Quasi-maximum likelihood estimation of Markov-switching bilinear time series model

Abdelouahab Bibi
Département de mathématiques, Université de Constantine (1), Algeria
e-mail address: a.bibi@umc.edu.dz

Abstract

In this talk, we consider the class of bilinear processes with Markov switching ($MS-BL$) that offers remarkably rich dynamics and may be considered as an alternative to models with constant coefficients non Gaussian data. So, firstly, some basic issues concerning this class of models including sufficient conditions ensuring the existence of stationarity (in strict sense) and ergodic solutions are given. Secondary, we illustrate the fundamental problems linked with $MS-BL$ models, i.e., parameters estimation by considering a quasi-likelihood (QML) approach. So, we provide the detail on the asymptotic properties of QML, in particular, we discuss conditions for its consistency and asymptotic normality for $MS-BL$.

Résumé

Dans cette communication, nous étudions la classe des modèles bilinéaires à changement de régimes markoviens. Nous donnons des conditions suffisantes de stationnarité stricte et au second ordre. Une approche par quasi-maximum de vraisemblance est proposée pour estimer les paramètres du modèle et ses propriétés asymptotiques.

Keywords: MS-bilinear processes, Quasi-maximum likelihood, Strong consistency, Asymptotic normality.

1 Introduction

Markov-switching time series models (MSM) have received recently a growing interest in macroeconomics because of their ability to adequately describe various observed time series subjected to change in regime. Flexibility is one of the main advantages of MSM which become an appealing tool for the modelling of business cycles and continue to gain popularity especially in financial time series which exhibits structural changes in regime.

In this talk we concern ourselves with a large class of models, the discrete-time bilinear model $(X_t, t \in \mathbb{Z})$ defined on some probability space $(\Omega, \Im, P)$ and generalized by the following stochastic difference equation

$$X_t = a_0(s_t) + \sum_{i=1}^{p} a_i(s_t)X_{t-i} + e_t + \sum_{j=1}^{q}\sum_{i=j}^{p} c_{ij}(s_t)X_{t-i}e_{t-j} \tag{1.1}$$

denotes by $MS-BL(p,0,p,q)$. In (1.1), the innovation process $(e_t, t \in \mathbb{Z})$ is supposed to be defined on the same probability space $(\Omega, \Im, P)$ with $E\{e_t\} = 0$ and $E\{\log^+ |e_t|\} < +\infty$ where for $x > 0$, $\log^+ x = \max(\log x, 0)$. The functions $a_i(s_t)$, $b_j(s_t)$ and $c_{ij}(s_t)$ depend upon an unobservable first order Markov chain $(s_t, t \in \mathbb{Z})$ that controls the dynamics of X_t and subject to the following assumption:

Assumption 1 *The Markov chain $(s_t, t \in \mathbb{Z})$ is irreducible, aperiodic (and hence stationary and ergodic), finite state space $\mathbb{S} = \{1, ..., d\}$, $n-$step transition probabilities matrix $\mathbb{P}^n = \left(p_{ij}^{(n)}, (i,j) \in \mathbb{S} \times \mathbb{S}\right)$ where $p_{ij}^{(n)} = P(s_t = j | s_{t-n} = i)$ with one-step transition probability matrix $\mathbb{P} := (p_{ij}, (i,j) \in \mathbb{S} \times \mathbb{S})$ where $p_{ij} := p_{ij}^{(1)} = P(s_t = j | s_{t-1} = i)$ for $i, j \in \mathbb{S}$, and stationary distribution $\underline{\pi} = (\pi(1), ..., \pi(d))'$ where $\pi(i) = P(s_0 = i)$, $i = 1, ..., d$. In addition, we shall assume that e_t and $\{(X_{s-1}, s_t), s \leq t\}$ are independent.*

It is convenient to represent (1.1) in a state-space representation $X_t = \underline{F}'\underline{Z}_t$ and

$$\underline{Z}_t = \Gamma_t \underline{Z}_{t-1} + \underline{\eta}_t. \tag{1.2}$$

where $\Gamma_t = \Gamma_0(s_t) + e_t\Gamma_1(s_t)$ with $\Gamma_i(s_t)$, $i = 0, 1$ and $\underline{F}, \underline{Z}_t, \underline{\eta}_t$ are appropriate matrices and vectors. For such representation the extended process $\left(\widetilde{\underline{Z}}_t := (\underline{Z}_t', s_t)', t \in \mathbb{Z}\right)$ is a Markov chain on $\mathbb{R}^s \times \mathbb{S}$.

2 Stationarity of $MS - BL$ processes

Now, since $((s_t, e_t), t \in \mathbb{Z})$ is stationary and ergodic process, then $(\Gamma_t, t \in \mathbb{Z})$ is also a stationary and ergodic process and both $E\{\log^+ \|\Gamma_t\|\}$ and $E\left\{\log^+ \left\|\underline{\eta}_t\right\|\right\}$ are finite. So, from Bougerol and Picard [1], the unique, causal, bounded in probability, strictly stationary and ergodic solution of (1.2) is given by almost surely $(a.s)$

$$\underline{Z}_t = \sum_{k=1}^{\infty}\left\{\prod_{i=0}^{k-1}\Gamma_{t-i}\right\}\underline{\eta}_{t-k} + \underline{\eta}_t \tag{2.1}$$

whenever the Lyapunov exponent $\gamma_L(\Gamma)$ associated with the sequence of random matrices $(\Gamma_t, t \in \mathbb{Z})$ and defined by

$$\gamma_L(\Gamma) := \inf\left\{\frac{1}{n}E\log\left\|\prod_{i=0}^{n-1}\Gamma_{t-i}\right\|, n \geq 1\right\} \overset{a.s}{=} \lim_{n\to\infty}\frac{1}{n}\log\left\|\prod_{i=0}^{n-1}\Gamma_{t-i}\right\| \tag{2.2}$$

is strictly negative. The second equality in (2.2) can be justified using Kingman's sub-additive ergodic theorem and the existence of $\gamma_L(M)$ is guaranteed by the fact that $E\left\{\log^+\|\Gamma_t\|\right\} \leq E\{\|\Gamma_t\|\} < +\infty$. So we have

Theorem 2.1 *Consider the $MS-BL(p,q,p,Q)$ process (1.1) with state−space representation (1.2) and suppose that $\gamma_L(\Gamma) < 0$. Then,*

1. *for all $t \in \mathbb{Z}$, the series (2.1) converges absolutely almost surely and the process $(\underline{F}'\underline{Z}_t, t \in \mathbb{Z})$ constitutes the unique, strictly stationary, ergodic and causal solution of (1.1).*
2. *almost surely, the sequence $\left(\prod_{i=0}^{n}\Gamma_{t-i}\underline{x}, n \geq 0\right)$ converge to the $\underline{0}$ for any $\underline{x} \in \mathbb{R}^r$.*

Proposition 2.1 *Assume that $E\left\{e_t^4\right\} < +\infty$ and set $\Phi_n(s_0) = E\left\{\left\|\prod_{j=0}^{n}\Gamma_{n-j}\right\| |s_0\right\}$, then*

1. *$\Phi_n(s_0)$ converges to 0 iff $\Phi_n(s_0)$ converges to 0 at an exponential rate as $n \to +\infty$.*

2. *The $MS-SBL(p,0,p,Q)$ process has a unique, strictly stationary, ergodic, causal and bounded in probability solution whose second-order moment exists if*

$$\lim_{n\to+\infty}\Phi_n(s_0) = 0. \tag{2.3}$$

3. *Under (2.3), the autocovariance matrix $\Sigma(m) = Cov\left\{\underline{Z}_{t+m}, \underline{Z}_t\right\}$ of the process $(\underline{Z}_t, t \in \mathbb{Z})$ decays at a geometric rate.*

3 Quasi-likelihood estimation for $MS-BL$

In this talk, we consider the model (1.1) in which, the innovation process $(e_t, t \in \mathbb{Z})$ is an $i.i.d$ sequence with zero mean and variance 1, the orders p, q and the regimes number d are assumed to be known and fixed, the $r-$unknown parameters, gathered in $\underline{\theta} := (\underline{\theta}_1', ..., \underline{\theta}_d')'$ belongs to a some Euclidian parameter space Θ where $\underline{\theta}_i = \left(\underline{a}_i', \underline{c}_i', \underline{\pi}', \underline{p}_i'\right)'$ with vectors coordinate projections $\underline{a}_i := (a_0(i), ..., a_p(i))'$, $\underline{c}_i := (c_{lk}(i), 1 \leq k \leq l \leq q)'$, $\underline{p}_i := (p_{ij}, ..., p_{id}, j \neq i)'$. The true parameter is denoted $\underline{\theta}_0$ and for any integers a and b, let $\underline{X}_{a:b}$ (resp. $\underline{\mathcal{X}}_{a:b}$) denotes the set $\{X_a, X_{a+1}\ldots, X_b\}$ (resp. $\{(X_a, e_a), (X_{a+1}, e_{a+1}), \ldots, (X_b, e_b)\}$) with possibly $a = -\infty$ in this cases we shall note $\underline{X}_b$ (resp. $\underline{\mathcal{X}}_b$). The problem of interest in this talk is the estimation of the parameter vector $\underline{\theta}$ governing Equation (1.1) from an observed sequence $\underline{X}_{1:n}$ and unobserved $((s_t, e_t), t \in \mathbb{Z})$, for this purpose, we shall denote the density function of observations by $g_{\underline{\theta}}(.)$ and that of innovations e_t by $f(.)$.. The quasi-likelihood $L_n(\underline{\theta})$ that we work with is given by

$$L_n(\underline{\theta}) = \sum_{(x_1,\dots,x_n)\in\mathbb{S}^n} \pi(x_1)\, g_{\underline{\theta}_{x_1}}\left(X_1|\underline{\mathcal{X}}_{1-p_0:0}\right) \prod_{t=2}^{n} p_{x_{t-1},x_t} g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)$$

. A quasi-maximum likelihood estimator ($QMLE$) of $\underline{\theta}$ is defined as any measurable solution $\widehat{\underline{\theta}}_n$ of

$$\widehat{\underline{\theta}}_n = \arg\max_{\theta\in\Theta} L_n(\underline{\theta}). \tag{3.1}$$

For the asymptotic purpose, it is convenient to approximate the process $g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)$ by its ergodic stationary version $g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)$ so we work with an approximate version $\widetilde{L}_n(\underline{\theta})$, i.e., $\widetilde{L}_n(\underline{\theta}) = \sum_{(x_1,\dots,x_n)\in\mathbb{S}^n} \pi(x_1)\, g_{\underline{\theta}_{x_1}}\left(X_1|\underline{\mathcal{X}}_0\right) \prod_{t=2}^{n} p_{x_{t-1},x_t} g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)$. Our approach is benefitted from the papers by [4], [2] and by [3].

3.1 Consistency of $QMLE$

Define $p_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)$ (resp. $q_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)$) the conditional density of X_t given $\underline{\mathcal{X}}_{1-p_0:t-1}$(resp. given $\underline{\mathcal{X}}_{t-1}$) and $p^*_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)$ (resp. $q^*_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)$) its logarithm and consider the following regularities conditions.

A1. $\underline{\theta}_0 \in \Theta$ and Θ is a compact subset of $\mathbb{R}^r$

A2. $\gamma_L(M^0) < 0$ for all $\underline{\theta} \in \Theta$ where M^0 denotes the sequence $(M_t, t\in\mathbb{Z})$ when the parameters $\underline{\theta}_i$ are replaced by their true values $\underline{\theta}_i^0$, $i = 1, \dots, d$.

A3.

a. For all $\underline{\theta} \in \Theta$, almost surely $0 < \min_k \left\{g_{\underline{\theta}_k}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)\right\} < \max_k \left\{g_{\underline{\theta}_k}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)\right\} < +\infty$

b. There exists a neighborhood $\mathcal{V}(\underline{\theta})$ of $\underline{\theta}$ such that $E_{\underline{\theta}_0}\left\{\sup_{\underline{\theta}'\in\mathcal{V}(\underline{\theta})} \left|q^*_{\underline{\theta}'}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)\right|\right\} < \infty$ for some $\delta > 0$.

A4. Identifiability Condition: For any $\underline{\theta}, \underline{\theta}' \in \Theta$, if almost surely $q_{\underline{\theta}}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right) = q_{\underline{\theta}'}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)$, then $\underline{\theta} = \underline{\theta}'$.

First we show the following general results.

Lemma 3.1 *Under **A2** and **A3**, almost surely, uniformly with respect to* $\underline{\theta} \in \Theta$, $\lim_{n\longrightarrow\infty} \frac{1}{n}\log \widetilde{L}_n(\underline{\theta}) = \lim_{n\longrightarrow\infty} \frac{1}{n}\log L_n(\underline{\theta}) = E_{\underline{\theta}_0}\left\{q^*_{\underline{\theta}}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)\right\}$.

Lemma 3.2 *Let* $Z_n(\underline{\theta}) = \frac{1}{n}\log\left(\frac{\widetilde{L}_n(\underline{\theta})}{\widetilde{L}_n(\underline{\theta}_0)}\right)$ *for all* $\underline{\theta} \in \Theta$. *Then under the conditions* ***A1-A4,*** *almost surely* $\lim_{n\longrightarrow\infty} Z_n(\underline{\theta}) \leq 0$ *with equality iff* $\underline{\theta} = \underline{\theta}_0$.

Lemma 3.3 *Under the assumptions* ***A1-A4.*** *For all* $\underline{\theta}' \neq \underline{\theta}_0$, *there exists a neighborhood* $\mathcal{V}(\underline{\theta}')$ *of* $\underline{\theta}'$ *such that almost surely* $\lim_{n\to+\infty}\sup \sup_{\underline{\theta}\in\mathcal{V}_m(\underline{\theta}')} Z_n(\underline{\theta}) < 0$.

Theorem 3.1 *For the* $MS - BL$ *model* ((1.1)), *let* $\hat{\underline{\theta}}_n$ *be the* $QMLE$ *sequence over* Θ *satisfying* (3.1). *Then under the conditions* ***A1-A4,*** $\hat{\underline{\theta}}_n \longrightarrow \underline{\theta}_0$ *a.s as* $n \longrightarrow \infty$.

3.2 Asymptotic normality

In the rest of the communication, we shall assume that the innovation process has normal distribution, i.e., $f(x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}$, $x \in \mathbb{R}$. On the other hand, since given a bath $s_t = x_t$, the Jacobean of the transformation from X_t to e_t is unity, then $g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right) = f\left(e_t\left(\underline{\theta}_{x_t}\right)|\underline{\mathcal{X}}_{1-p_0:t-1}\right)$ where $(e_t(\underline{\theta}), t \in \mathbb{Z})$ be the strict stationary process determined recursively in t as the solution of $e_t(\underline{\theta}) = X_t - a_0(s_t) - \sum_{i=1}^{p} a_i(s_t)X_{t-i} - \sum_{j=1}^{q}\sum_{i=j}^{p} c_{ij}(s_t)X_{t-i}e_{t-j}(\underline{\theta})$, so the likelihood function of $\underline{X}_{1:n}$ is the same as joint density function of $\underline{e}_{1:n}(\underline{\theta})$ summed over all possible path of the Markov chain and $L_n(\theta)$ is now a convex combination of $n-$multivariate Gaussian densities, i.e.,

$$L_n(\underline{\theta}) = \sum_{(x_1,\ldots,x_n)\in\mathbb{S}^n} \pi(x_1)\prod_{t=2}^{n} p_{x_{t-1},x_t}\prod_{t=1}^{n} f\left(e_{t+p_0-1}\left(\underline{\theta}_{x_t}\right)\right).$$

Remark 3.1 *The existence and the uniqueness of the process* $(e_t(\underline{\theta}), t \in \mathbb{Z})$ *is ensured by the invertibility of the model* (1.1) *in the sense that* $e_t \in \sigma((X_k, s_k), k \leq t)$. *Hence, the model* (1.1) *is invertible if the Lyapunov exponent* $\gamma_L(C)$ *associated with the sequence* $C = (C_t, t \in \mathbb{Z})$ *where* $C_t = [\beta_j(t)\delta_1(i) + \delta_i(j+1)]_{i,j=1,\ldots,q}$ *with* $\beta_j(t) = \sum_{i=j}^{p} c_{ij}(s_t)X_{t-i}$ *is such that* $\gamma_L(C) < 0$ *provided that there already exists a strictly stationary and ergodic process* $(X_t, t \in \mathbb{Z})$ *satisfying the Equation* (1.1).

To formulate the asymptotic normality of the parameter estimate, we have to introduce the gradient $\nabla_{\underline{\theta}}$ and Hessian $\nabla^2_{\underline{\theta}}$ operators with respect to parameter vector $\underline{\theta}$ and we added the following assumptions

A5. $\underline{\theta}_0 \in \overset{o}{\Theta}$ where $\overset{o}{\Theta}$ is the interior of Θ.

A6. $\gamma_L(C^0) < 0$ for all $\underline{\theta} \in \Theta$ where C^0 denotes the sequence $(C_t, t \in \mathbb{Z})$ when the parameters $\underline{\theta}_i$ are replaced by their true values $\underline{\theta}_i^0$, $i = 1, ..., d$.

A7. $E\{X_t^4\} < +\infty$

A8. $E\{e_t^4\} < +\infty$.

Lemma 3.4 *Under the conditions* ***A1-A8****, we have 1. The function* $\underline{\theta} \longrightarrow q_{\underline{\theta}}^*\left(X_t | \underline{\mathcal{X}}_{t-1}\right)$ *is of class* $\mathcal{C}^{(2)}$ *on* $\overset{o}{\Theta}$*, 2.* $\sup_{\underline{\theta}\in\Theta} \left\|\nabla_{\underline{\theta}} q_{\underline{\theta}}^*\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right\| < \infty$, $\sup_{\underline{\theta}\in\Theta} \left\|\nabla_{\underline{\theta}}^2 q_{\underline{\theta}}^*\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right\| < \infty.$*, 3.* $E\left\{\sup_{\underline{\theta}\in\Theta} \left\|\nabla_{\underline{\theta}} q_{\underline{\theta}}^*\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right\|\right\} < \infty$, $E\left\{\sup_{\underline{\theta}\in\Theta} \left\|\nabla_{\underline{\theta}}^2 q_{\underline{\theta}}^*\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right\|\right\} < \infty.$

Lemma 3.5 *Under the conditions* ***A1-A8****, there exist a functions* $h_i : \mathbb{R} \to \mathbb{R}_+$*,* $i = 0, 1, 2$ *satisfying* $E\{h_i(X_t)\} < +\infty$*,* $i = 0, 1, 2$ *such that* 1. $\sup_{\underline{\theta}} g_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right) \leq h_0(X_t)$, 2. $\left\|\nabla_{\underline{\theta}} g_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right\| \leq h_1(X_t)$*,* 3. $\left\|\nabla_{\underline{\theta}}^2 g_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right\| \leq h_2(X_t)$.

Lemma 3.6 *Let* $I_n(\underline{\theta})$ *be the covariance matrix of* $\frac{1}{\sqrt{n}} \nabla_{\underline{\theta}} \log L_n(\underline{\theta})$ *(*$I_n(\underline{\theta}_0)$ *plays the role of Fisher information matrix). Then* $I_n(\underline{\theta}) = \frac{1}{n} E_{\underline{\theta}_0}\left\{\frac{\partial \log L_n(\underline{\theta})}{\partial \underline{\theta}} \frac{\partial \log L_n(\underline{\theta})}{\partial \underline{\theta}'}\right\}$ $J_n(\underline{\theta}) = -\frac{1}{n} E_{\underline{\theta}_0}\left\{\frac{\partial^2 \log L_n(\underline{\theta})}{\partial \underline{\theta} \partial \underline{\theta}'}\right\} = Var_{\underline{\theta}_0}\left\{\nabla_{\underline{\theta}} \log q_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right\}$.

Lemma 3.7 *Under the conditions* ***A1-A8.,*** 1. $J(\underline{\theta}_0) = \lim_{n\to\infty} J_n(\underline{\theta}_0)$ *and* $I(\underline{\theta}_0) = \lim_{n\to\infty} I_n(\underline{\theta}_0)$ *exists and* 2. $J(\underline{\theta}_0)$ *is a positive definite matrix.*

Lemma 3.8 *Under* ***A1-A8****, almost surely* $\frac{1}{n} \nabla_{\underline{\theta}}^2 \log L_n(\underline{\theta}) \longrightarrow -J(\underline{\theta}_0)$.

Proposition 3.1 *Under conditions* ***A1-A8*** *we have 1.* $n^{-\frac{1}{2}} \nabla_{\underline{\theta}} L_n\left(\hat{\underline{\theta}}_n\right) \rightsquigarrow \mathcal{N}(0, I(\underline{\theta}_0))$ *and 2.*$\sqrt{n}\left(\hat{\underline{\theta}}_n - \underline{\theta}_0\right) \rightsquigarrow \mathcal{N}(0, \Sigma(\underline{\theta}_0))$ *where* $\Sigma(\underline{\theta}_0) = J^{-1}(\underline{\theta}_0) I(\underline{\theta}_0) J^{-1}(\underline{\theta}_0)$.

References

[1] Bougerol, P., Picard, N. (1992) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714 − 1730.

[2] Cappé O, E, Moulines, and T. Rydén (2005). Inference in Hidden Markov Models. Springer.

[3] Francq, C., M. Roussignol (1997) On white noises driven by hidden Markov chains. *J. of Time Ser. Anal.* 18, 553 − 578.

[4] Douc, R., Moulines, E., and Rydzn T. (2004). Asymptotic propertiesof the maximum likelihood estimator in autoregressive models with Markov regime. Ann. Statist., 32, 2254 − 2304.