

CONSISTENT AND ASYMPTOTICALLY NORMAL ESTIMATORS FOR PERIODIC BILINEAR MODELS

ABDELOUAHAB BIBI AND ANTONY GAUTIER

ABSTRACT. In this paper, a distribution free approach to the parameter estimation of a simple bilinear model with periodic coefficients is presented. The proposed method relies on minimum distance estimator based on the autocovariances of the squared process. Consistency and asymptotic normality of the estimator, as well as hypotheses testing, are derived. Numerical experiments on simulated data sets are presented to highlight the theoretical results.

1. Introduction

Periodically varying components are characteristic of many practical time series situations, including economic, hydrological and meteorological ones. Data of this type are frequently analyzed using a periodic autoregressive moving average (PARMA) model, defined by some linear dynamics which alternate in a regular cycle because of periodic coefficients (see Gautier [7] for references dealing with PARMA models). However, many real time series exhibit not only nonstationary structure, but also nonlinear features (see Bibi [1]). Therefore, bilinear time series models with time-varying parameters have been attracting a great deal of interest in the recent statistical literature. Bibi [1] and Bibi and Oyet [4] are references dealing with probabilistic properties and/or statistical inference for some subclasses of bilinear models with time-dependent coefficients. Our attention here is focused on the class of bilinear models with periodically varying coefficients. Bibi and Gautier [3] stated general results aimed to provide L^p -properties ($p \geq 1$) and the asymptotic behavior of moments estimates for strictly superdiagonal and completely bilinear process with periodic coefficients of period $s \in \mathbb{N}^* = \{1, 2, \dots\}$, denoted by $\text{PBL}(0, 0, P, 1)_s$, $P \geq 2$, and defined for all $v \in \{1, \dots, s\}$ by

$$(1) \quad X_t(v) = \sum_{i=2}^P b_i(v) X_t(v-i) \epsilon_t(v-1) + \epsilon_t(v),$$

Received April 15, 2008; Revised March 17, 2010.

2000 *Mathematics Subject Classification.* 62M10, 91B84.

Key words and phrases. bilinear time series, periodic coefficients, minimum distance estimator, asymptotic normality, hypotheses testing.

where $(\epsilon_t)_{t \in \mathbb{Z}}$ is a sequence of uncorrelated random variables with $E(\epsilon_t) = E(\epsilon_t^3) = 0$, $E(\epsilon_t^2) = 1$ and $E(\epsilon_t^4) < +\infty$. In the difference equation (1), $X_t(v)$ refers to as the variable X during the v -th ‘season’ or ‘regime’ of the t -th year with period s ; this means that $X_t(v)$ can also be written as X_{st+v} (respectively $\epsilon_t(v)$ as ϵ_{st+v}). The time-constant case (studied in Francq [6]) is obtained with $s = 1$. It is also worth noting that the process $(X_t)_{t \in \mathbb{Z}}$ has similar properties to those of an autoregressive conditionally heteroskedastic (ARCH) process (see Bibi and Gautier [3]), which plays an important role in financial econometrics, and thus can be used as a weak white noise in modelling series having certain specifications. Some applications of the model (1) we have in mind are, for instance, intra-year observations on most economic time series with different dynamics, quarterly or monthly electrical consumption and production functions being examples.

Based on Bibi and Gautier [3] and further extensions, the main purpose of this paper is to consider the problem of estimating the model (1) with $P = 2$ and $s = 2$, i.e.,

$$(2) \quad X_t(v) = b_v X_t(v-2) \epsilon_t(v-1) + \epsilon_t(v),$$

where $v \in \{1, 2\}$, and $(\epsilon_t)_{t \in \mathbb{Z}}$ is an iid $\mathcal{N}(0, 1)$ process. The Gaussianity assumption is only needed in order to facilitate technical computations involved in the next sections, but the asymptotic results presented below still hold without such an assumption. As mentioned before, one can notice that, for each $v \in \{1, 2\}$, we have

$$E \{ X_t^2(v) | X_t(v-2) \} = 1 + b_v^2 X_t^2(v-2),$$

which raises that the model (2) under consideration in this paper is conditionally heteroskedastic.

In this paper, we assume that all the random variables are defined on the same probability space $(\Omega, \mathfrak{F}, \mathbb{P})$ and that ϵ_t and X_u are independent for all $u < t$. The $k \times k$ identity matrix is denoted by $I_{(k)}$, the symbols ‘ \rightsquigarrow ’, ‘*a.s.*’ and ‘ $p \lim_{n \rightarrow \infty}$ ’ respectively mean convergence in law, almost surely and in probability. The notation $M^{\odot m} = M \odot \dots \odot M$ stands for the m -th Hadamard product for any matrix M , M' denotes the transpose of M , and $\|\cdot\|$ denotes any vectorial norm. The remainder of the paper is organized as follows. In Section 2, we give conditions ensuring existence of moments up to the p -th order ($p \geq 2$) for the process defined by (2) and the explicit form of the covariance function of the squared process. Further details for the proofs given in this section can be found in Bibi and Aknouche [2]. Section 3 presents results on asymptotic inference. Hypotheses testing is considered in Section 4. The finite sample properties are empirically evaluated in Section 5 by means of Monte Carlo simulations. Section 6 concludes the paper.

2. Existence of moments

Before dealing with the existence of moments for the model (2) and deriving the covariance structures, it is convenient to represent (2) in a vectorial form as follows. Let $\underline{X}_t = \{X_t(1), X_t(2)\}'$, $\underline{\epsilon}_t = \{\epsilon_t(1), \epsilon_t(2)\}'$ and the 2×2 random matrix $\Phi_t = \Phi(\underline{\epsilon}_t, \underline{\epsilon}_{t-1})$ defined by its (i, j) -th components

$$(\Phi_t)_{i,j} = b_i \epsilon_t(i-1) \mathbb{1}_{\{i=j\}},$$

where $\mathbb{1}_\Delta$ denotes the indicator function of set Δ . With these notations, we obtain the following generalized autoregressive representation for (2):

$$(3) \quad \underline{X}_t = \Phi_t \underline{X}_{t-1} + \underline{\epsilon}_t.$$

Such representation is potentially useful for deriving probabilistic properties for $(X_t)_{t \in \mathbb{Z}}$.

2.1. Stationarity of the vector process

Define the operator norm for any matrix A by

$$\|A\|_{\text{op}} = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

and the top-Lyapunov exponent by

$$\gamma_L = \inf_{t > 0} E \left(\frac{1}{t} \log \left\| \prod_{i=0}^{t-1} \Phi_{t-i} \right\|_{\text{op}} \right)$$

when $E(\log^+ \|\Phi_0\|_{\text{op}}) < +\infty$, where $\log^+ x = \max\{\log x, 0\}$ for $x > 0$. It follows from the subadditive ergodic theorem that

$$(4) \quad \gamma_L = \lim_{t \rightarrow \infty} \left\{ \frac{1}{t} \log \left\| \prod_{i=0}^{t-1} \Phi_{t-i} \right\|_{\text{op}} \right\} \quad \text{a.s.}$$

We call causal any solution of (3) such that \underline{X}_t is measurable with respect to the σ -field generated by $\{\underline{\epsilon}_u, u \leq t\}$. We have the following results.

Theorem 1. *Consider the process (2) and let $(\underline{X}_t)_{t \in \mathbb{Z}}$ be the associated vectorial process. Then, $\gamma_L < 0$ is a sufficient condition for (3) to have a unique, causal, strictly stationary and ergodic solution given by*

$$(5) \quad \underline{X}_t = \sum_{k \geq 0} \Pi_k(t) \underline{\epsilon}_{t-k},$$

where the above series converges a.s. and $\{\Pi_k(t)\}_{k \geq 1}$ are random matrices given by

$$\Pi_k(t) = \prod_{j=0}^{k-1} \Phi_{t-j}$$

with, as usual, products over empty sets equal to $I_{(2)}$.

Remark 1. By the strong law of large numbers and Jensen’s inequality, it is straightforward to show that $\gamma_L \leq \log |b_1 b_2|$. Hence, if $|b_1 b_2| < 1$, then for all $t \in \mathbb{Z}$, the series (5) converges a.s. and constitutes the unique strictly stationary and ergodic solution of the equation (3). However, the existence of explosive regimes (i.e., regimes with $|b_v| \geq 1$) does not preclude the strict stationarity.

Remark 2. The top-Lyapunov exponent criterion seems difficult to be obtained explicitly. However a potential method to verify whether $\gamma_L < 0$ or not is via Monte Carlo simulations using the equation (4). This fact heavily limits the interests on the criterion in statistical applications. Indeed, the solution needs to have some moments to make an estimation theory possible and the top-Lyapunov criterion does not guarantee the existence of such moments.

Remark 2 leads to search for conditions on $(b_v)_{v=1,2}$ ensuring the existence of moments for the stationary solution for which the top-Lyapunov exponent will be automatically negative. In that case we say that the equation (3) has a second-order stationary solution and the process $(X_t)_{t \in \mathbb{Z}}$ defined by (2) is said to be periodically stationary.

Theorem 2. *A sufficient condition for existence of a second-order stationary solution of the equation (3) is that $\det \{I_{(2)} - z\Phi\} \neq 0$ for all complex z such that $|z| \leq 1$, where $\Phi = E(\Phi_t^{\odot 2})$. The solution process has zero mean and variance-covariance matrix $\Sigma(h) = \text{diag} \{\underline{r}(h)\}$, where $\underline{r}(h) = E(\underline{X}_t \odot \underline{X}_{t-h})$ satisfies $\underline{r}(h) = \{I_{(2)} - \Phi\}^{-1} \delta_0(h)$ with $\delta_0(h) = (1, 1)'$ if $h = 0$, $(0, 0)'$ otherwise. Moreover, the solution process is unique, strictly stationary, causal and ergodic, given by (5). Conversely, a necessary condition for existence of a second-order stationary solution is that there exists a variance-covariance matrix $\Sigma(h) = \text{diag} \{\underline{r}(h)\}$ and $\underline{r}(h)$ is the solution of the equation $\{I_{(2)} - \Phi\} \underline{r}(h) = \delta_0(h)$.*

Proof. The proof is straightforward, since it is simply based on the computation of spectral radius of matrix Φ . □

Remark 3. As a consequence of the above theorem, the second-order stationary solution $(\underline{X}_t)_{t \in \mathbb{Z}}$ of (3) is a weak white noise ($\Sigma(h) = 0$ for $h \neq 0$), and so the process $(\underline{X}_t)_{t \in \mathbb{Z}}$ is serially uncorrelated, but not independent.

Remark 4. When causal, it is possible to relate $(X_t)_{t \in \mathbb{Z}}$ and $(\epsilon_t)_{t \in \mathbb{Z}}$ through the infinite order moving-average representation

$$(6) \quad X_t(v) = \sum_{k \geq 0} \psi_k(v) \xi_k^{(v)}(t),$$

where $\psi_k(v) = b_v^k$ and

$$\xi_k^{(v)}(t) = \left\{ \prod_{j=0}^{k-1} \epsilon_t(v - 2j - 1) \right\} \epsilon_t(v - 2k).$$

In (6), the ‘seasonal weights’ $\psi_k(v)$ satisfy

$$\sup_{v=1,2} \sum_{k \geq 0} |b_v|^k = \sup_{v=1,2} \{(1 - |b_v|)^{-1}\} < +\infty.$$

2.2. Existence of moments up to $2m$ -order for the vector process

We now investigate the existence of moments up to $2m$ -order for the vector process $(\underline{X}_t)_{t \in \mathbb{Z}}$. More precisely, we derive necessary and sufficient conditions for the finiteness of $E(\underline{X}_t^{\odot 2m})$ for any integer $m \geq 1$. In view of (3), for all $m \geq 1$, we have

$$\underline{X}_t^{\odot m} = \Phi_t^{\odot m} \underline{X}_{t-1}^{\odot m} + \epsilon_t^{\odot m} + \sum_{\ell=1}^{m-1} \frac{m!}{\ell!(m-\ell)!} (\Phi_t \underline{X}_{t-1})^{\odot \ell} \odot \epsilon_t^{\odot m-\ell}.$$

Since the vector ϵ_t is normally distributed, we have $E(\underline{X}_t^{\odot 2m+1}) = 0$ for all integers m . This property implies in particular that the coefficient of skewness $\tau(v) = 0$ for all $v \in \{1, 2\}$. The following theorem gives necessary and sufficient conditions for the existence of moments up to the $2m$ -th order for $(\underline{X}_t)_{t \in \mathbb{Z}}$.

Theorem 3. *Consider the model (3). Then, the process $(\underline{X}_t)_{t \in \mathbb{Z}}$ admits moments up to $2m$ -th order if and only if*

$$(2k - 1)!! \max_{v=1,2} \{b_v^{2k}\} < 1, \quad k = 1, \dots, m,$$

with $(2k - 1)!! = 1 \times 3 \times \dots \times (2k - 1)$.

Proof. The proof is straightforward, so we omit it. □

Remark 5. It is worth noting that, contrary to strict stationary condition given in Remark 2, the second-order stationary condition imposes that all regimes must be stationary.

2.3. Covariance structures

In this subsection, we concentrate on covariance structures for $\{X_t^2(v)\}_{t \in \mathbb{Z}}$ defined by (2). Such results will be crucially needed for applications to the estimation procedure we will consider in Section 3. Denote

$$\gamma^{(v)}(h) = E [\{X_t^2(v) - \mu(v)\} \{X_t^2(v-h) - \mu(v-h)\}], \quad h \geq 0,$$

the covariances of $X_t^2(v)$ for all $v \in \{1, 2\}$, where $\mu(v) = E\{X_t^2(v)\}$ and $\gamma^{(v)}(h) = \gamma^{(v-h)}(-h)$ when $h < 0$. The following results characterize the covariance function of the squared process $\{X_t^2(v)\}_{t \in \mathbb{Z}}$.

Theorem 4. *Assume that $(\underline{X}_t)_{t \in \mathbb{Z}}$ admits moments up to 4-th order. Then, for all $v \in \{1, 2\}$, we have:*

- (1) $\mu(v) = (1 - b_v^2)^{-1}$,
- (2) $\gamma^{(v)}(0) = 2(1 - 3b_v^4)^{-1} \mu^2(v)$,
- (3) $\gamma^{(v)}(1) = 2b_v^2(1 - b_v^2)^{-1} \mu(v + 1)$,

- (4) for all $h \geq 1$, $\gamma^{(v)}(2h) = b_v^{2h} \gamma^{(v)}(0)$,
- (5) for all $h \geq 2$, $\gamma^{(v)}(2h - 1) = b_v^{2(h-1)} \gamma^{(v)}(1)$.

Proof. The above formulas can be easily obtained using (6). □

Remark 6. A few comments can be made. One can see that the covariances of $\{X_t^2(v)\}_{t \in \mathbb{Z}}$ depend not only on the coefficients b_v , but also on the expectation of the squared process $\{X_t^2(v)\}_{t \in \mathbb{Z}}$. It is also worth mentioning that, under the assumptions of Theorem 4, the kurtosis $\kappa(v)$ of the process $\{X_t(v)\}_{t \in \mathbb{Z}}$ is given by

$$\kappa(v) = \frac{E\{X_t^4(v)\}}{[E\{X_t^2(v)\}]^2} = \frac{\gamma^{(v)}(0) + [E\{X_t^2(v)\}]^2}{[E\{X_t^2(v)\}]^2} = 1 + 2 \frac{1 - b_v^2}{1 - 3b_v^4}.$$

Hence the process is characterized by leptokurtosis in excess with respect to the normal distribution if

$$\frac{1}{3} < b_v^2 \leq \frac{1}{\sqrt{3}}$$

and, contrarily to the pure ARCH(1) model, we obtain $\kappa(v) \leq 3$ when

$$0 \leq b_v^2 \leq \frac{1}{3}.$$

This property can be used to distinguish between pure ARCH and periodic bilinear models.

3. Proposed minimum distance estimator and asymptotic properties

Let (X_1, \dots, X_{2n}) be a sample from a periodically stationary bilinear process generated by (2). This is equivalent to having a sample $(\underline{X}_1, \dots, \underline{X}_n)$ from a stationary generalized autoregressive process $(\underline{X}_t)_{t \in \mathbb{Z}}$ given by (3). In this section, our purpose is to use the minimum distance estimator (MDE) to estimate the periodic bilinear parameters, gathered in $\underline{b} = (b_1, b_2)'$. We suppose that \underline{b} belongs to a parameter space $\Theta \subset]-1, 1[\times]-1, 1[$. The true parameter value is unknown and is denoted by \underline{b}_0 . For practical estimation purpose, the moments of $(X_t^2)_{t \in \mathbb{Z}}$ can be approximated by their empirical sample moments, and so $\gamma^{(v)}(\cdot)$ in Theorem 4, can be replaced by quantities like

$$\tilde{\gamma}_n^{(v)}(2k) = b_v^{2k} \tilde{\gamma}_n^{(v)}(0) \quad \text{and} \quad \tilde{\gamma}_n^{(v)}(2k - 1) = b_v^{2(k-1)} \tilde{\gamma}_n^{(v)}(1),$$

where

$$\tilde{\gamma}_n^{(v)}(0) = \frac{2}{1 - 3b_v^4} \hat{\mu}_n(v) \quad \text{and} \quad \tilde{\gamma}_n^{(v)}(1) = \frac{2b_v^2}{1 - b_v^2} \hat{\mu}_n(v + 1)$$

with $\hat{\mu}_n(v) = n^{-1} \sum_{t=0}^{n-1} X_t^2(v)$. Here, the convention $\underline{X}_t = 0$ for $t > n$ is made.

By setting $\hat{\underline{\mu}}_n = \{\hat{\mu}_n(1), \hat{\mu}_n(2)\}'$, $\underline{\mu} = \{\mu(1), \mu(2)\}'$, $\tilde{\underline{\gamma}}_n(h) = \{\tilde{\gamma}_n^{(1)}(h), \tilde{\gamma}_n^{(2)}(h)\}'$

and $\underline{\gamma}(h) = \{\gamma^{(1)}(h), \gamma^{(2)}(h)\}'$, we have

$$\underline{\tilde{\gamma}}_n(h) = B(h)\underline{\hat{\mu}}_n \quad \text{and} \quad \underline{\gamma}(h) = B(h)\underline{\mu},$$

where

$$B(2h) = \begin{pmatrix} \frac{2b_1^{2h}}{1-3b_1^4} & 0 \\ 0 & \frac{2b_2^{2h}}{1-3b_2^4} \end{pmatrix}, \quad B(2h-1) = \begin{pmatrix} 0 & \frac{2b_1^{2h}}{1-b_1^2} \\ \frac{2b_2^{2h}}{1-b_2^2} & 0 \end{pmatrix}.$$

The asymptotic properties of the sequences $\{\hat{\mu}_n(\cdot)\}_{n \geq 1}$ and $\{\tilde{\gamma}_n^{(v)}(\cdot)\}_{n \geq 1}$ are given in the following theorem.

Theorem 5. *Assume that the process $(\underline{X}_t)_{t \in \mathbb{Z}}$ admits moments up to 4-th order. Then, for each $v, v' \in \{1, 2\}$,*

- (1) $\hat{\mu}_n(v)$ (respectively $\tilde{\gamma}_n^{(v)}(k)$) converges a.s. to $\mu(v)$ (respectively to $\gamma^{(v)}(k)$) as $n \rightarrow \infty$,
- (2) we have

$$\lim_{n \rightarrow \infty} n \text{Cov} \{ \hat{\mu}_n(v), \hat{\mu}_n(v') \} = (V)_{v,v'}$$

and

$$\lim_{n \rightarrow \infty} n \text{Cov} \{ \tilde{\gamma}_n^{(v)}(\ell), \tilde{\gamma}_n^{(v')}(k) \} = \{W(\ell, k)\}_{v,v'},$$

where

$$V = 2 \begin{pmatrix} \frac{1+b_1^2}{(1-3b_1^4)(1-b_1^2)^3} & \frac{(1-b_2^2)b_1^2 + (1-b_1^2)b_2^2}{(1-b_2^2)^2(1-b_1^2)^2} \\ \frac{(1-b_2^2)b_1^2 + (1-b_1^2)b_2^2}{(1-b_2^2)^2(1-b_1^2)^2} & \frac{1+b_2^2}{(1-3b_2^4)(1-b_2^2)^3} \end{pmatrix}$$

and $W(\ell, k) = B(\ell)VB'(k)$,

- (3) we have

$$\sqrt{n}(\underline{\hat{\mu}}_n - \underline{\mu}) \rightsquigarrow \mathcal{N}(0, V)$$

and

$$\sqrt{n} \{ \underline{\tilde{\gamma}}_n(h) - \underline{\gamma}(h) \} \rightsquigarrow \mathcal{N} \{ 0, W(h, h) \}.$$

Proof. Assertions 1–2 of Theorem 5 can be simply followed by exploiting the relationship between the periodic bilinear model (2) and its vectorial representation (3) and by expressing $\underline{Y}_t = \underline{X}_t^{\odot 2} - \underline{\mu}$ as a strictly stationary and ergodic 2-dimensional process. To show Assertion 3, define for any $m \geq 1$,

$$\underline{U}_t(v) = \sum_{k=0}^m \psi_k(v) \xi_k^{(v)}(t) \quad \text{and} \quad \underline{W}_t(v) = \sum_{k=m+1}^{\infty} \psi_k(v) \xi_k^{(v)}(t),$$

where $\xi_k^{(v)}(t)$ is given by (6). Consider the vectors $\underline{U}_t = \{U_t(1), U_t(2)\}'$ and $\underline{W}_t = \{W_t(1), W_t(2)\}'$. Then, $(\underline{X}_t)_{t \in \mathbb{Z}}$ can be expressed as

$$(7) \quad \underline{X}_t = \underline{U}_t + \underline{W}_t,$$

where $(U_t)_{t \in \mathbb{Z}}$ is an $(m + 1)$ -dependent stationary process and $(W_t)_{t \in \mathbb{Z}}$ converges in probability to zero as $m \rightarrow \infty$. Hence, the asymptotic distribution of $\sqrt{n}(\hat{\mu}_n - \mu)$ is the same as the one of $\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \{U_t^{\odot 2} - E(U_t^{\odot 2})\}$. Since $E\{U_t^{\odot 2} (U_t^{\odot 2})'\} < +\infty$, we have for any fixed m ,

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \{U_t^{\odot 2} - E(U_t^{\odot 2})\} \rightsquigarrow \mathcal{N} \left\{ 0, \sum_{k=-m}^m \text{Cov}(U_t^{\odot 2}, U_{t+k}^{\odot 2}) \right\}$$

(see Brockwell and Davis [5] Theorem 6.4.2). As $m \rightarrow \infty$, U_t converges to X_t in probability, and so the covariance matrix of the asymptotic distribution converges to $\sum_{k \in \mathbb{Z}} \text{Cov}(X_t^{\odot 2}, X_{t+k}^{\odot 2})$ which can be easily expressed by the matrix V . The rest of the proof immediately follows. \square

Let us now consider the MDE. For any $h \geq 1$, let

$$\begin{aligned} \Gamma(\underline{b}) &= \Gamma = \{\gamma'(0), \dots, \gamma'(h)\}', \\ \tilde{\Gamma}_n(\underline{b}) &= \tilde{\Gamma}_n = \{\tilde{\gamma}'_n(0), \dots, \tilde{\gamma}'_n(h)\}' \\ \text{and } \hat{\Gamma}_n &= \{\hat{\gamma}'_n(0), \dots, \hat{\gamma}'_n(h)\}' \end{aligned}$$

where $\hat{\gamma}'_n(i) = \{\hat{\gamma}_n^{(1)}(i), \hat{\gamma}_n^{(2)}(i)\}'$ with, for each $v = 1, 2$, the sample covariance of $X_t^2(v)$ evaluated at lag i

$$\hat{\gamma}_n^{(v)}(i) = \frac{1}{n} \sum_{t=0}^{n-1} [\{X_t^2(v) - \hat{\mu}_n(v)\} \{X_{t-i}^2(v) - \hat{\mu}_n(v-i)\}], \quad i = 0, \dots, h.$$

Note that, by Theorem 5, $\tilde{\Gamma}_n$ converges a.s. to Γ . Now, the MDE of the parameter \underline{b} is defined as any measurable solution \tilde{b}_n of

$$\arg \min_{\underline{b} \in \Theta} \tilde{F}'_n(\underline{b}) M_n \tilde{F}_n(\underline{b}) = \arg \min_{\underline{b} \in \Theta} \tilde{Q}_n(\underline{b}),$$

where $\tilde{Q}_n(\underline{b}) = \tilde{F}'_n(\underline{b}) M_n \tilde{F}_n(\underline{b})$, M_n is a $2(h + 1) \times 2(h + 1)$ weighting matrix, and $\tilde{F}_n(\underline{b}) = \hat{\Gamma}_n - \tilde{\Gamma}_n(\underline{b})$ is the score function. The dependence of $\tilde{F}_n(\cdot)$, $\tilde{Q}_n(\cdot)$ and M_n on X_t and h is generally deleted for notational convenience. To analyze the large sample properties of the proposed estimator, it is necessary to impose the following regularity conditions on the process $(X_t)_{t \in \mathbb{Z}}$, on the matrix M_n and on the parameter space Θ .

- A1:** The process $(X_t)_{t \in \mathbb{Z}}$ is strictly stationary and ergodic.
- A2:** (M_n) is a sequence of positive definite matrices and there exists a positive definite matrix of constants M such that $p \lim_{n \rightarrow \infty} M_n = M$.
- A3:** The matrix $\frac{\partial \Gamma'(\underline{b}_0)}{\partial \underline{b}} M \frac{\partial \Gamma(\underline{b}_0)}{\partial \underline{b}}$ is a finite nonsingular matrix of constants.
- A4:** The parameter Θ is compact and \underline{b}_0 is in the interior of Θ .

Under these assumptions, we can state the following result.

Theorem 6. Under A1 – A4 and if $(\underline{X}_t)_{t \in \mathbb{Z}}$ admits moments up to 4-th order, then $\tilde{\underline{b}}_n$ converges in probability to \underline{b}_0 .

Proof. From the first-order conditions (organized as column vector) for the minimization of $\tilde{Q}_n(\underline{b})$, we have

$$(8) \quad \frac{\partial \tilde{\Gamma}'_n(\tilde{\underline{b}}_n)}{\partial \underline{b}} M_n \tilde{F}_n(\tilde{\underline{b}}_n) = 0.$$

Taking the first-order Taylor-series expansion of the score vector $\tilde{F}_n(\tilde{\underline{b}})$ around \underline{b}_0 , we have

$$\tilde{F}_n(\tilde{\underline{b}}_n) = \tilde{F}_n(\underline{b}_0) - \frac{\partial \tilde{\Gamma}_n(\underline{b}_*)}{\partial \underline{b}} (\tilde{\underline{b}}_n - \underline{b}_0),$$

where \underline{b}_* is an intermediate point on the line segment joining $\tilde{\underline{b}}_n$ and \underline{b}_0 . Substituting for $\tilde{F}_n(\tilde{\underline{b}}_n)$ into (8) yields

$$\frac{\partial \tilde{\Gamma}'_n(\tilde{\underline{b}}_n)}{\partial \underline{b}} M_n \left\{ \tilde{F}_n(\underline{b}_0) - \frac{\partial \tilde{\Gamma}_n(\underline{b}_*)}{\partial \underline{b}} (\tilde{\underline{b}}_n - \underline{b}_0) \right\} = 0.$$

Rearranging the above expression gives a.s.

$$\tilde{\underline{b}}_n - \underline{b}_0 = \left\{ \frac{\partial \tilde{\Gamma}'_n(\tilde{\underline{b}}_n)}{\partial \underline{b}} M_n \frac{\partial \tilde{\Gamma}_n(\underline{b}_*)}{\partial \underline{b}} \right\}^{-1} \frac{\partial \tilde{\Gamma}'_n(\tilde{\underline{b}}_n)}{\partial \underline{b}} M_n \tilde{F}_n(\underline{b}_0).$$

Since the process $(\underline{X}_t)_{t \in \mathbb{Z}}$ is second-order stationary and ergodic, and since

$$\left\| \frac{\partial}{\partial \underline{b}} \tilde{\gamma}_n(i) \right\|_{\text{op}} = \left\| \frac{\partial}{\partial \underline{b}} \{B(i) \hat{\underline{\mu}}_n\} \right\|_{\text{op}} \leq K \|\hat{\underline{\mu}}_n\|$$

for each $i = 0, \dots, h$ and for some positive constant K , we can see that, under the conditions of Theorem 5,

$$\begin{aligned} p \lim_{n \rightarrow \infty} \frac{\partial \tilde{\Gamma}'_n(\tilde{\underline{b}}_n)}{\partial \underline{b}} M_n &= B = \frac{\partial \Gamma(\underline{b}_0)}{\partial \underline{b}} M, \\ p \lim_{n \rightarrow \infty} \frac{\partial \tilde{\Gamma}'_n(\tilde{\underline{b}}_n)}{\partial \underline{b}} M_n \frac{\partial \tilde{\Gamma}_n(\underline{b}_*)}{\partial \underline{b}} &= A = \frac{\partial \Gamma'(\underline{b}_0)}{\partial \underline{b}} M \frac{\partial \Gamma(\underline{b}_0)}{\partial \underline{b}}. \end{aligned}$$

Hence from the dominated convergence theorems,

$$p \lim_{n \rightarrow \infty} \left\{ \frac{\partial \tilde{\Gamma}'_n(\tilde{\underline{b}}_n)}{\partial \underline{b}} M_n \frac{\partial \tilde{\Gamma}_n(\underline{b}_*)}{\partial \underline{b}} \right\}^{-1} \frac{\partial \tilde{\Gamma}'_n(\tilde{\underline{b}}_n)}{\partial \underline{b}} M_n = A^{-1} B'$$

is finite, and since $p \lim_{n \rightarrow \infty} \tilde{F}_n(\underline{b}_0) = 0$, the consistency of $\tilde{\underline{b}}_n$ follows. □

Now, we consider the estimator

$$\hat{\underline{b}}_n = \arg \min_{\underline{b} \in \Theta} \hat{F}'_n(\underline{b}) M_n \hat{F}_n(\underline{b}) = \arg \min_{\underline{b} \in \Theta} \hat{Q}_n(\underline{b}),$$

where $\widehat{Q}_n(\underline{b}) = \widehat{F}'_n(\underline{b})M_n\widehat{F}_n(\underline{b})$, and the score function $\widehat{F}_n(\underline{b})$ is defined by $\widehat{F}_n(\underline{b}) = \widehat{\Gamma}_n - \underline{\Gamma}(\underline{b})$. The asymptotic properties of $\{\widehat{F}_n(\underline{b})\}_{n \geq 1}$ are given in the following theorem.

Lemma 1. *Let $(X_t)_{t \in \mathbb{Z}}$ be a process satisfying the model (2).*

- (1) *If $(\underline{X}_t)_{t \in \mathbb{Z}}$ admits moments up to 4-th order, then $\widehat{\Gamma}_n$ converges a.s. to $\underline{\Gamma}$.*
- (2) *If $(\underline{X}_t)_{t \in \mathbb{Z}}$ admits moments up to 8-th order, then*

$$\sqrt{n} \left(\widehat{\Gamma}_n - \underline{\Gamma} \right) \rightsquigarrow \mathcal{N} (0, \Sigma),$$

where Σ is a $2(h+1) \times 2(h+1)$ covariance matrix whose (i, j) -th block matrix $\Sigma_{i,j}$ is given by

$$\Sigma_{i,j} = \sum_{k \in \mathbb{Z}} \text{Cov} \left\{ \underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i), \underline{X}_{t+k}^{\odot 2} \odot \underline{X}_{t+k-j}^{\odot 2}(j) \right\}, \quad i, j = 0, \dots, h,$$

with $\underline{X}_t(i) = \{X_t(i+1), X_t(i+2)\}'$.

Proof. For any integer $i \geq 0$, let $\underline{X}_t(i) = \{X_t(i+1), X_t(i+2)\}'$, $\widehat{\underline{\mu}}_n(i) = \{\widehat{\mu}_n(i+1), \widehat{\mu}_n(i+2)\}'$ and $\underline{\mu}(i) = \{\mu(i+1), \mu(i+2)\}'$. Then the vector $\widehat{\underline{\gamma}}_n(i)$ can be written as

$$(9) \quad \widehat{\underline{\gamma}}_n(i) = \frac{1}{n} \sum_{t=0}^{n-1} \underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i) - \widehat{\underline{\mu}}_n \odot \widehat{\underline{\mu}}_n(i).$$

By ergodicity of \underline{X}_t , the first and second terms in (8) converge a.s. respectively to $E \{ \underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i) \}$ and to $\underline{\mu} \odot \underline{\mu}(i)$. Hence the first assertion is an immediate consequence of the strong law of large numbers. Since $\underline{\mu} \odot \underline{\mu}(i) = -\underline{\gamma}(i) + E \{ \underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i) \}$, then the asymptotic distribution of $\sqrt{n} \{ \widehat{\underline{\gamma}}_n(i) - \underline{\gamma}(i) \}$ is similar to the asymptotic distribution of

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} [\underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i) - E \{ \underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i) \}].$$

Let $(\underline{U}_t)_{t \in \mathbb{Z}}$ and $(\underline{W}_t)_{t \in \mathbb{Z}}$ be the processes defined as in (7). Since \underline{U}_t converges in probability to a stationary process and \underline{W}_t converges in probability to zero as $m \rightarrow \infty$, then we can show that the asymptotic distribution of $n^{-1/2} \sum_{t=0}^{n-1} [\underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i) - E \{ \underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i) \}]$ is the same as the one of

$$\frac{1}{\sqrt{n}} \left\{ \sum_{t=0}^{n-1} [\underline{U}_t^{\odot 2} \odot \underline{U}_{t-i}^{\odot 2}(i) - E \{ \underline{U}_t^{\odot 2} \odot \underline{U}_{t-i}^{\odot 2}(i) \}] \right\} \text{ as } m \rightarrow \infty.$$

Now, for any sequence of 2×1 vectors $(\Delta_i)_{0 \leq i \leq h}$, let $P_n = n^{-1/2} \sum_{t=0}^{n-1} Y_t(h)$, where

$$Y_t(h) = \sum_{i=0}^h \Delta_i' [\underline{U}_t^{\odot 2} \odot \underline{U}_{t-i}^{\odot 2}(i) - E \{ \underline{U}_t^{\odot 2} \odot \underline{U}_{t-i}^{\odot 2}(i) \}].$$

Clearly, $\{Y_t(h)\}_{t \in \mathbb{Z}}$ is a stationary $(m + 1)$ -dependent process with

$$E \{Y_t(h)Y_{t+k}(h)\} = \underline{\lambda}'W(k)\underline{\lambda} < +\infty,$$

where $\underline{\lambda} = (\underline{\lambda}'_0, \dots, \underline{\lambda}'_h)'$ and where $W(k)$ is the covariance matrix with (i, j) -th block matrix being

$$\{W(k)\}_{i,j} = \text{Cov} \left\{ \underline{U}_t^{\odot 2} \odot \underline{U}_{t-i}^{\odot 2}(i), \underline{U}_{t+k}^{\odot 2} \odot \underline{U}_{t+k-j}^{\odot 2}(j) \right\}, \quad i, j = 0, \dots, h.$$

Therefore, we have $n^{-1/2} \sum_{t=0}^{n-1} \underline{Y}_t(h) \rightsquigarrow \mathcal{N}(0, \underline{\lambda}'W\underline{\lambda})$, where the (i, j) -th block matrix of the covariance matrix W is $(W)_{i,j} = \sum_{k=-m}^m \{W(k)\}_{i,j}$. As $m \rightarrow \infty$, W converges to Σ whose (i, j) -th block matrix is

$$(\Sigma)_{i,j} = \sum_{k \in \mathbb{Z}} \text{Cov} \left\{ \underline{X}_t^{\odot 2} \odot \underline{X}_{t-i}^{\odot 2}(i), \underline{X}_{t+k}^{\odot 2} \odot \underline{X}_{t+k-j}^{\odot 2}(j) \right\}, \quad i, j = 0, \dots, h.$$

Finally the proof follows from the Cramer-Wold device. □

The following theorem states the asymptotic normality of $\widehat{\underline{b}}_n$.

Theorem 7. *Under A1–A4 and if $(\underline{X}_t)_{t \in \mathbb{Z}}$ admits moments up to 8-th order, then*

$$\sqrt{n} \left(\widehat{\underline{b}}_n - \underline{b}_0 \right) \rightsquigarrow \mathcal{N}(0, A^{-1}B\Sigma B' A'^{-1}).$$

Proof. The proof rests classically on a Taylor-series expansion of the score vector $F_n(\underline{b})$ around \underline{b}_0 . Thus, by the same argument used in Theorem 6, we have

$$(10) \quad \left\{ \frac{\partial \Gamma'(\widehat{\underline{b}}_n)}{\partial \underline{b}} M_n \frac{\partial \Gamma(\widehat{\underline{b}}_*)}{\partial \underline{b}} \right\} \left(\widehat{\underline{b}}_n - \underline{b}_0 \right) = \frac{\partial \Gamma'(\widehat{\underline{b}}_n)}{\partial \underline{b}} M_n \widehat{F}_n(\underline{b}_0).$$

From Lemma 1, we have $p \lim_{n \rightarrow \infty} \widehat{F}_n(\underline{b}_0) = 0$, and thus $p \lim_{n \rightarrow \infty} \widehat{\underline{b}}_n = \underline{b}_0$. On the other hand, in the expansion (10) we have the following limits

$$A = p \lim_{n \rightarrow \infty} \frac{\partial \Gamma'(\widehat{\underline{b}}_n)}{\partial \underline{b}} M_n \frac{\partial \Gamma(\widehat{\underline{b}}_*)}{\partial \underline{b}}, \quad B = p \lim_{n \rightarrow \infty} \frac{\partial \Gamma'(\widehat{\underline{b}}_n)}{\partial \underline{b}} M_n$$

since $\sqrt{n}\widehat{F}_n(\underline{b}) \rightsquigarrow \mathcal{N}(0, \Sigma)$. Then, the result simply follows from Slutsky's theorem. □

The result in Theorem 7 can be easily generalized in Theorem 8 to the estimator $\widetilde{\underline{b}}_n$. Indeed, we have $\widetilde{F}_n(\underline{b}) - \widehat{F}_n(\underline{b}) = \underline{\Gamma}(\underline{b}) - \underline{\Gamma}_n(\underline{b})$ and $\sqrt{n}\{\widetilde{F}_n(\underline{b}) - \widehat{F}_n(\underline{b})\}_{k,k} = -\sqrt{n}B(k)(\widehat{\underline{\mu}}_n - \underline{\mu})$. From Theorem 5, we have $\text{Var}(\widehat{\underline{\mu}}_n) = O(n^{-1})$ and thus $p \lim_{n \rightarrow \infty} \sqrt{n}(\widehat{\underline{\mu}}_n - \underline{\mu}) = 0$. Hence $p \lim_{n \rightarrow \infty} \sqrt{n}\{\widetilde{F}_n(\underline{b}) - \widehat{F}_n(\underline{b})\} = 0$ and consequently $\sqrt{n}\widetilde{F}_n(\underline{b})$ converges to the same limit distribution as the one of $\sqrt{n}\widehat{F}_n(\underline{b})$.

Theorem 8. *Under A1–A4 and if $(\underline{X}_t)_{t \in \mathbb{Z}}$ admits moments up to 8-th order, then*

$$\sqrt{n} \left(\widetilde{\underline{b}}_n - \underline{b}_0 \right) \rightsquigarrow \mathcal{N}(0, A^{-1}B\Sigma B' A'^{-1}).$$

Remark 7. We now discuss the optimal choice of the weighting matrix M . It is clear from Theorem 7 that the asymptotic variance of \widehat{b}_n depends on M_n via M . As it is the case for GMM estimation, under the conditions of Theorem 8, the choice of M matters for asymptotic efficiency. When appropriately choosing M , it is possible to minimize the asymptotic variance of \widehat{b}_n . Then the minimum variance that can be achieved is when $M = \Sigma^{-1}$. In this particular case, the asymptotic variance of \widehat{b}_n is

$$\left\{ \frac{\partial \underline{\Gamma}'(b_0)}{\partial \underline{b}} \Sigma^{-1} \frac{\partial \underline{\Gamma}(b_0)}{\partial \underline{b}} \right\}^{-1}$$

and $n\widehat{Q}_n(\underline{b})$ has an asymptotic chi-square distribution (see Hall [8, Theorem 3.4]). One can note that this choice is only sufficient for efficiency. Hence, estimating the matrix Σ by a consistent estimator $\widehat{\Sigma}_n$ is crucial since: i) it is the optimal weighting matrix of MDE ; ii) it is a part of the construction of \widehat{b}_n (see Hall [8, Chapter 3]) and its asymptotic variance (needed to construct confidence intervals and to make statistical tests based on \widehat{b}_n). In practice, the Newey-West estimator (see Newey and McFadden [9]) can be used:

$$\widehat{\Sigma}_n = \widehat{\Omega}_n(0) + \sum_{j=1}^q K\left(\frac{j}{q}\right) \left\{ \widehat{\Omega}_n(j) + \widehat{\Omega}'_n(j) \right\},$$

where $\widehat{\Omega}_n(j) = n^{-1} \sum_{t=0}^{n-j-1} W_t W'_{t+j}$ with $W_t = \{W'_t(0), \dots, W'_t(h)\}'$, $W_t(k) = (\underline{X}_t^{\odot 2} - \widehat{\underline{\mu}}_n) \odot \{\underline{X}_{t-k}^{\odot 2}(k) - \widehat{\underline{\mu}}_n(k)\} - \underline{\gamma}(k)$ for $k = 0, \dots, h$, the truncated lag q needs to go to infinity at some appropriate rate with respect to the sample, and the kernel weight $K(j/q)$ is assumed to satisfy $K(\cdot) \in \mathcal{K}$, where $\mathcal{K} = \{k : \mathbb{R} \rightarrow [-1, 1] \mid k(0) = 1, k(x) = k(-x), \forall x \in \mathbb{R}, \int |k(x)| dx < \infty, \text{ and } k \text{ is continuous but at some countable points}\}$. Examples of such kernel weights are the following:

name	expression
truncated	$k_T(x) = \begin{cases} 1 & \text{if } x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$
Bartlett	$k_B(x) = \begin{cases} 1 - x & \text{if } x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$
Parzen	$k_P(x) = \begin{cases} 1 - 6x^2 + 6 x ^3 & \text{if } x \leq 1/2, \\ 2(1 - x)^3 & \text{if } 1/2 < x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$
Tukey-Hanning	$k_H(x) = \begin{cases} (1 + \cos \pi x)/2 & \text{if } x \leq 1, \\ 0 & \text{otherwise,} \end{cases}$
quadratic spectral	$k_Q(x) = \frac{25}{12(\pi x)^2} \left\{ \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right\}$.

It can be shown that Bartlett, Parzen and quadratic spectral kernels all product positive semi-definite estimates of Σ while this is not necessarily the case for truncated and Tukey-Hanning kernels.

4. Hypotheses testing

As an application of Theorem 8, we consider the problem of testing a null hypothesis against an alternative hypothesis of the form

$$(11) \quad H_0 : R\underline{b} = \underline{b}^* \quad \text{v.s.} \quad H_1 : R\underline{b} \neq \underline{b}^*,$$

where R is a given 2×2 matrix of rank 2, and \underline{b}^* is a given vector. Under the null hypothesis H_0 in (11) and under the conditions of Theorem 8,

$$\sqrt{n} \left(R\tilde{\underline{b}}_n - \underline{b}^* \right) \rightsquigarrow \mathcal{N} \left(0, RA^{-1}B\Sigma B'A'^{-1}R' \right).$$

Moreover, if the matrix Σ is nonsingular, then the asymptotic variance matrix involved below is nonsingular. So we have the following result from the continuous mapping theorem.

Theorem 9. *Assume that the conditions of Theorem 8 hold and Σ is a non-singular matrix. Then, under the null hypothesis in (11) with R of rank 2, we have*

$$(12) \quad W_n = n \left(R\tilde{\underline{b}}_n - \underline{b}^* \right)' \left(R\hat{A}_n^{-1}\hat{B}_n\hat{\Sigma}_n\hat{B}'_n\hat{A}'_n{}^{-1}R' \right)^{-1} \left(R\tilde{\underline{b}}_n - \underline{b}^* \right) \rightsquigarrow \chi_2^2,$$

where \hat{A}_n and \hat{B}_n are some consistent estimates of A and B respectively. In addition, under the alternative hypothesis in (11), we have

$$(13) \quad p \lim_{n \rightarrow \infty} n^{-1}W_n = (R\underline{b} - \underline{b}^*)' (RA^{-1}B\Sigma B'A'^{-1}R')^{-1} (R\underline{b} - \underline{b}^*) > 0.$$

Note that the test statistics W_n is now the one of the Wald test of the null hypothesis in (11). Given the size $\alpha \in [0, 1]$, choose a critical value β so that under the null hypothesis in (11), $\mathbb{P}(W_n > \beta) \rightarrow \alpha$. Then the null hypothesis is accepted if $W_n \leq \beta$, and rejected in favor of the alternative hypothesis if $W_n > \beta$. This test is consistent due to (13). In the case when R is a row vector (so \underline{b}^* is a scalar), we can modify (12) to

$$t_n = \sqrt{n} \left(R\hat{A}_n^{-1}\hat{B}_n\hat{\Sigma}_n\hat{B}'_n\hat{A}'_n{}^{-1}R' \right)^{-1/2} \left(R\tilde{\underline{b}}_n - \underline{b}^* \right) \rightsquigarrow \mathcal{N} (0, 1),$$

whereas under the alternative hypothesis in (11), (13) becomes

$$p \lim_{n \rightarrow \infty} \frac{t_n}{\sqrt{n}} = (RA^{-1}B\Sigma B'A'^{-1}R')^{-1/2} (R\underline{b} - \underline{b}^*) \neq 0.$$

These results can be used to construct two-sided or one-sided tests. In particular, we have the following result.

Theorem 10. *Assume that the conditions of Theorem 9 hold. Consider the hypotheses*

$$H_0^{(v)} : b_v = b_v^* \quad \text{v.s.} \quad H_1^{(v)} : b_v \neq b_v^*, \quad v \in \{1, 2\},$$

where b_v^* is given. Let $(e_i, i = 1, 2)$ be the canonical basis of \mathbb{R}^2 . Then, under $H_0^{(v)}$,

$$t_n(v) = \frac{\sqrt{n} (\tilde{b}_{v,n} - b_v^*)}{\sqrt{e'_v \widehat{A}_n^{-1} \widehat{B}_n \widehat{\Sigma}_n \widehat{B}'_n \widehat{A}_n^{-1} e_v}} \rightsquigarrow \mathcal{N}(0, 1),$$

whereas under $H_1^{(v)}$,

$$p \lim_{n \rightarrow \infty} \frac{t_n(v)}{\sqrt{n}} = \frac{b_v - b_v^*}{\sqrt{e'_v A^{-1} B \Sigma B' A^{-1} e_v}} \neq 0.$$

Given the size $\alpha \in]0, 1[$, choose a critical value β so that, if the null hypothesis is true, we have $\mathbb{P}\{|t_n(v)| > \beta\} \rightarrow \alpha$. Then the null hypothesis is accepted if $|t_n(v)| \leq \beta$, and rejected in favor of the alternative hypothesis if $|t_n(v)| > \beta$.

Remark 8. One can notice that this test is obviously consistent. Moreover, it is worth mentioning that the value $b_v^* = 0$ is of special interest to test the Gaussianity.

5. Simulation results

In order to illustrate the performance of our asymptotic results described in previous sections, we now provide some numerical results from Monte Carlo experiments. We simulated 1000 independent trajectories via a periodic bilinear processes of length $n \in \{100, 500, 2000\}$ with period $s = 2$, standard normal errors distribution and parameter $\underline{b} = (b_1, b_2)'$ which satisfies the second order stationarity and existence of moments up to 4-th order conditions, that is to say $\max\{b_1^4, b_2^4\} < 1/3$. For the asymptotic normality, we assume that $\max\{b_1^8, b_2^8\} < 1/105$. For each trajectory, the parameter vector \underline{b} has first been estimated with MDE, noted as $\tilde{\underline{b}}_n^M = (\tilde{b}_{1,n}^M, \tilde{b}_{2,n}^M)'$. All the efficient Minimum Distance (MD) estimations have been performed with the parameter dependent truncated kernel weight k_T described in Remark 7. For purposes of comparison, the parameter vector \underline{b} has also been estimated by a usual Least Squares (LS) method, which produced the estimator noted as $\tilde{\underline{b}}_n^L$. In addition, in order to have an heavy-tailed distribution for the errors, we considered the Student t_5 -distribution as well (where 5 denotes the number of degrees of freedom), to replace the standard Gaussian assumption in MD and LS estimation. This additional experiment is made to emphasize that the proposed asymptotic theory is free from the Gaussianity assumption. In Tables 1–2 below, the rows “Means” correspond to the average of the parameters estimates over the 1000 simulations. We give into brackets the results obtained from the t_5 -distribution for the errors process (ϵ_t) . Replacing the unknown parameters by their estimates, we obtain the estimates \widehat{V}_n , $\widehat{W}_n(h, h)$ and $\widehat{\Sigma}_n$ respectively for V , $W(h, h)$ defined in Theorem 5 and for Σ defined in Lemma 1. We denote by $\sqrt{\text{Var}_{as}(\tilde{b}_{v,n})} = n^{-1/2} \sqrt{(\widehat{\Sigma}_n)_{v,v}}$ the estimator of the standard deviation.

TABLE 1. Characteristics of the empirical distribution for MDE and LS estimate, and rejection relative frequencies of the null hypotheses, for Design 1. The brackets give the estimations obtained from a t_5 -distribution for the errors. The number of replications is 1000.

Sample size n	100		500		2000	
Estimate	\tilde{b}_n^M	\tilde{b}_n^L	\tilde{b}_n^M	\tilde{b}_n^L	\tilde{b}_n^M	\tilde{b}_n^L
	Design 1: $\underline{b} = (0.5, -0.5)'$					
Means	0.4246 (0.4014)	0.4938 (0.5004)	0.4865 (0.4786)	0.4988 (0.4972)	0.4963 (0.4896)	0.4970 (0.4908)
	-0.4300 (-0.4100)	-0.4946 (-0.5067)	-0.4881 (-0.4761)	-0.4989 (-0.4951)	-0.4975 (-0.4914)	-0.4974 (-0.4920)
RMSE*	0.1893 (0.1781)	0.0650 (0.1144)	0.0752 (0.0753)	0.0264 (0.0572)	0.0349 (0.0402)	0.0257 (0.0632)
	0.1864 (0.1734)	0.0668 (0.1107)	0.0732 (0.0766)	0.0274 (0.0601)	0.0352 (0.0374)	0.0243 (0.0592)
RMSE	0.2037 (0.2027)	0.0662 (0.1143)	0.0764 (0.0782)	0.0264 (0.0572)	0.0351 (0.0415)	0.0259 (0.0639)
	0.1990 (0.1953)	0.0670 (0.1109)	0.0741 (0.0802)	0.0274 (0.0603)	0.0353 (0.0394)	0.0244 (0.0597)
$H_0^{(1)}$ rejection	68.9% (70.8%)	100% (98.2%)	99.9% (100%)	100% (99.2%)	100% (100%)	100% (99.1%)
$H_0^{(2)}$ rejection	71.2% (74.6%)	100% (98.5%)	99.8% (99.7%)	100% (99.3%)	100% (100%)	100% (99.1%)
$H_0^{(3)}$ rejection	84.8% (84.8%)	100% (99.5%)	100% (100%)	100% (99.9%)	100% (100%)	100% (99.6%)

In order to demonstrate that this estimate, although based on the asymptotic theory, can be successfully applied to finite samples of reasonable size, the average of $\sqrt{\text{Var}_{as}(\tilde{b}_{v,n})}$ over the 1000 simulations, denoted RMSE*, has been compared to the root of the mean of $(\tilde{b}_{v,n} - b_v)^2$ over the 1000 simulations, denoted by RMSE.

Now, let us consider the null hypotheses $H_0^{(1)} : b_1 = 0$, $H_0^{(2)} : b_2 = 0$ and $H_0^{(3)} : b_1 = b_2$. With MDE, $H_0^{(1)}$ (respectively $H_0^{(2)}$) is rejected when $|\tilde{b}_{1,n}^M|$ (respectively $|\tilde{b}_{2,n}^M|$) is greater than 1.96 times its estimated standard deviation $\sqrt{\text{Var}_{as}(\tilde{b}_{1,n}^M)}$ (respectively $\sqrt{\text{Var}_{as}(\tilde{b}_{2,n}^M)}$). If the asymptotic theory applies for such sample sizes, then the errors of first kind should be approximately 5%. More precisely $H_0^{(3)}$ is rejected when $|\tilde{b}_{1,n}^M - \tilde{b}_{2,n}^M|$ is greater than $1.96\sqrt{R\text{Var}_{as}(\tilde{b}_n^M)R'}$. We used the same procedure for the LS estimate \tilde{b}_n^L , as well as with a t_5 -distribution for the errors.

The simulation results are reported in Tables 1–2. The results related to Design 1 are in accordance with the asymptotic theory presented in the paper. It is clear that the RMSE’s are greater than the RMSE*’s, and the tests based on MDE are powerful for large sample sizes. The results reported for Design 2 also conclude to the performance of MDE for large sample sizes. The analysis of the simulation results for the MDE in the case $n = 100$ shows that the asymptotic theory should be used with caution for moderate sample sizes. In addition, even for large samples, the rejection relative frequency of $H_0^{(2)}$ is far from the theoretical 5% for $n = 2000$. Regarding the results given into brackets, but for the columns dealing with \tilde{b}_n^M , it is worth noting that, on the one hand, the effect of the t_5 -distribution is insignificant apart from the RMSE or RMSE*. On the other hand, the effect of the t_5 -distribution is meaningful in the remaining columns when the Gaussianity assumption is better than the Student distribution. In Design 2, with both estimators, the rejection relative frequency $H_0^{(3)}$ is deteriorated under both the Gaussian and the Student distributions. This is due to the fact that the standard deviation of the estimators are not well estimated by $\sqrt{\text{Var}_{as}(\cdot)}$, even for large samples. To solve this problem, some alternative estimators for Σ , like bootstrap or Newey-West (see Remark 7) estimators could be investigated, but it is not the scope of the paper. Finally, we can notice that, as expected, the LS estimate performs very well in each design, even for moderate sample sizes (see $n = 100$).

TABLE 2. As in Table 1, but for Design 2.

Sample size n	100		500		2000	
Estimate	\tilde{b}_n^M	\tilde{b}_n^L	\tilde{b}_n^M	\tilde{b}_n^L	\tilde{b}_n^M	\tilde{b}_n^L
	Design 2: $b = (0.5, 0.5)'$					
Means	0.4247 (0.4014)	0.4967 (0.4985)	0.4865 (0.4786)	0.4974 (0.4970)	0.4963 (0.4896)	0.4980 (0.4901)
	0.4359 (0.4169)	0.4981 (0.4986)	0.4873 (0.4761)	0.4975 (0.4965)	0.4965 (0.4913)	0.4983 (0.4902)
RMSE*	0.1890 (0.1766)	0.0647 (0.1129)	0.0752 (0.0753)	0.0319 (0.0525)	0.0349 (0.0402)	0.0223 (0.0632)
	0.1851 (0.1702)	0.0669 (0.1122)	0.0760 (0.0764)	0.0288 (0.0552)	0.0340 (0.0397)	0.0209 (0.0630)
RMSE	0.2034 (0.2022)	0.0647 (0.1129)	0.0764 (0.0792)	0.0321 (0.0525)	0.0351 (0.0415)	0.0224 (0.0639)
	0.1953 (0.1903)	0.0669 (0.1111)	0.0771 (0.0800)	0.0289 (0.0553)	0.0342 (0.0397)	0.0210 (0.0636)
$H_0^{(1)}$ rejection	69.4% (71.2%)	100% (98.4%)	100% (100%)	100% (99.6%)	100% (100%)	100% (98.9%)
$H_0^{(2)}$ rejection	73.3% (77.3%)	100% (98.4%)	100% (99.9%)	100% (99.6%)	100% (100%)	100% (98.7%)
$H_0^{(3)}$ rejection	3.7% (2.5%)	3.4% (3.9%)	2.6% (1.3%)	0.6% (1.0%)	2.2% (2.2%)	0.1% (1.0%)

6. Concluding remarks

This paper investigated the question of parameter estimation for a periodic bilinear model, which allows the coefficients to vary periodically with respect to time. This problem has been previously resolved in the statistical literature for the usual time-constant case. We considered a distribution-free approach based on MDE. We showed that MD (and LS) estimates perform very well for large sample sizes, not only with a common Gaussian assumption for the noise, but also with heavy-tailed distribution for the error, the Student distribution being an example. Consistency and asymptotic normality of the MDE, as well as hypotheses testing, have been derived. The behavior of the estimators has also been studied via simulations, showing satisfactory (and expected) results.

References

- [1] A. Bibi, *On the covariance structure of time varying bilinear models*, Stochastic Anal. Appl. **21** (2003), no. 1, 25–60.
- [2] A. Bibi and A. Aknouche, *Yule-Walker type estimators in periodic bilinear models: strong consistency and asymptotic normality*, Stat. Methods Appl. **19** (2010), 1–30.
- [3] A. Bibi and A. Gautier, *Propriétés dans L^2 et estimation des processus purement bilinéaires et strictement superdiagonaux à coefficients périodiques*, Canad. J. Statist. **34** (2006), no. 1, 131–148.
- [4] A. Bibi and A. J. Oyet, *Estimation of some bilinear time series models with time varying coefficients*, Stochastic Anal. Appl. **22** (2004), no. 2, 355–376.
- [5] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, Springer-Verlag, New York, 1996.
- [6] C. Francq, *ARMA models with bilinear innovations*, Comm. Statist. Stochastic Models **15** (1999), no. 1, 29–52.
- [7] A. Gautier, *Asymptotic inefficiency of mean-correction on parameter estimation for a periodic first-order autoregressive model*, Comm. Statist. Theory Methods **35** (2006), no. 10-12, 2083–2106.
- [8] A. Hall, *Generalized Method of Moments*, Oxford University Press, Oxford, 2005.
- [9] W. K. Newey and D. McFadden, *Large sample estimation and hypothesis testing*, Handbook of econometrics, Vol. IV, 2111–2245, Handbooks in Econom., 2, North-Holland, Amsterdam, 1994.

ABDELOUAHAB BIBI
 DÉPARTEMENT DE MATHÉMATIQUES
 UNIVERSITÉ MENTOURI, CONSTANTINE
 ALGÉRIE
E-mail address: `abd.bibi@gmail.com`

ANTONY GAUTIER
 UNIVERSITÉ LILLE 3
 LABORATOIRE EQUIPPE - UFR MSES - BP 60149
 59653 VILLENEUVE D'ASCQ CEDEX, FRANCE
E-mail address: `antony.gautier@univ-lille3.fr`