Identification of Singular Autoregressive Models Revisited

PHOEBUS J. DHRYMES*
Columbia University

November 2005

Abstract

In an earlier paper, Dhrymes (1994b), I examined the estimation problem presented by systems of singular autoregressive models owing to the singularity of the covariance matrix of their structural errors. At that time the issue of the identification of such models was not addressed directly; instead, the paper determined the conditions under which the estimation procedure suggessted therein could be implemented. In this paper I show that these conditions are implied by the indentifiability of the model based on the asymptotic Kullback information.

(JEL C10, C30, C43)
Key words: Singular Autoregressive Errors, Kullback Information, Identification.

1 The Model and its Identification Conditions

1.1 The Model

We consider the model

$$y_{t\cdot} = x_{t\cdot}B + u_{t\cdot}, \tag{1}$$

where $y_{t\cdot}$, $x_{t\cdot}$ are an m-element and G-element **row** vector, respectively; what is special about these models is the so called adding up property

$$\sum_{i=1}^{m} y_{ti} = x_{tG}, \quad \text{identically, for all } t. \tag{2}$$

As is well known this implies

$$Be = e_{\cdot G}, \quad u_{t\cdot}e = 0. \tag{3}$$

When the error process is autoregressive, i.e.

$$u_{t\cdot} = u_{t-1\cdot}H + \epsilon_{t\cdot}, \quad t = 1, 2, \ldots, T, \tag{4}$$

we also have the restriction

$$\epsilon_{t\cdot}e = 0, \tag{5}$$

where $e, e_{\cdot G}$ are m-element and G-element (column) vectors whose elements are all unity, and all zero except the Gth which is unity, respectively. These conditions are self enforced when we estimate B by regressing Y on X due to the latter's respective definitions. We denote the covariance matrices of $\epsilon_{t\cdot}$, $u_{t\cdot}$ by Σ_0, Ω_0, respectively and, in a departure from the conventions of Dhrymes (1994), we use explictly $U_{-1} = Y_{-1} - X_{-1}B$.

1.2 Identification

We adopt the criterion for identification given in the context of MC (minimum contrast) estimators, see Ducunha-Castelle and Duflo (1986, vol. I), or Dhrymes (1994a, ch. 5), as follows. Define

$$J_T(\theta) = \frac{1}{T}\left(y - [I_m \otimes X, I_m \otimes U_{-1}]\theta\right)'\left(y - [I_m \otimes X, I_m \otimes U_{-1}]\theta\right), \tag{6}$$

$$\theta = (\mathrm{vec}(B)', \mathrm{vec}(H)')', \tag{7}$$

and note that this is the function we wish to minimize with respect to θ, in order to obtain the MC estimator. It entails no loss of generality to write instead

$$J_T^*(\theta) = J_T(\theta) - J_T(\theta^0), \tag{8}$$

where θ^0 is the true parameter vector. In the theory of MC estimators, it is shown that under certain regularity conditions the estimator, $\hat{\theta}$, defined by the condition

$$J_T^*(\hat{\theta}) = \inf_{\theta \in \Theta} J_T^*(\theta), \tag{9}$$

where Θ is the admissible parameter space, converges to θ^*, such that

$$K(\theta^*, \theta^0) = \inf_{\theta \in \Theta} K(\theta, \theta^0), \quad K(\theta, \theta^0) = \operatorname*{plim}_{T \to \infty} J_T^*(\theta). \tag{10}$$

A useful property of the (asymptotic) Kullback information is that it is non-negative and assumes its global minimum, zero, at $\theta = \theta^0$, i.e.

$$K(\theta, \theta^0) \geq 0, \quad \text{for all } \theta \in \Theta, \quad K(\theta^0, \theta^0) = 0. \tag{11}$$

The condition for identification then is quite simple: the global minimizer is unique, i.e.

$$K(\theta^*, \theta^0) = K(\theta^0, \theta^0), \quad \text{implies } \theta^* = \theta^0. \tag{12}$$

In the present case, the model is identifiable if and only if

$$K(\theta, \theta^0) = 0, \quad \text{implies } \theta = \theta^0. \tag{13}$$

Thus, we consider the probability limit[1] of $J_T^*(\theta)$. For convenience in subsequent discussion, rewrite $J_T^*(\theta)$ as

$$J_T^*(\theta, \theta^0) = \frac{1}{T}\left(\operatorname{vec}(E) - P(\theta - \theta^0)\right)'\left(\operatorname{vec}(E) - P(\theta - \theta^0)\right) - \frac{1}{T}\operatorname{vec}(E)'\operatorname{vec}(E), \tag{14}$$

where

$$P = [I_m \otimes X), I_m \otimes U_{-1}]. \tag{15}$$

From Dhrymes (2000), p. 121, we can also write

$$J_T^*(\theta, \theta^0) = \operatorname{tr}\frac{1}{T}\left([(B - B^0)', (H - H^0)']\begin{bmatrix} X'X & X'U_{-1} \\ U_{-1}'X & U_{-1}'U_{-1} \end{bmatrix}[(B - B^0)', (H - H^0)']'\right). \tag{16}$$

Since it is evident in Eq. (16) that upon taking probability limits all cross products vanish, we find

$$\operatorname*{plim}_{T \to \infty} J_T^*(\theta, \theta^0) = \operatorname{tr}\begin{bmatrix} (B - B^0)'M_{xx}(B - B^0) & 0 \\ 0 & (H - H^0)'\Omega_0(H - H^0) \end{bmatrix} = K(\theta, \theta^0)$$

$$K(\theta, \theta^0) = \operatorname{tr}(B - B^0)'M_{xx}(B - B^0) + \operatorname{tr}(H - H^0)'\Omega_0(H - H^0), \tag{17}$$

where

$$M_{xx} = \operatorname*{plim}_{T \to \infty} \frac{X'X}{T}, \quad \Omega_0 = \operatorname*{plim}_{T \to \infty} \frac{U_{-1}'U_{-1}}{T}. \tag{18}$$

In view of Eq. (17) and the fact that M_{xx} is positive definite, we conclude that

$$K(\theta, \theta^0) = 0 \quad \text{implies}, \quad B = B^0, \tag{19}$$

but does not imply $H = H^0$, due to the fact that Ω_0 is **singular**. To see why this is so suppose that one equation of the model, say the kth, contains all lags and consider

[1] Under stronger conditions we may derive an a.c. limit, i.e. a limit with probability one.

$H = H^0 + c_k(e'_{\cdot k} \otimes e)$. This is an admissible H, provided we take c_k small enough to preserve stability. Note now that

$$\mathrm{tr}(H-H^0)'\Omega_0(H-H^0) = \mathrm{tr} c_k^2(e_{\cdot k} \otimes e')\Omega_0(e'_{\cdot k} \otimes e) = c_k^2 e'\Omega_0 e = 0. \tag{20}$$

But this shows that the system is not identified. More, generally, if H^* is such that $\theta^* = (b^{0\prime}, \mathrm{vec}(H^*)')'$ obeys

$$K(\theta^*, \theta^0) = \inf_{\theta \in \Theta} K(\theta, \theta^0) = 0, \tag{21}$$

then $H^{**} = H^* + c' \otimes e$, where c is an arbitrary vector, subject to the preservation of stability, is also a set of parameters such that,

$$K(\theta^{**}, \theta^0) = \mathrm{tr}(c \otimes e')\Omega_0(c' \otimes e) = \sum_{i=1}^{m} c_i^2 e'\Omega_0 e = 0, \tag{22}$$

with θ^{**}, defined *mutatis mutandis*. But this shows that the **system is not identified**, because the minimizer is not unique.

We have therefore proved

Theorem 1. Consider the model specified in Eqs. (1) through (5) above. A necessary and sufficient condition for the indentifiability of its parameters is that for all admissible H^*: $(H^* - H^0)$ is in the (row or column) null space of Ω_0 implies $H^* = H^0$, i.e. no column (of $(H^* - H^0)$) is of the form ce, for arbitrary c, (but sufficiently small to preserve stability).

Proof: Previous discussion.

Remark 1. Lack of identification is confined only to the elements of the matrix H, and does not affect those of the matrix B, which are always identifiable. This in effect means that no equation in the system $u_{t\cdot} = u_{t-1\cdot}H + \epsilon_{t\cdot}$, **is allowed to contain all the lags**; moreover, it also implies that if all lags appear in all equations, i.e. we have the standard vector autoregressive specification, the model's parameters are not identified!

Remark 2. This set of conditions is **precisely the same set of conditions** determined in Dhrymes (1994b) for the iterative estimation procedure to be carried out, as set forth therein. In that paper, p. 265, he obtains a sufficient condition for the inversion of the iteration matrix by considering the matrix W_{22}, Eq. (41), and the column null space of its three components. While two of these components are **symmetric**, the third is not, so the matrix W_{22} is not symmetric. But for non-symmetric matrices the representation of the column null space differs from that of the row null space. For example, for the matrices

$$C = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, C' = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \tag{23}$$

the column null space of C is generated by the vector $(0,1)'$, while that of C' is generated by the vector $(1,-1)'$; however, both describe the same space, since these two bases are connected by the (non-singular) linear transformation

$$\begin{bmatrix} 1 & 1 \\ 3 & -2 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{24}$$

The intersection of the three **row** spaces therein is given by the class of vectors $(c\otimes e)'$, where c is an arbitrary vector. But given the notation in that paper the solution $S_2\alpha = c\otimes e$ implies that $S_{i2}\alpha_i = c_i e$, $i = 1, 2, \ldots, m$, i.e. that the vector e is in the column space of S_{i2}; since S_{i2} is a permutation of $m_i < m$ columns of the identity matrix I_m (indicating which lags appear in the ith equation), this is impossible because some **rows** of the $m \times m_i$ matrix S_{i2} consist entirely of zeros! Thus, the conditions imposed by Dhrymes (1994b) on the matrix W_{22} in order to render the iterative procedure therein operable, **are implied by the indentifiability of the model**.

2 Implications for Estimation

In the discussion surrounding Eq. (16) we have derived the conditions for identification by considering the probability limit of the contrast function to be minimized. The estimation implied by the procedure above requires us to obtain an "initial consistent" estimator of U_{-1}, which we can always do by regressing Y on X, thus obtaining

$$\hat{B} = (X'X)^{-1}X'Y, \quad \hat{U} = Y - X\hat{B} = [I_T - X(X'X)^{-1}X']Y. \tag{25}$$

It is easily verified that these estimators obey the required conditions, i.e.

$$\hat{B}e = e_{\cdot G}, \quad \hat{U}e = 0, \quad \text{because} \quad Ye = x_{\cdot G}. \tag{26}$$

If in the first iteration we impose the identification conditions $\mathrm{vec}(H) = S_2\gamma$, as well as the optional (non-required) conditions $\mathrm{vec}(B) = S_1\beta$, and minimize

$$F_0 = \epsilon'\epsilon, \quad \mathrm{vec}(E) = \epsilon = (y - D\theta), \quad D = (I_m \otimes X)S_1, (I_m \otimes \hat{U}_{-1})S_2), \tag{27}$$

the resulting estimators are $\hat{B}_{(1)} = mat(S_1\hat{\beta}_1)$, $\hat{H}_{(1)} = mat(S_2\hat{\gamma}_{(1)})$, where the notation mat(a) means the inverse of the operation $\mathrm{vec}(A) = a$, and the subscript (1) indicates the first iterate. Unfortunately however, these estimators do not obey, for every sample size, the conditions

$$\hat{B}_{(1)}e = e_{\cdot G}, \quad \hat{U}_{(1)}e = 0, \quad I^*_{m-1}\hat{H}_{(1)} = 0 \tag{28}$$

as required by the specification of the model. Without loss of relevance and, as was done in Dhrymes (1994b), we take Σ_g as known, and we impose the restrictions above by means of Lagrange multipliers, so that we deal with

$$F_1 = \epsilon'(\Sigma_g \otimes I_T)\epsilon + 2\lambda'(R\theta - r), \quad R = \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix}, \quad r = (r_1', r_2')', \quad \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, \tag{29}$$

where

$$R_1 = (e' \otimes I_G)S_1, \quad R_2 = (e' \otimes I^*_{m-1})S_2, \quad r_1 = e_{\cdot G}, \quad r_2 = 0. \tag{30}$$

If we operate subject to these restrictions then, **at each iteration, and for every sample size,** we have

$$\hat{B}_{(k)}e = e_{\cdot G}, \quad I^*_{m-1}\hat{H}_{(k)} = 0, \quad \hat{U}_{-1(k)}e = 0, \quad \text{because,} \quad \hat{U}_{-1(k)} = Y_{-1} - X_{-1}\hat{B}_{(k)}, \tag{31}$$

where the subscript (k) indicates the kth iterate. Note further that

$$\operatorname{rank}(R_1) = G, \quad \operatorname{rank}(R_2) = m - 1, \tag{32}$$

i.e. they are both of full row rank.

The (first order conditions) estimating equations are

$$\begin{pmatrix} D'(\Sigma_g \otimes I_T)y \\ r \end{pmatrix} = \begin{bmatrix} A_{(k)} & R' \\ R & 0 \end{bmatrix} \begin{pmatrix} \theta \\ \lambda \end{pmatrix}, \quad A_{(k)} = D'_{(k)}(\Sigma_g \otimes I_T)D_{(k)} \tag{33}$$

$$Y - X\hat{B}_{(k)} = \hat{U}_{(k)}, \quad A_{(k)} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{34}$$

$$A_{11} = S_1'(\Sigma_g \otimes (X'X/T)S_1, \ A_{12} = S_1'(\Sigma_g \otimes [X'\hat{U}_{-1(k)}/T])S_2,$$

$$A_{22} = S_2'(\Sigma_g \otimes [\hat{U}'_{-1(k)}\hat{U}_{-1(k)}/T)S_2, \quad A_{21} = A_{12}'. \tag{35}$$

Noting further that

$$\operatorname{rank} A_{(k)} = \operatorname{rank}(D_{(k)}) = G^* + m^* \leq (G+m)m, \quad G^* = \sum_{i=1}^{m} G_i, \quad m^* = \sum_{i=1}^{m} m_i \tag{36}$$

where G^*, m^* are, respectively, the number of columns of S_1, S_2; it is clear by construction that, for every iteration, $D_{(k)}$ is of full column rank and thus

$$\operatorname{rank}(A_{(k)}) = G^* + m^*, \quad \text{because} \quad \operatorname{rank}(\Sigma_g \otimes I_T) = (m-1) \times T, \quad G^* + m^* < (m-1)T. \tag{37}$$

Consequently, for every iteration k, $A_{(k)}$ and $R[A_{(k)} + R'R]^{-1}R' = RV_{11}R'$ are **both non-singular**. Thus from Dhrymes (2000) p. 49 we conclude[2] that

$$\begin{bmatrix} A_{(k)} & R' \\ R & 0 \end{bmatrix}^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \quad V_{11} = (A_{(k)} + R'R)^{-1} \tag{38}$$

$$B_{11} = V_{11} - V_{11}R'(RV_{11}R')^{-1}RV_{11}, \quad B_{12} = B_{21}' \tag{39}$$

$$B_{21} = (RV_{11}R')^{-1}RV_{11}, \quad B_{22} = (RV_{11}R')^{-1} - I. \tag{40}$$

Remark 3. Notice that if the identifying conditions obtained in Theorem 1 do not hold, there exists a vector α (with m subvectors $\alpha_{\cdot i}$) such that

$$S_2\alpha = c \otimes e, \ \text{and thus,} \quad A_{(k)} \begin{pmatrix} 0 \\ \alpha \end{pmatrix} = 0, \tag{41}$$

because for at least one equation $S_{i2}\alpha_{\cdot i} = c_i \otimes e$, with c_i a scalar, so that $S_2\alpha = c \otimes e$, for an appropriate **non-zero** vector c. Such an α, in the absence of identification, is in the

[2] There is a misprint in the (2,2) diagonal block of the matrix at the end of Remark 2.27; it should be $(RV_{11}R')^{-1} - I$, **not** $I - (RV_{11}R')^{-1}$.

column nulll space of $S_2'(\Sigma_g \otimes \Omega_0)S_2$ as well as that of the matrix $\begin{pmatrix} A_{12} \\ A_{22} \end{pmatrix}$, as defined in Eq. (35); thus, we conclude that absent the identifying conditions of Theorem 1, **no estimator** exists at the kth iteration! Since, further, in Dhymes (1994b) it was shown that if such conditions hold, the iteration procedure can be carried out, the equivalence in the results of the present paper and those in Dhrymes (1994b) is evident.

3 Limiting Distribution and Inference

3.1 Limiting Distribution

The limiting distribution of $(\hat{\theta}', \hat{\lambda}')'$ is easily determined from Eq. (33). Rewrite it more conveniently, by substitution for y and writing $R\theta^0$ for r, as

$$\begin{pmatrix} \hat{\theta} - \theta^0 \\ \hat{\lambda} - \lambda^0 \end{pmatrix} = \begin{pmatrix} B_{11} \\ B_{21} \end{pmatrix} \frac{1}{T} \begin{pmatrix} S_1'(\Sigma_g \otimes X') \\ S_2'(\Sigma_g \otimes \hat{U}_{-1}') \end{pmatrix} \epsilon. \tag{42}$$

Consistency is then immediate. As for the limiting distribution, put

$$\xi_T = \frac{1}{\sqrt{T}} \begin{pmatrix} S_1'(\Sigma_g \otimes X') \\ S_2'(\Sigma_g \otimes \hat{U}_1') \end{pmatrix} \epsilon = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} S_1'(\Sigma_g \otimes x_{t\cdot}') \\ S_2'(\Sigma_g \otimes \hat{u}_{t-1\cdot}') \end{pmatrix} \epsilon_{t\cdot}'. \tag{43}$$

and note that the rightmost member is a sequence of martingale differences (MD) obeying the MD central limit theorem, see Dhrymes (1989), p. 337, so that

$$\xi_T \xrightarrow{\mathrm{d}} N(0, A^*), \quad A^* = \begin{bmatrix} S_1'(\Sigma_g \otimes M_{xx})S_1 & 0 \\ 0 & S_2'(\Sigma_g \otimes \Omega_0)S_2 \end{bmatrix} = \operatorname*{plim}_{T\to\infty} D'(\Sigma_g \otimes I_T)D, \tag{44}$$

where we have used the fact that $\Sigma_g \Sigma \Sigma_g = \Sigma_g$. Consequently, noting also that (under the null) $\lambda^0 = 0$, we have

$$\sqrt{T} \begin{pmatrix} \hat{\theta} - \theta^0 \\ \hat{\lambda} \end{pmatrix} \xrightarrow{\mathrm{d}} N(0, \Phi), \quad \Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \tag{45}$$

with

$$\Phi_{11} = V_{11}^* - V_{11}^* R'(RV_{11}^*R')^{-1}RV_{11}^*, \quad \Phi_{12} = \Phi_{21}' = 0, \tag{46}$$

$$\Phi_{22} = (RV_{11}^*R')^{-1} - I, \quad V_{11}^* = \operatorname*{plim}_{T\to\infty} V_{11} = (A^* + R'R)^{-1}. \tag{47}$$

We have therefore proved

Theorem 2. Consider the model specified in Eqs. (1) through (5) above; the MC estimator obtained by minimizing $\epsilon'(\Sigma_g \otimes I_T)\epsilon$, subject to the condition $R\theta - r = 0$ and implemented by Eqs. (33) through (36) obeys:

i. it is consisent;

ii. its limiting distribution is given by

$$\sqrt{T}\begin{pmatrix}\hat{\theta}-\theta^0\\ \hat{\lambda}\end{pmatrix} \xrightarrow{\mathrm{d}} N(0,\Phi), \tag{48}$$

where Φ is as defined in Eqs. (44) through (46).

Proof. Previous discussion.

3.2 Inference

Inference based on the limiting distribution is particularly simple. Thus, note that

i. Φ is **block** diagonal so that, in the limit, the distribution of $\hat{\theta}$ is independent of $\hat{\lambda}$. Thus, for example, if the restriction matrices S_1, S_2 contain more than what is necessary for the identification of the model, we can test these over-identifying restrictions by testing their associated Lagrange multipliers. Let L_i, $i = 1,2$ be a matrix such that it picks out of $\hat{\lambda}_i$, the over-identifying restrictions, put $L = \mathrm{diag}(L_1, L_2)$ and let it be desired to test (jointly) the hypotheses $L\lambda = 0$. A test of significance of such over-identifying restrictions can be carried out through the test statistic

$$\tau_1 = T(L\hat{\lambda})'(L\hat{\Phi}_{22}L')^{-1}(L\hat{\lambda}) \xrightarrow{\mathrm{d}} \chi^2_{\mathrm{rank}(L)}, \quad L = \mathrm{diag}(L_1, L_2). \tag{49}$$

It is clear that we are performing a **joint** test, i.e. that $L_1\lambda_1 = 0$ **and** $L_2\lambda_2 = 0$. In fact however, due to the diagonality of the covariance matrix, we have in Eq. (48) two individual tests carried out simultaneously and the degrees of freedom parameter applies to this **joint** test. But since the two estimators are asymptotically independent there is no compelling reason why a joint test should be carried out. If we **not interested in this joint test** but **only** in the individual tests $L_i\lambda_i = 0, i = 1, 2,$,then *mutatis mutandis* we can carry out such tests with the same expression as in Eq. (48), **but with** L **replaced by** $L = \mathrm{diag}(L_1, 0)$ and the degrees of freedom parameter replaced by $\mathrm{rank}(L_1)$, or L replaced by $L = \mathrm{diag}(0, L_2)$, and the degrees of freedom parameter replaced by $\mathrm{rank}(L_2)$.

ii. The covariance matrix of the limiting distribution of $\hat{\theta}$ is also diagonal, so that asymptotically $\hat{\beta}$ and $\hat{\gamma}$ are mutually independent. As in the discussion above if it is desired to test the (joint) hypothesis $C_1\beta = 0$ **and** $C_2\gamma = 0$ we can do so by the test statistic

$$\tau_2 = T(C\hat{\theta})'(C\hat{\Phi}_{11}C')^{-1}(C\hat{\theta}) \xrightarrow{\mathrm{d}} \chi^2_{\mathrm{rank}(C)}, \quad C = \mathrm{diag}(C_1, C_2). \tag{50}$$

The remarks made under i, are applicable here as well, so that if we are **only** interested in testing elements of β, we simply take, in Eq. (49), $C = \mathrm{diag}(C_1, 0)$ and replace the degrees of freedom parameter by $\mathrm{rank}(C_1)$, and, *mutatis mutandis* the same for γ.

REFERENCES

Dacunha-Castelle, D. and M. Duflo (1986), *Probability and Statistics*, vols. I and II, New York: Springer-Verlag

Dhrymes, P. J. (1989), *Topics in Advanced Econometrics; Probability Foundations*, New York: Springer-Verlag.

Dhrymes, P. J. (1994a), *Topics in Advanced Econometrics, vol. II, Linear and Nonlinear Simultaneous Equations*, New York: Springer-Verlag.

Dhrymes, P. J. (1994b), "Autoregressive Errors in Singular Systems of Equations", *Econometric Theory*, vol.10, pp.254-285.

Dhrymes, P. J. (2000), *Mathematics for Econometrics*, third edition, New York: Springer-Verlag.