A Class of Switching Regimes Autoregressive Driven Processes with Exogenous Components

Joseph Tadjuidje Kamgaing * Hernando Ombao †
Richard A. Davis ‡

July 3, 2008

Abstract

In this paper, we will develop a data-driven mixture of vector autoregressive models with exogenous components. The process is assumed to change regimes according to an underlying Markov process. In contrast to the hidden Markov setup, we allow the transition probabilities of the underlying Markov process to depend on past time series values and exogenous variables. Such processes have potential applications to modeling brain signals. For example, brain activity at time t (measured by electroencephalograms) will can be modeled as a function of both its past values as well as exogenous variables (such as visual or somatosensory stimuli). In this paper, we establish stationarity, geometric ergodicity and the existence of moments for these processes under suitable conditions on the parameters of the model. Such properties are important for understanding the stability properties of the model as well as deriving the asymptotic behavior of various statistics and model parameter estimators.

1 Introduction

In this paper, we develop a class of Markov switching models that are useful for modeling time series that are marked by potentially sudden changes in certain features. The motivation behind this work comes primarily from applications in neuroscience. For example, let Y_t be the electromagnetic activity at some location on the scalp of a subject exposed to an external stimulus indexed by U_t. The objective is to model brain activity at time t as a function of external stimuli U_t and past values Y_{t-1}, Y_{t-2}, and so on. In the standard Markov switching model setup, the dynamics of the model are assumed to change from one state to another as governed by an underlying Markov chain Q_t. Usually the number of states in the Markov chain is assumed known and the probability transition matrix is either constant or a function of previous values of the observations. In our case, we consider models in which the transition probabilities are functions of both lagged values of the process and the applied stimulus U_t. This adaptation allows for a more direct link between changes in the stimulus with the

*University of Kaiserslautern, Department of Mathematics, D-67653 Kaiserslautern, Germany. Email: tadjuidj@mathematik.uni-kl.de
†Brown University, Center for Statistical Sciences, 121 South Main Street, Providence, Rhode Island, USA. Email: ombao@stat.brown.edu
‡Department of Statistics, Columbia University, New York, New York, USA. Email: rdavis@stat.columbia.edu

changes in the underlying dynamics of the process Y_t. We highlight the contributions of this paper. We establish stationarity, geometric ergodicity and existence of moments for these processes under suitable conditions on the parameters of the model. Such properties are important for understanding stability properties of the model. Consequently, we derive the asymptotic behavior of various statistics and the asymptotic distribution of model parameter estimators.

For many time series, it is often found that linear and other stationary models do not provide adequate description of some of the key features in the data. This has spurred the development of new time series models that can capture a wider range of dynamics from the mean structure to other forms of dependence functions. Often, standard linear models serve as the building blocks in the specification of these new models. For example, Prado et al (2000) and Davis et al (2007) use autoregressive (AR) models in which the coefficients are piecewise constant. Priestley (1965) and Dahlhaus (1997) developed locally stationary time series models which include, as special cases, time-varying ARMA models. Other classes of non-stationary models use spectral representations based on localized functions as stochastic building blocks. See, for example, Nason, von Sachs and Kroisandt (1998) which used wavelets and Ombao, von Sachs and Guo (2005) which used the SLEX (smooth localized complex exponentials). Moreover, Chen and Tsay (1992) and Cai, Fan and Zhang (2002,2004) model ARMA-type processes where the ARMA parameters are modeled as some general functionals of time. In addition to the threshold autoregressive model (TAR) in Tong (1983) and a number of its variants such as the SETAR, there are also some linear models with Markov switching regimes, for example Smith and West (1983) as well as Gordon and Smith (1990) on monitoring renal transplants. Hamilton (1989) adapted the Markov switching models developed by Lindgren (1978) to detect changes between growth periods in the economy. Tadjuidje (2005) gives an application of a Markov switching model to financial data in the context of asset management and risk analysis where the trend and volatility functions are estimated by single layer neural networks. Geometric ergodicity was established for these models by Stockis, Tadjuidje, and Franke (2007).

As a starting point for the description of our model, consider the first-order vector autoregression with an a exogenous variable [VARX(1)] for the bivariate time series $Y_t = [Y_{1,t}, Y_{2,t}]'$ defined by

$$Y_{1,t} = \mu_1 + \alpha_1 Y_{1,t-1} + \beta_1 Y_{2,t-1} + \gamma_1 U_{t-1} + \epsilon_{1,t} \tag{1.1}$$
$$Y_{2,t} = \mu_2 + \alpha_2 Y_{1,t-1} + \beta_2 Y_{2,t-1} + \gamma_2 U_{t-1} + \epsilon_{2,t}. \tag{1.2}$$

The above can be compactly rewritten in a vector representation as

$$Y_t \quad = \quad \mu + M Y_{t-1} + \gamma U_{t-1} + \epsilon_t \tag{1.3}$$

where $\mu = (\mu_1, \mu_2)'$, $M = \begin{pmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{pmatrix}$, $\gamma = (\gamma_1, \gamma_2)'$, and $\epsilon(t) = (\epsilon_1(t), \epsilon_2(t))'$ is a white noise sequence with mean $\mathbf{0}$ and covariance matrix Σ (written $\mathrm{WN}(0, \Sigma)$). VARX models are extensively used to model macro-economic data (details are reported in Hannan and Deistler (1988)).

In this paper, we shall develop a process that is defined as a mixture of a finite number of VARX processes. Each one of these VARX models completely describes a particular dynamic for the mean and correlation structure and the effect of the exogenous process. As time evolves, changes in the mean structure as well as the auto-correlation and cross-correlation structures are governed by an underlying Markov "state" process Q_t with K states. At time t, only one VARX process is "activated" and *that* process is determined by the value of Q_t. One unique feature of our model is that, unlike other similar formulations of these models, we model the transition probability matrix for Q_t as also dependent on t with entries that are functions of lagged observations and the exogenous process. The formal specification of our model is as follows:

$$Y_t \quad = \quad \sum_{k=1}^{K} S_{tk}(\mu_k + M_k Y_{t-1} + \gamma_k U_{t-1}) + \epsilon_t \text{ with } S_{tk} = \begin{cases} 1, & \text{if } Q(t) = k, \\ 0, & \text{otherwise,} \end{cases}$$

where $\{\epsilon_t\} \sim \mathrm{WN}(\mathbf{0}, \Sigma)$. The *latent* process Q_t is the "hidden" process that is reflected by the auxiliary variables S_{tk}. This is similar to the setup described in Francq and Zakoïan (2001), Stockis, Tadjuidje and Franke (2007).

There are similarities between the model described here and the one introduced in Lai and Wong (2001) [LW]. Both our model and LW use the logistic function for the state conditional probability. Whereas LW *assumes* stationarity and strong mixing of the observed process and the exogenous processes, this paper *establishes* stationarity and α-mixing under more general conditions on the conditional transition probability of the state process.

The remainder of the paper is structured as follows. In Section 2, we give a more complete description of our model without an exogenous component and describe some of its properties. In particular, we show that under some restrictions on the model parameters, the process is geometrically ergodic. In Section 3, the model is extended to include an exogenous process $U(t)$ follows a Markov switching AR processes. Section 4 contains results on the moment structure of these processes. Section 5 considers estimation of the model parameters. A conditional likelihood approach is used and it is shown that the resulting estimators are consistent and asymptotically normal. The performance of these estimates are evaluated via a simulation study in Section 6.

2 A mixture of Autoregressive Driven Processes

For ease in presenting ideas, we shall consider a simplified model only for a univariate response (although the theory and estimation methodology also apply to a multivariate response) and without an exogenous variable. In this case the probability of the response being in one regime depends only on some past realizations of the process. This setting differs from some other approaches often considered in the literature by the fact that we do not impose any particular structure to the hidden process, e.g., a discrete stationary Markov structure such as those usually considered for HMM. For this model, we will derive some probabilistic properties such as the stability of the model, define the conditional likelihood and investigate the inference of the parameter estimates. Prior to establishing these, we shall demonstrate in this section that the proposed model is geometrically ergodic.

2.1 Model Definition and Basic Properties

We now study the model

$$Y_t \quad = \quad \sum_{k=1}^{K} S_{tk} M_k(Y_{t-1}, \cdots, Y_{t-p}) + \epsilon(t) \text{ with } S_{tk} = \begin{cases} 1, & \text{if } Q_t = k, \\ 0, & \text{otherwise,} \end{cases} \tag{2.1}$$

where

$$\mathbb{P}(S_{tk} = 1 \,|\, \mathcal{F}_{t-1}) = h_k(Y_{t-1}, \cdots, Y_{t-r}), \tag{2.2}$$

whereby $\mathcal{F}_{t-1} = \sigma\{Y_s : s \leq t-1\}$ is the σ-algebra generated by the past observations of the process up to the time $t-1$ and independent of ϵ_t. In the remainder of this paper we will consider without loss of generality $p = r$ and define $\mathcal{Y}_{t-1} = (Y_{t-1}, \cdots, Y_{t-p})'$. We can then derive the conditional expectation of Y_t given $\mathcal{F}_{t-1}$, i.e.,

$$\mathbb{E}(Y_t \,|\, \mathcal{F}_{t-1}) = \sum_{k=1}^{K} h_k(\mathcal{Y}_{t-1}) M_k(\mathcal{Y}_{t-1}) \tag{2.3}$$

and the conditional density of Y_t given $\mathcal{F}_{t-1}$ is given by

$$f(Y_t \,|\, \mathcal{F}_{t-1}) = \sum_{k=1}^{K} h_k(\mathcal{Y}_{t-1}) g(Y_t - M_k(\mathcal{Y}_{t-1})) \tag{2.4}$$

where g is the density function of ϵ_t. Note that we have not yet specified the functions $h_k, k = 1, \cdots, K$. In general, to prove the geometric ergodic property of the observed process, some regularity conditions on these functions will suffice.

An Example: A Mixture of First Order Autoregressive Processes. In this example, we consider the model (defined in equation (2.1)) above, but now the S_{tk} is not state indicator of a hidden Markov process and instead defined by its conditional probability given the σ-algebra $\mathcal{F}_{t-1}$ as follows:

$$\begin{aligned} \mathbb{P}(S_{tk} = 1 \,|\, \mathcal{F}_{t-1}) &= h_k(Y_s, s \leq t-1) \\ &= \frac{\xi_k}{1 + \exp\{\alpha + \gamma e^2_{t-1,k}\}} \end{aligned} \tag{2.5}$$

for some $0 < \xi_k < 1, \omega > 0, \gamma \leq 0$ and where

$$e^2_{t,k} = [Y_t - M_k(\mathcal{Y}_{t-1})]^2.$$

We then choose

$$M_k(\mathcal{Y}_{t-1}) = \sum_{i=1}^{p} \alpha_{ki} Y_{t-i},$$

i.e., the different regimes of the process can be regarded as linear autoregressive models of order p. Thus, the observed process can be written as

$$Y_t = \sum_{k=1}^{K} S_{tk} (\sum_{i=1}^{p} \alpha_{ki} Y_{t-i}) + \epsilon_t,$$

with

$$\mathbb{P}(S_{tk} = 1 \,|\, \mathcal{F}_{t-1}) = h_k(\mathcal{Y}_{t-1}) = 1 - \sum_{i \neq k} h_i(\mathcal{Y}_{t-1}).$$

Based on the model definition it is easy to see that $\mathcal{Y}_t = (Y_t, \cdots, Y_{t-p+1})'$ is a first order Markov chain for which we will establish geometric ergodicity. Usually, for the investigation the asymptotic behavior of the parameter estimates (consistency or asymptotic normality), e.g., of the log-likelihood function we require some moment assumptions for the observed process to make use of the ergodic theorem or central limit theorems for mixing processes.

That the above stochastic process defined in equation (2.1) satisfies e.g., β-mixing conditions can be regarded as a consequence of the asymptotic stability of the model, compare Davydov (1973) who establishes a closed relationship between the geometric ergodic property of a given stochastic process and the β-mixing property of a given stochastic process.

2.2 Stability Conditions

The goal of this section is to provide a set of conditions under which the model satisfies a geometric ergodic property. To achieve this goal, we apply some key results on the stability of Markov chains, from Meyn and Tweedie(1993).

2.2.1 Model Assumption

A. 2.1. Assumption on the residuals
The ϵ_t are i.i.d. random variables, independent of $\mathcal{F}_{t-1}$ and ϵ_t has a continuous positive probability density function g that is positive on $\mathbb{R}$. Furthermore, $\mathbb{E}\epsilon_t^2 < \infty$.

A. 2.2. Assumption on the conditional probability functions

1) for each $k \in \{1, \cdots, K\}, h_k : \mathbb{R}^p \longrightarrow [0, 1]$ is a continuous function.

2) There exist δ_L, δ_U such $0 < \delta_L \leq h_k(\mathcal{Y}_{t-1}) \leq \delta_U < 1$, $k = 1, \cdots, K$.

A. 2.3.

$$M_k(Y(t-1), \cdots, Y(t-p)) = \sum_{i=1}^{p} \alpha_{ki} Y_{t-i}$$

and

$$\delta_U \sum_{i=1}^{p} \sum_{k} \left(|\alpha_{ki}| \sum_{j=1}^{p} |\alpha_{kj}| \right) < 1. \tag{2.6}$$

2.2.2 Some Preliminary Results

Lemma 2.1. *Let us assume A.2.1 and A.2.2, then $\mathcal{Y}_t$ is a first order Markov chain, additionally it is a Feller chain, i.e., for each bounded continuous function $f_{bc} : \mathbb{R}^p \to \mathbb{R}$, the function of x given by $\mathbb{E}(f_{bc}(\mathcal{Y}_t) \,|\, \mathcal{Y}_{t-1} = x)$ is also bounded continuous.*

The proof is given in 7.1.

Once we have established that $\mathcal{Y}_t$ is a Feller chain, the topological considerations on our space ($\mathbb{R}^p$), compare Feigin and Tweedie (1985), imply that any compact set A with $\phi(A) > 0$ is a small set, whereby ϕ is the Lebesgue measure on $\mathbb{R}^p$. We have now defined a Markov chain for which we can use some stability results compare Meyn and Tweedie (1993) to derive its geometric ergodic property.

In order to prove our main results, we need to show that the assumptions of the following Theorem of Feigin and Tweedie (1985) hold.

Theorem 2.1. *(Feigin and Tweedie(1985), Theorem 1) Suppose $\{\Phi_t\}$ is a Feller Chain, that there exists a measure ϕ and a compact set A with $\phi(A) > 0$ such that*

i) $\{\Phi_t\}$ *is ϕ-irreducible*

ii) there exists a non-negative continuous function $V : E \to \mathbb{R}$ satisfying

$$V(x) \geq 1 \, for \, x \in A \tag{2.7}$$

and for some $\beta > 0$

$$\mathbb{E}[V(\Phi_t) \,|\, \Phi_{t-1} = x] \leq (1-\beta)V(x) \, for \, x \in A^c. \tag{2.8}$$

Then, $\{\Phi_t\}$ is geometrically ergodic.

2.2.3 Geometric Ergodicity

Theorem 2.2. *Let A.2.1 - A.2.3 hold. Then, the* $\mathcal{Y}_t = (Y_t, \cdots, Y_{t-p-1})'$ *is a geometrically ergodic Markov process.*

The theorem provides conditions for the existence of a unique strictly stationary ergodic solution of the model defined by equation 2.1, it also implies that the process converges to its stationary distribution at geometric rate even if it does not start from the stationary state. However, some of the assumptions used here can be weakened.

For example, the assumption on g can be relaxed to consider density functions that are only almost everywhere positive; we avoid this assumption here for sake of simplicity. For the later considerations, one can refer to Bhattacharya and Lee(1995) who present a proof in the autoregressive setting. Also, the moment condition $\mathbb{E}\epsilon_t^2 < \infty$ can be relaxed, compare Franke, Stockis and Tadjuidje (2007) who prove the geometric ergodicity for CHARME models, a type of generalized (Hidden Markov) mixture of nonlinear and non parametric AR-ARCH Models. Furthermore, the linearity assumption of the M_k is relaxed in the latter paper where they assume that the M_k are not linear but linearly dominated under some considerations. The technique of proof they use can be adapted here with little modification.

Corollary 2.1. *Under the assumption of Theorem 2.2*

$$\mathbb{E}_\pi Y_t^2 < \infty. \tag{2.9}$$

Moreover,

$$\int_{\mathbb{R}^p} \pi(dx) \left| \mathbb{E}(Y_t^2 \,|\, \mathcal{Y}_0 = x) - \int_{\mathbb{R}^p} \pi(dy) y^2 \right| = O(\rho^t), \; t \to \infty, \tag{2.10}$$

where π *is the stationary distribution of* $\mathcal{Y}_t$ *and* $0 < \rho < 1$.

The corollary follows from Theorem 2 in Feigin and Tweedie (1985) and the definition of the V function used in the proof of Theorem 2.2. It proves the existence of the second moment for the observed process Y_t and says that the $\mathbb{E}(Y_t^2 \,|\, \mathcal{Y}_0 = x)$ converges to $\int_{\mathbb{R}^p} \pi(dy)y^2$ with a geometric rate. The corollary gives us a flavor of the existence of moments. A full investigation of the existence of higher order moments is devoted to a separate section. The proof is given in Section 7.2.

3 A Mixture of Autoregressive Driven Processes with Exogenous Components

In this section we introduce the more general setting of our model, i.e., we consider the model with exogenous components, which here is assumed to follow an autoregressive model. However, we will not assume the exogenous variable to be stationary in all regimes although the main result of this section will imply the asymptotic stationarity of the observed series regarded as a mixture of processes. We now present the model and derive its asymptotic stability.

3.1 Model Definition

Let us define,

$$\begin{cases} Y_t = \sum_{k=1}^K S_{tk}(M_k(\mathcal{Y}_{t-1}) + U_t) + \epsilon_t \\ U_t = \sum_{k=1}^K S_{tk}\Gamma_k(\mathcal{U}_{t-1}) + \eta_t \end{cases} \quad \text{with } S_{tk} = \begin{cases} 1, & \text{if } Q_t = k, \\ 0, & \text{otherwise,} \end{cases} \tag{3.1}$$

where

$$\mathbb{P}(S_{tk} = 1 \,|\, \mathcal{G}_{t-1}, \eta_t) = h_k(\mathcal{Y}_{t-1}, \mathcal{U}_{t-1}), \tag{3.2}$$

$$M_k(\mathcal{Y}_{t-1}) = \sum_{i=1}^{p} \alpha_{ki} Y_{t-i}, \qquad \Gamma_k(\mathcal{U}_{t-1}) = \sum_{r=1}^{q} a_{kr} U_{t-r} \tag{3.3}$$

and $\mathcal{Y}_{t-1} = (Y_{t-1}, \cdots, Y_{t-p})', \mathcal{U}_{t-1} = (U_{t-1}, \cdots, U_{t-q})'$ and ϵ_t and η_t are independent of $\mathcal{G}_{t-1} = \sigma\{Y_s, U_s : \, s \leq t-1\}$, the σ-algebra generated by the realizations of the joint process $(\mathcal{Y}_s, \mathcal{U}_s)$ up to the time $t-1$. U_t is the exogenous component of the model, which can be regarded, for example, as a stimulus in the neuroscience framework. Another example is that of daily stock prices, for which the U_t can be regarded as the opening value of a stock index to which the observed process does not belong.

Additionally, we need that the conditional probabilities for Q_t depend on $\mathcal{G}_{t-1}$ and η_t ($\mathbb{P}(S_{tk} = 1 \,|\, \mathcal{G}_{t-1}, \eta_t)$ is not only conditioned on $\mathcal{G}_{t-1}$ but also on η_t) as technical assumption. This is different from the case of the mixture of autoregressive without exogenous component treated previously, since there the conditional probability functions are only conditioned on $\mathbf{Z}_{t-1}$.

Let us rewrite our model in the vector form, i.e.,

$$\mathbf{Z}_t = \sum_k S_{tk} A_k \mathbf{Z}_{t-1} + \zeta_t \tag{3.4}$$

where

$$\mathbf{Z}_t = (Y_t, Y_{t-1}, \cdots, Y_{t-p+1}, U_t, U_{t-1}, \cdots, U_{t-q+1})',$$

$$\zeta_t = (\epsilon_t + \eta_t, \underbrace{0, \cdots, 0}_{p-1}, \eta_t, \underbrace{0, \cdots, 0}_{q-1})', \quad A_k = \begin{bmatrix} A_{1k} & A_{2k} \\ A_{3k} & A_{4k} \end{bmatrix},$$

$$A_{1k}(p{*}p) = \begin{bmatrix} \alpha_{k1} & \cdots & \cdots & \cdots & \alpha_{kp} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \quad A_{2k}(p{*}q) = \begin{bmatrix} a_{k1} & \cdots & \cdots & \cdots & a_{kq} \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix},$$

$$A_{3k}(q{*}p) = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix}, \quad A_{4k}(q{*}q) = \begin{bmatrix} a_{k1} & \cdots & \cdots & \cdots & a_{kq} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.$$

Further, define

$$G = J \otimes \tilde{G} \tag{3.5}$$

whereby $\tilde{G} = \mathbb{E}(\zeta_t \zeta_t')$ and J is the $(p+q) \times (p+q)$ matrix with the only non zero element being a 1 at $J(1,1)$.

This vector representation of our model also covers the vector valued processes. The $(p,q)th$ order process with d-dimensional vector process Y_t is defined as

$$\begin{cases} Y_t = \sum_{k=1}^{K} S_{tk} (\sum_{i=1}^{p} \Lambda_{ki} Y_{t-1}) + U_t + \epsilon_t \\ U_t = \sum_{k=1}^{K} S_{tk} (\sum_{r=1}^{q} \Gamma_{kr} U_{t-1}) + \eta_t \end{cases} \quad \text{with } S_{tk} = \begin{cases} 1, & \text{if } Q(t) = k, \\ 0, & \text{otherwise,} \end{cases} \tag{3.6}$$

where for $i = 1, \cdots, p, r = 1, \cdots, q, k = 1, \cdots, K$ the $\Lambda_{ki}, \Gamma_{kr}$ are $d \times d$ matrices, ϵ_t, η_t are sequences of d-dimensional vector processes. In other words, in this setting we just need to take

$$Z_t = (Y_t', Y_{t-1}', \cdots, Y_{t-p+1}', U_t', U_{t-1}', \cdots, U_{t-q+1}')', \tag{3.7}$$

Further, for the state matrices the component $\alpha_{ki}, a_{kj}, 1$, resp. 0 of the state matrices A_k are replaced by the matrices $\Lambda_{ki}, \Gamma_{kr}, I_{d\times d}$, reps. $0_{d\times d}$ for $i = 1, \cdots, p, r = 1, \cdots, q$. Whereby the $I_{d\times d}$ resp. $(0_{d\times d})$ are the d-dimensional square identity resp. (Null) matrices.
Analogously,

$$\zeta_t = ((\epsilon_t + \eta_t)', \underbrace{0_{d\times 1}', \cdots, 0_{d\times 1}'}_{p-1}, \eta_t', \underbrace{0_{d\times 1}', \cdots, 0_{d\times 1}'}_{q-1})'$$

with ϵ_t, η_t, d-dimensional random vectors and $0_{d\times 1} = (\underbrace{0, \cdots, 0}_{d})'$.

However, for the sake of simplicity proof we will deal one dimensional case, for the observed process as well as the exogenous component, however the more general statement can be obtained in a similar way.

3.1.1 Geometric Ergodicity and Existence of Moments of Higher Order

A. 3.1. ϵ_t and η_t are i.i.d. random variables with $\mathbb{E}\epsilon_t = \mathbb{E}\eta_t = 0$, finite variances, independent of each other and have continuous density functions g_ϵ and g_η that are positive on $\mathbb{R}$.

A. 3.2. (a.) for each $k \in \{1, \cdots, K\}$, $h_k : \mathbb{R}^{p+q} \longrightarrow \mathbb{R}$ is a continuous function and there exist δ_L, δ_U for which $\delta_L \leq h_k \leq \delta_U, k = 1, 2, \cdots, K$;

(b.) the functions M_k and $\Gamma_k, k = 1, \cdots, K$, are defined as

$$M_k(\mathcal{Y}_{t-1}) = \sum_{i=1}^{p} \alpha_{ki} Y_{t-i} \text{ and } \Gamma_k(\mathcal{U}_{t-1}) = \sum_{i=1}^{q} a_{ki} U_{t-i} \text{ respectively.}$$

A. 3.3. $\delta_U \sum_{k=1}^{K} A_k \otimes A_k$ have all eigenvalues with moduli less than one.

Theorem 3.1. *If A.3.1 - A.3.3 hold, then the process Z_t is a geometrically ergodic Markov chain.*

Proof of Theorem 3.1 is given in Section 7.3. All the comments on Theorem 2.2 remain valid for the above theorem. In particular, Corollary 2.2.3 also holds, i.e., the existence of the second order moment is a direct consequence of the above theorem.

Additionally, the choice of the drift function based on the moduli of the eigenvalues to be less than one is quite similar to the choice made in Feigin and Tweedie 1985, choice made for a class of random coefficients autoregressive models that does not include the class of autoregressive driven models with exogenous component that is the object of the current paper.

To illustrate its usefulness, we apply it to a simple example . Indeed, let us consider the situation where $p = q = 1$, i.e.,

$$A_k = \begin{bmatrix} \alpha_k & a_k \\ 0 & a_k \end{bmatrix}.$$

Hence,

$$A_k \otimes A_k = \begin{bmatrix} \alpha_k^2 & \alpha_k a_k & \alpha_k a_k & a_k^2 \\ 0 & \alpha_k a_k & 0 & a_k^2 \\ 0 & 0 & \alpha_k a_k & a_k^2 \\ 0 & 0 & 0 & a_k^2 \end{bmatrix}.$$

We note that all eigenvalues of $\delta_U \sum_k A_k \otimes A_k$ have moduli less than one if and only if

$$\delta_U \sum_k \alpha_k^2 < 1 \tag{3.8}$$

$$\delta_U |\sum_k \alpha_k a_k| < 1 \tag{3.9}$$

$$\delta_U \sum_k a_k^2 < 1 \tag{3.10}$$

As one can observe this conditions can be easily checked. In particular, equations 3.8 and 3.10 are types of stability conditions for the processes Y_t and U_t if we were to study them separately, as one can derive from Theorem 2.2. Equation 3.9 has to be regarded as a type of cross condition. However, for processes of higher order or dimension one will rely on numerical estimation of the eigenvalues of the matrix $\delta_U \sum_k A_k \otimes A_k$, hence, of their moduli.

Now, we provide sufficient conditions for the existence of higher order moments of the process Z_t (as defined in equation (3.4)) that are important for developing the asymptotic normality of our model parameter estimators.

For a matrix A, we define $A^{\otimes n} = A \otimes A \otimes \cdots \otimes A \quad (n\,\text{terms})$, where $\otimes$ is the Kronecker product operator of matrices.

Theorem 3.2. *If A.3.1 and A.3.2 are satisfied and all the eigenvalues of*

$$\tilde{T} = \delta_u \sum_k A_k^{\otimes 2m} \tag{3.11}$$

have moduli less than one and $\mathbb{E}\|\zeta_t\|^{2m} < \infty$ *then*

$$\mathbb{E}_\pi \|Z_t\|^{2m} < \infty. \tag{3.12}$$

Additionally, any $2m$*th-order moment conditional on* $Z_0 = z$ *converges geometrically to the corresponding moment with respect to* $\mathbb{E}_\pi$*, where* π *is the stationary distribution of the process* Z_t*.*

The proof is given in Section 7.5.

4 Asymptotic of the Parameter estimates

Before we study the asymptotic behavior of the parameter given the conditional likelihood, let us first present the relationship between the likelihood and the weighted least squares.

4.1 Likelihood versus Weighted Least Squares

Consider now the observations $\mathcal{Y} = (Y_{-p+1}, \cdots, Y_1, \cdots, Y_n)$ and define the likelihood function as follows,

$$L(\theta, \mathcal{Y}) = \prod_{t=1}^{n} f(Y_t \,|\, \mathcal{Y}_{t-1}) \tag{4.1}$$

with

$$f(Y_t \,|\, \mathcal{Y}_{t-1}) = \sum_k h_k(\beta_k, \mathcal{Y}_{t-1}) g(Y_t - M_k(\mathcal{Y}_{t-1}))$$

whereby, $M_k(\mathcal{Y}_{t-1}) = \alpha_k' \mathcal{Y}_{t-1}$, $h_k = h_k(\beta_k, \mathcal{Y}_{t-1}), \theta = (\beta_1, \alpha_1, \cdots, \beta_K, \alpha_K)$ and recall that

$$\sum_k h_k(\beta_k, \mathcal{Y}_{t-1}) = 1.$$

Jensen inequality yields

$$\log f(Y_t \,|\, \mathcal{Y}_{t-1}) \geq \sum_k h_k(\mathcal{Y}_{t-1}) \log g(Y_t - M_k(\mathcal{Y}_{t-1})).$$

For sake of illustration, let us consider the example $K = 2$ and assume the residuals are i.i.d. $\mathcal{N}(0, 1)$ random variables. Then it holds

$$\log g(Y_t - M_k(\mathcal{Y}_{t-1})) \approx -\frac{(Y_t - M_k(\mathcal{Y}_{t-1}))^2}{2}$$

which implies

$$\log f(Y_t \,|\, \mathcal{Y}_{t-1}) \geq -h_1(\beta_1, \mathcal{Y}_{t-1}) \frac{(Y_t - M_1(\mathcal{Y}_{t-1}))^2}{2} - (1 - h_1(\beta_1, \mathcal{Y}_{t-1})) \frac{(Y_t - M_2(\mathcal{Y}_{t-1}))^2}{2}.$$

Therefore,

$$\begin{aligned} -\sum_t \log f(Y_t \,|\, \mathcal{Y}_{t-1}) \quad &\leq \quad \sum_t \Bigg(h_1(\beta_1, \mathcal{Y}_{t-1}) \frac{(Y_t - M_1(\mathcal{Y}_{t-1}))^2}{2} \\ &\qquad + (1 - h_1(\beta_1, \mathcal{Y}_{t-1})) \frac{(Y_t - M_2(\mathcal{Y}_{t-1}))^2}{2} \Bigg). \end{aligned}$$

Hence, a weighted type least squares approximation for which the Gaussianity of the residuals is implicitly assumed will almost always be sub-optimal(compared to the log-likelihood approximation). However, it might be numerically more efficient to solve a weighted least squares problem in some situations.

4.2 Conditional Likelihood Estimates

For sake of illustration we consider in this section the following model defined in equation (2.1), i.e.,

$$Y_t \quad = \quad \sum_{k=1}^{K} S_{tk} M_k(Y_{t-1}, \cdots, Y_{t-p}) + \epsilon(t) \text{ with } S_{tk} = \begin{cases} 1, & \text{if } Q_t = k, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\mathbb{P}(S_{tk} = 1 \,|\, \mathcal{F}_{t-1}) = h_k(Y_{t-1}, \cdots, Y_{t-q}).$$

Given $\mathbb{Y}_{t-1} = (Y_{t-1}, \cdots, Y_{t-p})$ and $Z_t = (Y_t, Y_{t-1}, \cdots, Y_{t-p})$, we introduce the notations

$$h_k(\theta, \mathbb{Y}_{t-1}) = h_k(Y_{t-1}, \cdots, Y_{t-p}) = h_k(\theta_k, Y_{t-1}, \cdots, Y_{t-p}),$$
$$M_k(\theta, \mathbb{Y}_{t-1}) = M_k(Y_{t-1}, \cdots, Y_{t-p}) = M_k(\alpha_k, Y_{t-1}, \cdots, Y_{t-p}),$$

where

$$\theta = (\alpha_1, \beta_1, \cdots, \alpha_K, \beta_K).$$

We now define

$$g_k(\theta, Z_t) = g(Y_t - M_k(\theta, \mathbb{Y}_{t-1}))$$

and in turn

$$\begin{aligned} f(\theta, Z_t) &= f_\theta(Y_t \,|\, Y_{t-1}, \cdots, Y_{t-p}) \\ &= \sum_k h_k(\theta_k, Y_{t-1}, \cdots, Y_{t-p}) g(Y_t - M_k(\alpha_k, Y_{t-1}, \cdots, Y_{t-p})). \end{aligned}$$

The log-likelihood function is then defined as

$$l_n(\theta) = \sum_{t=1}^{n} \log f(\theta, Z_t) = \sum_t q_t(\theta).$$

Our goal is to study the asymptotic behavior of the maximum likelihood estimate. For establishing consistency, we first introduce a Uniform Law of Large Numbers. Consider $B \subseteq \mathbb{R}_1^d$ a compact set and $\mathbb{C}(B, \mathbb{R}^{d_2})$ the space of continuous functions on B with values in $\mathbb{R}^{d_2}$). It is well known that $\mathbb{C}(B, \mathbb{R}^{d_2})$ equipped with the supremum norm is separable Banach space. Given this consideration we can make use of an almost sure uniform ergodic theorem for separable Banach space as introduce in Ronga Rao (1962) and for which a necessary condition as presented in Straumann and Mikosch (2006) is summarized in the following theorem.

Theorem 4.1. *Let $v_t(\theta)$ be a stationary ergodic random sequence with value in $\mathbb{C}(B, \mathbb{R}^{d_2})$ satisfying*

$$E \sup_{\theta \in B} |v_1(\theta)| < \infty.$$

Then

$$\sup_{\theta \in B} |\frac{1}{n} \sum_{t=1}^{n} v_t(\theta) - v(\theta)| \longrightarrow 0 \quad a.s. \ as \ n \longrightarrow \infty. \tag{4.2}$$

where $v(\theta) = \mathbb{E} v_1(\theta)$ for all $\theta \in B$.

The proof is due to Ronga Rao (1962)

A. 4.1. 1. Consider Y_t is the unique strictly stationary and ergodic solution of (2.1)

2. For all $k = 1, \cdots, K$, the h_k is twice continuously differentiable. Further, assume that g the probability distribution density of ϵ_t is also twice continuously differentiable.

A. 4.2. (Identifiability)

1. Let θ_0, which lives in the interior of a compact support parameter set B, be the unique minimizer of $-\mathbb{E} \log f(\theta, Z_1)$ (where the expectation is taken with respect to $f(\theta_0, Z_1)$.
2. Let $\mathbb{F}_\theta(Y)$ be the distribution of Y given θ. Then,

$$\mathbb{F}_{\theta_1}(Y) = \mathbb{F}_{\theta_2}(Y) \quad \text{iff} \quad \theta_1 = \theta_2$$

This assumption means that for two parameters θ and θ_0, the stationary distributions of Y_t given those parameters will not coincide unless the parameters coincide

A. 4.3. (Moment Conditions)

1. $\mathbb{E}|\log f(\theta_0, Z_1)| < \infty$

2. There exists M_h independent of θ, with $\mathbb{E}M_h < \infty$ such that for all $i = 1, \cdots, d$ and all $k = 1, \cdots, K$
$$\sup_{\theta \in B} \left| \frac{\partial h_k(\theta, \mathbb{Y}_{t-1})}{\partial \theta_i} \frac{1}{h_k(\theta, \mathbb{Y}_{t-1})} \right| \leq M_h$$

3. There exists M_g independent of θ, with $\mathbb{E}M_g < \infty$ such that for all $i = 1, \cdots, d$ and all $k = 1, \cdots, K$
$$\sup_{\theta \in B} \left| \frac{\partial g_k(\theta, Z_t)}{\partial \theta_i} \frac{1}{g_k(\theta, Z_t)} \right| \leq M_g$$

Theorem 4.2. *Assume A.4.1 to A.4.3 hold and define*
$$\hat{\theta}_n = \inf_{\theta \in B} -\frac{l_n(\theta)}{n}.$$
Then $\hat{\theta}_n$ is strongly consistent, i.e.,
$$\hat{\theta}_n \longrightarrow \theta_0 \quad a.s.(n \longrightarrow \infty).$$

Before embarking on the proof, let us present some preliminaries,
$$\frac{\partial f(\theta, Z_t)}{\partial \theta_i} = \sum_k^K \left(\frac{\partial h_k(\theta, \mathbb{Y}_{t-1})}{\partial \theta_i} g_k(\theta, Z_t) + h_k(\theta, \mathbb{Y}_{t-1}) \frac{\partial g_k(\theta, Z_t)}{\partial \theta_i} \right)$$
and
$$\frac{\partial \log f(\theta, Z_t)}{\partial \theta_i} = \frac{\partial f(\theta, Z_t)}{\partial \theta_i} \frac{1}{f(\theta, Z_t)}.$$

Proof: Since Y_t is stationary and ergodic it easy to see that $\log f(\theta, Z_t)$ is also stationary and ergodic and to prove the ULLN it suffices to prove that
$$\mathbb{E} \sup_{\theta \in B} |\log f(\theta, Z_t)| < \infty$$
Making use of a first order Taylor approximation, it follows
$$\log f(\theta, Z_t) \quad = \quad \log f(\theta_0, Z_t) + \left\langle \theta - \theta_0, \frac{\partial f}{\partial \theta}(\theta^*, Z_t) \frac{1}{f(\theta^*, Z_t)} \right\rangle$$
for some θ^* such that $\|\theta^* - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|$. Applying a Cauchy inequality, we have
$$\left| \left\langle \theta - \theta_0, \frac{\partial f}{\partial \theta}(\theta^*, Z_t) \frac{1}{f(\theta^*, Z_t)} \right\rangle \right| \leq \|\theta - \theta_0\| \left\| \frac{\partial f}{\partial \theta}(\theta^*, Z_t) \frac{1}{f(\theta^*, Z_t)} \right\|.$$
Recalling that $\frac{h_k(\theta^*, \mathbb{Y}_{t-1}) g_k(\theta^*, Z_t)}{f(\theta^*, Z_t)} \leq 1$ and applying A.4.3 yield
$$\begin{aligned}
\left\| \frac{\partial f}{\partial \theta}(\theta^*, Z_t) \frac{1}{f(\theta^*, Z_t)} \right\| &\leq \sum_k \sum_i \left| \frac{\partial h_k(\theta^*, \mathbb{Y}_{t-1})}{\partial \theta_i} g_k(\theta^*, Z_t) \frac{1}{f(\theta^*, Z_t)} \right| \\
&\quad + \sum_k \sum_i \left| h_k(\theta^*, \mathbb{Y}_{t-1}) \frac{\partial g_k(\theta^*, Z_t)}{\partial \theta_i} \right| \frac{1}{f(\theta^*, Z_t)} \\
&\leq \sum_k \sum_i \left| \frac{\partial h_k(\theta^*, \mathbb{Y}_{t-1})}{\partial \theta_i} \frac{1}{h_k(\theta^*, \mathbb{Y}_{t-1})} \frac{h_k(\theta^*, \mathbb{Y}_{t-1}) g_k(\theta^*, Z_t)}{f(\theta^*, Z_t)} \right| \\
&\quad + \sum_k \sum_i \left| \frac{\partial g_k(\theta^*, Z_t)}{\partial \theta_i} \frac{1}{g_k(\theta^*, Z_t)} \right| \frac{h_k(\theta^*, \mathbb{Y}_{t-1}) g_k(\theta^*, Z_t)}{f(\theta^*, Z_t)} \\
&\leq d * K(M_h + M_g)
\end{aligned}$$

and the claim follows since $\|\theta - \theta_0\|$ is bounded on B and $\mathbb{E}|\log f(\theta_0, Z_1)| < \infty$ is given.

The almost sure uniform law of large number and the Identifiability A.4.2 condition will imply the almost sure consistency of the parameter estimate; this a direct application of Lemma 3.1 of Pötscher and Prucha (1997). ■

4.2.1 Asymptotic Normality

In this section we establish the asymptotic normality of the parameter estimates under suitable conditions.

A. 4.4. 1. There exists an N_h, with $\mathbb{E}N_h < \infty$ such that for all $i, j = 1, \cdots, d$ and for all $k = 1, \cdots, K$

$$\sup_{\theta \in B} \left| \frac{\partial^2 h_k(\theta, \mathbb{Y}_{t-1})}{\partial\theta_i \partial\theta_j} \frac{1}{h_k(\theta, \mathbb{Y}_{t-1})} \right| \leq N_h$$

2. There exists N_g, with $\mathbb{E}N_g < \infty$ such that for all $i, j = 1, \cdots, d$ and all $k = 1, \cdots, K$

$$\sup_{\theta \in B} \left| \frac{\partial^2 g_k(\theta, Z_t)}{\partial\theta_i \partial\theta_j} \frac{1}{g_k(\theta, Z_t)} \right| \leq N_g$$

A. 4.5. Y_t is α-mixing with exponential decreasing rate.

A. 4.6. The matrix

$$\left(-\mathbb{E} \frac{\partial^2 \log f(\theta_0, Z_1)}{\partial\theta_i \partial\theta_j} \right)_{1 \leq i,j \leq d}$$

is positive definite.

Since

$$\begin{aligned}
\frac{\partial^2 f(\theta, Z_t)}{\partial\theta_i \partial\theta_i} &= \sum_{k=1}^{K} \left(\frac{\partial^2 h_k(\theta, \mathbb{Y}_{t-1})}{\partial\theta_i \theta_j} g_k(\theta, Z_t) + \frac{\partial h_k(\theta, \mathbb{Y}_{t-1})}{\partial\theta_i} \frac{\partial g_k(\theta, Z_t)}{\partial\theta_j} \right. \\
&+ \left. \frac{\partial h_k(\theta, \mathbb{Y}_{t-1})}{\partial\theta_j} \frac{\partial g_k(\theta, Z_t)}{\partial\theta_i} h_k(\theta, \mathbb{Y}_{t-1}) \frac{\partial^2 g_k(\theta, Z_t)}{\partial\theta_i \partial\theta_j} \right).
\end{aligned}$$

and

$$\frac{\partial^2 \log f(\theta, Z_t)}{\partial\theta_i \partial\theta_j} = \frac{\partial^2 f(\theta, Z_t)}{\partial\theta_i \partial\theta_j} \frac{1}{f(\theta, Z_t)} - \frac{\partial f(\theta, Z_t)}{\partial\theta_i} \frac{1}{f(\theta, Z_t)} + \frac{\partial f(\theta, Z_t)}{\partial\theta_j} \frac{1}{f(\theta, Z_t)},$$

The stationarity and mixing properties of Z_t imply that $\frac{\partial f(\theta, Z_t)}{\partial\theta} \frac{1}{f(\theta, Z_t)}$ and $\frac{\partial^2 f(\theta, (Z_t)}{\partial\theta\partial\theta'} \frac{1}{f(\theta, Z_t)}$ are sequences of stationary sequences with value in $\mathbb{C}(B, \mathbb{R}^d)$ and $\mathbb{C}(B, \mathbb{R}^{d \times d})$, respectively. Furthermore, these processes are α-mixing with the same rate as that of the process Y_t.

Using A.4.1 - A.4.3, for $\hat{\theta}_n = \inf_{\theta \in B} -\frac{l_n(\theta)}{n}$ we have $\hat{\theta}_n \stackrel{a.s.}{\to} \theta_0, n \longrightarrow \infty$. Additionally, θ_0 is an interior point of B and therefore for n large enough,

$$\begin{aligned}
\frac{\partial l_n(\hat{\theta}_n)}{\partial\theta} &= 0 \\
&= \frac{\partial l_n(\theta_0)}{\partial\theta} + \frac{\partial^2 l_n(\theta_n^{**})}{\partial\theta\partial\theta'} (\hat{\theta}_n - \theta_0).
\end{aligned}$$

for some θ_n^{**} satisfying $\|\theta_n^{**} - \theta_0\| < \|\hat{\theta}_n - \theta_0\|$. Hence,

$$\frac{\partial^2 l_n(\theta_n^{**})}{\partial\theta\partial\theta'}(\hat{\theta}_n - \theta_0) = -\frac{\partial l_n(\theta_0)}{\partial\theta}.$$

Now, using CLT for α-mixing processes with geometric decreasing rate, see e.g. Doukhan et al. (1994), we have

$$\frac{1}{\sqrt{n}}\frac{\partial l_n(\theta_0)}{\partial\theta} \xrightarrow{\mathcal{D}} \mathcal{N}(0, V)$$

where

$$V = \lim_{n\to} \frac{1}{n}\mathbb{E}\frac{\partial l_n(\theta_0)}{\partial\theta}\frac{\partial l_n(\theta_0)}{\partial\theta'}.$$

This limit exists by the assumption on the mixing rate but might be degenerate. On the other hand, applying the ULLN on $\mathbb{C}(B, \mathbb{R}^{d\times d})$ and since $\theta_n^{**} \xrightarrow{\mathcal{D}} \theta_0$ one obtains

$$-\frac{\partial^2 l_n(\theta_n^{**})}{\partial\theta\partial\theta'} \xrightarrow{a.s.} \left(-\mathbb{E}\frac{\partial^2 \log f(\theta_0, Z_1)}{\partial\theta_i\partial\theta_j}\right)_{1\le i,j\le d}.$$

We summarize the foregoing results in the following theorem.

Theorem 4.3. *If A.4.3 to A.4.6 hold then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, W),$$

where

$$W = \left(-\mathbb{E}\frac{\partial^2 \log f(\theta_0, Z_1)}{\partial\theta\partial\theta'}\right)^{-1} V \left(-\mathbb{E}\frac{\partial^2 \log f(\theta_0, Z_1)}{\partial\theta\partial\theta'}\right)^{-1}.$$

4.2.2 An Application of the Asymptotic Results

To illustrate the asymptotic results, we consider the case $K = 2$ and define

$$Y_t \quad = \quad \sum_{k=1}^{2} S_{tk}\alpha_k Y_{t-1} + \epsilon_t, \tag{4.3}$$

where $\{\epsilon_t\}$ are i.i.d. $\mathcal{N}(0, \sigma^2)$. The transition probability are defined via

$$\mathbb{P}(S_{t1} = 1 \mid \mathcal{F}_{t-1}) = h_k(Y_{t-1}) = \frac{\gamma}{1 + \exp(\omega + \beta|Y_{t-1}|)}$$

for some $0 < \gamma < 1, \omega > 0$ and $\beta \le 0$. Further, define the model parameter as $\theta = (\alpha_1, \alpha_2, \omega, \beta, \gamma)$.

Corollary 4.1. *Assuming the model parameter θ belongs to a compact set and and the optimal parameter of the likelihood for the model defined in (4.3) is identifiable. Then,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, W).$$

Proof: Since the parameter of the model belongs to the interior of a compact set, the h_1 is bounded away from zero and 1, therefore we can find α_1 and α such that our model satisfies the condition of Theorem 2.2, i.e., the process Y_t is strictly stationary and geometric ergodic, hence α-mixing with geometric rate. In addition by an application of Theorem 3.2 it follows the existence of the fourth moment of Y_t. Finally, applying Theorem 4.3, we have the consistency and asymptotic normality of the parameter estimates. ■

5 A Numerical Illustration

In this section we illustrate the performance of the conditional maximum likelihood estimator for the model specified in Section 4.2.1. The values of the parameters for this simulation are indicated in the table below. The maximum likelihood estimates were computed for samples of size $n = 500$ and $n = 5000$. The summary of the performance of the conditional maximum likelihood estimates is summarized in the table below. The columns contain the mean and the standard deviation of the respective estimates based on 10000 replications. The numerical solutions to the conditional likelihood equation are obtained using `fmincon` from the Matlab optimization toolbox.

True Values	$\alpha_1 = -0.71$	$\alpha_2 = 1.55$	$\omega = 1$	$\beta = -2$	$\gamma = 0.8$	$\sigma = 1$
Estimators	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\omega}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\sigma}$
n=500						
Mean	-0.6978	1.1475	1.1033	-2.1151	0.8206	0.9717
Standard Deviation	0.03	0.0616	0.8591	1.0945	0.0608	0.0387
n=5000						
Mean	-0.6997	1.1505	0.8287	-1.8101	0.8078	0.9739
Standard Deviation	0.01	0.02	0.3324	0.3524	0.0173	0.01

Table 5.1: Sumarry of results from a simulation study based on model given in 4.3. Results are based on 10000 replications.

Figure 1 and 2 illustrate the histogram of the parameter estimates for the sample size $n = 500$ and $n = 5000$ respectively .

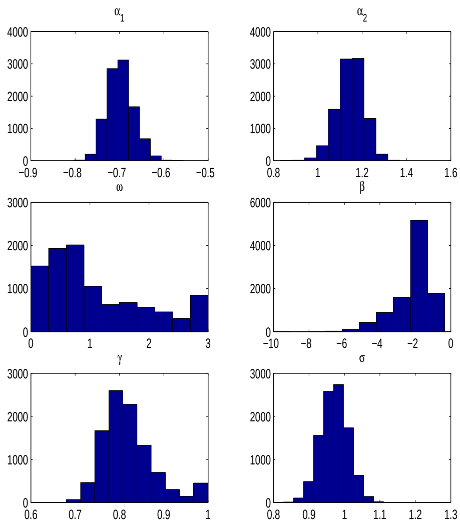

Figure 1: 500 Observations and 10000 Replications

One can observe that the estimates of $\alpha_1, \alpha_2, \gamma$ and σ perform reasonably well. appear to be reasonable. The estimation of ω, however, can still be improved. From the figures, we also see that the sampling distributions (especially for $\hat{\gamma}$) appear more normally distributed for $n = 5000$.

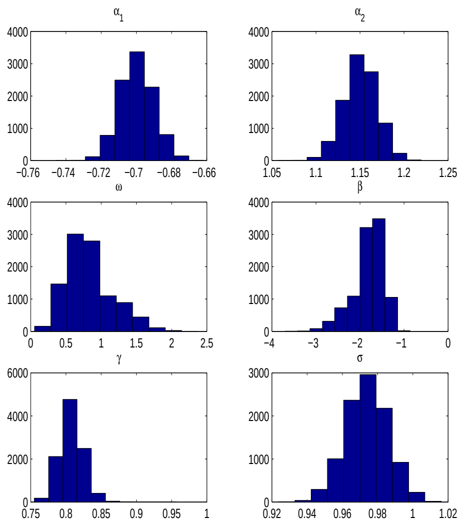

Figure 2: 5000 Observations and 10000 Replications

6 Conclusion and Future Work

We developed a process Y_t that is defined as a mixture of a finite number of VARX processes. Each one of these VARX models completely describes a particular dynamic for the mean and correlation structure and the effect of the exogenous process. However, as time evolves, changes in the mean structure as well as the auto-correlation and cross-correlation structures are governed by an underlying Markov "state" process Q_t, having K states so that at time t there is only one "active" VARX process which is determined by the value of Q_t. One unique feature of our model is that the transition probability matrix for Q_t is also dependent on t with entries that are functions of lagged observations and an exogenous process U_t. A conditional likelihood estimation procedure is used and it is shown that the resulting estimators are consistent and asymptotically normal. There are other several challenges as we further investigate the model and its applications. One issue is the choice of the number of states K - which often times need to be selected via information-based procedures (Akaike Information criterion, Bayesian-Schwarz information criterion). Moreover, for a selected value K, it is implicit that there are indeed K distinct and identifiable VARX mixtures present in the data. This leads to the second issue which is that of mixture identifiability, which is for example addressed for the hidden Markov mixture of Neural networks in Stockis, Tadjuidje, and Franke (2008). One approach to studying this problem is by adapting, e.g., the the conditions of irreducibility in Fariñas et al (2004) or Hwang and Ding (1997) to our model.

References

Bhattacharya, R. and Lee, C. (1995). On geometric ergodicity of nonlinear autoregressive models. *Statistics & Probability Letters* **2**2, 439–440

Cai, Z., Fan, J. and Li, R. (2000). Functional coefficient regression models for non-linear time series. *Journal of the American Statistical Association,* **95**, 941-956.

Cai, Z., Fan, J. and Yao, Q. (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society Series B,* **65**, 57-80.

Chen, R. and Tsay, R. (1993). Functional coefficient autoregressive models. *Journal of the American Statistical Association,* **88**, 298-308.

DAHLHAUS, R. (1997). Fitting Time Series Models to Nonstationary Processes. *Annals of Statistics*, **2**5, 1-37.

DAVIS, R.A., LEE, T.C.M. AND RODRIGUEZ-YAM, G.A. (2005). Structural Break Estimation for Non-stationary Time Series Models.

DAVYDOV Y. (1973) Mixing conditions for Markov chains. *Theor. Probab. Appl.*, **1**8, 2, 321-338

DOUKHAN, P., MASSART, P. AND RIO, E. (1994). The functional central limit theorem for strongly mixing processes. *Annales de l'Institut Henri Poincaré - Probablities et Statistiques*, **30**, 1, 63-82.

DUCKROW, R. AND SPENCER, S. (1992). Regional coherence and the transfer of ictal activity during seizure in the medial temporal lobe. *Electroencephalography and Clinical Neurophysiology,* **8**2, 415-422.

EICHLER, M. (2005). A graphical approach for evaluating effective connectivity in neural systems. *Philosophical Transactions of the Royal Society B,* **3**60, 953-967.

EICHLER, M. (2006). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, in press.

FARIÑAS, M.S., PEDREIRA, C.E., MEDEIROS, M.C. (2004), Local Global Neural Networks: A New Approach for Nonlinear time Series modeling, *J. Amer. Statist. Assoc.* **99**, 1092-1107.

FEIGIN, P.D. AND TWEEDIE, R.L. (1985). Random Coefficient Autoregressive Processes: A Markov Chain Analysis of stationarity and Finiteness of Moments. *Journal of Time Series Analysis*,**6**, 1-14.

FRANCQ, C. AND ZAKOÏAN, J.-M. (2001). Stationarity of multivariate Markov switching ARMA models. *Journal of Econometrics*, **1**02, 2, 339-364.

GORDON, K. AND SMITH, A. (1990). Modeling and monitoring biomedical time series. *Journal of the American Statistical Association,* **8**5, 328-337.

HAMILTON, J. (1989). A new approach to the economic analysis of non stationary time series and the business cycle. *Econometrica,* **5**7, 357-384.

HWANG, J.T.G., DING, A.A (1997). Prediction Intervals for Artificial Neural networks, *J. Amer. Statist. Assoc.* **9**2, 748-757.

KITAGAWA, G. (1987). Non-Gaussian state-space modeling of non-stationary time series (with discussion). *Journal of the American Statistical Association,* **8**2, 1041-1063.

LAI, T. L. AND WONG S. P-S. (2001). Stochastic Neural Networks with Application to Nonlinear Time Series. *Journal of the American Statistical Association,* **v**ol. 96,455, 968-981.

LINDGREN, G. (1978). Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics,* **5**, 81-91.

MEYN, S.P. AND TWEEDIE, R.L. (1993). *Markov Chain and Stochastic Stability,* Springer-Verlag, London.

NASON, G., VON SACHS, R. AND KROISANDT, G. (2000). Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum. *Journal of the Royal Statistical Society, Ser. B,* **62**, 271–292.

OMBAO, H., VON SACHS, R. AND GUO, W. (2005). The SLEX Analysis of Multivariate Non-Stationary Time Series. *Journal of the American Statistical Association,* **1**00, 519-531.

PÖTSCHER, B.M. AND PRUCHA, I.R. (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* New York: Springer-Verlag.

PRADO, WEST AND KRYSTAL (1999). Evaluation and Comparison of EEG Traces: Laten Structure in Nonstationary Time Series *Journal of the American Statistical Association,* **9**4, 446, 375-387

PRIESTLEY, M. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society, Series B,* **28**, 228–240.

RANGA RAO R. (1962). Relation between weak and uniform convergence of measures with applications. *Ann. Math. Stat.,* **3**3, 659-680.

SMITH, A. AND WEST, M. (1983). Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics,* **3**9, 867-878.

STOCKIS, J.-P., TADJUIDJE, K. J. AND FRANKE, J. (2007). On Geometric Ergodicity of CHMARME Models. *Manuscript, University of Kaiserslautern* .

STOCKIS, J.-P., TADJUIDJE, K. J. AND FRANKE, J. (2008). A note on the identifiability of the conditional expectation for the mixtures of neural networks. *Statistics and Probability Letters,* **7**8, 739-742.

STRAUMANN D AND MIKOSCH T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Annals of Statistics,* **3**4, 2449-2495.

TADJUIDJE, K. J. (2005). Competing Neural Networks as Models for Non stationary Financial Time Series *Ph.D. Dissertation, University of Kaiserslautern, http://kluedo.ub.uni-kl.de/volltexte/2005/1818/pdf/dfin.pdf*

TONG, H. (1983). *Threshold Models in Nonlinear Time Series Analysis.* New York: Springer-Verlag.

7 Appendix: Proofs

7.1 Proof of Lemma 2.1

We will proceed in three steps for the proof of this Lemma. For sake of clarity we will first present the conditional density, derive the Markov property of our chain and finally its Feller property. **P**roof: It is easy to show that $\mathcal{Y}_t = (Y_t, \cdots, Y_{t-p+1})$ is a Markov chain and for a given Borel set, $\mathbb{P}(\mathcal{Y}_t \in A \,|\, \mathcal{Y}_{t-1} = x)$, $(x = (x_{t-1}, \cdots, x_{t-p})')$ for a given Borel set A. Without loss of generality, we consider $A = A_p \times A_{p-1} \times \cdots \times A_1$. If there exists an $i \in \{1, \cdots, p-1\}$ such that $x_{t-i} \notin A_{p-i}$, it trivially follows that

$$\mathbb{P}(\mathcal{Y}_t \in A \,|\, \mathcal{Y}_{t-1} = x) = 0. \tag{7.1}$$

Therefore, in the remaining we consider only the Borel set for which $x_{t-i} \in A_{p-i}$, for all $i \in \{1, \cdots, p-1\}$. In this setting $\mathbb{P}(\mathcal{Y}_t \in A \,|\, \mathcal{Y}_{t-1} = x)$ is reduced to $\mathbb{P}(Y_t \in A_p \,|\, \mathcal{Y}_{t-1} = x)$. Furthermore, without loss of generality we only consider the Borel sets of the form defined previously with $A_p = (-\infty, y_p)$. Under these considerations

$$\mathbb{P}(Y_t \in A_p \,|\, \mathcal{Y}_{t-1} = x) = \sum_{k=1}^{K} h_k(x)\mathbb{G}(y_p - M_k(x)),$$

where $\mathbb{G}$ is the cumulative distribution function of the residuals ϵ_t. Hence, the conditional probability kernel is defined as

$$p(y \,|\, x) = \sum_{k=1}^{K} h_k(x) g(y_p - M_k(x)),$$

where g is the density function of ϵ_t.

Finally, that $\mathcal{Y}_t$ is a Feller chain follows directly from the assumptions made on the density of ϵ_t, h_k and m_k, $k \in \{1, \cdots, K\}$. More precisely, if we consider a bounded continuous function $f_{bc} : \mathbb{R}^p \longrightarrow \mathbb{R}$, then

$$\mathbb{E}(f_{bc}(\mathcal{Y}_t) \,|\, \mathcal{Y}_{t-1} = x) = \sum_k h_k(x) \int f_{bc}(y, x^*) g(y - M_k(x)) dy \tag{7.2}$$

$(x = (x_1, \cdots, x_p), x^* = (x_1, \cdots, x_{p-1}))$, which is obviously bounded and continuous. Thus, $\mathcal{Y}_t$ fulfills the Feller property. ■

7.2 Proof of Theorem 2.2

Proof: We have already shown that $\mathcal{Y}_t$ is a first order Markov chain that has the Feller property. Now, we need to show that under our assumptions this chain satisfies the conditions of Theorem 2.1.

We prove that it is λ-irreducibility with $\lambda =$ Lebesgue, it suffices to show that for any Borel set $A \in \mathcal{B}^p$ with positive Lebesgue measure, i.e., $\lambda(A) > 0$ implies, $P^2(\mathcal{Y}_0, A) > 0$. By definition, one has

$$\begin{aligned}
P^2(\mathcal{Y}_0, A) &= \mathbb{P}(\mathcal{Y}_2 \in A \,|\, \mathcal{Y}_0) \\
&= \iint_A q(Y_2, Y_1 \,|\, Y_0, Y_{-1}) dY_1 dY_2 \\
&= \iint_A \frac{r(Y_2, Y_1, Y_0, Y_{-1})}{g(Y_0, Y_{-1})} dY_1 dY_2) \\
&= \iint_A \frac{f(Y_2 \,|\, Y_1, Y_0) s(Y_1, Y_0, Y_{-1})}{g(Y_0, Y_{-1})} dY_1 dY_2 \\
&= \iint_A f(Y_2 \,|\, Y_1, Y_0) f(Y_1 \,|\, Y_0, Y_{-1})) dY_1 dY_2
\end{aligned}$$

with

$$f(Y_2 \,|\, Y_1, Y_0) = \sum_k h_k(\mathcal{Y}_1) g(Y_2 - M_k(\mathcal{Y}_1))$$

and

$$f(Y(1) \,|\, Y_0, Y_{-1})) = \sum_k h_k(\mathcal{Y}_0) g(Y_1 - M_k(\mathcal{Y}_0)).$$

Recall that there exists $\delta_L > 0$ such that $h_k \geq \delta_L$. Consequently we derive

$$P^2(\mathcal{Y}_0, A) \geq \delta_L^2 \iint_A \left(\sum_k g(Y_2 - M_k(\mathcal{Y}_1)) \right) \left(\sum_k g(Y_1 - M_k(\mathcal{Y}_0)) \right) dY_1 dY_2$$

Since g is positive everywhere it follows $\mathbb{P}(\mathcal{Y}_2 \in A \,|\, \mathcal{Y}_0) > 0, \forall\, \mathcal{Y}_0$ and this implies $\mathcal{Y}_t$ is λ-irreducible.

Next we find a function $V : \mathbb{R}^p \to [0, \infty)$, that satisfies the conditions (ii) of Theorem 2.1. Consider

$$V(\mathcal{Y}_t) = 1 + Y_t^2 + b_1 Y_{t-1}^2 + \cdots + b_{p-1} Y_{t-p+1}^2 \tag{7.3}$$

and recall that

$$M_k(Y(t-1), \cdots, Y(t-q)) = \sum_{i=1}^{p} \alpha_{ki} Y_{t-i}. \tag{7.4}$$

It follows by the definition of Y_t that

$$\begin{aligned} V(\mathcal{Y}_t) \;\; &= \;\; 1 + \sum_k S_{tk} (\sum_{i=1}^{p} \alpha_{ki} Y_{t-i})^2 + \epsilon_t^2 \\ &\quad + 2\epsilon_t \sum_k S_{tk} (\sum_{i=1}^{p} \alpha_{ki} Y_{t-i}) + b_1 Y_{t-1}^2 + \cdots + b_{p-1} Y_{t-p+1}^2. \end{aligned} \tag{7.5}$$

and yet,

$$\begin{aligned} (\sum_{i=1}^{p} \alpha_{ki} Y_{t-i})^2 \;\; &= \;\; \sum_{i=1}^{p} \alpha_{ki}^2 Y_{t-i}^2 + \sum_{i,j,i\neq j} \alpha_{ki} \alpha_{kj} Y_{t-i} Y_{t-j} \\ &\leq \;\; \sum_{i=1}^{p} \alpha_{ki}^2 Y_{t-i}^2 + \sum_{i,j,i\neq j} |\alpha_{ki}||\alpha_{kj}|(Y_{t-i}^2 + Y_{t-j}^2), \end{aligned}$$

since $2ab \leq a^2 + b^2$. Thus,

$$(\sum_{i=1}^{p} \alpha_{ki} Y_{t-i})^2 \;\; \leq \;\; \sum_{i=1}^{p} (|\alpha_{ki}| \sum_{j=1}^{p} |\alpha_{kj}|) Y_{t-i}^2. \tag{7.6}$$

It follows that

$$\sum_{k=1}^{K} S_{tk} (\sum_{i=1}^{p} \alpha_{ki} Y_{t-i}))^2 \;\; \leq \;\; \sum_{k=1}^{K} S_{tk} (\sum_{i=1}^{p} (|\alpha_{ki}| \sum_{j=1}^{p} |\alpha_{kj}|) Y_{t-i}^2), \tag{7.7}$$

and hence,

$$\begin{aligned} V(\mathcal{Y}_t) \;\; &\leq \;\; 1 + \sum_{k=1}^{K} S_{tk} \left(\sum_{i=1}^{p-1} (|\alpha_{ki}| \sum_{j=1}^{p} |\alpha_{kj}| + b_i) Y_{t-i}^2 \right) + \epsilon_t^2 \\ &\quad + \sum_{k=1}^{K} S_{tk} \left(|\alpha_{kp}| \sum_{j=1}^{p} |\alpha_{kj}| \right) Y_{t-p}^2 + 2\epsilon_t \sum_k S_{tk} \left(\sum_{i=1}^{p} \alpha_{ki} Y_{t-i}) \right). \end{aligned} \tag{7.8}$$

Since

$$\mathbb{E}\left(\epsilon_t \sum_k S_{tk}\left(\sum_{i=1}^{p} \alpha_{ki} Y_{t-i}\right) \mid \mathcal{Y}_{t-1} = x\right) = 0, \tag{7.9}$$

it follows that

$$\begin{aligned}
\mathbb{E}(V(\mathcal{Y}_t) \mid \mathcal{Y}_{t-1} = x) \leq 1 &+ \sum_{k=1}^{K} h_k(x)\left(\sum_{i=1}^{p-1}(|\alpha_{ki}| \sum_{j=1}^{p} |\alpha_{kj}| + b_i) x_{t-i}^2\right) \\
&+ \sum_{k=1}^{K} h_k(x)\left(|\alpha_{kp}| \sum_{j=1}^{p} |\alpha_{kj}|\right) x_{t-p}^2 + \sigma^2 \\
&\leq 1 + \sum_{i=1}^{p-1}\left[\delta_U\left(\sum_{k=1}^{K}(|\alpha_{ki}| \sum_{j=1}^{p} |\alpha_{kj}|\right) + b_i\right] x_{t-i}^2 \\
&+ \left[\delta_U \sum_{k=1}^{K}\left(|\alpha_{kp}| \sum_{j=1}^{p} |\alpha_{kj}|\right)\right] x_{t-p}^2 + \sigma^2.
\end{aligned} \tag{7.10}$$

Finally we need to find a small set for which the drift criterion defined in Theorem 2.1 is satisfied. We will proceed here with a two stage constructive proof. Let us first find conditions on the $b_i's$ such that

$$\begin{aligned}
x_{t-1}^2 + b_1 x_{t-2}^2 + \cdots + b_{p-1} x_{t-p}^2 \geq &\sum_{i=1}^{p-1}\left[\delta_U\left(\sum_{k=1}^{K}(|\alpha_{ki}| \sum_{j=1}^{p} |\alpha_{kj}|\right) + b_i\right] x_{t-i}^2 \\
&+ \left[\delta_U \sum_{k=1}^{K}\left(|\alpha_{kp}| \sum_{j=1}^{p} |\alpha_{kj}|\right)\right] x_{t-p}^2.
\end{aligned} \tag{7.11}$$

This inequality is satisfied if, for example,

$$\begin{aligned}
&\delta_U(\sum_{k=1}^{K}(|\alpha_{k1}| \sum_{j=1}^{p} |\alpha_{kj}|) + b_1 < 1 \\
&\delta_U(\sum_{k=1}^{K}(|\alpha_{k2}| \sum_{j=1}^{p} |\alpha_{kj}|) + b_2 < b_1 \\
&\vdots \\
&\delta_U(\sum_{k=1}^{K}(|\alpha_{k(p-1)}| \sum_{j=1}^{p} |\alpha_{kj}|) + b_{p-1} < b_{p-2} \\
&\delta_U \sum_{k=1}^{K}(|\alpha_{kp}| \sum_{j=1}^{p} |\alpha_{kj}|) < b_{p-1}.
\end{aligned}$$

Under the constraint

$$\delta_U \sum_{i=1}^{p} \sum_k (|\alpha_{ki}| \sum_{j=1}^{p} |\alpha_{kj}|) < 1, \tag{7.12}$$

we can then find b_i's satisfying equation (7.11) so that

$$\mathbb{E}(V(\mathcal{Y}_t) \mid \mathcal{Y}_{t-1} = x) \quad \leq K_1 V(x) + \sigma^2 \text{ for some } K_1 < 1. \tag{7.13}$$

Now, we need to find at least for a compact set with positive measure C, a K_2 satisfying $0 < K_1 < K_2 < 1$ and

$$\mathbb{E}(V(\mathcal{Y}_t)\,|\,\mathcal{Y}_{t-1} = x) \quad \leq \quad K_1 V(x) + \sigma^2 < K_2 V(x),\ \forall\, x \in C. \tag{7.14}$$

This inequality is satisfied if

$$V(x) > \frac{\sigma^2}{K_2 - K_1}.$$

Choosing,

$$C = \left\{ x : V(x) \leq \frac{\sigma^2}{K_2 - K_1} \right\},$$

and any K_2 that satisfies $0 < K_1 < K_2 < 1$, (7.14) follows, which completes the proof . ■

7.3 Proof of Theorem 3.1

Proof:

The Markov Property of Z_t can be derived similarly to that of $\mathcal{Y}_t$ in Lemma 2.1 to show that Z_t is a Markov chain. For its Feller property, let us consider a bounded continuous function $f_{bc} : \mathbb{R}^{p+q} \longrightarrow \mathbb{R}$ and compute

$$\begin{aligned}
\mathbb{E}(f_{bc}(Z_t)\,|\,Z_{t-1} = x) &= \int f_{bc}(Z_t) p(Z_t\,|\,Z_{t-1} = x) dZ_t \\
&= \int f_{bc}(Z_t) p(Y_t\,|\,U_t, Z_{t-1} = x) p(U_t\,|\,\mathbf{Z}_{t-1} = x) dZ_t
\end{aligned}$$

whereby

$$\begin{aligned}
p(Y_t\,|\,U_t, Z_{t-1} = x) &= p(Y_t\,|\,\eta_t, Z_{t-1} = x) \\
&= \sum_{k=1}^{K} h_k(Z_{t-1}) g_\epsilon(Y_t - M_k(\mathcal{Y}_{t-1}) - U_t)
\end{aligned} \tag{7.15}$$

and

$$p(U_t\,|\,\mathbf{Z}_{t-1} = x) = \sum_{k=1}^{K} h_k(Z_{t-1}) g_\eta(U_t - \Gamma_k(\mathcal{U}_{t-1})). \tag{7.16}$$

Hence, $\mathbb{E}(f_{bc}(Z_t)\,|\,Z_{t-1} = x)$ is bounded continuous by the continuity assumptions on the g_ϵ, g_η, boundedness and continuity assumptions on the $h_k, k = 1, \cdots, K$, and consequently the Feller property follows.

Following the proof presented for the case without exogenous component we once more consider for the sake of simplicity the case where $p = q = 2$. Thus, $Z_t = (Y_t, Y_{t-1}, U_t, U_{t-1})'$. We then need to investigate the strict positivity of, e.g., $P^2(A, x)$ for any given Borel set A with with positive Lebesgue measure $\lambda(A) > 0$. By definition,

$$\begin{aligned}
P^2(A, x) &= \mathbb{P}(Z_2 \in A\,|\,Z_0 = x) \\
&= \int_A p(Z_2\,|\,Z_0 = x) dZ_2 \\
&= \int_A \left(p(Y_2\,|\,U_2, \mathcal{U}_1, \mathcal{Y}_1) p(U_2\,|\,\mathcal{Y}_1, \mathcal{U}_1)\right) \left(p(Y_1\,|\,U_1, \mathcal{U}_0, \mathcal{Y}_0) p(U_1\,|\,\mathcal{Y}_0, \mathcal{U}_0)\right) dZ_2 \\
&= \int_A C \times D dZ_2
\end{aligned}$$

whereby

$$C=\left(\sum_{k=1}^{K}h_k(Z_1)g_\epsilon(Y_2-M_k\mathcal{Y}_1-U_2)\right)\left(\sum_{k=1}^{K}h_k(Z_1)g_\eta(U_2-\Gamma_k(\mathcal{U}_1))\right)$$

and

$$D=\left(\sum_{k=1}^{K}h_k(Z_0)g_\epsilon(Y_1-M_k\mathcal{Y}_0-U_1)\right)\left(\sum_{k=1}^{K}h_k(Z_0)g_\eta(U_1-\Gamma_k(\mathcal{U}_0))\right).$$

By the positivity of the densities function g_ϵ, g_η and the lower bound assumption on the $h_k, k=1,\cdots,K$, it follows that $P^2(A,x)>0$, hence, the irreducibility of the Markov chain Z_t.

Consider a positive definite matrix V of the form defined in the previous lemma and define

$$g(x)=1+x'Vx. \tag{7.17}$$

It follows that

$$Z_t'VZ_t \quad = \quad \sum_k S_{tk}Z_{t-1}'A_k'VA_kZ_{t-1}+\zeta_t'V\zeta_t+2\sum_k S_{tk}Z_{t-1}'A_k'V\zeta_t. \tag{7.18}$$

Since

$$\mathbb{E}(\sum_k S_{tk}Z_{t-1}'A_k'V\zeta_t\,|\,Z_{t-1}=x)=0,$$

we derive

$$\begin{aligned}1+\mathbb{E}(Z_t'VZ_t\,|\,Z_{t-1}=x) &= 1+\sum_k h_k(x)x'A_k'VA_kx+\mathbb{E}\zeta_t'V\zeta_t \\ &\le 1+\delta_U\sum_k x'A_k'VA_kx+\mathbb{E}\zeta_t'V\zeta_t \\ &\le 1+x'Vx-x'Wx+\mathbb{E}\zeta_t'V\zeta_t \\ &\le g(x)\left(1-\frac{x'Wx-\mathbb{E}\zeta_t'V\zeta_t}{g(x)}\right).\end{aligned} \tag{7.19}$$

For some $K>1$ define the compact set

$$C=\{x\in\mathbb{R}^q : x'Vx\le K\}.$$

It then follows for every $x\in C^c$

$$g(x)\le\frac{K}{K+1}x'Vx\le 2x'Vx \tag{7.20}$$

hence, after straightforward computations, we obtain

$$\begin{aligned}\frac{x'Wx-\mathbb{E}\zeta_t'V\zeta_t}{g(x)} &= \frac{x'Wx}{1+x'Vx}-\frac{\mathbb{E}\zeta_t'V\zeta_t}{1+x'Vx} \\ &\ge \frac{x'Wx}{2x'Vx}-\frac{\mathbb{E}\zeta_t'V\zeta_t}{1+K} \\ &\ge \frac{x'x\lambda_1(W)}{2x'x\lambda_q(V)}-\frac{\mathbb{E}\zeta_t'V\zeta_t}{K} \\ &\ge \frac{\lambda_1(W)}{2\lambda_q(V)}-\frac{\mathbb{E}\zeta_t'V\zeta_t}{K} \\ &\ge \epsilon,\end{aligned} \tag{7.21}$$

where $\lambda_1(W)$ is the smallest eigenvalue of W and $\lambda_q(V)$ is the largest eigenvalue of V(W and V are positive definite).
Given

$$0 < \epsilon < \frac{1}{2}\frac{\lambda_1(W)}{\lambda_q(V)},$$

one can finally choose

$$K \geq K(\epsilon) = \mathbb{E}\zeta_t' V \zeta_t \left[\frac{1}{2}\frac{\lambda_1(W)}{\lambda_q(V)} - \epsilon\right]^{-1}$$

and conclude that

$$\mathbb{E}(g(Z_t) \mid Z_{t-1} = x) \leq (1-\epsilon)g(x) \,\text{for all}\, x \in C^c.$$

Consequently, the drift conditions are satisfied. ■

7.4 Proof of Lemma 7.1

In this section we present a lemma that establishes the link between a given positive definite matrix and the drift function we need it for the proof of the geometric ergodicity. This lemma will be used to prove Theorem 3.2. Indeed, we make use of the connection between the scalar product, the Kronecker product and the vectorization.

Lemma 7.1. *Suppose W is $q \times q$ positive definite matrix and the eigenvalues of*

$$\delta_U \sum_{k=1} A_k \otimes A_k \tag{7.22}$$

have moduli less than the unity.

(i) If V is defined by

$$vec(V) = \left(I - \delta_U \sum_{k=1} A_k' \otimes A_k'\right)^{-1} vec(W) \tag{7.23}$$

then V is also positive definite.

(ii) For any x,

$$\delta_U \sum_k x' A_k' V A_k x = x'Vx - x'Wx \tag{7.24}$$

Proof: The result in i) follows as in Feigin and Tweedie(1985) from the identity

$$vec(V) = \sum_{j=0}^{\infty} \left(I - \delta_U \sum_{k=1} A_k' \otimes A_k'\right)^{j} vec(W). \tag{7.25}$$

For the results in ii) let us first recall the following identity that connects the Kronecker product $\otimes$ and the vectorization and present some other properties of the Kronecker product.

$$vec(ABC) = (C' \otimes A)vec(B) \tag{7.26}$$

$$(A \otimes B)' = A' \otimes B' \tag{7.27}$$

and for suitable matrices

$$(A \otimes B)(C \otimes D) = AC \otimes BD \tag{7.28}$$

From Lemma 7.1 *i*) it follows that

$$vec(V) - vec(W) = \delta_U \sum_k A_k' \otimes A_k' vec(V) \tag{7.29}$$

we then derive for every x

$$\begin{aligned} x'Vx - x'Wx &= x' \otimes x'(vec(V) - vec(W)) \\ &= x' \otimes x' \delta_U \sum_{k=1}^{K} A_k' \otimes A_k' vec(V) \end{aligned} \tag{7.30}$$

after some intermediate steps, which completes the proof. ■

7.5 Proof of Theorem 3.2

Proof: We present only the proof for the existence of the fourth moments, the other cases being a straightforward adaptation. Recall that for matrices A, B and V, with V symmetric and for vectors X, Y we have the following properties

$$\|AX\| \leq K\|X\| \text{ for some positive } K, \tag{7.31}$$

$$X'VY = Y'VX \leq \lambda_1(V)\|X\|\|Y\|, \tag{7.32}$$

with $\lambda_1(V)$ being the largest eigenvalue of a symmetric positive definite matrix V. Hence,

$$X'A'VBY \leq \lambda_1(V)K\|X\|\|Y\| \tag{7.33}$$

and

$$\|X \otimes Y\| = \|X\|\|Y\|. \tag{7.34}$$

Consider,

$$\begin{aligned} \tilde{Z}_t &= Z_t \otimes Z_t = \left[\sum_k S_{tk} A_k Z_{t-1} + \zeta_t\right] \otimes \left[\sum_k S_{tk} A_k Z_{t-1} + \zeta_t\right] \\ &= \sum_k S_{tk}(A_k Z_{t-1}) \otimes (A_k Z_{t-1}) + \sum_k S_{tk}(A_k Z_{t-1}) \otimes \zeta_t \\ &\quad + \sum_k S_{tk}(\zeta_t \otimes (A_k Z_{t-1})) + \zeta_t \otimes \zeta_t, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \tilde{Z}_t &= \sum_k S_{tk}(A_k \otimes A_k)(Z_{t-1} \otimes Z_{t-1}) + \sum_k S_{tk}(A_k \otimes I_{p+q})(Z_{t-1} \otimes \zeta_t) \\ &\quad + \sum_k S_{tk}(I_{p+q} \otimes A_k)(\zeta_t \otimes Z_{t-1}) + \zeta_t \otimes \zeta_t \\ &= A + B + C + D. \end{aligned}$$

By assumption, all the eigenvalues of $\tilde{T} = \delta_u \sum_k [(A_k \otimes A_k)] \otimes [(A_k \otimes A_k)]$ have moduli less than one and in Lemma 7.1 one can define a positive definite matrix $\tilde{W}$ such that

$$vec(\tilde{V}) = (I - \tilde{T})^{-1} vec(\tilde{W}).$$

Hence,

$$\tilde{T} vec(\tilde{V}) = vec(\tilde{V}) - vec(\tilde{W}). \tag{7.35}$$

Considering a positive definite matrix $\tilde{V}$ satisfying the above conditions, with

$$\tilde{T} = \delta_u \sum_k [(A_k \otimes A_k)] \otimes [(A_k \otimes A_k)] \tag{7.36}$$

and defining $\tilde{z} = z \otimes z$, it follows from an application of Lemma 7.1 that

$$\begin{aligned}
&\mathbb{E}(A' \tilde{V} A \,|\, Z_{t-1} = z) \\
&\quad \leq \delta_U \sum_k \tilde{Z}'_{t-1} \left[(A_k \otimes A_k) \otimes (A_k \otimes A_k))\right]' \tilde{V} \left[(A_k \otimes A_k) \otimes (A_k \otimes A_k))\right] \tilde{Z}_{t-1} \\
&\quad = \tilde{z}' \tilde{V} \tilde{z} - \tilde{z}' \tilde{W} \tilde{z}.
\end{aligned} \tag{7.37}$$

We can also trivially prove $\mathbb{E}(A' \tilde{V} B \,|\, Z_{t-1} = z) = 0$ and analogously

$$\mathbb{E}(B' \tilde{V} A \,|\, Z_{t-1} = z) = \mathbb{E}(A' \tilde{V} C \,|\, Z_{t-1} = z) = \mathbb{E}(C' \tilde{V} A \,|\, Z_{t-1} = z) = 0.$$

Now, using the matrix calculations presented at the beginning of this section, it can be shown that there exist positive constants K_1 and K_2 such that

$$\mathbb{E}(B' \tilde{V} B \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) K_1 \|z\|^2 \mathbb{E}\|\zeta_t\|^2,$$

$$\mathbb{E}(C' \tilde{V} C \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) K_2 \|z\|^2 \mathbb{E}\|\zeta_t\|^2,$$

and

$$\mathbb{E}(D' \tilde{V} D \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) \mathbb{E}\|\zeta_t\|^4.$$

Similarly, there exist positive constant $K_4, \ldots, K_9$ such that

$$\begin{aligned}
\mathbb{E}(A' \tilde{V} B \,|\, Z_{t-1} = z) &= \mathbb{E}(B' \tilde{V} A \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) K_4 \|z\|^3 \mathbb{E}\|\zeta_t\|, \\
\mathbb{E}(A' \tilde{V} C \,|\, Z_{t-1} = z) &= \mathbb{E}(C' \tilde{V} A \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) K_5 \|z\|^3 \mathbb{E}\|\zeta_t\|, \\
\mathbb{E}(A' \tilde{V} D \,|\, Z_{t-1} = z) &= \mathbb{E}(D' \tilde{V} A \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) K_6 \|z\|^2 \mathbb{E}\|\zeta_t\|^2, \\
B' \tilde{V} C \,|\, Z_{t-1} = z) &= \mathbb{E}(C' \tilde{V} B \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) K_7 \|z\|^2 \mathbb{E}\|\zeta_t\|^2, \\
\mathbb{E}(B' \tilde{V} D \,|\, Z_{t-1} = z) &= \mathbb{E}(D' \tilde{V} B \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) K_8 \|z\| \mathbb{E}\|\zeta_t\|^3,
\end{aligned}$$

and

$$\mathbb{E}(C' \tilde{V} D \,|\, Z_{t-1} = z) = \mathbb{E}(D' \tilde{V} C \,|\, Z_{t-1} = z) \leq \lambda_1(\tilde{V}) K_9 \|z\| \mathbb{E}\|\zeta_t\|^3.$$

Therefore, taking $g(\tilde{Z}_t) = 1 + \tilde{Z}_t'\tilde{V}\tilde{Z}_t$, we obtain

$$\begin{aligned}\mathbb{E}(g(\tilde{Z}_t) \mid Z_{t-1} = z) &\leq 1 + \tilde{z}'\tilde{V}\tilde{z} - \tilde{z}'\tilde{W}\tilde{z} \\ &+ \rho_1(\tilde{V})(C_1\|z\|^3\mathbb{E}\|\zeta_t\| + C_2\|z\|^2\mathbb{E}\|\zeta_t\|^2 \\ &\quad +C_3\|z\|\mathbb{E}\|\zeta_t\|^3 + \mathbb{E}\|\zeta_t\|^4),\end{aligned}$$

where $C_1 = K_5 + K_4$, $C_2 = K_1 + K_2 + K_6 + K_7$ and $C_3 = K_8 + K_9$. Now, Letting

$$s(z, \zeta_t) = \rho_1(\tilde{V})(C_1\|z\|^3\mathbb{E}\|\zeta_t\| + C_2\|z\|^2\mathbb{E}\|\zeta_t\|^2 + C_3\|z\|\mathbb{E}\|\zeta_t\|^3),$$

we have

$$\mathbb{E}(g(\tilde{Z}_t) \mid Z_{t-1} = z) \leq g(\tilde{z})\left(1 - \frac{\tilde{z}'\tilde{W}\tilde{z} - s(z, \zeta_t) - \rho_1(\tilde{V})\mathbb{E}\|\zeta_t\|^4}{1 + \tilde{z}'\tilde{V}\tilde{z}}\right). \tag{7.38}$$

Using the techniques in deriving equation (7.20) for the proof of geometric ergodicity, we observe that

$$\frac{\tilde{z}'\tilde{W}\tilde{z} - s(z, \zeta_t)}{1 + \tilde{z}'\tilde{V}\tilde{z}} \geq \frac{\lambda_{min}(\tilde{W})\|\tilde{z}\|^2 - s(z, \zeta_t)}{\lambda_{max}(\tilde{V})\|\tilde{z}\|^2} - \frac{\rho_{max}(\tilde{V})\mathbb{E}\|\zeta_t\|^4}{K} > \epsilon \tag{7.39}$$

and for $K(\epsilon)$ sufficiently large and large values of $\|z\|^4 = \|\tilde{z}\|^2$ (the dominating term in the above equation), one can conclude

$$\mathbb{E}(g(\tilde{Z}_t) \mid Z_{t-1} = z) \leq g(\tilde{z})(1 - \epsilon). \tag{7.40}$$

The existence of the of $4th$-order moment follows. ■