Filtering a Double Threshold Model With Regime Switching

Robert J. Elliott, Tak Kuen Siu, and John W. Lau

Abstract—**We introduce a new double threshold model with regime switches. New filtering equations are derived based on a reference probability approach. We also propose a new and practically useful method for implementing the filtering equations.**

Index Terms—**Expectation Maximization (EM), algorithm, financial data, hidden Markov chain, recursive filters, reference probability, threshold models.**

I. Introduction

Regime-switching models have a long history in engineering, in particular in signal processing, systems and control engineering. The key idea of regime switching models is that the coefficients of the models are modulated by an underlying state process which is usually a finite-state hidden Markov chain. For detail, interested readers may refer to the monograph by Yin and Zhu [15]. Though having a long history in engineering, the development of regime switching models may also be tracked back to some early works in statistics and econometrics. Quandt [11] and Goldfeld and Quandt [4] proposed the use of regime-switching regression models to describe nonlinearity exhibited in economic data. Applications on recent advance in engineering could also be found in Changuel *et al.* (2010), Chen *et al.* (2011) and Janczura (2010). These include analyzing video compression rate, forecasting electricity load, and modeling electricity spot price. The idea of regime switching also appeared in some earlier works on parametric nonlinear time series analysis, (see Tong and Lim [12] and Tong [13]), where one of the oldest classes of parametric nonlinear time series models, namely, threshold autoregressive (TAR) models, was introduced. An important subclass of TAR models is the self-exciting threshold autoregressive (SETAR) models, where regime switches of a time series depend on the past values of the time series. Hamilton [5] popularized the applications of Markovian regime-switching autoregressive (MRSAR) models in econometrics and economics, which may be considered a type of TAR models, (see Tong [13]).

In this paper, we introduce a double threshold model, where there are two kinds of regime switches, one modulated by a hidden Markov chain and another one determined by a self-exciting principle. The double threshold model is highly relevant for engineering applications. It incorporates both the class of self-exciting threshold models and the class of hidden Markov models. These two classes of models play significant roles in a number of important problems in engineering and physical sciences. Applications of hidden Markov models in engineering, in particular, electrical engineering, are well-known and were discussed in the monograph by Elliott *et al.* [2]. The applications of threshold models in various fields including engineering and physical sciences were discussed in the monograph by Tong [14]. In Section 2.3, Chapter 2 of the monograph by Tong [14], an example in electrical engineering was used to illustrate the motivation of the class of threshold models. In this example, a triode value, which is a thermionic value with three electrodes, namely a cathode, an anode and a grid, was used to illustrate the use of piecewise linearization to explain one of the key ideas in the theory of nonlinear oscillations, namely limit cycles. A piecewise linear diferential equation was used to approximate the true characteristic of the triode. In fact, a precursor of the threshold model in discrete time is a piecewise linear differential equation. Like a piecewise linear differential equation, one of the key motivations to introduce a threshold model in discrete time is to describe and explain limit cycles which not only manifest themselves in (electrical) engineering, but also in economics, where the notion of business cycles plays a vital role. Using an approach based on a reference probability and a version of the Bayes' rule, we derive recursive filters for the hidden Markov chain and some related quantities used in the Expectation Maximization, (EM), algorithm. There are at least three major contributions of the present paper. Firstly, we introduce a new model which is flexible enough to nest two important classes of models, namely the self-exciting threshold model and the hidden Markov models. Secondly, we introduce a theoretically sound method to filter and estimate the proposed model based on a reference probability approach. Last but not least, we introduce a new and practically useful approach to implement the filters and filter-based estimates. This new approach is based on a set of new recursive equations which grow much slower than the original filtering equations derived from the reference probability approach. It can avoid the overflow in the implementation of the original filtering equations and significantly speed up the computation of the filtering equations. Indeed these two problems are major issues which signficantly hinder the practical implementations and applications of many existing filtering methods developed using the reference probability approach.

Several good reviews about the EM algorithm do exist, see, for example, Meng & Rubin [9], Meng & van Dyk [10] and McLachlan & Krishnan [7]. There are different types of the EM algorithms. The EM algorithm considered here is a variant of previous adaptations. As usual, there are two steps. The E step computes the expectation of the log-likelihood ratio of the parameters in the previous iteration and the unknown parameters, that is $Q(\theta, \hat{\theta}) = E_{\hat{\theta}_k}[\log(dP_\theta / dP_{\hat{\theta}_k})]$ where $\hat{\theta}_k$ is the parameters in iteration k, θ is the unknown parameters vector. The M step concerns the maximization of the function $Q(\theta, \hat{\theta})$ with respect to θ and this results in obtaining a new set of parameters in iteration $k+1$, that is θ_{k+1}. In our case, the parameters of interest are multidimensional and this step can be decomposed into several separate maximizations over different components of the θ vector. Indeed, the EM algorithm considered here has a close connection to the ECM, AECM and SAGE algorithms described in Meng & Rubin [9], Meng & van Dyk [10], and Fesser & Hero [3] respectively. The convergence of the EM algorithm relies on the theory established in Meng & Rubin [9] and Fesser & Hero [3]. Particularly, as in the ordinal EM algorithm, the EM algorithm considered here and its variants described above do not guarantee convergence to a global maximum unless their objective functions are well-behaved in the M step.

The rest of this paper is organized as follows. In the next section, we give an introduction the double threshold model. In Section III, we

Manuscript received October 25, 2011; revised August 17, 2012; accepted April 10, 2013. Date of publication May 01, 2013; date of current version November 18, 2013. R. J. Elliott and T. K. Siu wish to acknowledge the Discovery Grant from the Australian Research Council (ARC) under Project DP1096243. Recommended by Associate Editor A. Chiuso.

R. J. Elliott is with the School of Mathematical Sciences, University of Adelaide, Adelaide, SA 5005, Australia, and also with the Haskayne School of Business, University of Calgary, Calgary, Alberta, T2N 1N4 Canada (e-mail: relliott@ucalgary.ca).

T. K. Siu is with Cass Business School, City University London, London, EC1Y 8TZ, U.K., and also with the Department of Applied Finance and Actuarial Studies, Faculty of Business and Economics, Macquarie University, Sydney, NSW 2109, Australia (e-mail: Ken.Siu.1@city.ac.uk; ktksiu2005@gmail.com).

J. W. Lau is with Department of Mathematics and Statistics, University of Western Australia, Perth, WA 2009, Australia (e-mail: john.w.lau@gmail.com).

Digital Object Identifier 10.1109/TAC.2013.2261186

first present a measure change that relates the reference probability to a real-world one. Then we derive the filtering equations based on the EM algorithm are derived in Section IV. The final section gives a summary. The proofs of the results are standard, so we only state the results.

II. The Double Threshold Model

Suppose $(\Omega, \mathcal{F}, P)$ is a complete probability space, where P is the real-world probability measure. Let $\{\mathbf{X}_t | t \in \mathcal{T}\}$ be a discrete-time, N-state, hidden Markov chain on $(\Omega, \mathcal{F}, P)$ with state space $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_N\}$, the set of standard unit vectors. Here $\mathcal{T} := \{0, 1, \cdots\}$ and the j^{th} component of $\mathbf{e}_i$ is the Kronecker delta function δ_{ij}, for each $i, j = 1, 2, \cdots, N$.

For each $i, j = 1, 2, \cdots, N$, let

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

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

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

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

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.

Then Elliott *et al.* [2] obtained the following semimartingale dynamics for the chain $\mathbf{X}$ under P:

$$\mathbf{X}_{t+1} = \mathbf{\Pi}\mathbf{X}_t + \mathbf{M}_{t+1}, \quad t \in \mathcal{T}, \tag{2}$$

where $\{\mathbf{M}_t | t \in \mathcal{T} \setminus \{0\}\}$ is an $\Re^N$-valued, $(\mathbb{F}^{\mathbf{X}}, P)$-martingale difference process.

We now describe the double threshold model. To simplify our discission and notation, we consider a simple case where there are two regimes in the self-exciting threshold part of the double threshold model. The filtering and estimation methods developed in this paper can be applied to the general case that there are M regimes in the self-exciting threshold part of the model.

For each $i = 1, 2$ and $t \in \mathcal{T}$, let

$$\widetilde{\mu}_{i,t} := \langle \boldsymbol{\mu}_i, \mathbf{X}_t \rangle, \quad \text{and} \quad \widetilde{\sigma}_{i,t} := \langle \boldsymbol{\sigma}_i, \mathbf{X}_t \rangle. \tag{3}$$

Here $\boldsymbol{\mu}_i := (\mu_{i1}, \mu_{i2}, \cdots, \mu_{iN})' \in \Re^N$ and $\boldsymbol{\sigma}_i := (\sigma_{i1}, \sigma_{i2}, \cdots, \sigma_{iN})' \in \Re^N$ with $\sigma_{ij} > 0$ for each $j = 1, 2, \cdots, N$; $\langle \cdot, \cdot \rangle$ is the scalar product in $\Re^N$.

Suppose r is the threshold parameter and d is the delay parameter, where r is a real number and d is a positive integer. Consider a time series $\{Y_t | t = -d+1, -d+2, \cdots, 0, 1, \cdots\}$, where the initial values $Y_{-d+1}, Y_{-d+2}, \cdots, Y_0$ are given in advance. We suppose the remaining terms of the time series $\{Y_t | t \in \mathcal{T} \setminus \{0\}\}$ follow an $(2, n)$-order double threshold model defined by:

$$\begin{aligned} Y_t = (\widetilde{\mu}_{1,t-1} + \widetilde{\sigma}_{1,t-1}\epsilon_t)&\left(1 - I_{\{Y_{t-d} > r\}}\right) \\ &+ (\widetilde{\mu}_{2,t-1} + \widetilde{\sigma}_{2,t-1}\epsilon_t) I_{\{Y_{t-d} > r\}}. \end{aligned} \tag{4}$$

Here we suppose, for simplicity, that the noise process $\{\epsilon_t | t \in \mathcal{T} \setminus \{0\}\}$ is a sequence of i.i.d. standard Gaussian random variables, (i.e., $\epsilon_t \sim N(0,1)$); $I_{\{Y_{t-d} > r\}}$ is the indicator function of the event $\{Y_{t-d} > r\}$. In general, we may enrich the above double threshold model by considering the situation where the autoregressive terms of the time series $\{Y_t | t \in \mathcal{T} \setminus \{0\}\}$ and the explanatory variables are included in the conditional mean terms $\widetilde{\mu}_{1,t-1}$ and $\widetilde{\mu}_{2,t-1}$. Furthermore, the coefficients of these autoregressive terms and the explanatory variables may depend on the "state" dependent threshold as well as the hidden Markov chain. This generalized model encompasses both the SETAR model and the MRSAR model.[1]

The reference probability approach we consider in the later sections of this paper may be used to develop filters and filter-based estimates based on the EM algorithm for this generalized version of the double threshold model. However, the notation and filtering equations could be complicated in this generalized situation. Also, a challenging problem when one considers the generalized model is model selection. In particular, one may wish to determine the order of the autoregressive part, the order of the hidden Markov chain and what explanatory variables to be included. A possible way to deal with these issues is to adopt some statistical methods for model selection such as the Akaike information criterion (AIC) and the Bayesian Information Criterion (BIC). However, the basic notions of the AIC and the BIC are based on likelihood functions which can be completely specified when there is no hidden information in the model. Due to the presence of the hidden Markov chain in our double threshold model, the AIC and BIC cannot be directly used for model selection. They need to be generalized by incorporating the presence of hidden information. A possible way to articulate the problem is to consider a generalized version of the AIC or the BIC based on a conditional expectation of a likelihood function. This situation is not unlike the use of the conditional expectation of a likelihood function in the EM algorithm. This may represent an interesting topic for further investigation. Here to simplify our discussion and to emphasize the key idea of our proposed model, we focus on developing the filters and the filter-based estimates for the basic double threshold model using the reference probability approach.

The double threshold model may entail some financial interpretations. For example, we can consider $\{Y_t | t \in \mathcal{T} \setminus \{0\}\}$ (4) as the return series of a financial asset and the hidden Markov chain $\mathbf{X}$ (2) as the hidden state of an economy. The self-exciting regime switches can incorporate the feedback effect of a past return of the asset on the expected return and volatility of the asset. For example, we may consider the threshold parameter r as a floor return level below which market participants expect that the future return of the asset will drop and the volatility of the asset will become higher. Another interesting case is when the threshold parameter $r = 0$ and the delay parameter $d = 1$. In this case, we can differentiate the impact of a positive current return of a financial asset on its future return dynamics from that of a negative current return.

III. Filters for Hidden Quantities

We first start with a reference probability $\bar{P}$ under which

1) $\{Y_t | t \in \mathcal{T} \setminus \{0\}\}$ (4) is a sequence of i.i.d. standard normal random variables;
2) the chain $\mathbf{X}$ (2) has a transition probability matrix $\mathbf{\Pi}$.

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:

$$\begin{aligned} \mathcal{Y}_t^0 &:= \sigma\{Y_1, Y_2, \cdots, Y_t\} \vee \mathcal{Y}_0^0, \\ \mathcal{G}_t^0 &:= \sigma\{Y_1, Y_2, \cdots, Y_t, \mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_t\} \vee \mathcal{G}_0^0, \\ &\quad t = 1, 2, \cdots, \quad \text{and} \\ \mathcal{Y}_0^0 &:= \sigma\{Y_{-d+1}, Y_{-d+2}, \cdots, Y_0\}, \\ \mathcal{G}_0^0 &:= \sigma\{\mathbf{X}_0\} \vee \mathcal{Y}_0^0, \end{aligned}$$

where $\mathcal{A} \vee \mathcal{B}$ is the minimal σ-field containing both $\mathcal{A}$ and $\mathcal{B}$; the σ-field $\sigma\{X, Y\}$ is the σ-field generated by two random variables X and Y.

[1]We would like to thank the Associate Editor for pointing out this interesting issue.

Write $\mathbb{Y}$ and $\mathbb{G}$ for the P-completion of $\mathbb{Y}^0$ and $\mathbb{G}^0$, respectively, so that $\mathbb{Y} := \{\mathcal{Y}_t | t \in \mathcal{T}\}$ and $\mathbb{G} := \{\mathcal{G}_t | t \in \mathcal{T}\}$, where

$$\mathcal{Y}_t := \mathcal{Y}_t^0 \vee \mathcal{N}, \quad \mathcal{G}_t := \mathcal{G}_t^0 \vee \mathcal{N},$$

with $\mathcal{N}$ being the collection of P-null sets in $\mathcal{F}$.

Consider now a $\mathbb{G}$-adapted process $\{\lambda_t | t \in \mathcal{T} \setminus \{0\}\}$ by:

$$\begin{aligned}\lambda_t &:= \frac{\phi\left(\frac{Y_t - \widetilde{\mu}_{1,t-1}}{\widetilde{\sigma}_{1,t-1}}\right)}{\widetilde{\sigma}_{1,t-1}\phi(Y_t)}\left(1 - I_{\{Y_{t-d}>r\}}\right) \\ &\quad + \frac{\phi\left(\frac{Y_t - \widetilde{\mu}_{2,t-1}}{\widetilde{\sigma}_{2,t-1}}\right)}{\widetilde{\sigma}_{2,t-1}\phi(Y_t)} I_{\{Y_{t-d}>r\}}, \\ &= \lambda_{t,1}\left(1 - I_{\{Y_{t-d}>r\}}\right) + \lambda_{t,2} I_{\{Y_{t-d}>r\}},\end{aligned} \tag{5}$$

where $\phi(z)$ is the density function of $N(0,1)$ and is given by: $\phi(z) := (1/\sqrt{2\pi})e^{-(z^2/2)}$.

Define another $\mathbb{G}$-adapted process $\{\Lambda_t | t \in \mathcal{T}\}$ by:

$$\Lambda_t := \prod_{i=1}^{t} \lambda_i, \quad t \in \mathcal{T} \setminus \{0\}, \quad \text{and } \Lambda_0 := 1. \tag{6}$$

It is not difficult to see that $\{\Lambda_t | t \in \mathcal{T}\}$ is a $(\mathbb{G}, \bar{P})$-martingale.

The following theorem gives the probability laws of the process ϵ and the chain $\mathbf{X}$ under P.

Theorem 3.1: Under P,

$$\begin{aligned}\epsilon_t &:= \left(\frac{Y_t - \widetilde{\mu}_{1,t-1}}{\widetilde{\sigma}_{1,t-1}}\right)\left(1 - I_{\{Y_{t-d}>r\}}\right) \\ &\quad + \left(\frac{Y_t - \widetilde{\mu}_{2,t-1}}{\widetilde{\sigma}_{2,t-1}}\right) I_{\{Y_{t-d}>r\}}, \quad t \in \mathcal{T} \setminus \{0\}, \\ &= \epsilon_{t,1}\left(1 - I_{\{Y_{t-d}>r\}}\right) + \epsilon_{t,2} I_{\{Y_{t-d}>r\}}, \quad \text{say},\end{aligned} \tag{7}$$

is a sequence of $N(0,1)$, i.i.d., random variables and the chain $\mathbf{X}$ has the rate matrix $\mathbf{\Pi}$.

Using the result in Theorem 3.1, we have the following corollary.

Corollary 3.1: Under P,

$$\begin{aligned}Y_t = (\widetilde{\mu}_{1,t-1} + \widetilde{\sigma}_{1,t-1}\epsilon_t)\left(1 - I_{\{Y_{t-d}>r\}}\right) \\ + (\widetilde{\mu}_{2,t-1} + \widetilde{\sigma}_{2,t-1}\epsilon_t) I_{\{Y_{t-d}<r\}},\end{aligned} \tag{8}$$

and

$$\mathbf{X}_t = \mathbf{\Pi}\mathbf{X}_{t-1} + \mathbf{M}_t. \tag{9}$$

This can be viewed as a state-space form of the double threshold model.

We now derive a recursive filter for the hidden Markov chain. As usual, we assume that $\mathbf{X}_0$ is known. Given $\mathcal{Y}_t$, we wish to estimate $\mathbf{X}_t$ as follows:

$$\widehat{\mathbf{X}}_t = \mathrm{E}[\mathbf{X}_t | \mathcal{Y}_t]. \tag{10}$$

Note that $\widehat{\mathbf{X}}_t$ is an optimal estimate of $\mathbf{X}_t$ given $\mathcal{Y}_t$ in the mean-square sense.

By a version of the Bayes' rule,

$$\mathrm{E}[\mathbf{X}_t | \mathcal{Y}_t] = \frac{\bar{\mathrm{E}}[\Lambda_t \mathbf{X}_t | \mathcal{Y}_t]}{\bar{\mathrm{E}}[\Lambda_t | \mathcal{Y}_t]}. \tag{11}$$

Define an unnormalized filter of $\mathbf{X}_t$ as follows:

$$\mathbf{q}_t := \bar{\mathrm{E}}[\Lambda_t \mathbf{X}_t | \mathcal{Y}_t]. \tag{12}$$

Note that $\mathbf{X}_0$ is assumed given, so $\mathbf{q}_0 = \mathbf{p}_0 = \mathbf{X}_0$, where $\mathbf{p}_0$ is the initial distribution of the chain $\mathbf{X}$.

The following theorem gives an exact recursive equation for the unnormalized filter $\mathbf{q}_t$.

Theorem 3.2: For each $j = 1, 2, \cdots, N$, let

$$\begin{aligned}\zeta_j(t, Y) := \frac{\phi\left(\frac{Y_t - \mu_{1j}}{\sigma_{1j}}\right)}{\sigma_{1j}\phi(Y_t)}\left(1 - I_{\{Y_{t-d}>r\}}\right) \\ + \frac{\phi\left(\frac{Y_t - \mu_{2j}}{\sigma_{2j}}\right)}{\sigma_{2j}\phi(Y_t)} I_{\{Y_{t-d}>r\}}.\end{aligned} \tag{13}$$

Write $\boldsymbol{\zeta}(t, Y) := (\zeta_1(t, Y), \zeta_2(t, Y), \cdots, \zeta_N(t, N))' \in \Re^N$, and $\mathbf{diag}(\boldsymbol{\zeta}(t, Y))$ for the diagonal matrix with diagonal elements being $\boldsymbol{\zeta}(t, Y)$. Then $\mathbf{q}_t$ satisfies the following exact recursive equation:

$$\mathbf{q}_t = \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t, Y)\right)\mathbf{q}_{t-1}, \quad t \in \mathcal{T} \setminus \{0\}. \tag{14}$$

Note that $\langle \mathbf{q}_t, \mathbf{1} \rangle = \bar{\mathrm{E}}[\Lambda_t \langle \mathbf{X}_t, \mathbf{1} \rangle | \mathcal{Y}_t] = \bar{\mathrm{E}}[\Lambda_t | \mathcal{Y}_t]$, so $\mathrm{E}[\mathbf{X}_t | \mathcal{Y}_t] = (\mathbf{q}_t / \langle \mathbf{q}_t, \mathbf{1} \rangle)$. Here $\mathbf{1} := (1, 1, \cdots, 1)' \in \Re^N$.

Consider the following quantities related to the hidden Markov chain $\mathbf{X}$:

1) The number of transitions of the chain $\mathbf{X}$ from state $\mathbf{e}_j$ to state $\mathbf{e}_k$ up to and including time t, for each $j, k = 1, 2, \cdots, N$:

$$N_t^{[k,j]} := \sum_{l=1}^{t} \langle \mathbf{X}_{l-1}, \mathbf{e}_j \rangle \langle \mathbf{X}_l, \mathbf{e}_k \rangle. \tag{15}$$

2) The occupation time of the chain $\mathbf{X}$ in state $\mathbf{e}_j$ up to and including time t, for each $j = 1, 2, \cdots, N$:

$$O_t^{[j]} := \sum_{l=1}^{t} \langle \mathbf{X}_{l-1}, \mathbf{e}_j \rangle. \tag{16}$$

3) The "generalized" occupation times of the chain $\mathbf{X}$ in state $\mathbf{e}_j$ up to and including time t, for each $j = 1, 2, \cdots, N$:

$$O_{t,1}^{[j]} := \sum_{l=1}^{t} \langle \mathbf{X}_{l-1}, \mathbf{e}_j \rangle \left(1 - I_{\{Y_{l-d}>r\}}\right), \text{ and} \tag{17}$$

$$O_{t,2}^{[j]} := \sum_{l=1}^{t} \langle \mathbf{X}_{l-1}, \mathbf{e}_j \rangle I_{\{Y_{l-d}>r\}}. \tag{18}$$

4) The "generalized" level processes associated with state $\mathbf{e}_j$ up to and including time t, for each $j = 1, 2, \cdots, N$:

$$L_{t,1}^{[j]}(f) := \sum_{l=1}^{t} f(Y_l) \langle \mathbf{X}_{l-1}, \mathbf{e}_j \rangle \left(1 - I_{\{Y_{l-d}>r\}}\right), \text{ and} \tag{19}$$

$$L_{t,2}^{[j]}(f) := \sum_{l=1}^{t} f(Y_l) \langle \mathbf{X}_{l-1}, \mathbf{e}_j \rangle I_{\{Y_{l-d}>r\}}. \tag{20}$$

Here $f : \Re \to \Re$ is any Borel-measurable test function. In Section IV, we shall suppose that $f(Y_l) = Y_l$ or $f(Y_l) = Y_l^2$.

The recursive filters of these quantities will be used to derive the filter-based estimates of the model parameters of the double threshold model.

For each $K_t := N_t^{[k,j]}, O_t^{[j]}, O_{t,i}^{[j]}$, or $L_{t,i}^{[j]}(f)$, $i = 1, 2$, $j = 1, 2, \cdots, N$, we define the following "unnormalized" quantity:

$$\sigma(K_t \mathbf{X}_t) := \bar{\mathrm{E}}[\Lambda_t K_t \mathbf{X}_t | \mathcal{Y}_t], \quad t \in \mathcal{T} \setminus \{0\}, \tag{21}$$

where, for each $t = -d+1, -d+2, \cdots, 0$,

$$\sigma(K_t \mathbf{X}_t) = \mathbf{0} \in \Re^N. \tag{22}$$

The following theorem gives a set of exact filtering equations which will be used for the filter-based estimates in the Expectation Maximization (EM) algorithm.

Theorem 3.3: For each $t \in \mathcal{T} \setminus \{0\}$ and $j, k = 1, 2, \cdots, N$ with $j \neq k$,

$$\begin{aligned}
\sigma\left(N_t^{[k,j]}\mathbf{X}_t\right) &= \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\sigma\left(N_{t-1}^{[k,j]}\mathbf{X}_{t-1}\right) \\
&\quad + \langle \boldsymbol{\zeta}(t,Y), \mathbf{e}_j\rangle \langle \mathbf{q}_{t-1}, \mathbf{e}_j\rangle \pi_{kj}\mathbf{e}_k, \\
\sigma\left(O_t^{[j]}\mathbf{X}_t\right) &= \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\sigma\left(O_{t-1}^{[j]}\mathbf{X}_{t-1}\right) \\
&\quad + \langle \mathbf{q}_{t-1}, \mathbf{e}_j\rangle \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\mathbf{e}_j, \\
\sigma\left(O_{t,1}^{[j]}\mathbf{X}_t\right) &= \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\sigma\left(O_{t-1,1}^{[j]}\mathbf{X}_{t-1}\right) \\
&\quad + (1 - I_{\{Y_{t-d}>r\}})\langle \mathbf{q}_{t-1}, \mathbf{e}_j\rangle \\
&\quad \times \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\mathbf{e}_j, \\
\sigma\left(O_{t,2}^{[j]}\mathbf{X}_t\right) &= \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\sigma\left(O_{t-1,2}^{[j]}\mathbf{X}_{t-1}\right) \\
&\quad + I_{\{Y_{t-d}>r\}}\langle \mathbf{q}_{t-1}, \mathbf{e}_j\rangle \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\mathbf{e}_j, \\
\sigma\left(L_{t,1}^{[j]}(f)\mathbf{X}_t\right) &= \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\sigma\left(L_{t-1,1}^{[j]}(f)\mathbf{X}_{t-1}\right) \\
&\quad + f(Y_t)(1 - I_{\{Y_{t-d}>r\}})\langle \mathbf{q}_{t-1}, \mathbf{e}_j\rangle \\
&\quad \times \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\mathbf{e}_j, \\
\sigma\left(L_{t,2}^{[j]}(f)\mathbf{X}_t\right) &= \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\sigma\left(L_{t-1,2}^{[j]}(f)\mathbf{X}_{t-1}\right) \\
&\quad + f(Y_t) I_{\{Y_{t-d}>r\}}\langle \mathbf{q}_{t-1}, \mathbf{e}_j\rangle \\
&\quad \times \mathbf{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t,Y)\right)\mathbf{e}_j.
\end{aligned}$$

where π_{kj} is the kj entry of the matrix $\mathbf{\Pi}$.

IV. Filter-Based Estimates and the EM Algorithm

In this section we derive filter-based estimates for the model parameters in the double threshold model and develop the corresponding EM algorithm. Firstly, we suppose that the threshold parameter r is given. For a given threshold parameter r, the set of parameters of interest is described by the set Θ_r defined as follows:

$$\mathbf{\Theta}_r := \{(\pi_{kl})_{k,l=1,2,\cdots,N}, (\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i), \quad i = 1, 2\} \tag{23}$$

where $\boldsymbol{\mu}_i := (\mu_{i1}, \mu_{i2}, \cdots, \mu_{iN})' \in \Re^N$ and $\boldsymbol{\sigma}_i := (\sigma_{i1}, \sigma_{i2}, \cdots, \sigma_{iN})' \in \Re^N$. To simplify the notation, we suppress the notation "r" and write $\mathbf{\Theta}$ for $\mathbf{\Theta}_r$.

Suppose now that the set of parameters $\mathbf{\Theta}$ is given and that the set of observed data $\{Y_1, Y_2, \cdots, Y_n\}$ is known. We wish to determine a new set of parameters $\mathbf{\Theta}(n)$ defined by:

$$\mathbf{\Theta}(n) := \left\{(\pi_{kl}(n))_{k,l=1,2,\cdots,N}, (\boldsymbol{\mu}_i(n), \boldsymbol{\sigma}_i(n)), \quad i = 1, 2\right\} \tag{24}$$

where $\boldsymbol{\mu}_i(n) := (\mu_{i1}(n), \mu_{i2}(n), \cdots, \mu_{iN}(n))'$ and $\boldsymbol{\sigma}_i(n) := (\sigma_{i1}(n), \sigma_{i2}(n), \cdots, \sigma_{iN}(n))'$.

Such a new set of parameters is determined by maximizing the conditional log-likelihoods defined below using all of the observed data $\{Y_1, Y_2, \cdots, Y_n\}$. The key idea is to update one set of parameters at a time. Here we start with $[\pi_{kl}]_{k,l=1,2,\cdots,N}$. Using this method, Elliott *et al.* [2], (see Chapter 2 therein), showed that a filter-based estimate $\widehat{\pi}_{kl}(n)$ for π_{kl} given observed data $\{Y_1, Y_2, \cdots, Y_n\}$ is given by:

$$\begin{aligned}
\widehat{\pi}_{kl}(n) &= \frac{\mathrm{E}\left[N_n^{[k,l]}|\mathcal{Y}_n\right]}{\mathrm{E}\left[O_n^{[k]}|\mathcal{Y}_n\right]} = \frac{\sigma\left(N_n^{[k,l]}\right)}{\sigma\left(O_n^{[k]}\right)} \\
&= \frac{\left\langle \sigma\left(N_t^{[k,j]}\mathbf{X}_t\right), \mathbf{1}\right\rangle}{\left\langle \sigma\left(O_t^{[k]}\mathbf{X}_t\right), \mathbf{1}\right\rangle}.
\end{aligned} \tag{25}$$

To simplify the notation, for each $i = 1, 2$, we write $\widehat{\boldsymbol{\mu}}_i$ and $\widehat{\boldsymbol{\sigma}}_i$ for $\widehat{\boldsymbol{\mu}}_i(n)$ and $\widehat{\boldsymbol{\sigma}}_i(n)$, respectively.

To change the set of parameters from $\boldsymbol{\mu}_i$ to $\widehat{\boldsymbol{\mu}}_i(n)$, $i = 1, 2$, we must consider the following factors, $(t = 1, 2, \cdots, n)$,

$$\begin{aligned}
\widehat{\lambda}_t(1) := &\exp\left[\frac{\langle \boldsymbol{\mu}_1, \mathbf{X}_{t-1}\rangle^2 - \langle \widehat{\boldsymbol{\mu}}_1, \mathbf{X}_{t-1}\rangle^2}{2\langle \boldsymbol{\sigma}_1, \mathbf{X}_{t-1}\rangle^2}\right] \\
&\times \exp\left[\frac{-2Y_t\langle \boldsymbol{\mu}_1, \mathbf{X}_{t-1}\rangle + 2Y_t\langle \widehat{\boldsymbol{\mu}}_1, \mathbf{X}_{t-1}\rangle}{2\langle \boldsymbol{\sigma}_1, \mathbf{X}_{t-1}\rangle^2}\right] \\
&\times \left(1 - I_{\{Y_{t-d}>r\}}\right) \\
&+ \exp\left[\frac{\langle \boldsymbol{\mu}_2, \mathbf{X}_{t-1}\rangle^2 - \langle \widehat{\boldsymbol{\mu}}_2, \mathbf{X}_{t-1}\rangle^2}{2\langle \boldsymbol{\sigma}_2, \mathbf{X}_{t-1}\rangle^2}\right] \\
&\times \exp\left[\frac{-2Y_t\langle \boldsymbol{\mu}_2, \mathbf{X}_{t-1}\rangle + 2Y_t\langle \widehat{\boldsymbol{\mu}}_2, \mathbf{X}_{t-1}\rangle}{2\langle \boldsymbol{\sigma}_2, \mathbf{X}_{t-1}\rangle^2}\right] \\
&\times I_{\{Y_{t-d}>r\}}.
\end{aligned}$$

Write, for each $t = 1, 2, \cdots, n$,

$$\widehat{\Lambda}_t(1) := \prod_{u=1}^{t} \widehat{\lambda}_u(1), \quad \text{and} \quad \widehat{\Lambda}_0(1) := 1.$$

It is easy to check that $\{\widehat{\Lambda}_t(1)|t = 1, 2, \cdots, n\}$ is a $(\mathbb{G}, P)$-martingale.

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

$$\left.\frac{d\widehat{P}^1}{dP}\right|_{\mathcal{G}_n} = \widehat{\Lambda}_n(1). \tag{26}$$

Then it is not difficult to show that

$$\begin{aligned}
&\mathrm{E}\left[\ln \widehat{\Lambda}_n(1)|\mathcal{Y}_n\right] \\
&= \sum_{j=1}^{N}\left[\frac{1}{2\sigma_{1j}^2}\left(-\widehat{\mu}_{1j}^2\widehat{O}_{n,1}^{[j]} + 2\widehat{\mu}_{1j}\widehat{L}_{n,1}^{[j]}(Y)\right)\right. \\
&\qquad \left. + \frac{1}{2\sigma_{2j}^2}\left(-\widehat{\mu}_{2j}^2\widehat{O}_{n,2}^{[j]} + 2\widehat{\mu}_{2j}\widehat{L}_{n,2}^{[j]}(Y)\right)\right] \\
&\quad + R(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, r)
\end{aligned}$$

where $R(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, r)$ does not involve $\widehat{\boldsymbol{\mu}}_1$ and $\widehat{\boldsymbol{\mu}}_2$.

The first-order condition for maximizing the above conditional log-likelihood with respect to $\widehat{\mu}_{1j}$, $j = 1, 2, \cdots, N$, gives:

$$\widehat{\mu}_{1j} = \frac{\widehat{L}_{n,1}^{[j]}(Y)}{\widehat{O}_{n,1}^{[j]}} = \frac{\sigma\left(L_{n,1}^{[j]}(Y)\right)}{\sigma\left(O_{n,1}^{[j]}\right)}. \tag{27}$$

Likewise, the first-order condition for maximizing the conditional log-likelihood with respect to $\widehat{\mu}_{2j}$, $j = 1, 2, \cdots, N$, gives:

$$\widehat{\mu}_{2j} = \frac{\widehat{L}_{n,2}^{[j]}(Y)}{\widehat{O}_{n,2}^{[j]}} = \frac{\sigma\left(L_{n,2}^{[j]}(Y)\right)}{\sigma\left(O_{n,2}^{[j]}\right)}. \tag{28}$$

To change the set of parameters from $\boldsymbol{\sigma}_i$ to $\widehat{\boldsymbol{\sigma}}_i$, $i = 1, 2$, we consider the following factors, $(t = 1, 2, \cdots, n)$,

$$\begin{aligned}\widehat{\lambda}_t(2) :=& \frac{\langle \boldsymbol{\sigma}_1, \mathbf{X}_{t-1}\rangle}{\langle \widehat{\boldsymbol{\sigma}}_1, \mathbf{X}_{t-1}\rangle} \frac{\exp\left[-\frac{(Y_t - \langle \boldsymbol{\mu}_1, \mathbf{X}_{t-1}\rangle)^2}{2\langle \widehat{\boldsymbol{\sigma}}_1, \mathbf{X}_{t-1}\rangle^2}\right]}{\exp\left[-\frac{(Y_t - \langle \boldsymbol{\mu}_1, \mathbf{X}_{t-1}\rangle)^2}{2\langle \boldsymbol{\sigma}_1, \mathbf{X}_{t-1}\rangle^2}\right]} \\ &\times \left(1 - I_{\{Y_{t-d} > r\}}\right) \\ &+ \frac{\langle \boldsymbol{\sigma}_2, \mathbf{X}_{t-1}\rangle}{\langle \widehat{\boldsymbol{\sigma}}_2, \mathbf{X}_{t-1}\rangle} \frac{\exp\left[-\frac{(Y_t - \langle \boldsymbol{\mu}_2, \mathbf{X}_{t-1}\rangle)^2}{2\langle \widehat{\boldsymbol{\sigma}}_2, \mathbf{X}_{t-1}\rangle^2}\right]}{\exp\left[-\frac{(Y_t - \langle \boldsymbol{\mu}_2, \mathbf{X}_{t-1}\rangle)^2}{2\langle \boldsymbol{\sigma}_2, \mathbf{X}_{t-1}\rangle^2}\right]} \\ &\times I_{\{Y_{t-d} > r\}}. \end{aligned} \tag{29}$$

Again, we write:

$$\widehat{\Lambda}_t(2) := \prod_{u=1}^{t} \widehat{\lambda}_u(2), \quad \text{and} \quad \widehat{\Lambda}_0(2) := 1. \tag{30}$$

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

$$\left.\frac{d\widehat{P}^2}{dP}\right|_{\mathcal{G}_n} := \widehat{\Lambda}_n(2). \tag{31}$$

Similarly, it can be shown that

$$\begin{aligned} &\mathrm{E}\left[\ln \widehat{\Lambda}_n^2 | \mathcal{Y}_n\right] \\ &= \sum_{i=1}^{N} \left[-\frac{1}{2} \ln \widehat{\sigma}_{1j}^2 \widehat{O}_{n,1}^{[j]} \right. \\ &\quad - \frac{1}{2\widehat{\sigma}_{1j}^2} \left(\widehat{L}_{n,1}^{[j]}(Y^2) - 2\widehat{\mu}_{1j} \widehat{L}_{n,1}^{[j]}(Y) + \widehat{\mu}_{1j}^2 \widehat{O}_{n,1}^{[j]} \right) \\ &\quad - \frac{1}{2} \ln \widehat{\sigma}_{2j}^2 \widehat{O}_{n,2}^{[j]} \\ &\quad \left. - \frac{1}{2\widehat{\sigma}_{2j}^2} \left(\widehat{L}_{n,2}^{[j]}(Y^2) - 2\widehat{\mu}_{2j} \widehat{L}_{n,2}^{[j]}(Y) + \widehat{\mu}_{2j}^2 \widehat{O}_{n,2}^{[j]} \right) \right] \\ &\quad + R'(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, r). \end{aligned} \tag{32}$$

where $R'(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, r)$ does not involve $\widehat{\boldsymbol{\sigma}}_1$ and $\widehat{\boldsymbol{\sigma}}_2$.

Again the first-order conditions for maximizing the conditional log-likelihood with respect to $\widehat{\sigma}_{1j}^2$ and $\widehat{\sigma}_{2j}^2$ give:

$$\begin{aligned}\widehat{\sigma}_{ij}^2 &= \frac{\widehat{L}_{n,i}^{[j]}(Y^2) - 2\widehat{\mu}_{ij} \widehat{L}_{n,i}^{[j]}(Y) + \widehat{\mu}_{ij}^2 \widehat{O}_{n,i}^{[j]}}{\widehat{O}_{n,i}^{[j]}} \\ &= \frac{\sigma\left(L_{n,i}^{[j]}(Y^2)\right)}{\sigma\left(O_{n,i}^{[j]}\right)} - \widehat{\mu}_{ij}^2, \quad \text{for } i = 1, 2 \end{aligned} \tag{33}$$

The estimation of μ_{kj}, σ_{kj}^2 and π_{lj} for $k = 1, 2$, and $l, j = 1, 2, \ldots, N$ requires to evaluate $(\sigma(L_{n,k}^{[j]}(Y^2))/\sigma(O_{n,k}^{[j]}))(\sigma(L_{n,k}^{[j]}(Y))/\sigma(O_{n,k}^{[j]}))$ and $\sigma(N_n^{[l,j]})/\sigma(O_n^{[l]})$. Computing the individual terms is still challenging since the overflow problem could happen even though n is less than 200 according to our experience. Here λ_t has a denominator $\phi(Y_t)$ where ϕ represents the standard normal density. If the realization of Y_t is very extreme, say 1000 or -1000, then the term $\phi(Y_t)$ will be very close to zero and eventually λ_t will lead to an overflow. Consequently, adjusting those terms by introducing a common factor could be a possible way to resolve the problem, noting that the estimates are related to the ratios only.

Let C_t be a finite and positive real number, i.e, $C_t \in \Re^+$ for $t \in \mathcal{T} \setminus \{0\}$. Let

$$\overline{\mathbf{q}}_0 = \mathbf{q}_0, \quad \text{and} \quad \overline{\mathbf{q}}_t = C_t \times \boldsymbol{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t, Y)\right) \overline{\mathbf{q}}_{t-1}. \tag{34}$$

A sensible choice of C_t is $1/\langle \boldsymbol{\Pi}\mathbf{diag}(\boldsymbol{\zeta}(t, Y))\overline{\mathbf{q}}_{t-1}, \mathbf{1}\rangle$, so

$$\overline{\mathbf{q}}_t = \frac{\boldsymbol{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t, Y)\right) \overline{\mathbf{q}}_{t-1}}{\left\langle \boldsymbol{\Pi}\mathbf{diag}\left(\boldsymbol{\zeta}(t, Y)\right) \overline{\mathbf{q}}_{t-1}, \mathbf{1}\right\rangle}. \tag{35}$$

This choice may be viewed as a normalization such that $\langle \overline{\mathbf{q}}_t, \mathbf{1}\rangle = 1$.

We introduce a new set of recursive equations. This set of equations grow much slower than those presented above, $\sigma(\cdot \mathbf{X}_t)$, since terms accumulated are all less than 1. Let $\mathbf{A}_t = \boldsymbol{\Pi}\mathbf{diag}(\boldsymbol{\zeta}(t, Y))$,

$$\begin{aligned} &\overline{\sigma}\left(N_t^{[l,j]} \mathbf{X}_t\right) \\ &= C_t \times \left(\mathbf{A}_t \overline{\sigma}\left(N_{t-1}^{[l,j]} \mathbf{X}_{t-1}\right) \right. \\ &\qquad \left. + \langle \boldsymbol{\zeta}(t, Y), \mathbf{e}_j\rangle \langle \overline{\mathbf{q}}_{t-1}, \mathbf{e}_j\rangle \pi_{lj} \mathbf{e}_l \right), \end{aligned} \tag{36}$$

$$\begin{aligned} &\overline{\sigma}\left(O_t^{[j]} \mathbf{X}_t\right) \\ &= C_t \times \left(\mathbf{A}_t \overline{\sigma}\left(O_{t-1}^{[j]} \mathbf{X}_{t-1}\right) + \langle \overline{\mathbf{q}}_{t-1}, \mathbf{e}_j\rangle \mathbf{A}_t \mathbf{e}_j \right), \end{aligned} \tag{37}$$

$$\begin{aligned} &\overline{\sigma}\left(O_{t,1}^{[j]} \mathbf{X}_t\right) \\ &= C_t \times \left(\mathbf{A}_t \overline{\sigma}\left(O_{t-1,1}^{[j]} \mathbf{X}_{t-1}\right) \right. \\ &\qquad \left. + \left(1 - I_{\{Y_{t-d} > r\}}\right) \langle \overline{\mathbf{q}}_{t-1}, \mathbf{e}_j\rangle \mathbf{A}_t \mathbf{e}_j \right), \end{aligned} \tag{38}$$

$$\begin{aligned} &\overline{\sigma}\left(O_{t,2}^{[j]} \mathbf{X}_t\right) \\ &= C_t \times \left(\mathbf{A}_t \overline{\sigma}\left(O_{t-1,2}^{[j]} \mathbf{X}_{t-1}\right) \right. \\ &\qquad \left. + I_{\{Y_{t-d} > r\}} \langle \overline{\mathbf{q}}_{t-1}, \mathbf{e}_j\rangle \mathbf{A}_t \mathbf{e}_j \right), \end{aligned} \tag{39}$$

$$\begin{aligned} &\overline{\sigma}\left(L_{t,1}^{[j]}(f) \mathbf{X}_t\right) \\ &= C_t \times \left(\mathbf{A}_t \overline{\sigma}\left(L_{t-1,1}^{[j]}(f) \mathbf{X}_{t-1}\right) \right. \\ &\qquad \left. + f(Y_t) \left(1 - I_{\{Y_{t-d} > r\}}\right) \langle \overline{\mathbf{q}}_{t-1}, \mathbf{e}_j\rangle \mathbf{A}_t \mathbf{e}_j \right), \end{aligned} \tag{40}$$

$$\begin{aligned} &\overline{\sigma}\left(L_{t,2}^{[j]}(f) \mathbf{X}_t\right) \\ &= C_t \times \left(\mathbf{A}_t \overline{\sigma}\left(L_{t-1,2}^{[j]}(f) \mathbf{X}_{t-1}\right) \right. \\ &\qquad \left. + f(Y_t) I_{\{Y_{t-d} > r\}} \langle \overline{\mathbf{q}}_{t-1}, \mathbf{e}_j\rangle \mathbf{A}_t \mathbf{e}_j \right). \end{aligned} \tag{41}$$

Then it is not difficult to check that

$$\widehat{\pi}_{lj} = \frac{\left\langle \overline{\sigma}\left(N_n^{[l,j]}\right), \mathbf{1}\right\rangle}{\left\langle \overline{\sigma}\left(O_n^{[l]}\right), \mathbf{1}\right\rangle} = \frac{\left\langle \sigma\left(N_n^{[l,j]}\right), \mathbf{1}\right\rangle}{\left\langle \sigma\left(O_n^{[l]}\right), \mathbf{1}\right\rangle}, \tag{42}$$

$$\widehat{\mu}_{kj} = \frac{\left\langle \overline{\sigma}\left(L_{n,k}^{[j]}(Y)\right), \mathbf{1}\right\rangle}{\left\langle \overline{\sigma}\left(O_{n,k}^{[j]}\right), \mathbf{1}\right\rangle} = \frac{\left\langle \sigma\left(L_{n,k}^{[j]}(Y)\right), \mathbf{1}\right\rangle}{\left\langle \sigma\left(O_{n,k}^{[j]}\right), \mathbf{1}\right\rangle}, \tag{43}$$

$$\begin{aligned}\widehat{\sigma}_{kj}^2 &= \frac{\left\langle \overline{\sigma}\left(L_{n,k}^{[j]}(Y^2)\right), \mathbf{1}\right\rangle}{\left\langle \overline{\sigma}\left(O_{n,k}^{[j]}\right), \mathbf{1}\right\rangle} - \widehat{\mu}_{kj}^2 \\ &= \frac{\left\langle \sigma\left(L_{n,k}^{[j]}(Y^2)\right), \mathbf{1}\right\rangle}{\left\langle \sigma\left(O_{n,k}^{[j]}\right), \mathbf{1}\right\rangle} - \widehat{\mu}_{kj}^2. \end{aligned} \tag{44}$$

For each given threshold parameter r, we estimate the unknown parameters Θ_r by going through the following steps of the EM algorithm.

Step I: Select $\widehat{\pi}_{ji}(0)$, $\widehat{\mu}_{kj}(0)$ and $\widehat{\sigma}_{kj}(0)$, $i, j = 1, 2, \cdots, N$ and $k = 1, 2$.
Step II: Using the filtering equations in Theorem 3.3, compute the MLEs, $\widehat{\pi}_{ji}(l+1)$, $\widehat{\mu}_{kj}(l+1)$ and $\widehat{\sigma}_{kj}(l+1)$.
Step III: Stop when $|\widehat{\pi}_{ji}(l+1) - \widehat{\pi}_{ji}(l)| < \epsilon_1$, $|\widehat{\mu}_{kj}(l+1) - \widehat{\mu}_{kj}(l)| < \epsilon_2$ and $|\widehat{\sigma}_{kj}(l+1) - \widehat{\sigma}_{kj}(l)| < \epsilon_3$; otherwise, continue from Step II, where ϵ_1, ϵ_2 and ϵ_3 are the desirable levels of accuracy and $\epsilon_1, \epsilon_2, \epsilon_3 > 0$.

Finally, we consider the estimation of the threshold parameter. For each given threshold parameter r, we obtain the filter-based estimates for the set of unknown parameters described by $\boldsymbol{\Theta}_r$. We define the total conditional log-likelihood evaluated at the filter-based estimates as follows:

$$L(r) := \mathrm{E}\left[\ln \widehat{\Lambda}_n(1)|\mathcal{Y}_n\right]\Big|_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)=(\widehat{\boldsymbol{\mu}}_1,\widehat{\boldsymbol{\mu}}_2)} + \mathrm{E}\left[\ln \widehat{\Lambda}_n(2)|\mathcal{Y}_n\right]\Big|_{(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2)=(\widehat{\boldsymbol{\sigma}}_1,\widehat{\boldsymbol{\sigma}}_2)}. \tag{45}$$

Then an estimate $\widehat{r}$ of the threshold parameter r can be obtained as follows:

$$\widehat{r} := \arg\sup_{r\in\Re} L(r). \tag{46}$$

We shall perform a grid search to find the estimate $\widehat{r}$.

From a practical perspective, one may wish to discuss properties of the estimates and algorithms, such as the error bounds on the estimates and the efficiency of the algorithms. These properties are related to the Cramer-Rao bound, or the Fisher information matrix, which also contributes to the speed of the convergence of ECM and some of its variations (see also Meng [8] and Liu & Rubin [6]). We believe that the same principle based on the reference probability approach as we used to derive the estimates and algorithms also applies to derive the Cramer-Rao bound, or the Fisher information matrix, though it would involve some tedious calculations and complicated notation. The derivations of these quantities and the investigation of their properties may represent interesting topics for future research.

V. Conclusion

New filtering equations and estimation algorithms were developed for a double threshold model which generalizes the self-exciting threshold model and the hidden Markov, regime-switching, threshold model. These new results will hopefully stimulate further developments and applications of the double threshold model in the engineering, statistics and econometrics literature.

Acknowledgment

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

References

[1] R. J. Elliott, "Exact adaptive filters for markov chains observed in gaussian noise," *Automatica*, vol. 30, pp. 1399–1408, 1994.
[2] R. J. Elliott, L. Aggoun, and J. Moore, *Hidden Markov Models: Estimation and Control*. New York: Springer, 1995.
[3] J. A. Fesser and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Processing*, vol. 42, pp. 2664–2677, 1994.
[4] S. M. Goldfeld and R. E. Quandt, "A Markov model for switching regressions," *J. Econometrics*, vol. 1, pp. 3–15, 1973.
[5] J. D. Hamilton, "A new approach to economic analysis of nonstationary time series and the business cycle," *Econometrica*, vol. 57, pp. 357–384, 1989.
[6] C. Liu and D. B. Rubin, "The ECME algorithm A simple extension of EM and ECM with faster monotone convergence," *Biometrika*, vol. 81, pp. 633–648, 1994.
[7] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
[8] X. L. Meng, "On the rate of convergence of the ECM algorithm," *The Annals of Statistics*, vol. 22, pp. 326–339, 1994.
[9] X. L. Meng and D. R. Rubin, "Maximum likelihood estimation via the ECM algorithm: A general framework," *Biometrika*, vol. 80, pp. 267–278, 1993.
[10] X. L. Meng and D. van Dyk, "The EM algorithm, an old folk song sung to afast new tune," *J. Royal Stat. Society: Series B*, vol. 59, pp. 511–567, 1997.
[11] R. E. Quandt, "The estimation of the parameters of a linear regression system obeying two separate regimes," *J. Amer. Stat. Assoc.*, vol. 53, pp. 873–880, 1958.
[12] H. Tong and K. S. Lim, "Threshold autoregression, limit cycles and cyclical data (with discussion)," *J. Royal Stat. Society: Series B*, vol. 42, pp. 245–292, 1980.
[13] H. Tong, *Threshold Models in Non-Linear Time Series Analysis*. New York: Springer, 1983.
[14] H. Tong, *Nonlinear Time Series: A Dynamical System Approach*. Oxford: Oxford Univ Press, 1990.
[15] G. Yin and C. Zhu, *Hybrid Switching Diffusions: Properties and Applications*. New York: Springer, 2010.

Stability Analysis of Quadratic MPC With a Discrete Input Alphabet

Ricardo P. Aguilera, *Member, IEEE*, and
Daniel E. Quevedo, *Member, IEEE*

***Abstract*—We study stability of Model Predictive Control (MPC) with a quadratic cost function for LTI systems with a discrete input alphabet. Since this kind of systems may present a steady-state error, the focus is on practical stability, i.e., ultimate boundedness of solutions. To derive sufficient conditions for practical stability and characterize the ultimately invariant set, we analyze the one-step horizon solution and adapt tools used for convex MPC formulations.**

***Index Terms*—Finite sets, practical stability, predictive control, quantized systems.**

I. Introduction

In a variety of applications, system inputs are restricted to belong to a discrete alphabet of allowed values [1]–[6]. There are several works related to stabilization of this kind of systems, mainly focusing on uniform quantization [7], [8], and logarithmic quantization [9]. However,

Manuscript received February 26, 2012; revised August 30, 2012 and December 09, 2012; accepted April 11, 2013. Date of publication May 21, 2013; date of current version November 18, 2013. This work was supported under the Australian Research Council's Discovery Projects funding scheme (project number DP110103074). Recommended by Associate Editor L. Zaccarian.

The authors are with the School of Electrical Engineering and Computer Science, The University of Newcastle, NSW, Australia (e-mail: raguilera@ieee.org; dquevedo@ieee.org).

Digital Object Identifier 10.1109/TAC.2013.2264551