side is equal to $T^{-1}\Sigma y_{t+1}^+ y_{t-1}^{-\prime}\hat{R}^{-1}\hat{\Omega}' = \hat{H}^A\hat{R}^{-1}\hat{\Omega}'$ where H^A is the Hankel matrix A shifted up by one submatrix row (p-rows) and the last p rows filled in the suitable $\hat{\Lambda}$'s.

The estimate is equal to

(9) $$\hat{A} = \hat{\Sigma}^{-1/2}\hat{U}'\hat{H}^{A}\hat{R}^{-1}\hat{\Omega}'\hat{\Pi}^{-1}.$$

This is to be compared with $\hat{\Sigma}^{-1/2}\hat{U}'\hat{H}^{A}\hat{V}\hat{\Sigma}^{-1/2}$ in Aoki (1987, p.121).

From the relation

$$(\hat{\Theta}\hat{A}-\Theta A)\hat{\Pi} = \Theta BT^{-1}\Sigma e_t z_t' + GT^{-1}\Sigma e_{t+1}^{+} z_t',$$

we see that the error matrix $\delta A = \hat{A} - A$ satisfies

$$\hat{\Theta}\delta A\hat{\Pi} + \delta\Theta\cdot A\hat{\Pi} = \Theta BT^{-1}\Sigma e_t z_t' + GT^{-1}\Sigma e_{t+1}^{+} z_t'.$$

The vectorized expression on the right of (8), when magnified by $T^{1/2}$, converges in distribution to a normal distribution with mean 0 and variance $Q = \Pi\otimes\left[\Theta B\Delta B'\Theta'+G(I\otimes\Delta)G'\right]$, from which follows

(10) $$T\left[\hat{\Pi}\otimes\Theta,(A\hat{\Pi})'\otimes I\right] \operatorname{cov}\begin{bmatrix}\mathrm{vec}\,\delta A\\ \mathrm{vec}\,\delta\Theta\end{bmatrix}\left[\Pi\otimes\Theta\ (A\Pi)'\otimes I\right]' \to Q.$$

A special case of (10) is $T[I\otimes C,\ \hat{A}'\otimes I] \operatorname{cov}\begin{pmatrix}\mathrm{vec}\,\delta A\\ \mathrm{vec}\,\delta C\end{pmatrix}[I\otimes C'\ \hat{A}\otimes I] \to \hat{\Pi}^{-1}\otimes[CB\Delta(CB)' + \Delta]$. To derive this directly, read off the 2nd p rows from (7) to note

$$(\widehat{CA})\hat{\Pi} = T^{-1}\Sigma y_{t+1}z_t'$$

and that

$$\delta(\hat{C}A)\hat{\Pi} = \hat{C}BT^{-1}\Sigma e_t z_t' + T^{-1}\Sigma e_{t+1}z_t'.$$

When $k = 1$, as a rough measure of the magnitude of cov(vecδA), we may use, on the assumption that δC is zero, and that C is invertible,

(10A) $$T\operatorname{cov}(\mathrm{vec}\,\delta A) \to \Pi^{-1}\otimes[B\Delta B'+C^{-1}\Delta(C^{-1})'].$$

This is a special case in which $\Theta^{+}G(I\otimes\Delta)G'\Theta'^{+}$, i.e., $\Sigma^{-1/2}U'G(I\otimes\Delta)G'U\Sigma^{-1/2}$ reduces to $C^{-1}\Delta(C^{-1})'$ since $H = \Lambda_1 = U\Sigma V'$ and the system estimaor of C is $U\Sigma^{1/2}$.

Vectorize the relation $\delta(CA) = \delta CA + C\delta A$ to obtain

$$(I\otimes\hat{C})\mathrm{vec}\delta A = \mathrm{vec}\delta(CA) - (\hat{A}'\otimes I)\mathrm{vec}\delta C$$

from which the desired expression follows.

In the applications of the estimated models, the dynamic matrix A does not appear by itself, even though its eigenvalues are of some intrinsic interest. The matrix A appears in the combinations as $CA^{i}B$, $i = 0, 1, \ldots$ since these are the impulse response matrices in conducting the dynamic multiplier analysis. The combination C(A-BC) appear in multiple-step-ahead forecast calculations;

$$y_{t+1|t} = \hat{E}(y_{t+1}|y_t^-)$$

$$= C(A-BC)z_t + CBy_t$$

$$y_{t+2|t} = C(A-BC)^2 z_t + C(A-BC)By_{t+1|t} + C(A-BC)By_t$$

for example. Therefore, the IV estimators for the matrices CA^{i}, which are derived above, $i = 1, 2, \ldots$ and CB are of more direct relevance. We later derive an IV estimator for $F = A - BC$.

<u>Matrix B</u>

From the relation $M = A\Pi C + B\Delta$, the matrix B is estimated by

(11) $\hat{B} = (\hat{M}-\hat{A}\hat{\Pi}\hat{C})\hat{\Delta}^{-1}$.

Multiplying the state transition equation from the left by the matrix C and repeating the above calculations with $\hat{e}_t = y_t - \hat{C}z_t$ as the instruments, we obtain

$$\widehat{(CB)}\hat{\Delta} = \hat{\Lambda}_1 - \hat{H}_{2.}\hat{R}_{-}^{-1}\hat{\Omega}'\hat{C}'$$

where $\hat{\Delta} = T^{-1}\Sigma\hat{e}_t\hat{e}_t'$, $\hat{e}_t = y_t - \hat{C}z_t$, by dropping the term $T^{-1}\sum e_{t+1}\hat{e}_t'$.

Analogous calculations yield the standard error expression for CB as

$$T\ \overline{\mathrm{vec}\delta(CB)\mathrm{vec}\delta(CB)'} \rightarrow n\Delta \otimes CB\Delta(CB)' + \Pi^{-1}\otimes\Delta$$

where n arises from

$$T\delta C\hat{\Pi}\delta C' = T^{-1}\ \overline{\Sigma e_t z_t'\hat{\Pi}^{-1}z_t e_t'}$$

$$\rightarrow n\Delta.$$

<u>The Matrix A-BC</u>

By substituting the observation equation into the dynamic (state transition) equation, we arrive at

$$z_{t+1} = Fz_t + By_t$$

where

$$F = A - BC.$$

In some sense the matrix F is more basic since it governs the speed of convergence of the forecast errors.

From the above define

$$T^{-1}\Sigma z_{t+1}z_t' = F\hat{\Pi} + BT^{-1}\Sigma y_t z_t'$$

where

$$T^{-1}\Sigma z_{t+1}z_t' = \hat{\Omega}[0I]'\hat{R}_-^{-1}\hat{\Omega}' = \hat{Z}$$

$$T^{-1}\Sigma y_t z_t' = \hat{H}_{1.}\hat{R}_t^{-1}\hat{\Omega}' = \hat{C}\hat{\Pi}.$$

Then an estimate for the matrix F is given by

$$\hat{F} = (Z-\hat{B}\hat{C}\hat{\Pi})\hat{\Pi}^{-1},$$

where $\hat{B}$ is given by (10).

<u>Matrix M</u>

In Aoki (1987), the matrix M is estimated by solving $H_{\cdot 1} = \Theta M$, i.e.,

$$\hat{M} = \hat{\Sigma}^{-1/2}\hat{U}'\hat{H}_{\cdot 1}. \tag{12}$$

Alternatively, from the definition

$$\begin{aligned} M &= T^{-1}\Sigma z_t y_{t-1}' \\ &= \Omega R_-^{-1} T^{-1}\Sigma y_{t-1}^- y_{t-1}' = \Omega R_-^{-1}(R_-)_{\cdot 1} \end{aligned} \tag{13}$$

since

$$(R_-)_{\cdot 1} = \begin{bmatrix} \hat{\Lambda}_0 \\ \hat{\Lambda}_1' \\ \vdots \\ \hat{\Lambda}_{k-1}' \end{bmatrix}$$

A third way is to minimize

$$(\mathrm{vec}H_{\cdot 1}-(I\otimes\Theta)\mathrm{vec}M)'W^{-1}(\mathrm{vec}H_{\cdot 1}-(I\otimes\Theta)\mathrm{vec}M)$$

for some weight matrix. For example, the first estimator of M is the case when W = I is used. A limited amount of Monte Carlo simulations for small number of observations (T=40 and 80) with 1000 and 750 replications, respectively, seems to indicate that (12) has better small sample properties.

Second-Round Estimators

The instrumental variable estimators are consistent. It is known also that any consistent estimator can be improved, under some technical conditions, to have the same asymptotic efficiency as the maximum likelihood estimators (mle) for static models by iterating once to produce a second-round estimator, see Rothenberg and Leenders (1964), and Bowden and Turkington (1984, Sec. 3.4) for example.[3] More specifically, let θ be the vector of parameters. Then letting $\ell n\ L$ be the logarithm of the likelihood function

$$\hat{\theta} = \theta_{1V} - \left(\frac{\partial^2\ \ell n\ L}{\partial\theta\partial\theta'}\right)^{-1}_{1V} \left(\frac{\partial\ \ell n\ L}{\partial\theta}\right)_{1V},$$

where $(\cdot)_{1V}$ denotes the derivatives evaluated at θ_{1V}. Then since θ_{1V} is consistent with sampling variance $O(T^{-\frac{1}{2}})$, where T is the sample size, $\hat{\theta}$ has the same asymptotic distribution as the mle.

We apply this general procedure to calculate the second-round estimate up the matrix C, A, and B. In practice $-\partial^2 \ell nL/\partial\theta\partial\theta'$ can be approximated by $T^{-1}\Sigma\left(\frac{\partial \ln L}{\partial\theta_t}\right)\left(\frac{\partial \ln L}{\partial\theta_t}\right)'$. Appendix 2 calculates the necessary derivatives.

We next show that the estimators of system theory origin and the IV estimators can be unified as special cases of the generalized method of moment estimators. Since $\hat{H}^A$ is $T^{-1}\Sigma y^+_{t+1}y^{-}_{t-1}{}'$ by definition, it is equal to $\Theta A\hat{\Omega} + \Theta BT^{-1}\Sigma e_t y^{-1}_{t-1} + GT^{-1}\Sigma e^+_{t+1}y^{-1}_{t-1}$. Taking note of (8), we calculate

$$\hat{H}^A - \hat{\Theta}A\hat{\Omega} = N$$

where

$$N = \Theta BT^{-1}\Sigma e_t y^{-}_{t-1}{}' + GT^{-1}\Sigma e^+_t(y^{-}_{t-2}{}' - y^{-}_{t-1}{}'\hat{R}^{-1}\hat{\Omega}'\hat{\Pi}^{-1}A\hat{\Omega})$$

except for the "edge" effects. (The term $e^+_{t+1}y^{-}_{t-1}{}'$ is replaced by $e^+_t y^{-}_{t-2}{}'$).

The covariance of vec N is denoted by W. It has a complicated structure and TW converges at T goes to infinty

(14) $$TW = T\text{cov}(\text{vec}N)$$

$$= R_{-} \otimes \Theta B\Delta B'\Theta$$

$$+ R_{-} \otimes G(I\otimes\Delta)G' + \hat{\Omega}'A'\hat{\Pi}^{-1}A\hat{\Omega}\otimes G(I\otimes\Delta)G' + \text{cross product terms.}$$

Consider the estimate of A which minimizes

$$\text{vec}(\hat{H}^A - \hat{\Theta}A\hat{\Omega})'\ W^{-1}\ \text{vec}(\hat{H}^A - \hat{\Theta}A\hat{\Omega}).$$

This is the generalized method of monment estimator. If W is replaced with I, then the system theoretic estimate of A

$$\hat{A} = \Sigma^{1/2}U'\hat{H}^A V\Sigma^{-1/2}$$

results and if W is replaced with $R_{-}\otimes I$, then the estimate of A with z_t as instrument, i.e., (9) results. An improved estimator of A may result if W is estimated by substituting $\hat{B}$, $\hat{A}$, $\hat{C}$ into the expression W in (14). For example, $R\_\otimes D$, where D is a diagonal matrix with Δ, $\Delta + H_1\Delta H_1'$, $\Delta + H_1\Delta H_1' + H_2\Delta H_2'$ ..., where $H_{k+1} = \hat{C}\hat{A}^k\hat{B}$ may be used.

Mutual Consistency Check

Since

$$\hat{\Delta} = T^{-1}\Sigma\hat{e}_t\hat{e}_t'$$

$$= \hat{\Lambda}_0 - \hat{C}T^{-1}\Sigma z_t y_t' - T^{-1}\Sigma y_t z_t'\hat{C} + \hat{C}\hat{\Pi}\hat{C}',$$

where

$$T^{-1}\Sigma z_t y_t' = \hat{\Pi} + T^{-1}\Sigma z_t\hat{e}_t',$$

and

$$T^{-1}\Sigma z_t\hat{e}_t' \to 0 \text{ a.s.},$$

as shown by Lai and Wei (1985), $\hat{\Lambda}_0 = \hat{\Delta} + \hat{C}\hat{\Pi}\hat{C}' + 0_p(T^{-\frac{1}{2}})$ is consistent with $\Lambda_0 = \Delta + C\Pi C'$.

Similar checks reveal that

$$\hat{\Lambda}_1 = \hat{C}\hat{M} + 0_p(T^{-\frac{1}{2}}),$$

where

$$\hat{M} = \hat{A}\hat{\Pi}\hat{C}' + \hat{B}\hat{\Delta} + 0_p(T^{-\frac{1}{2}}).$$

Monte Carlo Examples

This section reports a small scale Monte Carol experiments to demonstrate small sample behavior of the two types of the estimators.

The goodness of the model is greatly influenced by the signal-to-noise ratio to use the engineering terminology. As its measure, it is convenient to use the ratio of $\mathrm{tr}\Lambda_0$ over $\mathrm{tr}\Delta$. The former roughly measures the power of the signal in the data and the latter measure the power in the noise, or the innovation part. (More precisely $y_t = Cz_t + e_t$ implies that $C\Pi C'$ is the power of the signal and Δ is that of the innovation.) We take 2 particular cases of VAR(1) in the decreasing order of the signal-to-noise ratio.

The models are of the form $y_t = \Phi y_{t-1} + n_t$ where $\dim y_t = 2$. We can exactly calculate the system matrices in the balanced form and Λ_0. The two cases are:

Case 1: $\Phi = \begin{pmatrix} .7 & 1 \\ -.4 & 1 \end{pmatrix}$ and $\mathrm{cov}\ n_t = \begin{pmatrix} .1 & 0 \\ 0 & .1 \end{pmatrix} = N.$

In this case

$$\Lambda_0 = \begin{pmatrix} 1.549 & .026 \\ .026 & .653 \end{pmatrix}$$

and the balanced model has the parameters

$$A = \begin{pmatrix} .5357 & .6737 \\ -.6338 & .8643 \end{pmatrix}, \qquad C = \begin{pmatrix} 1.1089 & .2343 \\ -.3307 & .7856 \end{pmatrix}$$

$$M = \begin{pmatrix} 1.0678 & .4458 \\ -.6338 & .7565 \end{pmatrix}, \qquad \Pi = \begin{pmatrix} 1.0166 & .2814 \\ .2814 & .9533 \end{pmatrix}$$

and $\Delta = \mathrm{cov}\ n_t$.

$$S/N = \mathrm{tr}(\Lambda_0 - N)/\mathrm{tr}N = 9.9.$$

Case 2: $\Phi = \begin{pmatrix} .7 & .8 \\ -.4 & .6 \end{pmatrix}$, $\quad N = \begin{pmatrix} .1 & .05 \\ .05 & .1 \end{pmatrix}$

$$\Lambda_0 = \begin{pmatrix} .6110 & .0216 \\ * & .2928 \end{pmatrix}$$

$A = \begin{pmatrix} .5781 & .5454 \\ -.5916 & .7219 \end{pmatrix}$, $\quad C = \begin{pmatrix} .6943 & .1481 \\ -.2153 & .4774 \end{pmatrix}$

$M = \begin{pmatrix} .6790 & .2596 \\ -.1785 & .4669 \end{pmatrix}$, $\quad \Pi = \begin{pmatrix} .9446 & .1834 \\ .1834 & .8190 \end{pmatrix}$

and $\Delta = N$. $\quad S/N = 3.5$

In both cases, the number of sample points is 40 and 1000 replications are made.[4] The results are tabulated in Table 1 and 2.[5] (The values of Π given by (3) and the solutions of the Riccati equations are also compared. They are very close to each other and not tabulated.)

In each simulation run, n = 2 has been a priori imposed. However, the Monte Carlo simulations showed that when j = k = 2, the ratios of the average of the third singular values to that of the first is .010 in Case 1 and .021 in Case 2. Those of the average of the second singular values to the first is 0.84 and .70 in the two cases respectively. Because of these large discrepancies in the ratios, any automatic procedure to select the dimension is likely to choose two as the dimension. These simulation results seem to show that there is little to choose between the two based on the samll sample performances. If anything, the sample standard deviations for the IV estimator tends to be slightly smaller.

With the sample size T = 250 and 3600 replications, some statistics of the system estimators for the matrix A is as follows:

	Case 1				Case 2			
	a_{11}	a_{12}	a_{21}	a_{22}	a_{11}	a_{12}	a_{21}	a_{22}
mean	.530	.665	-.618	.844	.573	.540	-.584	.719
standard deviation	.041	.049	.053	.073	.068	.08	.109	.146
skewness	-.304	-.486	.714	-.586	-.100	-.282	.324	.201
kurtosis	1.44	2.44	1.38	2.11	.59	.61	1.19	2.54

Concluding Remarks

The recognition that z_t is the asymptotically most efficient instrumental variables for the matrix C leads to many new IV estimators for system matrices of system models for time series in (balanced) innovation presentations.

This note has shown how to construct some of them. In particular, the estimators of the matrice A have been shown to be special cases of the method of movement estimators with specific choices of the weight matrices. The use of the IV estimators avoid explicit solution of the Riccati equation. Small sample properties of these alternative estimators need be evaluated by more extensive Monte Carlo studies. Evidences from a limited amount of Monte Carlo simulations are that the IV estimators and the system theoretic estimators have about the same small sample properties, and the covariance matrix of the state vector given by (3) is about the same as that obtained by solving the Riccati equation explicitly in terms of mean values and sample variances. The results are best for the case $j = k = 1$. The state vector covariance matrices for the case $j = 1$, $k = 2$ tend to have smaller 2nd diagonal element and those for $j = k = 2$ tend to be larger than the true ones.

Footnotes

[1]By the nestedness we mean the orthogonality of estimated matrices, i.e., the properties that the appropriate submatrices of the estimates of matrices A, C, and M remain as the "correct" estimates when the state vector dimension is reduced, and when the state bector dimension is increased, the estimated matrices remain the same and newer estimated matrix elements are added to the existing ones, i.e., these submatrices are consistently estimated in the event of the dimension misspecification.

[2]The system theoretic estimator of the matrix C is $\hat{C} = H_{1.}V\Sigma^{-1/2}$ (Aoki 1987, p. 121). Note that the Moore-Penrose pseudo inverse of Ω is $V\Sigma^{-1/2}$.

[3]For dynamic systems, a sequence of such revisions is probably needed.

[4]Even though 10^3 sample paths are generated by the random number generator of MATLAB, some of them do not satisfy the regularity condition (Λ_0-CΠC'>0) or otherwise the Riccati equation solver fails. The percentage of failure ranges from 0 to a few percent of the number of samples generated by the random numbers. The usable samples thus vary from models to models and cases to cases. Since the number of failure is small, we have not made any correction for it.

[5]Sample standard deviations decrease in power of $1/\sqrt{T}$. For example with 240 sample points, the entries in sd (sample standard deviations) are expected to be reduced by $1/\sqrt{6}$.

With 250 replications, this observation is approximately confirmed. (Similarly, with T = 320 and 200 replications, the reduction by $1/\sqrt{8}$ is approximately confirmed.)

Table 1

Case 1

	j = 1 = k		j = 1, k = 2		j = 2 = k	
	S	IV	S	IV	S	IV
$\hat{A}$	$\begin{bmatrix} .520 & .645 \\ -.597 & .815 \end{bmatrix}$	$\begin{bmatrix} .519 & .646 \\ -.596 & .816 \end{bmatrix}$	$\begin{bmatrix} .507 & .710 \\ -.538 & .828 \end{bmatrix}$	$\begin{bmatrix} .507 & .716 \\ -.537 & .827 \end{bmatrix}$	$\begin{bmatrix} .516 & .606 \\ -.514 & .823 \end{bmatrix}$	$\begin{bmatrix} .519 & .609 \\ -.523 & .820 \end{bmatrix}$
sd	$\begin{bmatrix} .065 & .079 \\ .077 & .088 \end{bmatrix}$	$\begin{bmatrix} .063 & .084 \\ .090 & .085 \end{bmatrix}$	$\begin{bmatrix} .050 & .119 \\ .087 & .088 \end{bmatrix}$	$\begin{bmatrix} .053 & .125 \\ .086 & .091 \end{bmatrix}$	$\begin{bmatrix} .062 & .306 \\ .281 & .112 \end{bmatrix}$	$\begin{bmatrix} .065 & .317 \\ .268 & .105 \end{bmatrix}$
$\hat{C}$	$\begin{bmatrix} .940 & .192 \\ -.269 & .662 \end{bmatrix}$	$\begin{bmatrix} .942 & .192 \\ -.270 & .662 \end{bmatrix}$	$\begin{bmatrix} 1.036 & .264 \\ -.333 & .822 \end{bmatrix}$	$\begin{bmatrix} 1.030 & .263 \\ -.330 & .815 \end{bmatrix}$	$\begin{bmatrix} .925 & .054 \\ -.214 & .620 \end{bmatrix}$	$\begin{bmatrix} .926 & .077 \\ -.211 & .607 \end{bmatrix}$
sd	$\begin{bmatrix} .269 & .090 \\ .122 & .199 \end{bmatrix}$	$\begin{bmatrix} .264 & .094 \\ .125 & .199 \end{bmatrix}$	$\begin{bmatrix} .313 & .133 \\ .153 & .278 \end{bmatrix}$	$\begin{bmatrix} .297 & .125 \\ .147 & .264 \end{bmatrix}$	$\begin{bmatrix} .293 & .209 \\ .183 & .335 \end{bmatrix}$	$\begin{bmatrix} .321 & .216 \\ .193 & .336 \end{bmatrix}$
$\hat{M}$	$\begin{bmatrix} .898 & .390 \\ -.275 & .633 \end{bmatrix}$	$\begin{bmatrix} .899 & .390 \\ -.274 & .633 \end{bmatrix}$	$\begin{bmatrix} .821 & .342 \\ -.185 & .522 \end{bmatrix}$	$\begin{bmatrix} .816 & .340 \\ -.185 & .519 \end{bmatrix}$	$\begin{bmatrix} .832 & .455 \\ -.307 & .546 \end{bmatrix}$	$\begin{bmatrix} .829 & .455 \\ -.282 & .554 \end{bmatrix}$
sd	$\begin{bmatrix} .265 & .123 \\ .094 & .195 \end{bmatrix}$	$\begin{bmatrix} .262 & .122 \\ .095 & .196 \end{bmatrix}$	$\begin{bmatrix} .253 & .110 \\ .086 & .171 \end{bmatrix}$	$\begin{bmatrix} .239 & .110 \\ .089 & .159 \end{bmatrix}$	$\begin{bmatrix} .279 & .202 \\ .261 & .301 \end{bmatrix}$	$\begin{bmatrix} .307 & .231 \\ .303 & .299 \end{bmatrix}$
$\hat{\Delta}$	$\begin{bmatrix} .180 & .004 \\ * & .127 \end{bmatrix}$	$\begin{bmatrix} .180 & .003 \\ * & .127 \end{bmatrix}$	$\begin{bmatrix} .181 & .004 \\ * & .127 \end{bmatrix}$	$\begin{bmatrix} .180 & .004 \\ * & .127 \end{bmatrix}$	$\begin{bmatrix} .178 & .003 \\ * & .124 \end{bmatrix}$	$\begin{bmatrix} .182 & .004 \\ * & .128 \end{bmatrix}$
sd	$\begin{bmatrix} .066 & .025 \\ * & .039 \end{bmatrix}$	$\begin{bmatrix} .068 & .036 \\ * & .038 \end{bmatrix}$	$\begin{bmatrix} .069 & .025 \\ * & .038 \end{bmatrix}$	$\begin{bmatrix} .066 & .025 \\ * & .039 \end{bmatrix}$	$\begin{bmatrix} .069 & .027 \\ * & .039 \end{bmatrix}$	$\begin{bmatrix} .069 & .025 \\ * & .038 \end{bmatrix}$
$\hat{\Pi}$	$\begin{bmatrix} .972 & .262 \\ * & .882 \end{bmatrix}$	$\begin{bmatrix} .973 & .263 \\ * & .883 \end{bmatrix}$	$\begin{bmatrix} .788 & .207 \\ * & .590 \end{bmatrix}$	$\begin{bmatrix} .786 & .207 \\ * & .588 \end{bmatrix}$	$\begin{bmatrix} 1.043 & .278 \\ * & .867 \end{bmatrix}$	$\begin{bmatrix} 1.043 & .276 \\ * & .889 \end{bmatrix}$
sd	$\begin{bmatrix} .043 & .049 \\ * & .072 \end{bmatrix}$	$\begin{bmatrix} .041 & .048 \\ * & .069 \end{bmatrix}$	$\begin{bmatrix} .038 & .049 \\ * & .035 \end{bmatrix}$	$\begin{bmatrix} .038 & .047 \\ * & .036 \end{bmatrix}$	$\begin{bmatrix} .097 & .113 \\ * & .089 \end{bmatrix}$	$\begin{bmatrix} .095 & .113 \\ * & .095 \end{bmatrix}$

*Symmetric elements

Table 2

Case 2

	j = 1 = k S	j = 1 = k IV	j = 1, k = 2 S	j = 1, k = 2 IV	j = 2 = k S	j = 2 = k IV
$\hat{A}$	$\begin{bmatrix} .556 & .524 \\ -.577 & .683 \end{bmatrix}$	$\begin{bmatrix} .554 & .527 \\ -.589 & .684 \end{bmatrix}$	$\begin{bmatrix} .523 & .406 \\ -.371 & .719 \end{bmatrix}$	$\begin{bmatrix} .522 & .405 \\ -.368 & .712 \end{bmatrix}$	$\begin{bmatrix} .514 & .420 \\ -.436 & .737 \end{bmatrix}$	$\begin{bmatrix} .514 & .422 \\ -.441 & .741 \end{bmatrix}$
sd	$\begin{bmatrix} .072 & .117 \\ .120 & .151 \end{bmatrix}$	$\begin{bmatrix} .070 & .097 \\ .086 & .143 \end{bmatrix}$	$\begin{bmatrix} .066 & .416 \\ .412 & .145 \end{bmatrix}$	$\begin{bmatrix} .067 & .416 \\ .409 & .144 \end{bmatrix}$	$\begin{bmatrix} .060 & .341 \\ .408 & .155 \end{bmatrix}$	$\begin{bmatrix} .062 & .342 \\ .411 & .141 \end{bmatrix}$
$\hat{C}$	$\begin{bmatrix} .639 & .134 \\ -.198 & .431 \end{bmatrix}$	$\begin{bmatrix} .641 & .137 \\ -.270 & .437 \end{bmatrix}$	$\begin{bmatrix} .678 & .139 \\ -.290 & .371 \end{bmatrix}$	$\begin{bmatrix} .677 & .136 \\ -.287 & .369 \end{bmatrix}$	$\begin{bmatrix} .617 & .088 \\ -.248 & .353 \end{bmatrix}$	$\begin{bmatrix} .615 & .086 \\ -.245 & .349 \end{bmatrix}$
sd	$\begin{bmatrix} .149 & .063 \\ .079 & .123 \end{bmatrix}$	$\begin{bmatrix} .148 & .060 \\ .079 & .113 \end{bmatrix}$	$\begin{bmatrix} .168 & .191 \\ .097 & .378 \end{bmatrix}$	$\begin{bmatrix} .169 & .189 \\ .096 & .378 \end{bmatrix}$	$\begin{bmatrix} .151 & .154 \\ .092 & .275 \end{bmatrix}$	$\begin{bmatrix} .158 & .153 \\ .092 & .275 \end{bmatrix}$
$\hat{M}$	$\begin{bmatrix} .622 & .246 \\ -.169 & .415 \end{bmatrix}$	$\begin{bmatrix} .624 & .247 \\ -.170 & .425 \end{bmatrix}$	$\begin{bmatrix} .574 & .195 \\ -.069 & .260 \end{bmatrix}$	$\begin{bmatrix} .572 & .195 \\ -.670 & .258 \end{bmatrix}$	$\begin{bmatrix} .625 & .240 \\ -.105 & .346 \end{bmatrix}$	$\begin{bmatrix} .624 & .239 \\ -.105 & .345 \end{bmatrix}$
sd	$\begin{bmatrix} .142 & .054 \\ .074 & .117 \end{bmatrix}$	$\begin{bmatrix} .144 & .086 \\ .066 & .109 \end{bmatrix}$	$\begin{bmatrix} .128 & .078 \\ .054 & .288 \end{bmatrix}$	$\begin{bmatrix} .126 & .084 \\ .058 & .286 \end{bmatrix}$	$\begin{bmatrix} .137 & .099 \\ .082 & .299 \end{bmatrix}$	$\begin{bmatrix} .141 & .103 \\ .083 & .299 \end{bmatrix}$
$\hat{\Delta}$	$\begin{bmatrix} .129 & .046 \\ .046 & .106 \end{bmatrix}$	$\begin{bmatrix} .130 & .046 \\ .046 & .106 \end{bmatrix}$	$\begin{bmatrix} .129 & .045 \\ * & .105 \end{bmatrix}$	$\begin{bmatrix} .129 & .046 \\ * & .106 \end{bmatrix}$	$\begin{bmatrix} .128 & .045 \\ * & .104 \end{bmatrix}$	$\begin{bmatrix} .130 & .045 \\ * & .107 \end{bmatrix}$
sd	$\begin{bmatrix} .035 & .018 \\ * & .036 \end{bmatrix}$	$\begin{bmatrix} .034 & .018 \\ * & .025 \end{bmatrix}$	$\begin{bmatrix} .034 & .018 \\ * & .025 \end{bmatrix}$	$\begin{bmatrix} .035 & .018 \\ * & .026 \end{bmatrix}$	$\begin{bmatrix} .035 & .018 \\ * & .025 \end{bmatrix}$	$\begin{bmatrix} .035 & .038 \\ * & .026 \end{bmatrix}$
$\hat{\Pi}$	$\begin{bmatrix} .904 & .170 \\ * & .785 \end{bmatrix}$	$\begin{bmatrix} .903 & .169 \\ * & .783 \end{bmatrix}$	$\begin{bmatrix} .723 & .119 \\ * & .576 \end{bmatrix}$	$\begin{bmatrix} .722 & .114 \\ * & .577 \end{bmatrix}$	$\begin{bmatrix} .917 & .207 \\ * & .824 \end{bmatrix}$	$\begin{bmatrix} .914 & .204 \\ * & .812 \end{bmatrix}$
sd	$\begin{bmatrix} .050 & .057 \\ * & .091 \end{bmatrix}$	$\begin{bmatrix} .053 & .055 \\ * & .093 \end{bmatrix}$	$\begin{bmatrix} .039 & .147 \\ * & .055 \end{bmatrix}$	$\begin{bmatrix} .040 & .150 \\ * & .037 \end{bmatrix}$	$\begin{bmatrix} .076 & .171 \\ * & .101 \end{bmatrix}$	$\begin{bmatrix} .073 & .172 \\ * & .102 \end{bmatrix}$

*Symmetric element

Appendix 1

We start with a forward innovation model in which the state vector z_t is uncorrelated with the innovations of the data vector, e_s, $s \le t$,

$$z_{t+1} = Az_t + Be_t \tag{A.1}$$

$$y_t = Cx_t + e_t.$$

The dimension of the vector y_t is p. Lindquist et al. (1979) proved the existence of such a representation and that the state space of this model is a minimal splitting subspace if and only if the model is observable and A in invertible, or equivalently both observable and constructible.* Let $n = \dim z_t$.

Define a stacked future and past vectors by $y_t^+ = (y_t', y_{t+1}', \ldots)'$ and $y_t^- = (y_t', y_{t-1}', \ldots)'$.

We briefly describe how (A.1) arises from a common state space model such as the one in (A.2). A state space model consists of two equations; one describes how the state vector evolves with time, and the other specifies how that state vector is related to the data vector. The model involves a single lag for the state vector. A linear state space model is then of the form

---

*Note that the direction of time is reversed in these two related notions. Observability is the ability to reconstruct (or estimate consistently) the unobserved state vector from future observations. Constructibility has to do with the same ability using past data.

$$\text{(A.2)} \qquad x_{t+1} = Ax_t + u_t,$$

$$y_t = Cx_t + \nu_t.$$

The noises are assumed to be serially uncorrelated. This can be achieved by suitably augmenting the state vector to include noise dynamics in the state dynamic equation. This may be one of the ways that the state vector becomes not directly available for observation. Generally, only the vector y_t is directly observed. Let z_t be the estimate of x_t based on y_{t-1}^-,

$$\text{(A.3)} \qquad z_t = \hat{E}(x_t | y_{t-1}^-).$$

Then the innovation of y_t is defined to be

$$\text{(A.4)} \qquad e_t = y_t - \hat{E}(y_t | y_{t-1}^-)$$

$$= y_t - Cz_t$$

since u_t and ν_t are uncorrelated with y_{t-1}^- by assumption.

The vector z_t evolves with time as follows

$$\text{(A.5)} \qquad z_{t+1} = E(x_{t+1} | y_t^-)$$

$$= \hat{E}(x_{t+1} | y_{t-1}^-, e_t)$$

$$= \hat{E}(x_{t+1} | y_{t-1}^-) + Be_t$$

$$= Az_t + Be_t$$

where

$$\hat{E}(x_{t+1} | y_{t-1}^-) = \hat{E}(Ax_t | y_{t-1}^-)$$

$$= Az_t$$

and

$$B = \hat{E}(x_{t+1}|e_t)$$

$$= E(x_{t+1}e_t')\Delta^{-1}$$

with

$$\Delta = \mathrm{cov}\ (e_t).$$

Note that

$$B\Delta = E(x_{t+1}e_t') = E(z_{t+1}e_t')$$

$$= E\big(z_{t+1}(y_t - Cz_t)'\big)$$

$$= M - A\Pi C'$$

where

$$M = E(z_{t+1}y_t')$$

and

$$\Pi = \mathrm{cov}\ z_t.$$

The state space model (A.2) is thus put into the innovative representation which consists of (A.4) and (A.5). This is called a forward innovation model because time flows from past to future. Later a backward innovation model is introduced in which the time flow is reversed. From (A.2) we can relate y_t^+ to the state vector x_t by

(A.6) $$y_t^+ = \Theta x_t + Ge_t^+$$

where e_t^+ is defined similarly and O is the matrix, $[C'A'C'A'^2C'\ldots]'$. The matrix G is block lower triangular with with the main diagonal submatrices are the p-dimensional identity matrix, i.e., the determinant of the matrix G is one.

Consider predicting y_t^+ by its orthogonal projection on the manifold spanned by the data vectors. Using the notation $\hat{E}(u|v)$ to denote the orthogonal projection of the vector u on the manifold spanned by v, we derive

$$(A.7) \qquad \hat{E}(y_t^+|y_{t-1}^-) = HR_-^{-1}y_{t-1}^-$$

where H is the covariance matrix between the two stacked vector, called Hankel matrix, and

$$R_- = \mathrm{cov}(y_{t-1}^-).$$

Alternatively from (A.6), since e_t^+ is uncorrelated with the stacked data vector, we can write (A.7) in view of (A.3) as

$$(A.8) \qquad \hat{E}(y_t^+|y_{t-1}^-) = \Theta\hat{E}(x_t|y_{t-1}^-) = \Theta z_t$$

$$= \Theta\Omega R_-^{-1}y_{t-1}^-$$

where we define a matrix

$$(A.9) \qquad \Omega = E(x_t y_{t-1}^{-}{}') = E(z_t y_{t-1}^{'})$$

since $x_t - z_t$ is orthogonal to y_{t-1}^-.

From the two right-hand expressions in (A.7) and (A.8) and denoting the orthogonally projected image of the state vector by z_t in (A.3), we obtain its explicit relation in terms of the stacked data vector as

(A.10) $$z_t = \Omega R_{-}^{-1} y_{t-1}^{-}$$

if O is full rank, i.e., if (A.2) is observable. The state space model in (A.1) is observable by construction. Eq. (A.10) shows that the covariance matrix of the vector is

$$\Pi = \Omega R_{-}^{-1} \Omega'.$$

From the definition we also have a useful relation between the two state vectors x_t and z_t

$$\mathrm{cov}(z_t) \le \mathrm{cov}(x_t).$$

i.e., the model with z_t as its state vector has the smallest covariance matrix of all state vectors.

Appendix 2

In Koopmans (1950, p. 115) we find

$$\frac{\partial}{\partial\theta} \ln |D| = \mathrm{tr}\ (D^{-1} \frac{\partial D}{\partial\theta})$$

where θ is a scalar variable, and

$$\frac{\partial \mathrm{tr}(LMN)}{\partial M} = L'N'.$$

We apply these relations. When e_t is normally distributed with zero mean and finite variance, the concentrated log-likelihood function is

$$\ln L^* = k - \frac{T}{2} \ln|D|$$

where

$$\begin{aligned} D &= T^{-1}\Sigma\hat{e}_t\hat{e}_t' \\ &= \Lambda_0 - \hat{C}T^{-1}\Sigma z_t y_t' - T^{-1}\Sigma y_t z_t'\hat{C}' \\ &\quad + \hat{C}\hat{\Pi}\hat{C}'. \end{aligned}$$

From this, we obtain

$$dD = T^{-1}\sum(d\hat{e}_t\hat{e}_t' + \hat{e}_t d\hat{e}_t')$$

where

$$d\hat{e}_t = -dCz_t - Cdz_t,$$

i.e.,

$$dD = -dC\hat{X}' - \hat{X}dC' - CdU' - dUC'$$

where

$$\hat{X} = T^{-1}\sum \hat{e}_t z_t'$$

and

$$dU = T^{-1}\sum \hat{e}_t dz_t'.$$

From

$$d\ \ell n\ L^* = -\frac{1}{2}\,tr(D^{-1}dD)$$

we calculate

$$\text{(A.1)} \qquad \frac{\partial\ \ell n\ C^*}{\partial C} = D^{-1}\,\hat{X}.$$

Using

$$dz_t = dAz_{t-1},$$

$$\text{(A.2)} \qquad \frac{\partial\ \ell n\ L^*}{\partial A} = C'D^{-1}\,\hat{X}_{-1}$$

where

$$\hat{X}_{-1} = T^{-1}\sum \hat{e}_t z_{t-1}'.$$

Finally using $dz_t = dBe_{t-1}$

$$\text{(A.3)} \qquad \frac{\partial\ Ln\ L^*}{\partial B} = C'D^{-1}\,\hat{\Delta}_{-1}$$

where

$$\hat{\Delta}_{-1} = T^{-1}\sum \hat{e}_t \hat{e}_{t-1}'.$$

From (A.1)

$$TE\left(\frac{\partial\ \ell n\ L^*}{\partial C}\right)\left(\frac{\partial\ \ell n\ L^*}{\partial C}\ '\right) = (\mathrm{tr}\Pi)\ \Delta^{-1}$$

$$T\ E\left(\frac{\partial\ \ell n\ L^*}{\partial A}\right)\left(\frac{\partial\ \ell n\ L^*}{\partial A}\ '\right) = \mathrm{tr}(\pi)(C'\Delta^{-1}C)$$

$$TE\left(\frac{\partial\ \ell n\ L^*}{\partial B}\right)\left(\frac{\partial\ \ell n\ L^*}{\partial B}\ '\right) = (\mathrm{tr}\Delta)C'\Delta^{-1}C$$

$$TE\left(\frac{\partial\ \ell n\ C^*}{\partial C}\right)\left(\frac{\partial\ \ell n\ C^*}{\partial A}\ '\right) = (\mathrm{tr}\ A\pi)\Delta^{-1}C$$

$$TE\left(\frac{\partial\ \ell n\ C^*}{\partial C}\right)\left(\frac{\partial\ \ell n\ C^*}{\partial B}\ '\right) = \mathrm{tr}(B\Delta)\Delta^{-1}C$$

$$TE\left(\frac{\partial\ \ell n\ C^*}{\partial A}\right)\left(\frac{\partial\ \ell n\ C^*}{\partial B}\ '\right) = 0.$$

References

Aoki, M. 1987. State space modeling of time series. Springer-Verlag, Heidelber, New York.

_______________. 1989. Singular values, cannonical correlation coefficients and the dimension of state space models of time series. To be presented at the 1989 NSF/NBER Seminar, Madrid, September 1989.

_______________. 1988. State space models for vector-valued time series with real and complex random walk components. Invited lecture, 1988 Am. Stat. Assoc. Meeting, New Orleans, Aug. 1988.

Bowden, R. J., and Turkington, D. A. 1984. Instrumental variable. Cambridge University Press: New York.

Havenner, A., and Aoki, M. 1988. An instrumental variable interpretation of linear system theory estimation. J. of Econ. Dyn. Control 12.

Koopmans, T.C. (ed.) 1950. Statistical Inference in Dynamic Economic Models, Wiley, New York.

Lai, T. L., and Wei, C. Z. 1985. Asymptotic properties of multivariate weighted sumes with application to stochastic regression in linear dynamic systems," in D. R. Krishnaiah (ed.), Multivariate Analysis, North Holland, Amsterdam.

Lindquist, A.; Picci, G., and Ruckebush, G. 1979. On splitting subspaces and Markovian representations. Math Syst. Theory 12, 271-79.

Lindquist, A., and Pavon, M. 1984. On the structure of state-space models for discrete-time stochastic vector processes. IEEE AC-29, 418-32.

Rothenberg, T. J. and Leenders, C. T. 1964. Efficient estimation of simultaneous equation systems. Econometrica 32, 57-76.

Sargan, D. 1988. Lectures on Advanced Econometric Theory. Edited by M. Desai, Basil Blackwell, Oxford 1984.