Asymptotic Properties of the Maximum Likelihood Estimator for Markov-switching Observation-driven Models

Frederik Krabbe[a,*]

[a] *Aarhus University*

Abstract

In this paper, we study the asymptotic properties of the maximum likelihood estimator for a so-called Markov-switching observation-driven model. The Markov-switching observation-driven model contains several models proposed in the literature as special cases for which, to the best of our knowledge, no results for the asymptotic properties of the maximum likelihood estimator exist.

1. Introduction

A state space model is a stochastic process $((S_t, Y_t))_{t\in\mathbb{Z}}$ such that (i) $(S_t)_{t\in\mathbb{Z}}$ is an unobserved Markov chain taking values in $\mathrm{S} \subseteq \mathbb{R}$ and (ii) $(Y_t)_{t\in\mathbb{Z}}$ is an observed stochastic process taking values in $\mathrm{Y} \subseteq \mathbb{R}$ such that the conditional distribution of Y_t given $\mathbf{Y}_{-\infty}^{t-1}$ where $\mathbf{Y}_i^j := (Y_i, ..., Y_j)$ and $\mathbf{S}_{-\infty}^{t}$ where $\mathbf{S}_i^j := (S_i, ..., S_j)$ depends only on S_t. An important submodel is the hidden Markov model introduced by Baum and Petrie (1966) in which S is finite.[1] Hidden Markov models have been applied in numerous areas such as speech recognition (Juang and Rabiner (1991)), biology (Churchill (1989)), ecology (Langrock et al. (2012)), and meteorology (Zucchini and Guttorp (1991) and Bulla et al. (2012)), see the book by Zucchini et al. (2016) for more examples, so statistical inference for hidden Markov models is of significant practical importance. In the first paper on hidden Markov models, Baum and Petrie (1966) proved consistency and asymptotic normality of the maximum likelihood (ML) estimator in the case where Y is finite. Later, Leroux (1992) proved consistency of the ML estimator in the general case, and Bickel et al. (1998) proved asymptotic normality of the ML estimator in the general case.

The hidden Markov model is often insufficient to model economic and financial time series. Therefore, Hamilton (1989) introduced the linear Markov-switching autoregressive model to model

[*] I would like to thank Leopoldo Catania, Morten Ørregaard Nielsen, Christian Francq and Jean-Michel Zakoïan for useful discussions. This research is supported by the Danish National Research Foundation (DNRF Chair grant number DNRF154).

[1] This distinction is not made by all authors.

economic growth, see also the book by Krolzig (2013) and the references therein. In the Markov-switching autoregressive model of order $p \in \mathbb{N}$, the conditional distribution of Y_t given $\mathbf{Y}_{-\infty}^{t-1}$ and $\mathbf{S}_{-\infty}^{t}$ depends on both $\mathbf{Y}_{t-p}^{t-1}$ and S_t. Francq and Zakoıan (2001) proved stationarity and ergodicity for the linear Markov-switching autoregressive (moving-average) model and Douc et al. (2004) proved consistency and asymptotic normality of the ML estimator for the Markov-switching autoregressive model in the case where S is compact and not necessarily finite, see also Francq and Roussignol (1998) and Krishnamurthy and Ryden (1998).[2] More recently, Kasahara and Shimotsu (2019) relax some of the assumptions in Douc et al. (2004).

To model financial returns, Cai (1994) and Hamilton and Susmel (1994) introduced the Markov-switching autoregressive conditional heteroscedasticity model (ARCH) as a generalization of the ARCH model introduced by Engle (1982). It is, however, well-known that the generalized ARCH (GARCH) model introduced by Bollerslev (1986) often provides a better fit than the ARCH model when applied to real data. Therefore, Gray (1996) (see also Francq et al. (2001)) and Haas et al. (2004b) introduced (different) Markov-switching GARCH models. In the Markov-switching GARCH model proposed by Gray (1996), the conditional distribution of Y_t given $\mathbf{Y}_{-\infty}^{t-1}$ and $\mathbf{S}_{-\infty}^{t}$ depends on $\mathbf{Y}_{-\infty}^{t-1}$ and $\mathbf{S}_{-\infty}^{t}$ so ML estimation is infeasible whereas in the Markov-switching GARCH model proposed by Haas et al. (2004b), the conditional distribution of Y_t given $\mathbf{Y}_{-\infty}^{t-1}$ and $\mathbf{S}_{-\infty}^{t}$ depends 'only' on $\mathbf{Y}_{-\infty}^{t-1}$ and S_t.

In this paper, we prove stationarity and ergodicity as well as consistency and asymptotic normality of the ML estimator for a so-called Markov-switching observation-driven model. In the Markov-switching observation-driven model, the conditional distribution of Y_t given $\mathbf{Y}_{-\infty}^{t-1}$ and $\mathbf{S}_{-\infty}^{t}$ depends on both $\mathbf{Y}_{-\infty}^{t-1}$ and S_t. The Markov-switching observation-driven model can thus be thought of as a Markov-switching autoregressive model of infinite order, see Example 1 for more details. It contains the Markov-switching GARCH model proposed by Haas et al. (2004b) and many of its variants as well as other models proposed in the literature as special cases for which, to the best of our knowledge, only few results for the probabilistic properties of the model and no results for the asymptotic properties of the ML estimator exist. To do this, we use theory for stochastic recurrence equations as in Straumann and Mikosch (2006) for GARCH models, Blasques et al. (2022) for score-driven models (Creal et al. (2013) and Harvey (2013)), or, more generally, Blasques et al. (2018) for observation-driven models, see also Douc et al. (2017).

The paper is organized as follows. Section 2 introduces the Markov-switching observation-driven model and Section 3 gives some examples of Markov-switching observation-driven models. In Section 4, stationarity and ergodicity of the model is considered. Consistency and asymptotic normality of the maximum likelihood estimator is considered in Section 5. In Section 6, the theory is applied to some of the examples in Section 3. An application to real data is presented in Section 7. Section 8 concludes. All profs are collected in Section 9.

We adopt the following notation. Let $\mathbf{x}$ be an $n \times 1$ vector of real numbers, let $\mathbf{y}$ be an $m \times 1$

[2] The Markov-switching autoregressive conditional heteroscedasticity model discuss later is a special case hereof.

vector of real numbers, and let $\mathbf{X}$ be an $n \times m$ matrix of real numbers. Then, $||\mathbf{x}||_p := (\sum_{i=1}^{n} |x_i|^p)^{1/p}$ is the p-norm of $\mathbf{x}$ and $||\mathbf{X}||_{p,p} := (\sum_{i=1}^{n} \sum_{j=1}^{m} |x_{ij}|^p)^{1/p}$ is the p-norm $\mathbf{X}$. Moreover, let f be a $k \times 1$ vector of real valued functions defined on $\mathbb{R}^n \times \mathbb{R}^m$. Then, $\frac{\partial f}{\partial \mathbf{x}'} := \left(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n} \right)$ is the $k \times n$ matrix of first order derivatives of f and $\frac{\partial f}{\partial \mathbf{x}} := \left(\frac{\partial f}{\partial \mathbf{x}'} \right)'$. If $k = 1$, then $\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{y}'} := \frac{\partial}{\partial \mathbf{y}'} \frac{\partial f}{\partial \mathbf{x}}$ is the $n \times m$ matrix of second order derivatives of f.

2. The Markov-switching Observation-driven Model

In the following, all random variables are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $(S_t)_{t \in \mathbb{Z}}$ be an unobserved Markov chain taking values in $\{1, ..., J\}$ with transition probabilities

$$p_{ij} := \mathbb{P}(S_{t+1} = j \mid S_t = i), \quad i, j \in \{1, ..., J\},$$

and transition probability matrix

$$\mathbf{P} := \begin{bmatrix} p_{11} & \cdots & p_{1J} \\ \vdots & \ddots & \vdots \\ p_{J1} & \cdots & p_{JJ} \end{bmatrix}.$$

An observed stochastic process $(Y_t)_{t \in \mathbb{Z}}$ taking values in $\mathcal{Y} \subseteq \mathbb{R}$ is called a Markov-switching observation-driven model if the conditional distribution of Y_t given $\mathbf{Y}_{-\infty}^{t-1}, \mathbf{Y}_i^j := (Y_i, ..., Y_j)$ and $\mathbf{S}_{-\infty}^{t}, \mathbf{S}_i^j := (S_i, ..., S_j)$ depends only on $\mathbf{Y}_{-\infty}^{t-1}$ and S_t as follows

$$Y_t \mid (\mathbf{Y}_{-\infty}^{t-1}, \mathbf{S}_{-\infty}^{t}) \sim \mathcal{D}_{S_t}(X_{S_t,t}, \upsilon_{S_t}),$$

where $X_{j,t}, j \in \{1, ..., J\}$ is a time-varying parameter taking values in $\mathcal{X}_{\upsilon_j} \subseteq \mathbb{R}$ given by

$$X_{j,t+1} = \phi_j(Y_t, X_{j,t}; \upsilon_j),$$

$\upsilon_j, j \in \{1, ..., J\}$ is a vector of constant parameters taking values in $\Upsilon_j \subseteq \mathbb{R}^{d_j}$, $x \mapsto \phi_j(y, x; \upsilon)$ is a Lipschitz function for all $y \in \mathcal{Y}$, $\upsilon \in \Upsilon_j$ and $j \in \{1, ..., J\}$, $(y, x, \upsilon) \mapsto \phi_j(y, x; \upsilon)$ is continuous on $\mathcal{Y} \times \mathcal{X}_{\upsilon_j} \times \Upsilon_j$ for all $j \in \{1, ..., J\}$, and $x \mapsto \phi_j(y, x; \upsilon)$ is differentiable on $\mathcal{X}_{\upsilon_j}$ for all $y \in \mathcal{Y}$, $\upsilon \in \Upsilon_j$ and $j \in \{1, ..., J\}$. The Markov-switching observation-driven model is called a mixture observation-driven model if $(S_t)_{t \in \mathbb{Z}}$ is an i.i.d. chain.

Note that the time-varying parameter is given by

$$X_{j,t+1} = \phi_j(Y_t, X_{j,t}; \upsilon_j)$$

and *not* by

$$X_{t+1} = \phi_{S_t}(Y_t, X_t; \upsilon_{S_t}).$$

The reason is that the likelihood of a sample generated by a Markov-switching observation-driven model in which the time-varying parameter is given by the latter is impossible to compute for other than very small sample sizes because it depends on the entire history of $(S_t)_{t\in\mathbb{Z}}$. This is often referred to as the so-called path-dependence problem; see, for instance, the discussion in Haas et al. (2004b).

intro The one-step-ahead prediction for the Markov chain, $\pi_{j,t+1|t} := \mathbb{P}(S_{t+1} = j \mid \mathbf{Y}_{-\infty}^{t})$, is given by

$$\pi_{j,t+1|t} = \sum_{i=1}^{J} p_{ij}\pi_{i,t|t},$$

and the filter for the Markov chain, $\pi_{j,t|t} := \mathbb{P}(S_t = j \mid \mathbf{Y}_{-\infty}^{t})$, is given by

$$\pi_{j,t|t} = \frac{\pi_{j,t|t-1} f_{j,t|t-1}(Y_t; X_{j,t}, \upsilon_j)}{f_{t|t-1}(Y_t)},$$

where $f_{t|t-1}(y), y \in \mathcal{Y}$ is the conditional probability density function (pdf) of Y_t given $\mathbf{Y}_{-\infty}^{t-1}$ given by

$$f_{t|t-1}(y) = \sum_{k=1}^{J} \pi_{k,t|t-1} f_{k,t|t-1}(y; X_{k,t}, \upsilon_k),$$

and $f_{j,t|t-1}(y; X_{j,t}, \upsilon_j), y \in \mathcal{Y}$ is the conditional pdf of Y_t given $\mathbf{Y}_{-\infty}^{t-1}$ and $S_t = j$.[3] Hence, $\boldsymbol{\pi}_{t|t-1} := (\pi_{1,t|t-1}, ..., \pi_{J,t|t-1})'$ is given by

$$\boldsymbol{\pi}_{t+1|t} = \mathbf{P}'\boldsymbol{\pi}_{t|t},$$

and $\boldsymbol{\pi}_{t|t} := (\pi_{1,t|t}, ..., \pi_{J,t|t})'$ is given by

$$\boldsymbol{\pi}_{t|t} = \mathbf{F}_t(\boldsymbol{\pi}_{t|t-1})\boldsymbol{\pi}_{t|t-1},$$

where $\mathbf{F}_t(\boldsymbol{\pi}_{t|t-1})$ is a diagonal matrix with generic element

$$[\mathbf{F}_t(\boldsymbol{\pi}_{t|t-1})]_{ii} = \frac{f_{i,t|t-1}(Y_t; X_{i,t}, \upsilon_i)}{\sum_{k=1}^{J} \pi_{k,t|t-1} f_{k,t|t-1}(Y_t; X_{k,t}, \upsilon_k)}.$$

smoothing

The Markov-switching observation-driven model reduces to the hidden Markov model introduced by Baum and Petrie (1966) and studied, among others, by Leroux (1992) and Bickel et al. (1998) when $X_{j,t} \equiv X_j$, and to the Markov-switching autoregressive model introduced by Hamilton (1989) and studied by Douc et al. (2004) when $\phi_j(Y_t, X_{j,t}; \upsilon_j) = \phi_j(Y_t; \upsilon_j)$.

[3]More generally, the h-step-ahead prediction for the Markov chain, $\pi_{j,t+h|t} := \mathbb{P}(S_{t+h} = j \mid \mathbf{Y}_{-\infty}^{t})$, is given by

$$\pi_{j,t+h|t} = \sum_{i=1}^{J} p_{ij}^{(h)}\pi_{i,t|t},$$

where $p_{ij}^{(h)} := \mathbb{P}(S_{t+h} = j \mid S_t = i)$ is the (i, j)'th element of $\mathbf{P}^h := \prod_{i=1}^{h} \mathbf{P}$.

3. Examples

In this section, we give some examples of Markov-switching observation-driven models.

Example 1. *An example of a Markov-switching observation-driven model is*

$$Y_t = X_{S_t,t} + \sigma_{S_t}\varepsilon_t, \quad \sigma_{S_t} > 0, \tag{1}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. standard normal random variables independent of $(S_t)_{t\in\mathbb{Z}}$, and

$$X_{j,t+1} = \omega_j + \alpha_j Y_t + \beta_j X_{j,t},$$

where $\omega_j \in \mathbb{R}$, $\alpha_j \in \mathbb{R}$, and $\beta_j \in \mathbb{R}$. In this model, $\mathcal{Y} = \mathbb{R}$, $\mathcal{D}_{S_t}$ is the normal distribution with mean $X_{S_t,t}$ and variance $\sigma^2_{S_t}$, $\phi_j(y,x;\upsilon) = \upsilon_1 + \upsilon_2 y + \upsilon_3 x$, $\mathcal{X}_{\upsilon_j} = \mathbb{R}$, $\upsilon_j = (\omega_j, \alpha_j, \beta_j, \sigma^2_j)'$, and $\Upsilon_j = \mathbb{R}\times\mathbb{R}\times\mathbb{R}\times(0,\infty)$.

A related model is the Markov-switching autoregressive model of order 1 by Hamilton (1989). This model is given by

$$Y_t = a_{S_t} + b_{S_t}Y_{t-1} + \sigma_{S_t}\varepsilon_t, \quad a_{S_t} \in \mathbb{R}, b_{S_t} \in \mathbb{R}, \sigma_{S_t} > 0,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. standard normal random variables independent of $(S_t)_{t\in\mathbb{Z}}$. It can be shown that if $(Y_t)_{t\in\mathbb{Z}}$ is stationary and ergodic and $|\beta_j| < 1$ for all $j \in \{1,...,J\}$, then

$$X_{j,t} = \frac{\omega_j}{1-\beta_j} + \sum_{i=0}^{\infty} \alpha_j \beta_j^i Y_{t-1-i} \quad a.s.$$

The Markov-switching observation-driven model in Equation (1) can thus be thought of as a Markov-switching autoregressive model of order infinity given by

$$Y_t = a_{S_t} + \sum_{i=0}^{\infty} b^{(i)}_{S_t} Y_{t-1-i} + \sigma_{S_t}\varepsilon_t,$$

where

$$a_{S_t} = \frac{\omega_{S_t}}{1-\beta_{S_t}} \quad \text{and} \quad b^{(i)}_{S_t} = \alpha_{S_t}\beta^i_{S_t}.$$

Example 2. *The Markov-switching Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model by Haas et al. (2004b) is given by*

$$Y_t = \sqrt{X_{S_t,t}}\varepsilon_t,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. standard normal random variables independent of $(S_t)_{t\in\mathbb{Z}}$, and

$$X_{j,t+1} = \omega_j + \alpha_j Y_t^2 + \beta_j X_{j,t},$$

where $\omega_j > 0$, $\alpha_j \geq 0$, and $\beta_j \geq 0$. This is also an example of a Markov-switching observation-driven model where $\mathcal{Y} = \mathbb{R}$, $\mathcal{D}_{S_t}$ is the normal distribution with mean zero and variance $X_{S_t,t}$, $\phi_j(y, x; \upsilon) = \upsilon_1 + \upsilon_2 y^2 + \upsilon_3 x$, $\mathcal{X}_{\upsilon_j} = \left[\frac{\omega_j}{1-\beta_j}, \infty\right)$, $\upsilon_j = (\omega_j, \alpha_j, \beta_j)'$, and $\Upsilon_j = (0,\infty) \times [0,\infty) \times [0,\infty)$.

Note that the Markov-switching GARCH model reduces to the mixture GARCH model by Haas et al. (2004a) if $(S_t)_{t\in\mathbb{Z}}$ is an i.i.d. chain.

Example 3. *Let $y \mapsto F(y; x, \tilde{\upsilon})$ be a cumulative distribution function (cdf) with support $\mathcal{N} \subseteq [0, \infty)$ indexed by the mean x and a vector of parameters $\tilde{\upsilon}$ such that for all $u \in (0, 1)$,*

$$x \leq x^* \quad \Rightarrow \quad F^-(u; x, \tilde{\upsilon}) \leq F^-(u; x^*, \tilde{\upsilon}),$$

where $F^-(u; x, \tilde{\upsilon}) = \inf\{y \in \mathcal{N} : F(y; x, \tilde{\upsilon}) \geq u\}$.

The (present-regime dependent) Markov-switching Positive Linear Conditional Mean (POLI) model by Aknouche and Francq (2022) is given by

$$Y_t = F^-_{S_t}(U_t; X_{S_t,t}, \upsilon_{S_t}),$$

where $(U_t)_{t\in\mathbb{Z}}$ is a sequence of i.i.d. uniform random variables on $[0, 1]$ independent of $(S_t)_{t\in\mathbb{Z}}$, and

$$X_{j,t+1} = \omega_j + \alpha_j Y_t + \beta_j X_{j,t},$$

where $\omega_j > 0$, $\alpha_j \geq 0$, and $\beta_j \geq 0$ This is another example of a Markov-switching observation-driven model where $\mathcal{Y} = \mathcal{N}$, F_{S_t} is the cdf of $\mathcal{D}_{S_t}$, $\phi_j(y, x; \upsilon) = \upsilon_1 + \upsilon_2 y + \upsilon_3 x$, $\mathcal{X}_{\upsilon_j} = \left[\frac{\omega_j}{1-\beta_j}, \infty\right)$, $\upsilon_j = (\omega_j, \alpha_j, \beta_j, \tilde{\upsilon}_j)'$, and $\Upsilon_j = (0,\infty) \times [0,\infty) \times [0,\infty) \times \tilde{\Upsilon}_j$.

4. Probabilistic Properties of the Model

In the following, we restrict our attention to the Markov-switching observation-driven models that can be written as

$$Y_t = \mathbf{1}_{S_t} \mathbf{f}(\boldsymbol{\varepsilon}_t; \mathbf{X}_t, \boldsymbol{\upsilon}), \tag{2}$$

with $\mathbf{1}_{S_t} := (1_{\{S_t=1\}}, ..., 1_{\{S_t=J\}})$ where $(S_t)_{t\in\mathbb{Z}}$ is stationary, irreducible and aperiodic (thus ergodic), and $\mathbf{f}(\boldsymbol{\varepsilon}_t; \mathbf{X}_t, \boldsymbol{\upsilon}) := (f_1(\varepsilon_{1,t}; X_{1,t}, \upsilon_1), ..., f_J(\varepsilon_{J,t}; X_{J,t}, \upsilon_J))'$, $\boldsymbol{\varepsilon}_t := (\varepsilon_{1,t}, ..., \varepsilon_{J,t})' \in \mathcal{E}_1 \times \cdots \times \mathcal{E}_J =: \mathcal{E}$, where $(\varepsilon, x) \mapsto f_j(\varepsilon; x, \upsilon_j)$ is continuous on $\mathcal{E}_j \times \mathcal{X}_{\upsilon_j}$ for all $j \in \{1, ..., J\}$, $\varepsilon_{j,t} \overset{i.i.d.}{\sim} \mathcal{D}^\varepsilon_j(\upsilon^\varepsilon_j)$ for all $j \in \{1, ..., J\}$, $(\varepsilon_{i,t})_{t\in\mathbb{Z}}$ and $(\varepsilon_{j,t})_{t\in\mathbb{Z}}$ are independent for all $i, j \in \{1, ..., J\}$ with $i \neq j$, $\mathbf{X}_t := (X_{1,t}, ..., X_{J,t})'$ is given by

$$\mathbf{X}_{t+1} = \boldsymbol{\phi}^\varepsilon(S_t, \boldsymbol{\varepsilon}_t, \mathbf{X}_t; \boldsymbol{\upsilon}) \tag{3}$$

with

$$[\boldsymbol{\phi}^\varepsilon(S_t, \boldsymbol{\varepsilon}_t, \mathbf{X}_t; \boldsymbol{\upsilon})]_j = \phi_j(\mathbf{1}_{S_t} \mathbf{f}(\boldsymbol{\varepsilon}_t; \mathbf{X}_t, \boldsymbol{\upsilon}), X_{j,t}; \upsilon_j),$$

where $\mathbf{x} \mapsto \boldsymbol{\phi}^{\varepsilon}(s, \boldsymbol{\varepsilon}, \mathbf{x}; \boldsymbol{\upsilon})$ is a Lipschitz function for all $s \in \{1, ..., J\}$ and $\boldsymbol{\varepsilon} \in \mathcal{E}$, and $\boldsymbol{\upsilon} := (\upsilon_1, ..., \upsilon_J)'$. Finally, $(S_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_{j,t})_{t\in\mathbb{Z}}$ are independent for all $j \in \{1, ..., J\}$.[4]

4.1. Stationarity and Ergodicity

Theorem 4.1 gives conditions under which the Markov-switching observation-driven model is stationary and ergodic. It follows from an application of Theorem 3.1 in Bougerol (1993) (see also Theorem 2.8 in Straumann and Mikosch (2006)). To ease the notation, let $\mathcal{X}_{\boldsymbol{\upsilon}} := \mathcal{X}_{\upsilon_1} \times \cdots \times \mathcal{X}_{\upsilon_J}$. Moreover, let

$$\Lambda(\boldsymbol{\phi}_t^{\varepsilon}) := \sup_{\substack{\mathbf{x},\mathbf{y}\in\mathcal{X}_{\boldsymbol{\upsilon}} \\ \mathbf{x}\neq\mathbf{y}}} \frac{||\boldsymbol{\phi}_t^{\varepsilon}(\mathbf{x}) - \boldsymbol{\phi}_t^{\varepsilon}(\mathbf{y})||_2}{||\mathbf{x} - \mathbf{y}||_2},$$

where $\boldsymbol{\phi}_t^{\varepsilon}(\mathbf{x}) := \boldsymbol{\phi}^{\varepsilon}(S_t, \boldsymbol{\varepsilon}_t, \mathbf{x}; \boldsymbol{\upsilon})$.

Theorem 4.1. *Assume that*

(i) *there exists an* $\mathbf{x} \in \mathcal{X}_{\boldsymbol{\upsilon}}$ *such that* $\mathbb{E}[\log^+ ||\boldsymbol{\phi}_t^{\varepsilon}(\mathbf{x}) - \mathbf{x}||_2] < \infty$,

(ii) $\mathbb{E}[\log^+ \Lambda(\boldsymbol{\phi}_t^{\varepsilon})] < \infty$, *and*

(iii) *there exists an* $r \in \mathbb{N}$ *such that*

$$\mathbb{E}\left[\log \Lambda\left((\boldsymbol{\phi}_t^{\varepsilon})^{(r)}\right)\right] < 0,$$

where $(\boldsymbol{\phi}_t^{\varepsilon})^{(r)}(\mathbf{x}) := \boldsymbol{\phi}_t^{\varepsilon} \circ \cdots \circ \boldsymbol{\phi}_{t-r+1}^{\varepsilon}(\mathbf{x})$.

Then, $(Y_t)_{t\in\mathbb{Z}}$ *is stationary and ergodic.*

5. Asymptotic Properties of the Maximum Likelihood Estimator

We now turn to the maximum likelihood (ML) estimation of the Markov-switching observation-driven model discussed in the previous section.

The d-dimensional parameter vector, where $d := J^2 + \sum_{j=1}^{J} d_j$, is

$$\theta := (p_{ij}, i, j = 1, ..., J, \upsilon_j, j = 1, ..., J)'$$

and belongs to

$$\Theta \subset \left\{(p_{ij}, i, j = 1, ..., J, \upsilon_j, j = 1, ..., J) : p_{ij} > 0, \sum_{j=1}^{J} p_{ij} = 1, \upsilon_j \in \Upsilon_j\right\}.$$

[4]All examples in Section 3 can be written like this because if $\varepsilon_{i,t} \stackrel{d}{=} \varepsilon_{j,t}$ for all $i, j \in \{1, ..., J\}$, then let $\varepsilon_{i,t} = \varepsilon_t$ for all $i \in \{1, ..., J\}$ where $\varepsilon_t \stackrel{i.i.d.}{\sim} \mathcal{D}^{\varepsilon}(\upsilon^{\varepsilon})$, and $(S_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ are independent.

The true parameter vector is denoted θ_0. Assume that a realization, $(y_t)_{t=1}^T$, of the Markov-switching observation-driven model, $(Y_t)_{t\in\mathbb{Z}}$, satisfying Equation (2) and (3) for $\theta = \theta_0$ is observed. The average log-likelihood function is given by

$$\hat{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log \hat{f}_{t|t-1}(y_t;\theta),$$

where

$$\hat{f}_{t|t-1}(y_t;\theta) = \sum_{j=1}^{J} \hat{\pi}_{j,t|t-1}(\theta) f_{j,t|t-1}(y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j).$$

In the average log-likelihood function, $\hat{X}_{j,t}(\upsilon_j)$ is given by

$$\hat{X}_{j,t+1}(\upsilon_j) = \phi_j(y_t,\hat{X}_{j,t}(\upsilon_j);\upsilon_j)$$

for some initialization $\hat{X}_{j,1}(\upsilon_j) \in \mathcal{X}_{\upsilon_j}$, and $\hat{\boldsymbol{\pi}}_{t|t-1}(\theta)$ is given by

$$\hat{\boldsymbol{\pi}}_{t+1|t}(\theta) = \mathbf{P}'\hat{\boldsymbol{\pi}}_{t|t}(\theta)$$

with

$$\hat{\boldsymbol{\pi}}_{t|t}(\theta) = \hat{\mathbf{F}}_t(\hat{\boldsymbol{\pi}}_{t|t-1}(\theta);\theta)\hat{\boldsymbol{\pi}}_{t|t-1}(\theta)$$

for some initialization $\hat{\boldsymbol{\pi}}_{0|0}(\theta) \in \mathcal{S}$ with $\mathcal{S} := \{x \in \mathbb{R}^J : x_j \geq 0, \sum_{j=1}^J x_j = 1\}$ where

$$[\hat{\mathbf{F}}_t(\hat{\boldsymbol{\pi}}_{t|t-1}(\theta);\theta)]_{ii} = \frac{f_{i,t|t-1}(y_t;\hat{X}_{i,t}(\upsilon_i),\upsilon_i)}{\sum_{k=1}^J \hat{\pi}_{k,t|t-1}(\theta) f_{k,t|t-1}(y_t;\hat{X}_{k,t}(\upsilon_k),\upsilon_k)}.$$

The ML estimator is then given by

$$\hat{\theta}_T = \arg\max_{\theta\in\Theta} \hat{L}_T(\theta).$$

5.1. Consistency

The following assumptions are made.

Assumption 1. *The conditions in Theorem 4.1 hold for* $\theta = \theta_0$.

Assumption 2. Θ *is compact.*

Assumption 3. *For each* $j \in \{1, ..., J\}$,

(i) $(y, x, \upsilon) \mapsto f_{j,t|t-1}(y; x, \upsilon)$ *is continuous on* $\mathcal{Y} \times \mathcal{X}_{\Upsilon_j} \times \Upsilon_j$,
(ii) $x \mapsto f_{j,t|t-1}(y; x, \upsilon)$ *is differentiable on* $\mathcal{X}_{\Upsilon_j}$ *for all* $y \in \mathcal{Y}$ *and* $\upsilon \in \Upsilon_j$, *and*

(iii) $y \mapsto \frac{\partial f_{j,t|t-1}(y;x,\upsilon)}{\partial x}$ *is continuous on* $\mathcal{Y}$ *for all* $x \in \mathcal{X}_{\Upsilon_j}$ *and* $\upsilon \in \Upsilon_j$.

First, the asymptotic behaviour of the initialized sequences must be studied in order to ensure an appropriate form of convergence of the average log-likelihood function because the initialized sequences are non-stationary by construction.

The following lemma, which follows from an application of Theorem 3.1 in Bougerol (1993) (see also Theorem 2.8 in Straumann and Mikosch (2006)), gives conditions under which the non-stationary sequence $(\hat{X}_{j,t}(\upsilon_j))_{t\in\mathbb{N}}$ converges uniformly exponentially fast almost surely (e.a.s.) to a unique stationary and ergodic sequence $(X_{j,t}(\upsilon_j))_{t\in\mathbb{Z}}$ as $t \to \infty$.[5] Let

$$\bar{\Lambda}_{j,t}(\upsilon_j) := \sup_{x\in\mathcal{X}_{\Upsilon_j}} \left| \frac{\partial \phi_j(Y_t, x; \upsilon_j)}{\partial x} \right|,$$

where $\mathcal{X}_{\Upsilon_j} := \cup_{\upsilon_j \in \Upsilon_j} \mathcal{X}_{\upsilon_j}$.

Lemma 5.1. *Assume that Assumption 1-2 hold. Moreover, assume that*

(i) *there exists an* $x \in \mathcal{X}_{\Upsilon_j}$ *such that* $\mathbb{E}[\log^+ \sup_{\upsilon_j\in\Upsilon_j} |\phi_j(Y_t, x; \upsilon_j) - x|] < \infty$,

(ii) $\mathbb{E}[\log^+ \sup_{\upsilon_j\in\Upsilon_j} \bar{\Lambda}_{j,t}(\upsilon_j)] < \infty$, *and*

(iii) $\mathbb{E}[\log \sup_{\upsilon_j\in\Upsilon_j} \bar{\Lambda}_{j,t}(\upsilon_j)] < 0$

for each $j \in \{1, ..., J\}$. *Then, for each* $j \in \{1, ..., J\}$, *the sequence* $(X_{j,t}(\upsilon_j))_{t\in\mathbb{Z}}$ *given by*

$$X_{j,t+1}(\upsilon_j) = \phi_j(Y_t, X_{j,t}(\upsilon_j); \upsilon_j)$$

is stationary and ergodic for all $\upsilon_j \in \Upsilon_j$ *and*

$$\sup_{\upsilon_j\in\Upsilon_j} \left| \hat{X}_{j,t}(\upsilon_j) - X_{j,t}(\upsilon_j) \right| \overset{e.a.s.}{\to} 0 \quad \textit{as} \quad t \to \infty$$

for any initialization $\hat{X}_{j,1}(\upsilon_j) \in \mathcal{X}_{\upsilon_j}$.

Theorem 3.1 in Bougerol (1993) cannot be used for $(\hat{\boldsymbol{\pi}}_{t|t-1}(\theta))_{t\in\mathbb{N}}$ because $(\hat{\boldsymbol{\pi}}_{t|t-1}(\theta))_{t\in\mathbb{N}}$ depends on $(\hat{X}_{1,t}(\upsilon_1))_{t\in\mathbb{N}}, ..., (\hat{X}_{J,t}(\upsilon_J))_{t\in\mathbb{N}}$ which are non-stationary. The next lemma, which gives conditions under which the non-stationary sequence $(\hat{\boldsymbol{\pi}}_{t|t}(\theta))_{t\in\mathbb{N}}$ converges uniformly e.a.s. to a unique stationary and ergodic sequence $(\boldsymbol{\pi}_{t|t}(\theta))_{t\in\mathbb{Z}}$ as $t \to \infty$, follows instead from an application of Theorem 2.10 in Straumann and Mikosch (2006).

[5] A sequence of random elements $\{\hat{Z}_t\}_{t\in\mathbb{N}}$ taking values in $(\mathcal{Z}, ||\cdot||)$ is said to converge e.a.s. to another sequence of random elements $\{Z_t\}_{t\in\mathbb{Z}}$ taking values in $(\mathcal{Z}, ||\cdot||)$ if there exists a $\gamma > 1$ such that

$$\gamma^t ||\hat{Z}_t - Z_t|| \overset{a.s.}{\to} 0 \quad \text{as} \quad t \to \infty.$$

Lemma 5.2. *Assume that Assumption 1-3 and the conditions in Lemma 5.1 hold. Moreover, assume that for each $j \in \{1, ..., J\}$, there exists an $m_j > 0$ such that*

$$\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} \sup_{x \in \mathcal{X}_{\Upsilon_j}} \left| \frac{\partial \log f_{j,t|t-1}(Y_t; x, \upsilon)}{\partial x} \right|^{m_j}\right] < \infty.$$

Then, the sequence $(\boldsymbol{\pi}_{t|t}(\theta))_{t\in\mathbb{Z}}$ given by

$$\boldsymbol{\pi}_{t|t}(\theta) = \mathbf{F}_t(\mathbf{P}'\boldsymbol{\pi}_{t-1|t-1}(\theta); \theta)\mathbf{P}'\boldsymbol{\pi}_{t-1|t-1}(\theta)$$

is stationary and ergodic for all $\theta \in \Theta$ and

$$\sup_{\theta\in\Theta} \left|\left|\hat{\boldsymbol{\pi}}_{t|t}(\theta) - \boldsymbol{\pi}_{t|t}(\theta)\right|\right|_2 \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty$$

for any initialization $\hat{\boldsymbol{\pi}}_{0|0}(\theta) \in \mathcal{S}$.

The following corollary is a consequence hereof.

Corollary 5.1. *Under the assumptions in Lemma 5.2, the sequence $(\boldsymbol{\pi}_{t|t-1}(\theta))_{t\in\mathbb{Z}}$ given by*

$$\boldsymbol{\pi}_{t+1|t}(\theta) = \mathbf{P}'\boldsymbol{\pi}_{t|t}(\theta)$$

is stationary and ergodic for all $\theta \in \Theta$ and

$$\sup_{\theta\in\Theta} \left|\left|\hat{\boldsymbol{\pi}}_{t|t-1}(\theta) - \boldsymbol{\pi}_{t|t-1}(\theta)\right|\right|_2 \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty$$

for any initialization $\hat{\boldsymbol{\pi}}_{0|0}(\theta) \in \mathcal{S}$.

Remark 5.1. *Note that $\boldsymbol{\pi}_{t|t-1}(\theta) \in \mathcal{S}_\theta$ for all $t \in \mathbb{Z}$ where $\mathcal{S}_\theta := \{x \in \mathbb{R}^J : x_j \geq \min_{i\in\{1,...,J\}} p_{ij}, \sum_{j=1}^J x_j = 1\}$.*

The result in Lemma 5.2 is not surprising: the Markov chain itself forgets its initialization (in case it is initialized) as $p_{ij} > 0$ for all $i, j \in \{1, ..., J\}$, so it is not surprising that the filter also forgets its initialization provided that the time-varying parameters do the same.

Here, Theorem 2.10 in Straumann and Mikosch (2006) is used to give conditions under which the non-stationary sequence $(\hat{Z}_t(\theta))_{t\in\mathbb{N}}$, which depends on another non-stationary sequence $(\hat{\tilde{Z}}_t(\theta))_{t\in\mathbb{N}}$ that converges to a stationary and ergodic sequence $(\tilde{Z}_t(\theta))_{t\in\mathbb{Z}}$, converges to the stationary and ergodic sequence $(Z_t(\theta))_{t\in\mathbb{Z}}$. The theorem is often used differently to give conditions under which the non-stationary sequence $(\nabla^k_\theta \hat{Z}_t(\theta))_{t\in\mathbb{N}}$ converges to the stationary and ergodic sequence $(\nabla^k_\theta Z_t(\theta))_{t\in\mathbb{Z}}$ where $\nabla^k_\theta Z_t(\theta)$ denotes the k'th order derivative of $Z_t(\theta)$ with respect to θ; see, for instance, Straumann and Mikosch (2006) and Blasques et al. (2022). We believe that this application of the theorem can be useful in other situations as well.

Two important implications of Lemmata 5.1 and 5.2 are that the true time-varying parameters, $X_{1,t}(\upsilon_{1,0}), ..., X_{J,t}(\upsilon_{J,0})$, and the true filter, $\boldsymbol{\pi}_{t|t}(\theta_0)$, can be recovered asymptotically if θ_0 is known. In fact, they can be recovered asymptotically even if there is a sequence $(\hat{\theta}_n)_{n\in\mathbb{N}}$ such that $\hat{\theta}_n \stackrel{a.s.}{\to} \theta_0$ as $n \to \infty$ as the next results show.

Proposition 5.1. *work in progress*

Proposition 5.2. *work in progress*

The next corollary is a consequence hereof.

Corollary 5.2. *work in progress*

Lemmata 5.1 and 5.2 are thus not just technical conditions used to ensure an appropriate form of convergence of the average log-likelihood function.

Moreover, the following assumption is made.

Assumption 4. *For each $j \in \{1, ..., J\}$,*

$$\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j} |\log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)|\right] < \infty.$$

Finally, the following identification condition is imposed as in Francq and Roussignol (1998). We denote by $f_{t|t-m}(\mathbf{y}; \theta), \mathbf{y} \in \mathcal{Y}^m$ the conditional pdf of $(Y_t, ..., Y_{t-m+1})$ given $\mathbf{Y}_{-\infty}^{t-m}$.

Assumption 5. *There exists an $m \in \mathbb{N}$ such that*

$$f_{t|t-m}(Y_t, ..., Y_{t-m+1}; \theta) = f_{t|t-m}(Y_t, ..., Y_{t-m+1}; \theta_0) \quad a.s.$$

implies that

$$\theta = \theta_0.$$

Theorem 5.1 gives conditions under which the ML estimator is consistent.

Theorem 5.1. *Assume that Assumption 1-5 and the conditions in Lemmata 5.1 and 5.2 hold. Then,*

$$\hat{\theta}_T \stackrel{a.s.}{\to} \theta_0 \quad as \quad T \to \infty.$$

5.2. Asymptotic Normality

In addition to Assumption 1-5, the following assumptions are made.

Assumption 6. $\theta_0 \in int(\Theta)$

Assumption 7. *For each* $j \in \{1, ..., J\}$,

(i) $(x, \upsilon) \mapsto f_{j,t|t-1}(y; x, \upsilon)$ *is twice differentiable on* $\mathcal{X}_{\Upsilon_j} \times \Upsilon_j$ *for all* $y \in \mathcal{Y}$,

(ii) $(y, x, \upsilon) \mapsto \frac{\partial f_{j,t|t-1}(y;x,\upsilon)}{\partial(x,\upsilon)}$ *is continuous on* $\mathcal{Y} \times \mathcal{X}_{\Upsilon_j} \times \Upsilon_j$, *and*

(iii) $(y, x, \upsilon) \mapsto \frac{\partial^2 f_{j,t|t-1}(y;x,\upsilon)}{\partial(x,\upsilon)\partial(x,\upsilon)'}$ *is continuous on* $\mathcal{Y} \times \mathcal{X}_{\Upsilon_j} \times \Upsilon_j$.

Assumption 8. *For each* $j \in \{1, ..., J\}$, *the sequence* $\left(\frac{\partial X_{j,t}(\upsilon_j)}{\partial \theta}\right)_{t\in\mathbb{Z}}$ *is stationary and ergodic for all* $\upsilon_j \in \Upsilon_j$ *and*

$$\sup_{\upsilon_j \in \Upsilon_j} \left\| \frac{\partial \hat{X}_{j,t}(\upsilon_j)}{\partial \theta} - \frac{\partial X_{j,t}(\upsilon_j)}{\partial \theta} \right\|_2 \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty$$

for any initialization $\frac{\partial \hat{X}_{j,1}(\upsilon_j)}{\partial \theta} \in \mathbb{R}^d$.

Assumption 9. *For each* $j \in \{1, ..., J\}$, *the sequence* $\left(\frac{\partial^2 X_{j,t}(\upsilon_j)}{\partial \theta \partial \theta'}\right)_{t\in\mathbb{Z}}$ *is stationary and ergodic for all* $\upsilon_j \in \Upsilon_j$ *and*

$$\sup_{\upsilon_j \in \Upsilon_j} \left\| \frac{\partial^2 \hat{X}_{j,t}(\upsilon_j)}{\partial \theta \partial \theta'} - \frac{\partial^2 X_{j,t}(\upsilon_j)}{\partial \theta \partial \theta'} \right\|_{2,2} \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty$$

for any initialization $\frac{\partial^2 \hat{X}_{j,1}(\upsilon_j)}{\partial \theta \partial \theta'} \in \mathbb{R}^{d\times d}$.

Assumption 10. *For each* $j \in \{1, ..., J\}$, *there exists a* $k_j > 1$ *such that*

$$\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} \left\| \frac{\partial X_{j,t}(\upsilon)}{\partial \theta} \right\|_2^{2k_j}\right] < \infty \quad and \quad \mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} \left\| \frac{\partial^2 X_{j,t}(\upsilon)}{\partial \theta \partial \theta'} \right\|_{2,2}^{k_j}\right] < \infty.$$

Assumption 11. *For each* $j \in \{1, ..., J\}$, *the sequence* $\left(\frac{\partial \pi_{j,t|t-1}(\theta)}{\partial \theta}\right)_{t\in\mathbb{Z}}$ *is stationary and ergodic for all* $\theta \in \Theta$ *and*

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \hat{\pi}_{j,t|t-1}(\theta)}{\partial \theta} - \frac{\partial \pi_{j,t|t-1}(\theta)}{\partial \theta} \right\|_2 \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty$$

for any initialization $\frac{\partial \hat{\pi}_{j,1|0}(\theta)}{\partial \theta} \in \mathbb{R}^d$.

Assumption 12. *For each $j \in \{1, ..., J\}$, there exists an $m_j > 0$ such that*

$$\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} \sup_{x \in \mathcal{X}_{\Upsilon_j}} \left| \frac{\partial \log f_{j,t|t-1}(Y_t; x, \upsilon)}{\partial \upsilon_h} \right|^{m_j}\right] < \infty$$

for all $h \in \{1, ..., d_j\}$,

$$\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} \sup_{x \in \mathcal{X}_{\Upsilon_j}} \left| \frac{\partial^2 \log f_{j,t|t-1}(Y_t; x, \upsilon)}{\partial x^2} \right|^{m_j}\right] < \infty,$$

and

$$\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} \sup_{x \in \mathcal{X}_{\Upsilon_j}} \left| \frac{\partial^2 \log f_{j,t|t-1}(Y_t; x, \upsilon)}{\partial \upsilon_h \partial x} \right|^{m_j}\right] < \infty$$

for all $h \in \{1, ..., d_j\}$.

Assumption 13. *For each $j \in \{1, ..., J\}$, the sequence $\left(\frac{\partial^2 \pi_{j,t|t-1}(\theta)}{\partial \theta \partial \theta'}\right)_{t \in \mathbb{Z}}$ is stationary and ergodic for all $\theta \in \Theta$ and*

$$\sup_{\theta \in \Theta} \left|\left| \frac{\partial^2 \hat{\pi}_{j,t|t-1}(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \pi_{j,t|t-1}(\theta)}{\partial \theta \partial \theta'} \right|\right|_{2,2} \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty$$

for any initialization $\frac{\partial^2 \hat{\pi}_{j,1|0}(\theta)}{\partial \theta \partial \theta'} \in \mathbb{R}^{d \times d}$.

Assumption 14. *For each $j \in \{1, ..., J\}$,*

$$\mathbb{E}\left[\sup_{\theta \in \Theta} \left|\left| \frac{\partial \pi_{j,t|t-1}(\theta)}{\partial \theta} \right|\right|_2^2\right] < \infty \quad and \quad \mathbb{E}\left[\sup_{\theta \in \Theta} \left|\left| \frac{\partial^2 \pi_{j,t|t-1}(\theta)}{\partial \theta \partial \theta'} \right|\right|_{2,2}\right] < \infty.$$

Assumption 15. *For each $j \in \{1, ..., J\}$,*

$$\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} \left|\bar{\nabla}_x \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)\right|^{\frac{2k_j}{k_j - 1}}\right] < \infty \quad and \quad \mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} |\bar{\nabla}_{xx} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)|^{\frac{k_j}{k_j - 1}}\right] < \infty,$$

where k_j is given in Assumption 10. Moreover, for each $j \in \{1, ..., J\}$,

$$\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} ||\bar{\nabla}_{\upsilon x} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)||_2^{\frac{k_j}{k_j - 1}}\right] < \infty.$$

Finally, for each $j \in \{1, ..., J\}$,

$$\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} ||\bar{\nabla}_{\upsilon} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)||_2^2\right] < \infty \quad and \quad \mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} ||\bar{\nabla}_{\upsilon\upsilon} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)||_{2,2}\right] < \infty.$$

Theorem 5.2 gives conditions under which the ML estimator is asymptotic normal.

Theorem 5.2. *Assume that Assumption 1-15 and the conditions in Lemmata 5.1 and 5.2 hold. Then,*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0)) \quad \text{as} \quad T \to \infty,$$

where

$$I(\theta) := -\mathbb{E}\left[\frac{\partial^2 \log f_{t|t-1}(Y_t;\theta)}{\partial\theta\partial\theta'}\right].$$

6. Examples

work in progress

7. Empirical Illustration

It is well-known that the GARCH model often provides a better fit than the ARCH model when applied to real data. To study whether or not the Markov-switching GARCH model also provides a better fit than the Markov-switching ARCH model when applied to real data, we perform a small empirical application in which the Markov-switching GARCH model is compared to the Markov-switching ARCH model with the GARCH model as a benchmark.

More specifically, we use daily log-returns on the Standard & Poor's (S&P) 500 index, the Financial Times Stock Exchange (FTSE) 100 index, and the Cotation Assistée en Continu (CAC) 40 index from January 2, 2004 to December 29, 2023 retrieved from Yahoo Finance. Figure 1 reports the three series.

We estimate a two-state Markov-switching GARCH model given by

$$Y_t = \sqrt{X_{S_t,t}}\varepsilon_t \tag{4}$$

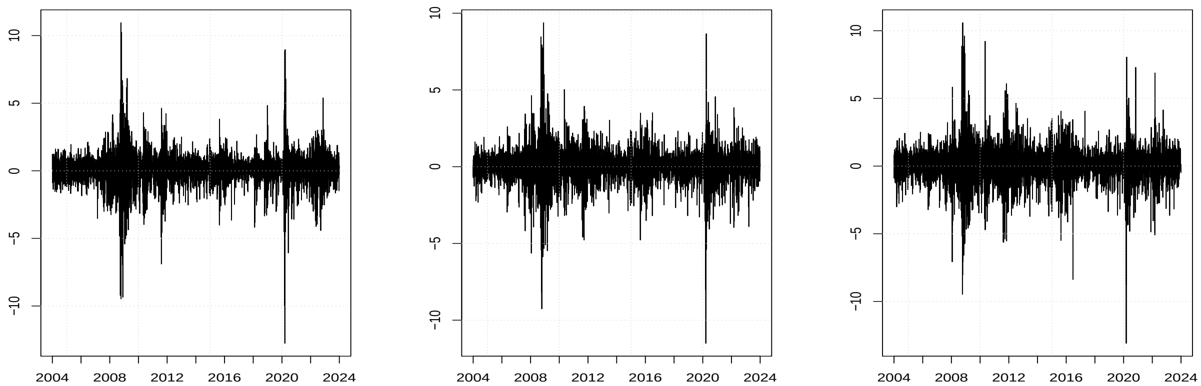
with $\varepsilon_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ where

$$X_{j,t+1} = \omega_j + \alpha_j Y_t^2 + \beta_j X_{j,t}, \tag{5}$$

a two-state Markov-switching ARCH model given by Equation (4) and (5) with $\beta_j \equiv 0$, and a GARCH model as a benchmark.[6] Table 1 and 2 report the estimated parameters and the information criteria, respectively.

For all series, the Markov-switching ARCH model and the Markov-switching GARCH model have a low-volatility state (state 1) and a high-volatility state (state 2) where, in the high-volatility state, the variance reacts more to the past squared return and, for the Markov-switching GARCH model, the variance is less persistent.

[6]Note that the Markov-switching GARCH model reduces to the GARCH model when $J = 1$.



(a) S&P 500 (b) FTSE 100 (c) CAC 100

Figure 1: Panel (a), (b), and (c) report the daily log-returns on the S&P 500 index, the FTSE 100 index, and the CAC 40 index from January 2, 2004 to December 29, 2023, respectively.

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are lower for the Markov-switching GARCH model than for the Markov-switching ARCH model as well as the GARCH model for all series showing that the Markov-switching GARCH model not only provides a better fit than the Markov-switching ARCH model but also the GARCH model when applied to this data. Note also that, interestingly, the AIC and BIC are lower for the GARCH model than for the Markov-switching ARCH model for all series.

8. Conclusion

In this paper, we study the asymptotic properties of the ML estimator for a so-called Markov-switching observation-driven model. The Markov-switching observation-driven model contains several models proposed in the litterature as special cases for which, to the best of our knowledge, no results for the asymptotic properties of the ML estimator exist. A possible extension could be to study the asymptotic properties of the ML estimator for a higher-order Markov-switching observation-driven model instead of a first-order Markov-switching observation-driven model as in this paper. We leave this for future research.

	S&P 500			FTSE 100			CAC 40		
	GARCH	MS-ARCH	MS-GARCH	GARCH	MS-ARCH	MS-GARCH	GARCH	MS-ARCH	MS-GARCH
$\hat{\omega}_1$	0.025	0.430	0.004	0.027	0.440	0.003	0.038	0.646	0.008
$\hat{\alpha}_1$	0.121	0.046	0.088	0.115	0.074	0.010	0.112	0.018	0.073
$\hat{\beta}_1$	0.856	-	0.886	0.859	-	0.978	0.865	-	0.900
$\hat{\omega}_2$	-	3.046	0.248	-	2.753	0.093	-	3.738	0.551
$\hat{\alpha}_2$	-	0.195	0.245	-	0.205	0.101	-	0.114	0.295
$\hat{\beta}_2$	-	-	0.750	-	-	0.895	-	-	0.701
$\hat{p}_{11}$	-	0.991	0.846	-	0.988	0.970	-	0.984	0.853
$\hat{p}_{22}$	-	0.978	0.198	-	0.967	0.901	-	0.964	0.105

Table 1: The estimated parameters for the GARCH model, the Markov-switching ARCH model, and the Markov-switching GARCH model for the S&P 500 index, the FTSE 100 index, and the CAC 40 index.

	S&P 500			FTSE 100			CAC 40		
	GARCH	MS-ARCH	MS-GARCH	GARCH	MS-ARCH	MS-GARCH	GARCH	MS-ARCH	MS-GARCH
AIC	13374.15	13688.84	**13183.86**	13339.24	13491.73	**13149.38**	15608.59	15755.88	**15389.25**
BIC	13393.72	13727.98	**13236.04**	13358.82	13530.89	**13201.60**	15628.21	15795.12	**15441.57**

Table 2: The AIC and BIC for the GARCH model, the Markov-switching ARCH model, and the Markov-switching GARCH model for the S&P 500 index, the FTSE 100 index, and the CAC 40 index. The bold values are the lowest ones.

9. Proofs

We adopt the following notation. Let f be a $k \times 1$ vector of real valued functions defined on $\mathbb{R}^n \times \mathbb{R}^m$, let $\mathbf{x}$ be an $n \times 1$ vector of real numbers, and let $\mathbf{y}$ be an $m \times 1$ vector of real numbers. Then, $\nabla_{\mathbf{x}'} f := \frac{\partial f}{\partial \mathbf{x}'}$ is the $k \times n$ matrix of first order derivatives of f and $\nabla_{\mathbf{x}} f := (\nabla_{\mathbf{x}'} f)'$. If $k = 1$, then $\nabla_{\mathbf{xy}} f = \nabla_{\mathbf{y}'} \nabla_{\mathbf{x}} f$ is the $n \times m$ matrix of second order derivatives of f.

9.1. Lemma 5.1

For each $j \in \{1, ..., J\}$, $(X_{j,t}(\upsilon_j))_{t\in\mathbb{Z}}$ is a stochastic process taking values in $\mathcal{X}_{\upsilon_j}$ equipped with $|\cdot|$ given by

$$X_{j,t+1}(\upsilon_j) = \phi_{j,t}(X_{j,t}(\upsilon_j); \upsilon_j)$$

for all $\upsilon_j \in \Upsilon_j$ where $(\phi_{j,t}(\cdot; \upsilon_j))_{t\in\mathbb{Z}}$ is a sequence of stationary and ergodic Lipschitz functions given by

$$\phi_{j,t}(x; \upsilon_j) = \phi_j(Y_t, x; \upsilon_j)$$

for all $\upsilon_j \in \Upsilon_j$ with Lipschitz coefficient

$$\bar{\Lambda}(\phi_{j,t}; \upsilon_j) = \sup_{\substack{x,y\in\mathcal{X}_{\Upsilon_j} \\ x\neq y}} \frac{|\phi_{j,t}(x; \upsilon_j) - \phi_{j,t}(y; \upsilon_j)|}{|x - y|} \leq \sup_{x\in\mathcal{X}_{\Upsilon_j}} \left| \frac{\partial \phi_{j,t}(x; \upsilon_j)}{\partial x} \right| = \bar{\Lambda}_{j,t}(\upsilon_j)$$

for all $\upsilon_j \in \Upsilon_j$.

The conclusion follows from Lemma Appendix A.1 since

(i) there exists an $x \in \mathcal{X}_{\Upsilon_j}$ such that $\mathbb{E}[\log^+ \sup_{\upsilon_j \in \Upsilon_j} |\phi_{j,t}(x;\upsilon_j) - x|] < \infty$,

(ii) $\mathbb{E}[\log^+ \sup_{\upsilon_j \in \Upsilon_j} \bar{\Lambda}_{j,t}(\upsilon_j)] < \infty$, and

(iii) $\mathbb{E}[\log \sup_{\upsilon_j \in \Upsilon_j} \bar{\Lambda}_{j,t}(\upsilon_j)] < 0$

for each $j \in \{1, ..., J\}$ by assumption.

9.2. Lemma 5.2

The proof consists of two parts. The first part considers the stochastic process $(\boldsymbol{\pi}_{t|t}(\theta))_{t\in\mathbb{Z}}$, and the second part considers the stochastic process $(\hat{\boldsymbol{\pi}}_{t|t}(\theta))_{t\in\mathbb{N}}$. Let $C \in \mathbb{R}$ be an arbitrary constant that can change throughout the proof.

First, $(\boldsymbol{\pi}_{t|t}(\theta))_{t\in\mathbb{Z}}$ is a stochastic process taking values in $\mathcal{S}$ equipped with $||\cdot||_2$ given by

$$\boldsymbol{\pi}_{t|t}(\theta) = \boldsymbol{\phi}_t(\boldsymbol{\pi}_{t-1|t-1}(\theta);\theta)$$

for all $\theta \in \Theta$ where $(\boldsymbol{\phi}_t(\cdot;\theta))_{t\in\mathbb{Z}}$ is a sequence of stationary and ergodic Lipschitz functions given by

$$\boldsymbol{\phi}_t(\mathbf{s};\theta) = \mathbf{F}_t(\mathbf{P}'\mathbf{s};\theta)\mathbf{P}'\mathbf{s}$$

for all $\theta \in \Theta$ with Lipschitz coefficient

$$\bar{\Lambda}(\boldsymbol{\phi}_t;\theta) = \sup_{\substack{\mathbf{x},\mathbf{y}\in\mathcal{S}\\ \mathbf{x}\neq\mathbf{y}}} \frac{||\boldsymbol{\phi}_t(\mathbf{x};\theta) - \boldsymbol{\phi}_t(\mathbf{y};\theta)||_2}{||\mathbf{x}-\mathbf{y}||_2}$$

for all $\theta \in \Theta$.

The first conclusion follows from Lemma Appendix A.2 if

(i) there exists an $\mathbf{s} \in \mathcal{S}$ such that $\mathbb{E}[\log^+ \sup_{\theta\in\Theta} ||\boldsymbol{\phi}_t(\mathbf{s};\theta) - \mathbf{s}||_2] < \infty$,

(ii) $\mathbb{E}[\log^+ \sup_{\theta\in\Theta} \bar{\Lambda}(\boldsymbol{\phi}_t;\theta)] < \infty$, and

(iii) there exists an $r \in \mathbb{N}$ such that $\mathbb{E}[\log \sup_{\theta\in\Theta} \bar{\Lambda}(\boldsymbol{\phi}_t^{(r)};\theta)] < 0$.

Condition (i) is trivial. Note that

$$\mathbb{P}_\theta(S_t = j \mid \mathbf{Y}_{-\infty}^u) = \sum_{i=1}^J \mathbb{P}_\theta(S_t = j \mid S_{t-1} = i, \mathbf{Y}_{-\infty}^u)\mathbb{P}_\theta(S_{t-1} = i \mid \mathbf{Y}_{-\infty}^u)$$

for all $t \leq u$ where

$$\mathbb{P}_\theta(S_t = j \mid S_{t-1} = i, \mathbf{Y}_{-\infty}^u) = \frac{\mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u, S_t = j \mid S_{t-1} = i, \mathbf{Y}_{-\infty}^{t-1})}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u, S_t = k \mid S_{t-1} = i, \mathbf{Y}_{-\infty}^{t-1})}$$

$$= \frac{\mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid S_t = j, S_{t-1} = i, \mathbf{Y}_{-\infty}^{t-1})\mathbb{P}_\theta(S_t = j \mid S_{t-1} = i, \mathbf{Y}_{-\infty}^{t-1})}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid S_t = k, S_{t-1} = i, \mathbf{Y}_{-\infty}^{t-1})\mathbb{P}_\theta(S_t = k \mid S_{t-1} = i, \mathbf{Y}_{-\infty}^{t-1})}$$

$$= \frac{\mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid \mathbf{Y}_{-\infty}^{t-1}, S_t = j)\mathbb{P}_\theta(S_t = j \mid S_{t-1} = i)}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid \mathbf{Y}_{-\infty}^{t-1}, S_t = k)\mathbb{P}_\theta(S_t = k \mid S_{t-1} = i)}.$$

Hence, $\pi_{j,t|u}(\theta) := \mathbb{P}_\theta(S_t = j \mid \mathbf{Y}_{-\infty}^u)$ is given by

$$\pi_{j,t|u}(\theta) = \sum_{i=1}^J \frac{\mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid \mathbf{Y}_{-\infty}^{t-1}, S_t = j)p_{ij}}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid \mathbf{Y}_{-\infty}^{t-1}, S_t = k)p_{ik}} \pi_{i,t-1|u}(\theta)$$

for all $t \leq u$ where

$$\mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid \mathbf{Y}_{-\infty}^{t-1}, S_t = j) = \sum_{s_{t+1}=1}^J \cdots \sum_{s_u=1}^J p_{js_{t+1}} \cdots p_{s_{u-1}s_u} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) \cdots f_{s_u,u|u-1}(Y_u; X_{s_u,u}(\upsilon_{s_u}), \upsilon_{s_u}),$$

and $\boldsymbol{\pi}_{t|u}(\theta) := (\pi_{1,t|u}(\theta), ..., \pi_{J,t|u}(\theta))'$ is given by

$$\boldsymbol{\pi}_{t|u}(\theta) = \mathbf{M}'_{t|u}(\theta)\boldsymbol{\pi}_{t-1|u}(\theta)$$

for all $t \leq u$ where $\mathbf{M}_{t|u}(\theta)$ is a stochastic matrix with generic element

$$[\mathbf{M}_{t|u}(\theta)]_{ij} = \frac{\mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid \mathbf{Y}_{-\infty}^{t-1}, S_t = j)p_{ij}}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_t^u \in d\mathbf{y}_t^u \mid \mathbf{Y}_{-\infty}^{t-1}, S_t = k)p_{ik}}.$$

Thus, we have that

$$\boldsymbol{\pi}_{t|t}(\theta) = \mathbf{M}'_{t|t}(\theta) \cdots \mathbf{M}'_{t-r+1|t}(\theta)\boldsymbol{\pi}_{t-r|t}(\theta) \tag{6}$$

for all $r \in \mathbb{N}$ where

$$[\boldsymbol{\pi}_{t-r|t}(\theta)]_j = \frac{\mathbb{P}_\theta(\mathbf{Y}_{t-r+1}^t \in d\mathbf{y}_{t-r+1}^t \mid \mathbf{Y}_{-\infty}^{t-r}, S_{t-r} = j)[\boldsymbol{\pi}_{t-r|t-r}(\theta)]_j}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_{t-r+1}^t \in d\mathbf{y}_{t-r+1}^t \mid \mathbf{Y}_{-\infty}^{t-r}, S_{t-r} = k)[\boldsymbol{\pi}_{t-r|t-r}(\theta)]_k}.$$

This observation leads to Lemma 9.2 proved using the following lemma.

Lemma 9.1. *If there exists an* $\varepsilon \in (0, 1)$ *such that*

$$p_{ij} \geq \varepsilon$$

for all $i, j \in \{1, ..., J\}$*, then, for all* $\tau \leq t$*, there exist a* $\mathbf{v}_{\tau|t}(\theta) \in \mathcal{S}$ *such that*

$$[\mathbf{M}_{\tau|t}(\theta)]_{ij} \geq \varepsilon[\mathbf{v}_{\tau|t}(\theta)]_j$$

for all $i, j \in \{1, ..., J\}$*.*

Proof 1. *Let* $\mathbf{v}_{\tau|t}(\theta)$ *be a stochastic vector with generic element*

$$[\mathbf{v}_{\tau|t}(\theta)]_j = \frac{\mathbb{P}_\theta(\mathbf{Y}_\tau^t \in d\mathbf{y}_\tau^t \mid \mathbf{Y}_{-\infty}^{\tau-1}, S_\tau = j)}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_\tau^t \in d\mathbf{y}_\tau^t \mid \mathbf{Y}_{-\infty}^{\tau-1}, S_\tau = k)}.$$

Then,

$$[\mathbf{M}_{\tau|t}(\theta)]_{ij} = \frac{\mathbb{P}_\theta(\mathbf{Y}_\tau^t \in d\mathbf{y}_\tau^t \mid \mathbf{Y}_{-\infty}^{\tau-1}, S_\tau = j)p_{ij}}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_\tau^t \in d\mathbf{y}_\tau^t \mid \mathbf{Y}_{-\infty}^{\tau-1}, S_\tau = k)p_{ik}} \geq \varepsilon \frac{\mathbb{P}_\theta(\mathbf{Y}_\tau^t \in d\mathbf{y}_\tau^t \mid \mathbf{Y}_{-\infty}^{\tau-1}, S_\tau = j)}{\sum_{k=1}^J \mathbb{P}_\theta(\mathbf{Y}_\tau^t \in d\mathbf{y}_\tau^t \mid \mathbf{Y}_{-\infty}^{\tau-1}, S_\tau = k)} = \varepsilon[\mathbf{v}_{\tau|t}(\theta)]_j.$$

Lemma 9.2. *Assume that there exists an* $\varepsilon \in (0, 1)$ *such that*

$$p_{ij} \geq \varepsilon$$

for all $i, j \in \{1, ..., J\}$*. Then, there exists an* $\alpha \in (0, 1)$ *such that*

$$\frac{||\boldsymbol{\phi}_t^{(r)}(\mathbf{x}; \theta) - \boldsymbol{\phi}_t^{(r)}(\mathbf{y}; \theta)||_2}{||\mathbf{x} - \mathbf{y}||_2} \leq \alpha^r C$$

for all $r \in \mathbb{N}$*,* $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ *with* $\mathbf{x} \neq \mathbf{y}$*, and* $\theta \in \Theta$*.*

Proof 2. *First,* $|| \cdot ||_2 \leq || \cdot ||_1$ *implies that*

$$||\boldsymbol{\phi}_t^{(r)}(\mathbf{x}; \theta) - \boldsymbol{\phi}_t^{(r)}(\mathbf{y}; \theta)||_2 \leq ||\boldsymbol{\phi}_t^{(r)}(\mathbf{x}; \theta) - \boldsymbol{\phi}_t^{(r)}(\mathbf{y}; \theta)||_1. \tag{7}$$

To ease the notation, let $p_{j,t} := \mathbb{P}_\theta(\mathbf{Y}_{t-r+1}^t \in d\mathbf{y}_{t-r+1}^t \mid \mathbf{Y}_{-\infty}^{t-r}, S_{t-r} = j)$*. Equation (6) shows that*

$$\boldsymbol{\phi}_t^{(r)}(\mathbf{z}; \theta) = \mathbf{M}'_{t|t}(\theta) \cdots \mathbf{M}'_{t-r+1|t}(\theta)\tilde{\mathbf{z}},$$

where

$$\tilde{z}_j = \frac{p_{j,t} z_j}{\sum_{k=1}^J p_{k,t} z_k}.$$

An application of Lemma Appendix B.1 together with Lemma 9.1 thus shows that there exists an $\alpha \in (0, 1)$ *such that*

$$||\boldsymbol{\phi}_t^{(r)}(\mathbf{x}; \theta) - \boldsymbol{\phi}_t^{(r)}(\mathbf{y}; \theta)||_1 \leq \alpha^r ||\tilde{\mathbf{x}} - \tilde{\mathbf{y}}||_1,$$

where

$$||\tilde{\mathbf{x}} - \tilde{\mathbf{y}}||_1 = \sum_{j=1}^J \left| \frac{p_{j,t} x_j}{\sum_{k=1}^J p_{k,t} x_k} - \frac{p_{j,t} y_j}{\sum_{k=1}^J p_{k,t} y_k} \right|.$$

Note that

$$\left| \frac{p_{j,t} x_j}{\sum_{k=1}^J p_{k,t} x_k} - \frac{p_{j,t} y_j}{\sum_{k=1}^J p_{k,t} y_k} \right| = \left| \frac{p_{j,t} x_j}{\sum_{k=1}^J p_{k,t} x_k} - \frac{p_{j,t} y_j}{\sum_{k=1}^J p_{k,t} x_k} + \frac{p_{j,t} y_j}{\sum_{k=1}^J p_{k,t} x_k} - \frac{p_{j,t} y_j}{\sum_{k=1}^J p_{k,t} y_k} \right|$$

$$
\begin{aligned}
&\leq \left| \frac{p_{j,t}x_j}{\sum_{k=1}^J p_{k,t}x_k} - \frac{p_{j,t}y_j}{\sum_{k=1}^J p_{k,t}x_k} \right| + \left| \frac{p_{j,t}y_j}{\sum_{k=1}^J p_{k,t}x_k} - \frac{p_{j,t}y_j}{\sum_{k=1}^J p_{k,t}y_k} \right| \\
&= \frac{p_{j,t}}{\sum_{k=1}^J p_{k,t}x_k} |x_j - y_j| + \frac{p_{j,t}y_j}{\sum_{k=1}^J p_{k,t}y_k} \frac{1}{\sum_{k=1}^J p_{k,t}x_k} \left| \sum_{k=1}^J p_{k,t}(x_k - y_k) \right| \\
&\leq \frac{p_{j,t}}{\sum_{k=1}^J p_{k,t}x_k} |x_j - y_j| + \sum_{k=1}^J \frac{p_{k,t}}{\sum_{k=1}^J p_{k,t}x_k} |x_k - y_k| \\
&\leq \frac{1}{\varepsilon} |x_j - y_j| + \frac{1}{\varepsilon} \sum_{k=1}^J |x_k - y_k|
\end{aligned}
$$

since

$$
p_{i,t} = \sum_{j=1}^J \mathbb{P}_\theta(\mathbf{Y}_{t-r+1}^t \in d\mathbf{y}_{t-r+1}^t \mid \mathbf{Y}_{-\infty}^{t-r}, S_{t-r+1} = j)p_{ij}.
$$

Hence,

$$
||\boldsymbol{\phi}_t^{(r)}(\mathbf{x};\theta) - \boldsymbol{\phi}_t^{(r)}(\mathbf{y};\theta)||_1 \leq \alpha^r \frac{J+1}{\varepsilon} ||\mathbf{x} - \mathbf{y}||_1. \tag{8}
$$

Finally, $|| \cdot ||_1 \leq \sqrt{J} || \ \cdot ||_2$ *implies that*

$$
\alpha^r \frac{J+1}{\varepsilon} ||\mathbf{x} - \mathbf{y}||_1 \leq \alpha^r \frac{J+1}{\varepsilon} \sqrt{J} ||\mathbf{x} - \mathbf{y}||_2. \tag{9}
$$

Together, Equation (7)-(9) *show that*

$$
\frac{||\boldsymbol{\phi}_t^{(r)}(\mathbf{x};\theta) - \boldsymbol{\phi}_t^{(r)}(\mathbf{y};\theta)||_2}{||\mathbf{x} - \mathbf{y}||_2} \leq \alpha^r \frac{J+1}{\varepsilon} \sqrt{J}.
$$

This concludes the proof.

Condition (ii) and (iii) follow straightforwardly from Lemma 9.2.

Moreover, $(\hat{\boldsymbol{\pi}}_{t|t}(\theta))_{t\in\mathbb{N}}$ is a stochastic process taking values in $\mathcal{S}$ equipped with $|| \cdot ||_2$ given by

$$
\hat{\boldsymbol{\pi}}_{t|t}(\theta) = \hat{\boldsymbol{\phi}}_t(\hat{\boldsymbol{\pi}}_{t-1|t-1}(\theta);\theta)
$$

for all $\theta \in \Theta$ where $(\hat{\boldsymbol{\phi}}_t(\cdot;\theta))_{t\in\mathbb{N}}$ is a sequence of non-stationary Lipschitz functions given by

$$
\hat{\boldsymbol{\phi}}_t(\mathbf{s};\theta) = \hat{\mathbf{F}}_t(\mathbf{P}'\mathbf{s};\theta)\mathbf{P}'\mathbf{s}
$$

for all $\theta \in \Theta$.

The second conclusion follows from Lemma Appendix A.2 as well if

(i) $\mathbb{E}[\log^+ \sup_{\theta \in \Theta} ||\boldsymbol{\pi}_{t|t}(\theta)||_2] < \infty$,

(ii) there exists an $\mathbf{s} \in \mathcal{S}$ such that $\sup_{\theta \in \Theta} ||\hat{\boldsymbol{\phi}}_t(\mathbf{s};\theta) - \boldsymbol{\phi}_t(\mathbf{s};\theta)||_2 \overset{e.a.s.}{\to} 0$ as $t \to \infty$, and

(iii) $\sup_{\theta \in \Theta} \bar{\Lambda}(\hat{\boldsymbol{\phi}}_t - \boldsymbol{\phi}_t;\theta) \overset{e.a.s.}{\to} 0$ as $t \to \infty$.

As above, Condition (i) is trivial. We have that

$$|[\hat{\boldsymbol{\phi}}_t(\mathbf{s};\theta) - \boldsymbol{\phi}_t(\mathbf{s};\theta)]_j| = \left| \frac{\sum_{i=1}^J \hat{f}_{j,t|t-1} p_{ij} s_i}{\sum_{k=1}^J \sum_{l=1}^J \hat{f}_{k,t|t-1} p_{lk} s_l} - \frac{\sum_{i=1}^J f_{j,t|t-1} p_{ij} s_i}{\sum_{k=1}^J \sum_{l=1}^J f_{k,t|t-1} p_{lk} s_l} \right|$$

for all $j \in \{1,...,J\}$ where $\hat{f}_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)$ and $f_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)$. Let $g_t^\theta : \mathcal{X}_{\upsilon_1} \times \cdots \times \mathcal{X}_{\upsilon_J} \to \mathbb{R}$ be given by

$$g_t^\theta(\mathbf{x}) = \frac{\sum_{i=1}^J f_{j,t|t-1}(Y_t;x_j,\upsilon_j) p_{ij} s_i}{\sum_{k=1}^J \sum_{l=1}^J f_{k,t|t-1}(Y_t;x_k,\upsilon_k) p_{lk} s_l}.$$

An application of the mean value theorem and the Cauchy–Schwarz inequality shows that there exists an $\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}) \in \{c\hat{\mathbf{X}}_t(\boldsymbol{\upsilon}) + (1-c)\mathbf{X}_t(\boldsymbol{\upsilon}) : c \in [0,1]\}$ such that

$$|g_t^\theta(\hat{\mathbf{X}}_t(\boldsymbol{\upsilon})) - g_t^\theta(\mathbf{X}_t(\boldsymbol{\upsilon}))| \leq ||\nabla_{\mathbf{x}} g_t^\theta(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))||_2 ||\hat{\mathbf{X}}_t(\boldsymbol{\upsilon}) - \mathbf{X}_t(\boldsymbol{\upsilon})||_2,$$

where, if $m \neq j$,

$$|[\nabla_{\mathbf{x}} g_t^\theta(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))]_m| \leq \frac{\sum_{i=1}^J \bar{f}_{j,t|t-1} p_{ij} s_i}{(\sum_{k=1}^J \sum_{l=1}^J \bar{f}_{k,t|t-1} p_{lk} s_l)^2} \sum_{l=1}^J |\nabla_x \bar{f}_{m,t|t-1}| p_{lm} s_l \leq C |\nabla_x \log \bar{f}_{m,t|t-1}|$$

and, if $m = j$,

$$\begin{aligned} |[\nabla_{\mathbf{x}} g_t^\theta(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))]_m| \leq & \frac{\sum_{i=1}^J |\nabla_x \bar{f}_{m,t|t-1}| p_{im} s_i}{\sum_{k=1}^J \sum_{l=1}^J \bar{f}_{k,t|t-1} p_{lk} s_l} \\ & + \frac{\sum_{i=1}^J \bar{f}_{m,t|t-1} p_{im} s_i}{(\sum_{k=1}^J \sum_{l=1}^J \bar{f}_{k,t|t-1} p_{lk} s_l)^2} \sum_{l=1}^J |\nabla_x \bar{f}_{m,t|t-1}| p_{lm} s_l \leq C |\nabla_x \log \bar{f}_{m,t|t-1}| \end{aligned}$$

where $\bar{f}_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t;\bar{X}_{j,t}(\upsilon_j),\upsilon_j)$ so

$$||\nabla_{\mathbf{x}} g_t^\theta(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))||_2 \leq C \sum_{m=1}^J \left| \nabla_x \log \bar{f}_{m,t|t-1} \right|.$$

Hence, we have that

$$\sup_{\theta \in \Theta} |[\hat{\boldsymbol{\phi}}_t(\mathbf{s};\theta) - \boldsymbol{\phi}_t(\mathbf{s};\theta)]_j| \leq C \sum_{m,n=1}^J \sup_{\upsilon \in \Upsilon_m} \sup_{x \in \mathcal{X}_{\Upsilon_m}} \left| \frac{\partial \log f_{m,t|t-1}}{\partial x} \right| \sup_{\upsilon_n \in \Upsilon_n} |\hat{X}_{n,t}(\upsilon_n) - X_{n,t}(\upsilon_n)|$$

for all $j \in \{1, ..., J\}$ where $\frac{\partial \log f_{j,t|t-1}}{\partial x}$ denotes $\frac{\partial \log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}$. Condition (ii) then follows from Lemmata 2.1 and 2.2 in Straumann and Mikosch (2006) since $(Y_t)_{t\in\mathbb{Z}}$ is stationary by Assumption 1, for each $j \in \{1, ..., J\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j} \sup_{x\in\mathcal{X}_{\Upsilon_j}} \left|\frac{\partial \log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}\right|^{m_j}\right] < \infty$ by assumption, and, for each $j \in \{1, ..., J\}$, $\sup_{\upsilon_j\in\Upsilon_j} |\hat{X}_{j,t}(\upsilon_j) - X_{j,t}(\upsilon_j)| \overset{e.a.s.}{\to} 0$ as $t \to \infty$ by Lemma 5.1. Moreover, we have that

$$\sup_{\substack{\mathbf{x},\mathbf{y}\in\mathcal{S}\\ \mathbf{x}\neq\mathbf{y}}} \frac{|[(\hat{\boldsymbol{\phi}}_t(\mathbf{x};\theta) - \boldsymbol{\phi}_t(\mathbf{x};\theta)) - (\hat{\boldsymbol{\phi}}_t(\mathbf{y};\theta) - \boldsymbol{\phi}_t(\mathbf{y};\theta))]_j|}{||\mathbf{x}-\mathbf{y}||_2} \leq \sup_{\mathbf{s}\in\mathcal{S}} \left|\left|\frac{\partial[\hat{\boldsymbol{\phi}}_t(\mathbf{s};\theta)]_j}{\partial \mathbf{s}} - \frac{\partial[\boldsymbol{\phi}_t(\mathbf{s};\theta)]_j}{\partial \mathbf{s}}\right|\right|_2$$

for all $j \in \{1, ..., J\}$ where

$$\begin{aligned}\left|\left[\frac{\partial[\hat{\boldsymbol{\phi}}_t(\mathbf{s};\theta)]_j}{\partial \mathbf{s}} - \frac{\partial[\boldsymbol{\phi}_t(\mathbf{s};\theta)]_j}{\partial \mathbf{s}}\right]_r\right| &\leq \left|\frac{\hat{f}_{j,t|t-1}p_{rj}}{\sum_{k=1}^J\sum_{l=1}^J \hat{f}_{k,t|t-1}p_{lk}s_l} - \frac{f_{j,t|t-1}p_{rj}}{\sum_{k=1}^J\sum_{l=1}^J f_{k,t|t-1}p_{lk}s_l}\right| \\ &+ \left|\frac{(\sum_{i=1}^J \hat{f}_{j,t|t-1}p_{ij}s_i)(\sum_{k=1}^J \hat{f}_{k,t|t-1}p_{rk})}{(\sum_{k=1}^J\sum_{l=1}^J \hat{f}_{k,t|t-1}p_{lk}s_l)^2} - \frac{(\sum_{i=1}^J f_{j,t|t-1}p_{ij}s_i)(\sum_{k=1}^J f_{k,t|t-1}p_{rk})}{(\sum_{k=1}^J\sum_{l=1}^J f_{k,t|t-1}p_{lk}s_l)^2}\right|\end{aligned}$$

for all $r \in \{1, ..., J\}$ where $\hat{f}_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j)$ and $f_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)$ as above. First, let $h^\theta_{1,t} : \mathcal{X}_{\upsilon_1} \times \cdots \times \mathcal{X}_{\upsilon_J} \to \mathbb{R}$ be given by

$$h^\theta_{1,t}(\mathbf{x}) = \frac{f_{j,t|t-1}(Y_t; x_j, \upsilon_j)p_{rj}}{\sum_{k=1}^J\sum_{l=1}^J f_{k,t|t-1}(Y_t; x_k, \upsilon_k)p_{lk}s_l}.$$

Then, there exists an $\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}) \in \{c\hat{\mathbf{X}}_t(\boldsymbol{\upsilon}) + (1-c)\mathbf{X}_t(\boldsymbol{\upsilon}) : c \in [0, 1]\}$ (not necessarily equal to the one above) such that

$$|h^\theta_{1,t}(\hat{\mathbf{X}}_t(\boldsymbol{\upsilon})) - h^\theta_{1,t}(\mathbf{X}_t(\boldsymbol{\upsilon}))| \leq ||\nabla_{\mathbf{x}} h^\theta_{1,t}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))||_2 ||\hat{\mathbf{X}}_t(\boldsymbol{\upsilon}) - \mathbf{X}_t(\boldsymbol{\upsilon})||_2,$$

where, if $m \neq j$,

$$|[\nabla_{\mathbf{x}} h^\theta_{1,t}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))]_m| \leq \frac{\bar{f}_{j,t|t-1}p_{rj}}{(\sum_{k=1}^J\sum_{l=1}^J \bar{f}_{k,t|t-1}p_{lk}s_l)^2}\sum_{l=1}^J |\nabla_x \bar{f}_{m,t|t-1}|p_{lm}s_l \leq C|\nabla_x \log \bar{f}_{m,t|t-1}|$$

and, if $m = j$,

$$\begin{aligned}|[\nabla_{\mathbf{x}} h^\theta_{1,t}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))]_m| &\leq \frac{|\nabla_x \bar{f}_{m,t|t-1}|p_{rm}}{\sum_{k=1}^J\sum_{l=1}^J \bar{f}_{k,t|t-1}p_{lk}s_l} \\ &+ \frac{\bar{f}_{m,t|t-1}p_{rm}}{(\sum_{k=1}^J\sum_{l=1}^J \bar{f}_{k,t|t-1}p_{lk}s_l)^2}\sum_{l=1}^J |\nabla_x \bar{f}_{m,t|t-1}|p_{lm}s_l \leq C|\nabla_x \log \bar{f}_{m,t|t-1}|\end{aligned}$$

where $\bar{f}_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t; \bar{X}_{j,t}(\upsilon_j), \upsilon_j)$ so

$$||\nabla_{\mathbf{x}} h_{1,t}^{\theta}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))||_2 \leq C \sum_{m=1}^{J} \left|\nabla_x \log \bar{f}_{m,t|t-1}\right|.$$

Moreover, let $h_{2,t}^{\theta} : \mathcal{X}_{\upsilon_1} \times \cdots \times \mathcal{X}_{\upsilon_J} \to \mathbb{R}$ be given by

$$h_{2,t}^{\theta}(\mathbf{x}) = \frac{(\sum_{i=1}^{J} f_{j,t|t-1}(Y_t; x_j, \upsilon_j) p_{ij} s_i)(\sum_{k=1}^{J} f_{k,t|t-1}(Y_t; x_k, \upsilon_k) p_{rk})}{(\sum_{k=1}^{J} \sum_{l=1}^{J} f_{k,t|t-1}(Y_t; x_k, \upsilon_k) p_{lk} s_l)^2}.$$

Again, there exists an $\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}) \in \{c\hat{\mathbf{X}}_t(\boldsymbol{\upsilon}) + (1-c)\mathbf{X}_t(\boldsymbol{\upsilon}) : c \in [0,1]\}$ (not necessarily equal to the ones above) such that

$$|h_{2,t}^{\theta}(\hat{\mathbf{X}}_t(\boldsymbol{\upsilon})) - h_{2,t}^{\theta}(\mathbf{X}_t(\boldsymbol{\upsilon}))| \leq ||\nabla_{\mathbf{x}} h_{2,t}^{\theta}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))||_2 ||\hat{\mathbf{X}}_t(\boldsymbol{\upsilon}) - \mathbf{X}_t(\boldsymbol{\upsilon})||_2,$$

where, if $m \neq j$,

$$\begin{aligned} |[\nabla_{\mathbf{x}} h_{2,t}^{\theta}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))]_m| \leq{} & \frac{(\sum_{i=1}^{J} \bar{f}_{j,t|t-1} p_{ij} s_i)(|\nabla_x \bar{f}_{m,t|t-1}| p_{rm})}{(\sum_{k=1}^{J} \sum_{l=1}^{J} \bar{f}_{k,t|t-1} p_{lk} s_l)^2} \\ & + 2\frac{(\sum_{i=1}^{J} \bar{f}_{j,t|t-1} p_{ij} s_i)(\sum_{k=1}^{J} \bar{f}_{k,t|t-1} p_{rk})}{(\sum_{k=1}^{J} \sum_{l=1}^{J} \bar{f}_{k,t|t-1} p_{lk} s_l)^3} \sum_{l=1}^{J} |\nabla_x \bar{f}_{m,t|t-1}| p_{lm} s_l \leq C |\nabla_x \log \bar{f}_{m,t|t-1}| \end{aligned}$$

and, if $m = j$,

$$\begin{aligned} |[\nabla_{\mathbf{x}} h_{2,t}^{\theta}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))]_m| \leq{} & \frac{(\sum_{i=1}^{J} |\nabla_x \bar{f}_{m,t|t-1}| p_{im} s_i)(\sum_{k=1}^{J} \bar{f}_{k,t|t-1} p_{rk})}{(\sum_{k=1}^{J} \sum_{l=1}^{J} \bar{f}_{k,t|t-1} p_{lk} s_l)^2} \\ & + \frac{(\sum_{i=1}^{J} \bar{f}_{m,t|t-1} p_{im} s_i)(|\nabla_x \bar{f}_{m,t|t-1}| p_{rm})}{(\sum_{k=1}^{J} \sum_{l=1}^{J} \bar{f}_{k,t|t-1} p_{lk} s_l)^2} \\ & + 2\frac{(\sum_{i=1}^{J} \bar{f}_{m,t|t-1} p_{im} s_i)(\sum_{k=1}^{J} \bar{f}_{k,t|t-1} p_{rk})}{(\sum_{k=1}^{J} \sum_{l=1}^{J} \bar{f}_{k,t|t-1} p_{lk} s_l)^3} \sum_{l=1}^{J} |\nabla_x \bar{f}_{m,t|t-1}| p_{lm} s_l \leq C |\nabla_x \log \bar{f}_{m,t|t-1}| \end{aligned}$$

where $\bar{f}_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t; \bar{X}_{j,t}(\upsilon_j), \upsilon_j)$ so

$$||\nabla_{\mathbf{x}} h_{2,t}^{\theta}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))||_2 \leq C \sum_{m=1}^{J} \left|\nabla_x \log \bar{f}_{m,t|t-1}\right|.$$

Hence, we have that

$$\sup_{\theta \in \Theta} \sup_{\mathbf{s} \in \mathcal{S}} \left| \left[\frac{\partial [\hat{\boldsymbol{\phi}}_t(\mathbf{s};\theta)]_j}{\partial \mathbf{s}} - \frac{\partial [\boldsymbol{\phi}_t(\mathbf{s};\theta)]_j}{\partial \mathbf{s}} \right]_r \right| \leq C \sum_{m,n=1}^{J} \sup_{\upsilon \in \Upsilon_m} \sup_{x \in \mathcal{X}_{\Upsilon_m}} \left| \frac{\partial \log f_{m,t|t-1}}{\partial x} \right| \sup_{\upsilon_n \in \Upsilon_n} |\hat{X}_{n,t}(\upsilon_n) - X_{n,t}(\upsilon_n)|$$

for all $j \in \{1, ..., J\}$ and $r \in \{1, ..., J\}$ where $\frac{\partial \log f_{j,t|t-1}}{\partial x}$ denotes $\frac{\partial \log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}$ as above. Condition (iii) then follows from Lemmata 2.1 and 2.2 in Straumann and Mikosch (2006) as well since $(Y_t)_{t\in\mathbb{Z}}$ is stationary by Assumption 1, for each $j \in \{1, ..., J\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial \log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}\right|^{m_j}\right] < \infty$ by assumption, and, for each $j \in \{1, ..., J\}$, $\sup_{\upsilon_j\in\Upsilon_j}|\hat{X}_{j,t}(\upsilon_j) - X_{j,t}(\upsilon_j)| \overset{e.a.s.}{\to} 0$ as $t \to \infty$ by Lemma 5.1.

9.3. Theorem 5.1

First, recall that $\hat{L}_T(\theta)$ is given by $\hat{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^T \log \hat{f}_{t|t-1}(Y_t;\theta)$ where $\hat{f}_{t|t-1}(Y_t;\theta) = \sum_{j=1}^J \hat{\pi}_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)$, and let $L_T(\theta)$ be given by $L_T(\theta) = \frac{1}{T}\sum_{t=1}^T \log f_{t|t-1}(Y_t;\theta)$ where $f_{t|t-1}(Y_t;\theta) = \sum_{j=1}^J \pi_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)$. Moreover, let $L(\theta) = \mathbb{E}[\log f_{t|t-1}(Y_t;\theta)]$. Finally, let $C \in \mathbb{R}$ be an arbitrary constant that can change throughout the proof.

The conclusion follows from Lemmata 3.1 and 4.1 in Pötscher and Prucha (1997) if

(i) Θ is compact,
(ii) $\sup_{\theta\in\Theta}|\hat{L}_T(\theta) - L(\theta)| \overset{a.s.}{\to} 0$ as $T \to \infty$,
(iii) $\theta \mapsto L(\theta)$ is continuous on Θ, and
(iv) $L(\theta) \le L(\theta_0)$ for all $\theta \in \Theta$ with equality if and only if $\theta = \theta_0$.

Condition (i) follows from Assumption 2. Condition (ii) follows from Lemmata 9.3 and 9.4 since

$$\sup_{\theta\in\Theta}|\hat{L}_T(\theta) - L(\theta)| \le \sup_{\theta\in\Theta}|\hat{L}_T(\theta) - L_T(\theta)| + \sup_{\theta\in\Theta}|L_T(\theta) - L(\theta)|,$$

and Condition (iii) is a by-product of the law of large numbers used in the proof of Lemma 9.4.

Lemma 9.3. *Assume that Assumption 1-3 and the conditions in Lemmata 5.1 and 5.2 hold. Then,*

$$\sup_{\theta\in\Theta}|\hat{L}_T(\theta) - L_T(\theta)| \overset{a.s.}{\to} 0 \quad as \quad T \to \infty.$$

Proof 3. *We have that*

$$|\hat{L}_T(\theta) - L_T(\theta)| \le \frac{1}{T}\sum_{t=1}^T |\log \hat{f}_{t|t-1} - \log \tilde{f}_{t|t-1}| + \frac{1}{T}\sum_{t=1}^T |\log \tilde{f}_{t|t-1} - \log f_{t|t-1}|,$$

where $\hat{f}_{t|t-1}$ denotes $\hat{f}_{t|t-1}(Y_t;\theta)$, $\tilde{f}_{t|t-1}$ denotes $\tilde{f}_{t|t-1}(Y_t;\theta) = \sum_{j=1}^J \pi_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)$, and $f_{t|t-1}$ denotes $f_{t|t-1}(Y_t;\theta)$. First, let $g^\theta_{1,t} : \mathcal{S}_\theta \to \mathbb{R}$ be given by

$$g^\theta_{1,t}(\mathbf{s}) = \log \sum_{j=1}^J s_j f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j).$$

An application of the mean value theorem and the Cauchy-Schwarz inequality shows that there exists an $\bar{\boldsymbol{\pi}}_{t|t-1}(\theta) \in \{c\hat{\boldsymbol{\pi}}_{t|t-1}(\theta) + (1-c)\boldsymbol{\pi}_{t|t-1}(\theta) : c \in [0,1]\}$ *such that*

$$|g^{\theta}_{1,t}(\hat{\boldsymbol{\pi}}_{t|t-1}(\theta)) - g^{\theta}_{1,t}(\boldsymbol{\pi}_{t|t-1}(\theta))| \leq ||\nabla_{\mathbf{s}} g^{\theta}_{1,t}(\bar{\boldsymbol{\pi}}_{t|t-1}(\theta))||_2 ||\hat{\boldsymbol{\pi}}_{t|t-1}(\theta) - \boldsymbol{\pi}_{t|t-1}(\theta)||_2,$$

where

$$\begin{aligned}
|[\nabla_{\mathbf{s}} g^{\theta}_{1,t}(\bar{\boldsymbol{\pi}}_{t|t-1}(\theta))]_k| &= \frac{f_{k,t|t-1}(Y_t; \hat{X}_{k,t}(\upsilon_k), \upsilon_k)}{\sum_{j=1}^{J} \bar{\pi}_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j)} \\
&= \frac{1}{\bar{\pi}_{k,t|t-1}(\theta)} \frac{\bar{\pi}_{k,t|t-1}(\theta) f_{k,t|t-1}(Y_t; \hat{X}_{k,t}(\upsilon_k), \upsilon_k)}{\sum_{j=1}^{J} \bar{\pi}_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j)} \\
&\leq C
\end{aligned}$$

by Remark 5.1 so

$$||\nabla_{\mathbf{s}} g^{\theta}_{1,t}(\bar{\boldsymbol{\pi}}_{t|t-1}(\theta))||_2 \leq C.$$

Moreover, let $g^{\theta}_{2,t} : \mathcal{X}_{\upsilon_1} \times \cdots \times \mathcal{X}_{\upsilon_J} \to \mathbb{R}$ *be given by*

$$g^{\theta}_{2,t}(\mathbf{x}) = \log \sum_{j=1}^{J} \pi_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t; x_j, \upsilon_j).$$

Another application of the mean value theorem and the Cauchy-Schwarz inequality shows that there exists an $\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}) \in \{c\hat{\mathbf{X}}_t(\boldsymbol{\upsilon}) + (1-c)\mathbf{X}_t(\boldsymbol{\upsilon}) : c \in [0,1]\}$ *such that*

$$|g^{\theta}_{2,t}(\hat{\mathbf{X}}_t(\boldsymbol{\upsilon})) - g^{\theta}_{2,t}(\mathbf{X}_t(\boldsymbol{\upsilon}))| \leq ||\nabla_{\mathbf{x}} g^{\theta}_{2,t}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))||_2 ||\hat{\mathbf{X}}_t(\boldsymbol{\upsilon}) - \mathbf{X}_t(\boldsymbol{\upsilon})||_2,$$

where

$$\begin{aligned}
|[\nabla_{\mathbf{x}} g^{\theta}_{2,t}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))]_k| &= \frac{\pi_{k,t|t-1}(\theta) |\nabla_x f_{k,t|t-1}(Y_t; \bar{X}_{k,t}(\upsilon_k), \upsilon_k)|}{\sum_{j=1}^{J} \pi_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t; \bar{X}_{j,t}(\upsilon_j), \upsilon_j)} \\
&= \frac{\pi_{k,t|t-1}(\theta) f_{k,t|t-1}(Y_t; \bar{X}_{k,t}(\upsilon_k), \upsilon_k)}{\sum_{j=1}^{J} \pi_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t; \bar{X}_{j,t}(\upsilon_j), \upsilon_j)} |\nabla_x \log f_{k,t|t-1}(Y_t; \bar{X}_{k,t}(\upsilon_k), \upsilon_k)| \\
&\leq |\nabla_x \log f_{k,t|t-1}(Y_t; \bar{X}_{k,t}(\upsilon_k), \upsilon_k)|
\end{aligned}$$

so

$$||\nabla_{\mathbf{x}} g^{\theta}_{2,t}(\bar{\mathbf{X}}_t(\boldsymbol{\upsilon}))||_2 \leq \sum_{k=1}^{J} \left|\nabla_x \log f_{k,t|t-1}(Y_t; \bar{X}_{k,t}(\upsilon_k), \upsilon_k)\right|.$$

Hence, we have that

$$\sup_{\theta \in \Theta} |\hat{L}_T(\theta) - L_T(\theta)| \leq C \frac{1}{T} \sum_{t=1}^{T} \sup_{\theta \in \Theta} ||\hat{\boldsymbol{\pi}}_{t|t-1}(\theta) - \boldsymbol{\pi}_{t|t-1}(\theta)||_2$$

$$+\frac{1}{T}\sum_{t=1}^{T}\sum_{k,l=1}^{J}\sup_{\upsilon\in\Upsilon_k}\sup_{x\in\mathcal{X}_{\Upsilon_k}}\left|\frac{\partial\log f_{k,t|t-1}}{\partial x}\right|\sup_{\upsilon_l\in\Upsilon_l}|\hat{X}_{l,t}(\upsilon_l)-X_{l,t}(\upsilon_l)|,$$

where $\frac{\partial\log f_{j,t|t-1}}{\partial x}$ *denotes* $\frac{\partial\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}$. *The conclusion thus follows from Lemmata 2.1 and 2.2 in Straumann and Mikosch (2006) since* $(Y_t)_{t\in\mathbb{Z}}$ *is stationary by Assumption 1, for each* $j\in\{1,...,J\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}\right|^{m_j}\right]<\infty$ *by assumption, for each* $j\in\{1,...,J\}$, $\sup_{\upsilon_j\in\Upsilon_j}|\hat{X}_{j,t}(\upsilon_j)-X_{j,t}(\upsilon_j)|\overset{e.a.s.}{\to}0$ *as* $t\to\infty$ *by Lemma 5.1, and* $\sup_{\theta\in\Theta}||\hat{\boldsymbol{\pi}}_{t|t-1}(\theta)-\boldsymbol{\pi}_{t|t-1}(\theta)||_2\overset{e.a.s.}{\to}0$ *as* $t\to\infty$ *by Corollary 5.1.*

Lemma 9.4. *Assume that Assumption 1-4 and the conditions in Lemmata 5.1 and 5.2 hold. Then,*

$$\sup_{\theta\in\Theta}|L_T(\theta)-L(\theta)|\overset{a.s.}{\to}0\quad as\quad T\to\infty.$$

Proof 4. *First,* $(\log f_{t|t-1}(Y_t;\theta))_{t\in\mathbb{Z}}$ *is stationary and ergodic for all* $\theta\in\Theta$ *since* $(Y_t)_{t\in\mathbb{Z}}$ *is stationary and ergodic by Assumption 1, for each* $j\in\{1,...,J\}$, $(X_{j,t}(\upsilon_j))_{t\in\mathbb{Z}}$ *is stationary and ergodic for all* $\upsilon_j\in\Upsilon_j$ *by Lemma 5.1, and* $(\boldsymbol{\pi}_{t|t-1}(\theta))_{t\in\mathbb{Z}}$ *is stationary and ergodic for all* $\theta\in\Theta$ *by Corollary 5.1. Note that*

$$|\log f_{t|t-1}(Y_t;\theta)|\leq\sum_{j=1}^{J}|\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|$$

since

$$\begin{aligned}\log f_{t|t-1}(Y_t;\theta)&\geq\log\min_{j\in\{1,...,J\}}f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)\\&=\min_{j\in\{1,...,J\}}\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)\\&\geq\min_{j\in\{1,...,J\}}-|\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|\geq-\sum_{j=1}^{J}|\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|\end{aligned}$$

and

$$\begin{aligned}\log f_{t|t-1}(Y_t;\theta)&\leq\log\max_{j\in\{1,...,J\}}f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)\\&=\max_{j\in\{1,...,J\}}\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)\\&\leq\max_{j\in\{1,...,J\}}|\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|\leq\sum_{j=1}^{J}|\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|.\end{aligned}$$

Hence, $\mathbb{E}\left[\sup_{\theta\in\Theta}|\log f_{t|t-1}(Y_t;\theta)|\right]<\infty$ *since, for each* $j\in\{1,...,J\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j}|\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon),\upsilon)|\right]<\infty$ *by Assumption 4. The conclusion thus follows from the uniform law of large numbers by Rao (1962).*

Finally, Condition (iv) follows from Lemma 9.5.

Lemma 9.5. *Assume that Assumption 1-5 and the conditions in Lemmata 5.1 and 5.2 hold. Then,*

$$L(\theta) \leq L(\theta_0)$$

for all $\theta \in \Theta$ with equality if and only if $\theta = \theta_0$.

Proof 5. *We have that*

$$L(\theta) - L(\theta_0) = \mathbb{E}\left[\log \frac{f_{t|t-1}(Y_t;\theta)}{f_{t|t-1}(Y_t;\theta_0)}\right].$$

Observe that

$$\begin{aligned}\mathbb{E}\left[\log \frac{f_{t|t-m}(Y_t,...,Y_{t-m+1};\theta)}{f_{t|t-m}(Y_t,...,Y_{t-m+1};\theta_0)}\right] &= \mathbb{E}\left[\log \prod_{i=1}^{m} \frac{f_{t-i+1|t-i}(Y_{t-i+1};\theta)}{f_{t-i+1|t-i}(Y_{t-i+1};\theta_0)}\right] \\ &= \sum_{i=1}^{m} \mathbb{E}\left[\log \frac{f_{t-i+1|t-i}(Y_{t-i+1};\theta)}{f_{t-i+1|t-i}(Y_{t-i+1};\theta_0)}\right] = m\mathbb{E}\left[\log \frac{f_{t|t-1}(Y_t;\theta)}{f_{t|t-1}(Y_t;\theta_0)}\right],\end{aligned}$$

so

$$\mathbb{E}\left[\log \frac{f_{t|t-1}(Y_t;\theta)}{f_{t|t-1}(Y_t;\theta_0)}\right] = \frac{1}{m}\mathbb{E}\left[\log \frac{f_{t|t-m}(Y_t,...,Y_{t-m+1};\theta)}{f_{t|t-m}(Y_t,...,Y_{t-m+1};\theta_0)}\right].$$

Note that $\log x \leq x - 1$ for all $x > 0$ with equality if and only if $x = 1$. Thus,

$$\mathbb{E}\left[\log \frac{f_{t|t-m}(Y_t,...,Y_{t-m+1};\theta)}{f_{t|t-m}(Y_t,...,Y_{t-m+1};\theta_0)}\right] \leq 0$$

for all $\theta \in \Theta$ with equality if and only if $f_{t|t-m}(Y_t,...,Y_{t-m+1};\theta) = f_{t|t-m}(Y_t,...,Y_{t-m+1};\theta_0)$ a.s. Hence, we have that

$$L(\theta) - L(\theta_0) \leq 0$$

for all $\theta \in \Theta$ with equality if and only if $\theta = \theta_0$ by Assumption 5.

9.4. Theorem 5.2

First, $\nabla_\theta \hat{L}_T(\theta)$ is given by

$$\nabla_\theta \hat{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{\hat{f}_{t|t-1}(Y_t;\theta)} \nabla_\theta \hat{f}_{t|t-1}(Y_t;\theta),$$

where

$$\nabla_\theta \hat{f}_{t|t-1}(Y_t;\theta) = \sum_{j=1}^{J}\Big(\nabla_\theta \hat{\pi}_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j) + \hat{\pi}_{j,t|t-1}(\theta)\nabla_\theta f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)\Big),$$

with $\nabla_\theta f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j) = \bar{\nabla}_x f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)\nabla_\theta \hat{X}_{j,t}(\upsilon_j) + \bar{\nabla}_\upsilon f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)$, $\bar{\nabla}_x f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j) = \frac{\partial f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}\Big|_{x=\hat{X}_{j,t}(\upsilon_j),\upsilon=\upsilon_j}$,

$$\bar{\nabla}_\upsilon f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j) = \Big(0, \bar{\nabla}_{\upsilon_1} f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j), ..., \bar{\nabla}_{\upsilon_{d_j}} f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j), 0\Big)',$$

and $\bar{\nabla}_{\upsilon_i} f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j) = \frac{\partial f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial \upsilon_i}\Big|_{x=\hat{X}_{j,t}(\upsilon_j),\upsilon=\upsilon_j}$ where $\nabla_\theta \hat{\pi}_{j,t|t-1}(\theta)$ and $\nabla_\theta \hat{X}_{j,t}(\upsilon_j)$ are given in Lemma X and Lemma X, respectively, and $\nabla_\theta L_T(\theta)$ is given by

$$\nabla_\theta L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \frac{1}{f_{t|t-1}(Y_t;\theta)} \nabla_\theta f_{t|t-1}(Y_t;\theta),$$

where

$$\nabla_\theta f_{t|t-1}(Y_t;\theta) = \sum_{j=1}^{J}\Big(\nabla_\theta \pi_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j) + \pi_{j,t|t-1}(\theta)\nabla_\theta f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)\Big),$$

with $\nabla_\theta f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j) = \bar{\nabla}_x f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)\nabla_\theta X_{j,t}(\upsilon_j) + \bar{\nabla}_\upsilon f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)$, $\bar{\nabla}_x f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j) = \frac{\partial f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}\Big|_{x=X_{j,t}(\upsilon_j),\upsilon=\upsilon_j}$,

$$\bar{\nabla}_\upsilon f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j) = \Big(0, \bar{\nabla}_{\upsilon_1} f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j), ..., \bar{\nabla}_{\upsilon_{d_j}} f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j), 0\Big)',$$

and $\bar{\nabla}_{\upsilon_i} f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j) = \frac{\partial f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial \upsilon_i}\Big|_{x=X_{j,t}(\upsilon_j),\upsilon=\upsilon_j}$ where $\nabla_\theta \pi_{j,t|t-1}(\theta)$ and $\nabla_\theta X_{j,t}(\upsilon_j)$ are given in Lemma X and Lemma X, respectively. Moreover, $\nabla_{\theta\theta}\hat{L}_T(\theta)$ is given by

$$\nabla_{\theta\theta}\hat{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\Big(-\frac{1}{\hat{f}^2_{t|t-1}(Y_t;\theta)}\nabla_\theta \hat{f}_{t|t-1}(Y_t;\theta)\nabla_{\theta'}\hat{f}_{t|t-1}(Y_t;\theta) + \frac{1}{\hat{f}_{t|t-1}(Y_t;\theta)}\nabla_{\theta\theta}\hat{f}_{t|t-1}(Y_t;\theta)\Big),$$

where

$$\begin{aligned}\nabla_{\theta\theta}\hat{f}_{t|t-1}(Y_t;\theta) = \sum_{j=1}^{J}\Big(&\nabla_{\theta\theta}\hat{\pi}_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j) + \nabla_\theta\hat{\pi}_{j,t|t-1}(\theta)\nabla_{\theta'} f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)\\ &+ \nabla_\theta f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)\nabla_{\theta'}\hat{\pi}_{j,t|t-1}(\theta) + \hat{\pi}_{j,t|t-1}(\theta)\nabla_{\theta\theta} f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)\Big),\end{aligned}$$

with $\nabla_{\theta\theta} f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j) = \bar{\nabla}_{xx} f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)\nabla_\theta\hat{X}_{j,t}(\upsilon_j)\nabla_{\theta'}\hat{X}_{j,t}(\upsilon_j) + \nabla_\theta\hat{X}_{j,t}(\upsilon_j)\bar{\nabla}_{x\upsilon} f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j) + \bar{\nabla}_x f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)\nabla_{\theta\theta}\hat{X}_{j,t}(\upsilon_j) +$

$\bar{\nabla}_{\upsilon x} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j) \nabla_{\theta'} \hat{X}_{j,t}(\upsilon_j) + \bar{\nabla}_{\upsilon\upsilon} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j)$, $\bar{\nabla}_{xx} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j) = \frac{\partial^2 f_{j,t|t-1}(Y_t; x, \upsilon)}{\partial x^2}\Big|_{x=\hat{X}_{j,t}(\upsilon_j), \upsilon=\upsilon_j}$,

$$\bar{\nabla}_{\upsilon\upsilon} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j) = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & \bar{\nabla}_{\upsilon_1\upsilon_1} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j) & \cdots & \bar{\nabla}_{\upsilon_1\upsilon_{d_j}} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \bar{\nabla}_{\upsilon_{d_j}\upsilon_1} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j) & \cdots & \bar{\nabla}_{\upsilon_{d_j}\upsilon_{d_j}} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j) & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

and $\bar{\nabla}_{\upsilon_i\upsilon_k} f_{j,t|t-1}(Y_t; \hat{X}_{j,t}(\upsilon_j), \upsilon_j) = \frac{\partial f_{j,t|t-1}(Y_t; x, \upsilon)}{\partial \upsilon_i \upsilon_k}\Big|_{x=\hat{X}_{j,t}(\upsilon_j), \upsilon=\upsilon_j}$ where $\nabla_{\theta\theta} \hat{\pi}_{j,t|t-1}(\theta)$ and $\nabla_{\theta\theta} \hat{X}_{j,t}(\upsilon_j)$ are given in Lemma X and Lemma X, respectively, and $\nabla_{\theta\theta} L_T(\theta)$ is given by

$$\nabla_{\theta\theta} L_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left(-\frac{1}{f_{t|t-1}^2(Y_t; \theta)} \nabla_\theta f_{t|t-1}(Y_t; \theta) \nabla_{\theta'} f_{t|t-1}(Y_t; \theta) + \frac{1}{f_{t|t-1}(Y_t; \theta)} \nabla_{\theta\theta} f_{t|t-1}(Y_t; \theta) \right),$$

where

$$\nabla_{\theta\theta} f_{t|t-1}(Y_t; \theta) = \sum_{j=1}^{J} \Big(\nabla_{\theta\theta} \pi_{j,t|t-1}(\theta) f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) + \nabla_\theta \pi_{j,t|t-1}(\theta) \nabla_{\theta'} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)$$
$$+ \nabla_\theta f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) \nabla_{\theta'} \pi_{j,t|t-1}(\theta) + \pi_{j,t|t-1}(\theta) \nabla_{\theta\theta} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) \Big),$$

with $\nabla_{\theta\theta} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) = \bar{\nabla}_{xx} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) \nabla_\theta X_{j,t}(\upsilon_j) \nabla_{\theta'} X_{j,t}(\upsilon_j) + \nabla_\theta X_{j,t}(\upsilon_j) \bar{\nabla}_{x\upsilon} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) + \bar{\nabla}_x f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) \nabla_{\theta\theta} X_{j,t}(\upsilon_j) + \bar{\nabla}_{\upsilon x} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) \nabla_{\theta'} X_{j,t}(\upsilon_j) + \bar{\nabla}_{\upsilon\upsilon} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)$, $\bar{\nabla}_{xx} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) = \frac{\partial^2 f_{j,t|t-1}(Y_t; x, \upsilon)}{\partial x^2}\Big|_{x=X_{j,t}(\upsilon_j), \upsilon=\upsilon_j}$,

$$\bar{\nabla}_{\upsilon\upsilon} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & \bar{\nabla}_{\upsilon_1\upsilon_1} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) & \cdots & \bar{\nabla}_{\upsilon_1\upsilon_{d_j}} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \bar{\nabla}_{\upsilon_{d_j}\upsilon_1} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) & \cdots & \bar{\nabla}_{\upsilon_{d_j}\upsilon_{d_j}} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

and $\bar{\nabla}_{\upsilon_i\upsilon_k} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) = \frac{\partial f_{j,t|t-1}(Y_t; x, \upsilon)}{\partial \upsilon_i \upsilon_k}\Big|_{x=X_{j,t}(\upsilon_j), \upsilon=\upsilon_j}$ where $\nabla_{\theta\theta} \pi_{j,t|t-1}(\theta)$ and $\nabla_{\theta\theta} X_{j,t}(\upsilon_j)$ are given in Lemma X and Lemma X, respectively. Finally, let $I(\theta) = -\mathbb{E}[\nabla_{\theta\theta} \log f_{t|t-1}(Y_t; \theta)]$. As above, let $C \in \mathbb{R}$ be an arbitrary constant that can change throughout the proof.

The conclusion follows from Lemma 8.1 in Pötscher and Prucha (1997) if

(i) $\theta_0 \in int(\Theta)$,

(ii) $\hat{\theta}_T \xrightarrow{\mathbb{P}} \theta_0$ as $T \to \infty$,

(iii) $\theta \mapsto \hat{L}_T(\theta)$ is twice continuously differentiable on $int(\Theta)$ a.s.,

(iv) $||\sqrt{T}\nabla_\theta \hat{L}_T(\hat{\theta}_T)||_2 \xrightarrow{\mathbb{P}} 0$ as $T \to \infty$,

(v) $\sqrt{T}\nabla_\theta \hat{L}_T(\theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0))$ as $T \to \infty$, and

(vi) $\sup_{\theta\in\Theta} ||\nabla_{\theta\theta}\hat{L}_T(\theta) - (-I(\theta))||_{2,2} \xrightarrow{\mathbb{P}} 0$ as $T \to \infty$.

First, Condition (i) follows from Assumption 6, Condition (ii) follows from Theorem 5.1, and Condition (iii) follows from Assumption 7. Note that Condition (iv) is a consequence of Condition (i) and (ii). Moreover, Condition (v) follows from Lemmata 9.6, 9.7, and 9.8.

Lemma 9.6. *Assume that ... Then,*

$$\sup_{\theta\in\Theta} ||\nabla_\theta \hat{L}_T(\theta) - \nabla_\theta L_T(\theta)||_2 \xrightarrow{a.s.} 0 \quad as \quad T \to \infty.$$

Proof 6. *We have that*

$$||\nabla_\theta \hat{L}_T(\theta) - \nabla_\theta L_T(\theta)||_2 \leq \frac{1}{T}\sum_{t=1}^{T} \left|\left| \frac{1}{\hat{f}_{t|t-1}} \nabla_\theta \hat{f}_{t|t-1} - \frac{1}{f_{t|t-1}} \nabla_\theta f_{t|t-1} \right|\right|_2,$$

where $\hat{f}_{t|t-1}$ denotes $\hat{f}_{t|t-1}(Y_t;\theta)$ and $f_{t|t-1}$ denotes $f_{t|t-1}(Y_t;\theta)$ and that

$$\begin{aligned}
\left|\left| \frac{1}{\hat{f}_{t|t-1}} \nabla_\theta \hat{f}_{t|t-1} - \frac{1}{f_{t|t-1}} \nabla_\theta f_{t|t-1} \right|\right|_2 \leq\ & C \sum_{j=1}^{J} \left|\left| \nabla_\theta \hat{\pi}_{j,t|t-1} - \nabla_\theta \pi_{j,t|t-1} \right|\right|_2 \\
&+ \sum_{j=1}^{J} \left|\left| \nabla_\theta \pi_{j,t|t-1} \right|\right|_2 \left| \frac{1}{\hat{f}_{t|t-1}} \hat{f}_{j,t|t-1} - \frac{1}{\tilde{f}_{t|t-1}} \hat{f}_{j,t|t-1} \right| \\
&+ \sum_{j=1}^{J} \left|\left| \nabla_\theta \pi_{j,t|t-1} \right|\right|_2 \left| \frac{1}{\tilde{f}_{t|t-1}} \hat{f}_{j,t|t-1} - \frac{1}{f_{t|t-1}} f_{j,t|t-1} \right| \\
&+ \sum_{j=1}^{J} \left| \bar{\nabla}_x \log \hat{f}_{j,t|t-1} \right| \left|\left| \nabla_\theta \hat{X}_{j,t} - \nabla_\theta X_{j,t} \right|\right|_2 \\
&+ \sum_{j=1}^{J} ||\nabla_\theta X_{j,t}||_2 \left| \frac{1}{\hat{f}_{t|t-1}} \hat{\pi}_{j,t|t-1} \bar{\nabla}_x \hat{f}_{j,t|t-1} - \frac{1}{\tilde{f}_{t|t-1}} \pi_{j,t|t-1} \bar{\nabla}_x \hat{f}_{j,t|t-1} \right| \\
&+ \sum_{j=1}^{J} ||\nabla_\theta X_{j,t}||_2 \left| \frac{1}{\tilde{f}_{t|t-1}} \pi_{j,t|t-1} \bar{\nabla}_x \hat{f}_{j,t|t-1} - \frac{1}{f_{t|t-1}} \pi_{j,t|t-1} \bar{\nabla}_x f_{j,t|t-1} \right|
\end{aligned}$$

$$
+\sum_{j=1}^{J}\sum_{h=1}^{d_j}\left|\frac{1}{\hat{f}_{t|t-1}}\hat{\pi}_{j,t|t-1}\bar{\nabla}_{\upsilon_h}\hat{f}_{j,t|t-1}-\frac{1}{\tilde{f}_{t|t-1}}\pi_{j,t|t-1}\bar{\nabla}_{\upsilon_h}\hat{f}_{j,t|t-1}\right|
$$

$$
+\sum_{j=1}^{J}\sum_{h=1}^{d_j}\left|\frac{1}{\tilde{f}_{t|t-1}}\pi_{j,t|t-1}\bar{\nabla}_{\upsilon_h}\hat{f}_{j,t|t-1}-\frac{1}{f_{t|t-1}}\pi_{j,t|t-1}\bar{\nabla}_{\upsilon_h}f_{j,t|t-1}\right|,
$$

where $\hat{\pi}_{j,t|t-1}$ denotes $\hat{\pi}_{j,t|t-1}(\theta)$, $\hat{f}_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)$, $\hat{X}_{j,t}$ denotes $\hat{X}_{j,t}(\upsilon_j)$, $\tilde{f}_{t|t-1}$ denotes $\tilde{f}_{t|t-1}(Y_t;\theta)=\sum_{j=1}^{J}\pi_{j,t|t-1}(\theta)f_{j,t|t-1}(Y_t;\hat{X}_{j,t}(\upsilon_j),\upsilon_j)$, $\pi_{j,t|t-1}$ denotes $\pi_{j,t|t-1}(\theta)$, $f_{j,t|t-1}$ denotes $f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)$, and $X_{j,t}$ denotes $X_{j,t}(\upsilon_j)$. By using the same arguments as in the proof of Lemma 9.3, we have that

$$
\sup_{\theta\in\Theta}||\nabla_\theta\hat{L}_T(\theta)-\nabla_\theta L_T(\theta)||_2
$$

$$
\leq C\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}\sup_{\theta\in\Theta}\left|\left|\nabla_\theta\hat{\pi}_{j,t|t-1}(\theta)-\nabla_\theta\pi_{j,t|t-1}(\theta)\right|\right|_2
$$

$$
+C\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}\sup_{\theta\in\Theta}\left|\left|\nabla_\theta\pi_{j,t|t-1}(\theta)\right|\right|_2\sup_{\theta\in\Theta}\left|\left|\hat{\boldsymbol{\pi}}_{t|t-1}(\theta)-\boldsymbol{\pi}_{t|t-1}(\theta)\right|\right|_2
$$

$$
+C\frac{1}{T}\sum_{t=1}^{T}\sum_{j,k,l=1}^{J}\sup_{\theta\in\Theta}\left|\left|\nabla_\theta\pi_{j,t|t-1}(\theta)\right|\right|_2\sup_{\upsilon\in\Upsilon_k}\sup_{x\in\mathcal{X}_{\Upsilon_k}}\left|\frac{\partial\log f_{k,t|t-1}}{\partial x}\right|\sup_{\upsilon_l\in\Upsilon_l}|\hat{X}_{l,t}(\upsilon_l)-X_{l,t}(\upsilon_l)|
$$

$$
+\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial\log f_{j,t|t-1}}{\partial x}\right|\sup_{\upsilon_j\in\Upsilon_j}\left|\left|\nabla_\theta\hat{X}_{j,t}(\upsilon_j)-\nabla_\theta X_{j,t}(\upsilon_j)\right|\right|_2
$$

$$
+C\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}\sup_{\upsilon\in\Upsilon_j}||\nabla_\theta X_{j,t}(\upsilon)||_2\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial\log f_{j,t|t-1}}{\partial x}\right|\sup_{\theta\in\Theta}\left|\left|\hat{\boldsymbol{\pi}}_{t|t-1}(\theta)-\boldsymbol{\pi}_{t|t-1}(\theta)\right|\right|_2
$$

$$
+\frac{1}{T}\sum_{t=1}^{T}\sum_{j,l=1}^{J}\sup_{\upsilon\in\Upsilon_j}||\nabla_\theta X_{j,t}(\upsilon)||_2\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial^2\log f_{j,t|t-1}}{\partial x^2}\right|\sup_{\upsilon_l\in\Upsilon_l}|\hat{X}_{l,t}(\upsilon_l)-X_{l,t}(\upsilon_l)|
$$

$$
+C\frac{1}{T}\sum_{t=1}^{T}\sum_{j,k,l=1}^{J}\sup_{\upsilon\in\Upsilon_j}||\nabla_\theta X_{j,t}(\upsilon)||_2\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial\log f_{j,t|t-1}}{\partial x}\right|\sup_{\upsilon\in\Upsilon_k}\sup_{x\in\mathcal{X}_{\Upsilon_k}}\left|\frac{\partial\log f_{k,t|t-1}}{\partial x}\right|\sup_{\upsilon_l\in\Upsilon_l}|\hat{X}_{l,t}(\upsilon_l)-X_{l,t}(\upsilon_l)|
$$

$$
+C\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{h=1}^{d_j}\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial\log f_{j,t|t-1}}{\partial\upsilon_h}\right|\sup_{\theta\in\Theta}\left|\left|\hat{\boldsymbol{\pi}}_{t|t-1}(\theta)-\boldsymbol{\pi}_{t|t-1}(\theta)\right|\right|_2
$$

$$
+\frac{1}{T}\sum_{t=1}^{T}\sum_{j,l=1}^{J}\sum_{h=1}^{d_j}\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial^2\log f_{j,t|t-1}}{\partial\upsilon_h\partial x}\right|\sup_{\upsilon_l\in\Upsilon_l}|\hat{X}_{l,t}(\upsilon_l)-X_{l,t}(\upsilon_l)|
$$

$$+ C\frac{1}{T}\sum_{t=1}^{T}\sum_{j,k,l=1}^{J}\sum_{h=1}^{d_j}\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial\log f_{j,t|t-1}}{\partial\upsilon_h}\right|\sup_{\upsilon\in\Upsilon_k}\sup_{x\in\mathcal{X}_{\Upsilon_k}}\left|\frac{\partial\log f_{k,t|t-1}}{\partial x}\right|\sup_{\upsilon_l\in\Upsilon_l}|\hat{X}_{l,t}(\upsilon_l)-X_{l,t}(\upsilon_l)|,$$

where $\frac{\partial\log f_{j,t|t-1}}{\partial x}$ *denotes* $\frac{\partial\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x}$, $\frac{\partial^2\log f_{j,t|t-1}}{\partial x^2}$ *denotes* $\frac{\partial^2\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x^2}$, $\frac{\partial\log f_{j,t|t-1}}{\partial\upsilon_h}$ *denotes* $\frac{\partial\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial\upsilon_h}$, *and* $\frac{\partial^2\log f_{j,t|t-1}}{\partial\upsilon_h\partial x}$ *denotes* $\frac{\partial^2\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial\upsilon_h\partial x}$. *The conclusion thus follows from Lemmata 2.1 and 2.2 in Straumann and Mikosch (2006) since, in addition to the arguments used in the proof of Lemma 9.3, for each* $j\in\{1,...,J\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial\upsilon_h}\right|^{m_j}\right]<\infty$ *for all* $h\in\{1,...,d_j\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial^2\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial x^2}\right|^{m_j}\right]<\infty$, *and* $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j}\sup_{x\in\mathcal{X}_{\Upsilon_j}}\left|\frac{\partial^2\log f_{j,t|t-1}(Y_t;x,\upsilon)}{\partial\upsilon_h\partial x}\right|^{m_j}\right]<\infty$ *for all* $h\in\{1,...,d_j\}$ *by assumption, for each* $j\in\{1,...,J\}$, $\sup_{\upsilon_j\in\Upsilon_j}||\nabla_\theta\hat{X}_{j,t}(\upsilon_j)-\nabla_\theta X_{j,t}(\upsilon_j)||_2\overset{e.a.s.}{\to}0$ *as* $t\to\infty$ *where* $(\nabla_\theta X_{j,t}(\upsilon_j))_{t\in\mathbb{Z}}$ *is stationary for all* $\upsilon_j\in\Upsilon_j$ *and* $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j}||\nabla_\theta X_{j,t}(\upsilon)||_2^{2k_j}\right]<\infty$ *by Assumption 8 and 10, and, for each* $j\in\{1,...,J\}$, $\sup_{\theta\in\Theta}||\nabla_\theta\hat{\pi}_{j,t|t-1}(\theta)-\nabla_\theta\pi_{j,t|t-1}(\theta)||_2\overset{e.a.s.}{\to}0$ *as* $t\to\infty$ *where* $(\nabla_\theta\pi_{j,t|t-1}(\theta))_{t\in\mathbb{Z}}$ *is stationary for all* $\theta\in\Theta$ *and* $\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_\theta\pi_{j,t|t-1}(\theta)||_2^2\right]<\infty$ *by Assumption 11 and 14.*

Lemma 9.7. *Assume that ... Then,*

$$\sqrt{T}\nabla_\theta L_T(\theta_0)\overset{d}{\to}\mathcal{N}\left(0,\mathbb{E}\left[\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0)\nabla_{\theta'}\log f_{t|t-1}(Y_t;\theta_0)\right]\right)\quad as\quad T\to\infty.$$

Proof 7. *First,* $(\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0))_{t\in\mathbb{Z}}$ *is a d-dimensional stationary and ergodic martingale difference sequence since, in addition to the arguments used in the proof of Lemma 9.4, for each* $j\in\{1,...,J\}$, $(\nabla_\theta X_{j,t}(\upsilon_j))_{t\in\mathbb{Z}}$ *is stationary and ergodic for all* $\upsilon_j\in\Upsilon_j$ *by Assumption 8, for each* $j\in\{1,...,J\}$, $(\nabla_\theta\pi_{j,t|t-1}(\theta))_{t\in\mathbb{Z}}$ *is stationary and ergodic for all* $\theta\in\Theta$ *by Assumption 11, and*

$$\mathbb{E}\left[\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0)\mid\mathbf{Y}_{-\infty}^{t-1}\right]=\int_{\mathcal{Y}}\nabla_\theta f_{t|t-1}(y;\theta)\big|_{\theta=\theta_0}dy=\nabla_\theta\int_{\mathcal{Y}}f_{t|t-1}(y;\theta)dy\bigg|_{\theta=\theta_0}=\mathbf{0}$$

because X. Let $\mathbf{c}\in\mathbb{R}^d$ *such that* $\mathbf{c}\neq\mathbf{0}$ *be given. Then,* $(\mathbf{c}'\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0))_{t\in\mathbb{Z}}$ *is trivially a stationary and ergodic martingale difference sequence. It thus follows from the central limit theorem by Billingsley (1961) that if*

$$\mathbb{E}\left[\left(\mathbf{c}'\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0)\right)^2\right]<\infty,$$

then

$$\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\mathbf{c}'\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0)\overset{d}{\to}\mathcal{N}\left(0,\mathbb{E}\left[\left(\mathbf{c}'\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0)\right)^2\right]\right).$$

First, $\mathbb{E}\left[\left(\mathbf{c}'\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0)\right)^2\right]<\infty$ *if* $\mathbb{E}\left[||\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0)||_2^2\right]<\infty$. *We have that*

$$||\nabla_\theta\log f_{t|t-1}(Y_t;\theta_0)||_2^2=\frac{1}{f_{t|t-1}^2(Y_t;\theta_0)}||\nabla_\theta f_{t|t-1}(Y_t;\theta_0)||_2^2$$

where

$$
\begin{aligned}
&||\nabla_\theta f_{t|t-1}(Y_t;\theta_0)||_2^2 \\
&\leq \sum_{i,j=1}^J ||\nabla_\theta \pi_{i,t|t-1}(\theta_0)||_2 f_{i,t|t-1}(Y_t; X_{i,t}(\upsilon_{i,0}),\upsilon_{i,0})||\nabla_\theta \pi_{j,t|t-1}(\theta_0)||_2 f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0}) \\
&+ 2\sum_{i,j=1}^J ||\nabla_\theta \pi_{i,t|t-1}(\theta_0)||_2 f_{i,t|t-1}(Y_t; X_{i,t}(\upsilon_{i,0}),\upsilon_{i,0})\pi_{j,t|t-1}(\theta_0)||\nabla_\theta f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2 \\
&+ \sum_{i,j=1}^J \pi_{i,t|t-1}(\theta_0)||\nabla_\theta f_{i,t|t-1}(Y_t; X_{i,t}(\upsilon_{i,0}),\upsilon_{i,0})||_2 \pi_{j,t|t-1}(\theta_0)||\nabla_\theta f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2,
\end{aligned}
$$

so

$$
\begin{aligned}
&\mathbb{E}\left[||\nabla_\theta \log f_{t|t-1}(Y_t;\theta_0)||_2^2\right] \\
&\leq C\left(\sum_{j=1}^J \left(\mathbb{E}\left[||\nabla_\theta \pi_{j,t|t-1}(\theta_0)||_2^2\right]\right)^{1/2}\right)^2 \\
&+ C\sum_{i=1}^J\sum_{j=1}^J \left(\mathbb{E}\left[||\nabla_\theta \pi_{i,t|t-1}(\theta_0)||_2^2\right]\right)^{1/2}\left(\mathbb{E}\left[||\nabla_\theta \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2^2\right]\right)^{1/2} \\
&+ \left(\sum_{j=1}^J \left(\mathbb{E}\left[||\nabla_\theta \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2^2\right]\right)^{1/2}\right)^2.
\end{aligned}
$$

Moreover, we have that

$$
||\nabla_\theta \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2^2 = \frac{1}{f_{j,t|t-1}^2(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})}||\nabla_\theta f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2^2
$$

where

$$
\begin{aligned}
&||\nabla_\theta f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2^2 \\
&\leq \left|\bar{\nabla}_x f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})\right|^2 ||\nabla_\theta X_{j,t}(\upsilon_{j,0})||_2^2 \\
&+ 2\left|\bar{\nabla}_x f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})\right| ||\nabla_\theta X_{j,t}(\upsilon_{j,0})||_2 ||\bar{\nabla}_\upsilon f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2 \\
&+ ||\bar{\nabla}_\upsilon f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2^2
\end{aligned}
$$

so

$$
\mathbb{E}\left[||\nabla_\theta \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}),\upsilon_{j,0})||_2^2\right]
$$

$$\begin{aligned}
&\leq \left(\mathbb{E}\left[\left|\bar{\nabla}_x \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}), \upsilon_{j,0})\right|^{2k_j/(k_j-1)}\right]\right)^{(k_j-1)/k_j} \left(\mathbb{E}\left[||\nabla_\theta X_{j,t}(\upsilon_{j,0})||_2^{2k_j}\right]\right)^{1/k_j} \\
&+ 2\left(\mathbb{E}\left[\left|\bar{\nabla}_x \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}), \upsilon_{j,0})\right|^{2k_j/(k_j-1)}\right]\right)^{(k_j-1)/2k_j} \left(\mathbb{E}\left[||\nabla_\theta X_{j,t}(\upsilon_{j,0})||_2^{2k_j}\right]\right)^{1/2k_j} \\
&\quad \cdot \left(\mathbb{E}\left[||\bar{\nabla}_\upsilon \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}), \upsilon_{j,0})||_2^2\right]\right)^{1/2} \\
&+ \mathbb{E}\left[||\bar{\nabla}_\upsilon \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_{j,0}), \upsilon_{j,0})||_2^2\right].
\end{aligned}$$

Hence, $\mathbb{E}\left[\left(\mathbf{c}'\nabla_\theta \log f_{t|t-1}(Y_t;\theta_0)\right)^2\right] < \infty$ *since, for each* $j \in \{1, ..., J\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j} ||\nabla_\theta X_{j,t}(\upsilon)||_2^{2k_j}\right] < \infty$ *by Assumption 10, for each* $j \in \{1, ..., J\}$, $\mathbb{E}\left[\sup_{\theta\in\Theta} ||\nabla_\theta \pi_{j,t|t-1}(\theta)||_2^2\right] < \infty$ *by Assumption 14, and, for each* $j \in \{1, ..., J\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j} \left|\bar{\nabla}_x \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)\right|^{2k_j/(k_j-1)}\right] < \infty$ *and* $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j} ||\bar{\nabla}_\upsilon \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)||_2^2\right] < \infty$ *by Assumption 15. Thus,*

$$\mathbf{c}'\sqrt{T}\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta \log f_{t|t-1}(Y_t;\theta_0) \xrightarrow{d} \mathbf{c}'\mathcal{N}\left(0, \mathbb{E}\left[\nabla_\theta \log f_{t|t-1}(Y_t;\theta_0)\nabla_{\theta'} \log f_{t|t-1}(Y_t;\theta_0)\right]\right).$$

An application of the Crámer-Wold theorem concludes the proof.

Lemma 9.8. *Assume that ... Then,*

$$I(\theta_0) = \mathbb{E}\left[\nabla_\theta \log f_{t|t-1}(Y_t;\theta_0)\nabla_{\theta'} \log f_{t|t-1}(Y_t;\theta_0)\right]$$

Proof 8. *We have that*

$$\begin{aligned}
\mathbb{E}\left[\nabla_{\theta\theta} \log f_{t|t-1}(Y_t;\theta_0) \mid \mathbf{Y}_{-\infty}^{t-1}\right] &= -\mathbb{E}\left[\frac{1}{f_{t|t-1}^2(Y_t;\theta_0)}\nabla_\theta f_{t|t-1}(Y_t;\theta_0)\nabla_{\theta'} f_{t|t-1}(Y_t;\theta_0) \mid \mathbf{Y}_{-\infty}^{t-1}\right] \\
&+ \int_{\mathcal{Y}} \nabla_{\theta\theta} f_{t|t-1}(y;\theta)\Big|_{\theta=\theta_0} dy \\
&= -\mathbb{E}\left[\nabla_\theta \log f_{t|t-1}(Y_t;\theta_0)\nabla_{\theta'} \log f_{t|t-1}(Y_t;\theta_0) \mid \mathbf{Y}_{-\infty}^{t-1}\right] \\
&+ \nabla_{\theta\theta}\int_{\mathcal{Y}} f_{t|t-1}(y;\theta)dy\Big|_{\theta=\theta_0} \\
&= -\mathbb{E}\left[\nabla_\theta \log f_{t|t-1}(Y_t;\theta_0)\nabla_{\theta'} \log f_{t|t-1}(Y_t;\theta_0) \mid \mathbf{Y}_{-\infty}^{t-1}\right]
\end{aligned}$$

because X and thus that

$$\mathbb{E}\left[\nabla_{\theta\theta} \log f_{t|t-1}(Y_t;\theta_0)\right] = -\mathbb{E}\left[\nabla_\theta \log f_{t|t-1}(Y_t;\theta_0)\nabla_{\theta'} \log f_{t|t-1}(Y_t;\theta_0)\right].$$

Finally, Condition (vi) follows from Lemmata 9.9 and 9.10 since

$$\sup_{\theta\in\Theta} ||\nabla_{\theta\theta}\hat{L}_T(\theta) - (-I(\theta))||_{2,2} \leq \sup_{\theta\in\Theta} ||\nabla_{\theta\theta}\hat{L}_T(\theta) - \nabla_{\theta\theta}L_T(\theta)||_{2,2} + \sup_{\theta\in\Theta} ||\nabla_{\theta\theta}L_T(\theta) - (-I(\theta))||_{2,2}.$$

Lemma 9.9. *Assume that ... Then,*

$$\sup_{\theta\in\Theta} ||\nabla_{\theta\theta}\hat{L}_T(\theta) - \nabla_{\theta\theta}L_T(\theta)||_{2,2} \overset{a.s.}{\to} 0 \quad as \quad T \to \infty.$$

Proof 9. *We have that*

$$||\nabla_{\theta\theta}\hat{L}_T(\theta) - \nabla_{\theta\theta}L_T(\theta)||_{2,2} \leq \frac{1}{T}\sum_{t=1}^{T} \left|\left|\nabla_\theta \log \hat{f}_{t|t-1}\nabla_{\theta'} \log \hat{f}_{t|t-1} - \nabla_\theta \log f_{t|t-1}\nabla_{\theta'} \log f_{t|t-1}\right|\right|_{2,2}$$
$$+ \frac{1}{T}\sum_{t=1}^{T} \left|\left|\frac{1}{\hat{f}_{t|t-1}}\nabla_{\theta\theta}\hat{f}_{t|t-1} - \frac{1}{f_{t|t-1}}\nabla_{\theta\theta}f_{t|t-1}\right|\right|_{2,2},$$

where $\hat{f}_{t|t-1}$ *denotes* $\hat{f}_{t|t-1}(Y_t;\theta)$ *and* $f_{t|t-1}$ *denotes* $f_{t|t-1}(Y_t;\theta)$*, that*

$$\left|\left|\nabla_\theta \log \hat{f}_{t|t-1}\nabla_{\theta'} \log \hat{f}_{t|t-1} - \nabla_\theta \log f_{t|t-1}\nabla_{\theta'} \log f_{t|t-1}\right|\right|_{2,2}$$
$$\leq \left|\left|\nabla_\theta \log \hat{f}_{t|t-1} - \nabla_\theta \log f_{t|t-1}\right|\right|_2^2 + 2\left|\left|\nabla_\theta \log f_{t|t-1}\right|\right|_2 \left|\left|\nabla_\theta \log \hat{f}_{t|t-1} - \nabla_\theta \log f_{t|t-1}\right|\right|_2$$

since $\mathbf{ab}' - \mathbf{cd}' = (\mathbf{a}-\mathbf{c})(\mathbf{b}'-\mathbf{d}') + (\mathbf{a}-\mathbf{c})\mathbf{d}' + \mathbf{c}(\mathbf{b}'-\mathbf{d}')$ *for all* $\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}\in\mathbb{R}^d$*, and that*

$$\left|\left|\frac{1}{\hat{f}_{t|t-1}}\nabla_{\theta\theta}\hat{f}_{t|t-1} - \frac{1}{f_{t|t-1}}\nabla_{\theta\theta}f_{t|t-1}\right|\right|_{2,2} \leq$$

where .... By using the same arguments as in the proof of 9.3, we have that

$$\sup_{\theta\in\Theta} ||\nabla_{\theta\theta}\hat{L}_T(\theta) - \nabla_{\theta\theta}L_T(\theta)||_{2,2}$$
$$\leq \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\left|\left|\nabla_\theta \log \hat{f}_{t|t-1} - \nabla_\theta \log f_{t|t-1}\right|\right|_2^2$$
$$+ 2\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\left|\left|\nabla_\theta \log f_{t|t-1}\right|\right|_2 \sup_{\theta\in\Theta}\left|\left|\nabla_\theta \log \hat{f}_{t|t-1} - \nabla_\theta \log f_{t|t-1}\right|\right|_2$$

work in progress

Lemma 9.10. *Assume that ... Then,*

$$\sup_{\theta\in\Theta} ||\nabla_{\theta\theta}L_T(\theta) - (-I(\theta))||_{2,2} \overset{a.s.}{\to} 0 \quad as \quad T \to \infty.$$

Proof 10. *First, $(\nabla_{\theta\theta} \log f_{t|t-1}(Y_t;\theta))_{t\in\mathbb{Z}}$ is stationary and ergodic for all $\theta \in \Theta$ since, in addition to the arguments used in the proofs of Lemmata 9.4 and 9.7, for each $j \in \{1, ..., J\}$, $(\nabla_{\theta\theta} X_{j,t}(\upsilon_j))_{t\in\mathbb{Z}}$ is stationary and ergodic for all $\upsilon_j \in \Upsilon_j$ by Assumption 9, and, for each $j \in \{1, ..., J\}$, $(\nabla_{\theta\theta}\pi_{j,t|t-1}(\theta))_{t\in\mathbb{Z}}$ is stationary and ergodic for all $\theta \in \Theta$ by Assumption 13. We have that*

$$||\nabla_{\theta\theta} \log f_{t|t-1}(Y_t;\theta)||_{2,2} \leq \frac{1}{f^2_{t|t-1}(Y_t;\theta)}||\nabla_\theta f_{t|t-1}(Y_t;\theta)||_2^2 + \frac{1}{f_{t|t-1}(Y_t;\theta)}||\nabla_{\theta\theta} f_{t|t-1}(Y_t;\theta)||_{2,2}$$

where

$$\begin{aligned}
&||\nabla_{\theta\theta} f_{t|t-1}(Y_t;\theta)||_{2,2} \\
&\leq \sum_{j=1}^{J} ||\nabla_{\theta\theta}\pi_{j,t|t-1}(\theta)||_{2,2} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j) \\
&+ 2\sum_{j=1}^{J} ||\nabla_\theta \pi_{j,t|t-1}(\theta)||_2 ||\nabla_\theta f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)||_2 \\
&+ \sum_{j=1}^{J} \pi_{j,t|t-1}(\theta)||\nabla_{\theta\theta} f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)||_{2,2},
\end{aligned}$$

so

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\theta\in\Theta} ||\nabla_{\theta\theta} \log f_{t|t-1}(Y_t;\theta)||_{2,2}\right] \\
&\leq \mathbb{E}\left[\sup_{\theta\in\Theta} ||\nabla_\theta \log f_{t|t-1}(Y_t;\theta)||_2^2\right] \\
&+ C\sum_{j=1}^{J} \mathbb{E}\left[\sup_{\theta\in\Theta} ||\nabla_{\theta\theta}\pi_{j,t|t-1}(\theta)||_{2,2}\right] \\
&+ C\sum_{j=1}^{J} \left(\mathbb{E}\left[\sup_{\theta\in\Theta} ||\nabla_\theta \pi_{j,t|t-1}(\theta)||_2^2\right]\right)^{1/2} \left(\mathbb{E}\left[\sup_{\theta\in\Theta} ||\nabla_\theta \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)||_2^2\right]\right)^{1/2} \\
&+ \sum_{j=1}^{J} \mathbb{E}\left[\sup_{\theta\in\Theta} ||\nabla_\theta \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)||_2^2\right] \\
&+ \sum_{j=1}^{J} \mathbb{E}\left[\sup_{\theta\in\Theta} ||\nabla_{\theta\theta} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)||_{2,2}\right].
\end{aligned}$$

Moreover, we have that

$$||\nabla_{\theta\theta} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)||_{2,2} \leq \frac{1}{f^2_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)}||\nabla_\theta f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon_j), \upsilon_j)||_2^2$$

$$+\frac{1}{f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)}||\nabla_{\theta\theta}f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_{2,2}$$

where

$$\begin{aligned}||\nabla_{\theta\theta}f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_{2,2} &\leq |\bar{\nabla}_{xx}f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|||\nabla_\theta X_{j,t}(\upsilon_j)||_2^2\\ &+|\bar{\nabla}_x f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|||\nabla_{\theta\theta}X_{j,t}(\upsilon_j)||_{2,2}\\ &+2||\bar{\nabla}_{\upsilon x}f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_2||\nabla_\theta X_{j,t}(\upsilon_j)||_2\\ &+||\bar{\nabla}_{\upsilon\upsilon}f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_{2,2},\end{aligned}$$

so

$$\begin{aligned}&\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_{\theta\theta}\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_{2,2}\right]\\ &\leq\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_\theta\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_2^2\right]\\ &+\left(\mathbb{E}\left[\sup_{\theta\in\Theta}|\bar{\nabla}_x\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|^{2k_j/(k_j-1)}\right]\right)^{(k_j-1)/k_j}\left(\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_\theta X_{j,t}(\upsilon_j)||_2^{2k_j}\right]\right)^{1/k_j}\\ &+\left(\mathbb{E}\left[\sup_{\theta\in\Theta}|\bar{\nabla}_{xx}\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|^{k_j/(k_j-1)}\right]\right)^{(k_j-1)/k_j}\left(\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_\theta X_{j,t}(\upsilon_j)||_2^{2k_j}\right]\right)^{1/k_j}\\ &+\left(\mathbb{E}\left[\sup_{\theta\in\Theta}|\bar{\nabla}_x\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|^{k_j/(k_j-1)}\right]\right)^{(k_j-1)/k_j}\left(\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_{\theta\theta}X_{j,t}(\upsilon_j)||_{2,2}^{k_j}\right]\right)^{1/k_j}\\ &+2\left(\mathbb{E}\left[\sup_{\theta\in\Theta}|\bar{\nabla}_x\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)|^{2k_j/(k_j-1)}\right]\right)^{(k_j-1)/2k_j}\\ &\quad\cdot\left(\mathbb{E}\left[\sup_{\theta\in\Theta}||\bar{\nabla}_\upsilon\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_2^2\right]\right)^{1/2}\left(\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_\theta X_{j,t}(\upsilon_j)||_2^{2k_j}\right]\right)^{1/2k_j}\\ &+2\left(\mathbb{E}\left[\sup_{\theta\in\Theta}||\bar{\nabla}_{\upsilon x}\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_2^{k_j/(k_j-1)}\right]\right)^{(k_j-1)/k_j}\left(\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_\theta X_{j,t}(\upsilon_j)||_2^{k_j}\right]\right)^{1/k_j}\\ &+\mathbb{E}\left[\sup_{\theta\in\Theta}||\bar{\nabla}_\upsilon\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_2^2\right]\\ &+\mathbb{E}\left[\sup_{\theta\in\Theta}||\bar{\nabla}_{\upsilon\upsilon}\log f_{j,t|t-1}(Y_t;X_{j,t}(\upsilon_j),\upsilon_j)||_{2,2}\right].\end{aligned}$$

Hence, $\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_{\theta\theta}\log f_{t|t-1}(Y_t;\theta)||_{2,2}\right]<\infty$ *since, in addition to the arguments used in the proof of Lemma 9.7, for each* $j\in\{1,...,J\}$, $\mathbb{E}\left[\sup_{\upsilon\in\Upsilon_j}||\nabla_{\theta\theta}X_{j,t}(\upsilon)||_{2,2}^{k_j}\right]<\infty$ *by Assumption 10, for each* $j\in\{1,...,J\}$, $\mathbb{E}\left[\sup_{\theta\in\Theta}||\nabla_{\theta\theta}\pi_{j,t|t-1}(\theta)||_{2,2}\right]<\infty$ *by Assumption*

14, and, for each $j \in \{1, ..., J\}$, $\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} |\bar{\nabla}_{xx} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)|^{k_j/(k_j-1)}\right] < \infty$, $\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} ||\bar{\nabla}_{\upsilon x} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)||_2^{k_j/(k_j-1)}\right] < \infty$, and $\mathbb{E}\left[\sup_{\upsilon \in \Upsilon_j} ||\bar{\nabla}_{\upsilon\upsilon} \log f_{j,t|t-1}(Y_t; X_{j,t}(\upsilon), \upsilon)||_{2,2}\right] < \infty$ by Assumption 15. The conclusion thus follows from the uniform law of large numbers by Rao (1962).

Appendix A. Technical Lemmas

Let $C \subset \mathbb{R}^d$ be a compact set. Let $(Y_t(c))_{t\in\mathbb{Z}}$ be a stochastic process taking values in $\mathcal{Y}_c \subseteq \mathbb{R}^{d'}$ equipped with $||Y(c)||_2 = (\sum_{i=1}^{d'} |Y_i(c)|^2)^{1/2}$ given by

$$Y_{t+1}(c) = \phi_t(Y_t(c); c)$$

for all $c \in C$, where $(\phi_t(\cdot; c))_{t\in\mathbb{Z}}$ is a sequence of stationary and ergodic Lipschitz functions for all $c \in C$ with Lipschitz coefficient

$$\Lambda(\phi_t; c) := \sup_{\substack{x,y\in\mathcal{Y}_c \\ x\neq y}} \frac{||\phi_t(x; c) - \phi_t(y; c)||_2}{||x - y||_2}$$

for all $c \in C$.

Let $\phi_t(Y_t(c); c) = \phi(Y_t(c), Z_t; c)$ where Z_t takes values in $\mathcal{Z}$. If $(y, c) \mapsto \phi(y, z; c)$ is continuous on $\mathcal{Y}_c \times C$ for all $z \in \mathcal{Z}$, then $(Y_t)_{t\in\mathbb{Z}}$ is a stochastic process taking values in $\mathbb{C}(C, \mathcal{Y}_C), \mathcal{Y}_C = \cup_{c\in C}\mathcal{Y}_c$, that is, the space of continuous functions from C to $\mathcal{Y}_C$, equipped with $||Y||_C = \sup_{c\in C} ||Y(c)||_2$ given by

$$Y_{t+1} = \Phi_t(Y_t)$$

with

$$\Phi_t(Y) = \phi_t(Y(\cdot); \cdot),$$

where $(\Phi_t(\cdot))_{t\in\mathbb{Z}}$ is a sequence of stationary and ergodic Lipschitz maps with Lipschitz coefficient

$$\Lambda(\Phi_t) := \sup_{\substack{X,Y\in\mathbb{C}(C,\mathcal{Y}_C) \\ X\neq Y}} \frac{||\Phi_t(X) - \Phi_t(Y)||_C}{||X - Y||_C}.$$

Note that $(\mathbb{C}(C, \mathcal{Y}_C), ||\cdot||_C)$ is a separable Banach space.

Finally, let

$$\bar{\Lambda}(\phi_t; c) := \sup_{\substack{x,y\in\mathcal{Y}_C \\ x\neq y}} \frac{||\phi_t(x; c) - \phi_t(y; c)||_2}{||x - y||_2}.$$

Lemma Appendix A.1. *Assume that*

(A) there exists a $y \in \mathcal{Y}_C$ *such that* $\mathbb{E}[\log^+ \sup_{c\in C} ||\phi_t(y; c) - y||_2] < \infty$*,*
(B) $\mathbb{E}[\log^+ \sup_{c\in C} \bar{\Lambda}(\phi_t; c)] < \infty$*, and*
(C) there exists an $r \in \mathbb{N}$ *such that* $\mathbb{E}[\log \sup_{c\in C} \bar{\Lambda}(\phi_t^{(r)}; c)] < 0$*.*

Then, the stochastic process $(Y_t)_{t\in\mathbb{Z}}$ is stationary and ergodic and

$$Y_t = \lim_{m\to\infty} \Phi_{t-1} \circ \cdots \circ \Phi_{t-m}(Y) \quad a.s,$$

for all $Y \in \mathbb{C}(C, \mathcal{Y}_C)$.

Let $(X_t)_{t\in\mathbb{N}}$ be another stochastic process taking values in $\mathbb{C}(C, \mathcal{Y}_C)$ equipped with $||X||_C = \sup_{c\in C} ||X(c)||_2$ given by

$$X_{t+1} = \Phi_t(X_t)$$

with

$$\Phi_t(X) = \phi_t(X(\cdot); \cdot),$$

where $(\Phi_t(\cdot))_{t\in\mathbb{N}}$ is a sequence of stationary and ergodic Lipschitz maps. Then,

$$||X_t - Y_t||_C \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty.$$

Proof 11. *The conclusions follow from Theorem 3.1 in Bougerol (1993) (see also Theorem 2.8 in Straumann and Mikosch (2006)) if*

(a) there exists a $Y \in \mathbb{C}(C, \mathcal{Y}_C)$ such that $\mathbb{E}[\log^+ ||\Phi_t(Y) - Y||_C] < \infty$,
(b) $\mathbb{E}[\log^+ \Lambda(\Phi_t)] < \infty$, and
(c) there exists an $r \in \mathbb{N}$ such that $\mathbb{E}[\log \Lambda(\Phi_t^{(r)})] < 0$.

(a): Let $Y \in \mathbb{C}(C, \mathcal{Y}_C)$ be given by $Y(c) = y$. Then,

$$\mathbb{E}[\log^+ ||\Phi_t(Y) - Y||_C] = \mathbb{E}[\log^+ \sup_{c\in C} ||\phi_t(y; c) - y||_2] < \infty$$

by (A).
We have that

$$\begin{aligned}
\Lambda(\Phi_t) &= \sup_{\substack{X,Y\in\mathbb{C}(C,\mathcal{Y}_C)\\ X\neq Y}} \frac{||\Phi_t(X) - \Phi_t(Y)||_C}{||X - Y||_C} \\
&= \sup_{\substack{X,Y\in\mathbb{C}(C,\mathcal{Y}_C)\\ X\neq Y}} \sup_{c\in C} \frac{||\phi_t(X(c); c) - \phi_t(Y(c); c)||_2}{||X(c) - Y(c)||_2} \frac{||X(c) - Y(c)||_2}{\sup_{c\in C} ||X(c) - Y(c)||_2} \\
&\leq \sup_{\substack{X,Y\in\mathbb{C}(C,\mathcal{Y}_C)\\ X\neq Y}} \sup_{c\in C} \frac{||\phi_t(X(c); c) - \phi_t(Y(c); c)||_2}{||X(c) - Y(c)||_2} \\
&\leq \sup_{\substack{X,Y\in\mathbb{C}(C,\mathcal{Y}_C)\\ X\neq Y}} \sup_{c\in C} \sup_{\substack{x,y\in\mathcal{Y}_C\\ x\neq y}} \frac{||\phi_t(x; c) - \phi_t(y; c)||_2}{||x - y||_2}
\end{aligned}$$

$$= \sup_{c \in C} \sup_{\substack{x,y \in \mathcal{Y}_C \\ x \neq y}} \frac{||\phi_t(x;c) - \phi_t(y;c)||_2}{||x-y||_2} = \sup_{c \in C} \bar{\Lambda}(\phi_t; c).$$

(b): Then,

$$\mathbb{E}[\log^+ \Lambda(\Phi_t)] \leq \mathbb{E}[\log^+ \sup_{c \in C} \bar{\Lambda}(\phi_t; c)] < \infty$$

by (B).
(c): Moreover, there exists an $r \in \mathbb{N}$ such that

$$\mathbb{E}[\log \Lambda(\Phi_t^{(r)})] \leq \mathbb{E}[\log \sup_{c \in C} \bar{\Lambda}(\phi_t^{(r)}; c)] < 0$$

by (C).

Lemma Appendix A.2. *Assume that*

(A) there exists a $y \in \mathcal{Y}_C$ such that $\mathbb{E}[\log^+ \sup_{c \in C} ||\phi_t(y;c) - y||_2] < \infty$,
(B) $\mathbb{E}[\log^+ \sup_{c \in C} \bar{\Lambda}(\phi_t; c)] < \infty$, and
(C) there exists an $r \in \mathbb{N}$ such that $\mathbb{E}[\log \sup_{c \in C} \bar{\Lambda}(\phi_t^{(r)}; c)] < 0$.

Then, the stochastic process $(Y_t)_{t \in \mathbb{Z}}$ is stationary and ergodic and

$$Y_t = \lim_{m \to \infty} \Phi_{t-1} \circ \cdots \circ \Phi_{t-m}(Y) \quad a.s.$$

for all $Y \in \mathbb{C}(C, \mathcal{Y}_C)$.

Assume that $\mathbb{E}[\log^+ ||Y_t||_C] < \infty$. Let $(X_t)_{t \in \mathbb{N}}$ be another stochastic process taking values in $\mathbb{C}(C, \mathcal{Y}_C)$ equipped with $||X||_C = \sup_{c \in C} ||X(c)||_2$ given by

$$X_{t+1} = \hat{\Phi}_t(X_t)$$

with

$$\hat{\Phi}_t(X) = \hat{\phi}_t(X(\cdot); \cdot),$$

where $(\hat{\Phi}_t(\cdot))_{t \in \mathbb{N}}$ is a sequence of Lipschitz maps. Assume that

(D) there exists a $y \in \mathcal{Y}_C$ such that $\sup_{c \in C} ||\hat{\phi}_t(y;c) - \phi_t(y;c)||_2 \overset{e.a.s.}{\to} 0$ as $t \to \infty$, and
(E) $\sup_{c \in C} \bar{\Lambda}(\hat{\phi}_t - \phi_t; c) \overset{e.a.s.}{\to} 0$ as $t \to \infty$.

Then,

$$||X_t - Y_t||_C \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty.$$

Proof 12. *The first conclusion follows from Lemma Appendix A.1. The second one follows from Theorem 2.10 in Straumann and Mikosch (2006) if*

(d) there exists a $Y \in \mathbb{C}(C, \mathcal{Y}_C)$ *such that* $||\hat{\Phi}_t(Y) - \Phi_t(Y)||_C \overset{e.a.s.}{\to} 0$ *as* $t \to \infty$*, and*
(e) $\Lambda(\hat{\Phi}_t - \Phi_t) \overset{e.a.s.}{\to} 0$ *as* $t \to \infty$*.*

(d): Let $Y \in \mathbb{C}(C, \mathcal{Y}_C)$ *be given by* $Y(c) = y$*. Then,*

$$||\hat{\Phi}_t(Y) - \Phi_t(Y)||_C = \sup_{c \in C} ||\hat{\phi}_t(y; c) - \phi_t(y; c)||_2 \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty$$

by (D).
As in the proof of Lemma Appendix A.1, we have that

$$\begin{aligned}
\Lambda(\hat{\Phi}_t - \Phi_t) &= \sup_{\substack{X,Y \in \mathbb{C}(C,\mathcal{Y}_C) \\ X \neq Y}} \frac{||(\hat{\Phi}_t(X) - \Phi_t(X)) - (\hat{\Phi}_t(Y) - \Phi_t(Y))||_C}{||X - Y||_C} \\
&= \sup_{\substack{X,Y \in \mathbb{C}(C,\mathcal{Y}_C) \\ X \neq Y}} \sup_{c \in C} \frac{||(\hat{\phi}_t(X(c); c) - \phi_t(X(c); c)) - (\hat{\phi}_t(Y(c); c) - \phi_t(Y(c); c))||_2}{||X(c) - Y(c)||_2} \frac{||X(c) - Y(c)||_2}{\sup_{c \in C} ||X(c) - Y(c)||_2} \\
&\leq \sup_{\substack{X,Y \in \mathbb{C}(C,\mathcal{Y}_C) \\ X \neq Y}} \sup_{c \in C} \frac{||(\hat{\phi}_t(X(c); c) - \phi_t(X(c); c)) - (\hat{\phi}_t(Y(c); c) - \phi_t(Y(c); c))||_2}{||X(c) - Y(c)||_2} \\
&\leq \sup_{\substack{X,Y \in \mathbb{C}(C,\mathcal{Y}_C) \\ X \neq Y}} \sup_{c \in C} \sup_{\substack{x,y \in \mathcal{Y}_C \\ x \neq y}} \frac{||(\hat{\phi}_t(x; c) - \phi_t(x; c)) - (\hat{\phi}_t(y; c) - \phi_t(y; c))||_2}{||x - y||_2} \\
&= \sup_{c \in C} \sup_{\substack{x,y \in \mathcal{Y}_C \\ x \neq y}} \frac{||(\hat{\phi}_t(x; c) - \phi_t(x; c)) - (\hat{\phi}_t(y; c) - \phi_t(y; c))||_2}{||x - y||_2} = \sup_{c \in C} \bar{\Lambda}(\hat{\phi}_t - \phi_t; c).
\end{aligned}$$

(e): Then,

$$\Lambda(\hat{\Phi}_t - \Phi_t) \leq \sup_{c \in C} \bar{\Lambda}(\hat{\phi}_t - \phi_t; c) \overset{e.a.s.}{\to} 0 \quad as \quad t \to \infty$$

by (E).

Appendix B. Other Lemmas

A $J \times 1$-dimensional vector $\mathbf{x}$ is called stochastic if $x_j \geq 0$ for all $j \in \{1, ..., J\}$ and $\sum_{j=1}^{J} x_j = 1$, and a $J \times J$-dimensional matrix $\mathbf{X}$ is called stochastic if $x_{ij} \geq 0$ for all $i, j \in \{1, ..., J\}$ and $\sum_{j=1}^{J} x_{ij} = 1$ for all $i \in \{1, ..., J\}$. The space of $J \times 1$-dimensional stochastic vectors is denoted by $\mathcal{S}$.

Lemma Appendix B.1. *Let $\mathbf{P}$ be a $J \times J$-dimensional stochastic matrix with generic element p_{ij}. Assume that there exist an $\varepsilon \in (0, 1)$ and a $\boldsymbol{\rho} \in \mathcal{S}$ such that*

$$p_{ij} \geq \varepsilon \rho_j$$

for all $i, j \in \{1, ..., J\}$. Then, there exists an $\alpha \in (0, 1)$ such that

$$||\mathbf{P}'\mathbf{x} - \mathbf{P}'\mathbf{y}||_1 \leq \alpha ||\mathbf{x} - \mathbf{y}||_1$$

for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ with $\mathbf{x} \neq \mathbf{y}$.

Proof 13. *Let $\tilde{\mathbf{P}}$ be a $J \times J$-dimensional matrix with generic element $\tilde{p}_{ij} = (1 - \varepsilon)^{-1}(p_{ij} - \varepsilon \rho_j)$. Note that $\tilde{\mathbf{P}}$ is a stochastic matrix. Thus,*

$$\begin{aligned}
||\mathbf{P}'\mathbf{x} - \mathbf{P}'\mathbf{y}||_1 &= \sum_{j=1}^{J} \left| \sum_{i=1}^{J} p_{ij}(x_i - y_i) \right| \\
&= (1 - \varepsilon) \sum_{j=1}^{J} \left| \sum_{i=1}^{J} \tilde{p}_{ij}(x_i - y_i) \right| \\
&\leq (1 - \varepsilon) \sum_{j=1}^{J} \sum_{i=1}^{J} \tilde{p}_{ij} |x_i - y_i| \\
&= (1 - \varepsilon) \sum_{i=1}^{J} |x_i - y_i| = (1 - \varepsilon) ||\mathbf{x} - \mathbf{y}||_1.
\end{aligned}$$

References

Aknouche, A. and Francq, C.: 2022, Stationarity and ergodicity of markov switching positive conditional mean models, *Journal of Time Series Analysis* **43**(3), 436–459.

Baum, L. E. and Petrie, T.: 1966, Statistical inference for probabilistic functions of finite state markov chains, *The annals of mathematical statistics* **37**(6), 1554–1563.

Bickel, P. J., Ritov, Y. and Ryden, T.: 1998, Asymptotic normality of the maximum-likelihood estimator for general hidden markov models, *The Annals of Statistics* **26**(4), 1614–1635.

Billingsley, P.: 1961, The lindeberg-levy theorem for martingales, *Proceedings of the American Mathematical Society* **12**(5), 788–792.

Blasques, F., Gorgi, P., Koopman, S. J. and Wintenberger, O.: 2018, Feasible invertibility conditions and maximum likelihood estimation for observation-driven models.

Blasques, F., van Brummelen, J., Koopman, S. J. and Lucas, A.: 2022, Maximum likelihood estimation for score-driven models, *Journal of Econometrics* **227**(2), 325–346.

Bollerslev, T.: 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of econometrics* **31**(3), 307–327.

Bougerol, P.: 1993, Kalman filtering with random coefficients and contractions, *SIAM Journal on Control and Optimization* **31**(4), 942–959.

Bulla, J., Lagona, F., Maruotti, A. and Picone, M.: 2012, A multivariate hidden markov model for the identification of sea regimes from incomplete skewed and circular time series, *Journal of Agricultural, Biological, and Environmental Statistics* **17**, 544–567.

Cai, J.: 1994, A markov model of switching-regime arch, *Journal of Business & Economic Statistics* **12**(3), 309–316.

Churchill, G. A.: 1989, Stochastic models for heterogeneous dna sequences, *Bulletin of mathematical biology* **51**(1), 79–94.

Creal, D., Koopman, S. J. and Lucas, A.: 2013, Generalized autoregressive score models with applications, *Journal of Applied Econometrics* **28**(5), 777–795.

Douc, R., Fokianos, K. and Moulines, E.: 2017, Asymptotic properties of quasi-maximum likelihood estimators in observation-driven time series models.

Douc, R., Moulines, E. and Rydén, T.: 2004, Asymptotic properties of the maximum likelihood estimator in autoregressive models with markov regime.

Engle, R. F.: 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica: Journal of the econometric society* pp. 987–1007.

Francq, C. and Roussignol, M.: 1998, Ergodicity of autoregressive processes with markov-switching and consistency of the maximum-likelihood estimator, *Statistics: A Journal of Theoretical and Applied Statistics* **32**(2), 151–173.

Francq, C., Roussignol, M. and Zakoian, J.-M.: 2001, Conditional heteroskedasticity driven by hidden markov chains, *Journal of Time Series Analysis* **22**(2), 197–220.

Francq, C. and Zakoıan, J.-M.: 2001, Stationarity of multivariate markov–switching arma models, *Journal of Econometrics* **102**(2), 339–364.

Gray, S. F.: 1996, Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* **42**(1), 27–62.

Haas, M., Mittnik, S. and Paolella, M. S.: 2004a, Mixed normal conditional heteroskedasticity, *Journal of financial Econometrics* **2**(2), 211–250.

Haas, M., Mittnik, S. and Paolella, M. S.: 2004b, A new approach to markov-switching garch models, *Journal of financial Econometrics* **2**(4), 493–530.

Hamilton, J. D.: 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica: Journal of the econometric society* pp. 357–384.

Hamilton, J. D. and Susmel, R.: 1994, Autoregressive conditional heteroskedasticity and changes in regime, *Journal of econometrics* **64**(1-2), 307–333.

Harvey, A. C.: 2013, *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Vol. 52, Cambridge University Press.

Juang, B. H. and Rabiner, L. R.: 1991, Hidden markov models for speech recognition, *Technometrics* **33**(3), 251–272.

Kasahara, H. and Shimotsu, K.: 2019, Asymptotic properties of the maximum likelihood estimator in regime switching econometric models, *Journal of econometrics* **208**(2), 442–467.

Krishnamurthy, V. and Ryden, T.: 1998, Consistent estimation of linear and non-linear autoregressive models with markov regime, *Journal of time series analysis* **19**(3), 291–307.

Krolzig, H.-M.: 2013, *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*, Vol. 454, Springer Science & Business Media.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D. and Morales, J. M.: 2012, Flexible and practical modeling of animal telemetry data: hidden markov models and extensions, *Ecology* **93**(11), 2336–2342.

Leroux, B. G.: 1992, Maximum-likelihood estimation for hidden markov models, *Stochastic processes and their applications* **40**(1), 127–143.

Pötscher, B. M. and Prucha, I.: 1997, *Dynamic nonlinear econometric models: Asymptotic theory*, Springer Science & Business Media.

Rao, R. R.: 1962, Relations between weak and uniform convergence of measures with applications, *The Annals of Mathematical Statistics* pp. 659–680.

Straumann, D. and Mikosch, T.: 2006, Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach.

Zucchini, W. and Guttorp, P.: 1991, A hidden markov model for space-time precipitation, *Water Resources Research* **27**(8), 1917–1923.

Zucchini, W., MacDonald, I. L. and Langrock, R.: 2016, *Hidden Markov models for time series: an introduction using R*, Chapman and Hall/CRC.