A Conditional Least Squares approach
to
bilinear time series estimation

by T. Grahn
Institute of Applied Mathematics
University of Heidelberg

Address:
Dr.T. Grahn
(z. Hd. Frau Neubauer)
Universität Heidelberg
Institut für Angewandte Mathematik
INF 294
D - 6900 Heidelberg

A Conditional Least Squares approach
to
bilinear time series estimation

by T. Grahn
University of Heidelberg

Abstract. In this paper we develop a Conditional Least Squares (CLS) procedure for estimating bilinear time series models. We apply this method to two general types of bilinear models. A model of type I is a special superdiagonal bilinear model which includes the linear ARMA model as a submodel. A model of type II is a standardized version of the popular bilinear BL(p, 0, p, 1) model (see e.g. Liu and Chen (1990), Sesay and Subba Rao (1991)). For both models we show that the limiting distribution of the resulting CLS estimates is Gaussian and the law of the iterated logarithm holds.

Keywords. Estimation; bilinear time series; central limit theorem; law of the iterated logarithm; conditional moments.

1   Introduction

In the last 20 years there has been a growing interest in the nonlinear modelling of time series data (for a review see e.g. Andel (1989), Tong (1990) and Tjøstheim (1992)). This interest is motivated mainly by the fact that a lot of real data sets exhibit features e.g. occasional sharp spikes, which cannot sufficiently explained by the theory of linear time series.

One of the approaches to nonlinear time series modelling is the bilinear one. In the context of time series, bilinear models have been introduced by Granger and Andersen (1978). A real valued stochastic process in discrete time $\{X_t\}_{t\in\mathbf{Z}}$ is said to be a *bilinear time series of type BL(p, q, k, r)* if it satisfies the difference equations

$$X_t = \alpha + \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} c_j \varepsilon_{t-j} + \sum_{i=1}^{k}\sum_{j=1}^{r} b_{ij}\, \varepsilon_{t-i} X_{t-j} + \varepsilon_t, \qquad (1.1)$$

for each $t \in \mathbf{Z}$, where $\{\varepsilon_t\}_{t\in\mathbf{Z}}$ is an iid zero mean white noise process.

Such a bilinear time series is particularly attractive in modelling processes with sample paths exhibiting occasional sharp spikes as they occur in vibration wave, econometrics and control theory (see e.g. Mohler (1973)).

2

Here we will present a new approach to solve the estimation problem for two submodels of the general bilinear model defined in (1.1). The model of our major interest is the special *bilinear extended ARMA model* of the type

$$\text{(I)} \qquad X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} c_j \varepsilon_{t-j} + \sum_{i=1}^{k} \sum_{j=w}^{r} b_{ij} \varepsilon_{t-i} X_{t-j} + \varepsilon_t, \qquad (1.2)$$

with $w \equiv \max \{q, k\} + 1$.

The other model which we consider here, is one of the most often considered bilinear models in the literature. It is defined by the difference equations

$$\text{(II)} \qquad X_t = \alpha + \sum_{i=1}^{p} a_i X_{t-i} + \varepsilon_{t-k} \sum_{j=k}^{r} b_{kj} X_{t-j} + \varepsilon_t, \ k \geq 1, \qquad (1.3)$$

with $b_{kj} \neq 0$ for at least one j, $k \leq j \leq r$, $\alpha \equiv - b_{kk}\sigma^2$ and $\varepsilon_t \sim N(0, \sigma^2)$.

For a stationary and ergodic bilinear time series many different ideas to solve the estimation problem have been proposed in the literature (see e.g. Subba Rao (1981), Pham and Tran (1981), Subba Rao and da Silva (1989), Liu (1990), Liu and Chen (1990), Sesay and Subba Rao (1992)), but the asymptotic properties of the deduced estimates are known in some very restrictive models only. The approaches differ not only in the method of estimation they use but also in the type of bilinearity they can capture and in the results which are attainable. The most frequently applied methods for estimating the parameter of the bilinear process are the Least Squares (LS) method (e.g. Pham and Tran (1981), Guegan and Pham (1989)) and the Method of Moments (MM) (e.g. Tang and Mohler (1988), Kim, Billard and Basawa (1990), Liu and Chen (1990)).

In the case of LS estimation up to now even in the simple bilinear model considered by Pham and Tran (1981) nothing is known about the limiting distribution of the LS estimates. Simulations of the distribution of the estimates sometimes indicate an asymptotic normal behaviour and sometimes they do not (cf. Kim and Billard (1990) and Subba Rao and da Silva (1989) respectively).

If we restrict our considerations to a Gaussian innovation process, in most of the cases where we fail to prove an asympotic normal behaviour of the LS estimates, the

application of the MM method yields to strong consistent and asymptotic normal estimates (e.g. Kim, Billard and Basawa (1990), Liu and Chen (1990)). But until now the bilinear BL(p, 0, 1, 1) model is the most general bilinear model which has been shown to be asymptotic normal estimable (in Liu and Chen (1990)).

Here we propose a new method for estimating the parameter of bilinear time series which is based mainly on a *Conditional Least Squares (CLS) approach* applied to the AR residuals of the bilinear process.

We will show that this procedure supplies estimates for the parameter of both models (1.2) and (1.3), where the method of estimating the model II is applicable also to the case of a more general innovation sequence $\{\varepsilon_t\}_{t \in \mathbf{Z}}$ whose first eight cumulants coincide with that of the Gaussian distribution $N(0, \sigma^2)$. We will prove that as well in model I as in model II the estimators defined by means of this CLS method satisfy both the Central Limit Theorem (CLT) and the Law of the Iterated Logarithm (LIL).

In view of these results it is possible for the first time that we can estimate the parameter of a bilinear time series which includes a non zero moving average part (i.e. $c_j \neq 0$ for some j in (1.2)) by an asymptotic normal estimate. Furthermore, in estimating the model (1.2) we do not need to impose the common assumption of Gaussian innovations neither for the definition of the estimation procedure nor for proving the asymptotic properties of the estimates. Up to now in the literature for proving an asymptotic normal behaviour of bilinear estimates it is usually required that the innovations follow a Gaussian law (e.g. Kim and Billard (1990), Kim, Billard and Basawa (1990)). Thus the results concerning the estimation of the bilinear model in (1.2) obtained here, are fundamental theoretical results for improving the general linear (Gaussian or nongaussian) ARMA approximation to time series data by applying this bilinear extension. It is our hope that basing on our results, in practice in many cases forecasts deduced from the usual ARMA fit can be improved by fitting the bilinear extended ARMA model (1.2) to the data.

Furthermore, we are not aware of any comparable results in estimating the model (1.3), which state the CLT or the LIL for the estimates under consideration, though this model is one of the most often considered bilinear models (mainly in the case $\alpha = 0$) in the literature (e.g. Subba Rao and Gabr (1984), Jou (1989), Wittwer (1989), Sesay and Subba Rao (1988, 1991)).

To prove our results we need to impose some appropriate moment conditions on the

process $\{X_t\}_{t \in \mathbf{Z}}$ and a kind of 'linear' invertibility of the underlying bilinear process. The required moment conditions are weak since we can prove the strong consistency of the suggested estimates under the assumption of the existence of the fourth moment of the process $\{X_t\}_{t \in \mathbf{Z}}$ only. In a certain sense this is a minimum requirement since even in the case of the simple bilinear model

$$X_t = b\, \varepsilon_{t-1} X_{t-2} + \varepsilon_t, \qquad \varepsilon_t \text{ iid} \sim N(0,\sigma^2), \ \ t \in \mathbf{Z}. \qquad (1.4)$$

the fourth order mixed moments are necessary to identify the parameter (b, σ^2) from the moments. Note the following result whose proof is given in chapter 4:

Lemma 1.1

Let $\{X_t\}_{t \in \mathbf{Z}}$ denote the bilinear time series defined by the equation $X_t = b\, \varepsilon_{t-1} X_{t-2} + \varepsilon_t$, where $\{\varepsilon_t\}_{t \in \mathbf{Z}}$ denotes an iid sequence with $E(\varepsilon_t^{2n}) < \infty$ for some $n \geq 1$, and $E(\varepsilon_t^m) = 0$ for all $1 \leq m \leq 2n - 1$, m odd. Then we have the following assertions:

(i) *If $b^2\, E(\varepsilon_t^2) < 1$, then $\{X_t\}_{t \in \mathbf{Z}}$ is a stationary, ergodic and causal process.*

(ii) *$E(X_t^{2n}) < \infty \quad \Leftrightarrow \quad b^{2n}\, E(\varepsilon_t^{2n}) < 1$.*

(iii) *In the special case of Gaussian innovations with $1/4 < b^4\sigma^4 < 1/3$ we have: The bilinear process $\{\widetilde{X}_t\}_{t \in \mathbf{Z}}$ defined by $\widetilde{X}_t = \widetilde{b}\, \widetilde{\varepsilon}_{t-1} \widetilde{X}_{t-2} + \widetilde{\varepsilon}_t$ with an iid Gaussian sequence $\{\widetilde{\varepsilon}_t\}_{t \in \mathbf{Z}}$ where $\widetilde{b} := (1/b\sigma^2) - b$, $\widetilde{\sigma^2} := (b\sigma^2)/\widetilde{b}$ and σ^2 and $\widetilde{\sigma^2}$ respectively denotes the variance of the innovations, is a stationary, ergodic and causal process. Moreover the fourth moment of the process exists and the first, second and third order moments of the processes $\{X_t\}_{t \in \mathbf{Z}}$ and $\{\widetilde{X}_t\}_{t \in \mathbf{Z}}$ coincide.*

The assertion (iii) makes the identification problem obvious. Even in the Gaussian innovation case we require the first four moments of the process to identify the parameter determining the bilinear process by its moments only. If we apply the assertion (ii) of this lemma to the bilinear model (1.4) we see that the requirement $E|X_t^4| < \infty$ is equivalent to $b^4\sigma^4 < 1/3$ and $E|X_t^8| < \infty$ corresponds to $b^8\sigma^8 < 1/105$.

The paper is organized as follows:

In chapter 2 the CLS procedure for estimating the parameter of the bilinear models (1.2) and (1.3) is presented explicitly.

In the third chapter the results on the asymptotic behaviour of these parameter estimates are summarized.

The main proofs are given in the fourth chapter then.

2 Conditional Least Squares estimation of bilinear time series

To motivate the application of the CLS method to bilinear models first we restrict our attention on the bilinear model (1.4).

One of the most often considered methods for estimating bilinear models is the method of Least Squares (LS) (e.g. Pham and Tran (1981), Wittwer (1989)). In the Gaussian innovation case this method coincides with the method of Conditional Maximum Likelihood estimation.

Given observations $X_1,\ldots, X_N$ generated by the difference scheme in (1.4) according to the concept of LS we try to minimize the sum

$$\sum_{t=3}^{N} \{\varepsilon_t(b)\}^2 = \sum_{t=3}^{N} \{X_t - b\,\varepsilon_{t\text{-}1}(b)\,X_{t\text{-}2}\}^2 \tag{2.1}$$

with respect to b which provides us with an estimator $\widehat{b}_{LS}$ of the parameter *b*.

In the literature it is well known that there exist at least three problems in applying this method. The first is that we need to ensure the invertibility of the bilinear process, because otherwise the residuals $\varepsilon_{t\text{-}1}(b)$ in (2.1) cannot be estimated from the observations $X_1,\ldots, X_N$ only. To guarantee the invertibility of the process we have to restrict the space of the admissable parameter values in a complicated manner (see e.g. Pham and Tran (1981)). A further problem is that to make inference on the limiting distribution of the standardized quantity $N^{1/2}\,(\widehat{b}_{LS} - b)$ we have to require that all moments of the bilinear process exist. In model (1.4) this requirement forces b to be zero (cf. Lemma 1.1)!

Apart from these problems arising in the deduction of the asymptotic behaviour of the LS estimate we are also faced with finite sample problems. For minimizing the sum of the least squares we need a numerical minimization procedure which requires an initial value placed in a neighbourhood of the value of the true parameter. To find such an initial value is also far from being trivial.

Minimizing the sum in (2.1) can be interpreted as searching for that parametervalue which minimizes the empirical expectation of the quadratic distance between X_t and its forecast $E(X_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ which is based on the information contained in the past innovations ε_{t-i}, $i \geq 1$. (If we assume the process $\{X_t\}_{t \in \mathbf{Z}}$ to be invertible, the σ - field generated by ε_{t-i}, $i \geq 1$, and that generated by X_{t-i}, $i \geq 1$, coincide.) If we relinquish the information contained in the recent innovations we can forecast X_t by $E(X_t \mid \varepsilon_{t-i}, \varepsilon_{t-i-1}, \ldots)$ for $i > 1$ too. But for each $i > 1$ we get $E(X_t \mid \varepsilon_{t-i}, \varepsilon_{t-i-1}, \ldots) = 0 \; (= E(X_t))$ such that each of these forecasts is only as good as the best forecast of X_t if no more information is available. Moreover they do not contain any information on the parameter b and σ^2 determining the bilinear series.

The idea now is to consider not only forecasting X_t but to study also forecasting products like $X_t X_{t-s}$, $s \geq 0$ with the hope that these forecasts contain more information on the underlying parameter. Applying this idea to model (1.4) we obtain

$$E(X_t X_{t-s} \mid \varepsilon_{t-2}, \varepsilon_{t-3}, \ldots) = \begin{cases} \sigma^2 + b^2 \sigma^2 X_{t-2}^2 , & s = 0 \\ b\sigma^2 X_{t-2} , & s = 1 \\ 0 , & s \geq 2 \end{cases} \tag{2.2}$$

This result now motivates the following procedure to estimate the parameter b and σ^2 determining the bilinear model in (1.4). First we apply the equation in (2.2) for $s = 0$ and try to minimize

$$\sum_{t=3}^{N} \{X_t^2 - (\beta_{11} + \beta_{12} X_{t-2}^2)\}^2$$

in $\underline{\beta}_1 = (\beta_{11}, \beta_{12})^t$ to get an estimate $\underline{\hat{\beta}}_1 = (\hat{\beta}_{11}, \hat{\beta}_{12})^t$ of $(\sigma^2, b^2 \sigma^2)^t$. And analogously we apply the equation (2.2) for $s = 1$ also and try to minimize

$$\sum_{t=3}^{N} \{X_t X_{t-1} - \beta_2 X_{t-2}\}^2$$

in β_2 to get an estimate $\hat{\beta}_2$ of $b\sigma^2$.

By taking these estimators into account in a second step we can define estimates of the parameter b and σ^2 by $\hat{b} \equiv \hat{\beta}_2 / \hat{\beta}_{11}$ and $\hat{\sigma^2} \equiv \hat{\beta}_{11}$ respectively. Due to the conditioning in (2.2) these estimates are called Conditional Least Squares (CLS) estimators.

We will show that these estimates are strong consistent estimates provided $E(X_t^4) < \infty$. Moreover, if also $E(X_t^8) < \infty$ then the limiting distribution of the vector $N^{1/2} (\hat{b} - b, \widehat{\sigma^2} - \sigma^2)^t$ follows a centred Gaussian law.

The idea to analyze the second order conditional moments of the bilinear process to define CLS estimators of the underlying parameter can also be applied to the more general bilinear models (1.2) and (1.3).

In these more general bilinear models first we estimate the AR coefficients $a_1,\ldots, a_p$ of the process separately and afterwards we apply the CLS approach to the process $\{v_t\}_{t \in \mathbf{Z}}$ of the AR residuals defined by $v_t \equiv X_t - \sum_{i=1}^{p} a_i X_{t-i}$.
So in the first step of the estimation procedure we take the Yule - Walker equations into account to define estimates for the AR coefficients $a_1,\ldots, a_p$. Then, in the second step we make use of the CLS method to estimate quantities which are uniquely determined by the conditional second order moments of the AR residual process and which can be applied to define an estimator of the parameter $\underline{\theta}^0$ which we are really interested in. The main task in applying this method is to prove the identifiability of the parameter $\underline{\theta}^0$. We have to check whether $\underline{\theta}^0$ is uniquely determined by the set of quantities, each of them minimizing the expected value of a conditional sum of squares, or not. E.g. in the case of model (1.4) this corresponds to checking whether $\underline{\theta}^0 = (b, \sigma^2)^t$ is uniquely determined by $\mathrm{argmin}_{\underline{\beta}_1} E[\{X_t^2 - (\beta_{11} + \beta_{12} X_{t-2}^2)\}^2]$ and $\mathrm{argmin}_{\underline{\beta}_2} E[\{X_t X_{t-1} - \beta_2 X_{t-2}\}^2]$ or not.

The proof of the identifiablity of the parameter is divided up into two parts. The first part is given in the proof of the asymptotic properties of the CLS estimates where it is shown that the minimizer of the expected value of the conditional sum of squares is uniquely determined for any conditional sum of squares considered. The second step which consists in proving that these minimizer uniquely determine the parameter of interest $\underline{\theta}^0$ will be given here.

To this purpose we consider the quantities which are estimated by the CLS estimates in more detail. We will need the following theorem which is applicable not only to the special bilinear models (1.2) and (1.3) but also to general standardized superdiagonal bilinear models. A bilinear model is defined to be of a superdiagonal type iff $b_{ij} = 0$ for i < j in (1.1). Moreover the model is called standardized iff $E(X_t) = 0$. (In contrast to the linear ARMA case the mean of a bilinear process usually depends on the parameter

8

determining the process. For a further discussion of the standardization of bilinear processes we refer to the appendix.)

Since we restrict ourselves to the study of stationary bilinear time series only, here we always assume

(A0) $\{X_t\}_{t\in\mathbf{Z}}$ is a stationary, ergodic, causal and square integrable solution of the bilinear equations (1.1) where $\{\varepsilon_t\}_{t\in\mathbf{Z}}$ is such that $E(\varepsilon_t^4) < \infty$ and $E(\varepsilon_t^3) = 0$.

Later, in Theorem 3.1 we will give conditions for this assumption to be satisfied for the types of bilinear models we consider.

In the following we use $\mathrm{cov}(\cdot\,,\cdot\,|\,$t-m$)$ as an abbrevation for $\mathrm{cov}(\cdot\,,\cdot\,|\,\varepsilon_{t-m},\varepsilon_{t-m-1},\ldots)$.

Theorem 2.1

Let $\{X_t\}_{t\in\mathbf{Z}}$ be a standardized superdiagonal bilinear time series with $E|X_t|^4 < \infty$. Under (A0) the following equations hold for $m \geq 1$ and $s \geq 0$:

(i) $cov(u_t,\,u_{t\text{-}s}\,|\,t\text{-}s\text{-}m)$

$$= \sum_{j=s}^{r} b_{sj}\,\sigma^2\,E(X_{t\text{-}j}\,|\,t\text{-}s\text{-}m) + \delta_{s,0}\,\sigma^2 + \sum_{\substack{i=s+1\\ i\leq j}}^{k\wedge(s+m\text{-}1)} \sum_{j=s+1}^{r} b_{ij}\left\{\,\delta_{i,j}\,(b_{i\text{-}s,i\text{-}s}\,(c_4^\varepsilon + \sigma^4)) + \right.$$

$$\left. + \sum_{n=i\text{-}s}^{r} b_{i\text{-}s,n}\,\sigma^2\,E(X_{t\text{-}j}\,X_{t\text{-}s\text{-}n}\,|\,t\text{-}s\text{-}m)\,\right\},$$

   where $b_{ij} \equiv 0$ for $i > k$ and c_4^ε denotes the fourth cumulant of ε_1.

(ii) $cov(v_t,\,v_{t\text{-}s}\,|\,t\text{-}s\text{-}m)$

$$= \sum_{i=s+1}^{q\wedge(s+m\text{-}1)} c_i\,c_{i\text{-}s}\,\sigma^2 + c_s\,\sigma^2\,I\{1 \leq s \leq q\wedge(s+m\text{-}1)\} +$$

$$+ \sum_{i=s+1}^{q\wedge(s+m\text{-}1)} \sum_{j=i\text{-}s}^{r} c_i\,b_{i\text{-}s,j}\,\sigma^2\,E(X_{t\text{-}s\text{-}j}\,|\,t\text{-}s\text{-}m) +$$

$$+ \sum_{i=1}^{q\wedge(m\text{-}1)} \sum_{j=i+s}^{r} c_i\,b_{i+s,j}\,\sigma^2\,E(X_{t\text{-}j}\,|\,t\text{-}s\text{-}m) + cov(u_t,\,u_{t\text{-}s}\,|\,t\text{-}s\text{-}m),$$

   with $c_0 := 1$

(iii) $cov(X_t, X_{t-s} \mid t\text{-}s\text{-}m)$

$$= \sum_{i=1}^{p} a_i\, cov(X_{t-i}, X_{t-s} \mid t\text{-}s\text{-}m) \;+\; cov(v_t, v_{t-s} \mid t\text{-}s\text{-}m) \;+$$

$$+ \sum_{n=1}^{\tau-1}\; \sum_{i_1=1}^{p\wedge(\tau-1)} \cdots \sum_{i_n=1}^{p\wedge(\tau-1-i_1-..-i_{n-1})} a_{i_1}\cdot...\cdot a_{i_n}\, cov(v_t, v_{t-s-i_1-...-i_n} \mid t\text{-}s\text{-}m),$$

where $\tau := ((q\vee k)+1-s)\wedge m$.

For the proof of this theorem we refer to Grahn (1993, chapter 3).

Moreover the corresponding results for the (unconditional) covariances of the several processes can be deduced from studying the a.s. behaviour of the conditional covariances for $m \to \infty$. This is the content of the following corollary.

For brevity here we additionally apply the notation $w \equiv (q \vee k) + 1$ and $w' \equiv (q + 1) \vee k$.

Corollary 2.2

Under the assumptions stated in Theorem 2.1 the following equations hold for $s \geq 0$:

(i) $cov(u_t, u_{t-s}) = \delta_{s,0}\, \sigma^2 + \displaystyle\sum_{\substack{i=s+1 \\ i\leq j}}^{k} \sum_{j=s+1}^{r} b_{ij}\left\{ \delta_{i,j}\left(b_{i-s,i-s}\left(c_4^{\varepsilon} + \sigma^4\right)\right) \;+\right.$

$$\left. + \sum_{n=i-s}^{r} b_{i-s,n}\, \sigma^2\, E(X_{t-j}\, X_{t-s-n}) \right\}$$

(ii) $cov(v_t, v_{t-s}) = \displaystyle\sum_{i=s+1}^{q} c_i\, c_{i-s}\, \sigma^2 \;+\; c_s\, \sigma^2\, I\{1 \leq s \leq q\} \;+\; cov(u_t, u_{t-s})$

(iii) $cov(X_t, X_{t-s}) = \displaystyle\sum_{i=1}^{p} a_i\, cov(X_{t-i}, X_{t-s}) \;+\; cov(v_t, v_{t-s}) \;+$

$$+ \sum_{n=1}^{w'-s-1}\; \sum_{i_1=1}^{p\wedge(w'-s-1)} \cdots \sum_{i_n=1}^{p\wedge(w'-s-1-i_1-..-i_{n-1})} a_{i_1}\cdot...\cdot a_{i_n}\, cov(v_t, v_{t-s-i_1-...-i_n}).$$

Now we apply Theorem 2.1 to the AR residuals $v_t \equiv X_t - \sum_{i=1}^{p} a_i X_{t-i}$ of the bilinear models defined in (1.2) and (1.3) and show their relevance for the definition of estimators for the parameter determining the bilinear models (1.2) and (1.3). First we discuss the case of the bilinear model (1.2). In model (1.2) we get:

$$cov(v_t, v_{t-s} \mid t\text{-}w) \ (= E(v_t \cdot v_{t-s} \mid t\text{-}w))$$

$$= \sum_{i=s}^{q} c_i\, c_{i\text{-}s}\ \sigma^2\ +\ \sum_{j=w}^{r} \left\{ \sum_{i=1}^{q} c_i\, b_{i+s,j}\ \sigma^2 + b_{sj}\ \sigma^2 \right\} X_{t\text{-}j}\ +\ \sum_{j=w}^{r} \sum_{i=s+1}^{q} c_i\, b_{i\text{-}s,j}\ \sigma^2\ X_{t\text{-}s\text{-}j}$$

$$+\ \sum_{j=w}^{r} \sum_{n=w}^{r} \sum_{i=s+1}^{k} b_{ij}\, b_{i\text{-}s,n}\ \sigma^2\ X_{t\text{-}j}\, X_{t\text{-}s\text{-}n}$$

$$= \gamma(s)\ +\ \sum_{j=w}^{r+s} d_j(s)\ X_{t\text{-}j}\ +\ \sum_{j=w}^{r} \sum_{n=w}^{r} h_{j,n}(s)\ X_{t\text{-}j}\, X_{t\text{-}s\text{-}n}\ , \tag{2.3}$$

where $\gamma(s)$ denotes the covariancefunction of a MA(q) process with the parameter c_1, .., c_q and σ^2,

$$d_j(s) := b_{sj}\ \sigma^2 + \sum_{i=s+1}^{w\text{-}1+s} (c_i\, b_{i\text{-}s,j\text{-}s} + c_{i\text{-}s}\, b_{ij})\ \sigma^2 \quad \text{and} \quad h_{j,n}(s) := \sum_{i=s+1}^{k} b_{ij}\, b_{i\text{-}s,n}\ \sigma^2,$$

and $c_i \equiv 0$ for $i > q$ and $b_{ij} \equiv 0$ for i, j out of the summation domain.

The representation of $c(v_t, v_{t-s} \mid t\text{-}w)$ in (2.3) shows that the conditional covariances depend on the parameter and a finite set of X_t's only!

The estimation procedure in model I

Step (i). In model (1.2) the Yule – Walker equations with $a_1,\ldots, a_p$ hold for $s \geq w'$ (cf. Corollary 2.3 applied to model (1.2)), and thus we can define an estimator $\hat{\underline{a}}$ for $\underline{a}$ by solving the empirical Yule - Walker equations

$$\widehat{C}_{p,w'\text{-}1}\ \hat{\underline{a}} = \hat{\underline{c}}, \tag{2.4}$$

where $\widehat{C}_{p,w'\text{-}1} \equiv \{ \hat{c}(w'\text{-}1+ \ i\text{-}j) \}_{1 \leq i,j \leq p}$, $\hat{\underline{c}} = (\hat{c}(w'),..,\ \hat{c}(w'+p))^t$ and $\hat{c}(s)$ denotes the empirical estimate of the covariancefunction $c(s)$ ($\equiv cov(X_t, X_{t-s})$) of the process $\{X_t\}_{t \in \mathbf{Z}}$. If the matrix $\widehat{C}_{p,w'\text{-}1}$ is invertible, at least asymptotically, the solution $\hat{\underline{a}}$ of (2.4) is asymptotically uniquely determined, and we can define the estimated AR residuals by $\widehat{v}_t := X_t - \sum_{i=1}^{p} \hat{a}_i\, X_{t\text{-}i}$.

Step (ii). In view of equation (2.3) now we try to minimize the conditional sum of squares

$$\sum_{t=(r+s)\vee p+1}^{N} \{ \widehat{v}_t\, \widehat{v}_{t\text{-}s} - E_{\underline{\beta}s}(v_t \cdot v_{t\text{-}s} \mid t\text{-}w) \}^2 \tag{2.5}$$

11

with respect to $\underline{\beta}_s$ to obtain an estimate of

$$\underline{\beta}_s^0 \;=\; (\gamma(s);\; d_j(s),\; w \leq j \leq r+s;\; h_{j,n}(s),\; w \leq j,\; n \leq r)^t\,, \tag{2.6}$$

for $0 \leq s \leq w\text{-}1$. If a minimum of the penalty function exists we denote it with $\widehat{\underline{\beta}}_s$ and call it the CLS estimator of $\underline{\beta}_s^0$.

Step (iii). Now in the third step of the estimation procedure we give the algorithm for identifying the parameter b_{ij}, c_j and σ^2 from the the minimizer $\underline{\beta}_s^0$ of $E[\{v_t\, v_{t-s} -$ $E_{\underline{\beta}_s}(v_t{\cdot}v_{t-s} \mid t\text{-}w)\}]^2$, $0 \leq s \leq w\text{-}1$.

The estimation of $\underline{\beta}_s^0$ includes the estimation of $\gamma(s)$, $0 \leq s \leq w\text{-}1$. The function $\gamma(s)$ can be interpreted as the covariancefunction of the MA(q) process $\{Z_t\}_{t \in \mathbf{Z}}$ defined by $Z_t = \sum_{j=0}^q c_i'\, e_t$, with $c_0' = 1$, $c_i' = c_i$, $1 \leq i \leq q$, and $E(e_t^2) = \sigma^2$. If the process $\{Z_t\}_{t \in \mathbf{Z}}$ is invertible, its covariancefunction $\gamma(s)$ uniquely determines the underlying parameter. Wilson (1969) gave an explicit algorithm for evolving $c_1,\ldots,\, c_q$ and σ^2 from the covariances $\gamma(0),\ldots,\, \gamma(q)$. Since $w \text{ - } 1 \geq q$ we can apply this algorithm here also. The estimates $\hat{c}_1,\ldots,\, \hat{c}_q$ and $\widehat{\sigma^2}$ resulting from the application of this algorithm to the estimated covariances $\hat{\gamma}(0),\ldots,\, \hat{\gamma}(q)$ are called the _Wilson - estimates with q covariances_. If $w\text{-}1$ is strictly greater than q moreover we can make use of the additional information involved in the estimated covariances $\hat{\gamma}(q+1),\ldots,\, \hat{\gamma}(w\text{-}1)$ for improving the asymptotic efficiency of the estimates of the parameter $c_1,\ldots,\, c_q$ and σ^2.

What remains now is to define an algorithm for identifying the quantities b_{ij} from $c_1,\ldots,$ c_q, σ^2 and $\underline{\beta}_s^0$, $0 \leq s \leq w\text{-}1$. This algorithm will be applied later to define estimates of the b_{ij}'s from the estimates $\widehat{\underline{\beta}}_s$, $\hat{c}_j$ and $\widehat{\sigma^2}$. To this aim we consider $d_j(s)$ for $w \leq j \leq r+s$, which is part of $\underline{\beta}_s^0$, $0 \leq s \leq w\text{-}1$, in more detail. By definition (see (2.3)) we have

$$d_j(s) := b_{sj} + \sum_{i=s+1}^{w-1+s} (c_i\, b_{i-s,j-s} + c_{i-s}\, b_{ij})\,, \tag{2.7}$$

where for simplicity we suppose $\sigma^2 \equiv 1$ (if not, we consider $d_j(s)/\sigma^2$).

Thus the task to identify the bilinear parameter b_{ij} reduces to proving the invertibility of the transformation from b_{ij} to $d_j(s)$, given c_i. Since this transformation is a linear one, we have to check whether or not the matrix T defined by

$$T\underline{b} = \underline{d}, \tag{2.8}$$

is an invertible matrix, where $\underline{b} \equiv (b_{0,\bullet}, b_{1,\bullet}, \ldots, b_{w-1,\bullet})^t$ and $\underline{d} \equiv (d_\bullet(0), d_\bullet(1), \ldots,$
$d_\bullet(w-1))^t$ with $b_{i,\bullet}$ denoting the components b_{ij}, $w \leq j \leq r+i$, and $d_\bullet(i)$ is defined
analogously. According to (2.7) the matrix T can be computed as

$$
T = \begin{pmatrix}
D_0 & U_{0,1} & \cdots\cdots & U_{0,w-1} \\
L_{1,0} & \ddots & \ddots & \vdots \\
& \ddots & \ddots & \ddots \\
\vdots & & \ddots & \ddots & U_{w-2,w-1} \\
\vdots & & \ddots & \ddots \\
L_{w-1,0} & \cdots & L_{w-1,w-2} & D_{w-1}
\end{pmatrix},
$$

$$
\text{with } D_i = \begin{pmatrix}
1 & & & & \\
0 & \ddots & & & \\
\vdots & & & & \\
0 & & \ddots & & \\
c_{2i} & & & \ddots & \\
& \ddots & & & \ddots \\
& & c_{2i} & 0 & \cdots & 0 & 1
\end{pmatrix} \begin{matrix} \\ \\ \\ i+1 \to \end{matrix} \in M\{(h+i)\times(h+i)\}, \; 0 \leq i \leq w-1,
$$

$$
U_{0,j} = \begin{pmatrix}
2c_j & & \\
& \ddots & \\
& & 2c_j
\end{pmatrix}, \quad
U_{i,j} = \begin{pmatrix}
c_{j-i} & & & \\
0 & \ddots & & \\
\vdots & & & \\
0 & & \ddots & \\
c_{j+i} & & & \ddots \\
& \ddots & & & \\
& c_{j+i} & 0 \cdots 0 & c_{j-i}
\end{pmatrix} \begin{matrix} \\ \\ i+1 \to \end{matrix} \in M\{(h+i)\times(h+j)\},
$$
$$
1 \leq i < j \leq w-1
$$

$$
L_{i,j} = \begin{pmatrix}
0 & & & \\
\vdots & \ddots & & \\
0 & & & \\
c_{i+j} & & \ddots & \\
& \ddots & & \\
& c_{i+j} & 0 & \cdots & 0
\end{pmatrix} \begin{matrix} \\ \\ i+1 \to \end{matrix} \in M\{(h+j)\times(h+i)\}, \; 1 \leq j < i \leq w-1,
$$

and $L_{i,0} \equiv \mathbf{0}$, where $h := r - (w-1)$.

For proving the invertibility of this matrix, we consider it as a special case of the more
general matrix

$$\widetilde{T} = \begin{pmatrix} \widetilde{D}_0 & \widetilde{U}_{0,1} & \cdots\cdots & & \widetilde{U}_{0,w-1} \\ \widetilde{L}_{1,0} & \ddots & & \ddots & \vdots \\ & & \ddots & \ddots & \ddots & \\ \vdots & & & & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & \widetilde{U}_{w-2,w-1} \\ \widetilde{L}_{w-1,0} & \cdots & & \widetilde{L}_{w-1,w-2} & \widetilde{D}_{w-1} \end{pmatrix},$$

where $\widetilde{D}_i = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ * & & 1 \end{pmatrix}$, $\widetilde{U}_{i,j} = \begin{pmatrix} * & & 0 \\ & \ddots & \\ * & & * \end{pmatrix}$, $\widetilde{L}_{i,j} = \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ * & & 0 \end{pmatrix}$ and $\widetilde{L}_{i,0} \equiv 0$.

The matrix $\widetilde{T}$ can easily be seen to be equivalent to a lower triangle matrix. Beginning with the last but r row up to the first row we can transform $\widetilde{T}$ by elementary row operations to a lower triangle matrix. During these operations the main diagonal remains unchanged, which implies det $|\widetilde{T}| = 1$ and hence the invertibility of T too.

Thus the coefficients b_{ij} are uniquely determined by equation (2.8), and we can define asymptotically unique estimators for b_{ij} by solving the equation

$$\widehat{T}\,\widehat{\underline{b}} = \widehat{\underline{d}} \tag{2.9}$$

with respect to $\widehat{\underline{b}}$. Here $\widehat{\underline{d}}$ denotes the estimate for $\underline{d}$ which we implicitly obtain by estimating the whole vector $\underline{\beta}_s^0$, $0 \le s \le w-1$, and $\widehat{T}$ denotes the estimate of the matrix T defined in (2.8) which corresponds to estimating the parameter of the MA part by the Wilson algorithm.

In a similar way we can proceed in estimating model (1.3). First we consider the quantities which can be estimated by the CLS approach and second we give an algorithm for identifying the true underlying parameter from these quantities.
As it was the case in analyzing model (1.2) the CLS approach heavily relies on the structure of the conditional covariances of the AR residuals of the bilinear process. According to Corollary 2.2 applied to the bilinear process defined in (1.3) the covariances fulfill the equation

14

$$cov(v_t, v_{t-s} \mid t\text{-}k\text{-}1) \quad (= E(v_t \cdot v_{t-s} \mid t\text{-}k\text{-}1))$$

$$= \delta_{s,k} \cdot \{ b_{kk} \, \sigma^2 \, E(X_{t-k} \mid t\text{-}k\text{-}1) + \sum_{j=k+1}^{r} b_{kj} \, \sigma^2 \, X_{t-j} \} +$$

$$+ \; \delta_{s,0} \cdot \{ \sigma^2 + b_{kk}^2 \, \sigma^4 + \sum_{j=k}^{r} \sum_{n=k}^{r} b_{kj} \, b_{kn} \, \sigma^2 \, E(X_{t-j} \, X_{t-n} \mid t\text{-}k\text{-}1) \}. \tag{2.10}$$

In contrast to the representation in (2.3) the representation in (2.10) still includes quantities which are difficult to estimate, e.g. $E(X_{t-k} \mid t\text{-}k\text{-}1)$. This will cause a problem in the estimation of the parameter of this kind of bilinear models and we have to introduce an additional step in the estimation procedure to avoid this difficulty.

The estimation procedure in model II

Step (i). Analogously to step (i) in model I we use the second order moment structure of the process to define the estimator of the AR parametervector $\underline{a}$. Here the Yule - Walker equations with AR parameter $\underline{a}$ even hold for $s \geq 1$ such that we can define an estimator $\hat{\underline{a}}$ for $\underline{a}$ by

$$\hat{C}_{p,0} \, \hat{\underline{a}} = \hat{\underline{c}}, \tag{2.11}$$

with $\hat{C}_{p,0} := \{ \hat{c}(i - j) \}_{1 \leq i,j \leq p}$, $\hat{\underline{c}} = (\hat{c}(1),.., \hat{c}(p + 1))^t$. The asymptotic invertibility of the matrix $\hat{C}_{p,0}$ is already guaranteed by the condition $c(0) > 0$ (Brockwell and Davis (1991, Prop. 5.1.1)) which always holds in the cases we consider here. Thus as well the estimator $\hat{\underline{a}}$ as $\hat{v}_t$ also are at least asymptotically uniquely determined.

Step (ii). While in the first step the definition of the estimate is independent of the distribution of the innovations, in this step we will make use of the supposed Gaussian distribution explicitely in the sense that we make use of the knowledge about the first eight moments of the distribution of the innovations. Analogously to step (ii) in model I we consider the conditional covariance $E(v_t \, v_{t-s} \mid t\text{-}w)$ in (2.10) to deduce CLS estimates. If we consider equation (2.10) in more detail the special difficulty in analyzing model II becomes obvious. Unlike to model I, in the case $b_{kk} \neq 0$ we cannot parametrize the conditional covariance in σ^2, b_{ij}, and $b_{kj} \, b_{kn} \, \sigma^2$.

We can avoid this problem if first we estimate $b_{kk}\sigma^2$ by $\widehat{b_{kk}\sigma^2}$, second we minimize

$$\sum_{t=(r+1)\vee p+1}^{N} \{\widehat{u}_t\, \widehat{u}_{t-k} - \widehat{b_{kk}\sigma^2}\, X_{t-k} - E_\beta(u_t\, u_{t-k} - b_{kk}\sigma^2\, X_{t-k} \mid t-k-1)\}^2 \qquad (2.12)$$

with respect to β, where β denotes the vector $(b_{kj}\, \sigma^2, k+1 \leq j \leq r)^t$, and third we deduce an estimator for σ^2 by using some higher order moment properties of the bilinear process.

To estimate $b_{kk}\sigma^2$ we make use of the relation

$$E(u_t \cdot u_{t-k}{}^2) = 2b_{kk}\sigma^2\, E(u_t{}^2) \qquad (2.13)$$

which can be proved easily.

Given a sample $X_1, \ldots, X_N$ according to this equation now we can estimate $b_{kk}\sigma^2$ by

$$\widehat{b_{kk}\sigma^2} \equiv \{\sum_{t=k+p+1}^{N} \widehat{u}_t\, \widehat{u}_{t-k}^2\}/2\{\sum_{t=k+p+1}^{N} \widehat{u}_t^2\}. \qquad (2.14)$$

Having estimated $b_{kk}\sigma^2$ we define an estimator $\widehat{\beta}$ for $\underline{\beta}^0 \equiv (b_{k,k+1}\sigma^2, \ldots, b_{k,r}\sigma^2)^t$ by trying to minimize the penalty function defined in (2.12). Here, in contrast to the procedure applied to model (1.2) we cannot make use of the equation (2.10) to deduce an estimator for σ^2 since in the above model for $b_{kk} \neq 0$ the troublesome product $b_{kk}\sigma^2\, E(X_{t-k} \mid t-k-1)$ appears which cannot be represented as a deterministic function of the sample $X_1, \ldots, X_N$ only. Hence we will suggest another procedure for estimating σ^2 in this model.

Step (iii). The crucial relation we need is the following one

$$E(v_t{}^4) = 3\mathrm{Var}\,\{(\sum_{j=k}^{r} b_{kj}\sigma\, X_{t-j})^2\} + 6b_{kk}\sigma^2\{E(v_t{}^3) - 2\,(b_{kk}\sigma^2)^3\} + 6\,(b_{kk}\sigma^2)^4. \qquad (2.15)$$

The proof of (2.15) is given at the end of this chapter.

The main utility of the equation (2.15) is that we can estimate $C := \mathrm{Var}\,\{(\sum_{j=k}^{r} b_{kj}\,\sigma\, X_{t-j})^2\}$ only by replacing $b_{kk}\sigma^2$ by its estimate defined in (2.14) and by replacing the theoretical moments of v_t with their empirical counterparts.

On the other hand also $\sigma^4 \cdot C = \mathrm{Var}\,\{(\sum_{j=k}^{r} b_{kj}\sigma^2\, X_{t-j})^2\}$ can be estimated by replacing the theoretical moments with their empirical counterparts and $b_{kj}\sigma^2$ by $\widehat{b_{kj}\sigma^2}$ only.

Provided that $C \neq 0$ thus an estimator for σ^2 is defined by

$$\widehat{\sigma^2} \equiv \mid \widehat{\sigma^4 C}/\, \widehat{C} \mid^{1/2} \qquad (2.16)$$

16

where $\hat{C}$ and $\widehat{\sigma^4 C}$ denotes the estimator of C and $\sigma^4 \cdot C$ respectively.

Studying the case $C = 0$ in detail we get

$$C = 0 \iff \sum_{j=k}^{r} b_{kj}\, \sigma\, X_{t-j} = 0 \ \text{ a.s. } \iff \ \forall\, j\colon b_{kj} = 0 \ \text{ by Lemma 4.6}$$

According to this equivalence now the additional assumption included in the bilinear model (1.3), namely that there exists at least one bilinear coefficient b_{kj}, $k \le j \le r$, with $b_{kj} \ne 0$, ensures $C \ne 0$. If we prove that $\hat{C}$ converge to C we may use $\widehat{\sigma^2}$ defined in (2.16) as an estimator for σ^2.

Then an estimator $\widehat{b_{kj}}$ for b_{kj}, $k \le j \le r$, can be defined by

$$\widehat{b_{kj}} \equiv \widehat{b_{kj}\sigma^2} \,/\, \widehat{\sigma^2} \tag{2.17}$$

with $\widehat{\sigma^2}$ defined in (2.16) and $\widehat{b_{kj}\sigma^2}$ being the component of $\hat{\beta}$.

PROOF OF (2.15).

By definition of v_t and taking the Gaussian assumption into account we obtain (we use 'c' as an abbrevation for 'cumulant' here)

$$E(v_t^4) = c(v_t,\, v_t,\, v_t,\, v_t)$$

$$= \sum_{i,j,l,m=k}^{r} b_{ki}\, b_{kj}\, b_{kl}\, b_{km}\ c(\varepsilon_{t-k} X_{t-i},\ \varepsilon_{t-k} X_{t-j},\ \varepsilon_{t-k} X_{t-l},\ \varepsilon_{t-k} X_{t-m})$$

and by applying the product theorem for cumulants (see e.g. Brillinger (1981)) further we obtain

$$= \sum_{i,j,l,m=k}^{r} b_{ki}\, b_{kj}\, b_{kl}\, b_{km}\ \{\ 3\sigma^4\ c(X_{t-i},\, X_{t-j},\, X_{t-l},\, X_{t-m}) + 6\sigma^4\ c(X_{t-i},\, X_{t-j}) \cdot$$

$$\cdot c(X_{t-l},\, X_{t-m}) + 36\sigma^2\ c(\varepsilon_{t-k},\, X_{t-i}) \cdot c(\varepsilon_{t-k},\, X_{t-j}) \cdot c(X_{t-l},\, X_{t-m}) +$$

$$+ 6\ c(\varepsilon_{t-k},\, X_{t-i}) \cdot c(\varepsilon_{t-k},\, X_{t-j}) \cdot c(\varepsilon_{t-k},\, X_{t-l}) \cdot c(\varepsilon_{t-k},\, X_{t-m})\ \}$$

$$= 3\sigma^4 \sum_{i,j,l,m=k}^{r} b_{ki}\, b_{kj}\, b_{kl}\, b_{km}\ c(X_{t-i},\, X_{t-j},\, X_{t-l},\, X_{t-m}) + 6\,\{ \sum_{i,j=k}^{r} b_{ki}\, b_{kj}\, \sigma^2 \cdot$$

$$\cdot c(X_{t-i},\, X_{t-j})\}^2 \ + \ 36\ (b_{kk}\,\sigma^2)^2 \sum_{i,j=k}^{r} b_{ki}\, b_{kj}\, \sigma^2\ c(X_{t-i},\, X_{t-j}) \ + \ 6\ (b_{kk}\,\sigma^2)^4$$

$$= 3Var\{(\sum_{j=k}^{r} b_{kj}\, \sigma\, X_{t\text{-}j})^2\} + 36\,(b_{kk}\,\sigma^2)^2 \sum_{i,j=k}^{r} b_{ki}\, b_{kj}\,\sigma^2\, c(X_{t\text{-}i}, X_{t\text{-}j}) + 6\,(b_{kk}\,\sigma^2)^4.$$

Thus it remains to prove, that

$$6\,(b_{kk}\,\sigma^2) \sum_{i,j=k}^{r} b_{ki}\, b_{kj}\,\sigma^2\, c(X_{t\text{-}i}, X_{t\text{-}j}) = E(v_t^3) - 2\,(b_{kk}\,\sigma^2)^3.$$

But this equation can immediately deduced from the product theorem for cumulants as follows:

$$E(v_t^3) = c(v_t, v_t, v_t)$$

$$= \sum_{i,j,l=k}^{r} b_{ki}\, b_{kj}\, b_{kl}\, c(\varepsilon_{t\text{-}k}\, X_{t\text{-}i}, \varepsilon_{t\text{-}k}\, X_{t\text{-}j}, \varepsilon_{t\text{-}k}\, X_{t\text{-}l})$$

$$= 2 \sum_{i,j,l=k}^{r} b_{ki}\, b_{kj}\, b_{kl}\, \{3\sigma^2\, c(X_{t\text{-}i}, X_{t\text{-}j})\cdot c(\varepsilon_{t\text{-}k}, X_{t\text{-}l}) + c(\varepsilon_{t\text{-}k}, X_{t\text{-}i})\cdot c(\varepsilon_{t\text{-}k}, X_{t\text{-}j})\cdot$$

$$\cdot c(\varepsilon_{t\text{-}k}, X_{t\text{-}l})\}$$

$$= 6\,(b_{kk}\,\sigma^2) \sum_{i,j=k}^{r} b_{ki}\, b_{kj}\,\sigma^2\, c(X_{t\text{-}i}, X_{t\text{-}j}) + 2\,(b_{kk}\,\sigma^2)^3. \blacktriangle$$

3   Results

Theorem 3.1 is the basic result for proving the asymptotic properties of the estimators deduced in chapter 2. We prove the CLT and the LIL for the empirical moments of the process $\{X_t\}_{t\in \mathbf{Z}}$. As a first application we obtain a result concerning the asymptotic behaviour of the Yule Walker estimator of the AR parameter in the general standardized superdiagonal bilinear model. This is the result of Theorem 3.2. Since both bilinear models under our consideration belong to this special class of bilinear models this result will be applied to the estimation as well in model (1.2) as in model (1.3) too.

Before we state Theorem 3.1 we give some preliminary computations which are useful not only for formulating the assertions of Theorem 3.1 but also for a better understanding of the proof of the theorem.

In Theorem 3.1 we establish conditions for both guaranteeing the existence of a stationary, ergodic and m - times integrable solution of the bilinear equations (1.1) and implying the CLT and the LIL for the empirical moments of the bilinear process to hold. These conditions shall be motivated by the following considerations.

First we rewrite the bilinear process in (1.1) as a vectorprocess by introducing the vectors $\underline{X}_t := (X_t, X_{t-1}, ..., X_{t-h+1})^t$, $\underline{\alpha} := (\alpha, 0, .., 0)^t$ and $D := (1, 0, ..., 0)^t \in \mathbf{R}^h$ with $h := \max\{p, r\}$, $\underline{\varepsilon}_t := (\varepsilon_t, \varepsilon_{t-1}, ..., \varepsilon_{t-q})^t \in \mathbf{R}^{q+1}$, and the matrices

$$A = \begin{pmatrix} a_1 & \cdots & a_p & 0 & \cdots & 0 \\ 1 & 0 & \cdots\cdots\cdots & 0 \\ & \ddots & & & & \vdots \\ & & \ddots\ddots & & & \vdots \\ 0 & & & \ddots & & \vdots \\ & & & & 1 & 0 \end{pmatrix}, \quad B_i = \begin{pmatrix} b_{i1} & \cdots & b_{ir} & 0 & \cdots & 0 \\ 0 & & \cdots\cdots\cdots & 0 \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & & \cdots\cdots\cdots & 0 \end{pmatrix} \in M(h, h) \text{ and}$$

$$C = \begin{pmatrix} 1 & c_1 & \cdots & c_q \\ 0 & \cdots\cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots\cdots & 0 \end{pmatrix} \in M(h, q+1).$$

With this notation the bilinear equations (1.1) can be represented in state space form as

$$\underline{X}_t = \underline{\alpha} + A \underline{X}_{t-1} + \sum_{i=1}^{k} \varepsilon_{t-i} B_i \underline{X}_{t-1} + C \underline{\varepsilon}_t$$

$$X_t = D \underline{X}_t .$$

(3.1)

Applying the notation used in (3.1) the idea underlying the proof of the first part of the Theorem 3.1 can be motivated easily. The proof is based on the study of the sequence of the vectorvalued stochastic processes defined by

$$\underline{S}_N(t) \equiv \begin{cases} 0 & N < 0 \\ C \underline{\varepsilon}_t & N = 0 \\ C \underline{\varepsilon}_t + \psi(t) \cdot \underline{S}_{N-1}(t-1) , & N > 0 \end{cases} \quad , t \in \mathbf{Z},$$

(3.2)

where $\Psi(t) = A + \sum_{i=1}^{k} \varepsilon_{t-i} B_i$.

For the results of Theorem 3.1 to hold we require some conditions which are imposed mainly for proving the following statements concerning the behaviour of $\underline{S}_N(t)$:

(i)　For each $t \in \mathbf{Z} : \underline{S}_N(t)$ converges in L^m for $N \to \infty$

19

(ii)   The process $\{Y_t\}_{t\in\mathbf{Z}}$ with $Y_t :=$ a.s. $- \lim\limits_{N\to\infty} \underline{S}_N(t)$ fulfills
the bilinear equations (1.1) with $\underline{\alpha} = 0$.

Here L^m denotes the space of all functions f such that $\int |f|^m \, dP < \infty$.

Now we show how the conditions required in Theorem 3.1 arise in the proof of the assertions (i) and (ii) above. As the conditions required for verifying (i) guarantee the validity of (ii) also, we restrict our attention on proving (i) here.

For $m = 2$ the proof of (i) has been worked out explicitly by Liu and Brockwell (1988, proof of Theorem 3.1). Analogously to the method they suggest, Liu and Liu (1990) extended the idea to the case $m > 2$. Thus we only give a sketch of the proof here.

Since for any $m \geq 2$ the space L^m is a complete space, (i) can be verified by proving that for each $t \in \mathbf{Z} : \{\underline{S}_N(t)\}_{N\in\mathbf{N}}$ is a Cauchy - sequence in L^m. To prove this property we consider the sequence of the first differences of $\underline{S}_N(t)$

$$\underline{\Delta}_N(t) \equiv \underline{S}_N(t) - \underline{S}_{N-1}(t) = \begin{cases} 0 & N < 0 \\ C\,\underline{\varepsilon}_t & N = 0 \\ \psi(t)\cdot\underline{\Delta}_{N-1}(t\text{-}1), & N > 0 \end{cases} \tag{3.3}$$

According to its definition $\{\underline{\Delta}_N(t)\}_{t\in\mathbf{Z}}$ is a strictly stationary process for each $N \geq 1$ with $\underline{\Delta}_N(t)$ being measurable with respect to the σ-field F_{t-1}, generated by ε_{t-i}, $i \geq 1$, for each $t \in \mathbf{Z}$.

Now the main effort is to prove that there exists an $\delta \in (0,1)$ such that

$$E\{\|\underline{\Delta}_N(t)\|_m^m\} \leq \text{const}\cdot\delta^{N/2} \quad \text{for } N \geq N_0 \tag{3.4}$$

holds, where $\| \underline{x} \|_m^m \equiv \sum_i |x_i|^m$ for $x = (x_1,\ldots, x_r)^t$, such that we immediately can deduce the Cauchy - convergence of $\{\underline{S}_N(t)\}_{N\in\mathbf{N}}$ in L^m then.

In view of the inequality

$$E\{\|\underline{\Delta}_N(t)\|_m^m\} \leq |E\{\underline{\Delta}_N(t)^{\otimes m}\}| \,,$$

where $| \underline{x} | \equiv \sum_i |x_i|$, for proving (3.4) it suffices to study the behaviour of $V_N := E\{\underline{\Delta}_N(t)^{\otimes m}\}$ only (here $\otimes$ denotes the tensor product and $\otimes^m$ its m - fold application).

Our aim now is to deduce a recursion formula for V_N such that by controlling the behaviour of the transition matrix appearing in the recursion formula we can gain the

geometrical decrease which we require in (3.4). By some tedious computations and under some additional assumptions on the behaviour of the higher order moments of the innovation process, we can deduce both a vector $\underline{Y}_N$ which contains V_N as a component and a matrix Γ_m such that for $N \geq N_0$ the following recursion is valid:

$$\underline{Y}_N = \Gamma_m \, \underline{Y}_{N-1} = \Gamma_m^2 \, \underline{Y}_{N-2} = \ldots = \Gamma_m^{N-N_0} \, \underline{Y}_{N_0}. \qquad (3.5)$$

(For $m = 2$ the vector $\underline{Y}_N$ and the matrix Γ_m has been computed explicitly in the bilinear model with $k = 2$ by Liu and Brockwell (1988, $\underline{Y}_N \equiv \underline{u}_N$ therein) and for $m = 4$ by Liu and Liu (1990).)

Let us now assume $\rho(\Gamma_m) < 1$. Then we can further conclude (cf. Proposition 2.1 in Liu and Brockwell (1988)) that

$$| \underline{V}_N | \leq | \underline{Y}_N | \leq \text{const}\cdot\delta^{N/2} \quad \text{for } N \geq N_0.$$

Since $E\{\|\underline{\Delta}_N(t)\|_m^m\} \leq | \underline{V}_N |$, by means of the inequality above and applying the recursion in (3.5) we obtain (3.4).

Obviously the condition $\rho(\Gamma_m) < 1$ is the crucial condition in this argumentation and thus we impose this condition in Theorem 3.1.

If we replace the matrix C in the definition of $\underline{S}_N(t)$ by another matrix $\tilde{C}$ we recognize that this causes no difference in the evaluation of the vector $\underline{Y}_N$ and the matrix Γ_m, since we only make use of the recursive definition of $\underline{\Delta}_N(t)$ and the measurability and stationarity properties of $\underline{\Delta}_0(t)$. Thus the matrix Γ_m is independent of the moving average part of the bilinear process which is represented by the matrix C!

After these preparations now we state the announced theorem.

(Notice that in contrast to the definition in (1.1) where we defined the bilinear BL(p, q, *k*, r) model, in formulating the subsequent result the quantity '*k*' is used with another meaning such that here we consider a bilinear BL(p, q, *s*, r) model!)

Theorem 3.1

Let $\{X_t\}_{t\in\mathbf{Z}}$ be the bilinear BL(p, q, s, r) process defined in (3.1) with $\underline{\alpha} = 0$. If for some $v \in \mathbf{N}$ the corresponding innovation sequence $\{\varepsilon_t\}_{t\in\mathbf{Z}}$ is a 2vs - th order symmetric innovation sequence and if moreover $\rho(\Gamma_{2v}) < 1$, where Γ_{2v} is implicitely defined in (3.5), then the following assertions hold:

(i) **There exists a strictly stationary, causal and ergodic solution** $\{X_t\}_{t\in Z}$ of (3.1) with $E[X_t^{2v}] < \infty$. The unique strictly stationary solution X_t is the first component of the vector

$$\underline{X}_t = \sum_{n=1}^{\infty} [\prod_{i=1}^{n} \Psi(t+1-i)] \, C \, \underline{\varepsilon}_{t-n} + C \, \underline{\varepsilon}_t \, ,$$

with $\Psi(t)$ defined in (3.2) and the infinite sum can be interpreted either as an almost sure or L^{2v} limit.

(ii) To give the result concerning the asymptotics of the empirical mixed moments of he process $\{X_t\}_{t\in Z}$ moreover we need the following definitions:

Let $K := \{(0, k_1,\ldots, k_i) : k_1 \leq\ldots\leq k_i, \, 0 \leq i \leq v - 1, \, k_j \in N_0 \text{ for } 0 \leq j \leq i\}$ denote the set of indices for identifying the several mixed moments of the process up to the order v. For $\underline{k} = (0, k_1,\ldots, k_i) \in K$ we further define $\mu_{\underline{k}} := E[X_0 \cdot X_{k_1} \cdot\ldots\cdot X_{k_i}]$ and $\eta_{\underline{k},t} := X_t X_{t+k_1} \cdot\ldots\cdot X_{t+k_i} - \mu_{\underline{k}}$.

Moreover for $u \in [0,1]$ we define the process of the partial sums $S_{N,\underline{k}}(u) := \sum_{t=1}^{[Nu]} \eta_{\underline{k},t}$ and a continuous version $\xi_{N,\underline{k}}(u) := S_{N,\underline{k}}(u) + (Nu - [Nu])\cdot\eta_{\underline{k},[Nu]+1}$.

Then for all $k \in N$ and for all $\gamma = (\gamma_1 ,.., \gamma_k)^t \in R^k$ and $\underline{k}_i \in K$, $1 \leq i \leq k$, we have

$$\left\{N^{-1/2} \sum_{i=1}^{k} \gamma_i \, \xi_{N,\underline{k}_i}(u) \right\}_{s\in[0,1]} \xrightarrow[C[0,1]]{D} \left\{c_\gamma B(u)\right\}_{u\in[0,1]}$$

where $\{B(u)\}_{u\in[0,1]}$ denotes the standard Brownian motion and

$$c_\gamma := \lim_{N\to\infty} E[(\sum_{i=1}^{k} \gamma_i \, \xi_{N,\underline{k}_i}(1))^2]/N \text{ exists with } 0 \leq c < \infty,$$

and the law of the iterated logarithm

$$\{N\cdot loglog(N\vee e)\}^{-1/2} \sum_{i=1}^{k} \gamma_i \, \xi_{N,\underline{k}_i} \text{ is relative compact in } (C[0,1], ||\cdot||_\infty)$$

holds.

REMARK.

Since the assumptions of Theorem 3.1 do not depend on the moving average part of the bilinear process (cf. the discussion before Theorem 3.1), the theorem remains valid even if we admit processes with $\underline{\alpha} \neq 0$, provided that the matrix $I - A$ is invertible. This is due to the fact that for $\underline{\alpha} \neq 0$ the process in (3.1) can be transformed equivalently to a process with $\underline{\alpha} = 0$ by changing the MA part of the model only if we suppose the matrix $I - A$ to be invertible (cf. the appendix). In the case of a general superdiagonal bilinear processes ($b_{ij} = 0$ for $i > j$) Liu (1992) proved that the condition $\rho(\Gamma_{2v}) < 1$ always implies $\rho(A) < 1$ and hence the invertibility of $I - A$. Thus in this case we do not need to impose any additional assumption even in the case $\underline{\alpha} \neq 0$.

In view of this remark, for applying Theorem 3.1 to the bilinear processes defined in (1.2) and (1.3), we impose the following assumption on these bilinear BL(p, q, k, r) processes:

(A1) $\{X_t\}_{t \in \mathbf{Z}}$ is a general superdiagonal bilinear process driven by a 4k-th order symmetric innovation sequence $\{\varepsilon_t\}_{t \in \mathbf{Z}}$ such that $\rho(\Gamma_4) < 1$.

According to assertion (i) of Theorem 3.1 this assumption implies our basic assumption (A0).

Furthermore we will make use of the following assumptions for proving the CLT and the LIL for the various estimates under consideration:

(A2) $C_{p,w'-1} := \{c(w'-1+i-j)\}_{1 \leq i,j \leq p}$ is invertible, where $c(i) := \operatorname{cov}(X_t, X_{t-i})$ and p corresponds to the order of the AR part of the bilinear process

(A3)   The MA - part of the bilinear model, defined by the MA coefficients $c_1,..., c_q$ and σ^2, is invertible

(A4) ε_t cannot take on only two values a. s.

Now we come to the study of the asymptotic properties of the estimates of the parameter determining the bilinear process. In chapter 2 we suggested first to fit an AR(p) process to the bilinear process. To this purpose we applied the Yule - Walker equations. Now we give the theoretical justification for proceeding in such a way.

Theorem 3.2 *(Estimation of the AR parameter)*

In the general standardized superdiagonal bilinear model which fulfills (A1) and (A2),

we obtain the following result

$N^{1/2}$ $(\hat{\underline{a}} - \underline{a})$ *is asymptotically normal distributed with expectation zero*

and

the LIL, i.e. $\hat{\underline{a}} - \underline{a} = O(L_N)$ *a.s. with* $L_N \equiv \{N/loglog(N \vee e)\}^{-1/2}$, *holds*

where $\hat{\underline{a}}$ *denotes the w'-th order Yule-Walker-estimator of* $\underline{a}$.

(Remind that $w' \equiv max\{q + 1, k\}$.*)*

Now we look for theoretical results attainable in the estimation of the remaining parameter of the bilinear model. To this purpose we consider each bilinear model ((1.2) and (1.3)) separately.

3.1   Asymptotics in model I

In this part we will prove that the estimates of the parameter determining the bilinear model I satisfy both the CLT and the LIL. In model I defined in (1.2) the parameter of our interest is $\underline{\theta}^0 = (a_1^0,..., \ a_p^0; b_{ij}^0, \ 1 \le i \le k, \ w+1 \le j \le r; \ c_1^0,..., \ c_q^0, \sigma_0^2)^t$, where we use the index '0' to denote the true parametervalue. The estimator of $\underline{\theta}^0$ is defined in chapter 2. Here we consider $\underline{\theta}^0 \in \Theta$ with

$$\Theta \equiv \big\{ (a_1,..., a_p; b_{ij}, \ 1 \le i \le k, \ w+1 \le j \le r; \ c_1,..., c_q, \sigma^2)^t \in \ \mathbf{R}^{p+k(r-w)+q} \times \mathbf{R}_+^0 :$$

$$\text{(A1), (A2) and (A3) are fulfilled}\big\}$$

For $\underline{\theta} \in \Theta$ as well the maximum eigenvalue of Γ_4 (see assumption (A1)) as also the zeroes of the polynomials occuring in (A2) and (A3) are continuous functions in the parameter $\underline{\theta}$. Thus for any $\underline{\theta} \in \Theta$ we always find an open neighbourhood of $\underline{\theta}$ such that each parameter in this neighbourhood belongs to Θ too. Hence Θ is an open subset of $\mathbf{R}^{p+k(r-w)+q} \times \mathbf{R}_+^0$.

Theorem 3.3

In a bilinear model of type I defined in (1.2) fulfilling the assumptions (A1) - (A4), the estimator $\hat{\underline{\theta}}$ of $\underline{\theta}^0$, where $\hat{\underline{\theta}}$ is defined by means of the estimation procedure worked out in the equations (2.4) - (2.9), has the following asymptotic properties:

(i) $\hat{\underline{\theta}} \to \underline{\theta}^0$ a.s.

(ii) if moreover $\rho(\Gamma_8) < 1$ holds and $\{\varepsilon_t\}_{t \in \mathbf{Z}}$ is an 8k-th order symmetric innovation sequence , then

$$N^{1/2} \, (\hat{\underline{\theta}} - \underline{\theta}^0) \text{ is asymptotically normal distributed with expectation zero}$$

and

$$\text{the LIL, i.e. } \hat{\underline{\theta}} - \underline{\theta}^0 = O(L_N) \text{ a.s., holds.}$$

REMARK.

From the proof of Theorem 3.3 it becomes obvious that we can define a moment estimator of $\underline{\theta}^0$ which has the same asymptotic distribution as the estimator $\hat{\underline{\theta}}$ since as well $\hat{\underline{a}}$ defined in (2.4) as $\hat{\underline{\beta}}_s$ from (2.5) also can be represented as differentiable functions of the empirical mixed moments of the process $\{X_t\}_{t \in \mathbf{Z}}$ only. The moment estimator which corresponds to the estimator $\hat{\underline{\theta}}$ is a minimal one in the sense that it depends on the first four moments of the bilinear process only.

The great advantage of the CLS procedure in comparison to the usual MM procedure consists in the ability to solve the identification problem by means of the uniqueness of the minimizer of the conditional variance. In the MM case the identification of the parameter determining the process from the moments is the main problem in applying MM to bilinear time series. (cf. Introduction)

According to this theorem the CLS procedure works well in the sense that it produces asymptotic normal estimates which moreover fulfill the LIL. In principle it is possible to compute the corresponding asymptotic covariance matrix also. But since even in the very simple bilinear model (1.2), the evaluation of this matrix needs extensive computations, we restrict ourselves on proving the asymptotic normal behaviour here.

3.2 Asymptotics in model II

Here we will prove as well the CLT as the LIL for the parameter estimates defined in chapter 2 part 2. In model II the parameter of interest is $\underline{\theta} = (a_1,..., a_p; b_{kj}, k \leq j \leq r; \sigma^2)^t$ $\in \mathbf{R}^{p+(r+1-k)} \times \mathbf{R}_+^0$.

The main results are summarized in the following theorem.

Theorem 3.4

In a bilinear model of type II defined in (1.3) fulfilling the assumptions (A1) and (A4) the estimator $\widehat{\underline{\theta}} = ((\widehat{\underline{a}}))^t, \widehat{b_{ij}}, \widehat{\sigma^2})^t$ of $\underline{\theta}^0$ with $\widehat{\underline{a}}$ denoting the first order Yule-Walker-estimator of $\underline{a}$ defined in (2.11), and $\widehat{b_{ij}}$, and $\widehat{\sigma^2}$ are defined in (2.14) - (2.16), has the following asymptotic properties:

(i) $\widehat{\underline{\theta}} \to \underline{\theta}^0$ a.s.

(ii) if moreover $\rho(\Gamma_8) < 1$ holds then

$N^{1/2} (\widehat{\underline{\theta}} - \underline{\theta}^0)$ *is asymptotically normal distributed with expectation zero*

and

the LIL, i.e. $\widehat{\underline{\theta}} - \underline{\theta}^0 = O(L_N)$ a.s., holds

REMARK.

As it was the case in the bilinear model of type I here also the deduced estimator $\widehat{\underline{\theta}}$ is asymptotically equivalent to a moment estimate. This becomes obvious from the proof of the Theorem which shows that approximately $\widehat{\underline{\theta}}$ can be represented as a differentiable function of the empirical mixed moments of the process only.

For a further discussion of this property we refer to the remark after Theorem 3.3.

3.3 Asymptotics in the unknown mean case

The results obtained in the sections 3.1 and 3.2 are stated for the case that we know the mean of the process $\{X_t\}_{t \in \mathbf{Z}}$ to be zero. Here we consider the case of an arbitrary unknown mean μ. Thus μ is an additional parameter which have to be estimated also.

Given the sample $X_1, \ldots, X_N$ an obvious estimate for μ is the empirical mean $\widehat{\mu} := N^{-1}$ $\sum_{t=1}^{N} X_t$. We will show that the results of the previous chapters remain valid even if we include the estimation of the mean.

If the mean of the process have to be estimated the Yule - Walker estimates defined in (2.4) and (2.11) respectively will be modified in the way that $\widehat{c}(i)$ is replaced with $\widehat{\widehat{c}}(ii) \equiv$ $N^{-1} \sum_{t=i+1}^{N} (X_t - \widehat{\mu})(X_{t-i} - \widehat{\mu})$. The modified Yule - Walker estimator shall be denoted with $\widehat{\widehat{a}}$. With these notations we get the following result, which is an extension of Theorem 3.2.

Theorem 3.5 *(Estimation of the AR coefficients in the unknown mean case)*
In a standardized superdiagonal bilinear model with expectation μ which fulfills (A1) and (A2), we obtain the following result

$N^{1/2} (((\widehat{\widehat{a}})^t, \widehat{\mu}) - ((\underline{a})^t, \mu))^t$ *is asymptotically normal distributed with expectation zero*

and
the LIL, i.e. $((\widehat{\widehat{a}})^t, \widehat{\mu}) - ((\underline{a})^t, \mu) = O(L_N)$ a.s., holds

In view of this generalization of Theorem 3.2 subsequently we will generalize the results of the parts 3.1 and 3.2 to the case of a standardized superdiagonal bilinear model with expectation μ, where μ is estimated by $\widehat{\mu}$.

Theorem 3.6
If instead of the zero - mean processes considered in the Theorem 3.3 and 3.4 now we consider the corresponding μ - mean processes with unknown μ and if we estimate μ by $\widehat{\mu}$, the deduced results concerning the estimation of $\underline{\theta}^0$ can be carried over to the estimation of the extended vector $\underline{\theta}_\mu^0 := (\mu, (\underline{\theta}^0)^t)^t$ too.

REMARK.
If we study the proof of Theorem 3.6 carefully we recognize that estimating μ has an influence on the asymptotic distribution of the parameterestimates which is not neglectable. The estimates are still asymptotic normal distributed but their asymptotic variance does increase if μ have to be estimated also.

4 Proofs

Proof of Lemma 1.1

Proof of (i). See e.g. Guegan (1981).

Proof of (ii).

From the causality property of $\{X_t\}_{t\in\mathbf{Z}}$ which is due to (i), we can conclude that not only the quantities $\varepsilon_{t-1}\cdot X_{t-2}$ and ε_t but also the variables ε_{t-1} and X_{t-2} are independent for each $t \in \mathbf{Z}$. Thus it is a straightforward matter to compute the *2n*-th moment of X_t as

$$E(X_t^{2n}) = \sum_{i=0}^{2n} \binom{2n}{i} b^i\, E(\varepsilon_t^i)\, E(\varepsilon_t^{2n-i})\, E(X_t^i)$$

$$= \sum_{\substack{i=0 \\ i\ \text{even}}}^{2(n-1)} \binom{2n}{i} b^i\, E(\varepsilon_t^i)\, E(\varepsilon_t^{2n-i})\, E(X_t^i) \;+\; b^{2n}\, E(\varepsilon_t^{2n})\, E(X_t^{2n})$$

where in the last equation we applied the assumptions on the distribution of the innovations. As a first consequence we get the equivalence

$$E(X_t^{2n}) < \infty \quad \Leftrightarrow \quad \{\, E(X_t^{2(n-1)}) < \infty \,\} \wedge \{\, b^{2n}\, E(\varepsilon_t^{2n}) < 1 \,\}$$

$$\Leftrightarrow \quad b^{2m}\, E(\varepsilon_t^{2m}) < 1 \;\; \text{for all} \;\; 1 \le m \le n$$

Taking Jensen's inequality into account moreover we remark that

$$b^{2m}\, E(\varepsilon_t^{2m}) \ge b^{2(m-1)}\, E(\varepsilon_t^{2(m-1)}) \;\; \text{for } m \in \mathbf{N}$$

such that the assertion is obvious now.

Proof of (iii).

From the equation

$$\widetilde{b}^2\widetilde{\sigma}^2 = 1 - b^2\sigma^2 \tag{1}$$

we conclude that $\widetilde{\sigma}^2 > 0$ and $\widetilde{b}^2\widetilde{\sigma}^2 < 1$, such that according to assertion (i) of the lemma the process $\{\widetilde{X}_t\}_{t\in\mathbf{Z}}$ is a stationary process. Moreover due to (1) we have

$$\widetilde{b}^4\widetilde{\sigma}^4 = (1 - b^2\sigma^2)^2 = b^4\sigma^4 - 2b^2\sigma^2 + 1.$$

From the requirement $1/4 < b^4\sigma^4 < 1/3$ further we deduce that $\widetilde{b}^4\widetilde{\sigma}^4 < 1/3$ which

according to assertion (ii) ensures the existence of the fourth moment of the process $\{\bar{X}_t\}_{t\in\mathbf{Z}}$. Thus it remains to prove that the first three mixed moments of the processes $\{X_t\}_{t\in\mathbf{Z}}$ and $\{\bar{X}_t\}_{t\in\mathbf{Z}}$ coincide. From Guegan (1984) and Kumar (1986) we know the explicit form of these moments. For $0 \le s \le u$ we have

$$E(X_t) = 0, \quad E(X_t X_{t-s}) = \delta_{s,o}\cdot\sigma^2/(1 - b^2\sigma^2), \quad E(X_t X_{t-s} X_{t-u}) = \delta_{s,1}\cdot\delta_{u,2}\cdot b\sigma^4/(1 - b^2\sigma^2).$$

Straightforward computations show that also

$$E(\bar{X}_t) = 0, \quad E(\bar{X}_t\bar{X}_{t-s}) = \delta_{s,o}\cdot\sigma^2/(1 - b^2\sigma^2), \quad E(\bar{X}_t\bar{X}_{t-s}\bar{X}_{t-u}) = \delta_{s,1}\cdot\delta_{u,2}\cdot b\sigma^4/(1 - b^2\sigma^2)$$

holds.▲

Proof of Theorem 3.1.

Proof of (i).

In the case $v = 1$ this result is due to Liu and Brockwell (1988). Applying the same method as Liu and Brockwell (1988), Liu and Liu (1990, cf. (3.1) - (3.10)) generalized this result to $v \ge 2$ also. In the case $v > 2$ Liu and Liu (1990) state the result without performing the proof explicitly. But the explicit proof is obvious as well from the way of proving the cases $v = 1$ and $v = 2$ as from some additional arguments [(2.21) - (2.25) in the reference above] also, though a lot of straightforward but tedious computations are necessary to write down the proof in a closed form.

Proof of (ii).

The first who recognized this result were Liu and Liu (1990, cf. Theorem 4.3 and Theorem 4.4). But they gave the explicit proof for $v = 1$ only. Since some more considerations are needed for proving the most general case $v \ge 1$ we will give the explicit proof for any arbitrary $v \in \mathbf{N}$ here.

Analoguously to the way in which Liu and Liu (1990) used to prove their Theorem 4.2 ($v = 1$) here we will verify the conditions for applying Corollary 5.4 and Theorem 5.5 of Hall and Heyde (1980) for proving the case $v > 1$. To show the assertion (i) the main step is to verify these conditions for γ being an unity vector in $\mathbf{R}^k$, $k\in\mathbf{N}$ arbitrary, but fixed. In view of the structure of the necessary conditions in Hall and Heyde (1980, Theorem 5.5) the generalization to any $\gamma \in \mathbf{R}^k$ is immediate then. Thus we restrict the proof to the case of γ being an unity vector in $\mathbf{R}^k$ here.

In this case the linear combination of processes considered in assertion (ii) reduces to a single process $\{\xi_{N,\underline{k}}(u)\}_{u \in [0,1]}$ with $\underline{k} \in K$,. According to Corollary 5.4 of Hall and Heyde (1980) the crucial condition to verify is

$$\sum_{m=1}^{\infty} \left\{ E[E(\eta_{\underline{k},0} \mid F_{-m})^2]^{1/2} + E[(\eta_{\underline{k},0} - E(\eta_{\underline{k},0} \mid F_m))^2]^{1/2} \right\} < \infty, \qquad (*)$$

where $F_m \equiv \sigma(\varepsilon_t, t \leq m)$.

According to the theory developed by Hall and Heyde (1980) this condition is used for approximating the sequence $\{\eta_{\underline{k},t}\}_{t \in Z}$ which is involved in $\{\xi_{N,\underline{k}}(u)\}_{u \in [0,1]}$ by a sequence of martingale difference schemes to which we can apply the classical martingale limit theory then.

If $\underline{k} = (0, k_1, \ldots, k_i)$, the definition of $\eta_{\underline{k},t}$ implies $E[\eta_{\underline{k},0} - E(\eta_{\underline{k},0} \mid F_m)] = 0$ for $m \geq k_i$, such that $(*)$ is already fulfilled if

$$\sum_{m=1}^{\infty} \left\{ E[E(\eta_{\underline{k},0} \mid F_{-m})^2]^{1/2} \right\} < \infty \qquad (1)$$

holds. For proving (1), the following decomposition of $\underline{X}_t$ turns out to be the basic tool. According to the proof of assertion (i) of Theorem 3.1, for $m > s$ it is possible to represent $\underline{X}_t$ as

$$\underline{X}_t = \underline{S}_0(t) + \sum_{n=1}^{m-s-1} \underline{\Delta}_n(t) + \sum_{n=m-s}^{\infty} \underline{\Delta}_n(t), \qquad (2)$$

with $\underline{S}_n(t)$ and $\underline{\Delta}_n(t)$ from (3.2) and (3.3) respectively, and where the infinite sum in (2) converges almost surely and in L^{2v} also. We want to remind that the variable 's' corresponds to the bilinear BL(p, q, s, r) model we consider here.

By the definition of $\underline{S}_n(t)$ and $\underline{\Delta}_n(t)$ we observe that for each $n \geq 0$ there exist measurable functions $\underline{f}_{n,1}$ and $\underline{f}_{n,2}$ such that for all $t \in Z$ we have

$$\underline{S}_n(t) = \underline{f}_{n,1}(\varepsilon_t, \ldots, \varepsilon_{t-s-n}) \text{ and } \underline{\Delta}_n(t) = \underline{f}_{n,2}(\varepsilon_{t-1}, \ldots, \varepsilon_{t-s-n}).$$

Hence, for $m > s$, the sum $\underline{U}_m(t) := \underline{S}_0(t) + \sum_{n=1}^{m-s-1} \underline{\Delta}_n(t)$ is independent of F_{-m} for all $t \in N_0$, whereas $\underline{\Sigma}_m(t) := \sum_{n=m-s}^{\infty} \underline{\Delta}_n(t)$ still depends on F_{-m}. In view of (2) these definitions yield a decomposition of $\underline{X}_t$ in

$$\underline{X}_t = \underline{U}_m(t) + \underline{\Sigma}_m(t). \qquad (2')$$

30

Now we take the vector form of $\underline{X}_t$ stated in (3.1) into account to obtain the following representation of $\eta_{\underline{k},0}$

$$\eta_{\underline{k},0} = \begin{cases} \underline{h}'\underline{X}_0\underline{X}'_{k_1}\underline{h}\underline{h}'\underline{X}_{k_2}\underline{X}'_{k_3}\underline{h} \,..\, \underline{h}'\underline{X}_{k_{i-1}}\underline{X}'_{k_i}\underline{h} - \mu_{k_i} & \text{if } i \text{ is even} \\[2ex] \underline{h}'\underline{X}_0\underline{X}'_{k_1}\underline{h}\underline{h}'\underline{X}_{k_2}\underline{X}'_{k_3}\underline{h} \,..\, \underline{h}'\underline{X}'_{k_i}\underline{h} - \mu_{k_i} & \text{if } i \text{ is odd} \end{cases} ,$$

where $\underline{h} \equiv (1, 0 ,\ldots, 0)^t \in \mathbf{R}^{p\vee r}$. Applying the decomposition (2') to $\eta_{\underline{k},0}$ for $m > s$ we obtain (we consider the case 'i is even' only, because the case 'i is odd' can be treated analogously):

$$\eta_{\underline{k},0} = \underline{h}^t\underline{U}_m(0)\underline{U}_m{}^t(k_1)\underline{h}\underline{h}^t\underline{U}_m(k_2)\underline{U}_m{}^t(k_3)\underline{h} \,..\, \underline{h}^t\underline{U}_m(k_{i-1})\underline{U}_m{}^t(k_i)\underline{h} -$$
$$- E[\text{the same expression}] +$$
$$\tag{3}$$
$$+ \text{meancorrected terms, where in each term at least one } \underline{\Sigma}_m(k_j) \text{ for } 0 \le j \le i \text{ occurs}$$

If now we consider $E(\eta_{\underline{k},0} \mid F_{-m})$ for $m > s$, the first term in (3) vanishes by applying the conditional expectation since it is independent of F_{-m}. Thus we restrict our attention to the other meancorrected terms, where in each term at least one $\underline{\Sigma}_m(k_j)$ for $0 \le j \le i$ occurs. Introducing the definition

$$V := \{\underline{V}_m : \underline{V}_m(k_j) \in \{\underline{U}_m(k_j), \underline{\Sigma}_m(k_j)\}, 0 \le j \le i, \text{ such that } \exists\, 0 \le j_0 \le i \text{ with}$$
$$\underline{V}_m(k_{j_0}) = \underline{\Sigma}_m(k_{j_0})\}$$

the conditional expectation $E(\eta_{\underline{k},0} \mid F_{-m})$ can be computed as

$$E(\eta_{\underline{k},0} \mid F_{-m})$$
$$= E\left[\sum_{\underline{V}_m \in V} \underline{h}^t\underline{V}_m(0)\underline{V}_m{}^t(k_1)\underline{h}\underline{h}^t\underline{V}_m(k_2)\underline{V}_m{}^t(k_3)\underline{h} \,..\, \underline{h}^t\underline{V}_m(k_{i-1})\underline{V}_m{}^t(k_i)\underline{h} \mid F_{-m}\right] -$$
$$- \{\text{the corresponding unconditioned expectation}\}.$$

By means of the inequality

$$E[\{E(X \mid F) - E(X)\}^2] \le \text{Var}(X) \le E(X^2)$$

applied to $X = \sum_{\underline{V}_m \in V} \underline{h}^t\underline{V}_m(0)\underline{V}_m{}^t(k_1)\underline{h}\underline{h}^t\underline{V}_m(k_2)\underline{V}_m{}^t(k_3)\underline{h} \,..\, \underline{h}^t\underline{V}_m(k_{i-1})\underline{V}_m{}^t(k_i)\underline{h}$ further we obtain

$$E[E(\eta_{\underline{k},0} \mid F_{-m})^2]$$

$$\leq E[\{ \sum_{V_m \in V} \underline{h}^t \underline{V}_m(0)\underline{V}_m{}^t(k_1)\underline{h}\underline{h}^t\underline{V}_m(k_2)\underline{V}_m{}^t(k_3)\underline{h} \ .. \ \underline{h}^t\underline{V}_m(k_{i-1})\underline{V}_m{}^t(k_i)\underline{h}\}^2],$$

$$\leq K \cdot \sum_{V_m \in V} E[\{\underline{h}^t\underline{V}_m(0)\underline{V}_m{}^t(k_1)\underline{h} \ .. \ \underline{h}^t\underline{V}_m(k_{i-1})\underline{V}_m{}^t(k_i)\underline{h}\}^2], \tag{4}$$

where K is a constant independent of m.

Hence for proving (1) it suffices to consider each summand in (4) seperately. Without loss of generality now we consider a summand with $\underline{V}_m(k_i) = \underline{\Sigma}_m(k_i)$. In this case we obtain

$$E[\{\underline{h}^t\underline{V}_m(0)\underline{V}_m{}^t(k_1)\underline{h} \ .. \ \underline{h}^t\underline{V}_m(k_{i-1})\underline{V}_m{}^t(k_i)\underline{h}\}^2]$$

$$= E[\sum_{n=m-s}^{\infty} \sum_{k=m-s}^{\infty} \underline{h}^t\underline{V}_m(0)\underline{V}_m{}^t(k_1)\underline{h} \ .. \ \underline{h}^t\underline{V}_m(k_{i-1})\underline{\Delta}_n{}^t(k_i)\underline{h}\underline{h}^t\underline{\Delta}_k(k_i)\underline{V}_m{}^t(k_{i-1})\underline{h} \ ..$$

$$.. \ \underline{h}^t\underline{V}_m(k_1)\underline{V}_m{}^t(0)\underline{h}]$$

$$\leq \sum_{n=m-s}^{\infty} \sum_{k=m-s}^{\infty} E[|\underline{h}^t\underline{V}_m(0)| \ |\underline{V}_m{}^t(k_1)\underline{h}|..|\underline{h}^t\underline{V}_m(k_{i-1})| \ |\underline{\Delta}_n{}^t(k_i)\underline{h}| \ |\underline{h}^t\underline{\Delta}_k(k_i)| \ |\underline{V}_m{}^t(k_{i-1})\underline{h}| \ ..$$

$$.. \ |\underline{h}^t\underline{V}_m(k_1)| \ |\underline{V}_m{}^t(0)\underline{h}|]$$

$$\leq \sum_{n=m-s}^{\infty} \sum_{k=m-s}^{\infty} E[||\underline{V}_m(0)||_{2v}^2 \, ||\underline{V}_m(k_1)||_{2v}^2 \cdots ||\underline{V}_m(k_{i-1})||_{2v}^2 \, ||\underline{\Delta}_n(k_i)||_{2v} \, ||\underline{\Delta}_k(k_i)||_{2v}],$$

$$\text{with } ||x||_{2v}^{2v} := \sum_{i=1}^{r} x_i^{2v} \text{ for } x \in \mathbf{R^r},$$

$$\leq \prod_{j=1}^{i-1} E[\, ||\underline{V}_m(k_j)||_{2v}^{2v}]^{1/(2v)} \cdot \{ \sum_{n=m-s}^{\infty} E[\, ||\underline{\Delta}_n(k_i)||_{2v}^{2v}]^{1/(2v)}\}^2 \tag{5}$$

by an application of Hölder's inequality.

Now we make use of a property of $\underline{\Delta}_n(t)$ which was used implicitely for proving our assertion (i), namely the validity of the inequality (3.4) which says

For each $t \in \mathbf{N}:$ $E\{||\underline{\Delta}_n(t)||_{2v}^{2v}\} \leq \text{const} \cdot \delta^{n/2}$, $\delta \in (0,1)$ for n large.

This result implies

$$\sum_{n=m-s}^{\infty} E\{||\underline{\Delta}_n(k_i)||_{2v}^{2v}\} \leq \text{const} \cdot \lambda^m \text{ for an } \lambda \in (0,1)$$

32

and

$$E\{\|\underline{V}_m(k_j)\|_{2v}^{2v}\} \leq \text{const} \quad \text{for } 0 \leq j \leq i-1 \text{ and } m > s,$$

where the constants can be chosen independently of $m > s$. Thus we can estimate (5) by

$$K'\cdot\lambda^{m/v}, \text{ with a constant } K' \text{ independent of } m > s.$$

Since the number of summands occuring in (4) is a finite number which is independent of m, and since each summand in (4) can be estimated by $\text{const}\cdot\lambda^{m/v}$ with $\lambda \in (0,1)$ (λ possibly changes from summand to summand), (1) and thus also ($*$) is proved now. Hence we can apply Theorem 5.5 and Corollary 5.4 of Hall and Heyde (1980) which yields the assertion (ii) for γ being an unity vector in $\mathbf{R}^k$ for any $k \in \mathbf{N}$.

In the general case of γ being any realvalued k–dimensional vector as well the validity of the crucial condition ($*$) as the validity of the other conditions required for applying the results of Hall and Heyde (1980) also, can be reduced easily to the basic case considered above.▲

The basic tool for proving the results concerning the asymptotic behaviour of the defined estimates is included in the following obvious result.

Lemma 4.1

Let $\{\underline{X}_N\}_{N \in N}$ be a sequence of random $k \times 1$ - vectors and $f = (f_1,..., f_m)^t : \mathbf{R}^k \to \mathbf{R}^m$ be a function which is continuously differentiable in a neighbourhood of $x_0 \in \mathbf{R}^k$.
Then we have

(i) *If $N^{1/2}(\underline{X}_N - x_0)$ is asymptotically (for $N \to \infty$) normal distributed, then $N^{1/2}\{f(\underline{X}_N) - f(x_0)\}$ is asymptotically normal distributed also*

(ii) *If $\underline{X}_N - x_0 = O(L_N)$ a.s., then $f(\underline{X}_N) - f(x_0) = O(L_N)$ a.s.*

In view of Theorem 3.1 and Lemma 4.1 above for proving the CLT and the LIL we will apply the following considerations.
Let $\underline{\theta}^0 = ((\underline{\theta}_1^0)^t, (\underline{\theta}_2^0)^t)^t$ denote the true parameter where $\underline{\theta}_i^0$ are allowed to be of different dimension and let $\widehat{\underline{\theta}}_i$ denote the corresponding estimators. Furthermore let us suppose

that $\underline{\theta}_i^0 = f_i(M_i)$, i = 1, 2, for some continuously differentiable functions f_i and vectors M_i whose components belong to the set of the theoretical mixed moments of the process $\{X_t\}_{t \in \mathbf{Z}}$. Often we can observe that within a small error term we can represent the estimates $\widehat{\underline{\theta}}_i$ as $\widehat{\underline{\theta}}_i = f_i(\widehat{M}_i)$, i = 1, 2, where $\widehat{M}_i$ denotes the empirical estimate of the vector M_i. If now for each i = 1, 2, $\widehat{M}_i - M_i$ fulfills the requirements of Lemma 4.1 we can conclude that both the multidimensional CLT and the LIL hold for the vector $\widehat{\underline{\theta}} = ((\widehat{\underline{\theta}}_1)^t, (\widehat{\underline{\theta}}_2)^t)^t$.

For the estimates we consider here the conditions for applying Lemma 4.1 can be satisfied by taking Theorem 3.1 into account which states the CLT and the LIL for the empirical mixed moments of the process $\{X_t\}_{t \in \mathbf{Z}}$.

Thus a desirable property for proving the asymptotic behaviour of an estimator $\hat{z}$ of z^0 in the way described above, is that *the pair $(z^0, \hat{z})$ fulfills the condition (CA) with moments of the order m*, i.e.

(i) there exists both an finite dimensional vector M_m whose components belong to the set of the theoretical mixed moments of the process $\{X_t\}_{t \in \mathbf{Z}}$ up to the order *m* and a function $f \in C^1 \{N(M_m)\}$ such that $z^0 = f(M_m)$

$$(4.1)$$

(ii) $|\hat{z} - f(\widehat{M}_m)| = \begin{cases} o_p(N^{-1/2}) \\ o(L_N) \text{ a.s.} \end{cases}$,where $\widehat{M}_m$ denotes the vector of the corresponding empirical estimates of the theoretical moments M_m.

Here we use $C^1 \{U\}$ as an abbrevation for the set of functions $\{f : U \rightarrow V, f$ is continuously differentiable in U$\}$ and N(x) denotes a suitable neighbourhood of x.

The condition (CA) is the crucial condition we will verify for the several estimates under our consideration.

Since often we will prove part (ii) of the condition (CA) for both error terms at the same time, for simplicity we introduce the following abbrevation.

NOTATION

Any sequence r_N with the properties $r_N = \begin{cases} o_p(N^{-1/2}) \\ o(L_N) \text{ a.s.} \end{cases}$ will be denoted by $\mathbf{o}_N$.

Now we apply this concept to the estimators of our interest.

Proposition 4.2

Under the assumptions of Theorem 3.2 the pair ($\underline{a}$, $\hat{\underline{a}}$) fulfills the condition (CA) with moments of the order 2.

Proof.

According to Corollary 2.2 the standardized superdiagonal bilinear process satisfies the Yule-Walker-equations with AR parametervector $\underline{a}$ for $s \geq w'$, i.e.

$$C_{p,w'-1} \, \underline{a} = \underline{c},$$

where $\underline{c} = (c(w'),\ldots, c(w'+p))^t$ and $C_{p,w'-1} = \{c(w'-1+i-j)\}_{1 \leq i, j \leq p}$.

Due to the assumption (A2) the vector $\underline{a}$ fulfills part (i) of the condition (CA) with $m = 2$ because both the vector $\underline{c}$ and the matrix $(C_{p,w'-1})^{-1}$ are continuously differentiable in its components in a neighbourhood of the second order mixed moments of the process. Since the Yule - Walker - estimator $\hat{\underline{a}}$ is defined by replacing the theoretical moments in the above equation with its empirical counterparts, the second part of the condition (CA) is satisfied too.▲

In view of this proposition now the proof of Theorem 3.2 is immediate.

Proof of Theorem 3.2

According to Theorem 3.1 the assumptions of the Theorem 3.2 imply that the conditions for applying Lemma 4.1 are fulfilled if we choose $\underline{X}_N$ as $\widehat{M}_2$ and x^0 as M_2, where M_2 denotes the vector of the first $w'+p$ second order mixed moments of the process and $\widehat{M}_2$ its empirical counterpart. In view of the Proposition 4.2 the result is obvious then.▲

For proving the Theorem 3.3 first we verify the condition (CA) in (4.1) for the pair ($\underline{\beta}_s^0$, $\widehat{\underline{\beta}}_s$), $0 \leq s \leq w-1$, where $\underline{\beta}_s^0$ and the corresponding estimate $\widehat{\underline{\beta}}_s$ are defined in (4.6) and (4.7) respectively.

Proposition 4.3

In a bilinear model of type I fulfilling the assumptions (A1) - (A4), we have for each $0 \leq s \leq w-1$

(i) $\widehat{\beta}_s \to \beta_s^0$ *a.s.*

(ii) *if moreover* $\rho(\Gamma_8) < 1$ *and* $\{\varepsilon_t\}_{t \in \mathbf{Z}}$ *is an 8k-th order symmetric innovation sequence, then the pair* $(\beta_s^0, \widehat{\beta}_s)$ *fulfills the condition (CA) with moments of the order 4.*

Proof.

In (2.5) we defined $\widehat{\beta}_s$, $0 \le s \le w-1$, to be the minimizer of the penalty function

$$\widehat{Q}_N^s (\beta_s) := \sum_{t=(r+s)\vee p+1}^N \{\widehat{v}_t \, \widehat{v}_{t-s} - E_{\beta_s}(v_t v_{t-s} \mid t\text{-w})\}^2 , \qquad (4.2)$$

if such a minimum exists. According to the linear representation of $E_{\beta_s}(v_t v_{t-s} \mid t\text{-w})$ in terms of β_s in (2.3), we recognize that $\widehat{Q}_N^s$ is a quadratic function in the components of β_s. Thus for minimizing $\widehat{Q}_N^s$ we consider the solutions of the equation

$$\nabla \widehat{Q}_N^s = 0 \qquad (1)$$

with $\nabla \widehat{Q}_N^s$ being a linear function of the parameter β_s. Hence, in general, the equation (1) has an unique solution which moreover can be computed explicitly. According to the linearity of $\nabla \widehat{Q}_N^s$ we have the expansion

$$\nabla \widehat{Q}_N^s (\widehat{\beta}_s) = \nabla \widehat{Q}_N^s (\beta_s^0) + (\widehat{\beta}_s - \beta_s^0)^t \, \nabla^2 \, \widehat{Q}_N^s (\beta_s^0). \qquad (2)$$

Thus, for proving the almost sure convergence of $\widehat{\beta}_s$ to β_s^0 we only have to verify the following conditions:

(C1) $\qquad N^{-1} \, \nabla \, \widehat{Q}_N^s (\beta_s^0) \to 0 \quad$ a.s.

(C2) $\qquad N^{-1} \, \nabla^2 \, \widehat{Q}_N^s (\beta_s^0) \to V^0 > 0 \quad$ a.s.

To prove part (ii) of the Proposition 4.3 moreover we need the condition

(C3) the pair $(0, N^{-1} \, \nabla \, \widehat{Q}_N^s (\beta_s^0))$ fulfills the condition (CA) in (4.1) with $m = 4$

to be satisfied.

According to Theorem 3.1 the general assumptions of Proposition 4.3 imply $E|X_t|^4 < \infty$. Hence we can take the ergodic theorem into account to deduce $\widehat{M}_4 - M_4 = o(1)$ a.s. So

the condition (C3) implies the condition (C1). Thus we restrict our effort on proving the conditions (C2) and (C3) only.

It should be noted here that for verifying these conditions we only apply the general assumptions of the proposition but not the additional assumption required for assertion (ii).

Check of (C2).

If we compute the second order derivatives of $\widehat{Q}_N^s$ with respect to $\underline{\beta}_s$ at $\underline{\beta}_s^0$ we obtain

$$\nabla^2 \widehat{Q}_N^s (\underline{\beta}_s^0) = 2 \sum_{t=(r+s)\vee p+1}^{N} (\nabla E_{\underline{\beta}_s^0})^t (\nabla E_{\underline{\beta}_s^0}).$$

This equation holds due to the linearity of $E_{\underline{\beta}_s}$ in $\underline{\beta}_s$ (cf. equation (2.3)) which implies $\nabla^2 E_{\underline{\beta}_s} = 0$ ($E_{\underline{\beta}_s}$ is used as an abbrevation for $E_{\underline{\beta}_s}(v_t v_{t-s} \mid t-w)$ here). Since $\nabla E_{\underline{\beta}_s}$ moreover is linear in $(1; X_{t-i} ; X_{t-j} X_{t-k} ; 1 \le i \le r+s, w \le j \le r, w+s \le k \le r+s)^t$ and the assumptions of Theorem 3.3 imply $E|X_t|^4 < \infty$, we can apply the ergodic theorem to deduce

$$N^{-1} \nabla^2 \widehat{Q}_N^s (\underline{\beta}_s^0) \rightarrow V := 2 E\{(\nabla E_{\underline{\beta}_s^0})^t (\nabla E_{\underline{\beta}_s^0})\} \text{ a.s.}$$

From the definition of the matrix V it is immediate that we always have $V \ge 0$. Thus for verifying the condition (C2) it remains to prove that if $\underline{\alpha}^t V \underline{\alpha} = 0$ for some $\underline{\alpha}$ then we have $\underline{\alpha} = 0$.

Since $E_{\underline{\beta}_s}$ is stationary in $t \in \mathbf{Z}$ we get

$$\underline{\alpha}^t V \underline{\alpha} = 0 \iff E[(\underline{\alpha}^t \nabla E_{\underline{\beta}_s^0})^2] = 0 \iff \underline{\alpha}^t \nabla E_{\underline{\beta}_s^0} = 0 \text{ a.s. for all } t \in \mathbf{Z}.$$

According to Lemma 4.6 now the last equivalence implies $\underline{\alpha} = 0$ such that we can deduce $V > 0$ which proves the condition (C2).

Check of (C3).

For verifying this condition we need to analyse the penalty function $\widehat{Q}_N^s$ in more detail. To this purpose we define

$$Q^s (\underline{\beta}_s , \underline{a}, M_4) := E[\{v_t(\underline{a}) v_{t-s}(\underline{a}) - E_{\underline{\beta}_s}(v_t v_{t-s} \mid t-w)\}^2] \tag{4.3}$$

where M_4 denotes a suitable vector of theoretical mixed moments of the process $\{X_t\}_{t \in \mathbf{Z}}$ up to the order four and $v_t(\underline{a}) \equiv X_t - \sum_{i=1}^p a_i X_{t-i}$ and $\underline{a} \equiv (a_1,\dots, a_p)^t$.

37

The fact that for $\underline{\beta}_s$ fixed the conditional expectation E_{β_s} belongs to span$\{1; X_{t-i}; X_{t-j} X_{t-k}$; $1 \le i \le r+s$, $w \le j \le r$, $w+s \le k \le r+s\}$ together with the linearity of $v_t(\underline{a})$ in X_{t-i}, $0 \le i \le p$ imply that Q^s is really a function of the moments up to the fourth order only. Thus Q^s is a polynomial in M_4. These features combined with the linearity of E_{β_s} in $\underline{\beta}_s$ moreover imply that Q^s and hence its partial derivatives $\partial/\partial\underline{\beta}_s \{Q^s(\underline{\beta}_s^0, \cdot, \cdot)\}$ are arbitrarily often differentiable functions also.

According to Proposition 4.2 now there exists a continuously differentiable function g with $\underline{a}^0 = g(M_2^0)$ ($= g(M_4^0)$, since M_4^0 can be chosen such that $M_2^0 \subset M_4^0$) and $\hat{\underline{a}} = g(\widehat{M}_4) + \mathbf{o}_N$, such that further we can conclude that there exists a continuously differentiable function f with the properties

$$f(M_4^0) = \frac{\partial}{\partial\underline{\beta}_s} \, Q^s(\underline{\beta}_s^0, \underline{a}^0, M_4^0)$$

and

$$f(\widehat{M}_4) = \frac{\partial}{\partial\underline{\beta}_s} Q^s(\underline{\beta}_s^0, \hat{\underline{a}}, \widehat{M}_4) + \mathbf{o}_N. \tag{3}$$

(In order to avoid misunderstandings caused by the notation, we denote the true underlying AR parametervector and the true theoretical moments of the process by $\underline{a}^0$ and M^0 respectively.)

On the other hand we can represent the quantity of interest, $N^{-1} \nabla \widehat{Q}_N^s (\underline{\beta}_s^0)$, by means of the function Q^s defined in (4.3) as follows

$$\frac{\partial}{\partial\underline{\beta}_s} Q^s(\underline{\beta}_s^0, \hat{\underline{a}}, \widehat{M}_4) = N^{-1} \nabla \widehat{Q}_N^s(\underline{\beta}_s^0) + \mathbf{o}_N. \tag{4}$$

According to its definition $\underline{\beta}_s^0$ minimizes $Q^s(\cdot, \underline{a}^0, M_4^0)$ and we get $\partial/\partial\underline{\beta}_s \{Q^s(\underline{\beta}_s^0, \underline{a}^0, M_4^0)\} = f(M_4^0) = 0$, which together with (3) and (4) proves (C3).

Hence the conditions (C1) - (C3) are verified (reminding that (C3) implies (C1) here) and it remains to show how these conditions can be used to deduce the assertions of the Proposition 4.3.

Since the proof of the assertion (i) is obvious we restrict our effort on proving the

assertion (ii) here.

Since $\nabla \widehat{Q}_N^s (\widehat{\beta}_s) = 0$ in view of the expansion in (2) we obtain

$$- N^{-1} \nabla \widehat{Q}_N^s (\beta_s^0) = (\widehat{\beta}_s - \beta_s^0)^t \, N^{-1} \nabla^2 \widehat{Q}_N^s (\beta_s^0).$$

By means of (C2) moreover we get

$$- (V^0)^{-1} N^{-1} \nabla \widehat{Q}_N^s (\beta_s^0) = (\widehat{\beta}_s - \beta_s^0)^t \, \{1 + o(1)\} \quad \text{a.s.} \tag{5}$$

At this stage we make use of the additional assumptions presumed in part (ii) of the proposition. According to Theorem 3.2 these assumptions imply that $\widehat{M}_4$ and M_4^0 fulfill the requirements of Lemma 4.1 with $\underline{X}_N = \widehat{M}_4$ and $x^0 = M_4^0$. Hence according to this Lemma and in view of the equation (5) the validity of the condition (C3) can be applied to deduce

$$(\widehat{\beta}_s - \beta_s^0)^t \cdot o(1) = \mathbf{o}_N$$

A second application of the condition (C3) moreover ensures the existence of an $f \in C^1$ $\{N(M_4^0)\}$ such that

$$(\widehat{\beta}_s - \beta_s^0)^t = (V^0)^{-1} \{f(\widehat{M}_4) - f(M_4^0)\} + \mathbf{o}_N.$$

If now we define

$$h(\cdot) \equiv (V^0)^{-1} \{f(\cdot) - f(M_4^0)\} + \beta_s^0,$$

then h is a continuously differentiable function with the properties $h(M_4^0) = \beta_s^0$ and $h(\widehat{M}_4)$ $= \widehat{\beta}_s + \mathbf{o}_N$ such that the pair $(\beta_s^0, \widehat{\beta}_s)$ fulfills the condition (CA) with m= 4, which proves (ii).▲

Proof of Theorem 3.3

According to the propositions 5.14 and 5.15 the pair $(\gamma^0, \widehat{\gamma}) := \{((\underline{a}^0)^t, (\beta_s^0)^t)^t , (\widehat{\underline{a}}^t ,$ $(\widehat{\beta}_s)^t)^t\}$ satisfies the condition (CA) in (4.1) with m = 4. Moreover, by means of Theorem 3.1, the assumptions of Theorem 3.3 imply that M_4^0 and $\widehat{M}_4$ hold the requirements for applying Lemma 4.1 to the pair $(\gamma^0, \widehat{\gamma})$. Thus the vector $N^{1/2} (\widehat{\gamma} - \gamma^0)^t$ is asymptotically jointly normal distributed and the LIL is valid also.

The asymptotic behaviour of the estimator of interest $\widehat{\underline{\theta}}$ now can be deduced from the asymptotics of the estimator $\widehat{\gamma}$ by applying continuously differentiable mappings to $\widehat{\gamma}$.

According to our procedure for estimating $\underline{\theta}^0$ we apply the Wilson algorithm to a selection of the components of the vector $\hat{\underline{\gamma}}$ first. Since in the neighbourhood of the true value this algorithm can be regarded as a continuously differentiable transformation, an application of Lemma 4.1 implies the desired asymptotic behaviour for the vector $((\hat{\underline{\gamma}} - \underline{\gamma}^0)^t , \hat{c}_j - c_j^0, 1 \leq j \leq q; \widehat{\sigma^2} - \sigma_0^2)^t$ to hold also. Furthermore the mapping from $(\hat{\underline{\gamma}}^t, \hat{c}_j, 1 \leq j \leq q; \widehat{\sigma^2})^t$ to $\hat{\underline{\theta}}$ (see (2.9)) is continuously differentiable in a neighbourhood of the true parameter. Again applying Lemma 4.1 now implies the assertion of the Theorem 3.3.▲

For the proof of Theorem 3.4 first we consider the estimates defined in the steps (ii) and (iii) of the estimation procedure seperately. This is the content of the following proposition.

Proposition 4.4

In a bilinear model of type II fulfilling the assumptions (A1) and (A4) we have

(i) *the pair $(b_{kk}\sigma^2, \widehat{b_{kk}\sigma^2})$ fulfills the condition (CA) with moments of the order 4, where $\widehat{b_{kk}\sigma^2}$ is defined in (2.14)*

(ii) *$\hat{\underline{\beta}} \rightarrow \underline{\beta}^0$ a.s, where $\hat{\underline{\beta}}$ minimizes the penalty function defined in (2.12). If moreover $\rho(\Gamma_8) < 1$ and $\{\varepsilon_t\}_{t\in\mathbf{Z}}$ is an 8k-th order symmetric innovation sequence , then the pair $(\underline{\beta}^0, \hat{\underline{\beta}})$ fulfills the condition (CA) with moments of the order 4.*

(iii) *the pair $(\sigma^2, \widehat{\sigma^2})$ fulfills the condition (CA) with moments of the order 4, where $\widehat{\sigma^2}$ is defined in (2.16).*

Proof.
Proof of (i).
First we observe that $\hat{u}_t \ (= X_t - \sum_{i=1}^p \hat{a}_i X_{t-i})$ is linear as well in $\hat{\underline{a}}$ as in X_{t-i} , $1 \leq i \leq p$ also. Since in the bilinear model of type II the Yule - Walker - equations already hold for $s \geq 1$, the assumption (A2) is always fulfilled here and we can apply Proposition 4.2 to deduce that the pair $(\underline{a}, \hat{\underline{a}})$ fulfills the condition (CA) in (4.1) with $m = 4$. If now we consider the equation (2.14) defining $\widehat{b_{kk}\sigma^2}$ we recognize that it is possible to find both a vector M_4 and a continuously differentiable function f such that $\widehat{b_{kk}\sigma^2} = f(\widehat{M_4}) + \mathbf{o}_N$ and $b_{kk}\sigma^2 = f(M_4)$ also. Thus the assertion is proved since both requirements of the condition

(CA) are satisfied.

Proof of (ii).

In (2.12) we defined $\hat{\underline{\beta}}$ to be the minimizer of

$$\hat{P}_N(\underline{\beta}) := \sum_{t=(r+l)\vee p+1}^{N} \left\{ \hat{u}_t\, \hat{u}_{t-k} - \widehat{b_{kk}\sigma^2}\, X_{t-k} - E_{\underline{\beta}}(u_t\, u_{t-k} - b_{kk}\sigma^2\, X_{t-k} \mid t\text{-}k\text{-}1) \right\}^2 .$$

In view of the explicit representation of $E_{\underline{\beta}}(\cdot)$ in terms of $\underline{\beta}$ in (4.15) we notice that $\hat{P}_N$ is a quadratic function in $\underline{\beta}$ and that it depends on the estimated values $\hat{\underline{a}}$ and $\widehat{b_{kk}\sigma^2}$. Since both the pair $(\underline{a}, \hat{\underline{a}})$ and the pair $(b_{kk}\sigma^2, \widehat{b_{kk}\sigma^2})$ fulfill the condition (CA) in (4.1) with m = 4 the main characteristics of the function $\hat{P}_N$ coincides with those of the penalty function $\hat{Q}_N^s$ defined in (4.2). Thus we can prove the stated assertion by going along the lines of the proof of Proposition 4.3 (ii).

Proof of (iii).

The proof is analogously to the proof of (i).▲

Proof of Theorem 3.4.

From the propositions 5.14 and 5.17 we conclude that the pair $(\underline{\gamma}^0,\ \hat{\underline{\gamma}}) := \{ (\underline{a}^t,\ b_{kk}\sigma^2,$ $(\underline{\beta}^0)^t,\ \sigma^2)^t\ ;\ (\hat{\underline{a}}^t,\ \widehat{b_{kk}\sigma^2},\ (\hat{\underline{\beta}})^t,\ \hat{\sigma}^2)^t \}$ fulfills the condition (CA) in (4.1) with m = 4. Due to Theorem 3.1, the assumptions of Theorem 3.4 imply that the theoretical moments gathered in M_4 and their empirical counterparts $\widehat{M}_4$ from the proof of Proposition 4.4 hold the requirements for applying Lemma 4.1 too. So we can conclude that the vector $N^{1/2}\,(\hat{\underline{\gamma}} - \underline{\gamma}^0)^t$ is asymptotically jointly normal distributed and the LIL holds. Obviously this is also tue for the transformed quantity $N^{1/2}\,(\hat{\underline{\theta}} - \underline{\theta}^0)^t$.▲

Proof of Theorem 3.5.

Analogously to the proof of Theorem 3.2 the following proposition is the crucial result we need for proving the Theorem.

Proposition 4.5

In a standardized superdiagonal bilinear model with expectation μ which fulfills (A1) and (A4), the following assertions hold:

(i)   *the pair ($\underline{a}$, $\hat{\underline{a}}$) fulfills the condition (CA) in (4.1) with moments of the order 2*

(ii) *the pair (μ, $\hat{\mu}$) fulfills the condition (CA) with moments of the order 1*

Proof.

Proof of (i).

The proof goes along the lines of the proof of Proposition 4.2 and thus is omitted.

Proof of (ii).

Choose f = id.▲

The proof of Theorem 3.5 now is immediate since the assumptions of the theorem imply that we can take Theorem 3.1 into account to verify the conditions (i) and (ii) of Lemma 4.1 by choosing $\underline{X}_N = \widehat{M}_2$ and $x_0 = M_2$. The assertion follows from the above proposition then.▲

Proof of Theorem 3.6.

The influence of estimating μ on the estimation of the AR coefficients of the bilinear model is studied in Theorem 3.5. Thus it remains to consider the effect of estimating μ on step (ii) and (iii) here. We have to replace the sample $\{X_1,\ldots, X_N\}$ with the sample $\{X_1 - \hat{\mu} ,\ldots, X_N - \hat{\mu}\}$. But in the deduction of the desired results this replacement causes no major difficulties. According to Proposition 4.5 we can make use of the same arguments as they were applied in the proof of the Theorems 5.3 to cope with the fact that the true parameter $\underline{a}$ is not available, but only its estimated version $\hat{\underline{a}}$. In both cases the influence of first estimating $\underline{a}$ and $((\underline{a})^t, \mu)^t$ respectively is not neglectable but treatable.▲

4.1. Auxiliary Results

The following Lemma is basic for proving the parametrization properties of the conditional expectations of the bilinear process (cf. chapter 2). Although our emphasis here is on analyzing bilinear processes only the result is given in a more general context.

Lemma 4.6

Let $X_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) + \varepsilon_t$ for $t \in \mathbf{Z}$ with a measurable function f and an iid sequence $\{\varepsilon_t\}_{t\in Z}$ such that the assumption (A4) is fulfilled. If for all $t \in \mathbf{Z}$

$$b + \sum_{i=1}^{r} c_i X_{t-i} + \sum_{j=1}^{m} \sum_{k=1}^{n} d_{jk} X_{t-j} X_{t-k} = 0 \quad a.s., \text{ where } b, c_i \text{ and } d_{jk} \in \mathbf{R}$$

then

$$b = c_i = d_{jk} = 0 \text{ for all } i, j, k \text{ in the summation domain.}$$

Proof.

A proof of this result can be found in Grahn (1993, proof of Lemma 5.20).▲

Appendix: Standardized bilinear time series

If $\{X_t\}_{t\in \mathbf{Z}}$ follows a stationary linear ARMA model we know that $E(X_t) = 0$ independent of the choice of the parameter a_i, c_j, and σ^2 determining the model. For modelling data sets with unknown mean $\mu \neq 0$ we extend the model class by admitting processes $\{X_t\}_{t\in \mathbf{Z}}$ such that $\{X_t - \mu\}_{t\in \mathbf{Z}}$ is a zero mean ARMA process also where μ is treated as an additional parameter then.

But in the bilinear case in general the mean of the bilinear process is a function of the parameter a_i, c_j, b_{ij} and σ^2. For example, let us consider the bilinear time series defined by $X_t = \alpha + b \varepsilon_{t-1} X_{t-1} + \varepsilon_t$. Then we obtain $E(X_t) = \alpha + b\sigma^2$. In most of the 'classical' literature concerning bilinear time series (e.g. Granger and Andersen (1978), Subba Rao and Gabr (1984), Liu and Brockwell (1988)) only the case $\alpha = 0$ is considered such that the expectation of the bilinear series is uniquely determined by the parameter of the process.

Here we introduce the definition of standardized bilinear time series which have the common property of a zero mean. Especially for fitting bilinear processes to real data sets we find it more natural to classify the bilinear models with respect to its mean than to the parameter α in (1.1) which would be possible also.

Definition (standardized bilinear time series)

A bilinear time series of type BL(p, q, k, r) defined by means of the equations (1.1) is said to be a <u>standardized bilinear time series</u> if $\alpha \in \mathbf{R}$ is such that $E(X_t) = 0$. Moreover in accordance with the ARMA case we call $\{X_t\}_{t \in \mathbf{Z}}$ a standardized bilinear time series with expectation μ if the process $\{X_t - \mu\}_{t \in \mathbf{Z}}$ is a standardized bilinear time series.

In a standardized bilinear time series with expectation μ the parameter of interest is the vector $\underline{\theta} = (\mu, a_i, c_j, b_{ij}, \sigma^2)^t$ with α excluded since due to the standardization of the bilinear process the value of α is uniquely determined by the parameter $\underline{\theta}$ (e.g. the standardized version of the bilinear time series $X_t = \alpha + b\,\varepsilon_{t-1}X_{t-1} + \varepsilon_t$ corresponds to $\alpha = -b\sigma^2$). The models defined in (1.2) and (1.3) are standardized bilinear models too.

Moreover both models belong to the more general class of superdiagonal bilinear models. In the literature there is some confusion concerning this naming (cf. Tong (1990), p. 112) but here we will call a general bilinear time series to be of the <u>*superdiagonal*</u> type iff $b_{ij} = 0$ for $i > j$. Analogously we define the <u>*subdiagonal*</u> bilinear model ($b_{ij} = 0$ for $i < j$) and the <u>*diagonal*</u> bilinear model ($b_{ij} = 0$ for $i \neq j$).

A general result of Liu and Brockwell (1988, Theorem 3.1) provides us with sufficient conditions guaranteeing the existence of a stationary, ergodic and causal solution of the bilinear equations (1.1) in the case $\alpha = 0$. In order to apply this result to the case $\alpha \neq 0$ also we make use of the following considerations where we apply the state space representation of the bilinear process given in (3.1).

If we assume the matrix $I - A$ defined therein to be invertible, which e.g. is fulfilled if $\rho(A) < 1$, we can apply the equation in (3.1) to $\underline{X}_t$ to conclude that

$$\{\underline{X}_t - (I-A)^{-1}\underline{\alpha}\ \}$$

$$= \underline{\alpha} + A\,\{\underline{X}_{t-1} - (I-A)^{-1}\underline{\alpha}\ \} + A(I-A)^{-1}\underline{\alpha} + \sum_{i=1}^{k} \varepsilon_{t-i}\,B_i\,\{\underline{X}_{t-1} - (I-A)^{-1}\underline{\alpha}\ \} +$$

$$+ \sum_{i=1}^{k} \varepsilon_{t-i}\,B_i\,(I-A)^{-1}\underline{\alpha} + C\,\underline{\varepsilon}_t - (I-A)^{-1}\underline{\alpha}$$

$$= A \{\underline{X}_{t-1} - (I-A)^{-1}\underline{\alpha} \} + \sum_{i=1}^{k} \varepsilon_{t-i} B_i \{\underline{X}_{t-1} - (I-A)^{-1}\underline{\alpha} \} + C\underline{\varepsilon}_t + \sum_{i=1}^{k} \varepsilon_{t-i} B_i(I-A)^{-1}\underline{\alpha}$$

since $\underline{\alpha} + A(I-A)^{-1}\underline{\alpha} - (I-A)^{-1}\underline{\alpha} = \underline{\alpha} + (A-I)(I-A)^{-1}\underline{\alpha} = 0,$

$$= A \{\underline{X}_{t-1} - (I-A)^{-1}\underline{\alpha} \} + \sum_{i=1}^{k} \varepsilon_{t-i} B_i \{\underline{X}_{t-1} - (I-A)^{-1}\underline{\alpha} \} + \widetilde{C} \underline{\varepsilon}_t ,$$

with $\underline{\varepsilon}_t := (\varepsilon_t , \varepsilon_{t-1},\dots, \varepsilon_{t-q\vee k})^t$ and $\widetilde{C} = \widetilde{C}(\alpha) = \begin{pmatrix} 1 & \widetilde{c}_1 & \cdots & \widetilde{c}_{q\vee k} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \in M(h, q\vee k+1),$

where the coefficients $\widetilde{c}_i$ are uniquely determined by the moving average coefficients c_j and the first components of the vectors $B_i (I-A)^{-1} \underline{\alpha}$, $1 \le i \le k$. (Here 'q∨k' denotes the maximum of q and k.) Hence we recognize that provided that the matrix $I - A$ is invertible, the process $\{\underline{X}_t - (I-A)^{-1}\underline{\alpha} \}_{t\in \mathbf{Z}}$ satisfies the state space equation

$$\underline{Y}_t = A \underline{Y}_{t-1} + \sum_{i=1}^{k} \varepsilon_{t-i} B_i \underline{Y}_{t-1} + \widetilde{C} \underline{\varepsilon}_t ,$$

$$Y_t = D\underline{Y}_t. \tag{1}$$

Thus a change of α in (3.1) effects the bilinear representation in (1) only in the moving average part.

Liu and Brockwell (1988) proved the existence of an ergodic and strictly stationary solution of (3.1) in the case $\underline{\alpha} = 0$. The crucial condition for their result to hold is $\rho(\Gamma) < 1$, where Γ is a matrix defined implicitly in their equation (3.8). Going along the lines of the proof of Liu and Brockwell's Theorem 3.1 (cf. the first part of chapter 3 also) we notice that the matrix Γ is independent of the moving average coefficients of the bilinear process under consideration. Hence the condition $\rho(\Gamma) < 1$ is sufficient for bilinear processes defined by means of the equations (1) too. If $\alpha \ne 0$ and $I - A$ is invertible, this result carries over to the transformed process $\{\underline{X}_t - (I-A)^{-1}\underline{\alpha} \}_{t\in \mathbf{Z}}$ also since the process fulfills (1) then. Thus for any $\alpha \in \mathbf{R}$ the value of α does not influence the validity of the condition $\rho(\Gamma) < 1$ as long as we only consider bilinear processes where the matrix $I - A$ is invertible.

Thus, analogously to the linear case in the bilinear case also the condition guaranteeing

the stationarity and ergodicity of the bilinear process does not depend on the moving average part of the process.

References

Andel, J. (1989) : On nonlinear models for time series. *Statistics* 20, 4, 615 - 632.

Brillinger, D. R. (1981) : *Time series: Data analysis and theory.* Holt, Rinehart & Winston, New York.

Brockwell, P. and Davis, R. A. (1991) : *Time series. Theory and Methods. (second edn).* New York, Springer

Grahn, T. (1993) : A new approach to bilinear time series estimation, *PhD thesis,* University of Heidelberg.

Granger, C. W. and Andersen, A. P. (1978) : *An introduction to bilinear time series models.* Vandenhoek and Ruprecht, Gottingen.

Guegan, D. (1981) : Etude d'un modèle non linéare, le modèle superdiagonal d' ordre 1. *C. R. Acad. Sci. Paris*, Serie I, Math. 448, 159 - 162.

—— (1984) : Test de modèles non linéares. *Alternative Approaches to Time Series Analysis* (J. P. Florens, M. Mouchart, J. P. Raoult and L. Simar, eds.). Facultes Universitaires Saint - Louis, Bruxelles, 45 - 77.

—— and Pham, D. T. (1989) : A note on the estimation of the parameters of the diagonal bilinear model by the method of least squares. *Scand J. Statist.* 16, 129 - 136.

Hall, P. and Heyde, C.C. (1980) : *Martingale limit theory and its Application.* Academic Press, New York.

Jou, Y. J. (1989) : On the bilinear time series model BL(p, 0, p, 1). Ph. D. thesis, University of Pennsylvania.

Kim, W. K. and Billard, L. (1990) : Asymptotic properties for the first - order bilinear time series model. *Commun. Stat.-Theory Meth.*, 19(4), 1171 - 1183.

—— and Basawa, I. V. (1990) : Estimation for the first - order diagonal bilinear time series model. *J. Time Ser. Anal.* 11, 215 - 227.

Kumar, K. (1986) : On the identification of some bilinear time series. *J. Time Ser. Anal.* 7, 117 - 122.

Liu, J. and Brockwell, P. (1988) : On the general bilinear time series model. *J. Appl. Prob.* 25, 553 - 564.

Liu, J. (1990) : Estimation for some bilinear time series. *Stochastic. Models*, 6(4), 649 - 665.

—— and Chen, Z. G. (1990) : Bilinear (p, 0, 1, 1) model and its consistent identification. *Tech. Report*, Dept. of Stat., Univ. of British Columbia.

—— and Liu, Z. (1990) : Higher order moments and limit theory of a general bilinear time series. *Tech. Report*, Dept. of Stat., Univ. of British Columbia.

—— (1992) : Spectral radius, Kronecker products and stationarity. *J. Time Ser.*

Anal. 13, 319 - 327.

Mohler, R. R. (1973) : *Bilinear control processes*. Academic Press, New York.

Pham, D. T. and Tran, T. L. (1981) : On the first order bilinear time series models. *J. Appl. Probab.* 18, 617 - 627.

Seasay, S. A. O. and Subba Rao, T. (1988) : Yule - Walker type difference equations for higher - order moments and cumulants for bilinear time series models. *J. Time Ser. Anal.* 9, 385 - 401.

—— (1991) : Difference equations for higher order moments and cumulants for the bilinear time series model BL(p, 0, p, 1). *J. Time Ser. Anal.* 12, 159 - 176.

—— (1992) : Frequency - domain estimation of bilinear time series models. *J. Time Ser. Anal.* 13, 521 - 545.

Subba Rao, T.. (1981) : On the Theory of Bilinear Time Series Models. *J. R. Stat. Soc.*, B43, 244 - 255.

—— and Gabr, M. M. (1984) : *An introduction to bispectral analysis and bilinear time series models.* Lecture Notes in Statistics, 24. Springer, New York.

—— and da Silva, M. E. A. (1989) : Identification and estimation of bilinear time series models. *Tech. Rep. 195*, Dep. of Stat., University of Manchester Institute of Science and Technology.

Tang, Z. and Mohler, R. R. (1988) : Bilinear time series: theory and applications. In *Nonlinear Time Series and Signal Processing* (ed. R. R. Mohler), Lecture notes in Control and Information Sciences, Springer.

Tjøstheim, D. (1992) : Nonlinear time series : A selective review. Preprint No. 694, SFB 123, University of Heidelberg.

Tong, H. (1990) : *Non - linear time series.* Oxford University Press, Oxford.

Wilson, G. T. (1969) : Factorization of the generating function of a pure moving average process. *SIAM J. Num. Analysis*, 6, 1 - 7.

Wittwer, G. (1989) : Some remarks on bilinear time series models. *Statistics* 20, 4, 521 - 529.