REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTÈRE DE L'ENSEIGNEMENT SUPÈRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

Département de Mathématiques

Faculté des Sciences Exactes

Université des Frères Mentouri-Constantine

N◦ d'ordre: .. / .. / ..
N◦ de série: .. / .. / ..

THÈSE

Pour l'obtention du diplôme de

DOCTORAT EN SCIENCES

Présentée par

AHMED GHEZAL

Intitulé

STATISTIQUE ASYMPTOTIQUE DANS LES MODÈLES BILINÉAIRES À CHANGEMENT DE RÉGIME MARKOVIEN

Soutenue le / / 2015

Composition du jury:

Mr. F. L Rahmani	Professeur, Université des Frères Mentouri-Constantine	**Président**
Mr. A. Bibi	Professeur, Université des Frères Mentouri-Constantine	**Rapporteur**
Mr. A. Aknouche	Professeur, USTHB	**Examinateur**
Mr. T. Kernane	MCA, USTHB	**Examinateur**

على

بركة الله

Université des Frères Mentouri-Constantine

Faculté des Sciences Exactes

Département de Mathématiques

Statistique Asymptotique Dans Les Modèles Bilinéaires À Changement De Régime Markovien

Contents

Introduction

0.1 Motivations

Les séries temporelles sont considérées à tort comme étant une branche exclusive de l'économétrie. Cette dernière est une discipline qui est relativement jeune alors que les séries temporelles ont été utilisées bien avant, par exemple en astronomie (1906) et en météorologie (1968).

L'objet des séries temporelles est l'étude des variables au cours du temps. Même s'ils n'ont pas été à l'origine de cette discipline, ce sont les économètres qui ont assuré les grandes avancées qu'a connues cette discipline (beaucoup de « Prix Nobel » d'économie sont des économètres).

Parmi ses principaux objectifs figurent la détermination de tendances au sein de ces séries ainsi que la stabilité des valeurs (et de leur variation) au cours du temps. Citons par exemple: le volume des ventes hebdomadaires d'un produit, le prix des actions de la banque de clôture du jour, le volume de la production quotidienne de pétrole brut en Algérie, le taux de chômage dans une période connue, ...Ces applications nécessitent beaucoup de recherches et d'expériences. Les techniques standards d'analyse de séries temporelles ont longtemps reposé sur les propriétés fondamentales de linéarité et stationnarité. L'essor considérable qu'a connu l'analyse statistique des séries chronologiques au cours des ces derniers trois décennies est lié essentiellement au développement de l'approche temporelle sous deux hypothèses remarquables:

Stationnarité et linéarité.

L'analyse des séries chronologiques a été considérablement développée depuis la publication de l'ouvrage dev Box et Jenkins [1] (1970) qui a été décisive. En effet, dans l'ouvrage les deux auteurs développent le très populaire modèle $ARMA$ (Auto Regressive Moving Average). Les modèles ($ARMA$) ont ainsi fait l'objet d'un intérêt croissant sur une vaste étendue disciplinaire allant de l'économétrie et la finance à la climatologie ou l'électrotechnique. Cependant, de nombreuses recherches ont démontré que les hypothèses de linéarité n'étaient qu'un pis-aller utopique apportant un confort appréciable dans l'étude probabiliste et statistique du modèle, et cette classe de processus ($ARMA$) jouera le rôle prépondérant dans notre modélisation concrète des processus stationnaires, tandis que la classe encore plus large des processus linéaires caractérisés par leurs flexibilités, facilités d'utilisation est ses utilisation pour prédire les valeurs futures, ils sont facile à interpréter parmi la plupart des autres modèles. Cependant, les méthodes d'analyse et d'inférence statistique sont mieux développés dans cette classe de processus, et de plus elles peuvent être employées comme une analyse préliminaire. Habituellement, elles donnent une représentation parcimonieuse et interprétable.

Comme il a été mentionné par Hallin (1978), dans de nombreux cas l'hypothèse qui se trouve à la base de l'utilisation des méthodes de Box et Jenkins (dans le cas non stationnaire) peut escamoter les problèmes de la non stationnarité plutôt que de les résoudre. Cependant, et à partir des années 70, et dans le but de résoudre certains problèmes liés à la non stationnarité, on trouve un intérêt croissant

[1] George BOX et Gwilym JENKINS sont deux statisticiens qui ont contribué, dans les années (1970), à populariser la théorie des séries temporelles univariées. Les procédures de modélisation sont présentées dans leur célèbre ouvrage intitulé « Time Series Analysis: Forecasting and control ».

Ils ont proposé une démarche générale de prévision pour les séries chronologiques. Cette démarche est fondée sur la notion de processus ($ARMA$) et elle comprend quatre phases: l'identification a priori, l'estimation du modèle ($ARMA$) identifié, l'identification a posteriori et la prévision.

aux modèles dont les coefficients eux mêmes sont susceptibles d'évoluer avec le temps. On peut distinguer deux classes de modèles à coefficients dépendant du temps selon que cette évolution est de nature déterministe ou stochastique

Evolution déterministe Dans cette classe de modèles qui visent à décrire des processus de nature linéaires mais non stationnaires, que l'on conviendra d'appeler évolutifs, où les trajectoires des coefficients s'expriment comme des combinaisons linéaires de fonction du temps, supposées connues a priori et en nombre fini. D'une part et du point de vue technique, l'idée de faire une projection des trajectoires des coefficients sur une base de fonctions était déjà sous-jacente aux travaux de Mendel (1973) et Rao (1970). Dans ces travaux, la représentation évolutive apparait plutôt comme une "astuce" de calcul. Liporace (1975) a été le premier à étudier l'estimation d'un modèle autorégressif suivi par Hall, Oppenheim et Willsky (1983). Grenier (1986) a travaillé sur ce point et a proposé un jeu d'algorithmes rapides adaptés à plusieurs variantes de modèles. D'autre part, Cramer (1961) a étendu le résultat fondamental de Wold au cas non stationnaire (avec variance finie) donc nous parlerons désormais de la décomposition de Wold-Cramer, et les mêmes arguments; prévision et modélisation, peuvent être obtenus à partir de la décomposition de Wold-Cramer en utilisant des modèles $ARMA$ à coefficients dépendant du temps. Dans cette perspective, nous trouvons Mélard (1985) et Hallin (1986, 1989), et autres (cf. Priestley (1988)), ont enrichi la littérature des séries chronologiques avec de nombreux travaux fondamentaux.

Evolution stochastique C'est la classe des modèles qui a reçu jusqu'a présent le plus d'attention surtout dans les domaines de l'économétrie, l'automatique, le traitement du signal, et des séries chronologiques. Ces modèles (et comme dans le cas linéaire à coefficients constants) visent à décrire principalement des processus de nature stationnaires, mais non linéaires. La littérature sur ces modèles étant assez dispersée. La monographie de

Nicolls et Quinn (1982) et la bibliographie qu'elle contient font d'excellentes références dans ce domaine. Plusieurs articles, citons Andel (1976) et Andel (1982) font appel non seulement à l'erreur associée à la spécification du modèle, mais aussi à la nature même des rapports économiques pour justifier l'existence d'une composante non-déterministe dans les coefficients de modèles. Cette hypothèse nous semble aussi plus naturellement adaptée lors de l'analyse des séries chronologiques multiples (cf. Hannan (1970)).

Ainsi, face à une réalisation $\{X_1, X_2, \ldots X_n\}$ d'un processus $(X_t)_{t\in\mathbb{Z}}$, $\mathbb{Z} = \{0, \pm1, \pm2, ...\}$, un statisticien, un économètre cherchent à identifier le "bon" modèle. Donc, il est devant deux choix fondamentaux de modèles: modèles linéaires et modèles non linéaires, bien que la frontière entre les concepts de linéarité et de non linéarité soit difficile à concevoir, Rao et Gabr (1984) ont observés que la série décrivant le taux de chômage dans l'Allemagne de l'ouest a un comportement non linéaire et que les prévisions obtenues en utilisant des modèles bilinéaires sont mieux que celles obtenues par des modèles $ARMA$. Par conséquent, et durant ces deux dernières décennies, les modèles non linéaires ont reçus plus d'attention. Une classe particulière de modèles stationnaires et non linéaires qui est introduite dans la littérature de la théorie du contrôle, et qui a trouvé d'autres champs d'applications (cf. Mohler (1988)) est la classe des modèles bilinéaires. Cette classe de modèles, qui peut être regardée comme une extension des modèles $ARMA$, a été suffisamment étudiée par Granger et Andersen (1978), Pham et Tran (1981), Subba Rao (1981), Subba Rao et Gabr (1984), Guégan (1994), Shu-Ing (1985), Priestley (1988), Liu et Brockwell (1988), Liu (1989) , 1992, et Terdik (2000) et autres. Notons aussi qu'une sous classe particulière de modèles bilinéaires peut être utilisée comme des résidus dans la représentation $ARMA$ (cf. Francq (1999), Francq et Zakoian (2000)). Dans plusieurs situations pratiques, nous sommes devant des données générées par un certain processus non seulement non linéaire mais aussi non stationnaire. Kendall en 1953 a déjà mentionné que *"No economic system yet observed has been stationary over long periods [...]. It seems natural [to consider] the case* *when the constants [in the model] are themselves slowly moving through time as the economy changes".* Ceci suggère de considérer des modèles à coefficients dépendant du temps. Autres exemples de coefficients dépendant du temps qui ont un intérêt particulier sont les coefficients périodiques (pour des données saisonnières), ou les coefficients de rupture à un instant connu t_0 (pour une série qui subit un changement en un instant t_0). Il existe cependant des séries qui sont soupçonnées d'être à la fois non linéaires et non stationnaires, par exemple, en théorie de l'économie, la plupart des indices d'un stock d'une marchandise sont des différences de martingales (donc non nécessairement un processus $i.i.d$), pour de telles séries les techniques des modèles linéaires habituelles sont inapplicables. Lorsque l'économie change, il est difficile de justifier l'utilisation du même modèle (non linéaire) sur une longue période. Donc le recours aux modèles non linéaires à coefficients dépendant du temps nous semble raisonnable.

Motivés par la précédente discussion, nous allons abondonner l'hypothèse de linéarité, bien que ces modèles ont été relativement utilisés, nous allons néanmoins consacrer notre travail dans un cadre probabiliste et statistique à l'étude (bien entendu non exhaustive) d'une classe de processus $(X_t)_{t\in\mathbb{Z}}$ définis sur un espace de probabilité $(\Omega, \Im, P)$ non linéaires.

Un modèle générale et est donnée par

$$X_t = f_{s_t}\left(X_{t-i}, e_{t-j}, 0 < i \leq P, 0 < j \leq Q\right) + e_t$$

pour une fonction mesurable $f_{.}$ et un processus d'innovation $(e_t, t \in \mathbb{Z})$ supposé être indépendant et identiquement distribuées ($i.i.d$). Ainsi, certains localement (c'est à dire, dans chaque «régime») linéaire ou non-linéaire des modèles particuliers ont été étudiés afin de capturer les propriétés probabilistes et statistiques de ces modèles. Par exemple, $MS-ARMA$: Francq et Zakoïan [22], $MS-$non-linear $ARMA$: Lee [44], Yao et Attali [78], $MS-GARCH$: Francq et Zakoïan [21] et d'autres. C'est la classe de processus générés par des modèles bilinéaires à changements de régimes markoviens définis par l'équation aux différences stochastiques suivante:

$$X_t = \sum_{i=1}^{p} a_i(s_t)X_{t-i} + \sum_{i=1}^{q} b_i(s_t)e_{t-i} + \sum_{i=1}^{P}\sum_{j=1}^{Q} c_{ij}(s_t)X_{t-i}e_{t-j} + e_t \tag{0.1}$$

noté $MS-BL\,(p,q,P,Q)$ où $(a_i(s_t))_{1\leq i\leq p}\,,(b_i(s_t))_{1\leq i\leq q}\,,(c_{ij}(s_t))_{1\leq i\leq P,1\leq j\leq Q}$ sont des fonctions bornées, déterministes, dépendant éventuellement d'une chaîne de Markov à éspace d'état fini, i.e., $\mathbb{S}=\{1,...,d\}$ et où $(e_t)_{t\in\mathbb{Z}}$ est un processus de bruit blanc fort centré de variance finie et satisfaisant l'hypothèse suivante

$$\left\{e_t \text{ et } X_s \text{ sont indépendants pour tout } s<t\right\}.$$

La classe de modèles (0.1) contient trois sous classes

1. La classe de modèles $MS-ARMA\,(p,q)$ peut être obtenue en posant $c_{ij}(.)=0$ pour tout i,j, donc (0.1) est une extension naturelle des processus $MS-ARMA$.

2. La classe de modèles superdiagonaux obtenue en supposant $c_{ij}(.)=0$ for $i\leq j$.

3. La classe de modèles sous diagonaux obtenue en supposant $c_{ij}(.)=0$ for $i>j$.

Du fait de la dépendance non-linéaire entre X_t et e_{t-k}, $k\geq 1$ ceci rend délicat la manipulation des termes de type $X_te_{t-j}, j>0$. Pour cette raison, seule la classe des modèles super-diagonaux a reçu plus d'attention. Notons que, si dans le modèle général (0.1) les coefficients sont constants et le bruit blanc $(e_t)_{t\in\mathbb{Z}}$ est stationnaire, nous trouvons ainsi une littérature abondante. Cette abondance est dûe aux conditions sous lesquelles le modèle devient stationnaire et ergodique. Cependant de nombreux travaux de recherche existent sur le développement des propriétés probabilistes et statistiques, l'identification, les tests et l'estimation des paramètres de certains modèles bilinéaires stationnaires (pour une bibliographie récente voir Terdik (2000)). En revanche, lorsque le modèle est non-stationnaire les méthodes classiques sont inapplicables.

Certes l'étude des modèles bilinéaires à coefficients dépendant du temps est loin d'être achevée. De nombreux problèmes restent ouverts. Néanmoins on peut se demander s'il est possible de résoudre, par exemple, le problème de l'identification de certains modèles bilinéaires à coefficients dépendant du temps comme ce fut le cas pour certains modèles de séries chronologiques bilinéaires stationnaires, dans la mesure où la classe des modèles considérés est très riche et très complexe. Par contre la théorie des tests qui jusqu'à présent a été peu étudiée (dans le cas stationnaire) doit permettre d'aboutir assez rapidement à quelques résultats: outre les tests de stationnarité (respectivement de linéarité) pour lesquels quelques procédures ont été proposées sous l'hypothèse de la linéarité (respectivement sous l'hypothèse de stationnarité) on a besoin de tests portant sur le choix de la nature des coefficients, autrement dit le choix des modèles.

0.2 Apport et présentation de la thèse

Notre thèse intitulée "Statistique Asymptotique Dans Les Modèles Bilinéaires À Changement De Régime Markovien" se compose en quatre chapitres principaux:

Chapitre 1 : we used some algebraic notations

Dans ce chapitre, nous utilisons quelques notations algébriques.

Chapitre 2 : Probabilistic proprieties of $MS-BL$ processes

Ce chapitre présente une représentation vectorielle qui est utilisé pour obtenir des conditions suffisantes pour le processus $MS-BL$ généré par l'équation $(II-1.1)$ de la stationarité (au sens fort et faible), la causalité, l'ergodicité et l'existence des moments d'orders supérieurs. Aussi, les conditions sont nécessaires pour certains cas particuliers. La structure de $\mathbb{L}_2$ est analysé et la fonction de covariance est obtenu qui nous permet de donner une représentation $ARMA$. Nous avons aussi discuté la structure de $\mathbb{L}_m$ à partir de $\mathbb{L}_2$ et demontré que le processus de puissance $(X_t^m, t \in \mathbb{Z})$ admet aussi une représentation $ARMA$ et donner quelques exemples illustratifs. Nous donnons aussi des conditions suffisantes qui garantissent l'inversibilité. Nous proposons des conditions garantissant l'ergodicité géometrique et $\beta-$mélange. Et nous avons fourni certaines applications à une famille de $MS-GARCH\,(1,1)$.

Chapitre 3 : QMV approach for $MS-BL$ models

Dans ce chapitre, étudier la consistance forte de l'estimateur du quasi-maximum de vraisemblance (QML) dans les modèles $MS-BL$.

Chapitre 4 : GMM approach for $MS-BL$ models

Ce chapitre étudier la consistance forte et les propriétés de la normalité asymptotique de l'estimateur de distance minimale (MDE). Un ensemble d'expériences numériques illustre l'importance pratique de nos résultats théoriques.

Nous terminons notre thèse par un chapitre additif comportant une conclusion générale, des remarques, quelques perspectives et nos occupations futures.

Chapter 1

Algebraic notations

Some notations are used throughout the thesis: $I_{(n)}$ is the $n \times n$ identity matrix and $\mathbb{I}'_{(n)} := (\underbrace{I_{(n)} \vdots ... \vdots I_{(n)}}_{d-block})$ is the $n \times nd$ matrix and $\mathbb{I}_{\Delta}$ denotes the indicator function of the set Δ. $p \lim\limits_{n \to \infty}$ signify the convergence in probability. $O_{(k,l)}$ denotes the matrix of order $k \times l$ whose entries are zeros, for simplicity we set $O_{(k)} := O_{(k,k)}$ and $\underline{O}_{(k)} := O_{(k,1)}$, $\widetilde{\underline{M}} := Vec(M)$ is the vector obtained from a matrix $M := (m_{ij})$ by setting down the column of M underneath each other, the spectral radius of squared matrix M is denoted by $\rho(M)$. Let $\|.\|$ denote any induced operator norm on the set of $m \times n$ and $m \times 1$ matrices, and for $\gamma \in]0,1]$, let $|M|^{\gamma} := (|m_{ij}|^{\gamma})$, then it is easy to see that $|.|^{\gamma}$ is submultiplicative, i.e., $|M_1 M_2|^{\gamma} \leq |M_1|^{\gamma} |M_2|^{\gamma}$, $|M\underline{X}|^{\gamma} \leq |M|^{\gamma} |\underline{X}|^{\gamma}$ for any appropriate vector $\underline{X}$ and also subadditive, i.e., $|\sum_i M_i|^{\gamma} \leq \sum_i |M_i|^{\gamma}$ where the inequality $M \leq N$ denotes the elementwise relation $m_{ij} \leq n_{ij}$ for all i and j. $\otimes$ is the usual Kronecker product of matrices and $M^{\otimes r} = M \otimes M \otimes ... \otimes M$, $r-$times. If $(M(i), i \in \mathbb{N})$ is $n \times n$ matrices sequence, we shall denote for any integers l and j, $\prod\limits_{i=l}^{j} M(i) = M(l)M(l+1)\ldots M(j)$ if $l \leq j$ and $I_{(n)}$ otherwise. When the matrices $\underline{M} = (M(i), i \in \mathbb{S})$ is a sequence of non random matrices, we shall denote

$$\mathbb{P}(\underline{M}) = \begin{pmatrix} p_{11}M(1), & \ldots, & p_{d1}M(1) \\ \vdots & \vdots & \vdots \\ p_{1d}M(d) & \ldots & p_{dd}M(d) \end{pmatrix}, \underline{\Pi}(\underline{M}) = \begin{pmatrix} \pi(1)M(1) \\ \vdots \\ \pi(d)M(d) \end{pmatrix}.$$

Chapter 2

Probabilistic proprieties of $MS - BL$ processes

Abstract: This chapter investigates some probabilistic properties and statistical applications of general Markov$-$switching bilinear processes ($MS - BL$) that offers remarkably rich dynamics and complex behavior to model non Gaussian data with structural changes. In these models, the parameters are allowed to depend on unobservable time$-$homogeneous and stationary Markov chain with finite state space. So, some basic issues concerning this class of models including necessary and sufficient conditions ensuring the existence of ergodic stationary (in some sense) solutions, existence of finite moments of any order and $\beta-$mixing are studied. As a consequence, we observe that the local stationarity of the underlying process is neither sufficient nor necessary to obtain the global stationarity. Also, the covariance functions of the process and its power are evaluated and it is shown that the second (resp. higher)$-$order structure is similar to a some linear processes, and hence admit $ARMA$ representation. We have also sufficient conditions ensuring the invertibility are studied. We establish also sufficient conditions for the $MS - BL$ model to be $\beta-$mixing and geometrically ergodic. We then use these results to give sufficient conditions for $\beta-$mixing of a family of $MS - GARCH(1,1)$ processes. A number of illustrative examples are given to clarify the theory and the variety of applications.[1]

[1]This chapter is published in the Journal of Stochastics: An International Journal of Probability and Stochastic Processes.

2.1 Introduction

Since the seminal works by Hamilton [36], Markov−switching models (MSM) have received a growing interest and becomes a powerful tool of modelling and describing asymmetric business cycles (as originally proposed by Hamilton [36]) and continue to gain a more popularity especially in financial data. This is due to its higher flexibility in capturing the persistence and/or the asymmetric effects on the shocks of volatility and their ability to model time series which are characterized by some features including excess kurtosis, asymmetry, turning or sudden change in regime.

A discrete-time MSM is a bivariate random process $((X_t, s_t), t \in \mathbb{Z})$, $\mathbb{Z} := \{0, \pm 1, \pm 2, ...\}$ such that (i): $(s_t, t \in \mathbb{Z})$ is an unobservable (referred henceforth as "regime"), finite state space, discrete−time and homogeneous Markov chain and (ii): the conditional distribution of X_k given $\{(X_{t-1}, s_t), t \leq k\}$ depends on $\{(X_{t-1}, s_k), t \leq k\}$ only. So, the changes in regimes can be abrupt, and they occur frequently or occasionally depending on the transition probability of the chain. However, some locally (i.e., in each "regime") linear or nonlinear models were investigated in order to capture the probabilistic and statistical properties of such models. For instance, $MS - ARMA$: Francq and Zakoïan (hereafter FZ) [22] and Stelzer [68], $MS-$nonlinear $ARMA$ and bilinear processes: Lee [44], Yao and Attali [78] and Bibi and Aknouche [7], $MS - GARCH$: FZ [21], Hass et *al.,* [30], Liu [55] among others.

In this chapter, we alternatively propose a $MS-$bilinear model in which the process follows locally from a bilinear representation. This is in order to cover many commonly used models in the literature, and to give a general flexible and parsimonious framework for data sets exhibiting occasional sharp spikes or involving at certain points a high amplitude oscillations and which cannot sufficiently explained by the theory of standard linear models. In this context, we say that a $\mathbb{R}$-valued process $(X_t, t \in \mathbb{Z})$ defined on some probability space $(\Omega, \Im, P)$ has a general Markov−switching bilinear representation (denoted by $MS - BL(p, q, P, Q)$) if it is a solution of the following stochastic difference equation

$$X_t = \sum_{i=1}^{p} a_i(s_t)X_{t-i} + \sum_{j=1}^{Q}\sum_{i=1}^{P} c_{ij}(s_t)X_{t-i}e_{t-j} + \sum_{j=0}^{q} b_j(s_t)e_{t-j}, \ t \in \mathbb{Z}. \quad \text{(II-1.1)}$$

In $(II - 1.1)$, $(e_t, t \in \mathbb{Z})$ is an independent and identically distributed ($i.i.d.$) sequence of random variables defined on the same probability space $(\Omega, \Im, P)$

with $E\left\{\log^{+}|e_t|\right\}<+\infty$ where $\log^{+}x=\max\{0,\log x\}$, $x>0$. The functions $a_i(s_t)$, $b_j(s_t)$ and $c_{ij}(s_t)$ depend upon an unobserved Markov chain $(s_t,t\in\mathbb{Z})$ subject to the following assumption:

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

In literature of non$-$linear models, the bilinear models have been receiving an increasing interest and were successfully applied for analyzing non-Gaussian data and can be proposed to model financial returns and other complex data set. Indeed, Subba Rao and Gaber [72] have used a bilinear model to fit and forecast the West German monthly unemployment data and showed that the obtained results are "better" than many linear $ARMA$ alternatives. Peel and Davidson [62] propose a bilinear errors correction mechanism and suggested its application for the models that displays "abrupt changes". Maravall [56] consider bilinear models for forecasting nonlinear processes and demonstrated its improvement over $ARMA$ forecasts. Recently Aknouche and Rabehi [2] propose a mixture bilinear model in order to absorb several features exhibited by Canadian Lynx data and IBM stock prices and show the importance of mixture in bilinear representation.

The $MS-BL\,(p,q,P,Q)$ model encompass many commonly used models in the literature, indeed,

(i) Standard $BL\,(p,q,P,Q)$ models: These models are obtained by assuming constant the functions $a_i(.)$, $b_j(.)$ and $c_{i,j}(.)$ in $(II-1.1)$ or equivalently by assuming that the chain (s_t) has a single regime (e.g., Granger and Anderson [25]).

(ii) Hidden$-$Markov models (HMM): In contrast with MSM, HMM are characterized by the fact that given s_t, $(X_t,t\in\mathbb{Z})$ is a sequence of independent random variables. This class is obtained by setting $X_t=\;b_0\,(s_t)\,e_t$, i.e., $a_i(.)=b_j(.)=c_{i,j}(.)=0$ for all i, j in $(II-1.1)$ except that $b_0\,(s_t)\neq0$ (e.g., Francq and Roussignol [24]).

(iii) Markov$-$switching $ARMA$ models ($MS-ARMA$): These models are obtained by setting $c_{i,j}(.) = 0$ for all i and j in $(II-1.1)$ (e.g., FZ [22] and Stelzer [68]).

(iv) Some classes of $MS-(G)\,ARCH\,(p,q)$: (e.g., Abramson and Cohen [1], FZ [21] and Liu [55]). (see also Kristensen [43] for the building of $GARCH\,(p,q)$ models as special case of $BL\,(p,q,P,Q)$).

(v) Independent-switching $BL\,(p,q,P,Q)$: In this specification, analyzed by Aknouche and Rabehi [2] in the bilinear models, $(s_t, t \in \mathbb{Z})$ is an $i.i.d.$ process.

(vi) Some classes of periodic models ($PARMA$, PBL, $PGARCH$,...) (e.g., Bibi and Lessak [9]): These models can be obtained by rewording and/or drooping some hypothesis in Assumption 2.1.

The main aim of the chapter is to investigate some theoretical properties for $MS-BL$ processes, and thus for the foregoing models. So, and for the statistical purpose, it is often desirable in practice that the solutions processes $(X_t, t \in \mathbb{Z})$ for $(II-1.1)$ should be stationary, ergodic and satisfy $X_t = f\,(e_t, s_t, e_{t-1}, s_{t-1}, ...)$ almost surely $(a.s.)$ where f is a measurable function from $\mathbb{R}^\infty$ to $\mathbb{R}$. Such solutions are called causal. The mentioned properties were studied recently for the $MS-ARMA$ by FZ [22], Stelzer [68], and by Lee [44]. For the $MS-GARCH$ by FZ [21], Liu [55] and Abramson and Cohen [1]. However, and beside the properties listed in our Note [7], we continue to investigate others properties with different approach of $MS-BL\,(p,q,P,Q)$.

2.2 Stationarity and existence of higher−order moments of $MS-BL$

In what follows, we shall assume, without loss of generality, that in $(II-1.1)$ $P = p$ since otherwise zeros of $a_i(.)$ or $c_{ij}(.)$ can be filled in. Define the $r = (p+q)$ −state vector $\underline{X}_t := (X_t, ..., X_{t-p+1}, e_t, ..., e_{t-q+1}\ )'$, the $r \times r$−matrices $(A_j\,(s_t)\,, 0 \le j \le Q)$

$$A_0(s_t) = \begin{pmatrix} M_0(s_t) & B\,(s_t) \\ O_{(q,p)} & J \end{pmatrix},$$

$$A_j(s_t) = \begin{pmatrix} M_j(s_t) & O_{(p,q)} \\ O_{(q,p)} & O_{(q)} \end{pmatrix}, 1 \le j \le Q$$

and the vectors $\underline{H} := (1, \underline{O}'_{(r-1)})'$, $\underline{b}_0(s_t) = \left(b_0(s_t), \underline{O}'_{(p-1)}\right)'$ and
$\underline{b}(s_t) = \left(\underline{b}'_0(s_t), 1, \underline{O}'_{(q-1)}\right)'$ where

$$\begin{aligned}
B(s_t) &= \begin{pmatrix} b_1(s_t) & \dots & b_q(s_t) \\ & O_{(p-1,q)} & \end{pmatrix}_{p\times q} \\
J &= \begin{pmatrix} \underline{O}'_{(q-1)} & 0 \\ I_{(q-1)} & \underline{O}_{(q-1)} \end{pmatrix}_{q\times q}, \\
M_0(s_t) &= \begin{pmatrix} a_1(s_t) & \dots & \dots & a_p(s_t) \\ 1 & \ddots & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}_{p\times p}, \\
M_j(s_t) &= \begin{pmatrix} c_{1j}(s_t) & \dots & c_{pj}(s_t) \\ & O_{(p-1,p)} & \end{pmatrix}_{p\times p}, j = 1, ..., Q.
\end{aligned}$$

So the Equation $(II-1.1)$ can be expressed in the following state−space representation i.e., for all $t \in \mathbb{Z}$, $X_t = \underline{H}'\underline{X}_t$ and

$$\underline{X}_t = A_t \underline{X}_{t-1} + e_t \underline{b}(s_t), \; a.s. \tag{II-2.1}$$

where $A_t = A_0(s_t) + \sum_{j=1}^{Q} A_j(s_t) e_{t-j}$. In the subdiagonal model $(MS-SBL(p,q,p,Q))$ for which $c_{ij}(.) = 0$ in $(II-1.1)$ for $i < j$ the following representation is considered

$$\underline{X}_t = A_0(s_t)\underline{X}_{t-1} + \sum_{j=1}^{Q} A_j(s_t)\underline{X}_{t-j} e_{t-j} + e_t \underline{b}(s_t), \; t \in \mathbb{Z} \tag{II-2.2}$$

in which $(M_j(s_t), 1 \leq j \leq Q)$ are the same as above except that their first lines are $(\underbrace{c_{jj}(s_t) \;\; \dots \;\; c_{pj}(s_t),}_{p-j+1} \underline{O}'_{(j-1)})$.

2.2.1 Strict stationarity

Since $((s_t, e_t), t \in \mathbb{Z})$ is a stationary and ergodic Markov process with state space $\mathbb{S} \times \mathbb{R}$, $((A_t, \underline{b}(s_t)e_t), t \in \mathbb{Z})$ is also a stationary and ergodic process and due to the finiteness of the state space of $(s_t, t \in \mathbb{Z})$, then $E\left\{\log^+ \|\underline{b}(s_t)e_t\|\right\}$ and $E\left\{\log^+ \|A_t\|\right\}$ are finite. By setting $\mathbf{A}_t(k) = \prod_{j=0}^{k-1} A_{t-j}$, it follows from Brandt [15] (see Bougerol and Picard [13]), that the unique, causal, bounded in probability, strictly stationary and ergodic solution of $(II-2.1)$ is given by

$$\underline{X}_t = \sum_{k=1}^{\infty} \mathbf{A}_t(k)\underline{b}(s_{t-k})e_{t-k} + \underline{b}(s_t)e_t \tag{II-2.3}$$

whenever the Lyapunov exponent $\gamma_L(A)$ associated with the sequence of random matrices $A = (A_t, t \in \mathbb{Z})$ defined by $\gamma_L(A) := \inf_{t>0} E\left\{\frac{1}{t}\log\|\mathbf{A}_t(t)\|\right\} \overset{a.s}{=} \lim_{t\to\infty}\left\{\frac{1}{t}\log\|\mathbf{A}_t(t)\|\right\}$ is strictly negative. Obviously, any strictly stationary solution $(X_t, t \in \mathbb{Z})$ of $(II-1.1)$ leads to a strictly stationary of $(II-2.1)$ via the above transformation. Conversely, we can see that, $\underline{H}'\underline{X}_t$ constitutes the unique solution of $(II-1.1)$ having the same properties as $\underline{X}_t$.

The needness of the condition $\gamma_L(A) < 0$ for the existence of the strictly stationary solution of $(II-2.1)$ was established by Bibi and Aknouche [7] under *controllability* concept. Recalling that the Representation $(II-2.1)$ or, equivalently the sequence $\{\underline{b}(s_t), A_j(s_t), 0 \leq j \leq Q\}$ is said to be controllable if the matrices $\mathcal{C}(s_t) := \left[c_1(s_t) \vdots c_2(s_t) \vdots \ldots \vdots c_r(s_t)\right]$ have almost surely a rank equal to r where the matrices $(c_j(s_t), 1 \leq j \leq r)$ are defined recursively by: $c_1(s_t) = \underline{b}(s_t)$ and $c_j(s_t) = \left[A_0(s_t)c_{j-1}(s_{t-j+1}) \vdots \ldots \vdots A_Q(s_t)c_{j-1}(s_{t-j+1})\right]$ for $j \geq 2$. The following theorem gives us the main result for stochastic difference equation $(II-2.1)$.

Theorem 1 *Consider the model $(II-1.1)$ with state space representation $(II-2.1)$. Then*

1. *$\gamma_L(A) < 0$ is a sufficient condition for $(II-2.1)$ to have a unique, strictly stationary, causal and ergodic solution, given by the Series $(II-2.3)$ which converges absolutely almost surely for all $t \in \mathbb{Z}$.*

2. *Conversely, suppose that $\{\underline{b}(s_t), A_j(s_t), 0 \leq j \leq Q\}$ is controllable and $(II-2.1)$ has a strictly stationary solution then $\gamma_L(A) < 0$.*

Proof. By the subadditive ergodic theorem (see Kingman [41]), almost surely, $\limsup_k \|\mathbf{A}_t(k)\|^{1/k} \leq \exp\{\gamma_L(A)\} < 1$. On the other hand, by the moment lemma (a.k.a. Markov's inequality) and the Borel-Cantelli lemma we have $P\left(\limsup_{k\to+\infty} |e_{t-k}|^{\frac{1}{k}} > e\right) = 0$ for all $e > 1$. So,

$$\limsup_{k\to+\infty} \|\mathbf{A}_t(k)\underline{b}(s_{t-k})e_{t-k}\|^{\frac{1}{k}} \leq \exp\{\gamma_L(A)\} < 1$$

and by the Cauchy's root test the series $(II - 2.3)$ converges absolutely almost surely. To prove the second assertion, we observe that if there exists a strictly stationary solution for $(II - 2.1)$, then $\lim_{t\to\infty} \|\mathbf{A}_t(t)\underline{b}(s_t)\| = 0$ in probability. By controllability, we obtain in probability $\lim_{t\to\infty} \|\mathbf{A}_t(t)\| = 0$. By simple modification of Lemma 3.4 in Bougerol and Picard [13] we deduce that $\gamma_L(A) < 0$. ■

Using the same arguments used by FZ [22], it is straightforward to see that $\gamma_L(A) = \gamma_L(M)$ where $\gamma_L(M)$ is the Lyapunov exponent associated with the random matrices $M = (M_t, t \in \mathbb{Z})$ with $M_t = M_0(s_t) + \sum_{j=1}^{Q} M_j(s_t)e_{t-j}$. This means that $\gamma_L(.)$ is independent of the moving average part. Hence

Corollary 2 *Consider the model $(II - 1.1)$ with state space representation $(II - 2.1)$. Then, we have*

1. $\gamma_L(M) < 0$ *is a sufficient condition for $(II - 2.1)$ to have a unique, strictly stationary, causal and ergodic solution, given by the Series $(II - 2.3)$ which converges absolutely almost surely for all $t \in \mathbb{Z}$.*

2. *Conversely, suppose that $\{\underline{b}_0(s_t), M_j(s_t), 0 \leq j \leq Q\}$ is controllable and $(II - 2.1)$ has a strictly stationary solution, then $\gamma_L(M) = \gamma_L(A) < 0$.*

Corollary 3 *Consider the $MS - BL(p, q, p, Q)$ process $(II - 1.1)$ with state space representation $(II - 2.1)$. If $E\left\{\log\left\|\prod_{j=0}^{p-1} M_{t-j}\right\|\right\} < 0$ then Equation $(II - 1.1)$ has a unique, strictly stationary and ergodic solution given by the series $(II - 2.3)$ whose first component $(X_t, t \in \mathbb{Z})$ is a strictly stationary, causal and ergodic solution for $(II - 1.1)$.*

Proof. The proof follows upon observing that

$$\gamma_L(A) = \gamma_L(M) \leq E\left\{\log\left\|\prod_{j=0}^{p-1} M_{t-j}\right\|\right\}.$$

■

Example 4 [*Non-necessity of local stationarity*] *Consider the first−order $MS-BL$ model, i.e.,*

$$X_t = (a_1(s_t) + c_{11}(s_t)e_{t-1})X_{t-1} + b_0(s_t)e_t.$$

The sufficient condition is $\gamma_L(A) = \sum_{i=1}^{d} \pi(i)E\{\log|a_1(i) + c_{11}(i)e_0|\} < 0$. On the other hand, if $b_0(i) \neq 0$ for all $i \in \mathbb{S}$, then $\gamma_L(A) < 0$ is also necessary *since* $rg\{c_1(s_t)\} = 1$*, otherwise,* $X_t = 0$ *is the unique strictly stationary solution without any constraint on* $\gamma_L(A)$*. This example shows the importance of the presence of moving average part for the necessity condition. It shows also that the local strict stationarity is not necessary, i.e., the existence of explosive regimes (i.e.,* $E\{\log|a_1(i) + c_{11}(i)e_0|\} > 0$*) does not preclude global strict stationarity.*

It is worth noting that the condition involving the strict stationarity for the first−order $MS - BL$ model depends only on the initial distribution of the chain $(s_t, t \in \mathbb{Z})$. This turns out not to be true in higher−order $MS - BL$ models as showed in the following example

Example 5 *Consider the following* $MS - AR(2)$ *model with two regimes*

$$X_t = a_1 X_{t-1}\mathbb{I}_{\{s_t=1\}} + (a_2 X_{t-1} + b_2 X_{t-2})\,\mathbb{I}_{\{s_t=2\}} + e_t$$

with $a_1 \neq 0$ *and state space representation* $\underline{X}_t = B_t \underline{X}_{t-1} + \underline{e}_t$ *where* $B_t = B_0\mathbb{I}_{\{s_t=1\}} + B_1\mathbb{I}_{\{s_t=2\}}$ *in which*

$$B_0 = \begin{pmatrix} a_1 & 0 \\ 1 & 0 \end{pmatrix}, B_1 = \begin{pmatrix} a_2 & b_2 \\ 1 & 0 \end{pmatrix}.$$

Since the matrix B_0 *is singular, then there exists an invertible matrix* Q *(can be obtained using a change of basis) such that* QB_0Q^{-1} *can be written as* $\widetilde{B}_0 = \begin{pmatrix} a_1 & 0 \\ 0 & 0 \end{pmatrix}$*. So, we consider the model*

$$Q\underline{X}_t = \underline{\widetilde{X}}_t = \widetilde{B}_t \underline{\widetilde{X}}_{t-1} + \underline{\widetilde{e}}_t$$

where $\underline{\widetilde{e}}_t = Q\underline{e}_t$ *and* $\widetilde{B}_t = QB_tQ^{-1} = \widetilde{B}_0\mathbb{I}_{\{s_t=1\}} + \widetilde{B}_1\mathbb{I}_{\{s_t=2\}}$ *with* $\widetilde{B}_1 = QB_1Q^{-1}$*. Since the Lyapunov exponent associated with* $\left(\widetilde{B}_t, t \in \mathbb{Z}\right)$ *and with* $(B_t, t \in \mathbb{Z})$ *is the same, then by writing* $\widetilde{B}_1^n = \begin{pmatrix} b_{11}(n) & b_{12}(n) \\ b_{21}(n) & b_{22}(n) \end{pmatrix}$*, then a result due to Lima an Rahibe [46], says that the exact Lyapunov exponent is given by*

$$\gamma_L(B) = \gamma_L\left(\widetilde{B}\right) = \frac{p_{21}}{p_{21} + p_{12}} \log|a_1| + \pi(1)\, p_{12}p_{21} \sum_{n=1}^{\infty} p_{22}^{n-1} \log|b_{11}(n)|,$$

showing that the Lyapunov exponent depends on the initial distribution of the chain $(s_t, t \in \mathbb{Z})$ *and their transition probabilities.*

Example 6 *[Non-sufficiency of local stationarity] Consider the model*

$$X_t = \begin{cases} X_{t-1} + c(1)X_{t-1}e_{t-1} + e_{t-1} \text{ if } s_t = 1 \\ X_{t-1} + c(2)X_{t-1}e_{t-1} + e_{t-1} \text{ if } s_t = 2. \end{cases}$$

First note that for each $i = 1, 2$, if $X_t = -\dfrac{1}{c(i)}$ for some t, then $X_{t+n} = -\dfrac{1}{c(i)}$ for all $n \geq 0$. On the other hand if $x \neq -\dfrac{1}{c(i)}$, then $P\left(x + (c(i)x + 1) = -\dfrac{1}{c(i)}\right) = 0$. So, $X_t = -\dfrac{1}{c(i)}$ is always the unique stationary solution for each regime regarding $\gamma_L^{(i)}(A) = \pi(i)E\{\log|1 + c(i)e_0|\} < 0$, $i = 1, 2$, or not and the conclusion follows.

Though the condition $\gamma_L(A) < 0$ could be used as a sufficient condition for the strict stationarity, it is of little use in practice since this condition involves the limit of products of infinitely many random matrices. Hence, some simple sufficient conditions ensuring the negativity of $\gamma_L(A)$ can be given.

Proposition 7 *Consider the $MS - BL(p, q, p, 1)$ model in which $E\left\{|e_0|^\delta\right\} < \infty$ for some $0 < \delta \leq 1$. Let $\underline{M}_\delta := \left(E\left\{|M_0(i) + M_1(i)e_0|^\delta\right\}, 1 \leq i \leq d\right)$. Then $\rho(\mathbb{P}(\underline{M}_\delta)) < 1$ implies that $\gamma_L(M) < 0$ and hence the statement of the first assertion of Theorem 1 hold.*

Proof. Because the Lyapunov exponent is independent of the norm, by choosing an absolute norm, i.e., a norm $\|.\|$ such that $\|N\|^\delta \leq \left\||N|^\delta\right\|$ (e.g., $\|N\| = \sum_{i,j}|n_{ij}|$). Therefore, since $\rho(\mathbb{P}(\underline{M}_\delta)) < 1$, there exists $0 < \lambda < 1$ such that $\limsup_t \|\mathbb{P}^t(\underline{M}_\delta)\|^{1/t} < \lambda$. By Jensen inequality and submultiplicativity of the operator $|.|^\delta$ we obtain

$$\begin{aligned} \gamma_L(M)\delta &= \lim_{t\to\infty} \frac{1}{t} E\left\{\log\left\|\prod_{j=0}^{t-1}(M_0(s_{t-j}) + M_1(s_{t-j})e_{t-j-1})\right\|^\delta\right\} \\ &\leq \lim_{t\to\infty} \frac{1}{t} \log E\left\{\left\|\prod_{j=0}^{t-1}(M_0(s_{t-j}) + M_1(s_{t-j})e_{t-j-1})\right\|^\delta\right\} \\ &\leq \lim_{t\to\infty} \frac{1}{t} \log E\left\{\left\|\prod_{j=0}^{t-1}|M_0(s_{t-j}) + M_1(s_{t-j})e_{t-j-1}|^\delta\right\|\right\} \\ &\leq \limsup_{t\to\infty} \log\left\|\mathbb{P}^t(\underline{M}_\delta)\right\|^{1/t} < 0, \end{aligned}$$

and the result follows from the Corollary 2. ∎

Corollary 8 *$[MS - ARMA]$ In the $MS-ARMA(p,q)$ model, i.e., when $M_1(s_t) = O_{(p)}$ in above proposition, then the sufficient condition reduces to $\rho\left(\mathbb{P}\left(\underline{M}_0\right)\right) < 1$ where $\underline{M}_0 = \left(\left|M_0(i)\right|, i = 1, ..., d\right)$.*

Proof. Straightforward and hence omitted. ∎

The Lyapunov exponent criterion γ_L seems difficult to obtain explicitly when $p > 1$. However, a potential method to verify whether or not $\gamma_L < 0$ is via Monte-Carlo simulation using Equation $(II - 2.1)$. This fact heavily limits the interests of the criterion in statistical applications which often suggests that the solution process must have some moments not ensured by the condition $\gamma_L < 0$. So, we need to search for conditions ensuring the existence of moments for the strict stationary solutions.

2.2.2 Second-order stationarity

The problem of finding conditions ensuring the existence of second$-$order stationarity solutions for weak $MS - ARMA(p,q)$ has been addressed recently in a series papers by FZ ([22], [21] and the references therein). Some results on the existence of second$-$order moments for strictly stationary $MS - BL$ processes have been obtained by Bibi and Aknouche [7] for some restrictive models. However, and since, it is difficult to handle in $(II - 1.1)$ the product terms, like $X_t e_{t-k}$, $k > 0$, because of the nonlinear dependence between X_t and e_{t-k}, $k > 0$, we shall restrict ourselves to the subdiagonal model $(II - 2.2)$ which can be easily transformed into a Markovian state$-$space representation. Indeed, let $s = r(Q+1)$ and define the $s-$vectors $\underline{Y}_t := (\underline{X}'_t, \underline{X}'_t e_t, ..., \underline{X}'_{t-Q+1} e_{t-Q+1})'$, $\underline{\eta}_{s_t}(e_t) := \left(\underline{b}'(s_t)e_t, \underline{b}'(s_t)e_t^2, \underline{O}'_{(r(Q-1))}\right)'$ and the $s \times s-$matrix

$$\Gamma_0\left(s_t\right) = \begin{pmatrix} A_0(s_t) & A_1(s_t) & A_2(s_t) & \dots & A_Q(s_t) \\ O_{(r)} & O_{(r)} & O_{(r)} & \dots & O_{(r)} \\ O_{(r)} & I_{(r)} & O_{(r)} & \dots & O_{(r)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ O_{(r)} & \dots & \dots & I_{(r)} & O_{(r)} \end{pmatrix},$$

$$\Gamma_1\left(s_t\right) = \begin{pmatrix} O_{(r)} & O_{(r)} & \dots & \dots & O_{(r)} \\ A_0(s_t) & A_1(s_t) & \dots & \dots & A_Q(s_t) \\ O_{(r)} & O_{(r)} & \dots & \dots & O_{(r)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ O_{(r)} & \dots & \dots & \dots & O_{(r)} \end{pmatrix},$$

then $(II-2.2)$ admits the following state−space representation $\underline{X}_t = F'\underline{Y}_t$ and

$$\underline{Y}_t = \Gamma_{s_t}(e_t)\underline{Y}_{t-1} + \underline{\eta}_{s_t}(e_t) \tag{II-2.4}$$

where $\Gamma_{s_t}(e_t) = \Gamma_0(s_t) + e_t\Gamma_1(s_t)$ and $F' = \left(I_{(r)} \vdots O_{(r)} \vdots ... \vdots O_{(r)}\right)$ is a $r \times s$ matrix. The advantage of the representation $(II-2.4)$ is that given $s_t = i$, the state vector $\underline{Y}_t$ is independent of $\underline{\eta}_{s_k=i}(e_k)$ for any $k > t$ and that the extended process $\left((\underline{Y}'_t, s_t)', t \in \mathbb{Z}\right)$ is a Markov chain on $\mathbb{R}^s \times \mathbb{S}$.

Theorem 9 *Consider the $MS-SBL$ process $(II-2.2)$ with state−space representation $(II-2.4)$ and assume that $\kappa_4 := E\{e_t^4\} < +\infty$.*

Let $\underline{\Gamma}^{(2)} := \left(\Gamma^{(2)}(i) = \Gamma_0^{\otimes 2}(i) + \sigma^2\Gamma_1^{\otimes 2}(i), 1 \leq i \leq d\right)$. If

$$\lambda_{(2)} := \rho\left(\mathbb{P}(\underline{\Gamma}^{(2)})\right) < 1, \tag{II-2.5}$$

then equation $(II-2.4)$ has a unique, causal, ergodic and strictly stationary solution given by

$$\underline{Y}_t = \sum_{k=1}^{\infty}\left\{\prod_{i=0}^{k-1}\Gamma_{s_{t-i}}(e_{t-i})\right\}\underline{\eta}_{s_{t-k}}(e_{t-k}) + \underline{\eta}_{s_t}(e_t) \tag{II-2.6}$$

with the above series converging absolutely almost surely and in $\mathbb{L}_2$.

Remark 10 *Using the same arguments used by FZ [22], it is straightforward to see that*

$$\lambda_{(2)} := \rho\left(\mathbb{P}(\underline{\widetilde{\Gamma}}^{(2)})\right) < 1, \tag{II-2.7}$$

where $\underline{\widetilde{\Gamma}}$ is the matrix obtained by replacing the matrices $A_j(.)$ by $M_j(.)$, $j = 0, ..., Q$. Hence the key element governing the second−order stationarity is independent of the moving average part.

Corollary 11 *When either (i) (s_t) has a single regime or (ii) (s_t) is an independent process, the condition $(II-2.7)$ reduces to*

$$\lambda_{(2)} := \rho\left(E\left\{\widetilde{\Gamma}_{s_t}^{\otimes 2}(e_t)\right\}\right) < 1.$$

Proof. In the first case, $\mathbb{P} = 1$, so $\rho\left(\mathbb{P}(\underline{\Gamma}^{(2)})\right) = \rho\left(E\left\{\Gamma_{s_t}^{\otimes 2}(e_t)\right\}\right)$. The second follows from the independence of the process $((e_t, s_t), t \in \mathbb{Z})$ and therefore the columns (by block) of the matrix $\underline{\widetilde{\Gamma}}^{(2)}$ are identical and however, the spectral radius of $\underline{\widetilde{\Gamma}}^{(2)}$ reduce to the spectral radius of the sum of block matrices of any column of such matrix. ■

Corollary 12 *For the model $MS-BL(1,q,1,1)$ with $b_0(s_t)\neq 0$. The condition $(II-2.7)$ reduces to* $\rho\left(\mathbb{P}(\underline{\gamma}^{(2)})\right)<1$ *where* $\underline{\gamma}^{(2)}=\left(a_1^2\left(i\right)+\sigma^2 c_{11}^2\left(i\right),i=1,...,d\right)$.

Proof. In this case $M_0(s_t)=a_1(s_t)$, $M_1(s_t)=c_{11}(s_t)$, so

$$\Gamma^{(2)}(i)=\begin{pmatrix} a_1^2(i) & a_1(i)c_{11}(i) & a_1(i)c_{11}(i) & c_{11}^2(i)\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ \sigma^2 a_1^2(i) & \sigma^2 a_1(i)c_{11}(i) & \sigma^2 a_1(i)c_{11}(i) & \sigma^2 c_{11}^2(i)\end{pmatrix}$$

a simple calculation shows that the non−zero eigenvalues of $\mathbb{P}(\underline{\Gamma}^{(2)})$ are the same of that of $\mathbb{P}(\underline{\gamma}^{(2)})$ so the result follows. Noting here that when $d=2$, with $p_{12}=p_{21}=p$, then the condition $(II-2.7)$ is equivalent to the following two conditions

$$\begin{cases}\begin{pmatrix}(2p-1)\left(a_1^2(1)+\sigma^2c_{11}^2(1)\right)\left(a_1^2(2)+\sigma^2c_{11}^2(2)\right)\\ +(1-p)\left(a_1^2(1)+\sigma^2c_{11}^2(1)+a_1^2(2)+\sigma^2c_{11}^2(2)\right)\end{pmatrix}<1,\\ (1-p)\left(a_1^2(1)+\sigma^2c_{11}^2(1)+a_1^2(2)+\sigma^2c_{11}^2(2)\right)\leq 2.\end{cases}$$

In particular, for $d=2$, $a_1(2)=0$, $c_{11}(1)=c_{11}(2), p=0.85$ and $e_t\rightsquigarrow N(0,1)$, the regions of strict and second-order stationarity are shown in $Fig1$.

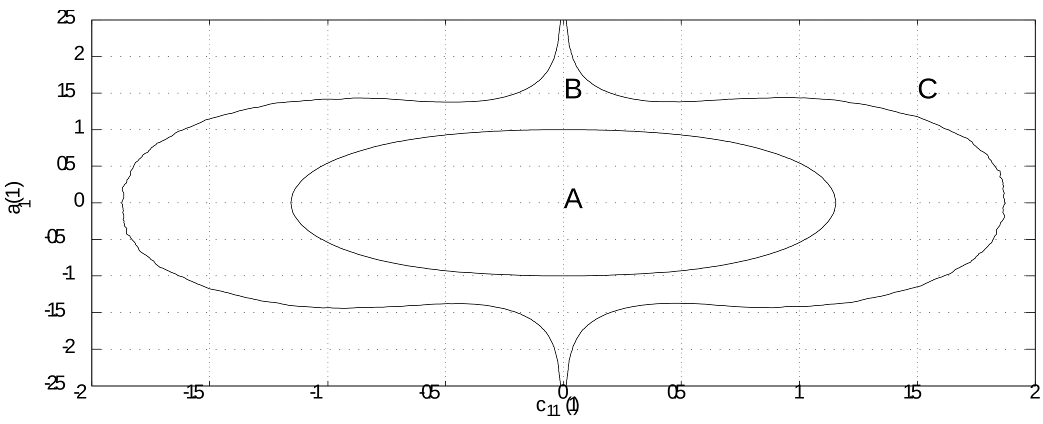

Fig1. Strict and second-order stationarity region for Example 4. A: Second-order stationarity, $A\cup B$: Strict stationarity, and C: non-stationary

■

Corollary 13 *When $Q=1$, the condition $(II-2.7)$ reduces to* $\rho\left(\mathbb{P}(\underline{M}^{(2)})\right)<1$ *where* $\underline{M}^{(2)}=\left(M_0^{\otimes 2}\left(i\right)+\sigma^2 M_1^{\otimes 2}\left(i\right),i=1,...,d\right)$

Proof. In this case the matrices $\Gamma^{(2)}(i)$ take the form

$$\Gamma^{(2)}(i)=\begin{pmatrix} M_0^{\otimes 2}(i) & M_0(i)\otimes M_1(i) & M_1(i)\otimes M_0(i) & M_1^{\otimes 2}(i) \\ O_{(r^2)} & O_{(r^2)} & O_{(r^2)} & O_{(r^2)} \\ O_{(r^2)} & O_{(r^2)} & O_{(r^2)} & O_{(r^2)} \\ \sigma^2 M_0^{\otimes 2}(i) & \sigma^2 M_0(i)\otimes M_1(i) & \sigma^2 M_1(i)\otimes M_0(i) & \sigma^2 M_1^{\otimes 2}(i) \end{pmatrix}$$

and consequently by Remark 10, the non-zero eigenvalues of $\mathbb{P}(\underline{\Gamma}^{(2)})$ are the same as those of $\mathbb{P}(\underline{M}^{(2)})$. ■

Corollary 14 [*The* $MS-ARMA$] *For the* $MS-ARMA$ *model, i.e., when the coefficients* $c_{ij}(.)$ *in* $(II-1.1)$ *are all zeros. The condition* $(II-2.7)$ *reduces to* $\rho\left(\mathbb{P}(\underline{M}_0^{(2)})\right)<1$, *where* $\underline{M}_0^{(2)}=\left(M_0^{\otimes 2}(i), i=1,...,d\right)$.

Proof. In this case; the matrices $M_j(s_t)=O_{(p)}$, $j=1,...,Q$, so $\rho\left(\mathbb{P}\left(\underline{\Gamma}^{(2)}\right)\right)=\rho\left(\mathbb{P}(\underline{M}_0^{(2)})\right)$. ■

Corollary 15 [$MS-GARCH$] *Without loss of generality, we consider the* $MS-GARCH\,(p,p)$ *model defined by*

$$\begin{cases} X_t=\sqrt{h_t}e_t \\ h_t=a_0(s_t)+\sum\limits_{j=1}^{p}a_j(s_t)h_{t-j}+\sum\limits_{i=1}^{p}c_{ii}(s_t)X_{t-i}^2 \end{cases}$$

in which for all $k\in\mathbb{S}$, *the sequences* $(a_i(k))_{0\leq i\leq p}$ *and* $(c_{ii}(k))_{1\leq i\leq p}$ *are positive with* $a_0(k)>0$. *Note that the volatility process* $(h_t, t\in\mathbb{Z})$ *can be regarded as a diagonal bilinear model without moving average part. So, the condition* $(II-2.7)$ *reduces to* $\rho\left(\mathbb{P}(\underline{M})\right)<1$, *where* $\underline{M}=(M_0(i)+\sigma^2 M_1(i), i=1,...,d)$.

Proof. Straightforward and hence omitted. ■

To verify that the series $(II-2.6)$ is well-defined in $\mathbb{L}_2$, it is sufficient to show that the sequence $\underline{Y}_t(k)=\left\{\prod\limits_{i=0}^{k-1}\Gamma_{s_{t-i}}(e_{t-i})\right\}\underline{\eta}_{s_{t-k}}(e_{t-k})$ converges to $\underline{O}_{(s)}$ in $\mathbb{L}_2$ at an exponential rate as $k\to+\infty$. From the independence between s_t and e_t, we have $E\left\{\underline{Y}_t^{\otimes 2}(k)\right\}=\mathbb{I}'_{(s^2)}\mathbb{P}^k(\underline{\Gamma}^{(2)})\underline{\Pi}(\underline{\Sigma}^{(2)})$ where $\underline{\Sigma}^{(2)}:=\left(E\left\{\underline{\eta}_{s_t=i}^{\otimes 2}(e_t)\right\}, 1\leq i\leq d\right)$. Under the condition $(II-2.5)$ and by Jordan's decomposition, we obtain

$$\|\underline{Y}_t(k)\|_{\mathbb{L}_2}\leq\left\|\mathbb{I}'_{(s^2)}\mathbb{P}^k(\underline{\Gamma}^{(2)})\underline{\Pi}(\underline{\Sigma}^{(2)})\right\|^{1/2}\leq\left\|\mathbb{I}'_{(s^2)}\right\|^{1/2}\left\|\mathbb{P}^k(\underline{\Gamma}^{(2)})\right\|^{1/2}\left\|\underline{\Pi}(\underline{\Sigma}^{(2)})\right\|^{1/2}\leq K\lambda_{(2)}^{k/2}$$

so $\underline{Y}_t(k)\to\underline{O}_{(s)}$ in $\mathbb{L}_2$ at an exponential rate as $k\to+\infty$. Hence, the series $(II-2.6)$ is the unique solution of $(II-2.4)$ which converges in $\mathbb{L}_2$ and absolutely almost surely. ■

Example 16 [*Non-sufficiency of local second-order stationarity*] *Consider the following* $MS-BL$

$$X_t = \begin{cases} c_1(1)X_{t-1}e_{t-1} + c_2(1)X_{t-2}e_{t-1} + e_t \text{ if } s_t = 1 \\ c_1(2)X_{t-1} + e_t \text{ if } s_t = 2 \end{cases}$$

in which $(e_t, t \in \mathbb{Z})$ *is an i.i.d process distributed as* $N(0,1)$. *If* $(X_t, t \in \mathbb{Z})$ *is a second-order stationary, then* $E\{X_t^2 | s_t = 1, s_{t-1} = 2\}$ *is finite and independent of* t. *Moreover,*

$$\begin{aligned} & E\left\{X_t^2 | s_t = 1, s_{t-1} = 2\right\} \\ & = E\left\{\left((c_1(1)c_1(2) + c_2(1))X_{t-2}e_{t-1} + c_1(1)e_{t-1}^2 + e_t\right)^2 | s_t = 1, s_{t-1} = 2\right\} \\ & \geq E\left\{(c_1(1)c_1(2) + c_2(1))^2 X_{t-2}^2 | s_t = 1, s_{t-1} = 2\right\} \\ & \geq \left\{ \begin{array}{l} (c_1(1)c_1(2) + c_2(1))^2 E\left\{X_{t-2}^2 | s_t = 1, s_{t-1} = 2, s_{t-2} = 1, s_{t-3} = 2\right\} \\ P(s_{t-2} = 1, s_{t-3} = 2 | s_t = 1, s_{t-1} = 2) \end{array} \right\} \\ & = (c_1(1)c_1(2) + c_2(1))^2 p_{12}p_{21} E\left\{X_{t-2}^2 | s_{t-2} = 1, s_{t-3} = 2\right\}. \end{aligned}$$

This is not possible when $\lambda_{(1)} = (c_1(1)c_1(2) + c_2(1))^2 p_{12}p_{21} > 1$. *Furthermore, it can be seen that the first regime is locally second-order stationary if* $\rho\left(E\left\{c^{\otimes 2}(e_t)\right\}\right) < 1$ *where* $c(e_t) = \begin{pmatrix} c_1(1)e_t & c_2(1)e_t \\ 1 & 0 \end{pmatrix}$, *hence, for*

$$c_1(1) = 0.35, c_2(1) = -0.9, c_1(2) = -0.9, p_{11} = 0.2, p_{22} = 0.1$$

we can check that both regimes are second-order stationary and that $\lambda_{(1)} > 1$. *Other examples for* $MS-ARMA$ *models can be found in FZ* [22] *and in Stelzer* [68].

Remark 17 *It is worth noting that the condition* $(II-2.7)$ *is only sufficient in general as already pointed by* FZ [22] *(see example 4) upon a* $MS-AR(2)$ *model with two regimes.*

2.2.3 Existence of higher-order moments

In this subsection, we shall interested for conditions ensuring the existence of higher−order moments for strictly stationary $MS-SBL$ processes having state−space representation $(II-2.4)$.

Theorem 18 *Consider the* $MS-SBL$ *process* $(II-2.2)$ *with state−space representation* $(II-2.4)$. *For any positive integer* m, *suppose that* $E\left\{e_t^{m+2}\right\} <$

$+\infty$ *and*

$$\lambda_{(m)} := \rho\left(\mathbb{P}(\underline{\Gamma}^{(m)})\right) < 1, \tag{II-2.8}$$

where $\underline{\Gamma}^{(m)} := \left(E\left\{\Gamma_{s_t=i}^{\otimes m}(e_t)\right\}, 1 \leq i \leq d\right)$. *Then the process* $(\underline{Y}_t, t \in \mathbb{Z})$ *defined by* $(II-2.4)$ *has a unique, causal, ergodic, strictly stationary solution given by* $(II-2.6)$ *and satisfies* $E\left\{\left|\underline{Y}_t^{\otimes m}\right|\right\} < +\infty$.

Proof. It is readily seen that $E\left\{\underline{Y}_t^{\otimes m}(k)\right\} = \mathbb{I}'_{(s^m)}\mathbb{P}^k(\underline{\Gamma}^{(m)})\underline{\Pi}(\underline{\Sigma}^{(m)})$ where $\underline{\Sigma}^{(m)} = \left(E\left\{\underline{\eta}_{s_t=i}^{\otimes m}(e_t)\right\}, i = 1, ..., d\right)$. Hence

$$\left\|\underline{Y}_t^{\otimes m}(k)\right\|_{\mathbb{L}_m} \leq \left\|\mathbb{I}'_{(s^m)}\right\|^{1/m}\left\|\mathbb{P}^k(\underline{\Gamma}^{(m)})\right\|^{1/m}\left\|\underline{\Pi}(\underline{\Sigma}^{(m)})\right\|^{1/m}.$$

So, under the condition $(II-2.8)$ and by Jordan decomposition, $\left\|\mathbb{P}^k(\underline{\Gamma}^{(m)})\right\|$ converges to zero at an exponential rate as $k \to \infty$. Consequently, for any t, $\sum_{k=1}^{n} \underline{Y}_t(k)$ converges as $n \to \infty$ to $\underline{Y}_t$ defined by $(II-2.6)$ both in $\mathbb{L}_m$ and absolutely almost surely. The rest of assertions are immediate and hence omitted.
■

In the following table, we summarize the sufficient conditions for the existence of $E\{X_t^m\}$ in some particular models.

Specification	Condition $(II-2.8)$	special cases: $p=Q=1$
Standard	$\lambda_{(m)}=\rho\left(E\left\{\Gamma^{(m)}(e_t)\right\}\right)<1,$	$E\left\{(a_1+c_{11}e_0)^m\right\}<1$
Independent-switching	$\lambda_{(m)}=\rho\left(E\left\{\Gamma_{s_t}^{(m)}(e_t)\right\}\right)<1,$	$E\left\{(a_1(s_t)+c_{11}(s_t)e_0)^m\right\}<1$
$MS-ARMA$	$\lambda_{(m)}=\rho\left(\mathbb{P}(\underline{M}_0^{(m)})\right)<1,$	$\rho\left(\mathbb{P}\left(\underline{a}^{(m)}\right)\right)<1^{(a)}$
$MS-GARCH$	$\lambda_{(m)}=\rho\left(\mathbb{P}(\underline{M}^{(m)})\right)<1,$	$\rho\left\{\mathbb{P}\left(\underline{\delta}^{(m)}\right)\right\}<1^{(b)}$

$^{(a)}\underline{a}^{(m)}=\{a_1^m(i), i=1,...,d\}$

$^{(b)}\underline{\delta}^{(m)}=\left\{E\left\{\left(a_1(i)+c_{11}(i)e_0^2\right)^m\right\}, i=1,...,d\right\}$

Table 1: Conditions $(II-2.8)$ for the existence of $E\{X_t^m\}$ for certain $MS-$models

2.2.4 Computation of the second−order moment and $ARMA$ representation

Once second−order stationary condition is established, it can be useful to compute the second−order moment of the process $(X_t, t \in \mathbb{Z})$. For the convenience, we shall consider the centered version of the state vector $\underline{Y}_t$, i.e.,

$$\widehat{\underline{Y}}_t = \Gamma_{s_t}(e_t)\widehat{\underline{Y}}_{t-1} + \widehat{\underline{\eta}}_{s_t}(e_t), \tag{II-2.9}$$

where $\widehat{\underline{Y}}_t = \underline{Y}_t - E\{\underline{Y}_t\}$ and $\widehat{\underline{\eta}}_{s_t}(e_t)$ is centred residual vector such that given $s_t = k$, $\widehat{\underline{\eta}}_k(e_t)$ is orthogonal to $\widehat{\underline{Y}}_{t'}$ for all $t' < t$. Let $\widehat{\underline{\Sigma}}_\eta^{(2)} := \left(E\left\{\widehat{\underline{\eta}}^{\otimes 2}_{s_t=i}(e_t)\right\}, i = 1, ..., d\right)$, $\underline{\Gamma}_0 = (\Gamma_0(i), i = 1, ..., d)$, $\underline{1} = (1, ..., 1)' \in \mathbb{R}^d$. To express the moments, we recall here that if for $i \geq 1$, Z_{t-i} be an integrable random variable belonging to $\sigma(e_{t-s}, s \geq i)$, then $\pi(k)E\{Z_{t-i}|s_t = k\} = \sum_{j=1}^{d} E\{Z_{t-i}|s_{t-i} = j\} p_{jk}^{(i)}\pi(j)$.

Proposition 19 *Consider the $MS-SBL(p, q, p, Q)$ process $(II-2.2)$ with state-space representation $(II-2.9)$. Then under the conditions of Theorem 9, we have*

$$\begin{aligned}\underline{\Sigma}_{\underline{X}}(h) &= Vec\left\{Cov\left(\underline{X}_t, \underline{X}_{t-h}\right)\right\}\\ &= \begin{cases} \left(\underline{1} \otimes F^{\otimes 2}\right)' \left(I_{(ds^2)} - \mathbb{P}\left(\underline{\Gamma}^{(2)}\right)\right)^{-1} \underline{\Pi}\left(\widehat{\underline{\Sigma}}_\eta^{(2)}\right) & \text{if } h = 0\\ \left(\underline{1} \otimes F^{\otimes 2}\right)' \mathbb{P}^h\left(\underline{\Gamma}_0 \otimes I_{(s)}\right)\left(I_{(ds^2)} - \mathbb{P}\left(\underline{\Gamma}^{(2)}\right)\right)^{-1} \underline{\Pi}\left(\widehat{\underline{\Sigma}}_\eta^{(2)}\right) & \text{if } h > 0\end{cases}\end{aligned}$$

Proof. Starting from $(II-2.9)$,

1. **a**. When $h = 0$ we have

$$\begin{aligned}\pi(k)E\left\{\widehat{\underline{Y}}_t^{\otimes 2}|s_t = k\right\} &= \pi(k)\widehat{\underline{\Sigma}}_\eta^{(2)}(k) + \pi(k)\Gamma^{(2)}(k)E\left\{\widehat{\underline{Y}}_{t-1}^{\otimes 2}|s_t = k\right\}\\ &= \pi(k)\widehat{\underline{\Sigma}}_\eta^{(2)}(k) + \Gamma^{(2)}(k)\sum_{j=1}^{d} E\left\{\widehat{\underline{Y}}_{t-1}^{\otimes 2}|s_{t-1} = j\right\} p_{jk}\pi(j).\end{aligned}$$

Set $\underline{\Sigma}(0) = \left(E\left\{\widehat{\underline{Y}}_t^{\otimes 2\prime}|s_t = k\right\}, k = 1, ..., d\right)'$, then we have $\underline{\Pi}(\underline{\Sigma}(0)) = \mathbb{P}\left(\underline{\Gamma}^{(2)}\right)\underline{\Pi}(\underline{\Sigma}(0)) + \underline{\Pi}\left(\widehat{\underline{\Sigma}}_\eta^{(2)}\right)$, so

$$\Pi(\underline{\Sigma}(0)) = \left(I_{(ds^2)} - \mathbb{P}\left(\underline{\Gamma}^{(2)}\right)\right)^{-1} \underline{\Pi}\left(\widehat{\underline{\Sigma}}_\eta^{(2)}\right).$$

b. When $h > 0$,

$$\pi(k)E\left\{\widehat{\underline{Y}}_t \otimes \widehat{\underline{Y}}'_{t-h}\middle| s_t = k\right\} = \sum_{j=1}^{d}\left\{\begin{array}{l}\left(\Gamma_0(k) \otimes I_{(s)}\right) p_{jk}\pi(j)\\ E\left\{\widehat{\underline{Y}}_{t-1} \otimes \widehat{\underline{Y}}'_{t-1-(h-1)}\middle| s_{t-1} = j\right\}\end{array}\right\}.$$

Let $\underline{\Sigma}(h)=\left(E\left\{\widehat{\underline{Y}}_t\otimes\widehat{\underline{Y}}_{t-h}|s_t=k\right\},k=1,...,d\right)$, then

$$\underline{\Pi}\left(\underline{\Sigma}(h)\right)=\mathbb{P}\left(\underline{\Gamma}_0\otimes I_{(s)}\right)\underline{\Pi}\left(\underline{\Sigma}(h-1)\right)=\mathbb{P}^h\left(\underline{\Gamma}_0\otimes I_{(s)}\right)\underline{\Pi}\left(\underline{\Sigma}(0)\right), \tag{II-2.10}$$

and hence $\underline{\Sigma}_X(h)=\left(\underline{1}\otimes F^{\otimes 2}\right)'\underline{\Pi}\left(\underline{\Sigma}(h)\right)$. So

$$\gamma(h)=\left(\underline{1}\otimes(F\underline{H})^{\otimes 2}\right)'\underline{\Pi}\left(\underline{\Sigma}(h)\right) \tag{II-2.11}$$

■

Remark 20 *For $MS-ARMA(p,q)$ process with $b_0(s_t)=1$ and under the conditions of Corollary 14, then we have $R^{(X)}(h)=\mathbb{P}(A_0)R^{(X)}(h-1)=\mathbb{P}^h(A_0)R^{(X)}(0)$ for any $h>0$. So $E\{X_tX_{t-h}\}=\left(\underline{e}\otimes\underline{H}^{\otimes 2}\right)'R^{(X)}(h)$ and the covariance function of $(X_t,t\in\mathbb{Z})$ follows.*

Remark 21 *For the $MS-GARCH(p,p)$ model defined in Corollary 15, we have $E\{X_t\}=0$ and covariance function*

$$Cov(X_t,X_{t-h})=\begin{cases}\underline{e}'\left(I_{(d)}-\sum_{j=1}^{p}\mathbb{P}^{(j)}\left(\underline{a}_j+\sigma^2\underline{c}_{jj}\right)\right)^{-1}\Pi\left(\sigma^2\underline{a}_0\right) \ \textit{if}\ h=0\\ 0\ \textit{for all}\ h>0\end{cases}$$

where $\underline{a}_j+\sigma^2\underline{c}_{jj}=\{a_j(i)+\sigma^2c_{jj}(i),i=1,...,d\}$, $j=1,...,d$.

$ARMA$ representation

The $ARMA$ representation play an important role in forecasting and in identification, so certain non-linear processes are already represented as an $ARMA$ model. Indeed, Bibi [6] showed that a superdiagonal bilinear process with time$-$varying coefficient admit a weak $ARMA$ representation, FZ [21] have established an $ARMA$ representation for $(MS-)\,GARCH$ and others non linear processes of interest. The following proposition establishes an $ARMA$ representation for $MS-SBL(p,q,p,Q)$.

Proposition 22 *Under the conditions of Proposition 19, the $MS-SBL(p,q,p,Q)$ process with state-space representation $(II-2.9)$ is an $ARMA$ process.*

Proof. To prove proposition 22, we use the same approach as FZ [22]. First, it is worth noting that the equation $(II-2.10)$ is equivalently to

$$W(h)=\mathbb{P}\left(\underline{\Gamma}_0\right)W(h-1) \tag{II-2.12}$$

where $W(h)=\underline{\Pi}\left(\underline{\Sigma}\left(h\right)\right)$ with $\underline{\Sigma}(h)=\left(E\left\{\underline{\widehat{Y}}_t\underline{\widehat{Y}}'_{t-h}|s_t=k\right\},k=1,...,d\right)$. Hence, the result follows essentially the same arguments as in FZ [22], in other words, there exists an uncorrelated sequence of random variables $(\xi(t))_{t\in\mathbb{Z}}$ with zero mean and finite variance such that and almost surely

$$X_t=\sum_{i=1}^{p^*}a_i^*X_{t-i}+\sum_{j=0}^{q^*}b_j^*\xi_{t-j},\ b_0^*=1, \tag{II-2.13}$$

where the coefficients $\left(a_j^*,\ b_j^*,1\leq j\leq p^*\vee q^*\right)$ are functions of $\left(a_j\left(.\right),1\leq j\leq p\right)$, $\left(b_j\left(.\right),1\leq j\leq Q\right)$ and $\left(c_{ij}\left(.\right),1\leq j\leq i\leq Q\vee p\right)$. The innovation process $\left(\xi_t\right)_{t\in\mathbb{Z}}$ is not Gaussian nor a martingale difference sequence when the $c_{ij}\left(.\right)$'s are not equal to zero. ■

Remark 23 *The representation $(II-2.13)$ may be used to compute the best linear predictor $\widehat{X}_{t+1|t}$ of $MS-SBL(p,q,p,Q)$ given $\{X_s,s\leq t\}$. Indeed, according Bibi [6] the best linear predictor of X_{t+1} when $\{X_s,s\leq t\}$ is given is*

$$\widehat{X}_{t+1|t}=\left(1-\frac{\Psi^*\left(B\right)}{\Phi^*\left(B\right)}\right)X_{t+1}$$

where $\Psi^*\left(z\right)=1-\sum\limits_{i=1}^{p^*}a_i^*z^i$ and $\Phi^*\left(z\right)=1+\sum\limits_{j=1}^{q^*}b_j^*z^j\neq 0$ for all $z\in\mathbb{C}:|z|<1$. Moreover, $E\left\{\left(\widehat{X}_{t+1|t}-X_{t+1}\right)^2\right\}=var\left\{\xi_t\right\}>var\left\{e_t\right\}=\sigma^2$, so, the mean square error of the best linear prediction is always greater than the innovation variance of the process $(X_t,t\in\mathbb{Z})$ satisfying $(II-2.2)$.

2.3 Covariance structure of higher−power for $MS-SBL$

For the identification purpose it is necessary to look at higher−power of the process in order to distinguish between different $ARMA$ representation. So, in this section we extend the result of the subsection 2.2.4 to power of $(X_t,t\in\mathbb{Z})$. For this purpose, we first establish the following lemma

Lemme 24 *Consider the $MS-SBL(p,q,p,Q)$ model $(II-2.2)$ with state space representation $(II-2.9)$. Let us define the following matrices $B_j^{(l)}(s_t,e_t)$, $j=0,...,l=0,...,m$, with appropriate dimension such that*

$$\forall l\in\mathbb{N}:\underline{\widehat{Y}}_t^{\otimes l}=\left(\widehat{\underline{\eta}}_{s_t}(e_t)+\Gamma_{s_t}(e_t)\underline{\widehat{Y}}_{t-1}\right)^{\otimes l}=\sum_{j=0}^{l}B_j^{(l)}\left(s_t,e_t\right)\underline{\widehat{Y}}_{t-1}^{\otimes j} \tag{II-3.1}$$

where by convention $B_j^{(l)}(.,.) = O$ *if* $j > l$ *or* $j < 0$, $\underline{\widehat{Y}}_t^{\otimes 0} = B_0^{(0)}(.,.) = 1$. *Then* $B_j^{(l)}(s_t, e_t)$ *are uniquely determined by the following recursion*

$$B_0^{(1)}(s_t, e_t) = \underline{\widehat{\eta}}_{s_t}(e_t), B_1^{(1)}(s_t, e_t) = \Gamma_{s_t}(e_t),$$
$$B_j^{(m+1)}(s_t, e_t) = \underline{\widehat{\eta}}_{s_t}(e_t) \otimes B_j^{(m)}(s_t, e_t) + \Gamma_{s_t}(e_t) \otimes B_{j-1}^{(m)}(s_t, e_t) \text{ for } m > 1$$

Proof. indeed, the formula $(II - 3.1)$ is obvious for $l = 1$. Assuming that $(II - 3.1)$ hold for every $l \geq 1$, then we have

$$\begin{aligned}
\underline{\widehat{Y}}_t^{\otimes l+1} &= \sum_{j=0}^{l} \left(\underline{\widehat{\eta}}_{s_t}(e_t) + \Gamma_{s_t}(e_t)\underline{\widehat{Y}}_{t-1} \right) \otimes B_j^{(l)}(s_t, e_t)\, \underline{\widehat{Y}}_{t-1}^{\otimes j} \\
&= \sum_{j=0}^{l} \underline{\widehat{\eta}}_{s_t}(e_t) \otimes \left\{ B_j^{(l)}(s_t, e_t)\, \underline{\widehat{Y}}_{t-1}^{\otimes j} \right\} + \left\{ \Gamma_{s_t}(e_t) \otimes B_j^{(l)}(s_t, e_t) \right\} \underline{\widehat{Y}}_{t-1} \otimes \underline{\widehat{Y}}_{t-1}^{\otimes j} \\
&= \sum_{j=0}^{l+1} \left\{ \underline{\widehat{\eta}}_{s_t}(e_t) \otimes B_j^{(l)}(s_t, e_t) \right\} \underline{\widehat{Y}}_{t-1}^{\otimes j} + \left\{ \Gamma_{s_t}(e_t) \otimes B_{j-1}^{(l)}(s_t, e_t) \right\} \underline{\widehat{Y}}_{t-1}^{\otimes j}
\end{aligned}$$

So the result follows. ■

Now, set $\underline{\Sigma}_{\underline{Y}}^{(m)} = \left(E\left\{ \underline{\widehat{Y}}_t^{\otimes m} | s_t = k \right\}, k = 1, ..., d \right)$,
$\underline{\Sigma}_{\underline{Y}}^{(l,m)}(h) = \left(E\left\{ \underline{\widehat{Y}}_t^{\otimes l} \otimes \underline{\widehat{Y}}_{t-h}^{\otimes m} | s_t = k \right\}, k = 1, ..., d \right)'$ and
$\underline{B}^{(m,j)} = \left(E\left\{ B_j^{(m)}(k, e_t) \right\}, k = 1, ..., d \right)$, then it is no difficult to see that

$$\begin{aligned}
\underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(m)}) &= \sum_{j=0}^{m} \mathbb{P}\left(\underline{B}^{(m,j)} \right) \underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(j)}) \text{ and} \\
\underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(l,m)}(h)) &= \sum_{j=0}^{l} \mathbb{P}\left(\underline{\widetilde{B}}^{(l,j)} \right) \underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(l-j,m)}(h-1)),\ l > 1.
\end{aligned}$$

if $\underline{\widehat{Y}}_t \in \mathbb{L}_{l+m}$ where $\underline{\widetilde{B}}^{(l,j)} = \underline{B}^{(l,j)} \otimes I_{(s^m)}$. Moreover, we have

$$W^{(m)}(h) = \begin{pmatrix} \underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(m,m)}(h)) \\ \underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(m-1,m)}(h)) \\ \vdots \\ \vdots \\ \underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(0,m)}(h)) \end{pmatrix}$$

$$= \begin{pmatrix} \mathbb{P}\left(\underline{\widetilde{B}}^{(m,0)}\right) & \mathbb{P}\left(\underline{\widetilde{B}}^{(m,1)}\right) & \dots & \mathbb{P}\left(\underline{\widetilde{B}}^{(m,m)}\right) \\ O & \mathbb{P}\left(\underline{\widetilde{B}}^{(m-1,0)}\right) & \dots & \mathbb{P}\left(\underline{\widetilde{B}}^{(m-1,m-1)}\right) \\ \vdots & \ddots & \dots & \vdots \\ \vdots & \ddots & \mathbb{P}\left(\underline{\widetilde{B}}^{(1,0)}\right) & \mathbb{P}\left(\underline{\widetilde{B}}^{(1,1)}\right) \\ O & \dots & O & \mathbb{P}\left(I_{(s^m)}\right) \end{pmatrix} W^{(m)}(h-1)$$

$$= \Lambda(m)\, W^{(m)}(h-1)$$

in which O is the null matrices with appropriate dimensions.

Proposition 25 *Consider the $MS-SBL(p,q,p,Q)$ model $(II-2.2)$ with state- space representation $(II-2.9)$. If $E\left\{e_t^{2(m+1)}\right\} < +\infty$ and $\lambda_{(2m)} := \rho\left(\mathbb{P}(\underline{\Gamma}^{(2m)})\right) < 1$, then $\left(\underline{\widehat{Y}}_t^{\otimes m}, t \in \mathbb{Z}\right)$ is second-order stationary process and*

$$\underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(m)}) = \begin{cases} \underline{\Pi}(1) \text{ if } m = 0 \\ \left(I_{(ds^m)} - \mathbb{P}\left(\underline{B}^{(m,m)}\right)\right)^{-1} \sum\limits_{j=0}^{m-1} \mathbb{P}\left(\underline{B}^{(m,j)}\right) \underline{\Pi}(\underline{\Sigma}_{X}^{(j)}) \text{ if } m > 1 \end{cases}$$

$$\underline{\Pi}\left(\underline{\Sigma}_{\underline{Y}}^{(l,m)}(1)\right) = \sum_{j=0}^{l} \mathbb{P}\left(\underline{\widetilde{B}}^{(l,j)}\right) \underline{\Pi}\left(\underline{\Sigma}_{\underline{Y}}^{(l+m)}\right) \text{ if } \underline{\widehat{Y}}_t \in \mathbb{L}_{l+m} \text{ and}$$

$$\underline{W}^{(m)}(h) = \Lambda(m)\, \underline{W}^{(m)}(h-1) \text{ and for any } h > 1$$

Remark 26 *The proposition 25, we allow to compute $W^{(m)}(h)$ recursively for all $h \geq 0$. The unconditional mean and covariance of $\left(\underline{X}_t^{\otimes m}, t \in \mathbb{Z}\right)$ are given by*

$$\underline{\mu}_{\underline{X}^m}^{(m)} = E\left\{\underline{\widehat{X}}_t^{\otimes m}\right\} = F^{\otimes m\prime} E\left\{\underline{\widehat{Y}}_t^{\otimes m}\right\} = \left(\underline{1} \otimes F^{\otimes m}\right)' \underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(m)})$$

$$\begin{aligned} \Sigma_{\underline{X}^m}(h) &= E\left\{\underline{\widehat{X}}_t^{\otimes m} \otimes \underline{\widehat{X}}_{t-h}^{\otimes m}\right\} - E\left\{\underline{\widehat{X}}_t^{\otimes m}\right\} \otimes E\left\{\underline{\widehat{X}}_{t-h}^{\otimes m}\right\} \\ &= \left(\underline{1} \otimes F^{\otimes 2m}\right)' \underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(m,m)}(h)) - E\left\{\underline{\widehat{X}}_t^{\otimes m}\right\} \otimes \left(\underline{1} \otimes F^{\otimes m}\right)' \underline{\Pi}(\underline{\Sigma}_{\underline{Y}}^{(0,m)}(h)) \\ &= \underline{v}' W^{(m)}(h) \end{aligned}$$

where $\underline{v}' = ((\underline{1} \otimes F^{\otimes 2m})' \vdots O \vdots \dots \vdots O \vdots - E\left\{\underline{\widehat{X}}_t^{\otimes m}\right\} \otimes (\underline{1} \otimes F^{\otimes m})')$.

We are now in a position to state the following results

Proposition 27 *Let* $(X_t, t \in \mathbb{Z})$ *be the causal second−order stationary of* $MS-SBL\,(p,q,p,Q)$ *with state−space representation* $(II - 2.9)$. *Then under the conditions of Proposition* (25), $(X_t^m, t \in \mathbb{Z})$ *is an* $ARMA$ *process.*

Proof. The proof follows essentially the same arguments as in proof of proposition 22. ■

2.3.1 Generic examples

In this subsection, we consider some important special cases aimed to illustrate the second−order stationarity conditions and explicit expansions of variance and covariance function and $ARMA$ representation for given process $(X_t, t \in \mathbb{Z})$. For this purpose, we consider the following models

$$\text{Model(1): } X_t = g_{s_t}(e_t)X_{t-1} + e_t, \text{ Model(2): } X_t = h_{s_t}(e_{t-1})X_{t-2} + e_t \quad \text{(II-3.2)}$$

in which $g_{s_t}(e_t) = a\,(s_t) + e_t c(s_t)$ (resp. $h_{s_t}(e_{t-1}) = c(s_t)e_{t-1}$) and $e_t \rightsquigarrow N(0,1)$ with $\mu_e^{(2p)} = E\left\{e_0^{2p}\right\} = (2p)!/2^p p!$ and $\mu_e^{(2p+1)} = E\left\{e_0^{2p+1}\right\} = 0$.

Moments

Table 2, summarizes the available results concerning $E\left\{|X_t|^m\right\}$ to be finite.

Specification	Model(1)	Model(2)
Standard	$E\left\{g^{m}\left(e_{0}\right)\right\}<1$	$c^{m}\mu_{e}^{(2m)}<1,$
Independent-switching	$E\left\{g_{s_t}^{m}\left(e_{0}\right)\right\}<1,$	$E\left\{h_{s_t}^{m}(e_{0})\right\}<1,$
$MS^{(*)}$	$\rho\left(\mathbb{P}(\underline{g}^{m}\left(e_{0}\right))\right)<1$	$\rho\left(\mathbb{P}^{(2)}(\underline{h}^{m}(e_{0}))\right)<1,$
$^{(*)}\underline{k}^{m}\left(e_{0}\right)=\left(E\left\{k_{s_t=i}^{m}\left(e_{0}\right)\right\},i=1,...,d\right)$, $k=g$ or h		

Table 2: Conditions for the existence of $E\left\{\left|X_{t}\right|^{m}\right\}$ for certain specifications

Second-order structure

The following proposition summarizes the available results concerning $\underline{\Sigma}_X^{(1,1)}(h)$

Proposition 28 *Under the appropriate condition ensuring the existence of second order stationarity solution of Models*(i)*,* $i=1$ *and* 2 *(see Table 2), we have*

Model(1): $\underline{\Pi}\left(\underline{\Sigma}_X^{(1,1)}(h)\right)=\begin{cases}\left(I_{(d)}-\mathbb{P}\left(\underline{g}^2(e_0)\right)\right)^{-1}\underline{\Pi}, & \text{if } h=0\\ \mathbb{P}\left(\underline{a}\right)\underline{\Pi}\left(\underline{\Sigma}_X^{(1,1)}(h-1)\right) & \text{if } h>0\end{cases}$

Model(2): $\underline{\Pi}\left(\underline{\Sigma}_X^{(1,1)}(h)\right)=\begin{cases}\left(I_{(d)}-\mathbb{P}^{(2)}\left(\underline{h}^2(e_0)\right)\right)^{-1}\underline{\Pi}, & \text{if } h=0\\ 0 & \text{otherwise}\end{cases}$

The covariance functions can be obtained by $\gamma_X(h)=\underline{1}'\underline{\Pi}\left(\underline{\Sigma}_X^{(1,1)}(h)\right)$*. Hence, the model*$(2)$ *may be considered as a weak white noise.*

Higher order structure

In order to derive the higher order structure, we shall note $\underline{B}_{(i)}^{(l,j)}$ the associated quantities for Models(i), $i=1,2$ respectively, i,e.;

$$\begin{aligned}\underline{B}_{(1)}^{(l,j)} &= \left(\frac{l!}{j!\,(l-j)!}E\left\{g_k^j(e_t)\,e_t^{l-j}\right\},k=1,...,d\right),\\ \underline{B}_{(2)}^{(l,j)} &= \left(\frac{l!}{j!\,(l-j)!}E\left\{h_k^j(e_{t-1})\,e_t^{l-j}\right\},k=1,...,d\right),\end{aligned}$$

so

$$\underline{\Pi}(\underline{\mu}_X^{(l)})=\begin{cases}\sum\limits_{j=0}^{l}\mathbb{P}\left(\underline{B}_{(1)}^{(l,j)}\right)\underline{\Pi}(\underline{\mu}_X^{(j)}) \text{ for Model (1)}\\ \sum\limits_{j=0}^{l}\mathbb{P}^{(2)}\left(\underline{B}_{(2)}^{(l,j)}\right)\underline{\Pi}(\underline{\mu}_X^{(j)}) \text{ for Model (2)}\end{cases}$$

$$\underline{\Pi}(\underline{\Sigma}_X^{(l,m)}(h))=\begin{cases}\sum\limits_{j=0}^{l}\mathbb{P}\left(\underline{B}_{(1)}^{(l,j)}\right)\underline{\Pi}(\underline{\Sigma}_X^{(j,m)}(h-1)), l\in\{0,...,m\} \text{ for Model (1)}\\ \sum\limits_{j=0}^{l}\mathbb{P}^{(2)}\left(\underline{B}_{(2)}^{(l,j)}\right)\underline{\Pi}(\underline{\Sigma}_X^{(j,m)}(h-2)), l\in\{0,...,m\} \text{ for Model (2)}\end{cases}$$

$$\underline{W}^{(m)}(h)=\begin{cases}\Lambda^{(1)}(m)\,W^{(m)}(h-1) \text{ for Model (1)}\\ \Lambda^{(2)}(m)\,W^{(m)}(h-2) \text{ for Model (2)}\end{cases}$$

where $\Lambda^{(i)}(m)$ has the same form, their entries are respectively $\mathbb{P}^{(i)}(.)$, $i=1,2$. Moreover we have

$$\begin{aligned}&E\left\{X_t^m\right\}=\underline{1}'\underline{\Pi}(\underline{\mu}_X^{(m)})\\ &\gamma_{X^m}(h)=\underline{1}'\underline{\Pi}(\underline{\Sigma}_X^{(m,m)}(h))-E\left\{X_t^m\right\}\underline{1}'\underline{\Pi}(\underline{\Sigma}_X^{(0,m)}(h))=\underline{V}'\underline{W}^{(m)}(h)\end{aligned}$$

where $\underline{V}' = (\underline{1}', \underline{O}', ..., \underline{O}', -E\{X_t^m\}\underline{1}')$. It is worth noting that due to diagonal form of $\Lambda^{(i)}(m)$, $i = 1, 2$, the characteristic polynomial is the product of characteristic polynomials of matrices $\mathbb{P}^{(i)}\left(\underline{B}_{(i)}^{(m,0)}\right)$. We have then the following proposition due to FZ [21].

Proposition 29 *Let $(X_t, t \in \mathbb{Z})$ be a solution of Model(1) or Model(2) and assume that $(X_t^m, t \in \mathbb{Z})$ is second-order stationary. Then $(X_t^m, t \in \mathbb{Z})$ is a solution of $ARMA$ equation of the form*

$$\prod_{i=0}^{m}\prod_{k=1}^{k_i}\left(1-\lambda_k^{(i)}L\right)\left(X_t^m - E\{X_t^m\}\right) = P^{(m)}(L)\,\xi_t^{(m)}$$

where $\left\{\lambda_1^{(i)}, \lambda_2^{(i)}, ..., \lambda_{k_i}^{(i)}\right\}$ (with $k_i \leq d$,$k_0 \leq d-1$) are the eigenvalues different from 0 and 1 in module of $\mathbb{P}^{(j)}\left(\underline{B}_{(j)}^{(l,0)}\right)$, $l = 0, ..., m$, $j = 0, 1$ and $P^{(m)}(L)$ is some polynomial such that $d^o g P^{(m)} \leq d(m+1)-1$, $P^{(m)}(0) = 1$ and $\left(\xi_t^{(m)}, t \in \mathbb{Z}\right)$ is some white noise.

ARMA Representation

Table(4), summarizes the result concerning the orders and the coefficient of the $ARMA$ representation of different power of Model$(i), i = 1, 2$.

Specification	Representation	AR polynomial
	$\underline{m=1}$	
Standard	$AR(1)$	$1-aL$
Independent-switching	$AR(1)$	$1-E\{a(s_t)\}L$
$MS^{(*)}$	$AR(d)$	$\prod_{k=1}^{d}(1-\lambda_k L)$
	$\underline{m>1}$	
Standard$^{(**)}$	$ARMA(m-1,m-1)$	$\prod_{k=1}^{m}\left(1-c_{(1)}^{(m,k)}L\right)$
Independent-switching$^{(***)}$	$ARMA(m-1,m-1)$	$\prod_{k=1}^{m}\left(1-E\left\{c_{(1)}^{(m,k)}(s_t)\right\}L\right)$
$MS^{(****)}$	$ARMA(d(m+1)-1,d(m+1)-1)$	$\prod_{k=1}^{d(m+1)-1}(1-\lambda_k L)$

$^{(*)}\lambda_k =$ eigenvalues $\neq 1$ of $\mathbb{P}(\underline{a})$

$^{(**)}c_{(1)}^{(m,j)}=\frac{m!}{j!(m-j)!}E\left\{g^l(e_t)\,e_t^{m-j}\right\}$

$^{(***)}c_{(1)}^{(m,j)}(s_t)=\frac{m!}{j!(m-j)!}g_{s_t}^l(e_t)\,e_t^{m-j}$

$^{(****)}\lambda_k =$ eigenvalues $\neq 1$ of $\mathbb{P}\left(\underline{B}_{(1)}^{(m,k)}\right)$, $k=0,...,m$.

Table 3:$ARMA$ representation for some specifications of model(1)

Specification	Representation	AR polynomial
	$\underline{m=1}$	
$Standard$, $Independent-switching$ and $MS\}$ is always a weak white noise		
	$\underline{m>1}$	
Standard$^{(*)}$	$ARMA\,(m-1,m-1)$	$\prod_{k=1}^{m}\left(1-c_{(2)}^{(m,k)}L\right)$
Independent-switching$^{(**)}$	$ARMA\,(m-1,m-1)$	$\prod_{k=1}^{m}\left(1-E\left\{c_{(2)}^{(m,k)}\,(s_t)\right\}L\right)$
$MS^{(***)}$	$ARMA\,(d\,(m+1)-1,d\,(m+1)-1)$	$\prod_{k=1}^{d(m+1)-1}(1-\lambda_k L)$

$^{(*)}c_{(2)}^{(m,j)}=\frac{m!}{j!(m-j)!}E\left\{h^l\,(e_{t-1})\,e_t^{m-j}\right\}$

$^{(**)}c_{(2)}^{(m,j)}\,(s_t)=\frac{m!}{j!(m-j)!}h_{s_t}^l\,(e_{t-1})\,e_t^{m-j}$

$^{(***)}\lambda_k=$ eigenvalues $\neq 1$ of $\mathbb{P}^{(2)}\left(\underline{B}_{(2)}^{(m,j)}\right)$, $i=0,...,m$

Table 4:$ARMA$ representation for some specifications of Model(2)

2.4 Invertibility of $MS-BL$ processes

The concept 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, among which, and according to Wittwer [75], model $MS-BL\left(p,q,p,q\right)$ with $b_0\left(.\right)=1$ is said to be strongly invertible if almost surely $e_n-\hat{e}_n\rightarrow 0$ as $n\rightarrow+\infty$ for any sequence $\left(\hat{e}_n\right)_{n\geq q+1}$ of estimate of e_n with initial values $\{\widehat{e}_1,..,\widehat{e}_q,X_1,...,X_q,s_1\}$ (see also Aknouche and Rabehi [2] for further discussion). So by setting $\beta_j(t)=b_j(s_t)+\sum\limits_{i=1}^{P}c_{ij}(s_t)X_{t-i}$ then we obtain

$$\xi_t=e_t-\widehat{e}_t=-\sum_{j=1}^{q}\beta_j(t)\xi_{t-j}$$

. The process $(\xi_t,t\in\mathbb{Z})$ may be rewritten in state space form as

$$\underline{\xi}_t=G\left(t\right)\underline{\xi}_{t-1}$$

where $\underline{\xi}_t=\left(\xi_t,...,\xi_{t-q+1}\right)'$ and $G\left(t\right)$ is an appropriate square random matrix.

Theorem 30 *Consider the model $MS-BL(p,q,p,q)$ with $\gamma_L\left(M\right)<0$, then it is strongly invertible if $\gamma_L\left(G\right)<0$.*

Proof. The proof follows upon the observation that under the condition $\gamma_L\left(M\right)<0$, the process $(X_t,t\in\mathbb{Z})$ is strictly stationary and admit a finite moment of order $\delta\in\left]0,1\right]$, so $E\left\{\log^+\left|X_t\right|\right\}<+\infty$ and hence the top Lyapunov exponent associated with the random matrices $\left(G\left(t\right),t\in\mathbb{Z}\right)$ is well defined. ■
Some simple sufficient condition for invertibility for some restrictive models can be given.

Corollary 31 *Consider the $MS-SBL(0,0,p,1)$ model generate by*

$$X_t=\sum_{j=2}^{p}c_j\left(s_t\right)X_{t-j}e_{t-1}+e_t \qquad \text{(II-4.1)}$$

Then, under the condition of Theorem 30, the model $(II-4.1)$ is invertible whenever $E\left\{\log\left|\beta\left(t\right)\right|\right\}<0$ where $\beta\left(t\right)=-\sum\limits_{j=2}^{p}c_j\left(s_t\right)X_{t-j}$.

Proof. We have

$$\begin{aligned}\frac{1}{t}\log|\xi_t| &= \frac{1}{t}\log\left|\prod_{i=0}^{t-1}\beta(t-i)\right| + \frac{1}{t}\log|\xi_0| \\ &= \frac{1}{t}\sum_{i=0}^{t-1}\log|\beta(t-i)| + \frac{1}{t}\log|\xi_0| \to E\{\log|\beta(t)|\}, a.s.\end{aligned}$$

So the result follows by the Cauchy's criterion. ■

It is worth noting that the evaluation of $E\{\log|\beta(t)|\}$ is difficult, because it depend on the distribution of $(e_t, t \in \mathbb{Z})$ within each regime. So in the following corollary, and easier sufficient condition is given.

Corollary 32 *Consider the model* $(II - 4.1)$ *and assume that* $\rho(A) < 1$ *where* $A = \sigma^2 \sum_{j=2}^{p} \mathbb{P}^{(j)}\left(\underline{c}_j^{(2)}\right)$ *where* $\underline{c}_j^{(2)} = (c_j^2(i),\ i = 1, ...d)$*. Then if*

$$\underline{1}' A \underline{1} < \frac{d}{\sigma^2 d + 1}. \tag{II-4.2}$$

with $\underline{1} = (1, 1, ..., 1) \in \mathbb{R}^d$ *, model* $(II - 4.1)$ *has a unique strictly, ergodic, second−order stationary and invertible solution.*

Proof. It can be shown that under the assumption of the corollary, model $(II - 4.1)$ admits an ergodic, strictly and second−order stationary solution (see Bibi and Ghazel [10]). Moreover, the stationary solution is a weak white noise with mean 0 and variance $R(0) = \sigma^2 \underline{1}'\left(I_{(d)} - A\right)^{-1}\underline{\pi}$. On the other hand,

$$\begin{aligned}E\{\log|\beta(t)|\} &= \frac{1}{2}E\left\{\log\beta^2(t)\right\} \\ &\leq \frac{1}{2}\log E\left\{\beta^2(t)\right\} \\ &= \frac{1}{2}\log E\left\{\sum_{j=2}^{p} c_j^2(s_t) X_{t-j}^2\right\} = \frac{1}{2}\log\left\{\sigma^2 \underline{e}'(I_{(d)} - A)^{-1} A\underline{\pi}\right\}\end{aligned}$$

so $\sigma^2\underline{1}'(I_{(d)} - A)^{-1}A\underline{\pi} < 1$ if $\sigma^2\underline{1}'(I_{(d)} - A)^{-1}A\underline{1} < 1$. The last inequality hold true if for example $\sigma^2 A\underline{1} \prec \frac{1}{d}(I_{(d)} - A)\underline{1}$ (where the inequality $M \prec N$ denotes the elementwise relation $m_{ij} \leq n_{ij}$ for all i and j) and hence $E\{\log|\beta(t)|\} < 0$ if $(II - 4.2)$ is satisfied or whether

$$\sigma^2 \sum_{j=2}^{p}\sum_{i=1}^{d} c_j^2(i) < \frac{d}{\sigma^2 d + 1}.$$

In particular for $p=2$, $d=2$,and $p_{11}=q, p_{12}=1-q, p_{21}=1-q$, $p_{22}=q=0.5$ with $e_t \rightsquigarrow N(0,(0.5)^2)$, the strict, second−order stationarity and invertibility zones are shown in the following figure

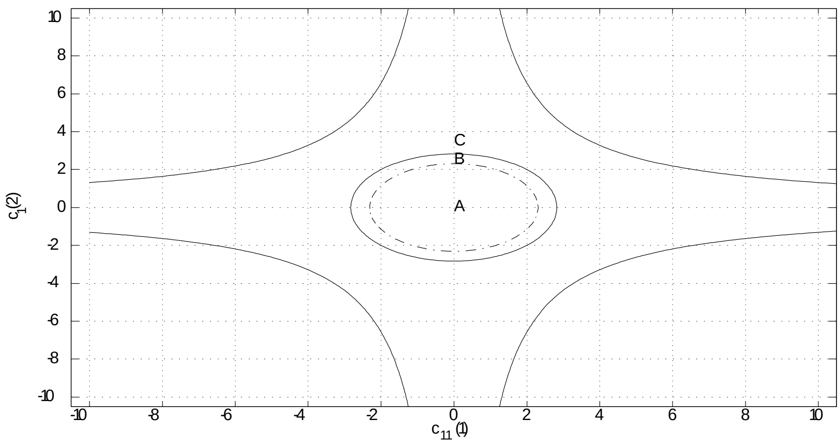

Fig 1: A:Invertibility,$A\cup B$: Second order stationarity and $A\cup B\cup C$: Strict stationarity

■

2.5 Geometric ergodicity and $\beta-$mixing properties

Mixing conditions describe some type of asymptotic independence, which can be useful in proving many limit theorems e.g. for central limit theorem, law of large numbers and for sample covariance function that can be employed to derive the consistency, asymptotic normality and the law of iterated logarithm of some estimation procedures. We first recall a few definitions which will be used throughout. Noting firstly that the extended process $\left(\underline{\widetilde{Y}}_t=(\underline{Y}_t', s_t)',\ t\in\mathbb{Z}\right)$ can be rewritten as

$$\underline{\widetilde{Y}}_t=\widetilde{\Gamma}_{s_t}(e_t)\underline{\widetilde{Y}}_{t-1}+\underline{\widetilde{\eta}}_{s_t}(e_t), \tag{II-5.1}$$

where $\underline{\widetilde{\eta}}_{s_t}(e_t):=\left(\underline{\eta}'_{s_t}(e_t), s_t\right)'$ and $\widetilde{\Gamma}_{s_t}(e_t)$ is an $(s+1)\times(s+1)-$random matrix. Let $\underline{\widetilde{Y}}_0$ be an arbitrarily specified random vector in $\mathbb{R}^{s+1}$ independent of $\left(\left(\widetilde{\Gamma}_{s_t}(e_t), \underline{\widetilde{\eta}}_{s_t}(e_t)\right)\right)_{t\in\mathbb{Z}}$, then $\left(\underline{\widetilde{Y}}_t, t\in\mathbb{Z}\right)$ is a Markov chain with a state space $\mathbb{R}^s\times\mathbb{S}$ and $n-$step transition probability $P^{(n)}\left(\underline{\widetilde{y}}, C\right)=P\left(\underline{\widetilde{Y}}_n\in C|\underline{\widetilde{Y}}_0=\underline{\widetilde{y}}\right)$ and invariant probability measure $\pi(C)=\int P\left(\underline{\widetilde{y}}, C\right)\pi\left(d\underline{\widetilde{y}}\right)$ for any Borelean set

$C\in\mathcal{B}_{\mathbb{R}^{s+1}}$. The chain is $\phi-$irreducible if, for some non trivial measure ϕ on $(\mathbb{R}^{s+1},\mathcal{B}_{\mathbb{R}^{s+1}})$ such that $\forall C\in\mathcal{B}_{\mathbb{R}^{s+1}}$, $\phi(C)>0\Longrightarrow\exists n>0$, $P^{(n)}\left(\underline{\widetilde{y}},C\right)>0$. It called a Feller Markov chain if the function $E\left\{g(\underline{\widetilde{Y}}_t)|\underline{\widetilde{Y}}_{t-1}=\underline{\widetilde{y}}\right\}$ is continuous for every bounded and continuous function g on $\mathbb{R}^{s+1}$. The chain is called geometrically ergodic if there exists some probability measure π on $\mathcal{B}_{\mathbb{R}^{s+1}}$ and a positive real number $c\in\,]0,1]$ such that $c^{-n}\left\|P^{(n)}\left(\underline{\widetilde{y}},.\right)-\pi\left(.\right)\right\|_V\to 0$ as $n\to+\infty$ where $\|.\|_V$ denotes the total variation norm. It is called exponential $\beta-$mixing if $\int\left\|P^{(n)}\left(\underline{\widetilde{y}},.\right)-\pi\left(.\right)\right\|_V\pi\left(d\underline{\widetilde{y}}\right)\leq kc^n$ for some constant $k>0$.

Recalling here, that one of the most popular mixing measure for stationary processes is $\alpha-$mixing coefficients defined by

$$\alpha_Y\left(k\right)=\sup_{A\in\sigma\left(\underline{\widetilde{Y}}_t,t\leq 0\right),B\in\sigma\left(\underline{\widetilde{Y}}_t,t\geq k\right)}\left|P\left(A\cap B\right)-P\left(A\right)P\left(B\right)\right|.$$

A closely mixing measure that we found in literature is $\beta-$mixing coefficients which are defined by

$$\beta_Y\left(k\right)=\sup_{A_i\in\sigma\left(\underline{\widetilde{Y}}_t,t\leq 0\right),B_j\in\sigma\left(\underline{\widetilde{Y}}_t,t\geq k\right)}\sum_{i=1}^{I}\sum_{j=1}^{J}\left|P\left(A_i\cap B_j\right)-P\left(A_i\right)P\left(B_j\right)\right|$$

where $\{A_i,i=1,...,I\}$ and $\{B_j,i=1,...,J\}$ are finite partition of the sample space Ω. The chain $\left(\underline{\widetilde{Y}}_t\right)_{t\in\mathbb{Z}}$ is called $\alpha-$mixing (resp. $\beta-$mixing) if $\lim\limits_{k\to\infty}\alpha_Y\left(k\right)=0$ (resp. $\lim\limits_{k\to\infty}\beta_Y\left(k\right)=0$). It can be seen that $\alpha_Y\left(k\right)\leq\beta_Y\left(k\right)$ so that $\beta-$mixing implies the $\alpha-$mixing. Consequently, and according to the above definitions, exponential $\beta-$mixing and geometric ergodicity are equivalent for Markov chains. Moreover, for any ergodic chain with invariant probability measure π we have $\beta_Y\left(k\right)=\int\left\|P^{(k)}\left(\underline{\widetilde{y}},.\right)-\pi\left(.\right)\right\|_V\pi\left(d\underline{\widetilde{y}}\right)$, so if the chain is geometrically ergodic $\beta_Y\left(k\right)=O\left(c^k\right)$, $c\in\,]0,1]$.

In what follows and in order to make the notation shorter, we shall set $\mathbf{\Gamma}_t(k):=\prod\limits_{i=0}^{k-1}\Gamma_{s_{t-i}}(e_{t-i})$ (rep. $\widetilde{\mathbf{\Gamma}}_t(k):=\prod\limits_{i=0}^{k-1}\widetilde{\Gamma}_{s_{t-i}}(e_{t-i})$) . For the representation $(II-5.1)$ we have the following results.

Lemme 33 *Let $\gamma_L(\Gamma)$ (resp. $\gamma_L(\widetilde{\Gamma})$) be the Lyapunov exponent associated with the respect to the sequence of random matrices $(\Gamma_{s_t}(e_t),t\in\mathbb{Z})$ (resp. $(\widetilde{\Gamma}_{s_t}(e_t),t\in\mathbb{Z})$). Then if $\gamma_L(\Gamma)<0\Longrightarrow\gamma_L(\widetilde{\Gamma})<0$ and hence the representation $(II-5.1)$ has a unique, causal, ergodic and strictly stationary solution given by*

$$\underline{\widetilde{Y}}_t=\sum_{k=1}^{\infty}\widetilde{\mathbf{\Gamma}}_t(k)\underline{\widetilde{\eta}}_{s_{t-k}}(e_{t-k})+\underline{\widetilde{\eta}}_{s_t}(e_t)\qquad\text{(II-5.2)}$$

Proof. Straightforward and hence omitted. ■

Lemme 34 *Let $\gamma_L(\Gamma)$ be the Lyapunov exponent associated with respect to the sequence of random matrices $(\Gamma_{s_t}(e_t), t \in \mathbb{Z})$ and assume that $\mu_\delta = E\left\{|e_0|^\delta\right\} < +\infty$, for some $\delta > 0$. Then if $\gamma_L(\Gamma) < 0$, there is $\delta^* \in]0,1[$ such that $E\left\{|X_t|^{\delta^*}\right\} < +\infty$.*

Proof. Noting that because $\gamma_L(\Gamma) < 0$, then equation $(II-2.4)$ has a unique, causal, strictly stationary and ergodic solution given by the series $(II-2.6)$. Hence, we have to show that if $\gamma_L(\Gamma) < 0$ there is $\delta^* \in]0,1[$ and $t_0 > 0$ such that $E\left\{\|\mathbf{\Gamma}_t(t_0)\|^{\delta^*}\right\} < 1$. Indeed, let us consider an absolute norm, then by definition of $\gamma_L(\Gamma)$, there is an integer $t_0 \geq 1$ such that $E\left\{\log\|\mathbf{\Gamma}_t(t_0)\|\right\} < 0$. On the other hand, by the submultiplicativity of $|.|^\delta$, we have

$$\begin{aligned}
E\left\{\|\mathbf{\Gamma}_t(t_0)\|^\delta\right\} &\leq E\left\{\left\|\prod_{i=0}^{t_0-1}\left|\Gamma_{s_{t-i}}(e_{t-i})\right|^\delta\right\|\right\} = \left\|E\left\{\prod_{i=0}^{t_0-1}\left|\Gamma_{s_{t-i}}(e_{t-i})\right|^\delta\right\}\right\| \\
&= \left\|E\left\{E\left\{\prod_{i=0}^{t_0-1}\left|\Gamma_{s_{t-i}}(e_{t-i})\right|^\delta |s_t, s_{t-1}, ..., s_{t-t_0+1}\right\}\right\}\right\| \\
&\leq \left\|\mathbb{I}'_{(s)}\mathbb{P}^{t_0}\left(\underline{\Gamma}_\delta\right)\underline{\Pi}(I_{(s)})\right\| < +\infty.
\end{aligned}$$

where $\underline{\Gamma}_\delta = \left(E\left\{|\Gamma_0(i) + e_0\Gamma_1(i)|^\delta\right\}, 1 \leq i \leq d\right)$. Now, for any $t > 0$, let $f(t) = E\left\{\|\mathbf{\Gamma}_t(t_0)\|^t\right\}$, since $f'(0) = E\left\{\log\|\mathbf{\Gamma}_t(t_0)\|\right\} < 0$, then $f(t)$ decrease in a neighborhood of 0 and since $f(0) = 1$, it follows that there exists $\delta^* \in]0,1[$ such that $E\left\{\|\Gamma_t(t_0)\|^{\delta^*}\right\} < 1$. Since $\|\underline{Y}_t\| \leq \sum_{k\geq 1}\left\|\mathbf{\Gamma}_t(k)\underline{\eta}_{s_{t-k}}(e_{t-k})\right\| + \left\|\underline{\eta}_{s_t}(e_t)\right\|$ and because $\delta^* \in]0,1[$, we obtain $\|\underline{Y}_t\|^{\delta^*} \leq \sum_{k\geq 1}\left\|\mathbf{\Gamma}_t(k)\underline{\eta}_{s_{t-k}}(e_{t-k})\right\|^{\delta^*} + \left\|\underline{\eta}_{s_t}(e_t)\right\|^{\delta^*}$ and by the dominated convergence theorem together with the fact that $\limsup_{k\to\infty}\left\|\mathbf{\Gamma}_t(k)\underline{\eta}_{s_{t-k}}(e_{t-k})\right\|^{\frac{1}{k}} \leq \exp\left\{\gamma_L(\Gamma)\right\} < 1$, the last series converges in $\mathbb{L}_{\delta^*}$ and hence the result follows. ■

A similar result to the Proposition 7 for the representation $(II-2.4)$ is given in following lemma

Lemme 35 *Assume that $\rho\left(\mathbb{P}\left(\underline{\Gamma}_\delta\right)\right) < 1$, then $\gamma_L(\Gamma) < 0$ (consequently $\gamma_L(\widetilde{\Gamma}) < 0$) and hence equation $(II-2.4)$ (rep. equation $II-5.1$) has a unique, strictly stationary and ergodic solution given by the Series $(II-2.6)$ (resp. by $II-5.2$). Moreover, for all $t \in \mathbb{Z}$, the sequence $\left(\mathbf{\Gamma}_t(k)\underline{y}\right)_{k\geq 1}$ (resp. $\left(\widetilde{\mathbf{\Gamma}}_t(k)\underline{\widetilde{y}}\right)_{k\geq 1}$ converges almost surely to $\underline{O}$ for any vector $\underline{y} \in \mathbb{R}^s$ (resp. $\underline{\widetilde{y}} \in \mathbb{R}^{s+1}$).*

Proof. Straightforward and hence omitted. ■

Noting that one of the most well-known condition used in establishing stationarity and geometric ergodicity for discrete-time, aperiodic and $\phi-$irreducible Markov chains is the drift condition developed by Meyn and Tweedie [59].

The chain $\left(\underline{\widetilde{Y}}_t, t \in \mathbb{Z}\right)$ is called hold the drift condition, if there exist a positive function V on $\mathbb{R}^s \times \mathbb{S}$, a compact set C of $\mathbb{R}^s \times \mathbb{S}$ with nonempty interior and real numbers $\alpha > 0$, $\gamma > 0$ and $\rho \in]0, 1[$ such that

(i) $E\left\{V(\underline{\widetilde{Y}}_t)|\underline{\widetilde{Y}}_{t-1} = \underline{\widetilde{y}}\right\} \leq \rho V(\underline{\widetilde{y}}) - \alpha$ if $\underline{\widetilde{y}} \in C$

(ii) $E\left\{V(\underline{\widetilde{Y}}_t)|\underline{\widetilde{Y}}_{t-1} = \underline{\widetilde{y}}\right\} \leq \gamma$ if $\underline{\widetilde{y}} \notin C$

In order to derive the geometric ergodicity of the chain $\left(\underline{\widetilde{Y}}_t, t \in \mathbb{Z}\right)$ we make the following assumptions.

The marginal distribution of e_t is absolutely continuous with respect to the Lebesgue measure λ. The support of e_t is defined by its strictly positive density f_e and contains an open set around zero.

$\mu_\delta = E\left\{|e_0|^\delta\right\} < +\infty$ for some $0 < \delta \leq 1$.

$\rho\left(\mathbb{P}\left(\underline{\Gamma}_\delta\right)\right) < 1$.

$\mu_s = E\left\{|e_0|^s\right\} < +\infty$ for some $s \geq 1$

$\rho\left(\mathbb{P}\left(\|\underline{\Gamma}\|^s\right)\right) < 1$ where $\|\underline{\Gamma}\|^s = \left(E\left\{\|\underline{\Gamma}_k\left(e_0\right)\|^s\right\}, k = 1, ..., d\right)$

Lemme 36 *Under* **A.0 – A.2**, $\left(\underline{\widetilde{Y}}_t, t \in \mathbb{Z}\right)$ *is Feller,* $(\lambda \otimes \nu)-$*irreducible and aperiodic chain where* ν *is a counting measure on* $\mathbb{S}$ *and therefore every compact set is a petit set.*

Proof. For any bounded and continuous function g on $\mathbb{R}^s \times \mathbb{S}$, and by the Lebesgue dominated convergence theorem, the function $E\left\{g(\underline{\widetilde{Y}}_t)|\underline{\widetilde{Y}}_{t-1} = \underline{\widetilde{y}}\right\} = E\left\{g\left(\widetilde{\Gamma}_k(e_t)\underline{\widetilde{y}} + \underline{\widetilde{\eta}}_k(e_t)\right)\right\}$, $k \in \mathbb{S}$ is continuous in $\underline{\widetilde{y}}' := \left(\underline{y}', k\right) \in \mathbb{R}^s \times \mathbb{S}$, thus the chain $\left(\underline{\widetilde{Y}}_t, t \in \mathbb{Z}\right)$ is Feller. $\left(\underline{\widetilde{Y}}_t, t \in \mathbb{Z}\right)$ is $(\lambda \otimes \nu)-$irreducible, indeed, let $B = B_1 \times B_2 \in \mathcal{B}_{\mathbb{R}^s} \times \mathbb{S}$ be such that $(\lambda \otimes \nu)(B) > 0$. Since B_2 contains at least one integer $k \in \mathbb{S}$, it is enough to prove that there exists t such that $P\left(\underline{\widetilde{Y}}_t' \in B_1 \times \{k\} |\underline{\widetilde{Y}}_0' = \left(\underline{y}', l\right)'\right) > 0$ for all $k, l \in \mathbb{S}$. By definition

$$
\begin{aligned}
P\left(\underline{\widetilde{Y}}_1' \in B_1 \times \{k\} |\underline{\widetilde{Y}}_0' = \left(\underline{y}', l\right)\right) &= P\left(\underline{Y}_1 \in B_1, s_1 = k|\underline{Y}_0 = \underline{y}, s_0 = l\right) \\
&= \left\{ \begin{array}{l} P\left(\underline{Y}_1 \in B_1|s_1 = k, s_0 = l, \underline{Y}_0 = \underline{y}\right) \\ P\left(s_1 = k|s_0 = l, \underline{Y}_0 = \underline{y}\right) \end{array} \right\} \\
&= p_{lk} P\left(\Gamma_k(e_1)\underline{y} + \underline{\eta}_k(e_1) \in B_1\right).
\end{aligned}
$$

Moreover, it can be easy seen that $\Gamma_k(e_1)\underline{y} + \underline{\eta}_k(e_1) \in B_1$ can be transformed into $g_k(e_1) = a_k e_1^2 + b_k e_1 \in C_k$ where a_k, b_k are real constants and C_k is some Borelean on which the equation $y = g_k(x)$ has a unique solution $x = g_k^{-1}(y)$. So

$$\begin{aligned} P\left(\widetilde{\underline{Y}}_1' \in B_1 \times \{k\} \,|\widetilde{\underline{Y}}_0' = \left(\underline{y}', l\right)\right) &= p_{lk} P\left(g_k(e_1) \ \in C_k\right) \\ &= \int_{C_k} p_{lk} f_e\left(g_k^{-1}(y)\right) \left|\frac{d}{dy} g_k^{-1}(y)\right| dy \end{aligned}$$

which is strictly greater than 0. Analogously, it can be easily seen that $\left(\widetilde{\underline{Y}}_t, t \in \mathbb{Z}\right)$ is aperiodic. ■

Lemme 37 *Under* **A**.**1** − **A**.**2** *the chain* $\left(\widetilde{\underline{Y}}_t, t \in \mathbb{Z}\right)$ *holds the drift condition with* $V\left(\widetilde{\underline{y}}\right) = \left\|\widetilde{\underline{y}}\right\|^{\delta} + 1$.

Proof. From $(II - 5.1)$ we have by recursion

$$\left\|\widetilde{\underline{Y}}_t\right\|^{\frac{1}{t}} \leq \sum_{k=0}^{t-1} \left\|\widetilde{\mathbf{\Gamma}}_t(k)\right\|^{\frac{1}{t}} \left\|\widetilde{\underline{\eta}}_{s_{t-k}}(e_{t-k})\right\|^{\frac{1}{t}} + \left\|\widetilde{\mathbf{\Gamma}}_t(t+1)\right\|^{\frac{1}{t}} \left\|\widetilde{\underline{Y}}_0\right\|^{\frac{1}{t}}, \qquad \text{(II-5.3)}$$

such that by **A**.**2** we have almost surely $\frac{1}{t}\log\left\|\widetilde{\mathbf{\Gamma}}_t(t)\right\| \to \gamma_L\left(\widetilde{\Gamma}\right) < 0$ and hence $\left\|\widetilde{\mathbf{\Gamma}}_t(t)\right\|^{\frac{1}{t}} \to e^{\gamma_L\left(\widetilde{\Gamma}\right)} < 1$. On the other hand, the random variables $\left\|\widetilde{\mathbf{\Gamma}}_t(t)\right\|^{\frac{1}{t}}$ being almost surely bounded by $\sup_t \left\|\Gamma_{s_t}(e_t)\right\|$, we have by dominated convergence theorem for all $x \in \mathbb{S}$: $\lim_{t\to\infty} E\left\{\left\|\widetilde{\mathbf{\Gamma}}_t(t)\mathbb{I}_{\{s_1=x\}}\right\|^{\frac{1}{t}}\right\} < 1$. It follows that there is a positive integer $p > \frac{1}{\delta}$ and a positive constant $K < 1$ such that

$$\alpha := \sup_{x\in\mathbb{S}} E\left\{\left\|\widetilde{\mathbf{\Gamma}}_p(p)\mathbb{I}_{\{s_1=x\}}\right\|^{\frac{1}{p}}\right\} \leq \rho\left(\mathbb{P}\left(\underline{\Gamma}_{\frac{1}{p}}\right)\right) \leq K.$$

Taking now the conditional expectation in both sides of inequality $(II - 5.3)$, we obtain $E\left\{\left\|\widetilde{\underline{Y}}_p\right\|^{\frac{1}{p}} |\widetilde{\underline{Y}}_0 = \widetilde{\underline{y}}\right\} \leq \alpha \left\|\widetilde{\underline{y}}\right\|^{\frac{1}{p}} + \gamma$

where $\gamma = E\left\{\sum_{k=0}^{p-1} \left\|\widetilde{\mathbf{\Gamma}}_p(k)\right\|^{\frac{1}{p}} \left\|\widetilde{\underline{\eta}}_{s_{p-k}}(e_{p-k})\right\|^{\frac{1}{p}} |\widetilde{\underline{Y}}_0 = \widetilde{\underline{y}}\right\}$. The last inequality implies we have $E\left\{V(\widetilde{\underline{Y}}_p)|\widetilde{\underline{Y}}_0 = \widetilde{\underline{y}}\right\} \leq \alpha V(\widetilde{\underline{y}}) + \gamma + 1 - \alpha$. Set $\rho = \alpha + \frac{1-\alpha}{2}$ and define the compact $C = \{\widetilde{\underline{y}}' = \left(\underline{y}', x\right) \in \mathbb{R}^s \times \mathbb{S} : \rho V(\widetilde{\underline{y}}) \leq \alpha V(\widetilde{\underline{y}}) + 1 + \delta\}$. Then, it is easily seen that $0 < \rho < 1$ and thus the drift condition hold true. ■

Lemme 38 *Under* **B**.**1** − **B**.**2** *the chain* $\left(\widetilde{\underline{Y}}_t, t \in \mathbb{Z}\right)$ *holds the drift condition with* $V\left(\widetilde{\underline{y}}\right) = \left\|\widetilde{\underline{y}}\right\|^{s} + 1$ *for some* $s \geq 1$.

Proof. We have

$$\left\|\underline{\widetilde{Y}}_t\right\|^s \leq \left(\sum_{k=0}^{t-1}\left\|\widetilde{\mathbf{\Gamma}}_t(k)\right\|\left\|\underline{\widetilde{\eta}}_{s_{t-k}}(e_{t-k})\right\| + \left\|\widetilde{\mathbf{\Gamma}}_t(t+1)\right\|\left\|\underline{\widetilde{Y}}_0\right\|\right)^s,$$

taking expectation for any $\underline{\widetilde{y}}_0 = (\underline{y}, x) \in \mathcal{B}_{\mathbb{R}^s} \times \mathbb{S}$ and from the $\mathbb{L}_s$ norm inequality we obtain

$$\begin{aligned}
\left(E\left\{\left\|\underline{\widetilde{Y}}_t\right\|^s\right\}\right)^{1/s} &\leq \left(E\left\{\left(\sum_{k=0}^{t-1}\left\|\widetilde{\mathbf{\Gamma}}_t(k)\right\|\left\|\underline{\widetilde{\eta}}_{s_{t-k}}(e_{t-k})\right\| + \left\|\widetilde{\mathbf{\Gamma}}_t(t+1)\right\|\left\|\underline{\widetilde{Y}}_0\right\|\right)^s\right\}\right)^{1/s} \\
&\leq \left\{\begin{array}{l}\left(E\left\{\sum_{k=0}^{t-1}\left\|\widetilde{\mathbf{\Gamma}}_t(k)\right\|\left\|\underline{\widetilde{\eta}}_{s_{t-k}}(e_{t-k})\right\|\right\}^s\right)^{1/s} \\ +\left(E\left\{\left\|\widetilde{\mathbf{\Gamma}}_t(t+1)\right\|^s\left\|\underline{\widetilde{Y}}_0\right\|^s\right\}\right)^{1/s}\end{array}\right\} \\
&\leq \alpha\left\|\underline{\widetilde{Y}}_0\right\| + \gamma
\end{aligned}$$

where $\alpha = \sup_{s\in\mathbb{S}}\left(E\left\{\left\|\widetilde{\mathbf{\Gamma}}_t(t+1)\right\|^s |s_t = x\right\}\right)^{1/s}$ and

$$\gamma = \left(E\left\{\sum_{k=0}^{t-1}\left\|\widetilde{\mathbf{\Gamma}}_t(k)\right\|\left\|\underline{\widetilde{\eta}}_{s_{t-k}}(e_{t-k})\right\|\right\}^s\right)^{1/s}.$$

Hence $E\left\{\left\|\underline{\widetilde{Y}}_t\right\|^s\right\} \leq \left(\alpha\left\|\underline{\widetilde{Y}}_0\right\| + \gamma\right)^s$. Since $\alpha \leq \rho\left(\mathbb{P}\left(\underline{\Gamma}_1^s\right)\right) < 1$, the conclusion is the same as in proof of Lemma 37. ■

Theorem 39 *Under conditions* $\mathbf{A.0}-\mathbf{A.2}$ *(rep.* $\mathbf{A.0}, \mathbf{B1}$ *and* $\mathbf{B.2}$*) the chain* $\left(\underline{\widetilde{Y}}_t, t\in\mathbb{Z}\right)$ *is geometrically ergodic. Moreover, if* $\underline{\widetilde{Y}}_t$ *is starting with the stationary distribution, then its first component process is strictly stationary and* $\beta-$*mixing with exponential decay rates.*

Proof. The result follows from Lemmas 36−38 and Theorem 15.0.1 by Meyn and Tweedie [59]. ■

Remark 40 *In standard case, with i.i.d. innovation, Pham [64] studied a bilinear Markovian representation in the form* $\underline{Y}_t = A(e_t)\underline{Y}_{t-1} + B(e_t)$ *of general bilinear model in which* $(A(e_t), B(e_t))$ *being i.i.d. pairs of random matrix and vector, independent of* $\underline{Y}_s$ *for* $s < t$. *He establishes the geometric ergodicity under condition involving the symmetric tensor power of matrix* $A(e_t)$ *and the rank of complicated matrix. However, our conditions are more flexible.*

Remark 41 *In independent switching bilinear model* $BL(1,0,1,1)$, *Aknouche and Rabehi [2] have established the geometric ergodicity under the strict stationarity assumption with drift condition* $V(\underline{x}) = \|\underline{x}\|^\delta + 1$, $\delta \in]0,1]$. *The case* $V(\underline{x}) = \|\underline{x}\|^\delta + 1$, $\delta > 1$, *is not established. So our result is more general.*

2.5.1 Application to a family of $MS - GARCH\,(1,1)$ processes

The result of Theorem 39 can be extended to the so−called augmented $MS - GARCH\,(1,1)$ processes defined by $\epsilon_t = \sqrt{h_t} e_t$ where $(e_t, t \in \mathbb{Z})$ is an $i.i.d$ process independent of h_0 with $E\{e_t\} = 0$, $E\{e_t^2\} = 1$ and satisfying the condition **A**0. The volatility process $(h_t, t \in \mathbb{Z})$ is assumed to verify

$$f(h_t) = h_{s_t}(\eta_t) f(h_{t-1}) + g_{s_t}(\eta_t) \tag{II-5.4}$$

where $f : \mathbb{R}_+ \to \mathbb{R}$ is strictly monotonic and continuous function and $\eta_t \in \Im_{t-1}^{(e)} = \sigma((e_{t-j}, s_t), j \geq 1)$. In this subsection, we shall derive sufficient condition for the existence of higher-order moment and $\beta-$mixing with exponential rate for several classes of augmented $MS - GARCH\,(1,1)$ processes. Note that all examples considered here were examined by Carrasco and Chen [16] in standard cases.

Proposition 42 *Consider the model $(II - 5.4)$, and assume that $g_{s_t}(\eta_t)$ and $h_{s_t}(\eta_t)$ are polynomials function in η_t and there is an integer $p \geq 1$ such that*

$$\rho(\mathbb{P}(\underline{h}^p)) < 1 \text{ and } \underline{1}'\Pi\left(\underline{g}^p\right) < +\infty \tag{II-5.5}$$

where $\underline{h}^p = (E\{h_k^p(\eta_t)\}, k = 1, ..., d)$ and $\underline{g}^p = (E\{g_k^p(\eta_t)\}, k = 1, ..., d)$. Then

(i) $\widetilde{\underline{h}}_t = (h_t, s_t)'$ is Markovian chain geometrically ergodic. If $\left(\widetilde{\underline{h}}_t, t \in \mathbb{Z}\right)$ is initialized from the invariant measure then $(h_t, t \in \mathbb{Z})$ and $(\epsilon_t, t \in \mathbb{Z})$ are strictly stationary and $\beta-$mixing with exponential rate and $E\{|f(h_t)|^p\} < +\infty$.

(ii) Moreover, if $f(h_t) = h_t^q$, $q > 0$ and $E\left\{|e_t|^{2pq}\right\} < +\infty$ then $E\left\{|\epsilon_t|^{2pq}\right\} < +\infty$.

Proof.

(i) The geometric ergodicity and mixing property of the chain $(f(h_t), s_t)'$ may be deduced from Theorem 39 with $X_t = f(h_t), \Gamma_{s_t}(e_t) = h_{s_t}(e_t)$ and $\underline{\eta}_{s_t}(e_t) = g_{s_t}(e_t)$. So, since f is invertible, then h_t is also $\beta-$mixing with exponential rate and so ϵ_t which can be regarded as a hidden Markov process.

(ii) This statement follows from the fact that $E\left\{|\epsilon_t|^{2qp}\right\} = E\{|h_t|^{qp}\} E\left\{|e_t|^{2qp}\right\}$.

■

Corollary 43 $[MS - LGARCH\,(1,1)]$ *Consider the model*

$$h_t = \omega\,(s_t) + \beta\,(s_t)\,h_{t-1} + \alpha\,(s_t)\,(e_{t-1} - c\,(s_t))^2\,h_{t-1}$$

in which $\omega\,(.) > 0$, $\beta\,(.) \geq 0$ and $\alpha(.) \geq 0$. The case $c\,(.) = 0$ correspond to classical $MS-GARCH(1,1)$ model. Assume that there is an integer $p \geq 1$ such that $\rho\left(\mathbb{P}\left(\underline{C}_1^{(p)}\right)\right) < 1$ where $\underline{C}_1^{(p)} = \left(E\left\{\left(\beta\,(k) + \alpha\,(k)\,(e_t - c\,(k))^2\right)^p\right\}, k = 1, ..., d\right)$. If $\left(\widetilde{\underline{h}}_t, t \in \mathbb{Z}\right)$ is initialized from the invariant measure, then $(h_t, t \in \mathbb{Z})$ and $(\epsilon_t, t \in \mathbb{Z})$ are strictly stationary and $\beta-$mixing with exponential rate. Moreover $E\,\{h_t^p\} < +\infty$ and $E\left\{|\epsilon_t|^{2p}\right\} < +\infty$.

Proof. The result follows from the proposition 42 with $f\,(h_t) = h_t, h_{s_t}\,(\eta_t) = \beta\,(s_t) + \alpha\,(s_t)\,(\eta_t - c\,(s_t))^2$ with $\eta_t = e_{t-1}$ and $g_{s_t}\,(\eta_t) = \omega\,(s_t)$, which satisfies the conditions of the proposition 42. ■

Corollary 44 $[MS - MGARCH\,(1,1)]$ *Consider the model*

$$\log\,(h_t) = \omega\,(s_t) + \beta\,(s_t)\log\,(h_{t-1}) + \alpha\,(s_t)\log e_{t-1}^2.$$

Assuming that there is an integer $p \geq 1$ such that $E\left\{|\log e_t^2|^p\right\} < +\infty$ and $\rho\left(\mathbb{P}\left(\underline{\beta}\right)\right) < 1$ where $\underline{\beta} = (\beta(k), k = 1, ..., d)$. If $\left(\widetilde{\underline{h}}_t, t \in \mathbb{Z}\right)$ is initialized from the invariant measure, then $(h_t, t \in \mathbb{Z})$ and $(\epsilon_t, t \in \mathbb{Z})$ are strictly stationary, $\beta-$mixing with exponential rate and $E\,\{|\log\,(h_t)|^p\} < +\infty$ and $E\left\{|\log \epsilon_t^2|^p\right\} < +\infty$.

Proof. The result follows from the proposition 42 with $f\,(h_t) = \log\,(h_t)$, $h_{s_t}\,(\eta_t) = \beta\,(s_t)$ and $g_{s_t}\,(\eta_t) = \omega\,(s_t) + \alpha\,(s_t)\log \eta_t^2$ with $\eta_t = e_{t-1}$. ■

Corollary 45 $[MS - EGARCH\,(1,1)]$ *Consider the model*

$$\log\,(h_t) = \omega\,(s_t) + \beta\,(s_t)\log\,(h_{t-1}) + \alpha\,(s_t)\,(|e_{t-1}| + \gamma\,(s_t)\,e_{t-1})$$

where $\gamma\,(.) \neq 0$. Assuming that there is an integer $p \geq 1$ such that $E\left\{|e_t^2|^p\right\} < +\infty$ and $\rho\left(\mathbb{P}\left(\underline{\beta}\right)\right) < 1$ where $\underline{\beta} = (\beta(k), k = 1, ..., d)$. If $\left(\widetilde{\underline{h}}_t, t \in \mathbb{Z}\right)$ is initialized from the invariant measure, then $(h_t, t \in \mathbb{Z})$ and $(\epsilon_t, t \in \mathbb{Z})$ are strictly stationary and $\beta-$mixing with exponential rate. Moreover, if $E\left\{|\log e_t^2|^p\right\} < +\infty$ then $E\left\{|\log \epsilon_t^2|^p\right\} < +\infty$.

Proof. The result follows from the proposition 42 with $f\,(h_t) = \log\,(h_t)$, $h_{s_t}\,(\eta_t) = \beta\,(s_t)$ and $g_{s_t}\,(\eta_t) = \omega\,(s_t) + \alpha\,(s_t)\,\eta_t$, with $\eta_t = |e_{t-1}| + \gamma\,(s_t)\,e_{t-1}$. ■

Corollary 46 $[MS-VGARCH\,(1,1)]$ *Consider the model*

$$h_t=\omega\,(s_t)+\beta\,(s_t)\,h_{t-1}+\alpha\,(s_t)\,(e_{t-1}-c\,(s_t))^2$$

in which $\omega\,(.)>0$, $\beta\,(.)\geq 0$ *and* $\alpha(.)\geq 0$. *Assuming that there is an integer* $p\geq 1$ *such that* $E\left\{|e_t|^{2p}\right\}<+\infty$ *and* $\rho\left(\mathbb{P}\left(\underline{\beta}\right)\right)<1$ *where* $\underline{\beta}=(\beta(k),k=1,...,d)$. *If* $\left(\underline{\widetilde{h}}_t,t\in\mathbb{Z}\right)$ *is initialized from the invariant measure, then* $(h_t,t\in\mathbb{Z})$ *and* $(\epsilon_t,t\in\mathbb{Z})$ *are strictly stationary,* $\beta-$*mixing with exponential rate,* $E\left\{|h_t|^p\right\}<+\infty$ *and* $E\left\{|\epsilon_t^2|^p\right\}<+\infty$.

Proof. The result follows from the proposition 42 with $f\,(h_t)=h_t, h_{s_t}\,(\eta_t)=\beta\,(s_t)$ and $g_{s_t}\,(\eta_t)=\omega\,(s_t)+\alpha\,(s_t)\,(\eta_t-c\,(s_t))^2$ with $\eta_t=e_{t-1}$. ■

Corollary 47 $[MS-GJR-GARCH\,(1,1)]$ *Consider the model*

$$h_t=\omega\,(s_t)+\beta\,(s_t)\,h_{t-1}+\alpha_1\,(s_t)\,e_{t-1}^2h_{t-1}+\alpha_2\,(s_t)\max(0,-e_{t-1})^2h_{t-1}$$

in which $\omega\,(.)>0$, $\beta\,(.)\geq 0$ *and* $\alpha_1(.)>0$ *and* $\alpha_1(.)+\alpha_2(.)\geq 0$. *Assuming that there is an integer* $p\geq 1$ *such that* $\rho\left(\mathbb{P}\left(\underline{C}_2^{(p)}\right)\right)<1$ *where* $\underline{C}_2^{(p)}=\left(E\left\{(\beta\,(k)+\alpha_1\,(k)\,e_t^2+\alpha_2\,(k)\max(0,-e_t)^2)^p\right\},k=1,...,d\right)$. *If* $\left(\underline{\widetilde{h}}_t,t\in\mathbb{Z}\right)$ *is initialized from the invariant measure, then* $(h_t,t\in\mathbb{Z})$ *and* $(\epsilon_t,t\in\mathbb{Z})$ *are strictly stationary,* $\beta-$*mixing with exponential rate,* $E\left\{|h_t|^p\right\}<+\infty$ *and* $E\left\{|\epsilon_t^2|^p\right\}<+\infty$.

Proof. The result follows from the proposition 42 with $f\,(h_t)=h_t, h_{s_t}\,(\eta_t)=\beta\,(s_t)+\eta_t$ and $g_{s_t}\,(\eta_t)=\omega\,(s_t)$ with $\eta_t=\alpha_1\,(s_t)\,e_{t-1}^2+\alpha_2\,(s_t)\max(0,-e_{t-1})^2$. ■

Corollary 48 $[MS-TSGARCH\,(1,1)]$ *Consider the model*

$$\sqrt{h_t}=\omega\,(s_t)+\beta\,(s_t)\sqrt{h_{t-1}}+\alpha_1\,(s_t)\,|e_{t-1}|\sqrt{h_{t-1}}+\alpha_2\,(s_t)\max(0,-e_{t-1})\sqrt{h_{t-1}}$$

in which $\omega\,(.)>0$, $\beta\,(.)\geq 0$ *and* $\alpha_1(.)+\alpha_2(.)\geq 0$. *Assuming that there is an integer* $p\geq 1$ *such that* $\rho\left(\mathbb{P}\left(\underline{C}_3^{(p)}\right)\right)<1$ *where*
$\underline{C}_3^{(p)}=\left(E\left\{(\beta\,(k)+\alpha_1\,(k)\,|e_t|+\alpha_2\,(k)\max(0,-e_t))^p\right\},k=1,...,d\right)$. *If* $\left(\underline{\widetilde{h}}_t,t\in\mathbb{Z}\right)$ *is initialized from the invariant measure, then* $(h_t,t\in\mathbb{Z})$ *and* $(\epsilon_t,t\in\mathbb{Z})$ *are strictly stationary,* $\beta-$*mixing with exponential rate,* $E\left\{|h_t|^{p/2}\right\}<+\infty$ *and* $E\left\{|\epsilon_t|^p\right\}<+\infty$.

Proof. The result follows from the proposition 42 with $f\,(h_t)=\sqrt{h_t}, h_{s_t}\,(\eta_t)=\beta\,(s_t)+\eta_t$ and $g_{s_t}\,(\eta_t)=\omega\,(s_t)$ with $\eta_t=\alpha_1\,(s_t)\,|e_{t-1}|+\alpha_2\,(s_t)\max(0,-e_{t-1})$. ■

Chapter 3

QMV approach for $MS - BL$ models

Abstract: In this chapter, we consider the class of Markov−switching bilinear processes ($MS - BL$). Analysis based on models with time varying coefficients has long suffered from the lack of an asymptotic theory except in very restrictive cases. So, we illustrate the fundamental problems linked with $MS - BL$ models, i.e., parameters estimation by considering a maximum likelihood (ML) approach. So, we provide the detail on the asymptotic properties of ML, in particular, we discuss conditions for its strong consistency.[1]

3.1 Introduction

The estimating of general bilinear model ($II - 1.1$) is quite challenging even when $d = 1$. So, in the literature many ideas have been established for estimating the parameters of some restrictive stationary and ergodic bilinear models. The most frequently used methods are the (generalized) method of moments $(G)(MM)$ and the (conditional) least squares $(C)(LS)$ method. The asymptotic properties of the $(G)\,MM$ and $(C)\,LS$ estimates have been also discussed under certain restrictions, see for example, Pham and Tran [63], Guégan and Pham [28], Liu [54], Kim et al. [39], Grahn [26], Wittwer [75] and among others.

[1]This chapter is published in the Journal of Statistics and Probability Letters.

3.1.1 Model and its parameters

In this chapter, we will investigate the estimating of $MS-SBL(p,0,p,q)$ i.e.,

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

In $(III-1.1)$, 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 $\underline{\theta} := \left(\underline{a}', \underline{c}', \underline{p}', \underline{\pi}'\right)'$ and its true values denoted $\underline{\theta}^0$, belongs to a parameter space $\Theta\subset\mathbb{R}^R$ where $\underline{a} = \left(\underline{a}_i', 1\le i\le d\right)', \underline{c} = \left(\underline{c}_i', 1\le i\le d\right)', \underline{p} = \left(\underline{p}_i', 1\le i\le d\right)'$ and $\underline{\pi} = (\pi(1), ..., \pi(d))'$ with vectors coordinate projections $\underline{a}_i := (a_0(i), ..., a_p(i))'$, $\underline{c}_i := (c_{lk}(i), 1\le k, l\le \max(p,q)')$, $\underline{p}_i := (p_{ij}, ..., p_{id}, j\neq i)'$ (due to the constraints $\sum_{j=1}^{d} p_{ij} = 1$ for all i). For any integers a and b, such that $a\le b$ let $\underline{X}_{a:b}$ (resp. $\underline{\mathcal{X}}_{a:b}$) denotes the set $\{X_a, X_{a+1}\dots, X_b\}$ (resp. $\{(X_a,e_a), (X_{a+1},e_{a+1}), \dots, (X_b,e_b)\}$) with possibly $a=-\infty$ in this cases we shall note $\underline{X}_b$ (resp. $\underline{\mathcal{X}}_b$). The problem of interest in this chapter is the estimation of the parameter vector $\underline{\theta}$ governing Equation $(III-1.1)$ from an observed sequence $\underline{X}_{1:n}$. For this purpose, we denote the density function of observations by $g_{\underline{\theta}}(.)$ and that of innovations e_t by $f_{\underline{\theta}}(.)$ and we use $p_{\underline{\theta}}(., ..., .)$ to denote the density with respect to probability measure on $\mathcal{B}_{\mathbb{S}^n\otimes\mathbb{R}^n}$. The corresponding conditional joint density given $s_1 = x_1$ and $\underline{\mathcal{X}}_{1-p_0:0}$ with $p_0 = \max\{p,q\}+1$ is

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

The likelihood $L_n(\underline{\theta})$ with respect to the measure $\lambda^n\otimes\mu^n$ (where λ denotes the Lebesgue measure and μ is the counting on $\mathbb{S}$) that we work with is given by summing over all possible path of the Markov chain the conditional density $p_{\underline{\theta}}\left(\underline{X}_{1:n}|\underline{\mathcal{X}}_{1-p_0:0}\right)$, i.e.,

$$L_n(\underline{\theta}) = \sum_{(x_1,\dots,x_n)\in\mathbb{S}^n} \pi(x_1)\, g_{\underline{\theta}_{x_1}}\left(X_1|\underline{\mathcal{X}}_{1-p_0:0}\right)\prod_{t=2}^{n} p_{x_{t-1},x_t} g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right) \tag{III-1.2}$$

which can be rewritten as a product of matrices due to this simplified structure,

$$L_n(\underline{\theta}) = \underline{\mathbf{1}}'\left\{\prod_{t=2}^{n}\mathbb{P}\left(g_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)\right)\right\}\underline{\pi}\left(g_{\underline{\theta}}(X_1|\underline{\mathcal{X}}_{1-p_0:0})\right) \tag{III-1.3}$$

where $g_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)=\left(g_{\underline{\theta}_k}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right),1\leq k\leq d\right)\right)$. A quasi−maximum likelihood estimator ($QMLE$) of $\underline{\theta}$ is defined as any measurable solution $\widehat{\underline{\theta}}_n$ of

$$\widehat{\underline{\theta}}_n=\arg\max_{\theta\in\Theta}L_n\left(\underline{\theta}\right). \tag{III-1.4}$$

It is worth noting that given s_t, the Jacobian of the transformation from X_t to e_t is unity, so $g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)=f_{\underline{\theta}_{x_t}}\left(e_t\left(\underline{\theta}\right)|\underline{\mathcal{X}}_{1-p_0:t-1}\right)$ where $\left(e_t\left(\underline{\theta}\right),t\in\mathbb{Z}\right)$ is the process determined recursively by

$$e_t\left(\underline{\theta}\right)=X_t-a_0\left(s_t\right)-\sum_{i=1}^{p}a_i(s_t)X_{t-i}-\sum_{j=1}^{q}\sum_{i=j}^{p}c_{ij}(s_t)X_{t-i}e_{t-j}\left(\underline{\theta}\right)$$

and hence the likelihood function of $\underline{X}_{1:n}$ is the same as the joint density function of $\underline{e}_{1:n}\left(\underline{\theta}\right)$ summed over possible paths of the chain $(s_t,t\in\mathbb{Z})$ i.e.,

$$\begin{aligned}L_n\left(\underline{\theta}\right)&=\sum_{(x_1,\ldots,x_n)\in\mathbb{S}^n}\pi\left(x_1\right)f_{\underline{\theta}_{x_1}}\left(e_1\left(\underline{\theta}\right)|\underline{\mathcal{X}}_{1-p_0:0}\right)\prod_{t=2}^{n}p_{x_{t-1},x_t}f_{\underline{\theta}_{x_t}}\left(e_t\left(\underline{\theta}\right)|\underline{\mathcal{X}}_{1-p_0:t-1}\right)\\&=\underline{\mathbf{1}}'\left\{\prod_{t=2}^{n}\mathbb{P}\left(f_{\underline{\theta}}\left(e_t\left(\underline{\theta}\right)|\underline{\mathcal{X}}_{1-p_0:t-1}\right)\right)\right\}\underline{\pi}\left(f_{\underline{\theta}}(e_1\left(\underline{\theta}\right)|\underline{\mathcal{X}}_{1-p_0:0})\right)\end{aligned}$$

in which $f_{\underline{\theta}}\left(e_t\left(\underline{\theta}\right)|\underline{\mathcal{X}}_{1-p_0:t-1}\right)=\left(f_{\underline{\theta}_k}(e_t\left(\underline{\theta}\right)|\underline{\mathcal{X}}_{1-p_0:t-1}),1\leq k\leq d\right)$. For the asymptotic purpose, it is convenient to approximate the process $g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)$ (resp. $f_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)$) by its ergodic stationary version $g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)$ (resp. $f_{\underline{\theta}_{x_t}}\left(e_t|\underline{\mathcal{X}}_{t-1}\right)$ so we work with the following approximate version $\widetilde{L}_n\left(\underline{\theta}\right)$ i.e.,

$$\begin{aligned}\widetilde{L}_n\left(\underline{\theta}\right)&=\sum_{(x_1,\ldots,x_n)\in\mathbb{S}^n}\pi\left(x_1\right)g_{\underline{\theta}_{x_1}}\left(X_1|\underline{\mathcal{X}}_0\right)\prod_{t=2}^{n}p_{x_{t-1},x_t}g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)\\&=\underline{\mathbf{1}}'\left\{\prod_{t=2}^{n}\mathbb{P}\left(g_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)\right)\right\}\underline{\pi}\left(g_{\underline{\theta}}(X_1|\underline{\mathcal{X}}_0)\right)\\&=\underline{\mathbf{1}}'\left\{\prod_{t=2}^{n}\mathbb{P}\left(f_{\underline{\theta}}\left(e_t\left(\underline{\theta}\right)|\underline{\mathcal{X}}_{t-1}\right)\right)\right\}\underline{\pi}\left(f_{\underline{\theta}}(e_1\left(\underline{\theta}\right)|\underline{\mathcal{X}}_0)\right).\end{aligned}$$

Remark 1 *The existence and the uniqueness of the process* $\left(e_t\left(\underline{\theta}\right),t\in\mathbb{Z}\right)$ *is ensured by the invertibility of the model* $(III-1.1)$*. Hence, from Theorem* 30*, a sufficient condition for the invertibility of the model* $(III-1.1)$ *is ensured by the negativity of the top Lyapunov exponent associated with the random matrices* $(G(t),t\in\mathbb{Z})$ *where* $G(t)=\left[\beta_j(t)\delta_1\left(i\right)+\delta_{j+1}\left(i\right)\right]_{i,j=1,\ldots,d}$ *with* $\delta_j\left(i\right)$ *is the Kronecker's function and* $\beta_j(t)=-\sum\limits_{i=j}^{p}c_{ij}(s_t)X_{t-i}$ *provide that there already exists a strictly stationary and ergodic process* $(X_t,t\in\mathbb{Z})$ *satisfying* $(III-1.1)$*.*

Remark 2 *When the innovation process* $(e_t, t\in\mathbb{Z})$ *is heteroscedastic, i.e.,* $e_t(\underline{\theta}) = h_t(\underline{\theta})\eta_t$, *then the conditional density function* $f_{\underline{\theta}_k}\left(e_t(\underline{\theta})\,|\underline{\mathcal{X}}_{t-1}\right)$ *should be replaced by* $\dfrac{1}{h_t(\underline{\theta})}f_{\underline{\theta}_k}\left(\dfrac{e_t(\underline{\theta})}{h_t(\underline{\theta})}|\underline{\mathcal{X}}_{t-1}\right)$.

Remark 3 *For instance, the initial values can be chosen as* $X_{1-p_0} = e_{1-p_0} = ,...,= X_0 = e_0 = a_0^0(k)$ *for any* $k\in\mathbb{S}$ *or*

$$X_{1-p_0} = e_{1-p_0} =,...,= X_0 = e_0 = 0. \tag{III-1.5}$$

It will be shown that the choice of the initial values does not matter for the asymptotic properties of the ML *estimator. However, it may have importance from a practical point of view.*

3.1.2 Consistency of $QMLE$

In this subsection, we will give conditions ensuring the strong consistency of MLE for $MS-BL$ model $(III-1.1)$. Our approach is benefitted from the papers by Xie et al., [76] and Krishnamurthy and Rydén [42] for $MS-AR$, Straumann and Mikosch [70] for general conditionally heteroscedastic time series, Leroux [45] for hidden Markov models. For this purpose 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, i.e., given $s_1 = x_1$ (see Leroux [45] for further discussions),

$$p_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right) = \sum_{(x_{t-1},x_t)\in\mathbb{S}^2} g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{1-p_0:t-1}\right)p_{x_{t-1},x_t}P\left(s_{t-1}=x_{t-1}|\underline{\mathcal{X}}_{1-p_0:t-1}\right),$$
$$q_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right) = \sum_{(x_{t-1},x_t)\in\mathbb{S}^2} g_{\underline{\theta}_{x_t}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)p_{x_{t-1},x_t}P\left(s_{t-1}=x_{t-1}|\underline{\mathcal{X}}_{t-1}\right),$$

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 $\underline{\theta}_i^0$, $i = 1,...,d$.

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 $\mathcal{V}(\underline{\theta}) = \{\underline{\theta}' : \|\underline{\theta} - \underline{\theta}'\| < \delta\}$ 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. 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 $\underline{\theta} = \underline{\theta}'$.

In Assumption **A1**., the compactness of Θ is assumed in order that several results from real analysis may be used. As seen in Theorem 1, Assumption **A2.,** ensure that the process $(X_t, t \in \mathbb{Z})$ defined by $(III-1.1)$ admits a strictly stationary, ergodic solution and the existence of a finite moments (see proposition 7). Assumption **A3.**, implies that there exists at least one subset $\mathcal{V}(\underline{\theta})$ of Θ containing $\underline{\theta}_0$ over which the expectation under $\underline{\theta}_0$ of $\left|q^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right|$ is uniformly bounded. While Assumption **A4.**, means that if Equation $(III-1.1)$ has two solutions processes associated with two different parameter, then their stationary laws do not coincide.
First we show the following general results.

Lemme 4 *Under **A2** and **A3**, almost surely, uniformly with respect to* $\underline{\theta} \in \Theta$

1. $\lim_{n \longrightarrow \infty} \frac{1}{n} \log \widetilde{L}_n(\underline{\theta}) = \lim_{n \longrightarrow \infty} \frac{1}{n} \log L_n(\underline{\theta}) = E_{\underline{\theta}^0}\left\{q^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right\}$

2. $\lim_{n \longrightarrow \infty} \frac{1}{n} \log \widetilde{L}_n(\underline{\theta}) = \lim_{n \longrightarrow \infty} \frac{1}{n} \log \left\| \prod_{t=2}^{n} \mathbb{P}\left(g_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right)\right\|$

Proof.

1. Using $L_n(\underline{\theta}) = \prod_{t=1}^{n} p_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{1-p_0:t-1}\right)$ and $\widetilde{L}_n(\underline{\theta}) = \prod_{t=1}^{n} q_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)$ and define the process

$$N_t(l) = \sup_{k \geq l}\left|p^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-k:t-1}\right) - q^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right|,$$

then for each fixed l, the process $(N_t(l), t \in \mathbb{Z})$ is stationary, ergodic and $E_{\underline{\theta}^0}\{N_t(l)\} < +\infty$. Moreover,

$$\begin{aligned}
&\lim\sup_{n \longrightarrow \infty}\left|\frac{1}{n}\sum_{t=1}^{n}\left(p^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{1-p_0:t-1}\right) - q^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right)\right| \\
&\leq \lim\sup_{n \longrightarrow \infty}\frac{1}{n}\sum_{t=1}^{n}\left|p^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{1-p_0:t-1}\right) - q^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)\right| \\
&\leq \lim\sup_{n \longrightarrow \infty}\frac{1}{n}\sum_{t=l-p_0+1}^{n} N_t(l) = E_{\underline{\theta}^0}\{N_t(l)\}.
\end{aligned}$$

Since, the process $\left(q^{*}_{\underline{\theta}}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right), t\in\mathbb{Z}\right)$ is stationary and ergodic and $\lim\limits_{l\longrightarrow\infty} E_{\underline{\theta}^0}\{N_t(l)\}=0$, then the result of the first assertion follows.

2. Noting that $\sum\limits_{(x_1,\ldots,x_n)\in\mathbb{S}^n}\pi(x_1)\prod\limits_{t=2}^{n}p_{x_{t-1},x_t}=1$ and choose a norm $\|.\|$ such that $\|A\|=\sum\limits_{i,j}|a_{ij}|$, then we have

$$\begin{aligned}&\min\left\{\pi(k)g_{\underline{\theta}_k}\left(X_1|\underline{\mathcal{X}}_0\right)\right\}\left\|\prod_{t=2}^{n}\mathbb{P}\left(g_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)\right)\right\|\\ \leq\;& \widetilde{L}_n(\underline{\theta})\leq\max\left\{\pi(k)g_{\underline{\theta}_k}\left(X_1|\underline{\mathcal{X}}_0\right)\right\}\left\{\left\|\prod_{t=2}^{n}\mathbb{P}\left(g_{\underline{\theta}}\left(X_t|\underline{\mathcal{X}}_{t-1}\right)\right)\right\|\right\},\end{aligned}$$

so the second assertion follows.

■

Lemme 5 *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* $\lim\limits_{n\longrightarrow\infty}Z_n(\underline{\theta})\leq 0$ *with equality iff* $\underline{\theta}=\underline{\theta}^0$.

Proof. Under conditions **A1-A4** almost surely $Z_n(\underline{\theta})$ is well defined. From the lemma 4 and Jensen's inequality, we have

$$\lim_{n\longrightarrow\infty}Z_n(\underline{\theta})=E_{\underline{\theta}^0}\left\{\log\frac{q_{\underline{\theta}}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)}{q_{\underline{\theta}^0}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)}\right\}\leq\log E_{\underline{\theta}^0}\left\{\frac{q_{\underline{\theta}}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)}{q_{\underline{\theta}^0}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)}\right\}\leq\log 1=0,$$

with equality iff almost surely $q_{\underline{\theta}}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)=q_{\underline{\theta}^0}\left(X_t|\,\underline{\mathcal{X}}_{t-1}\right)$. Moreover, by the condition **A4,** $Z_n(\underline{\theta})$ converge to Kullback–Leinbler information which equals to zero iff $\underline{\theta}=\underline{\theta}^0$. ■

Lemme 6 *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* $\lim\sup\limits_{n\rightarrow+\infty}\sup\limits_{\underline{\theta}\in\mathcal{V}_m(\underline{\theta}')}Z_n(\underline{\theta})<0$.

Proof. The proof follows essentially the same arguments as in Xie [76]. ■

Theorem 7 *For the* $MS-BL$ *model* $(III-1.1)$*, let* $\hat{\underline{\theta}}_n$ *be the* MLE *sequence over* Θ *satisfying* $(III-1.4)$*. Then under the conditions* ***A1-A4,*** $\hat{\underline{\theta}}_n\longrightarrow\underline{\theta}^0$ *a.s as* $n\longrightarrow\infty$.

Proof. Suppose that $\hat{\underline{\theta}}_n$ does not converge to $\underline{\theta}^0$ a.s. as $n \longrightarrow \infty$. This means that for any N (large enough), $\exists \delta > 0$ and $n > N$, such that $\left\| \hat{\underline{\theta}}_n - \underline{\theta}^0 \right\| \geq \delta$. By lemma 6, it follows that $L_n\left(\hat{\underline{\theta}}_n\right) < L_n\left(\underline{\theta}^0\right)$. However, by the definition of MLE given by $(III-1.4)$, we have $L_n\left(\hat{\underline{\theta}}_n\right) = \sup_{\underline{\theta} \in \Theta^*} L_n\left(\underline{\theta}\right) \geq L_n\left(\underline{\theta}^0\right)$ for any compact subset $\Theta^*$of Θ containing $\underline{\theta}^0$. This contradiction gives the result. ■

Chapter 4

GMM approach for $MS - BL$ models

Abstract: In this chapter we consider Markov−switching bilinear process ($MS - BL$). We illustrate the fundamental problems linked with $MS - BL$ models, i.e., parameters estimation by a minimum $\mathbb{L}_2$−distance estimator (MDE). So, we provide the detail on the asymptotic properties of MDE, in particular, we discuss conditions for its consistency and asymptotic normality. Numerical experiments on simulated data sets are presented to highlight the theoretical results.[1]

4.1 Introduction

In this chapter, we focus on the minimum distance estimation (MDE) of the parameters of Model ($II - 1.1$) with $c_{ij}(.) = 0$ if $i < j$,i.e.,

$$X_t = a_0(s_t) + \sum_{i=1}^{p} a_i(s_t)X_{t-i} + \sum_{j=0}^{q} b_j(s_t)e_{t-j} + \sum_{j=1}^{Q}\sum_{i=j}^{P} c_{ij}(s_t)X_{t-i}e_{t-j}, \quad \text{(IV-1.1)}$$

and use the models ($II - 3.2$) to examine the performance of the method. There are already established methods in the literature for estimating of some special cases of ($IV - 1.1$) with time−invariant parameters including [26], [38], [39], [40], [55], [63],[75] among other. In our best knowledge, only a few studies on bilinear models with time−varying coefficients can be found in the literature, except the works by [11], [7], [8] and by [12] concerning the bilinear models with

[1] This chapter is submit in the Journal of Statistics

periodic coefficients. So, in this chapter, we continue our effort in estimating the $MS-BL$. The proposed estimator is closely related to the MDE considered by Tieslau et al. [74] for long−memory processes, Bibi and Gautier [12] for simple periodic $BL(0,0,2,1)$ models and by Storti [69] and Baillie and Chung [5] for $GARCH(1,1)$ models based on the autocorrelation function of the squared observations.

4.1.1 Proposed MDE and asymptotic properties

Let $\{X_1,...,X_n\}$ be a sample from a second−order stationary $MS-BL$ process generated by $(IV-1.1)$ with l unknown parameters, gathered in $\underline{\theta} := \left(\underline{a}', \underline{b}', \underline{c}', \underline{p}', \underline{\pi}', \sigma^2\right)' \in \Theta \subset \mathbb{R}^l$ where $\underline{a} := \left(\underline{a}_0', \underline{a}_1', ..., \underline{a}_p'\right)'$, $\underline{b} := \left(\underline{b}_1', ..., \underline{b}_q'\right)'$, $\underline{c} := \left(\underline{c}_{11}', ..., \underline{c}_{PQ}'\right)'$, $\underline{p} := \left(\underline{p}_1', ..., \underline{p}_d'\right)'$, $\underline{\pi} := (\pi(j), 1 \leq j \leq d)'$ with $\underline{a}_i := (a_i(j), 1 \leq j \leq d)'$, $\underline{b}_i := (b_i(j), 1 \leq j \leq d)'$, $\underline{c}_{ij} := (c_{ij}(k), 1 \leq k \leq d)'$, $\underline{p}_i := (p_{ij}, 1 \leq j \leq d)'$. For any $h \geq l$, let

$$\begin{aligned} \underline{\Gamma}(\underline{\theta}) &: = \underline{\Gamma} = (\gamma(0), \ldots, \gamma(h))',\ \hat{\underline{\Gamma}}_n = (\hat{\gamma}_n(0), \ldots, \hat{\gamma}_n(h))', \\ \widetilde{\underline{\Gamma}}_n &= \widetilde{\underline{\Gamma}}_n(\underline{\theta}) = (\widetilde{\gamma}_n(0), \ldots, \widetilde{\gamma}_n(h))' \end{aligned}$$

where $\hat{\gamma}_n(i) = \frac{1}{n}\sum\limits_{t=i+1}^{n}\left(X_t - \widehat{X}_n\right)\left(X_{t-i} - \widehat{X}_n\right)$ with $\widehat{X}_n = \frac{1}{n}\sum\limits_{t=1}^{n} X_t$ and $\widetilde{\gamma}_n(i)$ is $\gamma(i), i = 0, ..., h$ defined by$(II-2.11)$ in which the mean and covariance functions are replaced by their consistent estimates. To derive the asymptotic properties of $\hat{\underline{\Gamma}}_n$ and $\widetilde{\underline{\Gamma}}_n$, we need the following assumptions.

A1. The chain $(s_t, t \in \mathbb{Z})$ is $q-$dependent.

A2. The process $(X_t, t \in \mathbb{Z})$ is strictly stationary, ergodic and admits moments up to $4-th$ order.

Lemme 1 *Let $(X_t, t \in \mathbb{Z})$ be a process satisfying model $(IV-1.1)$ with state-space representation $(II-2.4)$. Then under the conditions **A1** and **A2** we have,*

1. *almost surely $\widehat{X}_n$ converges to μ and $\sqrt{n}\left(\widehat{X}_n - \mu\right) \rightsquigarrow \mathcal{N}\left(0, \sum\limits_{h\in\mathbb{Z}}\gamma(h)\right)$ where $\gamma(h)$ is given in $(II-2.11)$*
2. *almost surely $\widehat{\underline{\Gamma}}_n$ converges to $\underline{\Gamma}$ and $\sqrt{n}\left(\widehat{\underline{\Gamma}}_n - \underline{\Gamma}\right) \rightsquigarrow \mathcal{N}(0, \Sigma)$ where Σ is an $(h+1)\times(h+1)$ covariance matrix whose (i,j)-th element $\sigma_{i,j}$ is given by*

$$\sigma_{i,j} = \sum_{k\in\mathbb{Z}} Cov\left(X_t X_{t-i-1}, X_{t-k} X_{t-j-k-1}\right),\ i,j = 0, \ldots, h,$$

Proof. From the representation $(II-2.4)$, define the processes $(U_t(m), t \in \mathbb{Z})$ and $(W_t(m), t \in \mathbb{Z})$ as $U_t(m) := \underline{F}'\underline{U}_t(m)$ and $W_t(m) := \underline{F}'\underline{W}_t(m)$ where

$$\underline{U}_t(m) = \sum_{k=1}^{m}\left\{\prod_{i=0}^{k-1}\Gamma_{s_{t-i}}(e_{t-i})\right\}\underline{\eta}_{s_{t-k}}(e_{t-k}) + \underline{\eta}_{s_t}(e_t),$$
$$\underline{W}_t(m) = \left\{\prod_{i=0}^{m}\Gamma_{s_{t-i}}(e_{t-i})\right\}\underline{Y}_{t-m-1}.$$

Then under the conditions **A1** and **A2,** $(U_t(m), t \in \mathbb{Z})$ is an $(m+1)-$dependent $(m > q)$ and second order stationary process and X_t can be expressed by

$$X_t = U_t(m) + W_t(m) \tag{IV-1.2}$$

1. The convergence of $\widehat{X}_n$ to μ follows immediately from the ergodic theorem. Next, by Theorem 9, $W_t(m)$ converges in probability to 0 as $m \to \infty$ and thus the asymptotic distribution of $\sqrt{n}\left(\widehat{X}_n - \mu\right)$ is the same as that of $n^{-\frac{1}{2}}\sum_{t=1}^{n}(U_t(m) - \mu_U)$ where $\mu_U = E\{U_t(m)\}$ and since $E\{U_t^2(m)\} < +\infty$, we have for m fixed

$$n^{-\frac{1}{2}}\sum_{t=1}^{n}(U_t(m) - \mu_U) \rightsquigarrow \mathcal{N}\left(0, \sum_{k=-m}^{m} Cov\left(U_t(m), U_{t-k}(m)\right)\right).$$

As $m \to \infty$, $U_t(m)$ converges in probability to X_t and hence the asymptotic variance converge to $\sum_{k\in\mathbb{Z}}\gamma(k)$ which can be expressed by $(II-2.11)$.

2. Firstly, the convergence of $\widehat{\underline{\Gamma}}_n$ to $\underline{\Gamma}$ follows from the ergodic theorem. Secondary, from the definition of $\widehat{\gamma}_n(.)$ we have

$$\widehat{\gamma}_n(i) = \frac{1}{n}\sum_{t=i+1}^{n}X_tX_{t-i} - \frac{\widehat{X}_n}{n}\sum_{t=i+1}^{n}(X_t + X_{t-i}) + \frac{(n-i)}{n}\widehat{X}_n^2.$$

By the ergodic theorem, the term $-\frac{\widehat{X}_n}{n}\sum_{t=i+1}^{n}(X_t + X_{t-i}) + \frac{(n-i)}{n}\widehat{X}_n^2$ converges to $-\mu^2$, so, the asymptotic distribution of $\sqrt{n}\left(\widehat{\gamma}_n(i) - \gamma(i)\right)$ is the same as that of $n^{-1/2}\sum_{t=i+1}^{n}(X_tX_{t-i} - E\{X_tX_{t-i}\})$. On the other hand, since $U_t(m)$ converges in probability to a stationary process and $W_t(m)$ converges in probability to zero as $m \to \infty$, then we can show that the asymptotic distribution of $n^{-1/2}\sum_{t=i+1}^{n}(X_tX_{t-i}-E\{X_tX_{t-i}\})$ is the same as the one of

$$n^{-1/2}\sum_{t=i+1}^{n}\left(U_t(m)U_{t-i}(m)-E\left\{U_t(m)U_{t-i}(m)\right\}\right) \text{ as } m\to\infty.$$

Now, for any real sequence $(\lambda_i, 0\le i\le h)$, let

$$P_n=n^{-1/2}\sum_{i=0}^{h}\sum_{t=i+1}^{n}\lambda_i\left(U_t(m)U_{t-i}(m)-E\left\{U_t(m)U_{t-i}(m)\right\}\right),$$

so $\lim\limits_{n\to\infty}P_n=\lim\limits_{n\to\infty}n^{-1/2}\sum_{t=i+1}^{n}Y_t(h)$ where

$$Y_t(h)=\sum_{i=0}^{h}\lambda_i\left(U_t(m)U_{t-i}(m)-E\left\{U_t(m)U_{t-i}(m)\right\}\right).$$

Clearly, $\{Y_t(h), t\in\mathbb{Z}\}$ is an $(m+h+1)-$dependent stationary process with $E\{Y_t(h)Y_{t-k}(h)\}=\underline{\lambda}'W(k)\underline{\lambda}<+\infty$ where $\underline{\lambda}=(\lambda_0,\ldots,\lambda_h)'$ and where $W(k)$ is the covariance matrix with $(i,j)-$th element being

$$(W(k))_{i,j}=Cov\left(U_t(m)U_{t-i-1}(m),U_{t-k}(m)U_{t-j-k-1}(m)\right),\ i,j=0,\ldots,h.$$

Therefore, we have $n^{-1/2}\sum_{t=i+1}^{n}Y_t(h)\rightsquigarrow\mathcal{N}(0,\underline{\lambda}'W\underline{\lambda})$ where the $(i,j)-th$ element of the asymptotic variance$-$covariance matrix W is $(W)_{i,j}=\sum_{k=-(m+h)}^{m+h}\{W(k)\}_{i,j}$. As $m\to\infty$, W converges to Σ whose $(i,j)-$th element is $\sigma_{i,j}=\sum_{k\in\mathbb{Z}}Cov\{X_tX_{t-i},X_{t-k}X_{t-k-j}\}, i,j=0,\ldots,h$. Finally the proof follows from the Cramer$-$Wold device.

■

Corollary 2 *Under the Condition of the above lemma, we have* $\widetilde{\underline{\Gamma}}_n-\underline{\Gamma}$ *converges to* $\underline{0}$ *almost surely and* $\sqrt{n}\left(\widetilde{\underline{\Gamma}}_n-\underline{\Gamma}\right)\rightsquigarrow\mathcal{N}(0,\Sigma)$.

Proof. Write $\widetilde{\underline{\Gamma}}_n-\underline{\Gamma}=\widetilde{\underline{\Gamma}}_n-\widehat{\underline{\Gamma}}_n+\widehat{\underline{\Gamma}}_n-\underline{\Gamma}$, since $\widetilde{\underline{\Gamma}}_n-\widehat{\underline{\Gamma}}_n$ converges almost surely to $\underline{0}$, then the asymptotic properties of $\widetilde{\underline{\Gamma}}_n-\underline{\Gamma}$ are the same as that of $\widehat{\underline{\Gamma}}_n-\underline{\Gamma}$. ■

We are now in a position to state the MD procedure. The MD estimator of $\underline{\theta}$ is defined as any measurable solution of

$$\widetilde{\underline{\theta}}_n:=\arg\min_{\underline{\theta}\in\Theta}\widetilde{Q}_n(\underline{\theta})$$

where $\widetilde{Q}_n(\underline{\theta}) = \underline{\widetilde{F}}_n'(\underline{\theta}) M_n \underline{\widetilde{F}}_n(\underline{\theta})$ with $\underline{\widetilde{F}}_n(\underline{\theta}) = \hat{\underline{\Gamma}}_n - \tilde{\underline{\Gamma}}_n(\underline{\theta})$ is the score function and M_n is a sequence of $(h+1)\times(h+1)$ random non−negative definite matrices introduced in order to improve the efficiency. 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.

A3. The sequence of matrices (M_n) converges in probability to a non random positive definite matrix M.

A4. The matrix $\dfrac{\partial\underline{\Gamma}'(\underline{\theta}_0)}{\partial\underline{\theta}} M \dfrac{\partial\underline{\Gamma}(\underline{\theta}_0)}{\partial\underline{\theta}}$ is a finite nonsingular matrix of constants.

A5. The parameter Θ is compact and $\underline{\theta}_0$ is in the interior of Θ.

Under these assumptions, we can state the following result.

Theorem 3 *Under **A1−A5**, $\underline{\widetilde{\theta}}_n$ converges in probability to $\underline{\theta}_0$.*

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

$$\frac{\partial\underline{\widetilde{\Gamma}}_n'(\underline{\widetilde{\theta}}_n)}{\partial\underline{\theta}} M_n \underline{\widetilde{F}}_n(\underline{\widetilde{\theta}}_n) = 0. \tag{IV-1.3}$$

Taking the first−order Taylor−series expansion of the score vector $\underline{\widetilde{F}}_n(\underline{\widetilde{\theta}})$ around $\underline{\theta}_0$, we have $\underline{\widetilde{F}}_n(\underline{\widetilde{\theta}}_n) = \underline{\widetilde{F}}_n(\underline{\theta}_0) - \dfrac{\partial\underline{\widetilde{\Gamma}}_n(\underline{\theta}_*)}{\partial\underline{\theta}}\left(\underline{\widetilde{\theta}}_n - \underline{\theta}_0\right)$ where $\underline{\theta}_*$ is an intermediate point on the line segment joining $\underline{\widetilde{\theta}}_n$ and $\underline{\theta}_0$. Substituting for $\underline{\widetilde{F}}_n(\underline{\widetilde{\theta}}_n)$ into $(IV-1.3)$ yields $\dfrac{\partial\underline{\widetilde{\Gamma}}_n'(\underline{\widetilde{\theta}}_n)}{\partial\underline{\theta}} M_n \left\{\underline{\widetilde{F}}_n(\underline{\theta}_0) - \dfrac{\partial\underline{\widetilde{\Gamma}}_n(\underline{\theta}_*)}{\partial\underline{\theta}}\left(\underline{\widetilde{\theta}}_n - \underline{\theta}_0\right)\right\} = 0$. Rearranging the above expression gives almost surely

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

Since the process $(X_t, t\in\mathbb{Z})$ is second−order stationary and ergodic, then under the conditions of Lemma 1,

$$\begin{aligned} p\lim_{n\to\infty} \frac{\partial\underline{\widetilde{\Gamma}}_n(\underline{\widetilde{\theta}}_n)}{\partial\underline{\theta}} M_n &= B = \frac{\partial\underline{\Gamma}(\underline{\theta}_0)}{\partial\underline{\theta}} M \text{ and} \\ p\lim_{n\to\infty} \frac{\partial\underline{\widetilde{\Gamma}}_n'(\underline{\widetilde{\theta}}_n)}{\partial\underline{\theta}} M_n \frac{\partial\underline{\widetilde{\Gamma}}_n(\underline{\theta}_*)}{\partial\underline{\theta}} &= A = \frac{\partial\underline{\Gamma}'(\underline{\theta}_0)}{\partial\underline{\theta}} M \frac{\partial\underline{\Gamma}(\underline{\theta}_0)}{\partial\underline{\theta}}. \end{aligned}$$

Hence from Slutsky's and the dominated convergence theorems

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

is finite, and since $p\lim_{n\to\infty}\widetilde{\underline{F}}_n(\underline{\theta}_0)=\underline{0}$, the consistency of $\widetilde{\underline{\theta}}_n$ follows. ■
Now, we consider the estimator

$$\widehat{\underline{\theta}}_n = \arg\min_{\underline{\theta}\in\Theta}\widehat{\underline{F}}_n'(\underline{\theta})M_n\widehat{\underline{F}}_n(\underline{\theta}) = \arg\min_{\underline{\theta}\in\Theta}\widehat{Q}_n(\underline{\theta})$$

where $\widehat{Q}_n(\underline{\theta}) = \widehat{\underline{F}}_n'(\underline{\theta})M_n\widehat{\underline{F}}_n(\underline{\theta})$ with $\widehat{\underline{F}}_n(\underline{\theta}) = \widehat{\underline{\Gamma}}_n - \underline{\Gamma}(\underline{\theta})$. The asymptotic properties of $\{\widehat{\underline{\theta}}_n\}_{n\geq 1}$ are given in the following theorem.

Theorem 4 *Under **A1**–**A5** we have*

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

Proof. The proof rests classically on a Taylor–series expansion of the score vector $\underline{F}_n(\underline{\theta})$ around $\underline{\theta}_0$. Thus, by the same argument used in Theorem 3, we have

$$\left\{\frac{\partial\underline{\Gamma}'(\widehat{\underline{\theta}}_n)}{\partial\underline{\theta}}M_n\frac{\partial\underline{\Gamma}(\widehat{\underline{\theta}}_*)}{\partial\underline{\theta}}\right\}\left(\widehat{\underline{\theta}}_n - \underline{\theta}_0\right) = \frac{\partial\underline{\Gamma}'(\widehat{\underline{\theta}}_n)}{\partial\underline{\theta}}M_n\widehat{\underline{F}}_n(\underline{\theta}_0). \tag{IV-1.4}$$

From Lemma 1, we have $p\lim_{n\to\infty}\widehat{\underline{F}}_n(\underline{\theta}_0) = 0$, and thus $p\lim_{n\to\infty}\widehat{\underline{\theta}}_n = \underline{\theta}_0$. On the other hand, in the expansion (IV-1.4) we have the following limits $A = p\lim_{n\to\infty}\frac{\partial\underline{\Gamma}'(\widehat{\underline{\theta}}_n)}{\partial\underline{\theta}}M_n\frac{\partial\underline{\Gamma}(\widehat{\underline{\theta}}_*)}{\partial\underline{\theta}}$, $B = p\lim_{n\to\infty}\frac{\partial\underline{\Gamma}'(\widehat{\underline{\theta}}_n)}{\partial\underline{\theta}}M_n$, since $\sqrt{n}\widehat{\underline{F}}_n(\underline{\theta}) \rightsquigarrow \mathcal{N}(0,\Sigma)$.
Then, the result simply follows from Slutsky's theorem. ■
The result in Theorem 4 can be easily generalized to the estimator $\widetilde{\underline{\theta}}_n$.

Theorem 5 *Under **A1**–**A5** we have*

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

Proof. We have $\widetilde{\underline{F}}_n(\underline{\theta}) - \widehat{\underline{F}}_n(\underline{\theta}) = \underline{\Gamma}(\underline{\theta}) - \widetilde{\underline{\Gamma}}_n(\underline{\theta})$, so by Corollary 2,

$$p\lim_{n\to\infty}\sqrt{n}\{\widetilde{\underline{F}}_n(\underline{\theta}) - \widehat{\underline{F}}_n(\underline{\theta})\} = \underline{0}$$

and consequently $\sqrt{n}\widetilde{\underline{F}}_n(\underline{\theta})$ converges to the same limit distribution as the one of $\sqrt{n}\widehat{\underline{F}}_n(\underline{\theta})$. ■

Discussion

We now discuss the optimal choice of the weighting matrix M. It is clear from Theorem 4 that the asymptotic variance of $\widehat{\underline{\theta}}_n$ depends on M_n via M. As it is the case for GMM estimation, under the conditions of Theorem 5, the choice of M matters for asymptotic efficiency. When appropriately choosing M, it is possible to minimize the asymptotic variance of $\widehat{\underline{\theta}}_n$. Then the minimum variance that can be achieved is when $M = \Sigma^{-1}$. In this particular case, the asymptotic variance of $\widehat{\underline{\theta}}_n$ is

$$\left\{ \frac{\partial \underline{\Gamma}'(\underline{\theta}_0)}{\partial \underline{\theta}} \Sigma^{-1} \frac{\partial \underline{\Gamma}(\underline{\theta}_0)}{\partial \underline{\theta}} \right\}^{-1}$$

and $n\widehat{Q}_n(\underline{\theta})$ has an asymptotic chi-square distribution (see Hall [32] Theorem 3.4). However, 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 $\widehat{\underline{\theta}}_n$ and its asymptotic variance (needed to construct confidence intervals and to make statistical tests based on $\widehat{\underline{\theta}}_n$). In practice, a heteroskedasticity and autocorrelation consistent (HAC) estimate of Σ can be used, i.e.,

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

where $\widehat{\Omega}_n(j) = n^{-1} \sum_{t=0}^{n-j-1} \widehat{\underline{W}}_t \widehat{\underline{W}}'_{t+j}$ with $\widehat{\underline{W}}_t = \left(\widehat{W}_t(0), \ldots, \widehat{W}_t(h)\right)'$, $\widehat{W}_t(k) = (X_t - \widehat{\mu}_n)(X_{t-k} - \widehat{\mu}_n) - \widetilde{\gamma}_n(k, \widehat{\underline{\theta}}^*_n)$, $k = 0, \ldots, h$, $\widehat{\underline{\theta}}^*_n$ any consistent estimator of $\underline{\theta}$. 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 $\mathcal{K} = \{k : \mathbb{R} \to [-1, 1] \mid k(0) = 1,\ k(x) = k(-x), \forall x \in \mathbb{R},\ \int |k(x)| dx < \infty$, and k is continuous but at some countable points$\}$. Examples of such kernel weights are the following

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

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.

4.1.2 Hypothesis testing

As an application of Theorem 5, we consider the problem of testing a null hypothesis against an alternative one of the form

$$H_0 : R\theta = \underline{\theta}^* \quad v.s \quad H_1 : R\underline{\theta} \neq \underline{\theta}^*, \tag{IV-1.5}$$

where R is a given $l \times l$ matrix of rank l, and $\underline{\theta}^*$ is a given vector. Under the null hypothesis H_0 in $(IV-1.5)$ and under the conditions of Theorem 5,

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

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

Theorem 6 *Assume that the conditions of Theorem 5 hold and Σ is a nonsingular matrix. Then, under the null hypothesis in $(IV-1.5)$ with R of rank l, we have*

$$W_n = n\left(R\underline{\widetilde{\theta}}_n - \underline{\theta}^*\right)'\left(R\widehat{A}_n^{-1}\widehat{B}_n\widehat{\Sigma}_n\widehat{B}_n'\widehat{A'}_n^{-1}R'\right)^{-1}\left(R\underline{\widetilde{\theta}}_n - \underline{\theta}^*\right) \rightsquigarrow \chi^2_l \tag{IV-1.6}$$

where $\widehat{A}_n$ and $\widehat{B}_n$ are some consistent estimates of A and B respectively. In addition, under the alternative hypothesis in ($IV-1.5$), we have

$$p\lim_{n\to\infty} n^{-1}W_n = (R\underline{\theta}-\underline{\theta}^*)'\left(RA^{-1}B\Sigma B'A'^{-1}R'\right)^{-1}(R\underline{\theta}-\underline{\theta}^*) > 0. \qquad \text{(IV-1.7)}$$

Note that the test statistics W_n is now the one of the Wald test of the null hypothesis in $(IV-1.5)$.

Given the size $\alpha \in [0,1]$, choose a critical value β so that under the null hypothesis in $(IV-1.5)$, $\mathbb{P}(W_n > \beta) \longrightarrow \alpha$. Then the null hypothesis is accepted if $W_n \le \beta$, and rejected in favor of the alternative hypothesis if $W_n > \beta$. This test is consistent due to $(IV-1.7)$. In the case when R is a raw vector (so $\underline{\theta}^*$ is a scalar), we can modify $(IV-1.6)$ to

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

whereas under the alternative hypothesis in $(IV-1.5)$, $(IV-1.7)$ becomes

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

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

Theorem 7 *Assume that the conditions of Theorem 6 hold. Consider the hypothesis*

$$H_0^{(\upsilon)}: \theta_\upsilon = \theta_\upsilon^* \quad v.s \quad H_1^{(\upsilon)}: \theta_\upsilon \neq \theta_\upsilon^*, \quad \upsilon \in \{1,...,l\},$$

where θ_υ^* is given. Let $(\underline{e}_i,\ i=1,...,l)$ be the canonical basis of $\mathbb{R}^l$. Then, under $H_0^{(\upsilon)}$,

$$t_n(\upsilon) = \frac{\sqrt{n}\left(\widetilde{\theta}_{\upsilon,n}-\theta_\upsilon^*\right)}{\sqrt{\underline{e}'_\upsilon\widehat{A}_n^{-1}\widehat{B}_n\widehat{\Sigma}_n\widehat{B}'_n\widehat{A'_n}^{-1}\underline{e}_\upsilon}} \rightsquigarrow \mathcal{N}(0,1),$$

whereas under $H_1^{(\upsilon)}$,

$$p\lim_{n\to\infty}\frac{t_n(\upsilon)}{\sqrt{n}} = \frac{\theta_\upsilon-\theta_\upsilon^*}{\sqrt{\underline{e}'_\upsilon A^{-1}B\Sigma B'A'^{-1}\underline{e}_\upsilon}} \neq 0.$$

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

Remark 8 *The likelihood ratio* (LR) *(resp. Lagrange Multiplier* (LM)*) style tests (despite* MDE *having no likelihood) are available. Indeed, let* $\widehat{\underline{\theta}}_n^{(LR)}$ *(resp.* $\widehat{\underline{\theta}}_n^{(LM)}$*) any measurable solutions of* $\widetilde{\underline{\theta}}_n^{(LR)} = Arg\min \widetilde{\underline{F}}_n'(\underline{\theta})\widehat{\Sigma}_n^{-1}\widetilde{\underline{F}}_n(\underline{\theta})$ *subject to* $R\underline{\theta} - \underline{\theta}^0 = \underline{0}$ *(resp.*$\widetilde{\underline{\theta}}_n^{(LM)} = Arg\min \widetilde{\underline{F}}_n'(\underline{\theta})\widehat{\Sigma}_n^{-1}\widetilde{\underline{F}}_n(\underline{\theta}) - \underline{\lambda}'(R\underline{\theta} - \underline{\theta}^0)$*) where* $\widehat{\Sigma}_n$ *is an estimate of the asymptotic variance defined in Lemma* 1*. Then from these estimates, a* LR *and* LM *−like tests statistics can be formed as*

$$\begin{aligned}
\frac{LR_n}{n} &= \widetilde{\underline{F}}_n'\left(\widetilde{\underline{\theta}}_n^{(LR)}\right)\widehat{\Sigma}_n^{-1}\widetilde{\underline{F}}_n\left(\widetilde{\underline{\theta}}_n^{(LR)}\right) - \widetilde{\underline{F}}_n'\left(\widetilde{\underline{\theta}}_n\right)\widehat{\Sigma}_n^{-1}\widetilde{\underline{F}}_n\left(\widetilde{\underline{\theta}}_n\right) \rightsquigarrow \chi_l^2 \\
\frac{LM_n}{n} &= \widetilde{\underline{F}}_n'\left(\widetilde{\underline{\theta}}_n^{(LM)}\right)\widehat{\Sigma}_n^{-1}\widetilde{F}\left(\widetilde{F}'\widehat{\Sigma}_n^{-1}\widetilde{F}\right)^{-1}\widetilde{F}'\widehat{\Sigma}_n^{-1}\widetilde{\underline{F}}_n\left(\widetilde{\underline{\theta}}_n^{(LM)}\right) \rightsquigarrow \chi_l^2
\end{aligned}$$

where in probability $\widetilde{F} := \lim_{n\to\infty} \dfrac{\partial \widetilde{\underline{F}}_{\mathbf{N}}(\underline{\theta})}{\partial \underline{\theta}}$.

4.2 Simulation results

4.2.1 Generalities

In order to illustrate the performance of our asymptotic results described in previous section, we now provide some numerical results from Monte Carlo experiments. We simulated 1000 independent trajectories via a $MS-BL$ Models $(II-3.2)$ of length $n \in \{500, 1000, 2000\}$ with $d=2$, standard normal errors distribution with parameter $\underline{\theta}$ which satisfies the second order stationarity and existence of moments up to 4−th order conditions. For each trajectory, the parameter vector $\underline{\theta}$ has been estimated with MDE, noted as $\widehat{\underline{\theta}}_n$. All the efficient Minimum Distance (MD) estimations have been performed with the parameter dependent truncated kernel weight k_T described in above discussion. In addition, and in order to have an heavy−tailed distribution for the errors, we consider the Student t_5−distribution as well (where 5 denotes the number of degrees of freedom), to replace the standard Gaussian assumption in MD. This additional experiment is made to emphasize that the proposed asymptotic theory is free from the Gaussianity assumption.

In Tables 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 (e_t). We denote by $\sqrt{Var_{as}(\widetilde{\theta}_{v,n})} = n^{-1/2}\sqrt{(\widehat{\Sigma}_n)_{v,v}}$ the estimator of the standard deviation. 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

$\sqrt{Var_{as}(\widetilde{\theta}_{v,n})}$ over the 1000 simulations, denoted $RMSE^*$, has been compared to the root of the mean of $(\widetilde{\theta}_{v,n}-\theta_v)^2$ over the 1000 simulations, denoted by $RMSE$.

4.2.2 $MS-BL(0,0,2,1)$ model

The model $MS-BL(0,0,2,1)$ was defined by $(II-3.2)$. When $d=2$ with $p_{11}=p_{22}=1-p, p_{12}=p_{21}=p$, the eigenvalues of $\mathbb{P}^2\left(\sigma^2\underline{c}^{(2)}\right)$ are

$$\frac{\sigma^2}{2}(1-\alpha)(c^2(1)+c^2(2)) \pm \frac{\sigma^2}{2}\sqrt{(1-2\alpha)(c^2(1)-c^2(2))^2+\alpha^2(c^2(1)+c^2(2))^2}$$

where $\alpha=2p(1-p)$. Hence, the condition $\lambda_{(2)}=\rho\left(\mathbb{P}^2\left(\sigma^2\underline{c}^{(2)}\right)\right)<1$ is equivalent to the following two conditions

$$\begin{cases} \sigma^2(1-\alpha)\left(c^2(1)+c^2(2)\right)-\sigma^4(1-2\alpha)\,c^2(1)\,c^2(2)<1 \\ \sigma^2(1-\alpha)\left(c^2(1)+c^2(2)\right)<2. \end{cases}$$

So, the stationarity region is shown in Figure $Fig2$.

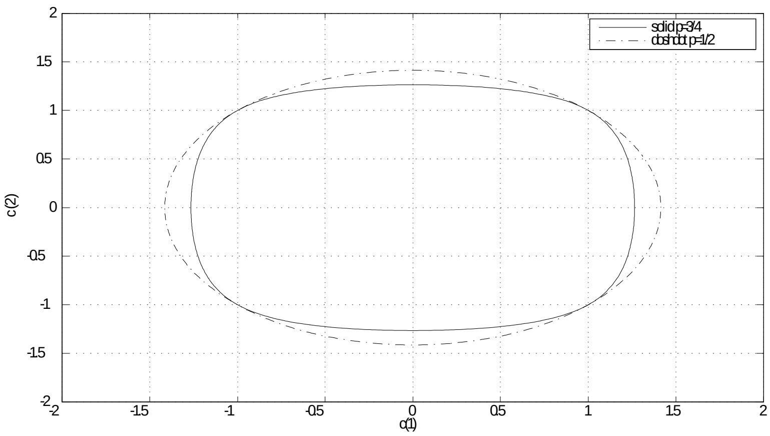

Fig 2. Plots of the boundary curves $\lambda_{(2)}=1$ for $MS-BL(0,0,2,1)$ Model with $e_t \rightsquigarrow N(0,1)$

The results of simulation

n	500		1000		2000	
Estimates	$\widehat{\underline{\theta}}_n$		$\widehat{\underline{\theta}}_n$		$\widehat{\underline{\theta}}_n$	
$Mean$	0.4675	(0.4425)	0.4875	(0.4795)	0.4975	(0.4795)
	−0.4625	(−0.4675)	−0.4885	(−0.4775)	−0.5005	(−0.4975)
	0.220	(0.216)	0.230	(0.236)	0.240	(0.237)
	0.710	(0.714)	0.730	(0.724)	0.740	(0.744)
$RMSE^*$	0.1568	(0.1468)	0.1468	(0.1461)	0.1168	(0.1178)
	0.1489	(0.1469)	0.1389	(0.1369)	0.1189	(0.1179)
	0.1790	(0.1690)	0.1290	(0.1272)	0.1190	(0.1158)
	0.1601	(0.1608)	0.1201	(0.1208)	0.1201	(0.1208)
$RMSE$	0.1588	(0.1458)	0.1488	(0.1458)	0.1188	(0.1158)
	0.1499	(0.1468)	0.1399	(0.1368)	0.1196	(0.1168)
	0.1794	(0.1691)	0.1294	(0.1291)	0.1196	(0.1194)
	0.1621	(0.1613)	0.1221	(0.1213)	0.1221	(0.1213)

Table (1): Model (1) with $\underline{\theta}=(\underline{c}', \pi(1), p)'$ with $\underline{c}' = (0.5, -0.5), \pi(1) = 1/4$ and $p = 3/4$

4.2.3 $MS-BL(1,0,1,1)$ model

The model $MS-BL(1,0,1,1)$ was defined by $(II-3.2)$. The eigenvalues of $\mathbb{P}\left(\underline{\gamma}\,(2)\right)$ are

$$\frac{1}{2}(1-\alpha)(x(1)+x(2))\pm\frac{1}{2}\sqrt{(1-2\alpha)(x(1)-x(2))^2+\alpha^2(x(1)+x(2))^2}$$

where $\alpha = p$ and $x(i) = a^2(i)+\sigma^2c^2(i),\ i = 1,2.$ The condition $\lambda_{(2)} := \rho\left(\mathbb{P}\left(\underline{\gamma}\,(2)\right)\right) < 1$ is therefore equivalent to the following conditions

$$\begin{cases} (1-\alpha)\,(x\,(1)+x\,(2))-(1-2\alpha)\,x\,(1)\,x\,(2)<1 \\ (1-\alpha)\,(x\,(1)+x\,(2))<2 \end{cases}$$

which reduce to the condition given by Francq and Zakoïan [23] for the $MS-AR$ model. The zone of stationarity for $MS-BL(0,0,1,1)$ is shown in Figure $Fig3$.

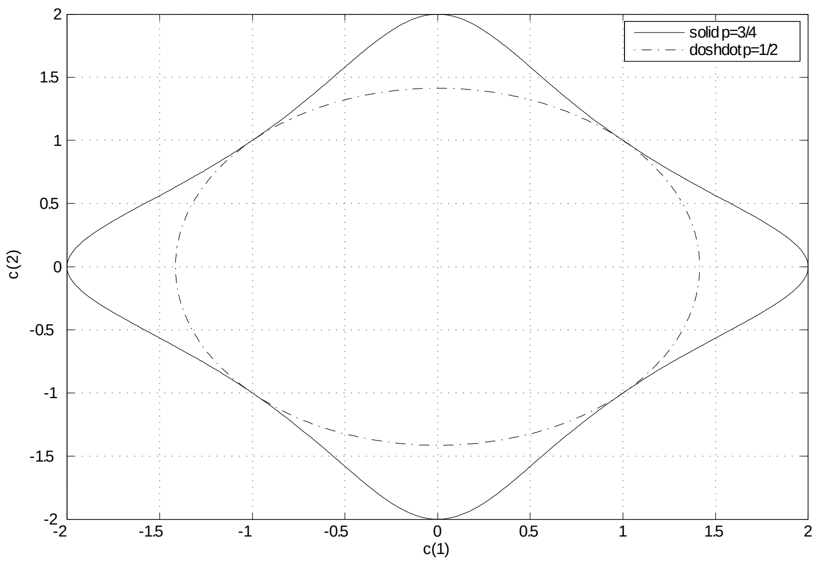

Fig 3. Plots of the boundary curves $\lambda_{(2)} = 1$ for $MS-BL(0,0,1,1)$ Model with $e_t \rightsquigarrow N(0,1)$

Remark 9 *It is worth noting that this example shows that the local second−order stationarity condition is neither sufficient nor necessary for the global second−order stationarity.*

The result of simulation

n	500		1000		2000	
Estimates	$\widehat{\underline{\theta}}_n$		$\widehat{\underline{\theta}}_n$		$\widehat{\underline{\theta}}_n$	
Mean	−0.3468	(−0.3458)	−0.3478	(−0.3458)	−0.3503	(−0.3508)
	0.3400	(0.3421)	0.3469	(0.3435)	0.3489	(0.3495)
	0.2403	(0.2402)	0.2453	(0.2452)	0.2503	(0.2502)
	0.4685	(0.4662)	0.4795	(0.4672)	0.4996	(0.4972)
	0.2300	(0.2305)	0.2310	(0.2345)	0.2485	(0.2495)
	0.7201	(0.7251)	0.7401	(0.7451)	0.7486	(0.7481)
$RMSE^*$	0.1568	(0.1558)	0.1468	(0.1548)	0.1268	(0.1240)
	0.1422	(0.1458)	0.1321	(0.1355)	0.1122	(0.1138)
	0.1423	(0.1448)	0.1222	(0.1238)	0.1123	(0.1146)
	0.1322	(0.1328)	0.1221	(0.1227)	0.1022	(0.1018)
	0.0230	(0.0258)	0.0131	(0.0208)	0.0110	(0.0108)
	0.0540	(0.0558)	0.0242	(0.0251)	0.0220	(0.0231)
$RMSE$	0.1561	(0.1559)	0.1478	(0.1541)	0.1267	(0.1241)
	0.1423	(0.1458)	0.1331	(0.1355)	0.1122	(0.1128)
	0.1423	(0.1448)	0.1222	(0.1237)	0.1124	(0.1148)
	0.1322	(0.1329)	0.1201	(0.1228)	0.1022	(0.1019)
	0.0231	(0.0257)	0.0134	(0.0207)	0.0111	(0.0107)
	0.0542	(0.0550)	0.0241	(0.0241)	0.0220	(0.0230)

Table (2): Model (2) with $\underline{\theta}=(\underline{a}',\underline{c}',\pi,p)'$ with $\underline{a}'=(-0.35,0.35)$, $\underline{c}'=(0.25,0.5)$, $\pi=1/4, p=3/4$

Chapter 5

General Conclusion: Remarks and some perspectives

The focus has been devoted on the extension of some results by FZ on $MS-GARCH$ for the MS-bilinear time series one. So, we have analyzed the probabilistic structure of several subclasses of $MS-$bilinear models especially subdiagonal. First, sufficient conditions for the existence of regular strictly stationary solutions are given for general MS-bilinear model. For the subdiagonal model, a Markovian bilinear (MB) representation is presented for which we are derived conditions ensuring the existence of second (resp. higher)-order stationary solutions. The main advantage of the MB representation is that besides its adaptation with nonlinear effects, it preserves the mathematically tractable $ARMA$ structure. In particular, it was seen in Sect. 4, that the second-order properties do not generally provide sufficient information about the structure. Moreover, it is shown that the power of $MS-$bilinear models having a $MB-$representation has also an $ARMA$ structure. Since, the mixing concept is often required in statistical applications, then we have examined sufficient conditions ensuring the $\beta-$mixing for $MS-$bilinear models having a $MB-$ representation. As an application, we then have established the geometric ergodicity of various $MS-GARCH(1,1)$ models including $MS-EGARCH(1,1)$, $TSGARCH\,(1,1)$, $VGARCH\,(1,1)$ among others. The forecasting problem using $MS-$bilinear models and its comparison with $ARMA$ models seems a problem of interest, so we leave this important issue for future researches.

This thesis investigated the question of parameter estimation for a stationary and ergodic bilinear process ($MS-BL$) in which we allows the coefficients to vary according an unobservable time-homogeneous Markov chain with finite state space. This problem has been previously resolved in the statistical literature for the usual time-constant case. To the best of our knowledge, the statistical inference has received a little attention in literature. So, this thesis, discusses some basic issues concerning the class of $MS - BL$ models including invertibility and the consistency of $QMLE$. We showed that QML estimates perform very well for large sample sizes, not only with a common Gaussian assumption for the noise, but also with general distribution for the error having some moment of finite order or with heteroscedastic error. On the other words and as already mentioned by Pham and Tran [63], the asymptotic normality of $QMLE$ seems to be difficult to establish. One of the most important difficulties is to know whether the partial derivatives of $q^*_{\underline{\theta}}\left(X_t | \underline{\mathcal{X}}_{t-1}\right)$ with respect to $\underline{\theta}$ are integrable. However, specific tools, for instance GMM method as an alternative should be adapted to estimate the $MS - Bl$ model. We leave this important issue for future researches. Also, we considered a distribution-free approach based on MDE. We showed that MD 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.

Bibliography

[1] Abramson, A., Cohen, I. (2007) On the stationarity of Markov-switching *GARCH* processes. Econometric Theory, 23, 485 − 500.

[2] Aknouche, A., Rabehi, N. (2010) On an independent and identically distributed mixture bilinear time series model. Journal of Time Series Analysis 31, 113 − 131.

[3] Andel Jiri. (1976) Autoregressive series with random parameters. Math. Operationsforsch. u. Statist. 7, 735 − 741.

[4] Andel Jiri. (1982) On autoregressive models with random parameters. Commun. Statist. Theory and Meth, 17 − 30.

[5] Baillie, R.T., Chung, H. (2001) Estimation of *GARCH* models from the autocorrelations of the squares of a process. Journal of Time Series Analysis 22 (6), 631 − 650.

[6] Bibi, A. (2003) On the covariance structure of time-varying bilinear models. Stoch. Anal. App. Vol. 21, 25 − 60.

[7] Bibi, A., Aknouche, A. (2010) Stationarity and $\beta-$mixing of general Markov−switching bilinear processes. Comptes Rendus Acad. Sciences Paris Ser. I 348(3),185 − 188.

[8] Bibi, A., Aknouche, A. (2010) Yule−Walker type estimators in periodic bilinear models: strong consistency and asymptotic normality. Statistical Methods and Applications, $19, 1-30$.

[9] Bibi, A., Lessak, R. (2009) On stationarity and $\beta-$mixing of periodic bilinear processes. Statistics & Probability Letters 79,79 − 87.

[10] Bibi, A., Ghezal, A. (2015) On the Markov−switching bilinear processes: stationarity, higher-order moments and $\beta-$mixing. To appear in Stochastics An international journal of probability and stochastic processes.

[11] Bibi, A., Oyet, A. J. (2004) Estimation of some bilinear time series models with time varying coefficients. Stochastic Analysis and Applications $22, 2, 355 - 376$.

[12] Bibi, A., Gautier, A. (2010) Consistent and asymptotically normal estimators for periodic bilinear models. Bulletin of the Korean Mathematical Society 47, 5, $889 - 905$.

[13] Bougerol, P., Picard, N. (1992) Strict stationarity of generalized autoregressive processes. Annals of Probability 20, $1714 - 1730$.

[14] Box, G. E., Jenkins, G. M.(1976) Time series analysis: forecasting and control, Holden-Day, San Francisco. (Revisited edition).

[15] Brandt, A. (1986) The stochastic Equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. Advances in Applied Probability 18, $211 - 220$.

[16] Carrasco, M., Chen, X. (2002) Mixing and moment properties of various $GARCH$ and stochastic volatility models. Econometric Theory 18, $17 - 39$.

[17] Cramer, H. (1961) On some classes of non-stationary stochastic processes, in Proceeding of the fourth Berkeley Symposium on Mathematical statistics and Probability. University of California Press, Berkeley and Los Angeles, Vol. 2, $57 - 78$.

[18] Francq, C. (1999) $ARMA$ models with bilinear innovations. Stochastic Models, 15, $29 - 52$.

[19] Francq, C., Zakoïan, J-M. (2000a) Multivariate $ARMA$ models with generalized autoregressive linear innovation. Stochastic Analysis and Applications, 18, $231 - 260$.

[20] Francq, C., Zakoïan, J-M. (2000b) Estimating weak $GARCH$ representations. Econometric theory, Vol.16, $692 - 728$.

[21] Francq, C., Zakoïan, J. M. (2005) L_2-structures of standard and switching-regime $GARCH$ models. Stochastic Processes and Their Applications 115, $1557 - 1582$.

[22] Francq, C., Zakoïan, J. M. (2001) Stationarity of multivariate Markov$-$switching $ARMA$ models. Journal of Econometrics 102, $339 - 364$.

[23] Francq, C., Zakoïan, J. M. (2002) Comments on the paper by Minxian Yang: "Some properties of vector autoregressive processes with Markov−switching coefficients". Econometric theory, $18, 815-818$.

[24] Francq, C., Roussignol, M. (1997) On white noises driven by hidden Markov chains. Journal of Time Series Analysis 18, $553-578$.

[25] Granger, C. W. J., Anderson, A. (1978) An introduction to bilinear time series models. Gottingen: Vandenhoeck and Ruprecht.

[26] Grahn, T. (1995). A conditional least squares approach to bilinear time series estimation. Journal of Time Series Analysis $16, 509-529$.

[27] Grenier, Y. (1986) Modèle $ARMA$ à coefficients dépendant du temps: estimateurs et applications, Traitement du signal, 3, n4 $-$ 5. $219-233$.

[28] Guégan, D., Pham, D. T. (1989) A note on the estimation of the parameters of the diagonal bilinear model by the method of least squares. Scandinavian Journal of Statistics $16, 129-136$.

[29] Guégan, D. (1994) Séries chronologiques non linéaires à temps discret. Economica.

[30] Haas, M., Mittnik, S., Paolella, M. S. (2004) A new approach to Markov−switching $GARCH$ models. Journal of Financial Econometrics 2, $493-530$.

[31] Hall. M., Oppenheim A. V., Willsky A. (1983) Time-varying parametric modelling of speech. Signal processing, vol 5, n3, $267-285$.

[32] Hall, A. (2005) Generalized method of moments, Oxford University Press.

[33] Hallin, M. (1978) Mixed autoregressive-moving average multivariate processes with time-dependent coefficients, J. Multivariate anal. 8, $567-572$.

[34] Hallin, M. (1986) Non-Stationary q-dependent processes and time-varying moving-average models: Invertibility properties and forecasting problem. Adv. Appl. Prob. 18, $170-210$.

[35] Hallin, M. (1989) Modèles non-stationnaires: séries univariées et multivariées $157-196$. In: Séries chronologiques. Economica.

[36] Hamilton, J. D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 2, 357−384.

[37] Hannan, E. J. (1970) Multiple time series. John Wiley & Sons, Inc.

[38] Hili, O. (2008) Hellinger distance estimation of general bilinear time series models. Statistical Methodology $5, 119 - 128$.

[39] Kim, W. Y., Billard, L. (1990) Asymptotic properties for the first−order bilinear time series model. Communications in Statistics−Theory and Methods $19, 1171 - 1183$.

[40] Kim, W.Y., Billard, L., Basawa, I. V. (1990) Estimation for the first−order diagonal bilinear time series model. Journal of Time Series Analysis 11, $215 - 229$.

[41] Kingman, J. F. C. (1968) The ergodic theory of subadditive stochastic processes. Journal of the Royal Statistical Society Series B 30, $499 - 510$.

[42] Krishnamurthy, V., Rydén, T. (1998) Consistent estimation of linear and nonlinear autoregressive models with Markov regime. Journal of Time Series Analysis 19, $291 - 307$.

[43] Kristensen, D. (2009) On stationarity and ergodicity of the bilinear model with applications to the $GARCH$ models. Journal of Time Series Analysis 30, $125 - 144$.

[44] Lee, O. (2005) Probabilistic properties of a nonlinear $ARMA$ process with Markov switching. Communications in Statistics: Theory and Methods, 34, $193 - 204$.

[45] Leroux, B. G. (1992) Maximum−likelihood estimation for hidden Markov models. Stochastic Processes and their Applications 40, $127 - 143$.

[46] Lima, R., Rahibe, M. (1994) Exact Lyapunov exponent for infinite products of random matrices. Journal of Physics A: Mathematical and General 27, $3427 - 3437$.

[47] Liporace. L.A. (1975) Linear estimation of non-stationary signals. J. Acoust. Soc. Amer., Vol 58, n6 $1288 - 1295$.

[48] Liu, J. and Brockwell, P. J. (1988) On the general bilinear time series model. J. Appl. Prob 25, $553 - 564$.

[49] Liu, J. (1989a) A simple condition for the existence of some stationary bilinear time series. J. Time ser. Anal., Vol. 8, 33 − 39.

[50] Liu, J. (1989b) On the existence of a general multiple bilinear time series. J. Time ser. Anal., Vol.10, 341 − 355.

[51] Liu, J. (1992a). Spectral radius Kronecker products and stationarity. J. Time ser. Anal., Vol. 13, 319 − 325.

[52] Liu, J. (1992b) Higher order moments and limit theory of a general time series. Techn. Report British Columbia Univ.

[53] Liu, J. (1988) On the general bilinear time series model. Journal of Applied Probability 25, 553 − 564.

[54] Liu, J. (1990) Estimation for some bilinear time series. Stochastic Models 6, 649 − 665.

[55] Liu, J. C. (2006) Stationarity of a Markov−switching $GARCH$ model. Journal of Financial Econometrics 4, 573 − 593.

[56] Maravall, A. (1983) An application of nonlinear time series forecasting. Journal of business & economic statistics 1, 66 − 74.

[57] Mélard, G. (1985) Analyse des données chronologiques. Les presses de l'université de Montreal.

[58] Mendel, J. M. (1973) Discrete techniques of parameter estimation. Marcel Dekker, INC. New York.

[59] Meyn, S. P., Tweedie, R.L. (2009) Markov chains and stochastic stability (2nd edition). Cambridge university press.

[60] Mohler.R. R. (1988) Nonlinear time series and signal processing. Lecture notes in control and information sciences N106. Berlin: Springer Verla.

[61] Nicholls, D.F. and Quinn, B.G. (1982) Random Coefficient Autoregressive Models: An Introduction, Springer-Verlag, New York.

[62] Peel, D., Davidson, J. (1998) A non−linear error correction mechanism based on the bilinear model. Economics letters $58, 165 - 170$.

[63] Pham, D. T., Tran, L. T. (1981) On the first−order bilinear time series model. Journal of Applied Probability 18, 617 − 627.

[64] Pham, D. T. (1986) The mixing properties of bilinear and generalized random coefficient autoregressive models. Stochastic Processes and their Applications 23, 291 – 300.

[65] Priestley, M. P. (1988) Non-linear and non-stationary time series analysis. Academic Press, London and New York.

[66] Rao, T. S. (1970) The fitting of the non-stationary time series models with time-dependent parameters. J. Royal.Statist. Soc.B.32, 310 – 322.

[67] Shu-Ing Liu. (1985) Theory of bilinear time series models. Commun. Statist. Theory Meth. 14(10), 2549 – 2561.

[68] Stelzer, R. (2009) On Markov−switching $ARMA$ processes: Stationarity, existence of moments and geometric ergodicity. Econometric Theory 25, 43 – 62.

[69] Storti, G. (2006) Minimum distance estimation of $GARCH\,(1,1)$ models. Computational Statistics & Data Analysis $51, 1803 - 1821$.

[70] Straumann, D., Mikosch, T. (2006) Quasi−maximum likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equation approach. The Annals of Statistics 34, 2449 – 2495.

[71] Subba Rao, T. (1981) On theory of bilinear time series models. J.R.Statist. Soc.B.43, 244 – 255.

[72] Subba Rao,T., Gaber, M. M. (1984) An introduction to bispectral analysis and bilinear time series models: Lecture Notes in Statistics No 24, Springer.

[73] Terdik, G. (2000) Bilinear stochastic models and related problems of nonlinear time series, A frequency domain approach. Technical report N98/2. Debrecen University, Hungary.

[74] Tieslau, M., Schmidt, P., Baillie, R. (1996) A minimum distance estimator for long−memory processes. Journal of Econometrics 71, 249 – 264.

[75] Wittwer, G. (1989) Some remarks on bilinear time series models. Statistics 20, 521 – 529.

[76] Xie, Y. (2009) Consistency of maximum likelihood estimators for the regime−switching $GARCH$ model. Statistics $43\,(2)$, 153 – 165.

[77] Yang, M. (2000) Some properties of vector autoregressive processes with Markov−switching coefficients. Econometric theory 16, 23 − 43.

[78] Yao, J. F., Attali, J.G. (2000) On stability of nonlinear *AR* processes with Markov switching. Advances in Applied Probability 32, 394 − 407.

الإحصاء التقاربي لنماذج السلاسل الزمنية ثنائية الخطية بالتغيرات الماركوفية

الـمـلـخص

هذه الأطروحة تتحرى بعض الخصائص الاحتمالية والتطبيقات الإحصائية للسيرورات ثنائية الخطية العامة مع التغيرات الماركوفية (MS-BL)، والتي توفر بشكل ملحوظ ديناميكية غنية ونمذجة لسلوكيات معقدة لنماذج البيانات غير الطبيعية مع التغيرات الهيكلية. في هذه النماذج، المعاملات تتعلق بسلسلة ماركوف لا يمكن ملاحظتها، متجانسة وثابتة ومعرفة على فضاء الحالة المنتهي. ندرس شروط لازمة وكافية للاستقرار (بمعانيه) ووجود العزوم المحدودة بمختلف الرتب و β-خليط لهذه النماذج. وكنتيجة، نلاحظ أن الاستقرار المحلي للسيرورات الأساسية ليست لازمة ولا كافية للحصول على الاستقرار الشامل. كذلك، يتم تقييم دوال التباين المشترك للعملية وقوى السيرورات ونبين أن بنية الرتب الثنائية (العليا على التوالي) مشابهة لبعض السيرورات الخطية وبالتالي تقبل التمثيل ARMA. وننشئ أيضا شروط كافية لنماذج MS-BL من أجل β-خليط والثبات الهندسي. وبعد ذلك نستخدم هذه النتائج لإعطاء شروط كافية لـ β- خليط لعائلة السيرورات MS-GARCH(1,1). ويتم إعطاء عدد من الأمثلة التوضيحية والتطبيقات المتنوعة لشرح النظرية. ثانيا، توضيح المشاكل الأساسية المرتبطة بنماذج MS-BL بمعنى تقدير المعلمات من خلال أسلوب الاحتمال الأقصى (ML). لذا نقدم تفاصيل على خصائص التقارب ML، وعلى وجه الخصوص، نناقش شروط الكفاءة القوية.

وفي الأخير، نستخدم أسلوب آخر لتوضيح المشاكل الأساسية المرتبطة مع نماذج MS-BL أي تقدير المعلمات بواسطة مقدر الحد الأدنى L_2 للمسافة (MDE). وبالتالي نقدم تفاصيل على خصائص التقارب لـ MDE، وعلى وجه الخصوص، نناقش شروط الكفاءة والتقارب الطبيعي. ويتم عرض التجارب العددية لمحاكاة مجموعات البيانات لتسليط الضوء على النتائج النظرية.

الكلمات المفتاحية

السيرورات الثنائية بتغيرات ماركوفية، الاستقرارية، الاستقرارية من الرتبة الثانية، التمثيل ARMA ، الثبات الهندسي، β-خليط، العكسية، شبه-الاحتمال الأقصى، الكفاءة القوية، q-تتعلق بسلسة ماركوف، تقدير الحد الأدنى للمسافة.

Asymptotic Statistics For Markov-Switching Bilinear Time Series Models

Abstract

This thesis investigates some probabilistic properties and statistical applications of general Markov-switching bilinear processes (MS-BL) that offers remarkably rich dynamics and complex behavior to model non Gaussian data with structural changes. In these models, the parameters are allowed to depend on unobservable time-homogeneous and stationary Markov chain with finite state space. So, some basic issues concerning this class of models including necessary and sufficient conditions ensuring the existence of ergodic stationary (in some sense) solutions, existence of finite moments of any order and β-mixing are studied. As a consequence, we observe that the local stationarity of the underlying process is neither sufficient nor necessary to obtain the global stationarity. Also, the covariance functions of the process and its power are evaluated and it is shown that the second (resp. higher)-order structure is similar to a some linear processes, and hence admit ARMA representation. We establish also sufficient conditions for the MS-BL model to be β-mixing and geometrically ergodic. We then use these results to give sufficient conditions for β-mixing of a family of MS-GARCH(1,1) processes. A number of illustrative examples are given to clarify the theory and the variety of applications. Secondary, we illustrate the fundamental problems linked with MS-BL models, i.e., parameters estimation by considering a maximum likelihood (ML) approach. So, we provide the detail on the asymptotic properties of ML, in particular, we discuss conditions for its strong consistency.

Finally, we used another approach for illustrate the fundamental problems linked with MS-BL models, i.e., parameters estimation by a minimum L_2-distance estimator (MDE). So, we provide the detail on the asymptotic properties of MDE, in particular, we discuss conditions for its consistency and asymptotic normality. Numerical experiments on simulated data sets are presented to highlight the theoretical results.

Keywords

Markov-switching bilinear processes, Stationarity, second-order stationarity, ARMA representation, Geometric ergodicity, β-Mixing, Invertibility, Quasi-maximum likelihood, Strong consistency, q-dependent Markov chain, minimum distance estimation.