Improved Estimates for Multivariate Complex–Normal Regression with Application to Analysis of Linear Time-Invariant Relationships

J. LILLESTØL

University of California, Berkeley and
The Norwegian School of Economics and Business Administration, Bergen, Norway

Communicated by E. J. Hannan

The objective of this paper is the estimation of linear time-invariant relationships for a stationary vector-valued time series using the Finite Fourier Transform as the basic statistic. Since this is asymptotically complex–Normal we are led to consider models of multivariate complex–Normal regression. We propose estimates of regression matrices in the tradition of Stein (shrunken estimates) which improve upon the usual estimates. Some experience with simulated time series is reported.

1. INTRODUCTION

To almost any result in real multivariate analysis there is a corresponding result for complex–Normal variates. The complex–Normal and the related complex–Wishart distribution were first explored by Goodman [6]. A survey of recent developments on complex multivariate distributions is given by Krishnaiah [7]. The multivariate complex–Normal distribution is of main importance to time-series analysts as it occurs as the limiting distribution of the finite Fourier transform of a stationary vector time series under suitable mixing conditions. A survey of this theory is given by Brillinger [2, 3].

The present author (Lillestøl [8]) has extended a large body of the theory in the tradition of Stein (shrunken estimates) on simultaneous estimation in Normal models to the complex–Normal case, and studied its potential application to time-series analysis. In Section 2 of this paper we report alternative estimates for multivariate complex–Normal regression. It extends the theory of Stein [11], Sclove [9, 10], and Baranchik [1] to the complex–Normal case. The results and

their proofs are similar and we only give a brief exposition. In Section 3 we consider the possible application of these results to the time-series problem of estimating linear time-invariant relationships. Thereafter we report some experience with the procedure on simulated time series.

2. THEORY

Consider independent $(r + 1)$-dimensional complex column vectors

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{X}_k \\ Y_k \end{bmatrix}, \qquad k = 1, 2, ..., n, \tag{2.1}$$

each distributed according to the complex–Normal distribution with zero mean

$$N^c_{r+1} \left(\begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \Phi_{XX} & \Phi_{XY} \\ \Phi_{YX} & \Phi_{YY} \end{bmatrix} \right) \tag{2.2}$$

so that

$$E(Y_k \mid \mathbf{X}_k) = \alpha \mathbf{X}_k \tag{2.3}$$

where α is the row vector given by

$$\alpha = \Phi_{YX} \Phi_{XX}^{-1}. \tag{2.4}$$

The corresponding conditional covariance is $\Psi = \Phi_{YY} - \Phi_{YX} \Phi_{XX}^{-1} \Phi_{XY}$. Consider the problem of estimating the complex regression vector α. As our loss function we take

$$L(\alpha, \hat{\alpha}) = \Psi^{-1}(\hat{\alpha} - \alpha) \, \Phi_{XX} (\hat{\alpha} - \alpha)^* \tag{2.5}$$

which is admittedly partly chosen for its nice invariance properties. (In general we denote $\mathbf{A}^* = \bar{\mathbf{A}}'$, that is $\mathbf{A}^*$ is the transposed conjugate of $\mathbf{A}$.)

Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)$, $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$, and define matrices $T = \mathbf{Y}\mathbf{Y}^* = \sum_k Y_k Y_k^*$, $\mathbf{U} = \mathbf{Y}\mathbf{X}^* = \sum_k Y_k \mathbf{X}_k^*$, and $\mathbf{V} = \mathbf{X}\mathbf{X}^* = \sum_k \mathbf{X}_k \mathbf{X}_k^*$ of dimension (1×1), $(1 \times r)$, and $(r \times r)$, respectively. The maximum likelihood estimates of the crucial parameters of the model are then given by

$$\hat{\alpha}^{(0)} = \mathbf{U}\mathbf{V}^{-1}, \qquad \hat{\Psi} = (1/n)W, \tag{2.6}$$

where $W = T - \mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*$. Among the properties of $\hat{\alpha}^{(0)}$ as an estimate of α we mention: It is unbiased (seen by conditioning on $\mathbf{X}$). With the given loss function it is fully invariant and minimax. However, in the case $r \geqslant 2$ and sufficiently large n (in fact $n \geqslant r + 1$) it is inadmissible in the sense that, with the given loss function, there exist other estimates with uniformly smaller risk (expected loss).

Below we will present a class of estimates $\hat{\alpha}^{(c)}$ having some members that dominate the common estimate $\hat{\alpha}^{(0)}$ and may suggest useful alternative procedures in the time-series context as well. The work done for the real case (see the references given in Section 1) are mainly focused on the situation where the means are unknown. However, contrary to this, it is the zero mean situation which is of practical importance in the complex case, when we have applications to time series in mind.

THEOREM 1. *The risk (expected loss) of the common estimate* $\hat{\alpha}^{(0)} = \mathbf{UV}^{-1}$ *is* $R(\alpha, \hat{\alpha}^{(0)}) = r(n - r)^{-1}$ *provided* $n \geqslant r + 1$.

Proof. A sufficient statistic for the model is $(\mathbf{UV}^{-1}, W, \mathbf{V})$. By arguments similar to the real case, and using well-known properties of the complex–Normal distribution it follows that $W \sim \frac{1}{2}\Psi\chi^2_{2(n-r)}$ independent of $(\mathbf{UV}^{-1}, \mathbf{V})$ where $\mathbf{V} \sim W_r^c(n, \Phi_{XX})$ and, conditionally given $\mathbf{V}$, $\mathbf{UV}^{-1} \sim N_r^c(\alpha, \Psi \cdot \mathbf{V}^{-1})$. Here $\sim$ denotes "is distributed as," χ_ν^2 denotes a chi-square-variate with ν degrees of freedom, and $W_r^c(n, \Phi)$ denotes a complex–Wishart variate of dimension r with covariance matrix Φ and n degrees of freedom. As in the real case one can use group-theoretic considerations together with invariance to show that in order to compute the risk of the common estimate $\hat{\alpha}^{(0)}$ it suffices to do it for $\Psi = 1$ and $\Phi_{XX} = \mathbf{I}$. Let E_0 denote expectation computed for this case. We then get $R(\hat{\alpha}, \hat{\alpha}^{(0)}) = EL(\alpha, \hat{\alpha}^{(0)}) = E\Psi^{-1}(\hat{\alpha}^{(0)} - \alpha) \Phi_{XX}(\hat{\alpha}^{(0)} - \alpha)^* = E_0(\hat{\alpha}^{(0)} - \alpha)(\hat{\alpha}^{(0)} - \alpha)^* = \mathrm{tr}\, E_0(\hat{\alpha}^{(0)} - \alpha)^*(\hat{\alpha}^{(0)} - \alpha) = \mathrm{tr}\, E_0(\mathbf{UV}^{-1} - \alpha)^* (\mathbf{UV}^{-1} - \alpha) = \mathrm{tr}\, E_0\mathbf{V}^{-1}$, where tr denotes the usual trace operator of a matrix. The last step follows by conditioning on $\mathbf{X}$ and using that for the given model $E(\mathbf{U} \mid \mathbf{X}) = \alpha\mathbf{V}$ and $E(\mathbf{U}^*\mathbf{U} \mid \mathbf{X}) = \Psi\mathbf{V} + (\alpha\mathbf{V})^*\alpha\mathbf{V}$. Since $\mathbf{V} \sim W_r^c(n, \Phi_{XX})$ it follows that (see Brillinger [3, p. 336]) $E\mathbf{V}^{-1} = (n - r)^{-1}\Phi_{XX}^{-1}$ provided $n \geqslant r + 1$ and the theorem follows. ∎

THEOREM 2. *For the class of estimates*

$$\hat{\alpha}^{(c)} = \left(1 - c\,\frac{W}{\mathbf{UV}^{-1}\mathbf{U}^*}\right)\mathbf{UV}^{-1} \tag{2.7}$$

we have in the case $r \geqslant 2$ *and* $n \geqslant r + 1$ *that*

$$R(\alpha, \hat{\alpha}^{(c)}) < R(\alpha, \hat{\alpha}^{(0)}) \tag{2.8}$$

provided the constant c *is chosen so that*

$$0 < c < 2(r - 1)(n - r + 1)^{-1}. \tag{2.9}$$

Remark. The estimates can be alternatively written as

$$\hat{\alpha}^{(c)} = \left(1 - c\,\frac{1 - R^2}{R^2}\right)\hat{\alpha}^{(0)}, \tag{2.10}$$

where $R^2 = \mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*/T$ is an estimate of the multiple coherence coefficient (the complex version of multiple correlation).

Proof. The proof is similar to the one given by Baranchik [1] for the real case. Again group-theoretic considerations and invariance will show that it suffices to compute the risk for the simpler case when $\Psi = 1$ and $\mathbf{\Phi}_{XX} = \mathbf{I}$, that is, $R(\alpha, \hat{\alpha}^{(c)}) = EL(\alpha, \hat{\alpha}^{(c)}) = E_0 L(\alpha, \hat{\alpha}^{(c)}) = E_0 \| \hat{\alpha}^{(c)} - \alpha \|^2$ where, in general, $\| \alpha \|^2 = \alpha \alpha^*$ and E_0 denotes expectation calculated for the simpler case. It is easily checked that

$$\| \hat{\alpha}^{(0)} - \alpha \|^2 - \| \hat{\alpha}^{(c)} - \alpha \|^2$$

$$= 2c \frac{W}{\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*} (\mathbf{U}\mathbf{V}^{-2}\mathbf{U}^* - \mathbf{U}\mathbf{V}^{-1}\alpha^*) - c^2 \frac{W^2}{(\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*)^2} \mathbf{U}\mathbf{V}^{-2}\mathbf{U}^* \quad (2.11)$$

and our objective is to prove that the expectation of the right-hand side is nonnegative and positive at least for some α. Using the fact that $W \sim \frac{1}{2}\chi^2_{2(n-r)}$ independent of $(\mathbf{U}\mathbf{V}^{-1}, \mathbf{V})$, and consequently $EW = n - r$, $EW^2 = (n - r + 1) \cdot (n - r)$, we see that the improvement of $\hat{\alpha}^{(c)}$ over $\hat{\alpha}^{(0)}$ can be written

$$\Delta(\hat{\alpha}^{(c)} : \hat{\alpha}^{(0)}) = E_0[\| \hat{\alpha}^{(0)} - \alpha \|^2 - \| \hat{\alpha}^{(c)} - \alpha \|^2]$$

$$= 2c(n - r)E_0 \left[\frac{\mathbf{U}\mathbf{V}^{-2}\mathbf{U}^*}{\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*} - \frac{\mathbf{U}\mathbf{V}^{-1}\alpha^*}{\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*} \right]$$

$$- c^2(n - r + 1)(n - r)E_0 \left[\frac{\mathbf{U}\mathbf{V}^{-2}\mathbf{U}^*}{(\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*)^2} \right]. \quad (2.12)$$

In order to be able to reduce this expression we need a computational lemma.

LEMMA 1.

$$E_0 \left(\frac{\mathbf{U}\mathbf{V}^{-1}\alpha^*}{\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*} \right) = d(\| \alpha \|, n) \sum_{k=0}^{\infty} a_k(\| \alpha \|, n, r)\gamma^k, \quad (2.13)$$

$$E_0 \left(\frac{\mathbf{U}\mathbf{V}^{-2}\mathbf{U}^*}{\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*} \right) = d(\| \alpha \|, n) \sum_{k=0}^{\infty} b_k(\| \alpha \|, n, r)\gamma^k, \quad (2.14)$$

$$E_0 \left(\frac{\mathbf{U}\mathbf{V}^{-2}\mathbf{U}^*}{(\mathbf{U}\mathbf{V}^{-1}\mathbf{U}^*)^2} \right) = d(\| \alpha \|, n) \sum_{k=0}^{\infty} c_k(\| \alpha \|, n, r)\gamma^k, \quad (2.15)$$

where $\gamma = \| \alpha \|^2 (1 + \| \alpha \|^2)^{-2}$ *and*

$$d(\| \alpha \|, n) = \frac{1}{\Gamma(n)} \cdot \frac{1}{(1 + \| \alpha \|^2)^{n-1}},$$

$$a_k(\| \alpha \|, n, r) = \frac{k\Gamma(n + k - 1)}{k!(r + k - 1)},$$

$$b_k(\| \alpha \|, n, r) = \frac{\Gamma(n + k - 1)}{k!(r + k)} \left[\frac{n - 1}{n - r} + k + \frac{(n + k - 1)(r - 1)}{(1 + \| \alpha \|^2)(n - r)} \right],$$

$$c_k(\| \alpha \|, n, r) = (r + k - 1)^{-1} b_k(\| \alpha \|, n, r).$$

A key to the proof of this lemma is provided by the fact that in the case of $\Psi = 1$ we have that, given $\mathbf{V}$, $\mathbf{UV}^{-1/2} \sim N_r^c(\alpha \mathbf{V}^{1/2}, \mathbf{I})$. One can first calculate the conditional expectations of the variates given $\mathbf{V}$ and then undo the conditioning. This calculation makes use of the complex versions of basic identities for normal variates similar to Lemma 1 of Baranchik [1]. Details are given in Lillestøl [8]. Using Lemma 1 we get the following expression for the improvement in the risk:

$$\Delta(\hat{\alpha}^{(c)} : \hat{\alpha}^{(0)}) = c(n - r)d \sum_{k=0}^{\infty} (2(b_k - a_k) - c(n - r + 1)c_k)\gamma^k. \quad (2.16)$$

Here we have suppressed the arguments of the functions a_k, b_k, c_k, and d. We see that this is positive provided

$$0 < c < 2(b_k - a_k) c_k^{-1}(n - r + 1)^{-1} = u_k. \quad (2.17)$$

Using the expressions given in Lemma 1 we obtain

$$u_k = 2 \left\{ (r + k - 1) - k(r + k) \left[\frac{n - 1}{n - r} + k + \frac{(n + k - 1)(r - 1)}{(1 + \| \alpha \|^2)(n - r)} \right]^{-1} \right\}. \quad (2.18)$$

A lower bound on u_k not depending on k and $\| \alpha \|$ is easily obtained provided $r \geqslant 2$. We have

$$u_k \geqslant 2\{(r + k - 1) - k(r + k)[(n - 1)/(n - r) + k]^{-1}\}$$
$$= 2(r - 1)(n + k - 1)[(n - r + 1)(n + k - 1) - (n - 1)(n - r) - k]^{-1}$$
$$> 2(r - 1)(n - r + 1)^{-1}.$$

We can therefore conclude that condition (2.17) holds at least when $0 < c < 2(r - 1)(n - r + 1)^{-1}$ in the case $r \geqslant 2$. We have assumed throughout that $n \geqslant r + 1$. This completes the proof of Theorem 2. ∎

It is instructive to examine the possible improvement in some special cases:

For $\alpha = 0$ the improvement in the risk becomes

$$\Delta(\hat{\alpha}^{(c)} : \hat{\alpha}^{(0)}) = 2c - c^2(r - 1)^{-1}(n - r + 1). \quad (2.19)$$

This is maximized for $c = c(0) = (r - 1)(n - r + 1)^{-1}$ corresponding to an improvement of

$$\Delta(\hat{\alpha}^{(c)} : \hat{\alpha}^{(0)}) = (r - 1)(n - r + 1)^{-1}. \quad (2.20)$$

For $\| \alpha \|^2 \to \infty$ it can be shown that

$$\Delta(\hat{\alpha}^{(c)} : \hat{\alpha}^{(0)}) = \| \alpha \|^{-2} \left[2c \frac{r - 1}{n - 2} - c^2 \frac{(n - r)(n - r + 1)}{(n - 1)(n - 2)} \right] + o(\| \alpha \|^{-2}). \quad (2.21)$$

Neglecting the lower-order term and maximizing with respect to c yields $c(\infty) = (n-1)(r-1)(n-r)^{-1}(n-r+1)^{-1}$. For this choice the improvement becomes

$$\Delta(\hat{\alpha}^{(c)} : \hat{\alpha}^{(0)}) = \| \alpha \|^{-2} \frac{(n-1)(r-1)^2}{(n-2)(n-r)(n-r+1)} + o(\| \alpha \|^{-2}). \quad (2.22)$$

It is seen that for any given $\| \alpha \|$ the optimal choice of c is one-half the maximum value satisfying the inequality. Call this $c(\| \alpha \|)$. We have obtained $c(\| \alpha \|) \geqslant (r-1)(n-r+1)^{-1} = c(0)$ with strict inequality for some $\| \alpha \|$. The evaluation of the improvement for moderate $\| \alpha \|$ is hampered by technical difficulties and it seems reasonable to judge the performance by doing some simulations. In practice one is most often interested in doing well for small $\| \alpha \|$, and $c(0)$ should therefore be a good choice for an all purpose estimate. Note also that this corresponds to the $c(\| \alpha \|)$ modifying the common estimate the least. We see that the modifying factor may become negative (with small probability). This seems to be an undesirable property of the estimate, and the estimate we are suggesting as a competitor to the usual estimate for practical use is therefore

$$\hat{\alpha}^{(+)} = \left(1 - \frac{r-1}{n-r+1} \cdot \frac{W}{UV^{-1}U^*}\right)^+ \cdot \hat{\alpha}^{(0)}, \quad (2.23)$$

where $a^+ = a$ for $a > 0$ and $a^+ = 0$ for $a \leqslant 0$. Using arguments similar to Stein [12] one can verify that $\hat{\alpha}^{(+)}$ further improves the risk in terms of the given loss function.

Estimates of the type presented here as an alternative are often referred to as shrunken estimates, obtained by shrinking the common estimate toward zero to an extent determined by the observations. There are numerous papers which discuss the application of such estimates in real multivariate analysis.

3. Applications to Time Series

Consider an $(r+1)$-variate stationary series

$$\mathbf{z}(t) = \begin{bmatrix} \mathbf{x}(t) \\ y(t) \end{bmatrix}, \quad t = 0, \pm 1, \pm 2,..., \quad (3.1)$$

with spectral density matrix given by

$$\mathbf{f}_{zz}(\lambda) = \begin{bmatrix} \mathbf{f}_{xx}(\lambda) & \mathbf{f}_{xy}(\lambda) \\ \mathbf{f}_{yx}(\lambda) & f_{yy}(\lambda) \end{bmatrix}. \quad (3.2)$$

Consider the problem of determining a linear time-invariant expression

$$\mu + \sum_{u=-\infty}^{\infty} \mathbf{a}(t-u)\,\mathbf{x}(u) \quad (3.3)$$

so that $E[y(t) - \mu - \sum_u \mathbf{a}(t - u) \mathbf{x}(u)]^2$ is minimized. Here μ is a scalar and $\{\mathbf{a}(u)\}$ is a $(1 \times r)$-vector-valued function of filter coefficients. The optimal choice is given by

$$\mu = Ey(t) - \left(\sum \mathbf{a}(u)\right) E\mathbf{x}(t), \tag{3.4}$$

$$\mathbf{a}(u) = (2\pi)^{-1} \int_0^{2\pi} \mathbf{A}(\alpha) \exp(iu\alpha) \, d\alpha, \tag{3.5}$$

where

$$\mathbf{A}(\lambda) = \mathbf{f}_{yx}(\lambda) \, \mathbf{f}_{xx}(\lambda)^{-1} \tag{3.6}$$

is the transfer function of the filter (see Brillinger [3]).

Our application of the theory in Section 2 will be to the estimation of $\mathbf{A}(\lambda)$ for a given $\lambda \not\equiv 0 \pmod{\pi}$: Suppose the series $\mathbf{z}(t)$ is observed at instants $t = 0, 1, 2,..., T - 1$ so that the finite Fourier transforms

$$\mathbf{d}_z^T(\lambda) = \begin{bmatrix} \mathbf{d}_x^T(\lambda) \\ d_y^T(\lambda) \end{bmatrix} = \sum_{t=0}^{T-1} \exp(-i\lambda t) \begin{bmatrix} \mathbf{x}(t) \\ y(t) \end{bmatrix} \tag{3.7}$$

are available for analysis. Following the suggestion of Goodman [6] we use the complex–Normal distribution

$$N_{r+1}^c \left(\mathbf{0}, \, 2\pi T \begin{bmatrix} \mathbf{f}_{xx}(\lambda) & \mathbf{f}_{xy}(\lambda) \\ \mathbf{f}_{yx}(\lambda) & f_{yy}(\lambda) \end{bmatrix} \right) \tag{3.8}$$

as an approximate distribution of $\mathbf{d}_z^T(\lambda)$ at $\lambda \not\equiv 0 \pmod{\pi}$ for large T. This suggests that

$$E(d_y^T(\lambda) \mid \mathbf{d}_x^T(\lambda)) \approx \mathbf{f}_{yx}(\lambda) \, \mathbf{f}_{xx}(\lambda)^{-1} \mathbf{d}_x^T(\lambda) = \mathbf{A}(\lambda) \, \mathbf{d}_x^T(\lambda) \tag{3.9}$$

so that the transfer function $\mathbf{A}(\lambda)$ may be interpreted as a complex regression coefficient vector. In order to produce estimates with better stability properties we follow the usual procedure to compute the finite Fourier transform at n frequencies $\lambda_1, \lambda_2,..., \lambda_n$ in the neighborhood of λ. Under suitable mixing conditions of the stationary series $\mathbf{z}(t)$ (dependencies have to die off sufficiently rapidly as the time span increases) it has been shown by various authors that $\mathbf{d}_z^T(\lambda_k)$, $k = 1, 2,..., n$ behave asymptotically as independent complex–Normal variates with the covariance structure given above. For precise statements see Brillinger [3] who proves a result of this kind by putting conditions on the joint cumulants of the series and considering neighboring frequencies of form $\lambda_k = \lambda_k(T) = 2\pi s_k(T)/T \to \lambda$ as $T \to \infty$ where $s_k(T)$ are integers and $2\lambda_j(T)$, $\lambda_j(T) \pm \lambda_k(T) \not\equiv 0 \pmod{2\pi}$ for $j \neq k$.

In order to make use of the theory of Section 2 we make the identification

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{X}_k \\ Y_k \end{bmatrix} = (2\pi T)^{-1/2} \begin{bmatrix} \mathbf{d}_x^T(\lambda_k) \\ d_y^T(\lambda_k) \end{bmatrix}, \qquad k = 1, 2,..., n \tag{3.10}$$

and, furthermore,

$$\alpha = \mathbf{A}(\lambda), \qquad \Phi = \mathbf{f}_{zz}(\lambda). \tag{3.4}$$

The common estimate of $\mathbf{f}_{zz}(\lambda)$ is

$$\mathbf{f}_{zz}^T(\lambda) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{d}_z^T(\lambda_k)\, \mathbf{d}_z^T(\lambda_k)^* \tag{3.12}$$

with partition similar to that of $\mathbf{f}_{zz}(\lambda)$. The common estimate of $\mathbf{A}(\lambda)$ is then

$$\mathbf{A}^T(\lambda) = \mathbf{f}_{yx}^T(\lambda)\, \mathbf{f}_{xx}^T(\lambda)^{-1}. \tag{3.13}$$

The complex–Normal theory of Section 2 suggests the following alternative estimate:

$$\mathbf{A}^{(+)}(\lambda) = \left(1 - \frac{r-1}{n-r+1} \cdot \frac{g^T(\lambda)}{\mathbf{f}_{yx}^T(\lambda)\, \mathbf{f}_{xx}^T(\lambda)^{-1}\, \mathbf{f}_{xy}^T(\lambda)}\right)^+ \mathbf{A}^T(\lambda), \tag{3.14}$$

where $g^T(\lambda) = f_{yy}^T(\lambda) - \mathbf{f}_{yx}^T(\lambda)\, \mathbf{f}_{xx}^T(\lambda)^{-1}\mathbf{f}_{xy}^T(\lambda)$ may be regarded as an estimate of the error spectrum $g(\lambda) = f_{yy}(\lambda) - \mathbf{f}_{yx}(\lambda)\, \mathbf{f}_{xx}(\lambda)^{-1}\mathbf{f}_{xy}(\lambda)$.

We close this section by making some comments on the presented application:

The theoretical results of Section 2 are derived under the assumption of complex–Normality and with a particular loss function. For time series the complex–Normality of the frequency domain "observations" (created by taking finite Fourier transforms) is only true in a limiting sense. However, a frequency analysis of a time series will usually not be undertaken unless T is large, say at least of order $T = 2^9 = 512$. The central limit effect will therefore come into effect very strongly and gives reason to believe that the approximation to complex–Normality is fairly good. This is also confirmed by practical experience. This contrasts with the situation for real multivariate data analysis where the multinormal distribution rarely gives a very good approximation to reality. Moreover, as pointed out by several authors in the real case, there seem to be potential gains by using the shrunken estimates also when observations are not exactly normal.

Furthermore, we point out that the loss function employed is to an extent arbitrary and partly chosen for convenience. Workers in the real case have pointed out that the issue is present for a wide class of reasonable loss functions and the suggested methods are likely to provide improvement in a more general context, but the mathematics involved gets much harder (see Brown [4]).

4. SIMULATIONS

In this section we report some of our experience with comparison of the common estimate and the alternative estimate on simulated time series. The main reasons for doing simulations are:

To demonstrate the extent of the improvement under different circumstances. The suggested procedures are based on asymptotic theory, and one may want to know how the extent of the improvement depends on the finiteness of T. This affects the normality and will also introduce slight dependencies among the frequency domain observations employed. One is typically interested in estimating the transfer function at a number of frequencies in the domain $(0, \pi)$. It would be of interest to see how smoothing over overlapping sets of neighboring frequencies affects the estimates. Simulations also give the opportunity to compare the estimates for alternative loss functions which may be harder to handle analytically.

An extensive simulation study was undertaken, where we generated a variety of different $(r + 1)$-variate series of length T. The main effort was spent on various pure noise Gaussian series with $r = 5$ and $T = 2^9 = 512$, and some details of our findings for this case are reported below. We tried out some non-Gaussian and higher-order schemes as well, but the results differed very little in character from those reported.

We used the Fast Fourier Transform Algorithm to compute Finite Fourier Transforms (see Cooley and Tukey [5]). With $T = 512$ we therefore have transforms available for all frequencies of form $\lambda_s = 2\pi s/512$, $s = 0, 1, 2,...,$ $T - 1$. We have chosen to compute estimates of $\mathbf{A}(\lambda)$ at the frequencies

$$\frac{2\pi \cdot 1}{32}, \frac{2\pi \cdot 2}{32},..., \frac{2\pi \cdot 15}{32},$$

which is a compromise between keeping down the amount of output and the desire to detect peculiarities between neighboring estimates, if any. The number n of neighboring frequencies of form $2\pi s/512$ to be used in smoothing the estimates is taken to be $n = 11$ and $n = 21$. In the former case the estimates at neighboring frequencies of form $2\pi k/32$ are based on disjoint sets of frequency domain observations. In the latter case 5 (5) out of the 21 "observations" overlap with those of the left (right) neighbor. This makes it possible to study both the effect on the improvement when the smoothing takes more distant frequencies into account and the effect on estimates at neighboring frequencies in the grid when the dependence between these estimates is increased. In order not to be misled by a single (perhaps peculiar) run one should make repeated runs for the same process parameters. Note that for the case of pure noise series, when the spectrum is constant, the estimation at 15 different frequencies is essentially 15 close to independent repetitions of the same problem, and thus provides some insight into the variability of the performance of the estimate in a given situation.

In order to simulate our pure noise series, we started in all cases with a stretch $\epsilon(t)$, $t = 0,..., T - 1$ of a 6-variate pure noise series with independent components each being pseudonormal with zero mean and unit variance. Each pure noise series was generated by the choice of a real (6×6)-matrix $\mathbf{B}$ and

computation of $z(t) = \mathbf{B}\boldsymbol{\epsilon}(t)$, $t = 0, 1,..., T - 1$. The autocovariance structure of the series is then given by the (6×6)-matrix $c(u) = Ez(t + u) z(t)' = \mathbf{BB}'$ for $u = 0$ and 0 for $u \neq 0$ and the spectral density matrix is $\mathbf{f}_{zz}(\lambda) = (2\pi)^{-1}\mathbf{BB}' = \boldsymbol{\Phi}$ not dependent on λ. The transfer function of the first series with respect to the other five series is then $\mathbf{A}(\lambda) = \boldsymbol{\Phi}_{yx}\boldsymbol{\Phi}_{xx}^{-1} = \boldsymbol{\alpha}$, also not dependent on λ. We had originally simulated 28 different pure noise structures in order to cover a variety of situations. However, many of these turned out to be too smooth to be challenging, and some had features very similar to others. We have chosen to focus our attention on 10 of these structures and we refer to them as runs PN-1 to PN-10. Their associated $\mathbf{B}$ matrices are given by

$$
\begin{pmatrix}
1.0 & .0 & \cdots & .0 \\
.0 & 1.0 & \cdots & .0 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix},
\qquad
\begin{pmatrix}
1.0 & 1.0 & \cdots & .0 \\
.0 & 1.0 & \cdots & .0 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix},
$$

$$\text{PN-1} \qquad\qquad\qquad \text{PN-2}$$

$$
\begin{pmatrix}
1.0 & 2.0 & \cdots & .0 \\
.0 & 1.0 & \cdots & .0 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix},
\qquad
\begin{pmatrix}
2.0 & 1.0 & \cdots & .0 \\
.0 & 1.0 & \cdots & .0 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix},
$$

$$\text{PN-3} \qquad\qquad\qquad \text{PN-4}$$

$$
\begin{pmatrix}
1.0 & .4 & \cdots & .4 \\
.4 & 1.0 & \cdots & .4 \\
\vdots & \vdots & & \vdots \\
.4 & .4 & \cdots & 1.0
\end{pmatrix},
\qquad
\begin{pmatrix}
1.0 & .4 & \cdots & .4 \\
.0 & 1.0 & \cdots & .4 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix},
$$

$$\text{PN-5} \qquad\qquad\qquad \text{PN-6}$$

$$
\begin{pmatrix}
2.0 & 2.0 & \cdots & .0 \\
.0 & 1.0 & \cdots & .0 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix},
\qquad
\begin{pmatrix}
5.0 & 1.0 & \cdots & .0 \\
.0 & 1.0 & \cdots & .0 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix},
$$

$$\text{PN-7} \qquad\qquad\qquad \text{PN-8}$$

$$
\begin{pmatrix}
5.0 & 5.0 & \cdots & .0 \\
.0 & 1.0 & \cdots & .0 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix},
\qquad
\begin{pmatrix}
1.0 & 5.0 & \cdots & .0 \\
.0 & 1.0 & \cdots & .0 \\
\vdots & \vdots & & \vdots \\
.0 & .0 & \cdots & 1.0
\end{pmatrix}.
$$

$$\text{PN-9} \qquad\qquad\qquad \text{PN-10}$$

The pure noise series are generated accordingly with different basic noise series $\boldsymbol{\epsilon}(t)$ each time. For each run our computations are: For each frequency

$2\pi k/32$, $k = 1, 2,..., 15$ we compute the transfer function coefficients, the usual estimate $\hat{\alpha}^{(0)}$, and our new estimate

$$\hat{\alpha}^{(+)} = \left(1 - \frac{r-1}{n-r+1} \cdot \frac{1-R^2}{R^2}\right)^+ \hat{\alpha}^{(0)}, \qquad (4.1)$$

together with their respective computed losses. We also compute the multiple coherence and the empirical multiple coherence (R^2). To avoid misinterpretation of the simulation results we point out again that our theory has provided an estimate which is better at *all* frequencies in terms of expected loss, at least asymptotically, and occasional flops should then be attributed to chance.

In Table I we give for each of the 10 runs the associated (constant) transfer function, and for both $n = 11$ and $n = 21$ we give the average of the computed loss over the 15 frequencies of form $2\pi k/32$, $k = 1, 2,..., 15$ for the estimates $\hat{\alpha}^{(0)}$ and $\hat{\alpha}^{(+)}$. For comparative purposes the computations for $n = 11$ and $n = 21$ are based on the same simulated series. Note that the plus modification comes into effect whenever $R^2 < 4/11 = 0.357$ in the case of $n = 11$ and whenever $R^2 < 4/17 = 0.191$ in the case of $n = 21$.

TABLE I

Averaged Computed Losses for Old and New Estimate

Run	α	$n = 11$		$n = 21$	
		$\hat{\alpha}^{(0)}$	$\hat{\alpha}^{(+)}$	$\hat{\alpha}^{(0)}$	$\hat{\alpha}^{(+)}$
PN-1	(.0, .0,..., .0)	1.04	0.31	0.27	0.03
PN-2	(1.0, .0,..., .0)	0.54	0.43	0.23	0.21
PN-3	(2.0, .0,..., .0)	0.79	0.71	0.31	0.27
PN-4	(1.0, .0,..., .0)	0.76	0.35	0.29	0.12
PN-5	(.19, .0,..., .0)	0.89	0.74	0.27	0.24
PN-6	(.40, .24,..., .05)	1.13	0.71	0.31	0.25
PN-7	(2.0, .0,..., .0)	0.81 ·	0.58	0.30	0.25
PN-8	(1.0, .0,..., .0)	0.54	0.13	0.23	0.06
PN-9	(5.0, .0,..., .0)	0.76	0.56	0.29	0.24
PN-10	(5.0, .0,..., .0)	1.06	1.07	0.30	0.31

Table I shows that the new estimate has performed well over the 15 frequencies chosen, for all 10 pure noise series considered. The savings over the usual estimate range from 70–80% for the least challenging pure noise series PN-1 to no savings for the most challenging series PN-10, and the savings for the other series are in most cases appreciable. The table shows no clear picture as to which of the cases $n = 11$ and $n = 21$ gives the largest percentage saving, but our general experience seems to indicate that it is the former case. However, in the latter case the losses are reduced anyway.

Another interesting aspect is summarized in Table II. It shows for each of the 10 runs the frequencies of form $2\pi s/32$ at which the new estimate got higher computed loss than the common estimate for both cases $n = 11$ and $n = 21$.

TABLE II

New Estimate Worse at Frequencies $2\pi s/32$ out of 15 Possible

Run	$s\ (n = 11)$	$s\ (n = 21)$
PN-1	None	None
PN-2	7, 13, 15	1, 3, 7, 9, 13, 14, 15
PN-3	2, 7, 15	1, 13, 15
PN-4	None	None
PN-5	9	3, 5, 6, 9
PN-6	6, 11	2, 11
PN-7	6, 13	4, 6, 7, 13
PN-8	None	None
PN-9	None	None
PN-10	2, 5, 6, 11, 13	2, 5, 6, 7, 10, 11, 13

In this sense the new estimate is favorable in all the cases studied, ranging from the least challenging situation PN-1 to the most challenging PN-10. Even for PN-10 the new estimate turned out better at a majority of frequencies. Typically, more is saved at those frequencies for which the new estimate is better than is lost at those where the new estimate is worse. Table II seems to indicate that the new estimate comes out best in this sense more often in the case $n = 11$ than $n = 21$, and this is also our general experience.

We have also simulated 21 different first-order moving average series and a few second-order moving average series. The character of our findings is no different from those above. Finally we have simulated 26 different first-order autoregressive series, for which our findings were extremely favorable to the new estimate; in most case the new estimate turned out best at all frequencies. However, these simulations were based on models which seemed to be realistic in practice, and it is of course possible to choose structures for which the savings are only moderate.

Further remarks and a number of suggestions for further research in this area are given in Lillestøl [8].

REFERENCES

[1] BARANCHIK, A. J. (1973). Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables. *Ann. Statist.* **1** 312–321.
[2] BRILLINGER, D. R. (1974). Fourier analysis of stationary processes. *Proc. IEEE* **62** 1628–1643.

[3] BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

[4] BROWN, L. D. (1966). On the admissibility of invariant estimates of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

[5] COOLEY, J. W. AND TUKEY, J. W. (1965). An algorithm for the machine computation of complex Fourier series. *Math. Comp.* **19** 297–301.

[6] GOODMAN, N. R. (1963). Statistical analysis based upon a certain multivariate complex Gaussian distribution (an introduction). *Ann. Math. Statist.* **34**, 152–177.

[7] KRISHNAIAH, P. R. (1976). Some recent developments on complex multivariate distributions. *J. Multivariate Analysis.* **6** 1–30.

[8] LILLESTØL, J. (1975). *Improved Estimates of Parameters in Compex-Normal Models with Application to Analysis of Several Stationary Time Series*. Ph.D. Dissertation, University of California, Berkeley.

[9] SCLOVE, S. L. (1970). Some remarks on normal multivariate regression. *Ann. Inst. Statist. Math.* **22** 319–326.

[10] SCLOVE, S. L. (1971). Improved estimation of parameters in multivariate regression. *Sankhyā Ser. A* **31** 61–66.

[11] STEIN, C. (1960). Multiple regression. In *Contributions to Probability and Statistics* (I. Olkin, Ed.), pp. 424–443. Stanford Univ. Press.

[12] STEIN, C. (1966). An approach to the recovery of interblock information in balanced incomplete block designs. In *Research Papers in Statistics* (F. N. David, Ed.), pp. 351–366. Wiley, New York.