Ann. Inst. Statist. Math.
Vol. 46, No. 4, 625–631 (1994)

CANONICAL CORRELATION AND REDUCTION OF MULTIPLE TIME SERIES

MOHSEN POURAHMADI

Division of Statistics and Statistical Consulting Laboratory, Northern Illinois University, DeKalb, IL 60115-2854, U.S.A.

(Received February 22, 1993; revised November 4, 1993)

Abstract. It is shown that a degenerate rank d-variate stationary time series can be reduced to a full rank time series of lower dimension via an orthogonal transformation T provided that ρ, the canonical correlation between past and future of the time series is strictly less than one. Procedures for estimation of rank of the multiple time series, T and testing $\rho = 1$ are outlined, the latter is related to testing the unit root hypothesis in ARMA models.

Key words and phrases: Rank of multiple time series, transformation, spectrum.

1. Introduction

The only source of redundancy in the multivariate data (when the observations are independent) is from inclusion of too many contemporaneously correlated variables in the study. In sharp contrast, multivariate time series data have two apparently distinct sources of redundancy. The first one is as above, while the second source is that of excessive temporal correlation. Notwithstanding this, most data reduction techniques used in multiple time series analysis seem to be complete analogues of those developed and used in multivariate statistics. That is, given data from a d-variate time series $\{X_t\}$, one seeks a (contemporaneous) linear transformation T, a $d \times d$ matrix so that the components of the transformed series $Y_t = TX_t$ have certain "desirable" properties. Two notable recent examples of such developments are (1) the class of MTV (multivariate time series variance component) models introduced by Kariya (1987, 1991) for which there exists an orthogonal transformation T so that components of $Y_t = TX_t$ are uncorrelated; (2) the class of vector ARMA models for which components of $Y_t = TX_t$ follow the scalar component model (SCM), Tiao and Tsay (1989). A notable exception is the spectral domain methods of principal component, factor analysis and canonical correlation, see Brillinger (1981), that lead to noncontemporaneous and noncausal transformations of the original multiple series.

Motivated by the foregoing discussion we may raise the following questions: For what classes of multivariate time series does there exist a transformation T

so that components of the transformed series $Y_t = TX_t$ have certain "desirable" properties? Given such a class, how can T be estimated? We address these questions for a particular situation. Namely for the class of degenerate rank d-variate time series $\{X_t\}$ we find conditions on $\{X_t\}$ so that there exists an orthogonal transformation T with

$$Y_t = TX_t = \begin{bmatrix} Z_t \\ 0 \end{bmatrix}, \tag{1.1}$$

and Z_t is an r-variate ($1 \leq r < d$) time series of full rank. More precisely, let $\{X_t\}$ be a mean zero, d-variate covariance stationary process with autocovariate function $\{\Gamma_k\}$ and innovation covariance matrix Σ, cf. Hannan (1970). For a matrix A, let $r(A)$ stand for its rank and $R(A)$ for its range or column space. A d-variate stationary process $\{X_t\}$ is said to be of *full rank* if $r(\Sigma) = d$, and of *degenerate rank* if $1 \leq r(\Sigma) < d$. Note that $r(\Sigma) = 0$ if and only if $\Sigma = 0$, such processes are called *singular processes*, and are excluded from further consideration in this paper.

We show in Section 2 that when ρ, the largest canonical correlation between past and future of the d-variate time series $\{X_t\}$ is *strictly less than one*, then there exists an orthogonal matrix T which reduces $\{X_t\}$ to an r-variate process of full rank as in (1.1) also it reduces Γ_0 to the form

$$\begin{gathered} T\Gamma_0 T' = \begin{pmatrix} \Gamma_{0,r} & 0 \\ 0 & 0 \end{pmatrix}, \\ r = r(\Gamma_0) = r(\Sigma), \end{gathered} \tag{1.2}$$

where $\Gamma_{0,r}$ is an $r \times r$ covariance matrix. We estimate r and T in Section 3 by performing a singular-value decomposition of the sample covariance matrix S_n of the data $X_1, \ldots, X_n$. This completely avoids the difficult problem of estimation of Σ, usually done by fitting parametric models to data which is inefficient in the presence of degeneracy because of redundant parameters, cf. Tiao and Tsay (1989).

2. Degenerate rank multivariate time series

In this section we review some basic concepts regarding the structure of d-variate stationary processes with particular attention to degenerate rank processes. The equivalence of time-domain and spectral-domain analysis of d-variate time series is used to motivate the class of *constant range* processes.

Consider a d-variate mean zero stationary process $X_t = (X_{1t}, X_{2t}, \ldots, X_{dt})'$, with autocovariance function $\Gamma_k = \mathrm{Cov}(X_{t+k}, X_t)$, $k = 0, \pm 1, \ldots$ and $d \times d$ spectral distribution $F(\lambda)$, let

$$dF(\lambda) = dF_{ac}(\lambda) + dF_s(\lambda) = f(\lambda)d\lambda + dF_s(\lambda), \tag{2.1}$$

be the unique Cramer-Lebesgue decomposition of F into its absolutely continuous (F_{ac}) and singular (F_s) parts. For more information on these and related concepts, see Hannan (1970) and Masani (1966). We use $\Gamma_{k,X}$, F_X, etc. when working with more than one multivariate process. Corresponding to (2.1) in the spectral domain, there is the Wold decomposition in the time domain which writes any *nondeterministic process* $\{X_t\}$ as sum of its purely *nondeterministic* and *deterministic* parts

$$X_t = U_t + V_t = \sum_{k=0}^{\infty} \Psi_k \epsilon_{t-k} + V_t, \tag{2.2}$$

where $\{\epsilon_t\}$ is a d-variate white noise process with $\Sigma = \text{cov}(\epsilon_t)$ and $\{V_t\}$ is a deterministic process uncorrelated with $\{U_t\}$, and $\sum_{k=0}^{\infty} \text{tr}\, \Psi_k \Psi_k' < \infty$. The *innovation covariance matrix* Σ carries considerable information about the dependence of components of $\{X_t\}$ and it is related to the spectral density $f(\lambda)$ of the process. Note that in general, $r[f(\lambda)]$ is a function of the frequency λ. Next we motivate the class of processes whose spectral densities have *constant rank*.

The implicit and well-accepted equivalence between time-domain and spectral-domain analysis of time series is usually justified by the fact that the autocovariance function $\{\Gamma_k\}$ and the spectral distribution F are Fourier pairs. Although this mathematical justification points to the right direction, the real reason for statistical equivalence of the two approaches to time series analysis is much deeper and relies on both the Cramer-Lebesgue decomposition (2.1) of F, and the Wold decomposition (2.2) of the process $\{X_t\}$ itself. Note that since $\{U_t\}$ and $\{V_t\}$ are uncorrelated, (2.2) gives an alternative decomposition of F:

$$dF = dF_U + dF_V = \Psi \Sigma \Psi' d\lambda + dF_V = f_U d\lambda + dF_V \tag{2.3}$$

where F_U and F_V are spectral distributions of $\{U_t\}$ and $\{V_t\}$, respectively, and $\Psi(\lambda) = \sum_{k=0}^{\infty} \Psi_k e^{ik\lambda}$, is the *transfer function* of the MA(∞) part of (2.2). Now, having two apparently different decompositions of the spectral distribution F of $\{X_t\}$, we need to know whether these decompositions are the same. From (2.1) and (2.3), our question is whether

$$f = f_U \ ? \tag{2.4}$$

Evidently, a positive answer to this question would provide a bona fide justification for the statistical equivalence of the two approaches to time series analysis.

The question (2.4) makes sense and is relevant even in the univariate case ($d = 1$). Note that from (2.2) f_U is the spectral density function of an MA(∞) and thus purely nondeterministic (regular) process. Therefore, for (2.4) to hold $f(\lambda)$ must satisfy $\int_{-\pi}^{+\pi} \log f(\lambda) d\lambda > -\infty$ which can be interpreted as $r[f(\lambda)] = 1$ for almost all λ, i.e. rank of $f(\lambda)$ does not depend on λ. In 1963, Robertson showed that the above observation regarding univariate processes can be generalized to d-variate processes, cf. Masani (1966).

THEOREM 2.1. (Robertson) *Let $\{X_t\}$ be a nondeterministic d-variate stationary process with $r(\Sigma) = r$. Then, $f = f_U$ if and only if $r[f(\lambda)] = r$, for almost all λ.*

This important result indicates the relevance to time series analysis of the class of processes whose spectral density matrices have constant ranks. It also indicates the relationship between ranks of Σ and $f(\lambda)$ for such degenerate rank processes. Unfortunately, this class is too large for our purposes in this paper, we give two simple examples of bivariate stationary processes where their spectral density matrices have constant ranks and yet (1.2) is not satisfied, indicating the need for imposing further restrictions on the density matrices than the constancy of their ranks. Let $\{\epsilon_t\}$ be a univariate mean zero white noise process with unit variance, that is $\text{cov}(\epsilon_t, \epsilon_s) = \delta_{t,s}$, the Kronecker's delta.

Example 2.1. (a) Consider the bivariate stationary process $\{\boldsymbol{X}_t\}$:

$$\boldsymbol{X}_t = (\epsilon_t, \epsilon_{t-1})', \quad t = 0, \pm 1, \ldots, \tag{2.5}$$

with autocovariances

$$\Gamma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Gamma_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \Gamma_{-1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \Gamma_{\pm k} = 0, \quad k \geq 2,$$

and innovation covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that $r(\Gamma_0) \neq r(\Sigma)$ and (1.2) is not satisfied even though

$$f(\lambda) = \begin{pmatrix} 1 & e^{-i\lambda} \\ e^{i\lambda} & 1 \end{pmatrix},$$

has constant rank one, for all λ. The failure of (1.2) can be explained on two accounts. First, the range or column space of $f(\lambda)$ depends on the frequency λ. Second, ρ the largest canonical correlation between the past and future of $\{X_t\}$, which in this case, is the same as the largest canonical correlation between the sets of random variables $P = \overline{\text{span}}\{\epsilon_t; t \geq 0\}$ and $F = \overline{\text{span}}\{\epsilon_t; t \leq 0\}$ is equal to one. For the sake of comparison with the next example, we rewrite (2.5) as an MA(1) process:

$$\boldsymbol{X}_t = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{bmatrix} \epsilon_{t-1} \\ \epsilon_{t-2} \end{bmatrix}.$$

(b) Let $\{X_t\}$ be a univariate AR(2) process defined by

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t,$$

and $\boldsymbol{X}_t$ be the bivariate process $\boldsymbol{X}_t = (X_t, X_{t-1})'$ with a bivariate AR(1) representation

$$\boldsymbol{X}_t = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} \boldsymbol{X}_{t-1} + \begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}.$$

This process has many features of the process defined in (2.5), in particular $\rho = 1$.

In view of these examples and our discussion so far, we need to further restrict the class of d-variate processes whose spectral density matrices have constant rank. A d-variate stationary process $\{X_t\}$ of rank r, $1 \leq r \leq d$, is said to have *constant range*, if the range (column space) of $f(\lambda)$ does not depend on λ, i.e. $R[f(\lambda)] =$ constant, for almost all λ. Such processes arise in theoretical problems dealing with invertibility of the MA(∞) part of (2.2) and autoregressive representation of degenerate rank processes. It is known that when the canonical correlation between past and future of $\{X_t\}$ is less than one, then $f(\lambda)$ has constant range, cf. Miamee and Pourahmadi (1987). For ease of reference, we state this result more precisely.

THEOREM 2.2. *Let $\{X_t\}$ be a d-variate stationary process with the spectral distribution F decomposed as in* (2.2). *If $\rho < 1$, then*

(a) $F_s = 0$, *i.e.* $dF = f d\lambda$.

(b) $R[f(\lambda)] =$ *constant, for almost all frequencies* λ.

In the rest of this section, we deal only with processes for which $\rho < 1$ so that their spectral densities have constant ranges. Since $R[f(\lambda)]$ is a constant subspace of R^d, let us denote this subspace by $R(f)$. By using the standard argument for relating two orthonormal bases of R^d, cf. Miamee and Salehi (1979), it follows that there exists an orthogonal matrix T such that

$$Tf(\lambda)T' = \begin{bmatrix} f_Z(\lambda) & 0 \\ 0 & 0 \end{bmatrix},$$

where $f_Z(\lambda)$ is an $r \times r$ spectral density matrix of a full-rank r-variate process. Now, by letting $Y_t = TX_t$ it follows that

$$\begin{aligned} \operatorname{cov}(Y_{t+k}, Y_t) &= T\operatorname{cov}(X_{t+k}, X_t)T' = T\int_{-\pi}^{\pi} e^{-ik\lambda} f(\lambda) d\lambda T' \\ &= \int_{-\pi}^{\pi} e^{-ik\lambda} Tf(\lambda)T' d\lambda = \int_{-\pi}^{\pi} e^{-ik\lambda} \begin{bmatrix} f_Z(\lambda) & 0 \\ 0 & 0 \end{bmatrix} d\lambda \\ &= \begin{bmatrix} \Gamma_{k,Z} & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

showing that the last d-r components of Y_t are zero, as stated in (1.1). Furthermore, this calculation shows that for any integer k,

$$T\Gamma_{k,X}T' = \begin{bmatrix} \Gamma_{k,Z} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad T\Sigma_X T' = \begin{bmatrix} \Gamma_Z & 0 \\ 0 & 0 \end{bmatrix}.$$

3. Estimation of r and testing $\rho = 1$

Reduction of Γ_0 in (1.2) provides some insight into the problem of estimation of r and T, since it identifies r as the rank of Γ_0 and T as the orthogonal matrix that reduces Γ_0 to a block diagonal form with the leading block being of full rank r and the other blocks are null matrices. Having time series data $X_1, \ldots, X_n$ from a d-variate stationary process $\{X_t\}$, a natural estimator of Γ_0 is $S_n = X(I_n - n^{-1}e_n e_n')X'$ where $X = (X_1, \ldots, X_n)$ is the $d \times n$ matrix of data and $e_n = (1, \ldots, 1)'$, its rank gives an estimator of the rank of Γ_0 as well as Σ, see (1.2). Let $\hat{r}_n = r(S_n)$, in the following we give conditions under which $\{\hat{r}_n\}$ is a strongly consistent estimator of r, see Okamoto (1973) for a proof of the following theorem.

THEOREM 3.1. *Let $\{X_t\}$ be a d-variate normal stationary process with innovation covariance matrix Σ of rank r and $\rho < 1$. Then, for $n > d$ we have,*
(a) $P\{\hat{r}_n = r\} = 1$.
(b) $P\{$*non zero eigenvalues of* S_n *are distinct*$\} = 1$.

Our procedure for estimation of r can be implemented and a reduction of dimension of $\{X_t\}$ may result if the process is of degenerate rank and $\rho < 1$. To render it fully satisfactory, one must infer from the time series data $X_1, \ldots, X_n$ as whether $\rho < 1$ or $\rho = 1$. Computation of canonical correlations and test of $\rho = 0$ for multiple time series are discussed in Tiao and Tsay ((1989), Section 4). However, the problem of testing $\rho = 1$ is a challenging problem and is related to the problem of testing unit roots for ARMA models. For information on the relationship between canonical correlations and location of roots of characteristic polynomials of ARMA models, see Hannan and Poskitt (1988). For a univariate AR(1) model, $X_t = \phi X_{t-1} + \epsilon_t$, it can be shown that $\rho = |\phi|$. For testing $\rho = \phi = 1$, Chan and Wei (1987), have proposed a double-array setting to reparametrize AR(1) models by

$$X_t = \left(1 - \frac{\gamma}{n}\right) X_{t-1} + \epsilon_t,$$

where γ is a fixed constant. They show that the limiting distribution of the least squares estimator of ϕ can be expressed as certain functional of Ornstein-Uhlenbeck process.

REFERENCES

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*, Holt, Rinehart and Winston, New York.

Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes, *Ann. Statist.*, **15**, 1050–1063.

Hannan, E. J. (1970). *Multiple Time Series*, Wiley, New York.

Hannan, E. J. and Poskitt, D. S. (1988). Unit canonical correlations between future and past, *Ann. Statist.*, **16**, 784–790.

Kariya, T. (1987). MTV model and its application to prediction of stock prices, *Proceedings of 2nd International Conference in Statistics* (ed. T. Pukkila), 161–176.

Kariya, T. (1991). An analysis of Yen-Dollar exchange rates and prediction of stock prices via MTV model, Manuscript, Hitotsubashi University, Tokyo.

Masani, P. R. (1966). Recent trends in multivariate prediction theory, *Multivariate Analysis* (ed. P. R. Krishnaiah), 351–382, Academic Press, New York.

Miamee, A. G. and Pourahmadi, M. (1987). Degenerate multivariate stationary processes: basicity, past and future, and autoregressive representation, *Sankhyā Ser. A*, **49**, 316–334.

Miamee, A. G. and Salehi, H. (1979). On the bilateral prediction error matrix of a multivariate stationary stochastic processes, *SIAM J. Math. Anal.*, **10**, 247–252.

Okamoto, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample, *Ann. Statist.*, **1**, 763–765.

Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate times series, *J. Roy. Statist. Soc. Ser. B*, **51**, 157–213.