…SITY
UNIVERSITEIT
BRABANT
POSTBOX 90153
5000 LE TILBURG
THE NETHERLANDS



[redacted]

[redacted]

DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

**HIGHER ORDER MOMENTS OF BILINEAR TIME SERIES PROCESSES WITH SYMMETRICALLY DISTRIBUTED ERRORS**

Jan G. de Gooijer and Ruud M.J. Heuts

**FEW 251** 510.9

April, 1987

# HIGHER ORDER MOMENTS OF BILINEAR TIME SERIES PROCESSES WITH SYMMETRICALLY DISTRIBUTED ERRORS

by

Jan G. de Gooijer[1)] and Ruud M.J. Heuts[2)]

[1)]Department of Economic Statistics
University of Amsterdam
Jodenbreestraat 23
1011 NH Amsterdam
Holland

[2)]Department of Econometrics
University of Brabant
P.O. Box 90153
5000 LE Tilburg
Holland

Summary

For the stationary superdiagonal, diagonal and subdiagonal bilinear time series model with symmetrically distributed errors formulas for the standardized third and fourth order central moments are obtained. It is shown that these results contain useful information to distinguish a bilinear process from a white noise process if the errors are Gaussian distributed. Application to situations where the distribution function of the errors is a member of the class of symmetric exponential power distributions is also briefly discussed. An empirical example is given for illustrative purpose.

Key words: Bilinear time series; diagonal, superdiagonal and subdiagonal models; kurtosis; skewness.

## 1. Introduction

Linear time series models such as the well-known family of autoregressive integrated moving average (ARIMA) models provide a remarkable effective and versatile range of possibilities to adequately approximate many time series observed in practice. However, despite reported evidence of the advantages of these models, it is increasingly recognized that there are time series in economics and operations research which are unlikely to be well represented by any linear model; see, for example, Maravall (1983), Hinich and Patterson (1985) and Subba Rao and Gabr (1984). A number of authors have studied various tractable classes of non-linear models. Among these are the so-called bilinear models discussed by Granger and Andersen (1978) and Subba Rao and Gabr (1984), the threshold autoregressive models of Tong and Lim (1980) and the exponential autoregressive models of Lawrance and Lewis (1985). Each of these models offer a useful avenue in representing nonlinearity in observed time series data.

In this paper we will concentrate on the class of bilinear autoregressive moving average models. The most general form of this model is given by

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + A_t + \sum_{j=1}^{q} \theta_j A_{t-j} + \sum_{\ell=1}^{r} \sum_{k=1}^{s} \beta_{\ell k} Y_{t-\ell} A_{t-k} \tag{1.1}$$

where $\{A_t\}$ is a sequence of i.i.d. random variables with mean zero and fixed variance $\mu_2$ and where $\{Y_t\}$ is a discrete stationary time series process. If the parameters $\beta_{\ell k}=0$ for all $\ell>k$ model (1.1) is usually referred to as the superdiagonal model. It is called the subdiagonal model if $\beta_{\ell k}=0$ for all $\ell<k$ and diagonal model if $\beta_{\ell k}=0$ for all $\ell \neq k$.

The successful application of a particular class of nonlinear models hinges heavily on the determination of the most appropriate model, or models, within its class with respect to some prespecified loss function. For bilinear models the use of Akaike's information criterion has been suggested to determine the orders p, q, r and s of (1.1). However, it is well-known that for linear

ARMA models this criterion asymptotically overestimates the "true" orders with a non-zero probability. This feature of AIC may also extend to the bilinear model (1.1).

Another way of identifying the structure of a bilinear model has been considered by Granger and Andersen (1978) and Maravall (1983). These authors suggest to make first a preliminary identification of the orders of the linear part of (1.1). In principle this can be done by any of the order determination methods proposed for linear time series modelling; see, for example, De Gooijer, Abraham, Gould and Robinson (1985) for a review of these methods. Then, a bilinear model is fitted to the residuals obtained in the first stage using the sample autocorrelations of the unsquared and squared residuals as identification tools. By relating these statistics to the known behaviour of their corresponding theoretical quantities inferences can be made about the most appropriate bilinear structure for the residuals.

A crucial element in this last approach is that knowledge should be available on the theoretical behaviour of the autocorrelations for different bilinear models. However, because of the rapid increase in algebraic complexity as the orders r and s in (1.1) become bigger, these results have only been obtained for a limited number of models (see, e.g., Granger and Andersen (1978) and Li (1984)). Moreover, one may wonder whether autocorrelations of unsquared and squared residuals provide enough information about the system under study. In particular, these statistics do not completely determine the structure of the process when the series are generated by a non-Gaussian bilinear model. In such a situation it is necessary to analyse the higher order moments of the process.

In this paper we study the theoretical properties of the third and fourth central moments of three stationary bilinear time series models having symmetrically distributed errors. These results can be used to distinguish a bilinear process from a white noise process. Moreover, they provide useful information about the most appropriate bilinear time series model.

The paper is organized as follows: The three bilinear time series models and some assumptions are given in Section 2. In Section 3 we derive the theoretical standardized third and fourth central moments for these models assuming that the errors are symmetrically distributed around mean zero. Estimators of these moments will be considered in Section 4 together with approximate expressions for their variances. Also some data simulations are performed in this section to study the sampling properties of the standardized third and fourth central moments. Finally, in Section 4 an example will be given illustrating the usefulness of the results obtained in the previous sections for detecting bilinear relations in empirical data.

## 2. Preliminaries

Assume that $\{Y_t\}$ is generated by one of the following three bilinear models (respectively, superdiagonal, diagonal and subdiagonal model)

$$Y_t = \beta Y_{t-\ell} A_{t-k} + A_t \qquad (0<k<\ell) \tag{2.1}$$

$$Y_t = \beta Y_{t-k} A_{t-k} + A_t \qquad (k>0) \tag{2.2}$$

$$Y_t = \beta Y_{t-k} A_{t-k-1} + A_t \qquad (k>0) \tag{2.3}$$

where for the sake of generality it is assumed that $\{A_t\}$ is distributed symmetrically around mean zero. For ease of reference, we define $E(A_t^i)=\mu_i$ if i is even and $\mu_i=0$ if i is odd. If the error distribution is specialised to a zero-mean Gaussian distribution then the even moments of $\{A_t\}$ up to order eight are given by $\mu_4=3\mu_2^2$, $\mu_6=15\mu_2^3$, $\mu_8=105\mu_2^4$ while all the odd moments are equal to zero.

Quinn (1982) gives the following necessary and sufficient condition for strict stationarity for time series generated by (2.1)-(2.3): $\ln|\beta|+E(\ln|A_t|)<0$. For zero-mean Gaussian distributed $\{A_t\}$ this condition reduces to $\mu_2|\beta|<1.8874$. We assume that this condition is satisfied in the rest of this paper. Also we assume that all moments up to order four exist. For the superdiagonal model (2.1) with k=1 and $\ell$=2, Granger and Andersen (1978, p. 40) show that this assumption becomes equivalent to checking the condition $\beta^4\mu_4<1$.

Instead of considering the behaviour of the autocorrelations of $\{Y_t^2\}$, $\rho(i)=\text{cov}(Y_t^2,Y_{t-i}^2)/\text{var}(Y_t^2)$, for different lags i as a characteristic pattern for bilinear model identification, we will investigate the behaviour of the standardized third and fourth central moments. These moments are, respectively, given by

$$\beta_1(i,j) = E[(Y_t-\mu)(Y_{t-i}-\mu)(Y_{t-j}-\mu)]/\{\text{var}(Y_t)\}^{3/2}$$

$$= [c(i,j) - \mu\{E(Y_{t-i}Y_{t-j})+E(Y_tY_{t-j})+E(Y_tY_{t-i})]\} + 2\mu^3]/\{\text{var}(Y_t)\}^{3/2} \tag{2.4}$$

and

$$\beta_2(0,i,j) = E[(Y_t-\mu)^2(Y_{t-i}-\mu)(Y_{t-j}-\mu)]/\{\text{var}(Y_t)\}^2$$

$$= [c(0,i,j) - \mu\{c(0,i)+c(0,j)+2c(i,j)\} + \mu^2\{E(Y_t^2)+E(Y_{t-i}Y_{t-j}) +2E(Y_tY_{t-j})+2E(Y_tY_{t-i})\} - 3\mu^4]/\{\text{var}(Y_t)\}^2 \tag{2.5}$$

where

$$c(i,j) = E(Y_tY_{t-i}Y_{t-j}), \tag{2.6}$$

$$c(u,i,j) = E(Y_tY_{t-u}Y_{t-i}Y_{t-j}) \tag{2.7}$$

and $\mu=E(Y_t)$.

These moments have several features over $\rho(i)$ useful for identification of a bilinear time series model. For instance, if $\{Y_t\}$ is generated by a linear model with symmetrically distributed errors then $\rho(i)$ can be nonzero whereas (2.4) will be zero for all lags i and j. Also the two dimensional pattern of (2.4) for different lags i an j should provide more information about the model under study than the one dimensional autocorrelation pattern of $\rho(i)$.

## 3. Higher order moments

In this section we summarize in a number of lemmas and theorems, the main results on the higher order moments of the bilinear time series models (2.1)-(2.3). For ease of notation we use throughout this section the definition $\lambda=\beta^2\mu_2$.

### 3.1 The superdiagonal model

*Lemma 3.1* If $\{Y_t\}$ is generated by (2.1) with $\{A_t\}$ distributed symmetrically around mean zero then $c(i,j)=\beta\mu_2^2/(1-\lambda)$ for $(i,j)=(k,\ell)$ or $(\ell,k)$ and $c(i,j)=0$ otherwise; see Guegan (1984, Table 1).

*Proof*. The proof is straightforward and is therefore omitted.

*Lemma 3.2* If $\{Y_t\}$ is generated by (2.1) with $\{A_t\}$ distributed symmetrically around mean zero then

$$\mathrm{cov}(Y_t^2,Y_{t-j}^2) = \begin{cases} \lambda^m[\mu_4-\mu_2^2-2\lambda(\mu_4-3\mu_2^2)(1-\lambda)]/(1-\lambda)^2(1-\beta^4\mu_4) & \text{for } j=m\ell \\ & (m=0,1,2,\ldots); \\ \lambda^{m+1}(\mu_4-\mu_2^2)/(1-\lambda)^2(1+\lambda) & \text{for } j=m\ell\pm k \text{ and } \ell\neq 2k \\ & (m=0,1,2,\ldots); \\ \lambda^{m+1}(\mu_4-\mu_2^2)/(1-\lambda)^2 & \text{for } j=m\ell+k \text{ and } \ell=2k \\ & (m=0,1,2,\ldots); \\ 0 & \text{otherwise.} \end{cases}$$

*Proof*. Only the first part of this lemma will be proved here. The proof of the second and third part is similar. The fourth part was proved by Li (1984, Lemma 2).

If $j=m\ell$ $(m=0,1,2,\ldots)$, then

$$\begin{aligned} c(0,m\ell,m\ell) &= E[(\beta^2Y_{t-\ell}^2A_{t-k}^2+2\beta Y_{t-\ell}A_{t-k}A_t+A_t^2)Y_{t-m\ell}^2] = \lambda E(Y_{t-\ell}^2Y_{t-m\ell}^2)+\mu_2E(Y_t^2) \\ &= \mu_2[1+\lambda+\lambda^2+\ldots+\lambda^{m-1}]E(Y_t^2)+\lambda^mE(Y_t^4). \end{aligned} \qquad (3.1)$$

Substituting (3.1), together with

$$E(Y_t^2) = \mu_2/(1-\lambda) \tag{3.2}$$

and

$$E(Y_t^4) = (\mu_4-\lambda\mu_4+6\lambda^2\mu_2)/(1-\lambda)(1-\beta^4\mu_4) \tag{3.3}$$

into $\mathrm{cov}(Y_t^2,Y_{t-m\ell}^2)=c(0,m\ell,m\ell)-E(Y_t^2)^2$ the result follows. □

Notice that for zero-mean Gaussian distributed $\{A_t\}$ and m=0 the second part of Lemma 3.2 is identical to the first part of Lemma 4 of Li (1984), while for m=1 the second part of Lemma 3.2 conforms to Lemma 3 of Li (1984). Notice also that for zero-mean Gaussian distributed $\{A_t\}$ and m=0 the third part of Lemma 3.2 is similar to the second part of Lemma 4 of Li (1984). Furthermore, it is easy to see that the results of Lemma 3.2 are in agreement with the results given by Guegan (1984, Table 2) for the superdiagonal model (2.1) with k=1, $\ell$=2 and zero-mean Gaussian distributed errors.

*Theorem 3.1* If $\{Y_t\}$ is generated by (2.1) with $\{A_t\}$ distributed symmetrically around mean zero then the standardized third and fourth central moments of $\{Y_t\}$, respectively, are given by

$$\beta_1(i,j) = c(i,j)/\{E(Y_t^2)\}^{3/2} \text{ and } \beta_2(u,i,j) = c(u,i,j)/\{E(Y_t^2)\}^2$$

where $E(Y_t^2)$ is given by (3.2), c(i,j) and c(0,j,j) follow from, respectively, Lemma 3.1 and 3.2, and where for $i \neq j$

$$c(u,i,j) = \begin{cases} 2\lambda^{m+2}\mu_2^2/(1-\lambda) & \text{for } u=0,\ i=m\ell+k,\ j=(m+2)\ell \text{ and } \ell=2k \\ & \qquad (m=-1,0,1,2,\ldots); \\ \lambda\mu_2^2/(1-\lambda) & \text{for } u=k,\ i=\ell+k \text{ and } j=2\ell; \\ 0 & \text{otherwise.} \end{cases}$$

*Proof*. Using the fact that $\mu=E(Y_t)=0$ the proof is straightforward and is omitted.

It is clear that the coefficient of skewness $\beta_1(0,0)$ is equal to zero for every $\beta\neq 0$. Hence, it will be difficult to distinguish the superdiagonal bilinear model (2.1) from pure white noise on the basis of this coefficient. The coefficient of kurtosis

$$\beta_2(0,0,0) = (\mu_4-\lambda\mu_4+6\lambda\mu_2^2)(1-\lambda)/\{\mu_2^2(1-\beta^4\mu_4)\} \tag{3.4}$$

is more useful for this purpose, provided $\beta^4\mu_4<1$. For zero-mean Gaussian distributed $\{A_t\}$ (3.4) reduces to

$$\beta_2(0,0,0) = 3(1-\lambda^2)/(1-3\lambda^2) \tag{3.5}$$

which is identical to relation (5.14) of Granger and Andersen (1978). In this case $\beta_2(0,0,0)>3$, for every $\beta\neq 0$, which indicates that this process has a higher degree of peakedness and thick-tailness than a pure Gaussian white noise process.

## 3.2 The diagonal model

For the diagonal model (2.2) the series $\{Y_t\}$ is a function of $A_t$, $A_{t-k}$, $A_{t-2k},\ldots$, so there are actually k independent series within $\{Y_t\}$. Therefore, unless j=mk, where m is an integer, $\text{cov}(Y_t,Y_{t-j})\equiv 0$ for errors distributed symmetrically around mean zero. Following the same reasoning as above, $\text{cov}(Y_t^2,Y_{t-j}^2)\equiv 0$, unless j=mk. Hence, Lemma 1 of Li (1984) is true even for errors distributed symmetrically around mean zero. Furthermore, under this assumption for $\{A_t\}$, the moments $E[(Y_t-\mu)(Y_{t-i}-\mu)(Y_{t-j}-\mu)]$ and $E[(Y_t-\mu)(Y_{t-u}-\mu)(Y_{t-i}-\mu)(Y_{t-j}-\mu)]$ are both identically zero, unless $u=m_1k$, $i=m_2k$ and $j=m_3k$ where $m_1$, $m_2$ and $m_3$ are integers. Thus the problem of deriving higher order moments of (2.2) reduces to obtaining moments of series $\{Y_t\}$ generated by the diagonal model

$$Y_t = \beta Y_{t-1}A_{t-1} + A_t. \qquad (3.6)$$

The following two lemmas are stated without any proof. Details of the proofs are, of course, available if required. However, since they do not possess any interesting features, it has been thought better to omit them.

*Lemma* 3.3 If $\{Y_t\}$ is generated by (3.6) with $\{A_t\}$ distributed symmetrically around mean zero and $\lambda<1$ then

$$c(i,j) = \begin{cases} 3\beta\mu_2^2+\beta^3\mu_6+3\beta^5\mu_4^2/(1-\lambda) & \text{for } i=j=0; \\ \beta\mu_2^2+3\beta\lambda\mu_4+\beta^3\lambda\mu_6+3\beta^5\lambda\mu_4^2/(1-\lambda) & \text{for } i=0,\ j=1; \\ \beta\mu_2^2+\lambda^{j-2}(\beta\lambda\mu_4+\beta^3\lambda^2\mu_6+3\beta\lambda^2\mu_4)+\beta\lambda\mu_4\{1+\lambda^{j-2}(3\beta^4\lambda\mu_4-1)\}/(1-\lambda) & \\ & \text{for } i=0,\ j>1; \\ \beta\mu_4(1+2\lambda)/(1-\lambda) & \text{for } i=j=1; \\ 4\beta\lambda\mu_2^2 & \text{for } i=1,\ j=2; \\ 2\lambda\mu_2^2 & \text{for } i=1,\ j>2; \\ \beta\mu_2 E(Y_{t-i}Y_{t-j}) & \text{for } i>1,\ j>1; \end{cases}$$

where $E(Y_tY_{t-i})=2\lambda\mu_2$ for $i=1$ and $E(Y_tY_{t-i})=E(Y_tY_{t+i})=\lambda\mu_2$ for $i>1$.

*Lemma* 3.4 If $\{Y_t\}$ is generated by (3.6) with $\{A_t\}$ distributed symmetrically around mean zero and $\beta^4\mu_4<1$ then

$$c(0,i,i) = \begin{cases} [\beta^6(-\mu_2\mu_8-5\mu_2\mu_4^2+6\mu_4\mu_6)+\beta^4(\mu_8-\mu_4^2)+5\lambda\mu_4+\mu_4]/(1-\lambda)(1-\beta^4\mu_4) & \\ & \text{for } i=0; \\ [\beta^8(7\mu_2\mu_4\mu_6-\mu_2^2\mu_8-6\mu_4^3)+\beta^6(\mu_2\mu_8-\mu_4\mu_6+\mu_2^3\mu_4-\mu_2\mu_4^2)+\beta^4(6\mu_4^2-\mu_2\mu_6- & \\ \mu_2^2\mu_4)+\beta^2(\mu_6+\mu_2\mu_4-\mu_2^3)+\mu_2^2]/(1-\lambda)(1-\beta^4\mu_4) & \text{for } i=1; \\ \lambda c(0,i-1,i-1)+(1-\lambda)E(Y_t^2)^2 & \text{for } i>1; \end{cases}$$

where $E(Y_t^2)=(\mu_2-\lambda\mu_2+\beta^2\mu_4)/(1-\lambda)$.

From Lemma 3.4 it is easy to see that for zero-mean Gaussian distributed $\{A_t\}$ we get

$$\mathrm{var}(Y_t^2) = 2\mu_2^2(1+4\lambda+40\lambda^2+18\lambda^3-54\lambda^4)/(1-\lambda)^2(1-3\lambda^2) \tag{3.7}$$

and

$$\mathrm{cov}(Y_t^2,Y_{t-1}^2) = 6\lambda\mu_2^2(2+3\lambda+5\lambda^2+\lambda^3-8\lambda^4)/(1-\lambda)^2(1-3\lambda^2) \tag{3.8}$$

provided $\lambda^2<1/3$. Substituting (3.7) and (3.8) into $\rho(1)=\mathrm{cov}(Y_t^2,Y_{t-1}^2)/\mathrm{var}(Y_t^2)$ leads to relation (6.35) of Granger and Andersen (1978). Furthermore, from the third part of Lemma 3.4, we have $\rho(i)=\lambda\rho(i-1)$, $i>1$, which is in agreement with relation (6.34) given by Granger and Andersen (1978, p. 55). Notice, however, that their statement saying that this result can be obtained "... without any assumptions being made about the distribution of $\{A_t\}$ ..." is not correct.

In principle the results of Lemma 3.4 can be generalized to any combination of the lags u, i and j in c(u,i,j). Since, however, we are only interested in the behaviour of $\beta_2(0,i,i)$ for lags $i\geq 0$ in Section 4 we will not pursue this matter here any further. From Lemma 3.3 and 3.4 the following theorem can be straightforwardly obtained.

*Theorem* 3.2 If $\{Y_t\}$ is generated by (3.6) with $\{A_t\}$ distributed symmetrically around mean zero then the standardized third and fourth central moments $\beta_1(i,j)$ and $\beta_2(0,i,i)$, respectively, are given by

$$\beta_1(i,j) = \begin{cases} \beta^3[2\mu_2^3+\mu_6+3\mu_4(\beta^2\mu_4-\mu_2)/(1-\lambda)]/\{\mathrm{var}(Y_t)\}^{3/2} & \text{for } i=j=0; \\ [c(i,1)-\beta\mu_2^2-2\beta\lambda\mu_2^2-\beta\lambda\mu_4/(1-\lambda)]/\{\mathrm{var}(Y_t)\}^{3/2} & \text{for } i=0,\ j=1 \text{ and } i=1,\ j=1; \\ [c(0,j)-\beta\mu_2^2-\beta\lambda\mu_4/(1-\lambda)]/\{\mathrm{var}(Y_t)\}^{3/2} & \text{for } i=0,\ j>1; \\ \beta\lambda\mu_2^2/\{\mathrm{var}(Y_t)\}^{3/2} & \text{for } i=1,\ j=2; \\ 0 & \text{for } i\geq 1,\ j>2, \end{cases}$$

and

$$\beta_2(0,i,i) = \begin{cases} [c(0,0,0)-4\beta\mu_2 c(0,0)+3\lambda\mu_2\{2\mu_2-\lambda\mu_2+2\beta^2\mu_4/(1-\lambda)\}]/\{var(Y_t)\}^2 & \text{for } i=0; \\ [c(0,1,1)-2\lambda\{c(0,1)+c(1,1)\}+\lambda\mu_2\{2\mu_2+5\lambda\mu_2+2\beta^2\mu_4/(1-\lambda)\}] \\ /\{var(Y_t)\}^2 & \text{for } i=1; \\ [c(0,i,i)-2\lambda\{c(0,i)+c(i,i)\}+\lambda\mu_2\{2\mu_2+\lambda\mu_2+2\beta^2\mu_4/(1-\lambda)\}] \\ /\{var(Y_t)\}^2 & \text{for } i>1, \end{cases}$$

where $var(Y_t)=(\mu_2^2-2\lambda\mu_2+\beta^2\mu_4+\lambda^2\mu_2)/(1-\lambda)$.

Using Theorem 3.2 the coefficients of skewness and kurtosis for zero-mean Gaussian distributed $\{A_t\}$ are, respectively, given by

$$\beta_1(0,0) = \frac{2\beta^3\mu_2^3(4+5\lambda^2)/(1-\lambda^2)}{\{\mu_2^2(1-\beta^2+\beta^4\mu_2)/(1-\lambda^2)\}^{3/2}} \tag{3.9}$$

provided $\lambda^2<1$, and

$$\beta_2(0,0,0) = \frac{\mu_2^2(3+9\lambda^2+57\lambda^4+93\lambda^6+117\lambda^8+135\lambda^{10})/(1-\lambda^2)(1-3\lambda^4)}{\{\mu_2^2(1+\beta^2+\beta^4\mu_2)/(1-\lambda^2)\}^2} \tag{3.10}$$

provided $3\lambda^4<1$.

Clearly the expression in the numerator of (3.9) is not in agreement with the third central moment given by Granger and Andersen (1978, p. 52). Also the numerator of (3.10) differs a factor 3 with the fourth central moment (6.26) given by these authors. This last result may explain the high values of the coefficient of kurtosis given in column 5 of Table 1 of Granger and Andersen (1978). Notice that for the diagonal model (2.2) we get a pattern of non-zero $\beta_1(i,j)$ at lags $(i,j)=(0,0)$, $(k,k)$, $(k,2k)$ and $(0,mk)$ for $m=1,2,3,\ldots$ . This in contrast to one dominating value of $\beta_1(i,j)$ at lag $(i,j)=(k,\ell)$ and $(\ell,k)$ and zero values elsewhere for the superdiagonal model (2.1).

### 3.3 The subdiagonal model

The moments of this model are much more difficult to derive than those of the two previously discussed bilinear time series models. Only for a few simple subdiagonal models the autocorrelations $\rho(i)$ of the squared time series process $\{Y_t^2\}$ have been obtained (see Granger and Andersen (1978, Chapter VII)). The reason is that it is often difficult to leave behind lagging errors upon repeatedly substituting the model equation for the lowest lagging $\{Y_t\}$ since $k<\ell$. For the subdiagonal model (2.3) no results have appeared in the literature for the standardized third and fourth central moments as far as we know. Without proof we first state the following two lemmas.

*Lemma* 3.5 If $\{Y_t\}$ is generated by (2.3) with $\{A_t\}$ distributed symmetrically around mean zero then

$$c(i,j) = \begin{cases} \beta\lambda\mu_4+\beta\mu_2^2 & \text{for } k=1 \text{ and } i=2,\ j=1 \text{ or } i=1, j=2; \\ \beta\mu_2^2 & \text{for } k=2 \text{ and } i=2,\ j=3 \text{ or } i=3,\ j=2; \\ \lambda\beta\mu_2^2+\lambda\mu_2^2/(1-\lambda) & \text{for } k=3,4,\ldots \text{ and } i=k,\ j=k+1 \text{ or } i=k+1,\ j=k; \\ 0 & \text{otherwise.} \end{cases}$$

*Lemma* 3.6 If $\{Y_t\}$ is generated by (2.3) with $\{A_t\}$ distributed symmetrically around mean zero and $\beta^4\mu_4<1$ then

$$c(0,i,i) = \begin{cases} \beta^8E(Y_t^4A_t^4A_{t-1}^4)+6\lambda^2\mu_4(1+\beta^4\mu_4)/(1-\lambda)+6\lambda\mu_2(\mu_2+\beta^4\mu_6)+\mu_4(1+\beta^4\mu_4) \\ \qquad \text{for } i=0,\ k=1; \\ \beta^8\lambda E(Y_t^4A_t^4A_{t-1}^4)+\lambda\mu_4\{1+6(\lambda+\beta^8\mu_4^2)\}/(1-\lambda)+\beta^6\mu_4\mu_6(6\lambda+1)+\mu_2^2(\lambda+1) \\ \qquad \text{for } i=1,\ k=1; \\ \beta^8\lambda^2E(Y_t^4A_t^4A_{t-1}^4)+\{\lambda^2\mu_4(1+\lambda)+2\beta^4\lambda\mu_4^2(1+3\beta^4\lambda\mu_4)\}/(1-\lambda)+ \\ \beta^6\lambda\mu_4\mu_6(6\lambda+1)+\mu_2^2(\lambda^2+\lambda+1)+\lambda(\beta^2\mu_6+\mu_4) \qquad \text{for } i=2,\ k=1; \\ \beta^6\mu_4^2\{6\lambda^4/(1-\lambda)+\beta^{10}\mu_4^2\}/(1-\beta^4\mu_4)+6\lambda\mu_2^2(1+\beta^4\mu_4+\beta^8\mu_4^2)/(1-\lambda)+\mu_4+ \\ \beta^4\mu_4^2(1+\beta^4\mu_4) \qquad \text{for } i=0,\ k\geq 2; \\ \lambda c(0,i-1,i-1)+\mu_2^2\{1+\lambda+\beta^4\mu_4/(1-\lambda)\}+\beta^4(\mu_4-\mu_2^2)E(Y_{t-i}^2A_{t-3}^2) \\ \qquad \text{for } i\geq 3,\ k=1; \\ \lambda c(0,i-k,i-k)+\mu_2^2/(1-\lambda) \qquad \text{for } i-1>k\geq 2, \end{cases}$$

where for k=1

$$E(Y_t^4A_t^4A_{t-1}^4) = \{6\beta^2\lambda\mu_4^2\mu_6/(1-\lambda)+6\lambda\mu_6^2+\mu_4\mu_8\}/(1-\beta^4\mu_4)$$

and

$$E(Y_{t-i}^2A_{t-3}^2) = \begin{cases} \mu_4+\lambda\mu_2^2+\lambda^2\mu_4/(1-\lambda) & \text{for } i=3; \\ \mu_2^2\{1+\lambda+\beta^4\mu_4/(1-\lambda)\} & \text{for } i\geq 4. \end{cases}$$

Using (2.4), (2.5) and $\mu=E(Y_t)=0$ the third and fourth central moments are given as follows:

*Theorem* 3.3 If $\{Y_t\}$ is generated by (2.3) with $\{A_t\}$ distributed symmetrically around mean zero then the standardized third and fourth central moments $\beta_1(i,j)$ and $\beta_2(0,i,i)$, respectively, are given by

$$\beta_1(i,j) = c(i,j)/\{E(Y_t^2)\}^{3/2} \quad \text{and} \quad \beta_2(0,i,i) = c(0,i,i)/\{E(Y_t^2)\}^2$$

where $c(i,j)$ and $c(0,i,i)$ follow from, respectively, Lemma 3.5 and 3.6, and where

$$E(Y_t^2) = \begin{cases} \mu_2+\lambda\mu_2+\beta^2\lambda\mu_4/(1-\lambda) & \text{for } k=1; \\ \mu_2/(1-\lambda) & \text{for } k\geq 2. \end{cases}$$

It is clear from Theorem 3.3 and Lemma 3.5 that the coefficient of skewness $\beta_1(0,0)$ is equal to zero for every $\beta\neq 0$ and $\{A_t\}$ distributed symmetrically around mean zero. Also, it is obvious from the last part of Lemma 3.6 that $\text{cov}(Y_t^2,Y_{t-i}^2)=\lambda\text{cov}(Y_t^2,Y_{t-i+k}^2)$, for $i-1>k\geq 2$, for symmetrically distributed errors. This generalizes relation (7.22) given by Granger and Andersen (1978) for Gaussian distributed $\{A_t\}$. Finally, for zero-mean Gaussian distributed $\{A_t\}$ we have for the coefficient of kurtosis $\beta_2(0,0,0)$ the representation

$$\beta_2(0,0,0) = \begin{cases} 3(1-\lambda)(-144\lambda^6+264\lambda^5+72\lambda^4+24\lambda^3+4\lambda^2+\lambda+1)/(1+2\lambda^2)^2(1-3\lambda^2) & \text{for } k=1; \\ 3(1-\lambda)\{(1+\lambda)(1-27\lambda^6)+9\beta^2\lambda^6(2+3\beta^2-3\beta^2\lambda)\}/(1-3\lambda^2) & \text{for } k\geq 2, \end{cases} \tag{3.11}$$

provided $3\lambda^2<1$.

4. Some simulation results

Let $\{Y_t: t=1,\ldots,n\}$ be a set of realizations of $\{Y_t\}$. Then the natural estimators of $\rho(i)$, $\beta_1(i,j)$ and $\beta_2(0,i,i)$, respectively, are given by

$$r^*(i) = \sum_{t=1}^{n-i}(Y_t^2-\tilde{Y})(Y_{t+i}^2-\tilde{Y})/\sum_{t=1}^{n}(Y_t^2-\tilde{Y})^2, \tag{4.1}$$

$$b_1(i,j) = n^{-1}\sum_{t=1}^{n-\max(i,j)}(Y_t-\bar{Y})(Y_{t+i}-\bar{Y})(Y_{t+j}-\bar{Y})/\{n^{-1}\sum_{t=1}^{n}(Y_t-\bar{Y})^2\}^{3/2}, \tag{4.2}$$

$$b_2(0,i,i) = n^{-1}\sum_{t=1}^{n-i}(Y_t-\bar{Y})^2(Y_{t+i}-\bar{Y})^2/\{n^{-1}\sum_{t=1}^{n}(Y_t-\bar{Y})^2\}^2 \tag{4.3}$$

where $\bar{Y}=n^{-1}\Sigma_{t=1}^{n}Y_t$ and $\tilde{Y}=n^{-1}\Sigma_{t=1}^{n}Y_t^2$.

If $\{Y_t\}$ is generated by a zero-mean Gaussian distributed white noise process then it follows from the central limit theorem proved by Sun (1963) that $r^*(i)$ is uncorrelated and asymptotically Gaussian distributed with mean zero and variance $n^{-1}$. Similarly, for Gaussian distributed white noise, the statistics (4.2) and (4.3) both are asymptotically Gaussian with mean zero. Since, however, the variance of $b_1(i,j)$ and $b_2(0,i,i)$ depend on the lags i and j, it is necessary to distinguish three different cases for the approximate variance expressions of these statistics.

First, if i=j=0 and $\{Y_t\}$ is zero-mean Gaussian white noise, exact expressions for the variance of (4.2) and (4.3), respectively, are given by

$$\mathrm{var}[b_1(i,j)] = 6(n-2)/(n+1)(n+3), \tag{4.4}$$

and

$$var[b_2(0,i,i)] = 24n(n-2)(n-3)/(n+1)^2(n+3)(n+5). \qquad (4.5)$$

These results can be easily obtained from Kendall and Stuart (1969, p. 297-298 and p. 305-306).

Second, if $i\neq0$, $j\neq0$, $i\neq j$ and $\{Y_t\}$ is zero-mean Gaussian white noise, we have $var[(Y_t-\bar{Y})(Y_{t-i}-\bar{Y})(Y_{t-j}-\bar{Y})]\simeq\mu_2^3$ and $var[(Y_t-\bar{Y})^2(Y_{t+i}-\bar{Y})^2]\simeq9\mu_2^4$. Hence, using the well-known Taylor series expansion for the variance of a ratio of two random variables (see, e.g., Kendall and Stuart (1969, p. 232)), we find

$$var[b_1(i,j)] \simeq n^{-1}\mu_2^3/\mu_2^3 = 1/n, \qquad (4.6)$$

$$var[b_2(0,i,i)] \simeq 9n^{-1}\mu_2^4/\mu_2^4 = 9/n. \qquad (4.7)$$

Finally, if $0=i\neq j$ or $0\neq i=j$ or $0=j\neq i$ and $\{Y_t\}$ is zero-mean Gaussian white noise, we have $var[(Y_t-\bar{Y})(Y_{t-i}-\bar{Y})(Y_{t-j}-\bar{Y})]\simeq E(Y_t^4)E(Y_t^2)=3\mu_2^3$ which gives

$$var[b_1(i,j)] \simeq 3n^{-1}\mu_2^3/\mu_2^3 = 3/n. \qquad (4.8)$$

To investigate the accuracy of the approximation (4.6)-(4.8) a simulation experiment was performed. For the number of replications N set equal to 18000/n, with n=100(50)300, the following two statistics were computed

$$sd[b_1(i,j)] = [\sum_{u=1}^{N}\{b_{1u}(i,j)-\bar{b}_1(i,j)\}^2/(N-1)]^{1/2} \quad (i=0,1,2;\ j=1,2) \qquad (4.9)$$

$$sd[b_2(0,i,i)] = [\sum_{u=1}^{N}\{b_{2u}(0,i,i)-\bar{b}_2(0,i,i)\}^2/(N-1)]^{1/2} \quad (i=1) \qquad (4.10)$$

where $\bar{b}_1(i,j)=\Sigma_1^N b_{1u}(i,j)/N$ and $\bar{b}_2(0,i,i)=\Sigma_1^N b_{2u}(0,i,i)/N$ with $b_{1u}(i,j)$ and $b_{2u}(0,i,i)$, respectively, the value of $b_1(i,j)$ and $b_2(0,i,i)$ obtained from the uth replication. These results are reported in Table 1 together with $\{var[b_1(i,j)]\}^{1/2}$ and $\{var[b_2(0,i,i)]\}^{1/2}$ obtained from (4.6)-(4.8).

Table 1. A comparison between approximated and simulated values of $\{var[b_1(i,j)]\}^{1/2}$ and $\{var[b_2(0,i,i)]\}^{1/2}$ for series generated by a zero-mean Gaussian white noise process and sample sizes n=100(50)300.

| n | $sd[b_1(i,j)]$ | | | | $\{Appr.(4.6)\}^{\frac{1}{2}}$ | $sd[b_1(1,2)]$ | $\{Appr.(4.8)\}^{\frac{1}{2}}$ | $sd[b_2(0,1,1)]$ | $\{Appr.(4.7)\}^{\frac{1}{2}}$ |
|---|---|---|---|---|---|---|---|---|---|
| | i=0, j=1 | i=0, j=2 | i=j=1 | i=j=2 | | | | | |
| 100 | .142 | .139 | .144 | .145 | .173 | .092 | .1 | .175 | .3 |
| 150 | .102 | .118 | .108 | .113 | .141 | .086 | .082 | .152 | .245 |
| 200 | .099 | .088 | .116 | .114 | .123 | .074 | .071 | .169 | .212 |
| 250 | .079 | .083 | .083 | .091 | .110 | .060 | .063 | .113 | .190 |
| 300 | .075 | .088 | .066 | .074 | .1 | .060 | .058 | .125 | .173 |

Legend: $\{Appr.(4.6)\}^{\frac{1}{2}}=(1/n)^{\frac{1}{2}}$ as an approximation of $\{var[b_1(i,j)]\}^{\frac{1}{2}}$ for $i\neq 0$, $j\neq 0$, $i\neq j$;

$\{Appr.(4.7)\}^{\frac{1}{2}}=(9/n)^{\frac{1}{2}}$ as an approximation of $\{var[b_2(0,i,i)]\}^{\frac{1}{2}}$ for $i\neq 0$, $j\neq 0$, $i\neq j$;

$\{Appr.(4.8)\}^{\frac{1}{2}}=(3/n)^{\frac{1}{2}}$ as an approximation of $\{var[b_1(i,j)]\}^{\frac{1}{2}}$ for $0=i\neq j$ or $0\neq i=j$ or $0=j\neq i$.

The tentative conclusion that emerges from this table is that for $n \geq 200$ approximation (4.6) and (4.8) can be considered as satisfactory. Hence $b_1(i,j)$, in conjunction with the variance expressions (4.4), (4.6) and (4.8), may be used as an alternative statistic to (4.1) for detecting nonlinear types of dependence in the residuals of fitted time series models. This also suggests that the statistic $n\Sigma_{(i,j)}\{b_1(i,j)\}^2/c_{ij}$, where $c_{ij}=1$ if $i\neq 0$, $j\neq 0$, $i\neq j$; $c_{ij}=3$ if $0=i\neq j$ or $0\neq i=j$ or $0=j\neq i$; and $c_{ij}=6$ if $0=i=j$, may be a potentially useful diagnostic statistic for testing the dependence amongst the residuals. Since, however, we do not know its approximate distribution we shall not pursue this matter here further. Approximation (4.7) gives values which differ more substantially from the simulated results. Clearly, $b_2(0,i,i)$ with $i>0$ is a less useful statistic for testing residual dependencies.

To study the behaviour of $b_1(i,j)$ and $b_2(0,i,i)$ for bilinear models with standard normally distributed errors a large scale simulation experiment was carried out. Table 2 and 3, respectively, contain results of $\bar{b}_1(i,j)$ and $\bar{b}_2(0,i,i)$, based on 100 replications of length n=200, for the bilinear model $Y_t=.5Y_{t-\ell}A_{t-k}+A_t$ with $(k,\ell)=(1,2)$, $(1,1)$ and $(2,1)$. These results form a small though representative subset of many other bilinear models we simulated. In each simulation run 100 observations were discarded as a precaution to avoid possible "start-up" difficulties in the simulated series.

The simulated results exhibit a pattern similar to that of $\beta_1(i,j)$ and $\beta_2(0,i,i)$. From Theorem 3.1 and 3.3 it follows directly that the only non-zero values of $\beta_1(i,j)$ are at lag $(i,j)=(j,i)=(1,2)$ for the super and subdiagonal model. For standard normally distributed errors and $\beta=.5$ they are, respectively, given by $\beta_1(1,2)=.433$ for the superdiagonal model and $\beta_1(1,2)=.408$ for the subdiagonal model. The non-zero values of $\beta_1(i,j)$ for the diagonal model follow from Theorem 3.2. Among the non-zero are $\beta_1(0,0)=.756$, $\beta_1(0,1)=\beta_1(1,0)=.486$, $\beta_1(1,1)=.756$, $\beta_1(0,2)=\beta_1(2,0)=.148$ and $\beta_1(1,2)=\beta_1(2,1)=.054$. The values of the kurtosis for the super and subdiagonal model are, respectively, 3.462 and 5.205. From (3.10) we have for the diagonal model $\beta_2(0,0,0)=5.816$.

Table 2. Mean standardized third central sample moment $\bar{b}_1(i,j)$ from 100 length 200 simulations of the bilinear model $Y_t = .5Y_{t-\ell}A_{t-k} + A_t$ with $(k,\ell)=(1,2)$, $(1,1)$ and $(2,1)$.

| (k,ℓ) | Lag j | Lag i: 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|
| (1,2) | 0 | -.061 | | | | |
| | 1 | -.026 | -.026 | | | |
| | 2 | -.012 | .394* | -.035 | | |
| | 3 | -.093 | .022 | .001 | .034 | |
| | 4 | -.007 | .065 | -.009 | .051 | -.009 |
| (1,1) | 0 | .531* | | | | |
| | 1 | .695* | .362* | | | |
| | 2 | .120 | .121 | .068 | | |
| | 3 | .034 | -.021 | .013 | .021 | |
| | 4 | .018 | .001 | -.018 | -.064 | -.020 |
| (2,1) | 0 | -.073 | | | | |
| | 1 | -.054 | .057 | | | |
| | 2 | .006 | .546* | -.128 | | |
| | 3 | .007 | -.050 | -.008 | -.049 | |
| | 4 | -.089 | .034 | .123 | -.007 | .105 |

Note: * denotes statistically significant at the 5% level.

Table 3. Mean standardized fourth central sample moment $\bar{b}_2(0,i,i)$ from 100 length 200 simulations of the bilinear model $Y_t = .5Y_{t-\ell}A_{t-k} + A_t$ with $(k,\ell)=(1,2)$, $(1,1)$ and $(2,1)$.

| (k,ℓ) | Lag i: 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| (1,2) | 3.246 | 1.336 | 1.429 | 1.053 | .999 |
| (1,1) | 4.030 | 2.195 | 1.254 | .996 | .952 |
| (2,1) | 3.499 | 1.708 | 1.495 | 1.111 | .907 |

Clearly the estimates of the skewness and kurtosis are much smaller than might be expected from the theoretical results for the three simulated models. This was also noticed for all other simulated bilinear time series. It illustrates a characteristic of these series noticed earlier by, for example, Granger and Andersen (1978) and Maravall (1983). They found that stationary bilinear time series usually behave much like a linear series for a long period of time. However, on occasion, a different regime seems to operate on the series resulting in increased activity for short periods. This increased activity will show up in the estimated values of $\beta_1(0,0)$ and $\beta_2(0,0,0)$ only if the series is extremely long. In our case a sample size of length 200 is too short to adequately capture the full bilinear structure of the series. Indeed, when we reestimated $\beta_1(0,0)$ and $\beta_2(0,0,0)$ for series of length 800, using 30 replications, we obtained much better results. For instance, the estimates of the kurtosis for the super and subdiagonal model with $\beta=.5$ are then, respectively, given by 3.565 and 5.488. We may conclude from this that it will be difficult to accurately identify a particular model on the basis of the statistics (4.2) and (4.3) unless the series is much longer than is usually the case in economic time series analysis. In particular, it will be hard to distinguish superdiagonal models from subdiagonal models using these statistics for moderately large time series.

This last point was also confirmed by a blind discrimination experiment between 60 simulated bilinear time series. Each series of length 200 was generated by one of the three previously discussed bilinear models with $\beta=.5$. The choice of a model was determined by an unseen sequence of numbers 1, 2 and 3 which we randomly obtained at the start of the experiment. The mean and variance of the generated $\{A_t\}$ was set respectively to zero and one throughout the simulations.

Using the characteristic pattern of $\rho(i)$, $\beta_1(i,j)$ and $\beta_2(0,i,i)$ we tried to discriminate between the three bilinear models. Of the 60 simulated series, we correctly identified 43 series which represents about 72% success. Of the

18 series generated by the diagonal model $Y_t=.5Y_{t-1}A_{t-1}+A_t$, 16 were correctly identified, whilst the other 2 were incorrectly thought to be generated by the superdiagonal model $Y_t=.5Y_{t-2}A_{t-1}+A_t$. This successful identification was based on the typical pattern of $\beta_1(i,j)$ and the fact that $E(Y_t)\neq 0$ for diagonal models. By comparing the pattern of the sample statistics (4.1)-(4.3) with their theoretical counter parts we obtained 7 out of 19 correct identifications of the subdiagonal model $Y_t=.5Y_{t-1}A_{t-2}+A_t$. For the remaining 12 series, 10 were incorrectly identified as being generated by the superdiagonal model. Finally, for the superdiagonal model 20 correct decisions were achieved and 3 were incorrectly identified as coming from the subdiagonal model.

Although the scope of this simulation experiment is rather limited the results seem to suggest that correct discrimination between series of length 200 generated by a super and subdiagonal model, using (4.2) and (4.3), is doubtful. There is hardly any marked difference in the behaviour of these statistics for these two bilinear models which makes them not very useful for model discrimination. Perhaps by estimating both a super and subdiagonal model one could more easily tell which model generated the data. On the other hand, the results of the simulation experiment also indicate that a quick and relative accurate method for distinguishing a diagonal model from a non-diagonal model can be based on the set of non-zero values of the statistic $b_1(i,j)$.

In economics the assumption that the errors of a time series model follow a Gaussian distribution with fixed mean and variance is often not very reasonable. For instance, it has been pointed out in the literature that stock price data are generated by models with errors coming from leptokurtic distributions. We now investigate the effect of non-Gaussian distributed $\{A_t\}$ on the value of the kurtosis $\beta_2(0,0,0)$ for various bilinear models. For this purpose we assume that the distribution function of the errors is a member of the family of symmetric exponential power distributions centered around mean zero. The probability density function of this class of distributions may be written in the general form

$$p(A) = \omega(\gamma)\sigma^{-1}\exp\{-c(\gamma)\sigma^{-2/(1+\gamma)}|A|^{2/(1+\gamma)}\} \qquad (4.11)$$

$$-\infty < A < \infty;\ -1<\gamma\leq 1;\ \sigma>0$$

where

$$\omega(\gamma) = \frac{\{\Gamma[\frac{3}{2}(1+\gamma)]\}^{\frac{1}{2}}}{(1+\gamma)\{\Gamma[\frac{1}{2}(1+\gamma)]\}^{3/2}} \quad \text{and} \quad c(\gamma) = \left\{\frac{\Gamma[\frac{3}{2}(1+\gamma)]}{\Gamma[\frac{1}{2}(1+\gamma)]}\right\}^{1/(1+\gamma)}.$$

Its ith moment is given by

$$\mu_i = E(A_t^i) = \begin{cases} 0 & \text{if } i \text{ is odd,} \\ \dfrac{\sigma^i\{\Gamma[\frac{1}{2}(1+\gamma)]\}^{(i-2)/2}\Gamma[\frac{1}{2}(i+1)(1+\gamma)]}{\{\Gamma[\frac{3}{2}(1+\gamma)]\}^{i/2}} & \text{if } i \text{ is even.} \end{cases} \qquad (4.12)$$

Here, the parameter $\sigma$ denotes the standard deviation of the population whereas the parameter $\gamma$ can be regarded as a measure of the kurtosis indicating the extent of non-Gaussianity of the parent distribution.

Many well-known symmetric distributions are a member of the class of symmetric exponential power distributions. It includes the uniform distribution when $\gamma \to -1$, the normal distribution when $\gamma=0$ and the double exponential distribution when $\gamma=1$. Hence, it covers symmetric leptokurtic ($\gamma>0$) as well as symmetric platikurtic ($\gamma<0$) distributions. Various shapes of the symmetric exponential power distribution are given by Box and Tiao (1973, Fig. 3.2.3) who use this family of distributions extensively in a Bayesian context. Since the shape of the distribution of many economic time series is leptokurtic we will only consider the range of parameter values $0\leq\gamma\leq 1$.

Table 4 displays the values of $\beta_2(0,0,0)$ for the model $Y_t=\beta Y_{t-\ell}A_{t-k}+A_t$ with $(k,\ell)=(1,2)$, $(1,1)$ and $(2,1)$ for various values of $|\beta|$ and $\gamma$. In all computations the parameter $\sigma$ was, without loss of generality, set at unity. Notice that for the range of values $0\leq\gamma\leq .4$ the symmetrically distributed $\{A_t\}$ produces a series $\{Y_t\}$ which is distributed not too far from Gaussianity

Table 4. Values of the kurtosis $\beta_2(0,0,0)$ of the series $\{Y_t\}$ generated by the bilinear model $Y_t=\beta Y_{t-\ell}A_{t-k}+A_t$ with $(k,\ell)=(1,2)$, (1,1) and (2,1) and with $\{A_t\}$ independent drawings from the family of symmetric exponential power distributions with characteristic parameter $\gamma$.

| $\gamma$ | $(k,\ell)$ | $\|\beta\|$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 0 | .12 | .24 | .36 | .48 | .60 |
| 0 | (1,2) | 3 | 3.00 | 3.02 | 3.11 | 3.38 | 4.27 |
| | (1,1) | 3 | 3.01 | 3.15 | 3.74 | 5.36 | 9.79 |
| | (2,1) | 3 | 3.00 | 3.03 | 3.29 | 4.69 | 10.88 |
| .2 | (1,2) | 3.42 | 3.41 | 3.40 | 3.47 | 3.77 | 5.02 |
| | (1,1) | 3.42 | 3.41 | 3.63 | 4.66 | 7.36 | 14.91 |
| | (2,1) | 3.42 | 3.41 | 3.42 | 3.78 | 6.23 | 18.28 |
| .4 | (1,2) | 3.94 | 3.92 | 3.87 | 3.91 | 4.29 | 6.12 |
| | (1,1) | 3.94 | 3.90 | 4.32 | 6.16 | 10.71 | 23.86 |
| | (2,1) | 3.94 | 3.91 | 3.89 | 4.48 | 9.15 | 34.43 |
| .6 | (1,2) | 4.53 | 4.48 | 4.41 | 4.44 | 4.92 | 7.83 |
| | (1,1) | 4.53 | 4.47 | 5.30 | 8.61 | 16.18 | 40.29 |
| | (2,1) | 4.53 | 4.48 | 4.45 | 5.54 | 14.78 | 70.74 |
| .8 | (1,2) | 5.21 | 5.15 | 5.04 | 5.06 | 5.73 | 10.82 |
| | (1,1) | 5.21 | 5.15 | 6.82 | 12.65 | 25.16 | 73.47 |
| | (2,1) | 5.21 | 5.14 | 5.11 | 7.29 | 26.05 | 160.29 |
| 1 | (1,2) | 6 | 5.92 | 5.77 | 5.81 | 6.83 | 18.92 |
| | (1,1) | 6 | 5.99 | 9.36 | 19.63 | 40.90 | 170.80 |
| | (2,1) | 6 | 5.92 | 5.93 | 10.41 | 49.39 | 424.02 |

depending on the value of $|\beta|$. Hence, for these parameter values no severe problems will arise, due to the non-Gaussianity of the residuals, when estimating the parameters of a bilinear model by a maximum likelihood procedure. Also it can be noticed that for values of $\beta\leq .48$ and $\gamma\leq .4$ the departure from Gaussianity is more significant for the diagonal model than for the other two models.

From (4.12) it follows that the coefficient of kurtosis of the $\{A_t\}$ distribution is given by

$$\beta_2(0,0,0) = \Gamma[2\tfrac{1}{2}(1+\gamma)]\Gamma[\tfrac{1}{2}(1+\gamma)]/\{\Gamma[1\tfrac{1}{2}(1+\gamma)]\}^{\frac{1}{2}}. \tag{4.13}$$

Knowing the value of $\beta$ one can approximately deduce from the results in Table 4 the appropriate value of the kurtosis (4.13) that need to be imposed on the errors in order to generate a bilinear time series $\{Y_t\}$ with a given kurtosis. This can be done along similar lines as has been proposed by Davies, Spedding and Watson (1980) for ARMA models with non-Gaussian residuals.

## 5. An example

Consider the coal production series given by Pankratz (1983, Case 3). The series represents the monthly bituminous coal production in the United States from January 1952 through December 1959, a total of 96 observations. The data have been seasonally adjusted and from the results presented by Pankratz it can be concluded that they do not contain a seasonal pattern. When reanalysing this series we arrived at the AR(2) model specification

$$Y_t = \underset{(3.13)}{7576.84} + \underset{(4.90)}{.49}Y_{t-1} + \underset{(3.01)}{.31}Y_{t-2} + A_t, \qquad \hat{\sigma}_a^2 = 9,054,181 \tag{5.1}$$

where $\hat{\sigma}_a^2$ is the residual variance and where t-ratios are standing between parentheses. The parameter values are reasonably close to those presented by Pankratz for this model. The predictability of this model can be expressed as $R^2=1-(\hat{\sigma}_a^2/\hat{\sigma}_y^2)$, where $\hat{\sigma}_y^2$ is the sample variance of the series $\{Y_t\}$, and is equal to .547.

Table 5. Autocorrelations $r_a(i)$ and $r_a^*(i)$ of, respectively, $\{\hat{A}_t\}$ and $\{\hat{A}_t^2\}$ from model (5.1) fitted to the coal production series.

| | Lag i | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| $r_a(i)$ | -.078 | -.015 | .155 | -.017 | .016 | -.012 |
| $r_a^*(i)$ | .361 | -.018 | .097 | .086 | -.041 | -.044 |

In Table 5 the autocorrelations of the residuals $\{\hat{A}_t\}$ and $\{\hat{A}_t^2\}$, respectively, denoted by $r_a(i)$ and $r_a^*(i)$, are presented for lags $i=1,2,\ldots,6$. The values of $r_a(i)$ suggest that the residuals $\{\hat{A}_t\}$ are uncorrelated. This is also supported by the modified $\chi^2(M-p-q)$ diagnostic statistic $Q_a=n(n+2)\Sigma_{i=1}^{M}r_a^2(i)/(n-i)$ which for M=20 is equal to 7.42. Looking at the values of $r_a^*(i)$ we notice that, with the exception of $r_a^*(1)$, all residual autocorrelations of $\{\hat{A}_t^2\}$ are not significantly different from zero at the 5% level. The value of $r_a^*(1)$ may indicate evidence of some nonlinear relation in the $\{\hat{A}_t\}$. However, this is not confirmed by the value of the diagnostic statistic $Q_a^*=n(n+2)\Sigma_{i=1}^{M}\{r_a^*(i)\}^2/(n-i)$ suggested by McLeod and Li (1983) for detecting autocorrelations in the squared residuals. For M=20 the value of this statistic is 16.07 which is less than the 5% significance point of the $\chi^2(M)$ distribution. One reason for this could be that $Q_a^*$ is computed on the basis of only 94 residuals which, as explained in the previous section, may be insufficient for identifying bilinear relations in the data. Therefore it is worthwhile to take a further look at the values of the statistics $b_1(i,j)$ and $b_2(0,i,i)$ to see whether they can provide some useful information about the process underlying the residuals.

The estimated values of $b_1(i,j)$ of the residuals $\{\hat{A}_t\}$ from model (5.1) are presented in Table 6. Using the variance expressions (4.4), (4.6) and (4.8) we notice significant values of $b_1(i,j)$ at lag $(i,j)=(0,1)$, $(0,3)$, $(1,2)$ and $(3,3)$ at the 5% level. The coefficient of kurtosis is equal to 4.097 while for lag i=1,2,3 and 4 the values of $b_2(0,i,i)$ are, respectively, given by 2.330, .999, 1.317 and 1.060. In general the distribution of $\{\hat{A}_t\}$ has a rather symmetric though slightly leptokurtic shape.

Table 6. Estimated values of $b_1(i,j)$ of $\{\hat{A}_t\}$ from model (5.1) fitted to the coal production series.

| Lag | Lag i | | | | |
|---|---|---|---|---|---|
| j | 0 | 1 | 2 | 3 | 4 |
| 0 | .209 | | | | |
| 1 | -.441* | .281 | | | |
| 2 | .309 | -.245* | -.075 | | |
| 3 | .372* | -.120 | .139 | -.386* | |
| 4 | -.065 | .081 | -.128 | .228 | -.345 |

Note: * denotes statistically significant at the 5% level.

It is evident from the results in Table 6 that the residuals have significant nonlinear properties. This suggests that model (5.1) could be improved by adding a bilinear term to the linear system. The super and subdiagonal model are both characterized by a theoretical pattern of $\beta_1(i,j)$ which is non-zero for lag $(i,j)=(k,\ell)$ and zeroes elsewhere. On the basis of the significant values of $b_1(i,j)$ a diagonal model seems more likely. However, this conclusion is somewhat tentative since the pattern of significant values of $b_1(i,j)$ does not fully conform the theoretical pattern of $\beta_1(i,j)$ for the diagonal model discussed in Subsection 3.2.

In order to check this last conjecture we decided to estimate the model $A_t=\beta A_{t-\ell}U_{t-k}+U_t$ with $\{k,\ell=1,2$ and $3\}$ and where $\{U_t\}$ is a sequence of i.i.d. random variables with mean zero and fixed variance. A close approximation to the maximum likelihood estimate of the parameter $\beta$ in this model can be obtained by minimizing the sum of squares $S(\beta)=\Sigma_t U_t^2$ over a grid of admissible parameter values. After standardizing the series $\{A_t\}$ and setting the initial values of $\{U_t\}$ equal to zero to avoid starting up problems, the minimum sum of squares was reached for the diagonal model $A_t=-.19A_{t-1}U_{t-1}+U_t$. In terms of the original series $\{Y_t\}$, model (5.1) augmented by this fitted diagonal model gives rise to a predictability measure $R^2$ equal to .589, which is an increase of more than 4% compared with the forecasting ability of model (5.1). The autocorrelations of $\{\hat{U}_t\}$ and $\{\hat{U}_t^2\}$ for lags 1 through 6 were not significantly different from zero at the 5% level which was also confirmed by the McLeod-Li (1983) diagnostic statistic. The values of $b_1(i,j)$ of the new residuals $\{\hat{U}_t\}$ were all very small, except from significant values at lag $(i,j)=(0,0)$ and $(i,j)=(0,3)$. Hence, practically all of the nonlinearity can be explained by the above simple diagonal bilinear model.

From this example it may be concluded that there are situations in practice where the autocorrelations of the squared and unsquared residuals do not provide sufficient information about the existence and type of nonlinearity present in the data. We feel that the results presented here demonstrate that the statistics $b_1(i,j)$ and $b_2(0,i,i)$ may be used as a useful additional tool to identify a particular bilinear time series model. However, more work is needed to solve the problem of discriminating between super and subdiagonal models on the basis of these statistics. Also more attention should be directed to the problem that bilinear time series models can only be well identified when the sample size is extremely large.

## References

BOX, G.E.P., and TIAO, G.C. (1973), *Bayesian Inference in Statistical Analysis*, Reading: Addison-Wesley.

DAVIES, N., SPEDDING, T., and WATSON, W. (1980), "Autoregressive moving average processes with non-normal residuals", *Journal of Time Series Analysis*, 2, 103-109.

GOOIJER, J.G. de, ABRAHAM, B., GOULD, A., and ROBINSON, L. (1985), "Methods for determining the order of an autoregressive moving average process: A survey", *International Statistical Review*, 53, 301-329.

GRANGER, C.W.J., and ANDERSEN, A.P. (1978), *An Introduction to Bilinear Time Series Models*, Göttingen: Vandenhoeck & Ruprecht.

GUEGAN, D. (1984), "Test de modèles non linéaires", in *Altenative Approaches to Time Series Analysis*, eds. J.P. Florens, M. Mouchart, J.P. Raoult and L. Simar, Bruxelles: Facultés Universitaires Saint-Louis, 45-77.

HINICH, M.J., and PATTERSON, D.M. (1985), "Evidence of nonlinearity in daily stock returns", *Journal of Business & Economic Statistics*, 3, 69-77.

KENDALL, M.G., and STUART, A. (1969), *The Advandced Theory of Statistics*, Vol. 1, London: Charles Griffin & Company Ltd., third ed.

LAWRANCE, A.J., and LEWIS, P.A.W. (1985), "Modelling and residual analysis of nonlinear autoregressive time series in exponential variables", *Journal of the Royal Statistical Society*, B, 47, 165-202.

LI, W.K. (1984), "On the autocorrelation structure and identification of some bilinear time series", *Journal of Time Series Analysis*, 5, 173-181.

MARAVALL, A. (1983), "An application of nonlinear time series forecasting", *Journal of Business & Economic Statistics*, 1, 66-74.

McLeod, A.I. and LI, W.K. (1983), "Diagnostic checking ARMA time series models using squared-residual autocorrelations", *Journal of Time Series Analysis*, 4, 269-273.

PANKRATZ, A. (1983), *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*, New York: John Wiley.

QUINN, B.G. (1982), "Stationarity and invertibility of simple bilinear models", *Stochastic Processes and their Applications*, 12, 225-230.

SUBBA RAO, T., and GABR, M.M. (1984), *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, New York: Springer-Verlag.

SUN, T-C (1963), "A central limit theorem for non-linear functions of a normal stationary process", *Journal of Mathematics & Mechanics*, 12, 945-977.

TONG, H., and LIM, K.S. (1980), "Threshold auto-regression, limit cycles and cyclical data", *Journal of the Royal Statistical Society*, B, 42, 245-292.

**IN 1986 REEDS VERSCHENEN**

202 J.H.F. Schilderinck
Interregional Structure of the European Community. Part III

203 Antoon van den Elzen and Dolf Talman
A new strategy-adjustment process for computing a Nash equilibrium in a noncooperative more-person game

204 Jan Vingerhoets
Fabrication of copper and copper semis in developing countries. A review of evidence and opportunities

205 R. Heuts, J. van Lieshout, K. Baken
An inventory model: what is the influence of the shape of the lead time demand distribution?

206 A. van Soest, P. Kooreman
A Microeconometric Analysis of Vacation Behavior

207 F. Boekema, A. Nagelkerke
Labour Relations, Networks, Job-creation and Regional Development. A view to the consequences of technological change

208 R. Alessie, A. Kapteyn
Habit Formation and Interdependent Preferences in the Almost Ideal Demand System

209 T. Wansbeek, A. Kapteyn
Estimation of the error components model with incomplete panels

210 A.L. Hempenius
The relation between dividends and profits

211 J. Kriens, J.Th. van Lieshout
A generalisation and some properties of Markowitz' portfolio selection method

212 Jack P.C. Kleijnen and Charles R. Standridge
Experimental design and regression analysis in simulation: an FMS case study

213 T.M. Doup, A.H. van den Elzen and A.J.J. Talman
Simplicial algorithms for solving the non-linear complementarity problem on the simplotope

214 A.J.W. van de Gevel
The theory of wage differentials: a correction

215 J.P.C. Kleijnen, W. van Groenendaal
Regression analysis of factorial designs with sequential replication

216 T.E. Nijman and F.C. Palm
Consistent estimation of rational expectations models

217 P.M. Kort
The firm's investment policy under a concave adjustment cost function

218 J.P.C. Kleijnen
Decision Support Systems (DSS), en de kleren van de keizer ...

219 T.M. Doup and A.J.J. Talman
A continuous deformation algorithm on the product space of unit simplices

220 T.M. Doup and A.J.J. Talman
The 2-ray algorithm for solving equilibrium problems on the unit simplex

221 Th. van de Klundert, P. Peters
Price Inertia in a Macroeconomic Model of Monopolistic Competition

222 Christian Mulder
Testing Korteweg's rational expectations model for a small open economy

223 A.C. Meijdam, J.E.J. Plasmans
Maximum Likelihood Estimation of Econometric Models with Rational Expectations of Current Endogenous Variables

224 Arie Kapteyn, Peter Kooreman, Arthur van Soest
Non-convex budget sets, institutional constraints and imposition of concavity in a flexible household labor supply model

225 R.J. de Groof
Internationale coördinatie van economische politiek in een twee-regio-twee-sectoren model

226 Arthur van Soest, Peter Kooreman
Comment on 'Microeconometric Demand Systems with Binding Non-Negativity Constraints: The Dual Approach'

227 A.J.J. Talman and Y. Yamamoto
A globally convergent simplicial algorithm for stationary point problems on polytopes

228 Jack P.C. Kleijnen, Peter C.A. Karremans, Wim K. Oortwijn, Willem J.H. van Groenendaal
Jackknifing estimated weighted least squares

229 A.H. van den Elzen and G. van der Laan
A price adjustment for an economy with a block-diagonal pattern

230 M.H.C. Paardekooper
Jacobi-type algorithms for eigenvalues on vector- and parallel computer

231 J.P.C. Kleijnen
Analyzing simulation experiments with common random numbers

232 A.B.T.M. van Schaik, R.J. Mulder
On Superimposed Recurrent Cycles

233 M.H.C. Paardekooper
Sameh's parallel eigenvalue algorithm revisited

234 Pieter H.M. Ruys and Ton J.A. Storcken
Preferences revealed by the choice of friends

235 C.J.J. Huys en E.N. Kertzman
Effectieve belastingtarieven en kapitaalkosten

236 A.M.H. Gerards
An extension of König's theorem to graphs with no odd-$K_4$

237 A.M.H. Gerards and A. Schrijver
Signed Graphs - Regular Matroids - Grafts

238 Rob J.M. Alessie and Arie Kapteyn
Consumption, Savings and Demography

239 A.J. van Reeken
Begrippen rondom "kwaliteit"

240 Th.E. Nijman and F.C. Palmer
Efficiency gains due to using missing data. Procedures in regression models

241 Dr. S.C.W. Eijffinger
The determinants of the currencies within the European Monetary System

IN 1987 REEDS VERSCHENEN

242 Gerard van den Berg
Nonstationarity in job search theory

243 Annie Cuyt, Brigitte Verdonk
Block-tridiagonal linear systems and branched continued fractions

244 J.C. de Vos, W. Vervaat
Local Times of Bernoulli Walk

245 Arie Kapteyn, Peter Kooreman, Rob Willemse
Some methodological issues in the implementation
of subjective poverty definitions

246 J.P.C. Kleijnen, J. Kriens, M.C.H.M. Lafleur, J.H.F. Pardoel
Sampling for Quality Inspection and Correction: AOQL Performance Criteria

247 D.B.J. Schouten
Algemene theorie van de internationale conjuncturele en strukturele afhankelijkheden

248 F.C. Bussemaker, W.H. Haemers, J.J. Seidel, E. Spence
On $(v,k,\lambda)$ graphs and designs with trivial automorphism group

249 Peter M. Kort
The Influence of a Stochastic Environment on the Firm's Optimal Dynamic Investment Policy

250 R.H.J.M. Gradus
Preliminary version
The reaction of the firm on governmental policy: a game-theoretical approach