Robert J. Elliott^{1,2} / Tak Kuen Siu³ / John W. Lau⁴

A hidden Markov regime-switching smooth transition model

¹ School of Commerce, University of South Australia, Australia

² Haskayne School of Business, University of Calgary, Calgary, Alberta, Canada

³ Department of Applied Finance and Actuarial Studies, Faculty of Business and Economics, Macquarie University, Sydney, Australia, Phone: (+61-2) 9850 8589; Fax: (+61-2) 9850 9481, E-mail: Ken.Siu@mq.edu.au

⁴ Department of Mathematics and Statistics, University of Western Australia, Perth, Australia

Abstract:

In this paper, we develop a new class of parametric nonlinear time series models by combining two important classes of models, namely smooth transition models and hidden Markov regime-switching models. The class of models is general and flexible enough to incorporate two types of switching behavior: smooth state transitions and abrupt changes in hidden states. The estimation of the hidden states and model parameters is performed by applying filtering theory and a filter-based expectation-maximization (EM) algorithm. Applications of the model are illustrated using simulated data and real financial data. Other potential applications are mentioned.

Keywords: filtering, Laplace series expansion, nonlinear time series, regime switching model, smooth transition model

DOI: 10.1515/snde-2016-0061

1 Introduction

The objective of this paper is to develop a new class of parametric nonlinear time series models which is general and flexible enough to incorporate two types of switching behavior, namely smooth state transitions and abrupt changes in hidden states. We call our class of nonlinear time series models hidden Markov regime-switching smooth transition (HMRS-STAR) models. It is developed by combining two important classes of parametric nonlinear time series models, namely smooth transition (STAR) models and hidden Markov regime-switching (HMRS) models. The reason behind combining the two models to develop the class of HMRS-STAR models is similar to that of combining first-generation models to produce second-generation models, third-generation models, and so on, as proposed in Tong (1990), (see Chapter 3, Section 13.3, therein). The class of HMRS-STAR models proposed here is general and may have potential applications in diverse fields such as economics, finance, actuarial science, biology, engineering and physical sciences, ecology and population dynamics, where nonlinear time series modelling or nonlinear dynamical systems may be highly relevant. Concrete applications of the HMRS-STAR models to financial modeling are presented below. Particularly, the class of HMRS-STAR models can describe the impacts of both abrupt regime switches and smooth regime switches on asset price dynamics or returns. Abrupt regime switches are those sudden changes in asset price dynamics or returns which may be attributed to major economic or market events such as financial crises. Some observed market news or events may be digested more slowly by market participants and their impacts may be more slowly reflected in asset prices or returns. These changes in asset prices or returns are gradual rather than abrupt and are referred to as smooth regime switches. Smooth state transitions in asset prices or returns may also have some implications on market efficiency and may be related to a price trend model in the financial econometrics literature.

Standard filtering theory based on the reference probability approach in the literature and its corresponding filter-based expectation-maximization (EM) algorithm, [see, for example, Elliott, Aggoun, and Moore (1995)], is applied to estimate the hidden states and unknown parameters in the proposed model. To simplify some nonlinear relationships in the filter-based based estimation due to the nonlinearity in the smoothing function, we employ a Laplace series expansion to simplify the estimation procedure and derive analytical formulas for some parameter estimates. Simulated data are used to illustrate the practical implementation of the estimation method and its accuracy. Real data examples based on two financial data sets, namely the Hang Seng Index and NASDAQ Composite index, are provided to illustrate potential applications of the proposed class of models to real financial data. Other potential applications of the proposed class of models are briefly discussed. In

Tak Kuen Siu is the corresponding author.

©2018 Walter de Gruyter GmbH, Berlin/Boston.

the course of our real data analysis, we have found that for some financial datasets, the estimation algorithm based on filtering theory does not converge. This reflects that there is still room for improving the filter-based estimation method. This may pose a challenging problem for further research. Given the important role played by hidden Markov models in signal processing, it is hoped that the proposed class of HMRS-STAR models may provide some insights into how the link between two important fields, namely nonlinear time series analysis and signal processing, may be consolidated. In what follows, some of the relevant literature and background of the proposed class of models are discussed.

Nonlinear time series analysis is an important topic in econometrics, statistics and dynamical systems theory. Some early developments of nonlinear time series models may be traced back to the late 1970s where limitations of linear time series modelling to fit real datasets in diverse fields were realised. See the monographs by Tong (1983, 1990). Some important classes of parametric nonlinear time series models proposed in the early stage of developments of the discipline are, for example, threshold autoregressive (TAR) models by Tong (1977, 1978) and Tong and Lim (1980), bilinear models by Granger and Anderson (1978) [see also Subba Rao and Gabr (1984)], autoregressive conditional heteroscedastic (ARCH) models by Engle (1982), stochastic volatility models of Taylor (1982) and state-dependent models by Priestley (1980). The monograph by Tong (1990) provides an excellent and authoritative account of parametric nonlinear time series models developed before the 1990s. Nonparametric and semi-parametric approaches to nonlinear time series analysis have also been studied in the literature. See the monographs by Fan and Yao (2003) and Gao (2007).

Regime-switching models represent an important class of parametric nonlinear time series models which have wide ranging applications in diverse fields such as engineering, economics, finance and actuarial science, amongst others. The basic principle of regime switching was conceived some time ago in economics, particularly in econometrics, and engineering, particularly in signal processing and control engineering. For discussions on applications of regime switching models and related models to engineering, one may refer to the monographs by Elliott, Aggoun, and Moore (1995) and Yin and Zhu (2010). Early works in econometrics where the idea of regime switching has been adopted are, for example, Quandt (1958) and Goldfeld and Quandt (1973), where regression models with switching parameters were introduced to discuss nonlinearity in economic data. The principle of regime switching also appeared in pioneering works on parametric nonlinear time series analysis, [see, for example, Tong and Lim (1980) and Tong (1983)]. Since the seminal article by Hamilton (1989) on Markov regime-switching autoregressive models for econometrics, regime switching models have become popular in economics, econometrics and finance. The basic idea of regime switching models is that the model parameters can change over time according to an underlying state process which could be a finite-state hidden Markov chain.

Chan and Tong (1986) and Teräsvirta (1994) introduced a class of time series models, namely, smooth transition autoregressive models. This class of time series models allow “smooth”, or gradual transitions, in regimes and may be thought of as a type of regime switching models in a wide sense, see, for example, Tong (1990), for further discussions. The key feature of this class of models is that “smooth” regime switching is described by introducing a “smooth” transition function, which is a non-decreasing function. Some typical examples of this “smooth” transition function are the cumulative normal distribution function and the logistic function. Like regime-switching autoregressive models, smooth transition autoregressive models can capture cyclical behavior in time series. For a survey on the smooth transition autoregressive models and their applications, interested readers may refer to Granger and Teräsvirta (1993), Potter (1999), and Teräsvirta (1998).

In Elliott, Siu, and Lau (2013), a double threshold model was considered, where two types of regime-switching, namely the regime switching governed by the threshold principle of Tong (1983, 1990) and the regime switching described by transitions of a hidden Markov chain. In Elliott, Liew, and Siu (2011), the filtering of a threshold stochastic volatility model was considered using the reference probability approach for hidden Markov models in, for example, Elliott, Aggoun, and Moore (1995). In a recent paper by Zhu et al. (2017), a hidden Markov models with threshold effects was considered and applied to oil price forecasting, where the threshold regime switching effect was present in transition probabilities of the hidden Markov chain. Note that the smooth transitions models may be more general than the threshold regime-switching models and may be used to test whether regime switching is attributed to smooth transitions or threshold regime switches, [see, for example, Tong (1990), for related discussions]. In this sense, the HMRS-STAR model may be more general than the double threshold model in Elliott, Siu, and Lau (2013) and the former may also be used to generalize threshold-type models in, for example, Elliott, Liew, and Siu (2011) and Zhu et al. (2017).

This paper is organized as follows. The next section gives an introduction to the HMRS-STAR model. Filters for the hidden Markov chain and filter-based estimates of the unknown parameters based on the EM algorithm are presented in Section 3. Simulation experiments and results are discussed in Section 4. Real data examples are presented in Section 5. Section 6 provides discussions on some potential economic applications of the proposed HMRS-STAR model. The final section gives some concluding remarks.

2 The HMRS-STAR model

We consider a complete probability space $(\Omega, \mathcal{F}, P)$ on which all random variables are defined, where P is a real-world probability. Let $\{\mathbf{X}_t | t \in \mathcal{T}\}$ be a discrete-time, N -state, hidden Markov chain defined on $(\Omega, \mathcal{F}, P)$ with state space being the set of unit vectors $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$. Here $\mathcal{T} := \{0, 1, \dots\}$ and the j^{th} component of $\mathbf{e}_i$ is the Kronecker delta function δ_{ij} for each $i, j = 1, 2, \dots, N$. Indeed the state space we considered here is called the canonical state space of the hidden Markov chain. The state space was adopted in, for example, Elliott, Aggoun, and Moore (1995).

Suppose that the chain $\mathbf{X}$ is time-homogeneous so that its probability law is completely determined by its transition probability and initial distribution. For each $i, j = 1, 2, \dots, N$, let

$$\begin{aligned}\pi_{ji} &:= P(\mathbf{X}_{t+1} = \mathbf{e}_j | \mathbf{X}_t = \mathbf{e}_i) \\ &= P(\mathbf{X}_1 = \mathbf{e}_j | \mathbf{X}_0 = \mathbf{e}_i),\end{aligned}$$

so $[\pi_{ji}]_{i,j=1,2,\dots,N}$ is the transition probability matrix of the chain $\mathbf{X}$ under P and it is denoted by Π .

Consider the P -completion of the natural filtration $\mathbb{F}^{\mathbf{X}} := \{\mathcal{F}_t^{\mathbf{X}} | t \in \mathcal{T}\}$ generated by the hidden Markov chain, where

$$\mathcal{F}_t^{\mathbf{X}} := \sigma\{\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_t\} \vee \mathcal{N},$$

the minimal σ -algebra generated by information about the values of the chain $\mathbf{X}$ up to and including time t and the collection $\mathcal{N}$ of P -null sets of $\mathcal{F}$. Here for σ -algebras $\mathcal{A}$ and $\mathcal{B}$, we denote by $\mathcal{A} \vee \mathcal{B}$ the minimal σ -algebra generated by $\mathcal{A}$ and $\mathcal{B}$.

With the canonical state space of the chain, Elliott, Aggoun, and Moore (1995) gave the following semi-martingale dynamics for the chain $\mathbf{X}$ under P :

$$\mathbf{X}_{t+1} = \Pi \mathbf{X}_t + \mathbf{M}_{t+1}, \quad t \in \mathcal{T}, \quad (1)$$

where $\{\mathbf{M}_t | t \in \mathcal{T} \setminus \{0\}\}$ is an $\mathbb{R}^N$ -valued, $(\mathbb{F}^{\mathbf{X}}, P)$ -martingale difference process. The above semimartingale dynamics will be used when deriving filters for the hidden Markov chain and filter-based estimates of the unknown parameters in the HMRS-STAR model.

We now describe the HMRS-STAR model. For illustration, we consider a simple case where there are two regimes in the smooth transition part of the model. Let

$$\begin{aligned}\mu(t) &:= \langle \boldsymbol{\mu}, \mathbf{X}_t \rangle, \\ \theta(t) &:= \langle \boldsymbol{\theta}, \mathbf{X}_t \rangle, \\ \sigma(t) &:= \langle \boldsymbol{\sigma}, \mathbf{X}_t \rangle.\end{aligned}$$

Here $\boldsymbol{\mu} := (\mu_1, \mu_2, \dots, \mu_N)' \in \mathbb{R}^N$, $\boldsymbol{\theta} := (\theta_1, \theta_2, \dots, \theta_N)' \in \mathbb{R}^N$ and $\boldsymbol{\sigma} := (\sigma_1, \sigma_2, \dots, \sigma_N)' \in \mathbb{R}^N$ with $\sigma_i > 0$ for each $i = 1, 2, \dots, N$; $\langle \cdot, \cdot \rangle$ is the scalar product in $\mathbb{R}^N$.

Let $F : (-\infty, \infty) \rightarrow [0, 1]$ be a continuous and non-decreasing function. This function plays the role of smoothing transitions in regimes. Suppose r is the threshold parameter and δ is the scale parameter, where $r \in \mathbb{R}$ and $\delta > 0$. Let d be a positive integer representing the delay parameter. Consider a time series $\{Y_t | t = -d+1, -d+2, \dots, 0, 1, \dots\}$, where the initial values $Y_{-d+1}, Y_{-d+2}, \dots, Y_0$ are given in advance. We suppose the remaining terms of the time series $\{Y_t | t \in \mathcal{T} \setminus \{0\}\}$ follow a HMRS-STAR $(2, N)$ model defined by:

$$Y_t = \mu(t) + \theta(t)F\left(\frac{Y_{t-d}-r}{\delta}\right) + \sigma(t)\epsilon_t.$$

Here we suppose, for simplicity, that the noise process $\{\epsilon_t | t \in \mathcal{T} \setminus \{0\}\}$ is a sequence of independent and identically distributed (i.i.d.) standard Gaussian random variables, (i.e., $\epsilon_t \sim N(0, 1)$).

When the scale parameter δ tends to zero, the function $F\left(\frac{Y_{t-d}-r}{\delta}\right)$ tends to the indicator function $I_{\{Y_{t-d} > r\}}$. In the limiting case, the transition in the smooth transition component becomes abrupt. The HMRS-STAR $(2, N)$ model becomes an $(2, N)$ -double threshold model in Elliott, Siu, and Lau (2013). One may ask, given that the smooth transition component may be able to model abrupt or smooth transitions for different values of the parameters δ , why we need to incorporate abrupt changes using a hidden Markov chain. There are at least two

reasons. Firstly, the nature of the abrupt changes in the proposed HMRS-STAR model and that incorporated in the hidden Markov chain are different. The abrupt change in the smooth transition component is a self-exciting change in the sense that it depends on a past value of the time series itself, while the abrupt change incorporated by the hidden Markov chain is generated exogenously. This partly motivated the double threshold model in Elliott, Siu, and Lau (2013). Secondly, the HMRS-STAR model can incorporate both abrupt changes and smooth changes simultaneously. The HMRS-STAR model could be applied to financial modelling. For example, we can consider $\{Y_t | t \in \mathcal{T} \setminus \{0\}\}$ as the return series of a financial asset. Then the HMRS-STAR model can incorporate two types of regime switches, namely, “abrupt” regime switches and “smooth” regime switches, in the appreciation rate of the asset. “Abrupt” regime switches may be attributed to sudden changes in hidden economic fundamentals which may occur during financial crises. “Smooth” regime switches may be attributed to observed news and events which are digested slowly by the market and gradually reflected in market prices or returns. In the modern finance theory, the theory of market efficiency is pertinent. One of the versions of this theory basically states that if a capital market is efficient, market prices react almost instantaneously to market news and information. However, the theory of market efficiency is questioned by some empirical studies for different markets such as equity, currency and commodity markets. Good references for these empirical studies are, for example, Taylor (1986, 2005). Indeed, in Taylor (1986), (see Chapter 7 therein), a price trend model was introduced as an alternative to the random walk model which is often assumed under market efficiency hypothesis. The key idea of the price trend model is that trend occurs when some market information is reflected in several consecutive asset returns. This reflects a kind of slow adjustments of market prices to market news and information.

In general, one could also consider a situation where both the threshold parameter r and the scale parameter δ are modulated by the hidden Markov chain $\mathbf{X}$. However, to simplify our analysis, we assume they are constants. The following two examples give two special cases of the HMRS-STAR model.

Example 1: Suppose the smooth transition function $F(x)$ is the standard Gaussian distribution $\Phi(x)$, (i.e., the probability distribution function of a zero-mean, unit-variance, normal distribution). Then the HMRS-STAR model becomes:

$$Y_t = \mu(t) + \theta(t)\Phi\left(\frac{Y_{t-d} - r}{\delta}\right) + \sigma(t)\epsilon_t.$$

Example 2: Suppose the smooth transition function $F(x)$ is the logistic distribution given by $[1 + \exp(-x)]^{-1}$. Then the HMRS-STAR model becomes:

$$Y_t = \mu(t) + \theta(t)\left[1 + \exp\left(\frac{r - Y_{t-d}}{\delta}\right)\right]^{-1} + \sigma(t)\epsilon_t.$$

The quantity $1/\delta$ is called the speed of transition parameter. It could be challenging to estimate using MLE directly, and the EM algorithm circumvents this step.

In the next section, we shall illustrate how to estimate the parameters r and δ using the Expectation Maximization (EM) algorithm and an approximation method for the case of the Gaussian smooth transition function. Indeed, this method also works for the case of the logistic smooth transition function when the following standard linear approximation method for the logistic function F is used:

$$F(x) = \frac{1}{1 + e^{-x}} \approx \frac{1}{2} + F'(0)x + O(x^2).$$

3 Filtering and estimation

The basic idea of the reference probability approach is to start with a reference probability $\bar{P}$ under which the observed process Y has simpler dynamics, (i.e. independent of the hidden Markov chain). Then the reference probability $\bar{P}$ is related to the real-world probability P by a measure change. Under P , the observed process is governed by the original dynamics. However, filters for the hidden quantities are derived under the reference probability $\bar{P}$. The reference probability approach for filtering hidden Markov models was discussed in Elliott, Aggoun, and Moore (1995). The mathematical techniques for deriving the filters and filter-based estimates here follow those in Elliott, Aggoun, and Moore (1995). We only present results which are relevant to the simulation and real-data analyses in the following two sections. For some technical details of the results, one may refer to, for example, Elliott, Aggoun, and Moore (1995). In what follows, we start by presenting essential concepts and notations. Then the filters and filter-based estimates are given.

3.1 Filters for the hidden chain and related quantities

Let $\bar{P}$ be a reference probability under which

1. $\{Y_t | t \in \mathcal{T} \setminus \{0\}\}$ is a sequence of i.i.d. standard normal random variables;
2. the chain $\mathbf{X}$ has a transition probability matrix Π .

Define $\mathbb{Y}^0 := \{\mathcal{Y}_t^0 | t \in \mathcal{T}\}$ and $\mathbb{G}^0 := \{\mathcal{G}_t^0 | t \in \mathcal{T}\}$ by:

$$\begin{aligned}\mathcal{Y}_t^0 &:= \sigma\{Y_1, Y_2, \dots, Y_t\} \vee \mathcal{Y}_0^0, \\ \mathcal{G}_t^0 &:= \sigma\{Y_1, Y_2, \dots, Y_t, \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t\} \vee \mathcal{G}_0^0, \quad t = 1, 2, \dots,\end{aligned}$$

and

$$\begin{aligned}\mathcal{Y}_0^0 &:= \sigma\{Y_{-d+1}, Y_{-d+2}, \dots, Y_0\}, \\ \mathcal{G}_0^0 &:= \sigma\{\mathbf{X}_0\} \vee \mathcal{Y}_0^0.\end{aligned}$$

Write $\mathbb{Y}$ and $\mathbb{G}$ for the P -completion of $\mathbb{Y}^0$ and $\mathbb{G}^0$, respectively. Given observed information $\mathcal{Y}_t$ up to time t , we wish to estimate $\mathbf{X}_t$ as:

$$\hat{\mathbf{X}}_t = E[\mathbf{X}_t | \mathcal{Y}_t],$$

where E is an expectation with respect to the measure P .

By a version of the Bayes' rule [see, for example, Elliott, Aggoun, and Moore (1995)],

$$E[\mathbf{X}_t | \mathcal{Y}_t] = \frac{\bar{E}[\Lambda_t \mathbf{X}_t | \mathcal{Y}_t]}{\bar{E}[\Lambda_t | \mathcal{Y}_t]},$$

where $\bar{E}$ is an expectation with respect to the measure $\bar{P}$. Write

$$\mathbf{q}_t := \bar{E}[\Lambda_t \mathbf{X}_t | \mathcal{Y}_t]. \quad (2)$$

For each $i = 1, 2, \dots, N$, let

$$\zeta_i(t, Y_t) := \frac{\phi\left(\frac{Y_t - \mu_i - \theta_i F\left(\frac{Y_t - d - r}{\delta}\right)}{\sigma_i}\right)}{\sigma_i \phi(Y_t)},$$

$$\boldsymbol{\zeta}(t, Y_t) := (\zeta_1(t, Y_t), \zeta_2(t, Y_t), \dots, \zeta_N(t, Y_t))' \in \mathfrak{R}^N,$$

and write $\mathbf{diag}(\boldsymbol{\zeta}(t, Y_t))$ for the diagonal matrix with diagonal elements being the components in the vector $\boldsymbol{\zeta}(t, Y_t)$. Then the unnormalized filter $\mathbf{q}_t$ satisfies the following recursive equation:

$$\mathbf{q}_t = \mathbf{diag}(\boldsymbol{\zeta}(t, Y_t)) \Pi \mathbf{q}_{t-1}, \quad t \in \mathcal{T} \setminus \{0\}.$$

Consequently,

$$E[\mathbf{X}_t | \mathcal{Y}_t] = \frac{\mathbf{q}_t}{\langle \mathbf{q}_t, \mathbf{1} \rangle}.$$

Here $\mathbf{1} := (1, 1, \dots, 1)' \in \mathfrak{R}^N$.

The following quantities will be used to derive the filter-based estimates of some unknown parameters in the HMRS-STAR model.

1. The number of transitions of the chain $\mathbf{X}$ from state $\mathbf{e}_i$ to state $\mathbf{e}_j$ up to time t , for each $i, j = 1, 2, \dots, N$, is:

$$J_t^{ij} := \sum_{l=1}^t \langle \mathbf{X}_{l-1}, \mathbf{e}_i \rangle \langle \mathbf{X}_l, \mathbf{e}_j \rangle.$$

2. The occupation time of the chain $\mathbf{X}$ in state $\mathbf{e}_i$ up to time t , for each $i = 1, 2, \dots, N$, is:

$$O_t^i := \sum_{l=1}^t \langle \mathbf{X}_l, \mathbf{e}_i \rangle.$$

3. The “generalized” level process associated with state $\mathbf{e}_i$ up to time t , for each $i = 1, 2, \dots, N$, is:

$$L_{d+1,t}^i := \sum_{l=1}^t f_{d+1}(Y_l, Y_{l-1}, \dots, Y_{l-d}) \langle \mathbf{X}_l, \mathbf{e}_i \rangle.$$

Here $f_{d+1} : \mathfrak{R}^{d+1} \rightarrow \mathfrak{R}$ is a Borel-measurable function.

As in Elliott, Aggoun, and Moore (1995) we define the following ‘unnormalized’ vector quantities:

$$\begin{aligned} \sigma(J_t^{ij} \mathbf{X}_t) &:= \bar{\mathbb{E}}[\Lambda_t J_t^{ij} \mathbf{X}_t | \mathcal{Y}_t] \\ \sigma(O_t^i \mathbf{X}_t) &:= \bar{\mathbb{E}}[\Lambda_t O_t^i \mathbf{X}_t | \mathcal{Y}_t] \\ \sigma(L_{d+1,t}^i \mathbf{X}_t) &:= \bar{\mathbb{E}}[\Lambda_t L_{d+1,t}^i \mathbf{X}_t | \mathcal{Y}_t], \quad t \in \mathcal{T} \setminus \{0\}, \end{aligned}$$

where

$$\sigma(J_0^{ij} \mathbf{X}_0) = \sigma(O_0^i \mathbf{X}_0) = \sigma(L_{d+1,0}^i \mathbf{X}_0) = \mathbf{0} \in \mathfrak{R}^N.$$

Then the exact recursive equations for $\sigma(J_t^{ij} \mathbf{X}_t)$, $\sigma(O_t^i \mathbf{X}_t)$ and $\sigma(L_{d+1,t}^i \mathbf{X}_t)$ are, respectively, given by:

$$\sigma(J_t^{ij} \mathbf{X}_t) = \mathbf{diag}(\boldsymbol{\zeta}(t, Y_t)) \Pi \sigma(J_{t-1}^{ij} \mathbf{X}_{t-1}) + \langle \boldsymbol{\zeta}(t, Y_t), \mathbf{e}_j \rangle \langle \mathbf{q}_{t-1}, \mathbf{e}_i \rangle \pi_{ji} \mathbf{e}_j, \quad (3)$$

$$\sigma(O_t^i \mathbf{X}_t) = \mathbf{diag}(\boldsymbol{\zeta}(t, Y_t)) \Pi \sigma(O_{t-1}^i \mathbf{X}_{t-1}) + \langle \boldsymbol{\zeta}(t, Y_t), \mathbf{e}_i \rangle \langle \Pi \mathbf{q}_{t-1}, \mathbf{e}_i \rangle \mathbf{e}_i, \quad (4)$$

and

$$\begin{aligned} &\sigma(L_{d+1,t}^i \mathbf{X}_t) \\ &= \mathbf{diag}(\boldsymbol{\zeta}(t, Y_t)) \Pi \sigma(L_{d+1,t-1}^i \mathbf{X}_{t-1}) + f_{d+1}(Y_t, Y_{t-1}, \dots, Y_{t-d}) \langle \boldsymbol{\zeta}(t, Y_t), \mathbf{e}_i \rangle \\ &\quad \times \langle \Pi \mathbf{q}_{t-1}, \mathbf{e}_i \rangle \mathbf{e}_i, \end{aligned} \quad (5)$$

for each $i = 1, 2, \dots, N$. See, for example, Elliott, Aggoun, and Moore (1995).

Note that

$$\begin{aligned} \sigma(J_t^{ij}) &:= \bar{\mathbb{E}}[\Lambda_t J_t^{ij} | \mathcal{Y}_t] = \langle \sigma(J_t^{ij} \mathbf{X}_t), \mathbf{1} \rangle, \\ \sigma(O_t^i) &:= \bar{\mathbb{E}}[\Lambda_t O_t^i | \mathcal{Y}_t] = \langle \sigma(O_t^i \mathbf{X}_t), \mathbf{1} \rangle, \\ \sigma(L_{d+1,t}^i) &:= \bar{\mathbb{E}}[\Lambda_t L_{d+1,t}^i | \mathcal{Y}_t] = \langle \sigma(L_{d+1,t}^i \mathbf{X}_t), \mathbf{1} \rangle. \end{aligned}$$

3.2 Filter-based estimates and the EM algorithm

The filter-based estimates for the unknown parameters in the HMRS-STAR model are derived using the Expectation Maximization, or EM, algorithm. Here we briefly present the idea of the EM algorithm. For details, one may refer to, for example, Elliott, Aggoun, and Moore (1995). Again the mathematical techniques used here follow from those in Elliott, Aggoun, and Moore (1995).

Let $\{P_\theta | \theta \in \Theta\}$ be a family of probability measures on a measurable space $(\Omega, \mathcal{F})$ all absolutely continuous with respect to a fixed probability measure P_0 , where θ is an unknown model parameter and Θ is the parameter space. Write $\mathcal{Y}$ for a sub- σ -field of $\mathcal{F}$. Then the likelihood function for computing an estimate of the unknown parameter θ given the available information in $\mathcal{Y}$ is:

$$L(\theta) := E_0 \left[\frac{dP_\theta}{dP_0} | \mathcal{Y} \right].$$

Here E_0 is the expectation under P_0 . Then the maximum likelihood estimate (MLE) $\hat{\theta}$ of θ is given by:

$$\hat{\theta} := \operatorname{argmax}_{\theta \in \Theta} L(\theta).$$

It may be challenging to compute the MLE $\hat{\theta}$ directly. The EM algorithm provides an iterative approximation method to compute the MLE $\hat{\theta}$. It consists of the following four steps, [see, for example, Elliott, Aggoun, and Moore (1995)]:

Step I: Set the counter $p = 0$ and choose $\hat{\theta}_0$.

Step II: (E-step) Set $\theta^* = \hat{\theta}_p$ and compute $Q(\cdot, \theta^*)$, where

$$Q(\theta, \theta^*) := E_{\theta^*} \left[\ln \left(\frac{dP_\theta}{dP_{\theta^*}} \right) | \mathcal{Y} \right],$$

where E_{θ^*} is the expectation under P_{θ^*} .

Step III: (M-step) Find

$$\theta_{p+1} := \operatorname{argmax}_{\theta \in \Theta} Q(\theta, \theta^*).$$

Step IV: Replace p by $p + 1$ and repeat beginning with Step II until a certain stopping criterion is satisfied.

Note that the EM algorithm we applied here recursively converges only to a local maximum of log-likelihood. For a discussion of the EM algorithm and its convergent properties, one may refer to Baum et al. (1970), Dembo and Zeitouni (1986), and Elliott, Aggoun, and Moore (1995).

The set of parameters of interest can be described by the set Θ defined as follows:

$$\Theta := \{(\pi_{kl})_{k,l=1,2,\dots,N}, (\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}), r, \delta\},$$

where $\boldsymbol{\mu} := (\mu_1, \mu_2, \dots, \mu_N)' \in \mathbb{R}^N$; $\boldsymbol{\theta} := (\theta_1, \theta_2, \dots, \theta_N)' \in \mathbb{R}^N$; $\boldsymbol{\sigma} := (\sigma_1, \sigma_2, \dots, \sigma_N)' \in \mathbb{R}^N$.

Since π_{ji} 's are transition probabilities, we also require that

$$\sum_{j=1}^N \pi_{ji} = 1, \quad i = 1, 2, \dots, N.$$

Suppose now that the set of parameters Θ is given and that the set of observed data described by the σ -field $\mathcal{Y}_n$ is known. We wish to determine a new set of parameters $\Theta(n)$ defined by:

$$\Theta(n) := \{(\pi_{kl}(n))_{k,l=1,2,\dots,N}, (\boldsymbol{\mu}(n), \boldsymbol{\theta}(n), \boldsymbol{\sigma}(n)), r(n), \delta(n)\},$$

where

$$\begin{aligned}\boldsymbol{\mu}(n) &:= (\mu_1(n), \mu_2(n), \dots, \mu_N(n))' \in \mathfrak{R}^N, \\ \boldsymbol{\theta}(n) &:= (\theta_1(n), \theta_2(n), \dots, \theta_N(n))' \in \mathfrak{R}^N, \\ \boldsymbol{\sigma}(n) &:= (\sigma_1(n), \sigma_2(n), \dots, \sigma_N(n))' \in \mathfrak{R}^N.\end{aligned}$$

Such a new set of parameters is determined by maximizing the conditional log-likelihoods defined below. The basic idea is that we update one set of parameters at a time starting with the $[\pi_{kl}]_{k,l=1,2,\dots,N}$. Using this method, it is known that a filter-based estimate $\widehat{\pi}_{kl}(n)$ for π_{kl} given observed data described by the σ -field $\mathcal{Y}_n$ is given by:

$$\widehat{\pi}_{kl}(n) = \frac{E[J_n^{kl}|\mathcal{Y}_n]}{E[O_n^l|\mathcal{Y}_n]} = \frac{\sigma(J_n^{kl})}{\sigma(O_n^l)}.$$

For derivation, interested readers may refer to Elliott, Aggoun, and Moore (1995), (see Chapter 2 therein).

Consider now another set of parameters $(\boldsymbol{\mu}, \boldsymbol{\theta})$. To change the set of parameters from $(\boldsymbol{\mu}, \boldsymbol{\theta})$ to $(\widehat{\boldsymbol{\mu}}(n), \widehat{\boldsymbol{\theta}}(n))$ while keeping other parameters $(\boldsymbol{\sigma}, r, \delta)$ constant, we must consider the following factors, $(t = 1, 2, \dots, n)$,

$$\begin{aligned}\widehat{\lambda}_t^1 &:= \exp \left\{ \frac{1}{2\langle \boldsymbol{\sigma}, \mathbf{X}_t \rangle^2} \left[\left\langle \boldsymbol{\mu} + \boldsymbol{\theta}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle^2 - \left\langle \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\theta}}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle^2 \right. \right. \\ &\quad \left. \left. - 2Y_t \left\langle \boldsymbol{\mu} + \boldsymbol{\theta}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle + 2Y_t \left\langle \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\theta}}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle \right] \right\}.\end{aligned}$$

Note that, to simplify the notation, we write $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\theta}}$ for $\widehat{\boldsymbol{\mu}}(n)$ and $\widehat{\boldsymbol{\theta}}(n)$, respectively.

Write

$$\begin{aligned}\widehat{\Lambda}_t^1 &:= \prod_{u=1}^t \widehat{\lambda}_u^1, \\ \widehat{\Lambda}_0^1 &:= 1.\end{aligned}$$

Then a new probability measure $\widehat{P}^1$ is defined so that the restriction of its Radon-Nikodym derivative $\frac{d\widehat{P}^1}{dP}$ to $\mathcal{Y}_n$ is given by:

$$\left. \frac{d\widehat{P}^1}{dP} \right|_{\mathcal{Y}_n} = \widehat{\Lambda}_n^1.$$

It is then not difficult to see that under $\widehat{P}^1$, the sequence defined by:

$$\frac{Y_t - \left\langle \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\theta}}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle}{\langle \boldsymbol{\sigma}, \mathbf{X}_t \rangle}, \quad t = 1, 2, \dots, n,$$

is a sequence of $N(0, 1)$ i.i.d. random variables.

Now

$$\begin{aligned}\ln \widehat{\Lambda}_n^1 &= \sum_{t=1}^n \frac{1}{2\langle \boldsymbol{\sigma}, \mathbf{X}_t \rangle^2} \left(\left\langle \boldsymbol{\mu} + \boldsymbol{\theta}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle^2 - \left\langle \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\theta}}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle^2 \right. \\ &\quad \left. - 2Y_t \left\langle \boldsymbol{\mu} + \boldsymbol{\theta}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle + 2Y_t \left\langle \widehat{\boldsymbol{\mu}} + \widehat{\boldsymbol{\theta}}F\left(\frac{Y_{t-d}-r}{\delta}\right), \mathbf{X}_t \right\rangle \right) \\ &= \sum_{t=1}^n \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[\left(\mu_i + \theta_i F\left(\frac{Y_{t-d}-r}{\delta}\right) \right)^2 - \left(\widehat{\mu}_i + \widehat{\theta}_i F\left(\frac{Y_{t-d}-r}{\delta}\right) \right)^2 \right. \\ &\quad \left. - 2Y_t \left(\mu_i + \theta_i F\left(\frac{Y_{t-d}-r}{\delta}\right) \right) + 2Y_t \left(\widehat{\mu}_i + \widehat{\theta}_i F\left(\frac{Y_{t-d}-r}{\delta}\right) \right) \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle \\ &= \sum_{t=1}^n \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[- \left(\widehat{\mu}_i + \widehat{\theta}_i F\left(\frac{Y_{t-d}-r}{\delta}\right) \right)^2 + 2Y_t \left(\widehat{\mu}_i + \widehat{\theta}_i F\left(\frac{Y_{t-d}-r}{\delta}\right) \right) \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle \\ &\quad + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta),\end{aligned}$$

where $R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta)$ does not involve $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\theta}}$ and it represents a quantity which may change from line to line. Consequently,

$$\begin{aligned} \ln \hat{\Lambda}_n^1 &= \sum_{t=1}^n \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[-\hat{\mu}_i^2 - \hat{\theta}_i^2 F^2\left(\frac{Y_{t-d}-r}{\delta}\right) - 2\hat{\mu}_i \hat{\theta}_i F\left(\frac{Y_{t-d}-r}{\delta}\right) \right. \\ &\quad \left. + 2\hat{\mu}_i Y_t + 2\hat{\theta}_i Y_t F\left(\frac{Y_{t-d}-r}{\delta}\right) \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta). \end{aligned}$$

Define the following functions:

$$\begin{aligned} f_{d+1}^1(Y_t, Y_{t-1}, \dots, Y_{t-d}) &:= Y_t, \\ f_{d+1}^2(Y_t, Y_{t-1}, \dots, Y_{t-d}) &:= F\left(\frac{Y_{t-d}-r}{\delta}\right), \\ f_{d+1}^3(Y_t, Y_{t-1}, \dots, Y_{t-d}) &:= Y_t F\left(\frac{Y_{t-d}-r}{\delta}\right), \\ f_{d+1}^4(Y_t, Y_{t-1}, \dots, Y_{t-d}) &:= F^2\left(\frac{Y_{t-d}-r}{\delta}\right), \end{aligned}$$

Consider the following functionals, for $i = 1, 2, \dots, N$ and $j = 1, 2, 3, 4$:

$$L_{d+1,n}^i(f_{d+1}^j) := \sum_{t=1}^n f_{d+1}^j(Y_t, Y_{t-1}, \dots, Y_{t-d}) \langle \mathbf{X}_t, \mathbf{e}_i \rangle.$$

Then

$$\begin{aligned} \ln \hat{\Lambda}_n^1 &= \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[-\hat{\mu}_i^2 O_n^i - \hat{\theta}_i^2 L_{d+1,n}^i(f_{d+1}^4) - 2\hat{\mu}_i \hat{\theta}_i L_{d+1,n}^i(f_{d+1}^2) + 2\hat{\mu}_i L_{d+1,n}^i(f_{d+1}^1) \right. \\ &\quad \left. + 2\hat{\theta}_i L_{d+1,n}^i(f_{d+1}^3) \right] + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta). \end{aligned}$$

Again $R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta)$ is a quantity which does not depend on $\hat{\mu}_i$ or $\hat{\theta}_i$ and can change from line to line. Conditioning on $\mathcal{Y}_n$ under P gives:

$$\begin{aligned} E[\ln \hat{\Lambda}_n^1 | \mathcal{Y}_n] &= \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[-\hat{\mu}_i^2 \hat{O}_n^i - \hat{\theta}_i^2 \hat{L}_{d+1,n}^i(f_{d+1}^4) - 2\hat{\mu}_i \hat{\theta}_i \hat{L}_{d+1,n}^i(f_{d+1}^2) + 2\hat{\mu}_i \hat{L}_{d+1,n}^i(f_{d+1}^1) \right. \\ &\quad \left. + 2\hat{\theta}_i \hat{L}_{d+1,n}^i(f_{d+1}^3) \right] + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta). \end{aligned}$$

Differentiating $E[\ln \hat{\Lambda}_n^1 | \mathcal{Y}_n]$ with respect to $\hat{\mu}_i$ and setting the derivative equal to zero gives:

$$\hat{\mu}_i \hat{O}_n^i + \hat{\theta}_i \hat{L}_{d+1,n}^i(f_{d+1}^2) = \hat{L}_{d+1,n}^i(f_{d+1}^1).$$

Solving for $\hat{\mu}_i$ then gives:

$$\hat{\mu}_i = \frac{\hat{L}_{d+1,n}^i(f_{d+1}^1) - \hat{\theta}_i \hat{L}_{d+1,n}^i(f_{d+1}^2)}{\hat{O}_n^i} = \frac{\sigma(L_{d+1,n}^i(f_{d+1}^1)) - \hat{\theta}_i \sigma(L_{d+1,n}^i(f_{d+1}^2))}{\sigma(O_n^i)}, \quad (6)$$

Note that the formula for $\hat{\mu}_i$ depends on $\hat{\theta}_i$.

Differentiating $E[\ln \hat{\Lambda}_n^1 | \mathcal{Y}_n]$ with respect to $\hat{\theta}_i$ and setting the derivative equal to zero gives:

$$\hat{\theta}_i \hat{L}_{d+1,n}^i(f_{d+1}^4) + \hat{\mu}_i \hat{L}_{d+1,n}^i(f_{d+1}^2) = L_{d+1,n}^i(f_{d+1}^3).$$

and this gives

$$\hat{\theta}_i = \frac{\hat{L}_{d+1,n}^i(f_{d+1}^3) - \hat{\mu}_i \hat{L}_{d+1,n}^i(f_{d+1}^2)}{\hat{L}_{d+1,n}^i(f_{d+1}^4)} = \frac{\sigma(L_{d+1,n}^i(f_{d+1}^3)) - \hat{\mu}_i \sigma(L_{d+1,n}^i(f_{d+1}^2))}{\sigma(L_{d+1,n}^i(f_{d+1}^4))}, \quad (7)$$

where the exact recursive formulae for evaluating $\sigma(L_{d+1,n}^i(f_{d+1}^j))$, $j=1, 2, 3, 4$, and $\sigma(O_n^i)$ are given in Section 3, say Theorem 3.4 and Theorem 3.5. Note that the formula for $\hat{\theta}_i$ depends on $\hat{\mu}_i$.

Consider now the set of parameters σ_i , $i=1, 2, \dots, N$. To change the parameters from σ_i to $\hat{\sigma}_i(n)$, $i=1, 2, \dots, N$, while keeping (μ, θ, r, δ) fixed, we must consider factors, $(t=1, 2, \dots, n)$:

$$\hat{\lambda}_t^2 := \frac{\langle \sigma, \mathbf{X}_t \rangle}{\langle \hat{\sigma}, \mathbf{X}_t \rangle} \frac{\exp \left[-\frac{1}{2\langle \hat{\sigma}, \mathbf{X}_t \rangle^2} \left(Y_t - \langle \mu + \theta F \left(\frac{Y_{t-d}-r}{\delta} \right), \mathbf{X}_t \rangle \right)^2 \right]}{\exp \left[-\frac{1}{2\langle \sigma, \mathbf{X}_t \rangle^2} \left(Y_t - \langle \mu + \theta F \left(\frac{Y_{t-d}-r}{\delta} \right), \mathbf{X}_t \rangle \right)^2 \right]}.$$

Here we assume that $\hat{\sigma}_i(n) > 0$.

Again, we write

$$\begin{aligned} \hat{\Lambda}_t^2 &:= \prod_{u=1}^t \hat{\lambda}_u^2, \\ \hat{\Lambda}_0^2 &:= 1. \end{aligned}$$

A new probability measure $\hat{P}^2$ can then be defined so that the restriction of its Radon-Nikodym derivative $\frac{d\hat{P}^2}{dP}$ to $\mathcal{Y}_n$ is given by:

$$\left. \frac{d\hat{P}^2}{dP} \right|_{\mathcal{Y}_n} = \hat{\Lambda}_n^2.$$

Now,

$$\begin{aligned} \ln \hat{\Lambda}_n^2 &= \sum_{t=1}^n \left[-\ln \langle \hat{\sigma}, \mathbf{X}_t \rangle - \frac{1}{2\langle \hat{\sigma}, \mathbf{X}_t \rangle^2} \left(Y_t - \left\langle \mu + \theta F \left(\frac{Y_{t-d}-r}{\delta} \right), \mathbf{X}_t \right\rangle \right)^2 \right] \\ &\quad + R(\mu, \theta, \sigma, r, \delta) \\ &= \sum_{t=1}^n \sum_{i=1}^N \left[-\ln \hat{\sigma}_i - \frac{1}{2\hat{\sigma}_i^2} \left(Y_t - \mu_i - \theta_i F \left(\frac{Y_{t-d}-r}{\delta} \right) \right)^2 \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle \\ &\quad + R(\mu, \theta, \sigma, r, \delta) \\ &= \sum_{t=1}^n \sum_{i=1}^N \left[-\ln \hat{\sigma}_i - \frac{1}{2\hat{\sigma}_i^2} \left(Y_t^2 - 2Y_t \mu_i - 2\theta_i Y_t F \left(\frac{Y_{t-d}-r}{\delta} \right) + \mu_i^2 \right. \right. \\ &\quad \left. \left. + \theta_i^2 F^2 \left(\frac{Y_{t-d}-r}{\delta} \right) + 2\mu_i \theta_i F \left(\frac{Y_{t-d}-r}{\delta} \right) \right) \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle \\ &\quad + R(\mu, \theta, \sigma, r, \delta). \end{aligned}$$

Define

$$f_{d+1}^5(Y_t, Y_{t-1}, \dots, Y_{t-d}) := Y_t^2,$$

and

$$L_{d+1,n}^i(f_{d+1}^5) = \sum_{t=1}^n f_{d+1}^5(Y_t, Y_{t-1}, \dots, Y_{t-d}) \langle \mathbf{X}_t, \mathbf{e}_i \rangle.$$

Again conditioning on $\mathcal{Y}_n$ under P gives:

$$\begin{aligned} & \mathbb{E}[\ln \widehat{\Lambda}_n^2 | \mathcal{Y}_n] \\ &= \sum_{i=1}^N \left[-\ln \widehat{\sigma}_i \widehat{O}_n^i - \frac{1}{2\widehat{\sigma}_i^2} \left(\widehat{L}_{d+1,n}^i(f_{d+1}^5) - 2\mu_i \widehat{L}_{d+1,n}^i(f_{d+1}^1) - 2\theta_i \widehat{L}_{d+1,n}^i(f_{d+1}^3) + \mu_i^2 \widehat{O}_n^i \right. \right. \\ & \quad \left. \left. + \theta_i^2 \widehat{L}_{d+1,n}^i(f_{d+1}^4) + 2\mu_i \theta_i \widehat{L}_{d+1,n}^i(f_{d+1}^2) \right) \right] + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta). \end{aligned}$$

Differentiating $\mathbb{E}[\ln \widehat{\Lambda}_n^2 | \mathcal{Y}_n]$ with respect to $\widehat{\sigma}_i$ and setting the derivative equal to zero gives:

$$\begin{aligned} \widehat{\sigma}_i^2 &= \frac{1}{\widehat{O}_n^i} \left(\widehat{L}_{d+1,n}^i(f_{d+1}^5) - 2\mu_i \widehat{L}_{d+1,n}^i(f_{d+1}^1) - 2\theta_i \widehat{L}_{d+1,n}^i(f_{d+1}^3) + \mu_i^2 \widehat{O}_n^i + \theta_i^2 \widehat{L}_{d+1,n}^i(f_{d+1}^4) \right. \\ & \quad \left. + 2\mu_i \theta_i \widehat{L}_{d+1,n}^i(f_{d+1}^2) \right) \\ &= \frac{1}{\sigma(O_n^i)} \left(\sigma(L_{d+1,n}^i(f_{d+1}^5)) - 2\mu_i \sigma(L_{d+1,n}^i(f_{d+1}^1)) - 2\theta_i \sigma(L_{d+1,n}^i(f_{d+1}^3)) + \mu_i^2 \sigma(O_n^i) \right. \\ & \quad \left. + \theta_i^2 \sigma(L_{d+1,n}^i(f_{d+1}^4)) + 2\mu_i \theta_i \sigma(L_{d+1,n}^i(f_{d+1}^2)) \right). \end{aligned} \quad (8)$$

By construction, these estimates are greater than zero. Note that the formula for $\widehat{\sigma}_i^2$ depends on μ_i and θ_i .

Finally, we consider the set of parameters (r, δ) . In this case, we consider a specific form of the smooth transition function F . For illustration, we consider the “normal” smooth transition function given in Example 1. That is,

$$F(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy.$$

In this case, we consider a Laplace series expansion for $\Phi(x)$ as follows:

$$\begin{aligned} \Phi(x) &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{n! 2^n (2n+1)} \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(x - \frac{x^3}{6} + \frac{x^5}{40} - \dots \right) \\ &= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} x + R(x^3). \end{aligned}$$

To change the parameters (r, δ) to $(\widehat{r}(n), \widehat{\delta}(n))$ while keeping other parameters $(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma})$ fixed, we consider the following factors, $(t = 1, 2, \dots, n)$:

$$\begin{aligned} \widehat{\lambda}_t^3 &:= \exp \left\{ \frac{1}{2 \langle \boldsymbol{\sigma}, \mathbf{X}_t \rangle^2} \left[\left\langle \boldsymbol{\mu} + \boldsymbol{\theta} \Phi \left(\frac{Y_{t-d} - r}{\delta} \right), \mathbf{X}_t \right\rangle^2 - \left\langle \boldsymbol{\mu} + \boldsymbol{\theta} \Phi \left(\frac{Y_{t-d} - \widehat{r}}{\widehat{\delta}} \right), \mathbf{X}_t \right\rangle^2 \right. \right. \\ & \quad \left. \left. - 2Y_t \left\langle \boldsymbol{\mu} + \boldsymbol{\theta} \Phi \left(\frac{Y_{t-d} - r}{\delta} \right), \mathbf{X}_t \right\rangle + 2Y_t \left\langle \boldsymbol{\mu} + \boldsymbol{\theta} \Phi \left(\frac{Y_{t-d} - \widehat{r}}{\widehat{\delta}} \right), \mathbf{X}_t \right\rangle \right] \right\}. \end{aligned}$$

Write

$$\begin{aligned} \widehat{\Lambda}_t^3 &:= \prod_{u=1}^t \widehat{\lambda}_u^3, \\ \widehat{\Lambda}_0^3 &:= 1. \end{aligned}$$

Again a new probability measure $\widehat{P}^3$ is defined so that the restriction of its Radon-Nikodym derivative $\frac{d\widehat{P}^3}{dP}$ to $\mathcal{Y}_n$ is given by:

$$\left. \frac{d\widehat{P}^3}{dP} \right|_{\mathcal{Y}_n} = \widehat{\Lambda}_n^3 .$$

Consequently,

$$\begin{aligned} & \ln \widehat{\Lambda}_n^3 \\ &= \sum_{i=1}^N \sum_{t=1}^N \frac{1}{2\sigma_i^2} \left[\left(\mu_i + \theta_i \Phi \left(\frac{Y_{t-d} - r}{\delta} \right) \right)^2 - \left(\mu_i + \theta_i \Phi \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right) \right)^2 \right. \\ & \quad \left. - 2Y_t \left(\mu_i + \theta_i \Phi \left(\frac{Y_{t-d} - r}{\delta} \right) \right) + 2Y_t \left(\mu_i + \theta_i \Phi \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right) \right) \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle \\ &= \sum_{t=1}^n \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[- \left(\mu_i + \theta_i \Phi \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right) \right)^2 + 2Y_t \theta_i \Phi \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right) \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle \\ & \quad + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta) \\ &= \sum_{t=1}^n \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[- \theta_i^2 \Phi^2 \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right) + 2(Y_t - \mu_i) \theta_i \Phi \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right) \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle \\ & \quad + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta) \\ &\approx \sum_{t=1}^n \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left\{ - \theta_i^2 \left[\frac{1}{4} + \frac{1}{\sqrt{2\pi}} \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right) + \frac{1}{2\pi} \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right)^2 \right] \right. \\ & \quad \left. + 2(Y_t - \mu_i) \theta_i \left[\frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left(\frac{Y_{t-d} - \hat{r}}{\hat{\delta}} \right) \right] \right\} \langle \mathbf{X}_t, \mathbf{e}_i \rangle \\ & \quad + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta) \\ &= \sum_{t=1}^n \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[\left(-\frac{1}{4} \theta_i^2 + \frac{\theta_i^2 \hat{r}}{\sqrt{2\pi} \hat{\delta}} - \frac{\theta_i^2 \hat{r}^2}{2\pi \hat{\delta}^2} - \mu_i \theta_i + \frac{2\mu_i \theta_i \hat{r}}{\sqrt{2\pi} \hat{\delta}} \right) \right. \\ & \quad + \left(\frac{\theta_i^2 \hat{r}}{\pi \hat{\delta}^2} - \frac{\theta_i^2}{\sqrt{2\pi} \hat{\delta}} - \frac{2\mu_i \theta_i}{\sqrt{2\pi} \hat{\delta}} \right) Y_{t-d} + \left(\theta_i - \frac{2\hat{r} \theta_i}{\sqrt{2\pi} \hat{\delta}} \right) Y_t + \frac{2Y_t Y_{t-d} \theta_i}{\sqrt{2\pi} \hat{\delta}} \\ & \quad \left. - \frac{\theta_i^2}{2\pi \hat{\delta}^2} Y_{t-d}^2 \right] \langle \mathbf{X}_t, \mathbf{e}_i \rangle + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta) , \end{aligned}$$

where $R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta)$ does not involve $\hat{r}$ and $\hat{\delta}$.

Write

$$\begin{aligned} f_{d+1}^6(Y_t, Y_{t-1}, \dots, Y_{t-d}) &= Y_{t-d} , \\ f_{d+1}^7(Y_t, Y_{t-1}, \dots, Y_{t-d}) &= Y_{t-d}^2 , \\ f_{d+1}^8(Y_t, Y_{t-1}, \dots, Y_{t-d}) &= Y_t Y_{t-d} , \end{aligned}$$

and for $j = 6, 7, 8$,

$$L_{d+1,n}^i(f_{d+1}^j) = \sum_{t=1}^n f_{d+1}^j(Y_t, Y_{t-1}, \dots, Y_{t-d}) \langle \mathbf{X}_t, \mathbf{e}_i \rangle .$$

Consequently,

$$\begin{aligned}
& \ln \hat{\Lambda}_n^3 \\
&= \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[\left(-\frac{1}{4}\theta_i^2 + \frac{\theta_i^2 \hat{r}}{\sqrt{2\pi}\hat{\delta}} - \frac{\theta_i^2 \hat{r}^2}{2\pi\hat{\delta}^2} - \mu_i\theta_i + \frac{2\mu_i\theta_i\hat{r}}{\sqrt{2\pi}\hat{\delta}} \right) \widehat{O}_n^i \right. \\
&\quad + \left(\frac{\theta_i^2 \hat{r}}{\pi\hat{\delta}^2} - \frac{\theta_i^2}{\sqrt{2\pi}\hat{\delta}} - \frac{2\mu_i\theta_i}{\sqrt{2\pi}\hat{\delta}} \right) \widehat{L}_{d+1,n}^i(f_{d+1}^6) + \left(\theta_i - \frac{2\hat{r}\theta_i}{\sqrt{2\pi}\hat{\delta}} \right) \widehat{L}_{d+1,n}^i(f_{d+1}^1) \\
&\quad \left. + \frac{2\theta_i}{\sqrt{2\pi}\hat{\delta}} \widehat{L}_{d+1,n}^i(f_{d+1}^8) - \frac{\theta_i^2}{2\pi\hat{\delta}^2} \widehat{L}_{d+1,n}^i(f_{d+1}^7) \right] + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta) .
\end{aligned}$$

Conditioning on $\mathcal{Y}_n$ under P then gives:

$$\begin{aligned}
& E[\ln \hat{\Lambda}_n^3 | \mathcal{Y}_n] \\
&= \sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[\left(-\frac{1}{4}\theta_i^2 + \frac{\theta_i^2 \hat{r}}{\sqrt{2\pi}\hat{\delta}} - \frac{\theta_i^2 \hat{r}^2}{2\pi\hat{\delta}^2} - \mu_i\theta_i + \frac{2\mu_i\theta_i\hat{r}}{\sqrt{2\pi}\hat{\delta}} \right) \widehat{O}_n^i \right. \\
&\quad + \left(\frac{\theta_i^2 \hat{r}}{\pi\hat{\delta}^2} - \frac{\theta_i^2}{\sqrt{2\pi}\hat{\delta}} - \frac{2\mu_i\theta_i}{\sqrt{2\pi}\hat{\delta}} \right) \widehat{L}_{d+1,n}^i(f_{d+1}^6) + \left(\theta_i - \frac{2\hat{r}\theta_i}{\sqrt{2\pi}\hat{\delta}} \right) \widehat{L}_{d+1,n}^i(f_{d+1}^1) \\
&\quad \left. + \frac{2\theta_i}{\sqrt{2\pi}\hat{\delta}} \widehat{L}_{d+1,n}^i(f_{d+1}^8) - \frac{\theta_i^2}{2\pi\hat{\delta}^2} \widehat{L}_{d+1,n}^i(f_{d+1}^7) \right] + R(\boldsymbol{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}, r, \delta) .
\end{aligned}$$

Differentiating with respect to $\hat{r}$ and setting the derivative equal to zero gives:

$$\begin{aligned}
\hat{r} &= \frac{\sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[\left[\frac{\theta_i^2}{\sqrt{2\pi}\hat{\delta}} + \frac{2\mu_i\theta_i}{\sqrt{2\pi}\hat{\delta}} \right] \widehat{O}_n^i + \frac{\theta_i^2}{\pi\hat{\delta}^2} \widehat{L}_{d+1,n}^i(f_{d+1}^6) - \frac{2\theta_i}{\sqrt{2\pi}\hat{\delta}} \widehat{L}_{d+1,n}^i(f_{d+1}^1) \right]}{\sum_{i=1}^N \frac{1}{2\sigma_i^2} \frac{\theta_i^2}{\pi\hat{\delta}^2} \widehat{O}_n^i} \\
&= \frac{\sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[\left[\frac{\theta_i^2}{\sqrt{2\pi}\hat{\delta}} + \frac{2\mu_i\theta_i}{\sqrt{2\pi}\hat{\delta}} \right] \sigma(O_n^i) + \frac{\theta_i^2}{\pi\hat{\delta}^2} \sigma(L_{d+1,n}^i(f_{d+1}^6)) - \frac{2\theta_i}{\sqrt{2\pi}\hat{\delta}} \sigma(L_{d+1,n}^i(f_{d+1}^1)) \right]}{\sum_{i=1}^N \frac{1}{2\sigma_i^2} \frac{\theta_i^2}{\pi\hat{\delta}^2} \sigma(O_n^i)} .
\end{aligned} \tag{9}$$

Note that the formula for $\hat{r}$ depends on $\hat{\delta}$, θ_i , μ_i and σ_i .

Differentiating with respect to $\hat{\delta}$ and setting the derivative equal to zero gives:

$$\begin{aligned}
\hat{\delta} &= \frac{\sum_{i=1}^N \frac{\theta_i^2}{2\pi\sigma_i^2} \left[-\widehat{L}_{d+1,n}^i(f_{d+1}^7) + 2\hat{r}\widehat{L}_{d+1,n}^i(f_{d+1}^6) - \hat{r}^2\widehat{O}_n^i \right]}{\sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[a_i\widehat{O}_n^i + b_i\widehat{L}_{d+1,n}^i(f_{d+1}^6) + c_i\widehat{L}_{d+1,n}^i(f_{d+1}^1) + d_i\widehat{L}_{d+1,n}^i(f_{d+1}^8) \right]} \\
&= \frac{\sum_{i=1}^N \frac{\theta_i^2}{2\pi\sigma_i^2} \left[-\sigma(L_{d+1,n}^i(f_{d+1}^7)) + 2\hat{r}\sigma(L_{d+1,n}^i(f_{d+1}^6)) - \hat{r}^2\sigma(O_n^i) \right]}{\sum_{i=1}^N \frac{1}{2\sigma_i^2} \left[a_i\sigma(O_n^i) + b_i\sigma(L_{d+1,n}^i(f_{d+1}^6)) + c_i\sigma(L_{d+1,n}^i(f_{d+1}^1)) + d_i\sigma(L_{d+1,n}^i(f_{d+1}^8)) \right]} ,
\end{aligned} \tag{10}$$

where

$$\begin{aligned}
a_i &:= -\frac{\theta_i^2 \hat{r}}{\sqrt{2\pi}} - \frac{2\mu_i\theta_i \hat{r}}{\sqrt{2\pi}} , \\
b_i &:= \frac{\theta_i^2}{\sqrt{2\pi}} + \frac{2\mu_i\theta_i}{\sqrt{2\pi}} , \\
c_i &:= \frac{2\hat{r}\theta_i}{\sqrt{2\pi}} , \\
d_i &:= -\frac{2\theta_i}{\sqrt{2\pi}} .
\end{aligned}$$

Note that these coefficients depend on $\hat{r}$, μ_i , θ_i and σ_i .

4 Simulation study

We illustrate the implementation of the proposed filtering and estimation algorithm for the Hidden Markov Regime-Switching Smooth Transition Model in practice using simulated data. The accuracy of the filters of the hidden Markov chain and filter-based estimates of unknown parameters as well as the convergence of the parameters estimates of the filter-based EM algorithm are also studied using the simulated data. Here we consider a simple, two-state and two-regime, HMRS-STAR model, namely an HMRS-STAR(2, 2)-model, following with lag $d = 1$ and sample size $n = 1000$. That is, the following model is considered.

$$Y_t = \begin{cases} -1 - 1 \times \Phi\left(\frac{Y_{t-1} - 0}{10}\right) + 2 \times \epsilon_t, & \text{for } \mathbf{X}_t = \mathbf{e}_1 \\ 1 + 1 \times \Phi\left(\frac{Y_{t-1} - 0}{10}\right) + 1 \times \epsilon_t, & \text{for } \mathbf{X}_t = \mathbf{e}_2 \end{cases}$$

It is clear that $\mu_1 = -1, \mu_2 = 1, \theta_1 = -1, \theta = 1, \sigma_1 = 2, \sigma_2 = 1, \delta = 10$ and $r = 0$, and that ϵ_t , for $t = 1, \dots, n$, are independent standard normal random variables. We consider the following “hypothetical” transition probability matrix $\begin{pmatrix} .6 & .3 \\ .4 & .7 \end{pmatrix}$ for the hidden Markov chain. The plot of a simulated series is depicted in Figure 1.

Figure 1: Plots of simulated data.

With the initial setting, $\mathbf{q}_0 = \mathbf{p}_0 = (0.5, 0.5)'$, $\mathbf{\Pi} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$; σ_1 is equal to the standard deviation of those Y 's which are less than its sample average; σ_2 is equal to the standard deviation of those Y 's which are greater than its sample average; $\mu_1 = \theta_1$ is equal to the average of those Y 's which are less than its sample average; $\mu_2 = \theta_2$ is equal to the average of those Y 's which are greater than its sample average, and $\delta = r = 1$. The algorithm is terminated when differences in the current parameters estimates and the previous parameters estimates are all less than 0.005. In this setup, we obtain our final estimation in iteration 87, see Table 1.

Table 1: Parameter estimates from the simulation.

Parameter	True value	Estimate	Parameter	True value	Estimate
π_{11}	0.6	0.56	π_{21}	0.4	0.44
π_{12}	0.3	0.27	π_{22}	0.7	0.73
μ_1	-1	-1.24	μ_2	1	0.88
θ_1	-1	-0.95	θ_2	1	1.15
σ_1	2	1.94	σ_2	1	1.08
r	0	0.0025	δ	10	13.52

We also provide figures to illustrate the convergence of the parameters estimates. From the estimation results displayed in the above table, we see that the parameters estimates from the filter-based EM algorithm are reasonably accurate for most of the parameters. However, compared with other parameters estimates, the parameter estimate $\hat{\delta}$ for the scale parameter δ is less accurate. Note that a sample size $n = 1000$ is required to obtain these parameters estimates. One possible explanation for the less accurate estimation result for the scale parameter delta is that the parameter estimate is based on an approximate formula derived from taking the first term only in the Laplace series expansion. The accuracy of the estimation for the scale parameter may be improved if higher-order terms in the Laplace series expansion are included. Of course, there is a trade-off

between the accuracy of the estimation and the tractability of the computation of the parameter estimates. Furthermore, in general, it seems to be difficult to achieve an accurate estimation for a scale parameter and hence a large sample size may be required to achieve an accurate estimate at an intuitive level. In practice, most of financial series are large and easily exceed 1000 observations.¹ Figure 2 plots the values of Π against iteration. Figure 3 plots the values of μ_1, μ_2 against iteration. Figure 4 plots the values of θ_1, θ_2 against iteration. Figure 5 plots the values of σ_1, σ_2 against iteration. Figure 6 plots the values of r and δ against iteration. Figure 7 presents the filtered estimate of the paths of the hidden Markov chain (Red line: p_1 of $\mathbf{p} = \mathbf{q} / < \mathbf{q}, \mathbf{1} >$, Blue line: p_2 of $\mathbf{p} = \mathbf{q} / < \mathbf{q}, \mathbf{1} >$) and the simulated paths of the hidden Markov chain based on the hypothetical parameters (black line). From Figure 3–Figure 6, we see that the parameters estimates converge at reasonable rates with most of them converging at 87 iterations. Furthermore, from Figure 7, it seems that the filtered estimate of the path of the hidden Markov chain matches quite reasonably well with the simulated path of the hidden Markov chain.

Figure 2: Convergence of Π .

Figure 3: Convergence of μ_1 & μ_2 .

Figure 4: Convergence of θ_1 & θ_2 .

Figure 5: Convergence of σ_1 & σ_2 .

Figure 6: Convergence of r & δ .

Figure 7: Estimate of $\mathbf{p} = \mathbf{q} / \langle \mathbf{q}, \mathbf{1} \rangle$.

5 Real data illustration

We implement our HMRS-STAR model and procedure on two sets of indexes, Hang Seng Index (HSI) and NASDAQ Composite (IXIC) from 3rd Jan 2012 to 30th Dec 2016. There are a total of 1246 and 1258 days for the Hang Seng Index and NASDAQ Composite respectively. Data are achieved from Yahoo finance via the link finance.yahoo.com. Let P_t be the index at day t and we consider the log return defined by: $Y_t = 100 \log(P_t / P_{t-1})$. Figure 8 shows the return data.

Figure 8: plots of $Y_t = 100 \times \log\text{-return}$ (left: HSI, Right: IXIC).

The returns are fitted to the two-state and two-regime, HMRS-STAR model, namely an HMRS-STAR(2, 2)-model, following with lag $d = 1$, and the initial values are identical to the one used in the simulation study, that is $\mathbf{q}_0 = \mathbf{p}_0 = (0.5, 0.5)'$, $\mathbf{\Pi} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$; σ_1 is equal to the standard deviation of those Y 's which are less than the sample average; σ_2 is equal to the standard deviation of those Y 's which are greater than its sample average; $\mu_1 = \theta_1$ is equal to the average of those Y 's which are less than its sample average; $\mu_2 = \theta_2$ is equal to the average of those Y 's which are greater than its sample average, and $\delta = r = 1$. The algorithm is terminated when differences in the current parameters estimates and the previous parameters estimates are all less than 0.005. We obtain our final estimations in iterations 394 and 60 for the Hang Seng Index and NASDAQ Composite respectively. The estimation algorithm follows the general algorithm, (i.e. Steps I–IV), presented in Section 3.2. The estimated values with respect to the log returns are given in Table 2. Again we also provide figures to illustrate the convergence of the parameter estimates. The following figures are plotted based on the estimation of $100 \times \log\text{-return}$. Figure 9 plots the values of $\mathbf{\Pi}$ against iteration. Figure 10 plots the values of μ_1, μ_2 against iteration. Figure 11 plots the values of θ_1, θ_2 against iteration. Figure 12 plots the values of σ_1, σ_2 against iteration. Figure 13 plots the values of r and δ against iteration. Figure 14 presents the filtered estimate of the paths of the hidden Markov chain (Red line: p_1 of $\mathbf{p} = \mathbf{q} / \langle \mathbf{q}, \mathbf{1} \rangle$, Blue line: p_2 of $\mathbf{p} = \mathbf{q} / \langle \mathbf{q}, \mathbf{1} \rangle$). From Figure 9–Figure 13, we see that the parameters estimates converge at reasonable rates with most of them converging at 394 and 60 iterations respectively. The convergence of the parameters are rather stable, except for the two parameters r and δ of HSI data significantly unstable around iterations 100–400 as shown in Figure 13. Also for the two real

data sets, it seems that the parameters estimates can disentangle the two regimes for μ s and σ s, but not for θ s for the HSI data.

Table 2: Parameter estimates and summary statistics from the real data.

Parameter	HSI	IXIC	Parameter	HSI	IXIC
π_{11}	0.59	0.88	π_{21}	0.41	0.12
π_{12}	0.31	0.10	π_{22}	0.69	0.90
μ_1	-0.36	0.12	μ_2	-0.15	-0.22
θ_1	0.62	-0.43	θ_2	0.47	0.86
σ_1	1.45	1.23	σ_2	0.67	0.58
r	7.58	-0.15	δ	0.41	-3.12
Sample mean	0.0123	0.0564	Sample standard deviation	1.0836	0.9392

Figure 9: Convergence of Π (left: HSI, right: IXIC).

Figure 10: Convergence of μ_1 & μ_2 (left: HSI, right: IXIC).

Figure 11: Convergence of θ_1 & θ_2 (left: HSI, right: IXIC).

Figure 12: Convergence of σ_1 & σ_2 (left: HSI, right: IXIC).

Figure 13: Convergence of r & δ (left: HSI, right: IXIC).

Figure 14: Estimate of $p = q / \langle q, 1 \rangle$ (left: HSI, right: IXIC).

6 Discussions on other potential economic applications

The proposed class of HMRS-STAR models may be applied to study some important problems in econometrics and economics as well as real-world problems which may be of socio-economic importance. Here we provide some suggestions for potential applications of the proposed class of models. There may be interesting applications beyond those mentioned below.

As noted in, for example, Hansen (2011), one of the important problems in econometrics is testing for linearity in economic time series against a certain nonlinear alternative which may be described by a parametric nonlinear time series model. This problem has significant economic implications since many economic models were developed under the premise of linearity. However, real-world data may reveal otherwise. A challenging issue is what nonlinear economic models may be used if a linear economic model is not appropriate.² Indeed, different conclusions may be drawn from testing for linearity if different parametric nonlinear time series models are used as the nonlinear alternative. The testing of a linear model against a threshold autoregressive model has been considered by statisticians and econometricians, see, for example, Chan (1990), Chan and Tong (1990), and Hansen (1996). See also Hansen (2011) and Tong (2011) and the relevant references therein. The testing of a linear model against a Markovian regime-switching autoregressive model has been considered in the literature. See, for example, Hamilton (2016) and the relevant references therein. The testing of a linear model against a smooth transition autoregressive model has also been considered. See, for example, van Dijk, Teräsvirta, and Franses (2002) and the related literature therein. However, it seems that the testing of linearity against a composite alternative model such as a second-generation nonlinear time series model may have received a relatively less attention in the literature. Intuitively, it appears that considering a composite alternative model in testing for linearity may provide a more general and flexible way to detect nonlinearity. The proposed class of HMRS-STAR models may be used to form a composite alternative model in testing for linearity in economic time series. Besides testing for linearity in economic time series, another potential application of the proposed class of HMRS-STAR models is testing for the unit root which is an important topic in econometrics. As noted in Hansen (2011), testing for the unit root has been investigated using a nonlinear stationarity threshold autoregressive model. It may perhaps be interesting to investigate to use of the proposed class of HMRS-STAR models to study the testing of the unit root. Time delay is also an important feature of economic time series. One potential application of the proposed class of HMRS-STAR models is modeling of time delay of economic time series. A flexibility that is provided by the HMRS-STAR models is that abrupt regime switches and time delay can be disentangled under the HMRS-STAR models. Indeed, inherent from the STAR models, time delay is related to smooth regime switches.

Some major classes of parametric nonlinear time series models such as the threshold autoregressive models have been adopted to investigate the law of one price and transactions costs. See, for example, Sarno, Taylor, and Chowdhury (2004) and Taylor (2001) for the use of threshold autoregressive models to investigate the purchasing power parity puzzle, the law of one price and transaction costs. See also Hansen (2011) for some

related discussions. It was noted in Taylor (2001) that a linear time series model such as an autoregressive model cannot provide a realistic description for nonlinear adjustments of prices attributed to the presence of transaction costs. In Taylor (2001), a two-regime threshold autoregressive model was employed to describe such nonlinearity, where the two threshold parameters are used to describe a “band of inaction” which may cause the nonlinearity. Instead of using the threshold autoregressive models, one may explore the use of the proposed class of HMRS-STAR models to study the law of one price and transaction costs. For example, one may investigate the nonlinearity due to transaction costs using both the abrupt and smooth regime switches.

7 Conclusion

We introduced a hidden, regime-switching, smooth transition model and discussed its filtering and estimation issues. A reference probability approach and a version of the Bayes’ rule were used to derive filters for the hidden Markov chain and some related quantities. These related quantities were used to derive filter-based estimates for the unknown parameters using the EM algorithm. For the threshold parameter and the scale parameter we employed a Laplace series expansion for the cumulative normal probability distribution to derive approximations to their filter-based estimates. Simulation experiments were presented to illustrate the practical implementation of the HMRS-STAR model as well as the filtering and estimation algorithm. The simulation results reveal that the parameters estimates converge at reasonable rates, say 87 iterations for most of the parameters, and that the parameters estimates are reasonably accurate for most of the parameters. Furthermore, real financial data were used to illustrate the practical implementation of the model. However, we have experienced a challenging issue that some other datasets that our estimation algorithm does not converge. So, there is still room to further improve the estimation method in particular the convergence, stability and robustness. This may represent interesting topics for further research. Other potential applications of the proposed model may be to apply the model for fitting some economic series such as unemployment data which is an important area in economics and econometrics. For this application, one may refer to, for example, Koop and Potter (1999) and Hamilton (2005). It may also be interesting to explore the use of the proposed model to fit financial time series other than equity indices, such as foreign exchange rates series. One may also explore some applications of the proposed model to study time series data in other fields such as climate science, ecology, population dynamics, biological, engineering and physical sciences.³

Acknowledgement

The authors would like to thank the Associate Editor and referees for their valuable and helpful comments.

Notes

- 1 We would like to thank one of the referees for stimulating this discussion.
- 2 As noted in Tong (1990), Chapter 2, when one leaves a linear world, there are infinitely many parametric nonlinear alternatives. A challenging issue is which nonlinear alternative may be used.
- 3 We would like to thank one of the referees for stimulating our discussions on some potential applications of the proposed model.

Code and Datasets: The author(s) published code and data associated with this article is on Code Ocean, a computational reproducibility platform. We recommend Code Ocean to SNDE contributors who wish share, discover, and run code in published research articles. (See: <https://doi.org/10.24433/CO.dd77cb0f-e54e-4693-905b-493a86cfd345>).

References

- Baum, L. E., T. Petrie, G. Soules, and N. Weiss. 1970. “A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains.” *The Annals of Mathematical Statistics* 41: 164–171.
- Chan, K. S., and H. Tong. 1986. “On Estimating Thresholds in Autoregressive Models.” *Journal of Time Series Analysis* 7: 178–190.
- Chan, K. S. 1990. “Testing for Threshold Autoregression.” *The Annals of Statistics* 18: 1886–1894.
- Chan, K. S., and H. Tong. 1990. “On Likelihood Ratio Tests for Threshold Autoregression.” *Journal of the Royal Statistical Society. Series B (Methodological)* 52: 469–476.

- Dembo, A., and O. Zeitouni. 1986. "Parameter Estimation of Partially Observed Continuous Time Stochastic Processes." *Stochastic Processes and Their Applications* 23: 91–113.
- Engle, R. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50: 987–1007.
- Elliott, R. J., L. Aggoun, and J. Moore. 1995. *Hidden Markov Models: Estimation and Control*. New York: Springer.
- Elliott, R. J., C. C. Liew, and T. K. Siu. 2011. "On Filtering and Estimation of a Threshold Stochastic Volatility Model." *Applied Mathematics and Computation* 218: 61–75.
- Elliott, R. J., T. K. Siu, and J. W. Lau. 2013. "Filtering a Double Threshold Model with Regime Switching." *IEEE Transactions on Automatic Control* 58: 3185–3190.
- Fan, J., and Q. Yao. 2003 *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.
- Gao, J.. 2007 *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. London: Chapman and Hall/CRC.
- Goldfeld, S. M., and S. M. Quandt. 1973. "A Markov Model for Switching Regressions." *Journal of Econometrics* 1: 3–15.
- Granger, C. W., and A. P. Anderson. 1978. *An Introduction to Bilinear Time Series Model*. Cuttingen: Vandenhoeck and Ruprecht.
- Granger, C. W. J., and T. Teräsvirta. 1993. *Modelling Non-linear Econometric Relationships*. Oxford: Oxford University Press.
- Hamilton, J. D. 1989. "A New Approach to Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57: 357–384.
- Hamilton, J. D. 2005. "What's Real About the Business Cycle?" *Federal Reserve Bank of St. Louis Review* 87: 435–452.
- Hamilton, J. D. 2016. "Macroeconomic Regimes and Regime Shifts." In *Handbook of Macroeconomics*, edited by H. Uhlig and J. Taylor, Volume 2A, 163–201. Amsterdam: Elsevier.
- Hansen, B. E. 1996. "Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis." *Econometrica* 64: 413–430.
- Hansen, B. E. 2011. "Threshold Autoregression in Economics." *Statistics and Its Interface* 4: 123–127.
- Koop, G., and S. M. Potter. 1999. "Dynamic Asymmetries in U.S. unemployment." *Journal of Business & Economic Statistics* 17: 298–312.
- Potter, S. M. 1999. "Nonlinear Time Series Modelling: An Introduction." *Journal of Economic Surveys* 13: 505–528.
- Priestley, M. B. 1980. "State-Dependent Models: A General Approach to Non-linear Time Series Analysis." *Journal of Time Series Analysis* 1: 47–71.
- Quandt, R. E. 1958. "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes." *Journal of the American Statistical Association* 53: 873–880.
- Sarno, L., M. P. Taylor, and I. Chowdhury. 2004. "Nonlinear Dynamics in Deviations from the Law of One Price: A Broad-Based Empirical Study." *Journal of International Money and Finance* 23: 1–25.
- Subba Rao, T., and M. M. Gabr. 1984. *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. New York: Springer.
- Taylor, S. J. 1982. "Financial Returns Modelled by the Product of Two Stochastic Processes, a Study of Daily Sugar Prices, 1961–79." In *Time Series Analysis: Theory and Practice* 1, edited by O. D. Anderson, 203–226. Amsterdam: North Holland.
- Taylor, S. J. 1986. *Modeling Financial Time Series*. Chichester: Wiley.
- Taylor, A. M. 2001. "Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price." *Econometrica* 69: 473–498.
- Taylor, S. J. 2005. *Asset Price Dynamics, Volatility, and Prediction*. Princeton: Princeton University Press.
- Teräsvirta, T. 1994. "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models." *Journal of the American Statistical Association* 89: 208–218.
- Teräsvirta, T. 1998. "Modelling Economic Relationships with Smooth Transition Regressions." In *Handbook of Applied Economic Statistics*, edited by D. E. A. Giles and A. Ullah, 507–552. New York: Marcel Dekker.
- Tong, H.. 1977. "Contribution to the Discussion of the Paper Entitled "Stochastic Modelling of Riverflow Time Series" by A. J. Lawrance and N. T. Kottegoda." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 140: 34–35.
- Tong, H. 1978. "On a Threshold Model, Pattern Recognition and Signal Processing." In *NATO ASI Series E: Applied Sc. No. 29*, edited by C. H. Chen, 575–586. The Netherlands: Sijthoff & Noordhoff.
- Tong, H. 1983. *Threshold Models in Non-Linear Time Series Analysis*. New York: Springer.
- Tong, H. 1990. *Nonlinear Time Series Analysis: A Dynamical System Approach*. Oxford: Oxford University Press.
- Tong, H. 2011. "Threshold Models in Time Series Analysis – 30 years on." *Statistics and Its Interface* 4: 107–118.
- Tong, H., and K. S. Lim. 1980. "Threshold Autoregression, Limit Cycles and Cyclical Data (with Discussion)." *Journal of the Royal Statistical Society. Series B (Methodological)* 42: 245–292.
- van Dijk, D., T. Teräsvirta, and P. H. Franses. 2002. "Smooth Transition Autoregressive Models – A Survey of Recent Developments." *Econometric Reviews* 21: 1–47.
- Yin, G., and C. Zhu. 2010. *Hybrid Switching Diffusions: Properties and Applications*. New York: Springer.
- Zhu, D. M., W. K. Ching, R. J. Elliott, T. K. Siu, and L. Zhang. 2017. "Hidden Markov Models with Threshold Effects and Their Applications to Oil Price Forecasting." *Journal of Industrial and Management Optimization* 13: 757–773.

Supplementary Material: The online version of this article offers supplementary material (DOI: <https://doi.org/10.1515/snde-2016-0061>).