A note on the properties of some time varying bilinear models

Abdelouahab Bibi\textsuperscript{a,}\textsuperscript{*}, Alwell J. Oyet\textsuperscript{b}

\textsuperscript{a}Département de Mathématiques, Université Mentouri de Constantine, 25000 Algeria
\textsuperscript{b}Department of Mathematics and Statistics, Memorial University of Newfoundland, Canada

Received August 2001; received in revised form April 2002

Abstract

In this note, a sufficient condition is given for the existence and uniqueness of a stable causal solution for bilinear time series with time-varying coefficients; also some conditions for invertibility and the optimal prediction procedure are given. The notions of controllability, observability and minimality are discussed. © 2002 Elsevier Science B.V. All rights reserved.

MSC: primary 62M10; secondary 62M05

Keywords: Time-varying bilinear models; Quasi-stationary processes; Causality; Invertibility; Controllability; Observability; Minimality

1. Introduction

Many time series exhibit certain characteristics which cannot be explained using standard autoregressive (AR), moving average (MA) or mixed autoregressive moving average (ARMA) models. Indeed, in many practical situations, the stochastic processes under study are nonstationary. For instance, the economies of many developing countries show signs of steady growth which indicates that there is a typical economic upward trend through time. Thus, the use of ARMA models with constant coefficients is inappropriate for these series. Several approaches have been studied in the literature for dealing with nonstationary time series. A popular approach is to plot the series, observe and extract the components of the series that are responsible for the nonstationarity such as trend and seasonality. Then the stationary component is modelled by one of the standard AR, MA or ARMA models. It is well known that this method may not always work. Thus, Hallin (1986) proposed using some suitable linear model with time-varying coefficients for describing nonstationary time series. There application is however limited to linear processes and there are time series

\textsuperscript{*} Corresponding author. Tel.: +213-31-66-37-86; fax: +213-31-66-37-86.

\textit{E-mail addresses:} bibi@wissal.dz (A. Bibi), aoyet@math.mun.ca (A.J. Oyet).
which are known to be both nonstationary and nonlinear. For instance, from economic theory, most of the stock market indexes are known to be martingale difference sequences (but not independent identically distributed sequences). Moreover, since the economy changes due to unforeseen interventions, it is difficult to justify using the same (nonlinear) model over a long period. For such series, a nonlinear model with time-varying coefficients may be more realistic. We believe that bilinear time series with time-varying coefficients may be a useful tool in describing the behavior of a wide class of dynamical time series. We consider a bilinear model with time-varying coefficients because it seems to be the simplest extension of the linear model, defined by adding terms to a classical ARMA model. Recently, Subba Rao (1997) briefly discussed special cases of the general bilinear time series model with time-varying coefficients, based on second and higher order evolutionary spectra. Bibi (2001) then extended the probabilistic properties of some stationary bilinear models to time-varying ones. The objectives of this note are to give conditions for the existence and uniqueness of a stable causal solution of the bilinear time series model with time-varying coefficients; and to extend the result of Guégan and Pham (1987), concerning the invertibility of stationary bilinear time series, to time-varying ones. For this purpose, relevant state-space representations are used.

In this note, we consider a time-varying bilinear (TVBL) time series \((X_t)_{t \in \mathbb{Z}}\), \(\mathbb{Z} = \{0, \pm 1, \pm 2 \ldots\}\) defined on a probability space \((\Omega, \mathcal{F}, P)\) which admits the general bilinear representation

\[
X_t = \sum_{i=1}^{p} a_i(t)X_{t-i} + \sum_{i=1}^{q} b_i(t)\xi_{t-i} + \sum_{k=1}^{R} \sum_{l=1}^{K} c_{lk}(t)X_{t-l}\xi_{t-k} + \zeta_t
\]

denoted by TVBL\((p, q, K, R)\), where \((a_i(t))_{1 \leq i \leq p}\), \((b_i(t))_{1 \leq i \leq q}\), and \((c_{lk}(t))_{1 \leq l \leq K, 1 \leq k \leq R}\) are the time-varying coefficients of the model; \((\xi_t)_{t \in \mathbb{Z}}\) is a strong white noise, i.e. a sequence of independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance \(\sigma^2 = 1\). By virtue of the bilinear representation above, \(\xi_t\) is independent of \(X_s\) for all \(s < t\). We call \((X_t)_{t \in \mathbb{Z}}\) stable if \(\text{Var}\{X_t\}\) is bounded. Notice that the time-varying ARMA\((p, q)\) model can be deduced from the TVBL\((p, q, K, R)\) model by setting \(c_{ij}(t) = 0\) for all \(i, j\). Thus a study of the TVBL\((p, q, K, R)\) model automatically includes a study of the ARMA\((p, q)\) model with time-varying coefficients.

Due to the nonlinear dependence between \(X_t\) and \(\xi_{t-k}\), \(k \geq 1\) it is difficult to handle the multiplicative terms with nonzero coefficients in a general representation, so that we will consider only the superdiagonal model TVBL\((p, q, p, q)\) with \(p \geq q\), defined by

\[
X_t = \sum_{i=1}^{p} a_i(t)X_{t-i} + \sum_{i=1}^{q} b_i(t)\xi_{t-i} + \sum_{k=1}^{q} \sum_{l=k}^{p} c_{lk}(t)X_{t-l}\xi_{t-k} + \xi_t. \tag{1.1}
\]

Throughout this paper, we use the following notations. For any square matrix \(A\), the spectral radius of \(A\) is denoted by \(\rho(A) = \max_i \{ |\lambda_i(A)| \} \) where \(\lambda_i(A)\) is the \(i\)th eigenvalue of \(A\). We denote by \(A \otimes B\) the Kronecker product of two matrices \(A\) and \(B\). The Kronecker product of \(A\) by itself is denoted by the symbol \(A^{\otimes 2}\). For any process \((Y_t)_{t \in \mathbb{Z}}\), we denote by \(\mathcal{F}(Y, t)\) the \(\sigma\)-field generated by \(\{Y_s, s \leq t\}\).

We begin Section 2 with a discussion of the Markovian representation of the TVBL\((p, q, p, q)\) model (1.1). This representation is then used in Section 3 to derive conditions under which a causal relationship exist between the white noise process \((\xi_t)_{t \in \mathbb{Z}}\) and the series \((X_t)_{t \in \mathbb{Z}}\). For a model to be
useful in forecasting future values, it has to be invertible. Thus, Section 4 deals with the invertibility conditions of model (1.1). The concepts of “controllability”, “observability” and “minimality” which are closely related to the concept of “causality” and “invertibility” are discussed in Section 5. The relationship between these concepts is also discussed at the end of Section 5.

2. Markovian bilinear representation and its properties

The concept of Markovian state space representation plays an important role in the analysis of time series models, in particular in connection with Kalman filter. It is well known that time-varying ARMA models admit a linear Markovian state space representation (see Grenier, 1987). It would be helpful to generalize the linear Markovian representation to include bilinear time series models with time-varying coefficients. The procedure proposed in Pham (1985) for building a Markovian state space representation in the case of time invariant coefficients can be easily extended to the present one. Let \( n = p + q \). For \( j = 1, \ldots, q \) let \( \beta_j(t) = b_j(t + j) + \sum_{i=j}^p c_{ij}(t + j)X_{t+i-j} \) and define the \( n \)-dimensional vector \( \eta_i = (\eta_i^1, \eta_i^p, \eta_i^{p+1}, \ldots, \eta_i^q)' \) where \( \eta_i^j = \eta_{t-j+i} \) for \( i = 1, \ldots, p \) and

\[
\eta_i^{p+i} = \begin{cases} 
X_{t+1}, & \quad i = 1, \\
\sum_{k=1}^p a_k(t+i)X_{t+i-k} + \sum_{k=i}^q \beta_k(t+i-k)\xi_{t+i-k}, & \quad 2 \leq i \leq q. 
\end{cases}
\]

Then \( X_t = H'\eta_t \), where the observation matrix \( H' = (O_{1 \times p}, 1, O_{1 \times (q-1)}) \). It can be verified that

\[
\eta_i^{p+i} = \begin{cases} 
\eta_{i-1}^{p+2} + a_1(t+1)X_t + \beta_1(t)\xi_t + \xi_{t+1}, & \quad i = 1, \\
\eta_{i-1}^{p+i+1} + a_i(t+i)X_t + \beta_i(t)\xi_t, & \quad 2 \leq i \leq (q-1), \\
\sum_{k=i}^q a_k(t+q)X_{t+q-k} + \beta_q(t)\xi_t, & \quad i = q.
\end{cases}
\]

Note that based on these notations, \( X_t \equiv \eta_{t-1}^{p+1} \) and \( X_{t+i-j} \equiv \eta_{t-1}^{p+i-1-j} \). Define the \( n \)-dimensional state vector \( x_t = \eta_t - H\xi_{t+1} \) and

\[
c(t) = (O_{1 \times (p-1)}, 1, a_1(t+1) + b_1(t+1), \ldots, a_q(t+q) + b_q(t+q))',
\]

\[
d(t) = C(t)H = (O_{1 \times p}, c_{11}(t+1), \ldots, c_{qq}(t+q))'.
\]

Then, in terms of the matrices

\[
C_{(p+q) \times (p+q)}(t) = \begin{pmatrix} O_{p \times p} & O_{p \times q} \\ \hat{C}_{q \times p}(t) & \hat{C}_{q \times q}(t) \end{pmatrix}, \quad A_{(p+q) \times (p+q)}(t) = \begin{pmatrix} \hat{A}_{p \times p} & \hat{A}_{p \times q} \\ \hat{A}_{q \times p}(t) & \hat{A}_{q \times q}(t) \end{pmatrix}
\]
with
\[
\hat{C}_{q \times p}(t) = \begin{pmatrix}
0 & c_p(t + 1) & \cdots & c_{21}(t + 1) \\
0 & 0 & c_{p2}(t + 2) & \cdots & c_{32}(t + 2) \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & c_{pq}(t + q) & \cdots & c_{q+1}(t + q)
\end{pmatrix},
\hat{I}_{p \times p} = \begin{pmatrix} 0 & I_{p-1} & 0 \\
0 & \cdots & 0 
\end{pmatrix}
\]

\[
\hat{C}_{q \times q}(t) = \begin{pmatrix}
c_{11}(t + 1) \\
c_{22}(t + 2) \\
\vdots \\
c_{qq}(t + q)
\end{pmatrix},
\hat{O}_{p \times q} = \begin{pmatrix} O_{(p-1) \times q} \\
1 & O_{1 \times (q-1)} 
\end{pmatrix}
\]

\[
\hat{A}_{q \times p}(t) = \begin{pmatrix}
0 & O_{(q-1) \times p} \\
0 & a_p(t + q) & \cdots & a_{q+1}(t + q)
\end{pmatrix},
\hat{A}_{q \times q}(t) = \begin{pmatrix}
a_1(t + 1) & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 1 \\
a_q(t + q) & 0 & \cdots & 0
\end{pmatrix}
\]

the state-space representation of (1.1) can be written as
\[
\begin{cases}
\dot{x}_t = (A(t) + \xi(t)C(t))x_{t-1} + c(t)\xi_t + d(t)\xi_t^2, \\
X_t = H'x_{t-1} + \xi_t.
\end{cases}
\tag{2.1}
\]

Let \( \mu_t \) be the mean vector of \( x_t \). Taking expectations of both sides of (2.1) we obtain \( \mu_t = A(t)\mu_{t-1} + d(t) \). This is a linear first-order difference equation in \( \mu_t \). The solution to this equation can be written as \( \mu_t = \sum_{k=0}^{\infty} v_k(t) \), where \( v_0(t) = d(t) \) and \( v_k(t) = \{ \prod_{j=0}^{k-1} A(t - j) \} d(t - k) \), for \( k \geq 1 \). A sufficient condition for the convergence of such series (see Miller, 1968, Chapter 2) is that for each fixed \( t \) : \( \det(I_{(p+q)} - \tau A(t)) \neq 0 \) for all \( \tau \in \mathbb{C} : |\tau| \leq \lambda \) or equivalently the operator \( \phi_t(B) = 1 - \sum_{i=1}^{p} a_i(t)B^i \) associated with the autoregressive (AR) part of model (1.1) satisfy a regularity condition, namely, the zeros of polynomial \( \phi_t(z) \) lie in the region \( |z| > \lambda \), where \( \lambda > 1 \) and \( B \) is a lag operator \( (B'X_t = X_{t-1}) \). This regularity condition ensures the finiteness of the first-order moment. Now, assume that this condition is satisfied and let \( \tilde{x}_t = x_t - \mu_t \) be the mean deleted state vector, then we obtain the following state-space representation
\[
\begin{cases}
\dot{\tilde{x}}_t = D(t)\tilde{x}_{t-1} + w(t)\tilde{\xi}_t + d(t)(\tilde{\xi}_t^2 - 1), \\
\tilde{X}_t = H'\tilde{x}_{t-1} + \tilde{\xi}_t.
\end{cases}
\tag{2.2}
\]
where \( D(t) = (A(t) + \xi C(t)) \), \( \tilde{X}_t = X_t - H'\mu_{t-1} \), and where \( w(t) = C(t)\mu_{t-1} + c(t) \). The above discussion is summarized in the following theorem.

**Theorem 2.1.** The model described by the TVBL\((p, q, p, q)\) representation (1.1) admits the Markovian state-space representation given by (2.2) if \( 1 - \sum_{i=1}^{p} a_i(t)z^i \neq 0 \) for all \( z \in \mathbb{C} : |z| \leq 1 \).

In what follows, we outline sufficient conditions for the existence of a causal relationship between the processes \( X_t \) and \( \tilde{\xi}_t \).

### 3. Causality

Model (1.1) is said to be causal if there exists a measurable function \( f : IR^\infty \rightarrow IR \) such that \( \forall t \in \mathbb{Z} : X_t = f(\tilde{\xi}_t, \tilde{\xi}_{t-1}, ...) \) a.s. By successive substitution in (2.2) we find that a formal series solution to (1.1) has the form

\[
\tilde{X}_t = \sum_{k=0}^{\infty} g_k(t),
\]

where \( g_0(t) = \tilde{\xi}_t \) and \( g_k(t) = H'\{(\prod_{j=1}^{k-1} D(t-j))Z(t-k)\} \) for \( k \geq 1 \), with \( Z(t) = w(t)\tilde{\xi}_t + d(t)\tilde{\xi}_t^2 - 1 \). Note that the process \( (g_k(t), k \geq 0)_{t \in \mathbb{Z}} \) is an orthogonal process. Thus, the problem is to find conditions on \( (g_k(t), k \geq 0)_{t \in \mathbb{Z}} \) such that formal series (3.1) converges in mean square sense for all \( t \) and thus defines a stable process solution to (1.1).

**Theorem 3.1.** Let \((X_t)_{t \in \mathbb{Z}}\) be a process generated by the TVBL\((p, q, p, q)\) model and admitting the representation (2.2). Then, model (1.1) admits a unique stable solution given by (3.1) if and only if

\[
\sum_{k \geq 0} E\{g_k^2(t)\} < +\infty \text{ for all } t \geq 0,
\]

where \( E\{g_k^2(t)\} = (H^{\otimes 2})'\prod_{j=1}^{k-1}\{D^{(2)}(t-j)\}(w^{\otimes 2}(t-k) + 2d^{\otimes 2}(t-k)) \) for \( k \geq 1 \) with \( D^{(2)}(t) = A^{\otimes 2}(t) + C^{\otimes 2}(t) \).

**Proof.** Follows from the Riesz–Fisher Theorem, cf. Doob (1953, p. 155). \(\Box\)

In the theorems that follow, we outline sufficient conditions for (3.2) to hold under the assumption that \( E\log|g_k(t)|\) exists for all \( t \) and for \( k \geq 1 \). Notice that if model (1.1) admits a stable solution given by (3.1), then \( \lim_{k \rightarrow \infty} \sup_{t \in \mathbb{Z}} \frac{1}{k} E\{\log|g_k(t)|\} < 0 \). Indeed, since \( E\{g_k^2(t)\} \rightarrow 0 \) as \( k \rightarrow \infty \) for all \( t \), then the results follow immediately from Jensen’s inequality.

**Theorem 3.2.** Model (1.1) admits a stable solution given by (3.1) if

\[
\lambda = \lim_{k \rightarrow \infty} \sup_{t \in \mathbb{Z}} \frac{1}{k} \log \left| \prod_{j=1}^{k-1} D^{(2)}(t-j) \right| < 0.
\]
Proof. The results follow from Cauchy’s criterion for convergence of a series. □

Example 3.1. Consider the TVBL(1,0,1,1) model defined by

\[ X_t = (a(t-1) + c(t-1)\xi_{t-1})X_{t-1} + \xi_t. \]  

This model admits the following Markovian representation:

\[ \tilde{x}_t = D(t)\tilde{x}_{t-1} + (a(t) + c(t)\mu_{t-1})\tilde{\xi}_t + c(t)(\tilde{\xi}_t^2 - 1), \]

where \( D(t) = a(t) + c(t)\tilde{\xi}_t \) with \( \tilde{X}_t = \tilde{x}_{t-1} + \tilde{\xi}_t \).

The sufficient condition given in the above theorem becomes \( \lim_{k \to \infty} \sup_{t \in \mathbb{Z}} \frac{1}{k} \sum_{i=1}^{k-1} \log(a^2(t-i) + c^2(t-i)) < 0 \). This condition can be interpreted as the second-order moment Lyapunov uniform exponents for the time-varying bilinear model (3.4).

In practice, Condition (3.3) is of little use for checking the existence of a unique stable solution for model (1.1) since this condition involves the limit of products of infinitely many matrices. In time-invariant case, this condition reduces to \( \lambda = \lim_{k \to \infty} \sup_{t \in \mathbb{Z}} \frac{1}{k} \log \|\sum_{i=1}^{k-1} D^{(2)}(t)\| \leq 0. \) Assume that there exists a constant \( \lambda^{(2)} > 0 \) such that \( \|D^{(2)}(t)\| \leq \lambda^{(2)}. \) In this case we can show that

\[ \lambda = \lim_{k \to \infty} \sup_{t \in \mathbb{Z}} \frac{1}{k} \log \left\{ \prod_{j=1}^{k-1} D^{(2)}(t-j) \right\} \leq \log \lambda^{(2)}. \]

This leads to the following theorem.

Theorem 3.3. Assume that there exists a constant \( \lambda^{(2)} > 0 \) such that \( \|D^{(2)}(t)\| \leq \lambda^{(2)}. \) Then, if \( \lambda^{(2)} < 1 \) model (1.1) admits a unique stable solution given by (3.1).

4. Invertibility

The notion of invertibility is very useful for statistical applications, such as the prediction of \( X_t \) given its past, or the use of algorithms for computing estimates of the parameters. Several definitions of this notion have been proposed in the literature. Granger and Andersen (1978), Guégan and Pham (1987), Pham and Tran (1981), Subba Rao and Gabr (1984), and Liu (1990) have derived invertibility conditions for some particular stationary bilinear models. Most of these conditions are based on the stationarity and ergodicity assumptions; ergodicity does not apply in the time-varying case, since any functional form of \( X_t \) changes when \( t \) vary in \( \mathbb{Z} \). In what follows, we shall restrict ourselves to processes which satisfy the following definition.
**Definition.** A stable time series \((X_t)_{t \in \mathbb{Z}}\) is said to be a quasi-stationary process, if there exists a function \(r(\cdot)\) such that for each \(h \geq 0\),

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N-h} m(t + h, t) := \lim_{N \to \infty} r_N(h) = r(h),
\]

where \(m(t, s) = E\{X_t X_s\}\).

We now state the following “ergodic-type” theorem for quasi-stationary processes.

**Theorem 4.1.** Let \((X_t)_{t \in \mathbb{Z}}\) be a stable quasi-stationary process with uniformly bounded fourth-order moments, then for each \(h \geq 0\):

(a) \(\hat{r}_N(h)\) converges in mean square to \(r(h)\) \(\Leftrightarrow\) \(\lim_{t \to \infty} C_t(h) = 0\)

(b) if there exists a positive number \(q\) such that \(\lim_{t \to \infty} t^q C_t(h) = 0\), then \(\hat{r}_N(h)\) converges with probability one to \(r(h)\) where \(\hat{r}_N(h) = \frac{1}{N} \sum_{t=1}^{N-h} X_{t+h} X_t\) and where \(C_t(h) = \frac{1}{t} \sum_{s=1}^{t-h} \text{cov}(X_{t+h} X_t, X_{s+h} X_s)\).

**Proof.**

(a) We have

\[
E\{\hat{r}_N(h) - r(h)\}^2 = \text{Var}\{\hat{r}_N(h)\} + \{\hat{r}_N(h) - r(h)\}^2.
\]

Since \((X_t)_{t \in \mathbb{Z}}\) is supposed to be quasi-stationary, it is clear that \(\hat{r}_N(h)\) converges in mean square to \(r(h)\) if and only if \(\lim_{N \to \infty} \text{Var}\{\hat{r}_N(h)\} = 0\). Now, note that

\[
\text{Var}\{\hat{r}_N(h)\} = \frac{2}{N^2} \sum_{t=1}^{N-h} tC_t(h) - \frac{1}{N^2} \sum_{t=1}^{N-h} \text{Var}\{X_{t+h} X_t\} \leq \frac{2}{N} \sum_{t=1}^{N} |C_t(h)|.
\]

Thus if \(C_t(h)\) converges to 0, then the sequence \(\frac{2}{N} \sum_{t=1}^{N} |C_t(h)|\) converges to 0 and \(\lim_{N \to \infty} \text{Var}\{\hat{r}_N(h)\} = 0\). Conversely for all \(h \geq 0\) we have

\[
C_t(h) = \text{Cov}\left(X_{t+h} X_t, \frac{1}{l} \sum_{s=1}^{l-t-h} X_{s+h} X_s\right) + \frac{1}{l} \sum_{s=t-h+1}^{l} \text{Cov}(X_{t+h} X_t, X_{s+h} X_s).
\]

By the Cauchy–Schwarz inequality there exists a nonnegative constant \(K\) such that

\[
|C_t(h)| \leq K \sqrt{\text{Var}\{\hat{r}_N(h)\}} + \frac{K^2 h}{t},
\]

and as \(t \to \infty\). Therefore, if \(\lim_{N \to \infty} \text{Var}\{\hat{r}_N(h)\} = 0\), \(|C_t(h)| \to 0\) as \(t \to \infty\).

(b) Using the method used to prove the Strong Law of Large Numbers in Parzen (1960, p. 420) one can prove that this statement hold.

Assume that \((X_t)_{t \in \mathbb{Z}}\) is a stable process generated by (1.1) which admits the Markovian representation given by (2.1). When the parameters of (1.1) are known completely, following Subba Rao

\[
\hat{r}_N(h) = \frac{1}{N} \sum_{t=1}^{N-h} m(t + h, t) := \lim_{N \to \infty} r_N(h) = r(h),
\]

where \(m(t, s) = E\{X_t X_s\}\).
and Gabr (1984), model (1.1) is said to be invertible if \( \lim_{t \to \infty} E(\xi_t - \hat{\xi}_t)^2 = 0 \) for each sequence \((\hat{\xi}_t)_{t \in \mathbb{Z}}\) of estimate \( \hat{\xi}_t \) with initial values of \( \hat{\xi}_t \) set equal to zero. Using (2.1) we can rewrite \( x_t \) as

\[
x_t = \tilde{I}_{p+q} x_{t-1} + k_t(x) \hat{\xi}_t + r_t(x) = K_t(x)x_{t-1} + \hat{r}_t(x),
\]

where the matrix \( \tilde{I}_{p+q} = (\delta_{j,k-1})_{1 \leq j,k \leq p+q} \), and \( \delta_{k,t} \) is the Kronecker symbol,

\[
k_t(x) = (O_{1 \times (p-1)}, 1, \beta_1(t), \ldots, \beta_q(t))_{1 \times (p+q)},
\]

\[
r_t(x) = \left( O_{1 \times p}, a_1(t+1)X_t, \ldots, a_{q-1}(t+q-1)X_t, \sum_{j=q}^{p} a_j(t+q)X_{t+q-j} \right)_{1 \times (p+q)}
\]

and where \( \hat{r}_t(x) = r_t(x) + k_t(x)X_t, K_t(x) = \tilde{I}_{p+q} - k_t(x)H' \). Iterating expression (4.1), and by substituting \( x_{t-1} \) from (2.1) we obtain for \( t > p \)

\[
\hat{\xi}_t = \hat{\xi}_t - H' \left\{ \prod_{j=1}^{t-p} K_{t-j}(x) \right\} x_{p-1}
\]

where \( \hat{\xi}_t = X_t - \sum_{l=0}^{t-p} H' \left\{ \prod_{j=1}^{l} K_{t-j}(x) \right\} \hat{r}_{t-l}(x) \) hence, for \( q = 1 \), we obtain \( (\hat{\xi}_t - \hat{\xi}_t)^2 = \left( \prod_{j=1}^{t-p} \beta^2_1(t-j) \right) X^2_p \). Using the well-known inequality that the geometric mean is less than the arithmetic mean, we obtain

\[
(\hat{\xi}_t - \hat{\xi}_t)^2 \leq \left\{ \frac{1}{t-p} \sum_{j=1}^{t-p} \beta^2_1(t-j) \right\} \left( \sum_{j=1}^{p} a_j(t-l)X_{t-l-j} + X_{t-l-1} \beta_1(t-l-1) \right)^2.
\]

Assuming that \( (\beta_1(t))_{t \in \mathbb{Z}} \) is a quasi-stationary process satisfying Theorem 4.1(b) and let \( r(0) = \lim_{t \to \infty} \frac{1}{t} \sum_{j=1}^{t} E \{ \beta^2_1(j) \} \), then \( r(0) < 1 \) is a sufficient condition for \( \lim_{t \to \infty} E(\xi_t - \hat{\xi}_t)^2 = 0 \). This result is summarized in the following theorem.

**Theorem 4.2.** Let \((X_t)_{t \in \mathbb{Z}}\) be a stable process satisfying model (1.1) with \( q = 1 \) and admitting the Markovian representation (2.1). If \( (\beta_1(t))_{t \in \mathbb{Z}} \) is quasi-stationary process and satisfies the conditions of Theorem 4.1(b) then model (1.1) is invertible if \( r(0) = \lim_{t \to \infty} \frac{1}{t} \sum_{j=1}^{t} E \{ \beta^2_1(j) \} < 1 \) and we have

\[
\hat{\xi}_t = X_t - \sum_{l=0}^{\infty} (-1)^l \left\{ \prod_{j=1}^{l} \beta_1(t-j) \right\} \times \left( \sum_{j=1}^{p} a_j(t-l)X_{t-l-j} + X_{t-l-1} \beta_1(t-l-1) \right),
\]

where the above series converges in mean square sense.

**Example 4.1.** Assume that the model given by

\[
X_t = \sum_{i=1}^{p} a_i(t)X_{t-i} + c_1(t)X_{t-1} \hat{\xi}_{t-1} + \xi_t
\]
defines a stable process and that the process $\beta_1(t) = c_{11}(t + 1)X_t$ is quasi-stationary and satisfies the conditions of Theorem 4.1(b). Then, the model is invertible if $r(0) = \lim_{t \to \infty} \frac{1}{t} \sum_{j=1}^{t} c_{11}(j + 1)E\{X_t^2\} < 1$. This reduces to $c_{11}E\{X_t^2\} < 1$, the condition for invertibility given in Subba Rao and Gabr (1984), in the stationary and ergodic case.

The following theorem, provides sufficient conditions for invertibility of (1.1).

**Theorem 4.3.** Let $(X_t)_{t \in \mathbb{Z}}$ be a stable process satisfying model (1.1) with $q = 1$ and has uniformly bounded fourth-order moments. Assume that

1. $\lim_{t \to \infty} \frac{1}{t} \sum_{j=1}^{t} b_{11}^2(j) \text{ exists}$,
2. $\lim_{t \to \infty} c_{11}(t) = 0, \forall i \in \{1, \ldots, p\}$,

then model (1.1) is invertible if $\lim_{t \to \infty} \frac{1}{t} \sum_{j=1}^{t} b_{11}^2(j) < 1$.

**Proof.** Straightforward, hence omitted. □

### 4.1. The optimal prediction

Consider the TVBL$(p,q,p,q)$ model defined by (1.1) which admits the bilinear Markovian representation (2.2). Without loss of generality, we shall focus our attention on the subclass of superdiagonal models with $d(t) = 0$ defined by

$$x_t = A(t)x_{t-1} + (C(t)x_{t-1} + c(t))\xi_t,$$

$$X_t = H'x_{t-1} + \bar{\xi}_t.$$  \hspace{1cm} (4.2)

Under the conditions of Theorem 3.1 solution (3.1) becomes

$$X_t = \bar{\xi}_t + \sum_{k=1}^{\infty} H' \left\{ \prod_{j=1}^{k-1} D(t-j) \right\} c(t-k)\bar{\xi}_{t-k}.$$  \hspace{1cm} (4.3)

In this subclass, we are interested in predicting $X_{t+k}$, $k > 0$ from knowledge of the past $\mathcal{F}(X_{t+k-1})$. If $\hat{X}_{t+k}$ is the predicted value of $X_{t+k}$, we shall take $E\{(X_{t+k} - \hat{X}_{t+k})^2\}$ as the criterion to minimize. This criterion is minimized if $\hat{X}_{t+k}$ coincides with the conditional expectation of $X_{t+k}$ given the past, i.e., $E\{X_{t+k} | \mathcal{F}(X_{t+k-1})\}$.

**Theorem 4.4.** Let $(X_t)_{t \in \mathbb{Z}}$ be a superdiagonal stable process satisfying model (1.1), then if the process is causal and invertible we have

1. for any $k > 0$

$$\hat{X}_{t+k} = H' \sum_{i=k}^{\infty} \left\{ \prod_{j=1}^{i-1} D(t+k-j) \right\} c(t+k-i)\bar{\xi}_{t+k-i}.$$  \hspace{1cm} (4.4)
2. The prediction error is

\[ e_t(k) = X_{t+k} - \hat{X}_{t+k} = \zeta_{t+k} + H' \sum_{i=1}^{k-1} \left\{ \prod_{j=1}^{i-1} D(t + k - j) \right\} c(t + k - i) \tilde{\zeta}_{t+k-i}, \]

and its variance is \( \text{Var}(e_t(k)) = H'Q_k(t)H + 1 \) where \( Q_k(t) = \sum_{i=1}^{k-1} P_i(t + k) \) with

\[
P_i(t) = \begin{cases} 
A(t + 1 - i)P_{i-1}(t)A'(t + 1 - i) \\
+ C(t + 1 - i)P_{i-1}(t)C'(t + 1 - i), & i \geq 1, \\
c(t)c'(t), & i = 0.
\end{cases}
\]

Proof.

1. From (3.1) we have

\[
X_{t+k} = H' \sum_{i=1}^{k-1} \left\{ \prod_{j=1}^{i-1} D(t + k - j) \right\} c(t + k - i) \tilde{\zeta}_{t+k-i} + H' \sum_{i=k}^{\infty} \left\{ \prod_{j=1}^{i-1} D(t + k - j) \right\} c(t + k - i) \tilde{\zeta}_{t+k-i}
\]

thus \( \hat{X}_{t+k} = H' \sum_{i=k}^{\infty} \left\{ \prod_{j=1}^{i-1} D(t + k - j) \right\} c(t + k - i) \tilde{\zeta}_{t+k-i} \).

2. The prediction error is however given by

\[ e_t(k) = X_{t+k} - \hat{X}_{t+k} = \zeta_{t+k} + H' \sum_{i=1}^{k-1} \left\{ \prod_{j=1}^{i-1} D(t + k - j) \right\} c(t + k - i) \tilde{\zeta}_{t+k-i}, \]

and since the process \( (g_k(t), k \geq 0)_{t \in \mathbb{Z}} \) is an orthogonal process, then the variance of the process \( (e_t(k))_{t \geq 0} \) is \( \text{var} \{ e_t(k) \} = 1 + H' \sum_{i=1}^{k-1} P_i(t + k)H \) where

\[
P_i(t + k) = E \left\{ \left( \prod_{j=1}^{i-1} D(t + k - j) \right) \Sigma(t + k - i) \left( \prod_{j=1}^{i-1} D(t + k - j) \right)' \right\}
\]

with \( \Sigma(t) = c(t)c'(t) \). Then, we can verify easily that the sequence \( (P_i(t))_{i \geq 0} \) satisfied recursion (4.3). \( \square \)
5. Controllability and observability

In this section we shall focus our attention on the concepts of controllability and observability of the Markovian representation (4.2) which provides a useful criterion for determining whether or not the state vector $x_t$ has the smallest possible dimension.

5.1. Controllability

Representation (4.2) is said to be controllable from the set $D_0 \subset \mathbb{R}^n$ to the set $D_1 \subset \mathbb{R}^n$ in the interval $[0 N] \equiv \{0, \ldots, N\}$ if for any two vectors $u_0, u_1 \in \mathbb{R}^n$, there exists $k \in [0 N]$ and a sequence of noises $\{\xi_1, \ldots, \xi_k\}$ such that $x_k = u_1$ when $x_0 = u_0$. There is no rule for determining the controllability and observability of a general nonlinear representation, but a remarkable property of the Makovian bilinear state space representation (2.3) is that the notion of controllability and observability can be defined in an analogous way as for time-varying ARMA models. Let $\Phi(t,j)$ be the transition matrix of the representation (4.2) defined by $\Phi(t,j) = \prod_{l=0}^{j-1} A(t-l)$ if $t \geq j$ and $I_n$ (the \(n \times n\) identity matrix) for $t < j$. Taking $k$ iterations successively in (4.2), we obtain

$$x_k = \Phi(k,1)x_0 + \sum_{j=1}^{k} \Phi(k,j+1)\{C(j)x_{j-1} + c(j)\}\xi_j.$$  \hspace{1cm} (5.1)

We set $V(i) = C(i)\Phi(i-1,1)x_0 + c(i)$ and rewrite (5.1) as $x_k = \Phi(k,1)x_0 = c(k)\xi_{(k-1)}$ where $c(k)$ is a matrix of $n$ rows and $2^k - 1$ columns defined in the following recursive way:

$$c(k) := \begin{cases} [V(k); A(k)c(k-1); C(k)c(k-1)] & k \geq 2 \\ V(1), & k = 1 \end{cases}$$  \hspace{1cm} (5.2)

and the vector $\xi_{(k-1)}$ is defined by $\xi_{(k-1)} = [\xi_{k}; \xi'_{(k-2)}; \xi'_{(k-2)}]'$ with $\xi_{(0)} = \xi_1$. Let us introduce the so-called controllability matrix defined as follows:

$$W(k) := c(k)c'(k) = N(k) + A(k)W(k-1)A'(k) + C(k)W(k-1)C'(k),$$

where $N(j) := V(j)V'(j)$. From the definition of $W(k)$ we can see that controllability is in fact a property of three matrices $A(t), C(t)$ and $c(t)$. We therefore say the triple $(A(t), C(t), c(t))$ is controllable if and only if representation (4.2) is controllable.

**Theorem 5.1.** The state space representation (4.2) or, equivalently, the triple $(A(t), C(t), c(t))$ is controllable in $\{0, \ldots, k\}$ (for some $k \geq n$) if and only if $\text{rank}\{W(k)\} = n$.

**Proof.** If $\text{rank}\{W(k)\} = n$, then $W(k)$ is nonsingular and thus the state can be made to pass from $u_0$ to $u_1$ by choosing $\xi_{(k-1)} = c'(k)W^{-1}(k)(u_1 - \Phi(k,1)u_0)$. Suppose that the representation (4.2) is controllable in $\{0, \ldots, k\}$ but $\text{rank}\{W(k)\} < n$. Thus there exists a nonzero vector $y \in \mathbb{R}^n$ such that $y'W(k)y = 0$. Then $y'W(k)y = \|c'(k)y\| = 0$, and thus $c'(k)y = 0$. Representation (2.3) is by assumption controllable in $\{0, \ldots, k\}$. Hence we can choose the vectors $u_1, u_0$ so that $y = u_1 - \Phi(k,1)u_0$. Thus we conclude that $y'y = 0$ this implies that the vector $y$ is equal to 0, which contradicts the assumption.  \(\square\)
5.2. Observability

Representation (4.2) is said to be observable if the knowledge of the sequence of observations \((X_j)_{1 \leq j \leq k}\) over the finite-time interval \([1, k]\) where \(k \geq n\) suffices to determine uniquely the initial state vector \(x_0\) when \(\xi_1 = \xi_2 = \cdots = \xi_k = \xi = \text{nonzero constant}\). In such a case, let \(\mathcal{C}(l) = H'\), and define

\[
\mathcal{C}_k(l) = \begin{pmatrix} H' \\ \mathcal{C}_{k-1}(l+1)A(l) \\ \mathcal{C}_{k-1}(l+1)C(l) \end{pmatrix}_{(2^k-1) \times n}
\]

and let \(\mathcal{O}(k) = \mathcal{C}_k(1)\). With this notation we can write \(\tilde{X}_k = T\mathcal{C}(k)x_0 + y_k\) where \(\tilde{X}_k = (X_1, \ldots, X_k)'\), \(y_k\) is a known constant vector which is independent of \(x_0\) and \(T\) is an appropriate \(k \times (2^k - 1)\) matrix of rank \(k\). Let us define the observability matrix by the following recursion

\[
\mathcal{O}(k) = HH' + A'(1)\mathcal{O}(k-1)A(1) + C'(1)\mathcal{O}(k-1)C(1),
\]

where \(\mathcal{O}(k) = \mathcal{C}'(k)\mathcal{C}(k)\) which depends on three matrices \(H, A(t)\) and \(C(t)\). We shall say that the triple \((A(t), C(t), H)\) is observable if and only if representation (4.2) is observable. Using an approach similar to the proof of Theorem 5.1, we can prove the following theorem.

**Theorem 5.2.** The state space representation (4.2) is observable in \(\{1, \ldots, k\}\) if and only if \(\text{rank}\{\mathcal{O}(k)\} = n\).

Thanks to the properties of matrices \(\mathcal{C}(k)\) and \(\mathcal{O}(k)\), we can also state the following theorem.

**Theorem 5.3.** Representation (2.3) is minimal (i.e., of least dimension) if and only if the \(n\) rows of \(\mathcal{C}(k)\) and, respectively, the \(n\) columns of \(\mathcal{O}(k)\) are linearly independent on \(\mathbb{R}^n\).

A consequence of Theorem 5.3 is that the lowest possible dimension of the state vector in representation (4.2) is \(n\) ensuring that the state vector will not contain redundant components. The theorem provides a connection between the notion of “minimal or irreducible realizations” and the concept of controllability and observability. Suppose we think of \(\xi_t\) as a physical input, rather than an unobservable strict white noise process, with \(X_t\) as the corresponding output and we can find a causal relationship between \(X_t\) and \(\xi_t\) in the sense that condition (3.3) holds. Then, by controllability we mean that starting from some initial state \(x_0\) at time \(t_0\), we can find an input which steers the system into any desired state \(x_1\) at some subsequent time \(t_1\) if the state vector has minimal dimension. The minimality condition in Theorem 5.3 also ensures that the state vector \(x_t\) can be determined explicitly from a sequence of observations on \(\{X_t\}\) and \(\{\xi_t\}\).

6. Concluding remarks

The theoretical results we have discussed provide the background for further studies into time-varying bilinear models. For instance, this work does not cover the important problem of parameter
estimation but provides sufficient conditions for expressing \( X_t \) in terms of its past history, which is useful for prediction and estimation purposes. The ideas of controllability and observability provides some incentive for a discussion on Kalman filter estimates of the state vectors.

**Acknowledgements**

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada. The authors are grateful for suggestions from the referee.

**References**


