Journal of Stochastic Analysis

Volume 3 | Number 2 Article 6

June 2022

The Construction and Estimation of Hidden Semi-Markov Models

Kurdstan Abdullah
University of Sulaimani, Sulaymaniyah, Iraq, kurdstan.abdullah@univsul.edu.iq

John van der Hoek
University of South Australia, Adelaide, Australia, john.vanderhoek@unisa.edu.au

Journal of Stochastic Analysis
Vol. 3, No. 2 (2022) Article 6 (22 pages)
DOI: 10.31390/josa.3.2.06

THE CONSTRUCTION AND ESTIMATION OF HIDDEN SEMI-MARKOV MODELS

KURDSTAN ABDULLAH* AND JOHN VAN DER HOEK

Abstract. In this article we construct new formulae and algorithms for Hidden semi-Markov models using Regime Switching Models. We shall include the steps and all necessary lemmas. The formulation of the semi-Markov chain generalizes the one used by Ferguson, by allowing the transition probabilities to be duration dependent. However Ferguson supposed that the transition matrix does not depend on the sojourn times. We assume that the transition matrix does depend on the sojourn time.

1. Introduction

We shall focus on vector autoregressive (VAR) time series. We note that real valued autoregressive time series models can be represented as VAR(1) time series with (one time lag). We include a regime switching feature where the parameters of the VAR(1) model change according to the state of a hidden semi-Markov process. Hidden semi-Markov models were first used in [6], (called variable duration models), in speech recognition. In a hidden semi-Markov models the sojourn times can have general distributions [19].

A hidden semi-Markov chain has also been called a variable-duration hidden Markov chain or an explicit-duration hidden Markov chain [1]. According to [19], hidden semi-Markov models have been used in nearly thirty different areas, such as human activity recognition, handwriting recognition or printed text recognition, rain event time series models, and many others areas. Overcoming the constraint of having geometrically distributed sojourn time distributions is the main advantage of hidden semi-Markov models [18]. The structure of this paper is as follow: Preliminaries with some notation are given in section 2. We then highlighted the formulas of the vector auto-regressive time series model. New algorithms for hidden semi-Markov model joined with vector auto-regressive time series is constructed in section 5. A numerical results of vector auto-regressive with hidden semi-Markov model is presented in section 7 and finally, the Conclusion is in the last section.

Received 2022-1-10; Accepted 2022-6-26; Communicated by the editors.
2020 *Mathematics Subject Classification.* 60G35; 60K15.
Key words and phrases. Discrete-time, semi-Markov models, filters, smoothers, EM algorithms.
* Corresponding author.

2. Some Notation

Let $\{X_t, t = 0, 1, 2, ...\}$ will denote a discrete time unobserved process which is a semi-Markov chain on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in a state space $\mathcal{S} = \{e_1, e_2, ..., e_N\} \subset \mathbb{R}^N$ which consists of the standard unit vectors : $e_i = (0, 0, ..., 1, 0, ...0)^\top$ ($^\top$ denotes transpose) with 1 in the i-th spot for $1 \leq i \leq N$.

$\{Y_t, t = 0, 1, 2, ...\}$ denote an observed process and $\{y_t, t = 0, 1, 2, ...\}$ will denote a sample path of $\{Y_t, t = 0, 1, 2, ...\}$. Write $Y_{0:t} = \{Y_0, Y_1, ..., Y_t\}$ for which we could have observations $y_{0:t} = \{y_0, y_1, ...y_t\} = y_t$, Write $\mathcal{Y}_n = \sigma(Y_0, Y_1, Y_2, ..., Y_n)$, so $\mathbf{E}[X \mid Y_{0:n}] = \mathbf{E}[X \mid \mathcal{Y}_n]$ for a suitable random variable X. The index t on a random variable will always denote the time. By the Doob-Dynkin Lemma ([15],page 174), $\mathbf{E}[X \mid \mathcal{Y}_n] = \mathbf{E}[X \mid Y_{0:n}] = h(Y_0, Y_1, \ldots, Y_n)$ for some Borel function $h : \mathbb{R}^n \to \mathbb{R}$. Further for this X, $\mathbf{E}\left[X \,|\, Y_{0:n} = y_{0:n}\right] = \mathrm{h}(y_0, ..., y_n)$.

Write $\otimes$ for the usual Kronecker product. So if A is a matrix of size $m \times n$ and B is a matrix of size $p \times q$ then the Kronecker product of A and B is the $mp \times nq$ matrix given

by $A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}$. We define $\mathrm{vec}(A)$ for a matrix A.

If A is $m \times 1$ then $\mathrm{vec}(A) = A$. If $A = [A_1, A_2, ..., A_n]$ with columns A_i for $1 \leq i \leq n$, then $\mathrm{vec}(A) = (A_1^\top, A_2^\top, \ldots A_n^\top)^\top$ is an $mn \times 1$ vector. We shall write $\mathbb{R}^n$ for $n \geq 1$, to denote the Euclidean n-space whose elements will be written as column vectors.

3. Vector Autoregressive Models

Time series analysis and its applications have been used widely in different areas, including medicine, business, engineering and economics. A vector autoregressive model of order p (VAR(p)), is a time series model whose dynamics can be expressed as [9, 11, 10, 16]. If $Y_t = (Y_{1t}, Y_{2t}, ..., Y_{kt})^\top \in \mathbb{R}^k$ is a $k \times 1$ random vector ($^\top$ denotes transpose); $Y_t = A_0 + A_1 Y_{t-1} + ... + A_p Y_{t-p} + u_t\,, \quad t = 0, \pm 1, \pm 2, ..$ For each t; t; $t = p, p+1, p+2, ...,$ $A_0 \in \mathbb{R}^k$, the coefficients A_i, for $i = 1, 2, ..., p$, are constant $k \times k$ matrices, $\{u_t\}$ is a centered (mean zero) k-dimensional white noise process, which has independent terms. Also $E(u_t\, u_t^\top) = \Sigma_u$ for all t and $E(u_t u_s^\top) = 0_{k \times k}$ when $s \neq t$.

We assume that $\{u_t\}$ is a Gaussian white noise process.

4. Vector Autoregressive with Regime Switching

This section combines the notion of a vector autoregressive time series and regime switching models. The vector autoregressive parameters $\{A_i\}, \Sigma_u$, will be time-varying, they will be constant in any regime specified by a process X [7, 9].

Let N be the number of possible regimes and $\{X_\ell\}$ for $t \geq 0$. Then for a given regime of X the time series $\{Y_t\}$, will be generated by a vector autoregressive process of order $p \geq 1$ (written as VAR(p)) and the model can be expressed as [7, 9, 8]: $Y_t = A_0(X_{t-1}) + A_1(X_{t-1})Y_{t-1} + ... + A_p(X_{t-1})\, Y_{t-p} + u_t\,, \quad$ for $t \geq p$ where X_{t-1} is replaced by e_i when X_{t-1} is in the i-th regime and $\{u_t\}$ is Gaussian white noise with $u_t|X_{t-1} \sim \mathcal{N}(0, \Sigma(X_{t-1}))$. The other symbols are defined in section 3.

We may not know the appropriate number of regimes to be used. In our analysis we shall make various choices for the value of N. The question of an unbiased estimation of N for a given set of observations will fall outside the scope of this paper, but for hidden Markov models this has been studied by various authors including [14]. Some authors, for example [7], choose $N = 2$. Of course, given any choice of N, we shall provide estimators to determine which regime we are in at any time, based on the observation process Y. For hidden semi-Markov models, the consistent estimation of N has not been studied to our knowledge.

5. Hidden Semi-Markov Models

A homogenous semi-Markov chain is defined via a (Markov) renewal process $\{(Z_n, T_n) : n \geq 0\}$ with state space $\mathcal{S}$ as follows (see [4], chapters 9 and 10):

(a) $Z_n \in \mathcal{S} = \{e_1, e_2, ..., e_N\}$ for each $n \geq 0$;

(b) There are random times $T_n \in \{0, 1, 2, ...\}$ with $T_0 = 0$ and $T_{n+1} > T_n$ for each $n \geq 0$;

(c) whenever $i \neq j$ $P(Z_{n+1} = e_j, T_{n+1} - T_n = d \,|\, Z_n = e_i, Z_{0:n-1}, T_{0:n})$
$= P(Z_1 = e_j, T_1 - T_0 = d \,|\, Z_0 = e_i)$.

Define a function q on $\mathcal{S} \times \mathcal{S} \times \{1, 2, ...\} \to [0, 1]$ by

$$q(e_j, e_i, d) = P(Z_{n+1} = e_j, T_{n+1} - T_n = d \,|\, Z_n = e_i)\,.$$

A semi-Markov chain $\{X_\ell : \ell \geq 0\}$ is defined using this renewal process Z by setting

$$X_\ell = Z_n \quad \text{where } T_n \leq \ell < T_{n+1} \text{ for each } \ell \geq 0 \tag{5.1}$$

where the sequence $\{T_n\}$ gives the random jump times of the renewal process.

To establish that a stochastic process $\{X_\ell : \ell \geq 0\}$ with state space $\mathcal{S}$ on $(\Omega, \mathcal{F}, P)$ is a semi-Markov process, we must show that there is a renewal process $\{(Z_n, T_n) : n \geq 0\}$, so that (5.1) holds. We now discuss hidden semi-Markov models which use a hidden semi-Markov chain.

5.1. Specification of the Semi-Markov Model.

Below we shall use the result that for $E, F, G \in \mathcal{F}$, then

$$P(E \cap F|G) = P(E|F \cap G)\, P(F|G) \tag{5.2}$$

and we now give two specifications for a semi-Markov chain:

The first specification: In this specification for $i \neq j$ and $d \geq 1$, the $\{\tau(e_j, e_i, d),\, j \neq i,\, d = 1, 2, 3, ...\}$ and $\{p_{j\,i},\, j \neq i\}$ are the parameters of the model which must be estimated. Now, based on equation 5.2,

$$\begin{aligned} q(e_j, e_i, d) &= \mathrm{P}(Z_{n+1} = e_j, T_{n+1} - T_n = d \,|\, Z_n = e_i) \\ &= \tau(e_j, e_i, d)\, p_{ji}. \end{aligned} \tag{5.3}$$

Note that $p_{ii} = 0$ and for $i \neq j$,

$p_{ji} = \sum_{d=1}^{\infty} \mathrm{P}(Z_{n+1} = e_j, T_{n+1} - T_n = d \,|\, Z_n = e_i) = \sum_{d=1}^{\infty} q(e_j, e_i, d)$.

This specification has been used by many authors, for example [1, 13, 12].

The second specification: In this specification for $i \neq j$ and $d \geq 1$, the $\{A_{j\,i}(d), j \neq i,\, d = 1,2,3,...\}$ and $\{p_i(d),\, d = 1,2,3,...\}$ are the parameters of the model which must be estimated. Then based on equation 5.2,

$$q(e_j, e_i, d) = P\left(Z_{n+1} = e_j \,|\, T_{n+1} - T_n = d, Z_n = e_i\right) = A_{ji}(d)\, p_i(d). \tag{5.3a}$$

In terms of $\{q(e_j, e_i, d) \;:\; 1 \leq i, j \leq N \;\text{ and }\; d \geq 1\}$, we have
$p_i(d) = \sum_{j=1}^{N} A_{ji}(d)\, p_i(d) = \sum_{j=1}^{N} q\,(e_j, e_i, d)$.

In this specification, it is clear that the elements of semi-Markov transition matrix (or kernel) depend on the sojourn times. This will be called the full model. Ferguson [6], Bulla [2], Bulla and Bulla [3] and many other authors used this specification but in their work the transition matrix $A_{ji}(d)$ does not depend on sojourn time d. That model will be called a partial (specified) model.

Obviously, given a renewal process $\{(T_n, Z_n) : n = 0,1,2,...\}$, we can construct a semi-Markov chain by $X_t = Z_n$ if $T_n \leq t < T_{n+1}$. Vice-versa, given $\{X_t : t = 0,1,2,...\}$ we can construct a renewal process $\{(T_n, Z_n) : n = 0,1,2,...\}$.

5.2. Properties of Semi-Markov Chains. We suppose that our semi-Markov chain $X = \{X_k; k = 0,1,2,...\}$ on $(\Omega, \mathcal{F}, \mathrm{P}\,)$ is described in term of the second specification of section 5.1, using the quantities π, $\{p_i(d); i = 1,2,...,N,\, d \geq 1\}$ and $\{A_{ji}(d) : 1 \leq i,j \leq N,\, d \geq 1\}$. We now establish properties of X in terms of these quantities. Using results from [17], assuming $X_\ell = e_i$ and $P(h_\ell(X_\ell) = d \,|\, X_\ell = e_i) > 0$, then $P(X_{\ell+1} \neq e_i \,|\, X_\ell = e_i,\, h_\ell(X_\ell) = d) = \frac{p_i(d)}{F_i(d)}$ where $F_i(d) = \sum_{\ell=d}^{\infty} p_i(\ell)$. also $P(X_{\ell+1} = e_i \,|\, X_\ell = e_i,\, h_\ell(X_\ell) = d) = \frac{F_i(d+1)}{F_i(d)}$. Also for $j \neq i$, $A(d)_{ji} = \mathrm{P}(Z_{n+1} = e_j \,|\, Z_n = e_i, T_{n+1} - T_n = d) = \frac{q(e_j,\, e_i,\, d)}{p_i(d)}$

when $p_i(d) = \sum_{j=1}^{N} q(e_j,\, e_i.\, d) \neq 0$. Write $\frac{p(d)}{F(d)} = \left(\frac{p_1(d)}{F_1(d)}, \frac{p_2(d)}{F_2(d)}, \ldots, \frac{p_N(d)}{F_N(d)}\right)'$
and

$$\frac{F(d+1)}{F(d)} = \left(\frac{F_1(d+1)}{F_1(d)}, \frac{F_2(d+1)}{F_2(d)}, \ldots, \frac{F_N(d+1)}{F_N(d)}\right)'.$$

Lemma 5.1. $\mathbf{E}\left[X_{\ell+1} \,\middle|\, \mathcal{F}_\ell\right] = \mathfrak{B}_\ell(X_\ell)\, X_\ell$
where $\mathfrak{B}_\ell(X_\ell) = A(h_\ell(X_\ell)) \left\langle X_\ell, \frac{p(h_\ell(X_\ell))}{F(h_\ell(X_\ell))}\right\rangle + \left\langle X_\ell, \frac{F(h_\ell(X_\ell)+1)}{F(h_\ell(X_\ell))}\right\rangle \mathrm{I}.$

Proof. We have $\mathbf{E}\left[X_{\ell+1} \,\middle|\, \mathcal{F}_\ell\right] = \mathbf{E}\left[\sum_{j=1}^{N} \langle e_j, X_{\ell+1}\rangle e_j \,\middle|\, \mathcal{F}_\ell\right] = \mathbf{E}\left[X_{\ell+1} \,\middle|\, X_\ell, h_\ell(X_\ell)\right]$ see Theorem 1 in ([17]). Let us now suppose that $X_\ell = e_i$ and $h_\ell(X_\ell) = d$. Then $\mathbf{E}\left[X_{\ell+1} \,\middle|\, X_\ell = e_i, h_\ell(X_\ell) = d\right] = \sum_{j=1}^{N} e_j \,\langle \mathcal{B}_\ell(X_\ell) X_\ell, e_j\rangle$ for if $X_\ell = e_i$, $h_\ell(X_\ell) = d$, then $\langle \mathcal{B}_\ell(X_\ell) X_\ell, e_j\rangle = A_{ji}(d) \frac{p_i(d)}{F_i(d)} + \delta_{ji} \frac{F_i(d+1)}{F_i(d)}$ and so $\mathbf{E}\left[X_{\ell+1} \,\middle|\, \mathcal{F}_\ell\right] = \mathfrak{B}_\ell(X_\ell)\, X_\ell$. The lemma is proved. □

5.3. Construction of Semi-Markov Chains. Suppose $(\Omega, \mathcal{F}, \overline{\mathrm{P}}\,)$ is a reference probability space and $X = \{X_k; k = 0,1,2,...\}$, $X = \{X_k; k = 0,1,2,...,M\})$ the uniform process. We now introduce a new probability P on $(\Omega, \mathcal{F})$ so that under p, X is a semi-Markov chains with specification $\{q(e_j, e_i, d), 1 \leq i, j \leq N,\, d \geq 1\}$ and state space $\mathcal{S} = \{e_1, e_2, ..., e_N\}$. This analysis will be used to construct the likelihood function for the estimation of a semi-Markov chain. Write $q(e_j, e_i, d) = A_{ji}(d)\, p_i(d)$ as above for $1 \leq i, j \leq N$, $d \geq 1$. Let$\lambda_0 = N\langle \pi, X_0\rangle$ and

$\lambda_{\ell+1} = N\langle \mathfrak{B}_\ell(X_\ell)\, X_\ell, X_{\ell+1}\rangle$, for each $\ell \geq 0$. Write $\Lambda_n = \prod_{\ell=0}^{n} \lambda_\ell$ for $n \geq 0$ and $\frac{dP}{d\overline{P}}\Big|_{\mathcal{F}_n} = \Lambda_n$ for all $0 \leq n \leq M$, and for some $M > 0$.

Lemma 5.2. $\overline{\mathbf{E}}\left[\, \lambda_{\ell+1} \,\middle|\, \mathcal{F}_\ell \right] = 1$ *for* $l \geq 0$ *and* $\overline{\mathbf{E}}\left[\, \lambda_0\right] = 1\,.$

Proof. We first prove the second identity. In fact

$$\overline{\mathbf{E}}[\lambda_0] = \overline{\mathbf{E}}\left[N\langle \pi, X_0\rangle\right] = N\sum_{j=1}^{N}\langle \pi, e_j\rangle \overline{P}(X_0 = e_j) = 1$$

as

$$\overline{P}(X_0 = e_j) = \frac{1}{N}$$

for each $1 \leq j \leq N$ and

$$\sum_{j=1}^{N}\langle \pi, e_j\rangle = \sum_{j=1}^{N}\pi_j = 1.$$

For the first expression,

$$\begin{aligned}\overline{\mathbf{E}}\left[\, \lambda_{\ell+1} \,\middle|\, \mathcal{F}_\ell \right] =& \overline{\mathbf{E}}\left[N\langle \mathfrak{B}(X_\ell)X_\ell, X_{\ell+1}\rangle \,\middle|\, \mathcal{F}_\ell\right] \\ =& N\langle \mathfrak{B}(X_\ell)X_\ell, \overline{\mathbf{E}}\left[X_{\ell+1}\right]\rangle\,.\end{aligned}$$

where $\overline{\mathbf{E}}\left[X_{\ell+1}\right] = \frac{1}{N}\,\mathbf{I}$ and hence $\mathbf{E}\left[\, \lambda_{\ell+1} \,\middle|\, \mathcal{F}_\ell \right] = \langle \mathfrak{B}(X_\ell)X_\ell, \mathbf{1}\rangle\,.$ Now as $\mathfrak{B}_\ell(X_\ell) = A(h_\ell(X_\ell))\left\langle X_\ell, \frac{p(h_\ell(X_\ell))}{F(h_\ell(X_\ell))}\right\rangle + \left\langle X_\ell, \frac{F(h_\ell(X_\ell)+1)}{F(h_\ell(X_\ell))}\right\rangle \mathrm{I}$, we have

$$\begin{aligned}\langle \mathfrak{B}(X_\ell)X_\ell, \mathbf{1}\rangle \;&= \mathbf{1}^\top\left\langle X_\ell, \frac{p(h_\ell(X_\ell))}{F(h_\ell(X_\ell))}\right\rangle X_\ell + \left\langle X_\ell, \frac{F(h_\ell(X_\ell)+1)}{F(h_\ell(X_\ell))}\right\rangle \mathbf{1}^\top X_\ell \\ &= \langle X_\ell, \mathbf{1}\rangle = 1\,.\end{aligned}$$

This is because $p_i(k) + F_i(k+1) = p_i(k) + \sum_{d=k+1}^{\infty} p_i(d) = \sum_{d=k}^{\infty} p_i(d) = F_i(k)$ for any $k \geq 1$. □

The next Lemma is similar to Theorem 4 in [5], but we provide an alternate proof.

Lemma 5.3. *Under probability* P*, we have* $\mathbf{E}\left[X_{\ell+1} \,\middle|\mathcal{F}_\ell\right] = \mathfrak{B}_\ell(X_\ell)\, X_\ell \in \mathbb{R}^N$ *and* $P(X_0 = e_j) = \pi(j)\,.$

Proof. We have, $\mathbf{E}\left[X_{\ell+1} \,\middle|\mathcal{F}_\ell\right] = \frac{\overline{\mathbf{E}}\left[\lambda_{\ell+1}X_{\ell+1} \,\middle|\, \mathcal{F}_\ell\right]}{\overline{\mathbf{E}}\left[\lambda_{\ell+1} \,\middle|\, \mathcal{F}_\ell\right]} = \overline{\mathbf{E}}\left[\lambda_{\ell+1}X_{\ell+1} \,\middle|\, \mathcal{F}_\ell\right]$ by Lemma 5.2 $\overline{\mathbf{E}}\left[\, \lambda_{\ell+1} \,\middle|\, \mathcal{F}_\ell\right] = 1\,.$ Further $\overline{\mathbf{E}}\left[\lambda_{\ell+1}X_{\ell+1}|\mathcal{F}_\ell\right]$
$= \overline{\mathbf{E}}\left[N\langle \mathfrak{B}_\ell(X_\ell)X_\ell, X_{\ell+1}\rangle X_{\ell+1} \,\middle|\mathcal{F}_\ell\right] = \sum_{i=1}^{N}\langle \mathfrak{B}_\ell(X_\ell)X_\ell, e_i\rangle e_i = \mathfrak{B}_\ell(X_\ell)X_\ell.$

For the second part $P(X_0 = e_j) = \pi(j)$, for all $1 \leq j \leq N$ and $A \in \mathcal{F}$.

$$\overline{\mathbf{E}}[I(X_0 = e_j)] = \overline{\mathbf{E}}[\Lambda_0\, I(X_0 = e_j)] = \pi(j).$$

□

Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $A, B \in \mathcal{F}$ we say that $M = P(A|B)$ if $0 < M \leq 1$ and $MP(B) = P(A \cap B)$. This definition occurs in various books, in particular in Çinlar 1975 ([4], p14). In the case that $P(B) > 0$, this definition agrees with the elementary definition.

Lemma 5.4. *Suppose A, B are elements of $\mathcal{F}$ and $\{D_\ell\}$ is a disjoint sequence in $\mathcal{F}$. Suppose further that*

$$B \subset \bigcup_\ell D_\ell \tag{5.4}$$

then $P(A|B) = \sum_\ell P(A|B \cap D_\ell)P(D_\ell|B)$.

Proof. Let
$M = \sum_\ell P(A|B \cap D_\ell)P(D_\ell|B)$ then $0 \leq M \leq \sum_\ell P(D_\ell|B) = P\left(\bigcup_\ell D_\ell |\, B\right) \leq 1$ and $MP(B) = \sum_\ell P(A|B \cap D_\ell)P(D_\ell|B)P(B) = P(A \cap B)$ as $B \subset \bigcap_\ell D_\ell$ and equation 5.4 holds. □

Lemma 5.5. *Suppose that $X = \{X_\ell : \ell \geq 0\}$ is a stochastic process on the probability space $(\Omega, \mathcal{F}, P)$ with a state space $\mathcal{S} = \{e_1, e_2, ..., e_N\} \subset \mathbb{R}^N$ and on which $\mathbf{E}[\, X_{\ell+1} \mid \mathcal{F}_\ell\,] = \mathcal{B}_\ell(X_\ell)X_\ell$ where $\mathcal{F}_\ell = \sigma\{X_0, X_1, ..., X_\ell\}$ for all $\ell \geq 0$. Then $\{X_\ell : \ell \geq 0\}$ is a semi-Markov chain, with specification $\{q(e_j, e_i, d) = A_{ji}(d)p_i(d) : 1 \leq i, j \leq N,\ d \geq 1\}$.*

Proof. Define a renewal process $\{(Z_n, T_n) : n \geq 0\}$ with specification $\{q(e_j, e_i, d) : 1 \leq i, j \leq N \,\text{and}\, d \geq 1\}$. Let $Z_0 = X_0, T_0 = 0$, Set
$T_{n+1} = \min\{\ell > T_n : X_\ell \neq X_{T_n}\}$ and $Z_{n+1} = X_{T_{n+1}}$ for each $n \geq 0$. Then $X_\ell = Z_n$ if $T_n \leq \ell < T_{n+1}$ for each $\ell \geq 0$. For all $j \neq i$ and $d \leq 1$, compute $P(\, Z_{n+1} = e_j, T_{n+1} - T_n = d \,|\, Z_n = e_i, T_{0:n}, Z_{0:n-1})$ where $T_{0:n} = \{T_0, T_1, ..., T_n\}$ and $Z_{0:n-1} = \{Z_0, Z_1, ..., Z_{n-1}\}$. We claim this quantity will equal $q(e_j, e_i, d)$ and also equal to $P(\, Z_{n+1} = e_j, T_{n+1} - T_n = d \,|\, Z_n = e_i)$.
Now, by using equation 5.2

$$P(A|B) = \sum_{\ell \geq 0} P(A|B \cap D_\ell)P(D_\ell|B) \tag{5.5}$$

with choices $A = \{Z_{n+1} = e_j, T_{n+1} - T_n = d\}$, $B = \{Z_n = e_i, T_{0:n}, Z_{0:n-1}\}$ and $D_\ell = \{T_n = \ell\}$ for $\ell \geq 0$. We note that $B \subset \bigcup_{\ell \geq 0} D_\ell = \Omega$ and

$$\sum_{\ell \geq 0} P(D_\ell|B) = 1\,. \tag{5.6}$$

Further, using $\sigma\{Z_{0:n-1}, T_{0:n-1}\} = \sigma\{X_{0:n-1}\}$

$$\begin{aligned} P(A|B \cap D_\ell) &= P(\, Z_{n+1} = e_j, T_{n+1} - T_n = d|\, Z_n = e_i, T_n = \ell, T_{0:n-1}, Z_{0:n-1}) \\ &= P(E \cap F|G) \end{aligned}$$

with $E = \{X_{d+\ell} = e_j\}, F = \{X_{\ell+1:d+\ell-1} = e_i\}, G = \{X_\ell = e_i, X_{\ell-1} \neq e_i, X_{0:\ell-1}\}$
and by using Lemma 5.4 $P(E \cap F|G) = P(E|F \cap G)P(F|G)$. So
$P(E|F \cap G) = P(X_{d+\ell} = e_j|X_\ell = e_i, X_{\ell+1} = e_i, ..., X_{\ell+d-1} = e_i, X_{\ell-1} \neq e_i, X_{0:\ell-1})$ which does not depend on ℓ. In fact,

$$\begin{aligned} &P(X_{d+\ell} = e_j|X_{\ell+d-1} = e_i, h_{d+\ell-1}(X_{d+\ell-1}) = d, X_{0:\ell-1}) \\ &= \mathbf{E}[\, I(X_{d+\ell} = e_i)|\, X_{d+\ell-1} = e_i, h_{d+\ell-1}(X_{d+\ell-1}) = d, X_{0:\ell-1}] = A_{ji}(d)\frac{p_i(d)}{F_i(d)} \end{aligned}$$

as $j \neq i$. Using Lemma 5.4 again, we have

$P(F|G) = P(X_{d+\ell-1} = e_i, ..., X_{\ell+1} = e_i | X_\ell = e_i, X_{\ell-1} \neq e_i, X_{0:\ell-1})$
$= P(X_{d+\ell-1} = e_i | X_{d+\ell-2} = e_i, ..., X_{\ell+1} = e_i, X_\ell = e_i, X_{\ell-1} \neq e_i, X_{0:\ell-1}) \times$
$P(X_{d+\ell-2} = e_i, ..., X_{\ell+1} = e_i | X_\ell = e_i, X_{\ell-1} \neq e_i, X_{0:\ell-1})$
where

$$P(X_{d+\ell-1} = e_i | X_{d+\ell-2} = e_i, ..., X_{\ell+1} = e_i, X_\ell = e_i, X_{\ell-1} \neq e_i, X_{0:\ell-1})$$
$$P(X_{d+\ell-2} = e_i, ..., X_{\ell+1} = e_i | X_\ell = e_i, X_{\ell-1} \neq e_i, X_{0:\ell-1})$$
$$= \langle e_i, \mathbf{E}[\, X_{d+\ell-1} |\, X_{d+\ell-2} = e_i, h_{d+\ell-2}(X_{d+\ell-2}) = d, X_{0:\ell-1}] \rangle = \frac{F_i(d)}{F_i(d-1)}\,.$$

Repeating the argument, we have

$$P(F|G) = \frac{F_i(d)}{F_i(d-1)} . \frac{F_i(d-1)}{F_i(d-2)} ... \frac{F_i(3)}{F_i(2)}\; P(X_{\ell+1} = e_i | X_\ell = e_i, X_{\ell-1} \neq e_i, X_{0:\ell-1}))$$
$$= \frac{F_i(d)}{F_i(2)} . \frac{F_2(d)}{F_i(1)} = F_i(d)$$

as $F_i(1) = 1$. Also $P(F|G) = F_i(d)$, which does not depend on ℓ. So $P(A|B \cap D_\ell) = P(E \cap F|G) = A_{ji}(d)p_i(d)$ which does not depended on $\ell \geq 0$. substituting into equation (5.5) and using equation (5.6) we obtain $P(A|B) = \sum_{\ell=0}^{\infty} A_{ji}(d)p_i(d)P(D_\ell|B) = A_{ji}(d)p_i(d)$. We have now established $P(\, Z_{n+1} = e_j, T_{n+1} - T_n = d|\, Z_n = e_i, T_{0:n}, Z_{0:n-1}) = q(e_j, e_i, d)$
for any $i \neq j,\ d \geq 1$, using (5.5) and (5.6). The same argument can be applied to

$$P(\, Z_{n+1} = e_j, T_{n+1} - T_n = d|\, Z_n = e_i)$$

to establish the same result. This means that $\{(Z_n, T_n) : n \geq 0\}$ is indeed a renewal process for X and so X is a semi-Markov chain. □

We note that $P(Z_0 = e_j) = P(X_0 = e_j) = \pi(j)$.

Remark 5.6. $\{X_k; k \geq 0\}$ could be simulate under P using lemma 5.3 or lemma 5.5.

6. Estimation

In this section, we shall discuss the estimation process for semi-Markov model and providing all necessary lemmas. Let $\{X_\ell\}$ be a semi-Markov chain with the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ which we shall express in terms of a reference probability space $(\Omega, \mathcal{F}, \overline{\mathrm{P}})$. We shall consider time series of the form.

$$Y_\ell = B(X_{\ell-1}) + C(X_{\ell-1})Y_{\ell-1} + D(X_{\ell-1})\, W_\ell \tag{6.1}$$

or

$$Y_\ell = B(X_\ell) + C(X_\ell)Y_{\ell-1} + D(X_\ell)\, W_\ell \tag{6.2}$$

where (B, C, D) are the parameters of the model and must be estimated along with the parameters of the semi-Markov chain. For $j = 1, ..., N$ for which we have $\pi(j) \geq 0$ for each j and these probabilities must sum to 1. Also $B_r(e_j)$ is the r–th component $(1 \leq r \leq m)$ of $B(X_{\ell-1})$ when $X_{\ell-1} = e_j$ and there are no restrictions on its value. The parameters $C_{r,s}(e_i)$ for $1 \leq r, s \leq m$ are the components of the $m \times m$ matrix in model (6.1) when $X_{\ell-1} = e_j$ and there are no restrictions on its values. The parameters $D_{r,s}(e_i)$ for $1 \le r, s \le m$ are the components of the $m \times m$ matrix in model (6.1) when $X_{\ell-1} = e_j$. We require that the matrices $D(e_j)$ are invertible for each $1 \le j \le N$. We shall denote by Θ the collection of all these admissible parameters.

We shall now define a new probability $\mathrm{P} = \mathrm{P}_\theta$ on $(\Omega, \mathcal{F})$ so that $\frac{d\mathrm{P}}{d\overline{\mathrm{P}}}\Big|_{\mathcal{G}_n} = \overline{\Lambda}_n$ for $0 \le n \le M$. In model (6.1) it will be assumed that under p_θ the terms of the sequence $\{W_\ell\}$ are independent and identically distributed Gaussian $\mathcal{N}(\mathbf{0}, \mathrm{I}_m)$ random variables for each $\ell \ge 1$.

We suppose, $\theta = (\pi, q, B, C, D)$ and $\theta' = (\pi', q', B', C', D')$.

Given parameters θ and θ', we define

$$Q(\theta', \theta) = \mathbf{E}_\theta\left[\log \frac{dP_\theta{}'}{dP_\theta}\middle| Y_{0:n} = y_{0:n}\right] = \mathbf{E}_\theta\left[\log \Lambda_{\theta'} - \log \Lambda_\theta \middle| Y_{0:n} = y_{0:n}\right]$$

and for a given θ, we seek $\widehat{\theta}$ to maximize $\theta' \to \mathbf{E}_\theta\left[\log \Lambda_{\theta'} \middle| Y_{0:n} = y_{0:n}\right]$. We now compute $\mathbf{E}_\theta\left[\log \Lambda_{\theta'} \middle| Y_{0:n} = y_{0:n}\right]$ which is the expectation step.

6.1. Initialization: To estimate the hidden semi-Markov model parameters, let some initial estimation of the parameters be denoted by $\theta = \theta_0$. We could take the uniform distribution over $\mathcal{S}$ for π, take $q = \{q(e_j, e_i, d) = A_{ji}(d)p_i(d);\ 1 \le i, j \le N,\ d \ge 1\}$ with

$$A = \frac{1}{N-1}\left(\mathbf{1}\,\mathbf{1}^\top - \mathrm{I}_N\right) \quad \text{for all} \quad d \ge 1 \text{ and} p_i(d) = \frac{1}{1-e^{-\lambda}}\,\frac{\lambda^d}{d\,!}\,e^{-\lambda}$$

for all i and $d \ge 1$ for same $\lambda > 0$ (example $\lambda = 1$), $B = 0, C = 0$ and $D = \mathrm{I}_k$.

6.2. The Expectation Step: We obtain an explicit expression for $Q(\theta', \theta)$.

$$\begin{aligned} Q(\theta', \theta) &= \sum_{j=1}^{N} \log \pi'(j)\, \mathbf{E}_\theta\left[\langle X_0, e_j\rangle \middle| Y_{0:n} = y_{0:n}\right] \\ &+ \sum_{i,j=1}^{N} \mathbf{E}_\theta\left[\sum_{\ell=1}^{n} \log\langle \mathfrak{B}'_{\ell-1}(X_{\ell-1})X_{\ell-1}, X_\ell\rangle\, \langle X_{\ell-1}, e_i\rangle\, \langle X_\ell, e_j\rangle \middle| Y_{0:n} = y_{0:n}\right] \\ &- \sum_{i=1}^{N} \log\left|\det(D'(e_i))\right| \mathbf{E}_\theta\left[\sum_{\ell=1}^{n}\langle X_{\ell-1}, e_i\rangle \middle| Y_{0:n} = y_{0:n}\right] \\ &- \frac{1}{2}\sum_{\ell=1}^{n}\sum_{i=1}^{N} \|f'_{i\ell}(y_\ell, y_{\ell-1})\|^2\, \mathbf{E}_\theta\left[\langle X_{\ell-1}, e_i\rangle \middle| Y_{0:n} = y_{0:n}\right] + \text{constant} \end{aligned} \tag{6.3}$$

where the constant does not depend on θ'. Given $\theta = (\pi, q, B, C, D)$, we provide expressions for quantities:

$$\breve{Z}^{j}_{\ell,n} = \mathbf{E}_\theta\left[\langle X_\ell, e_j\rangle \middle| Y_{0:n} = y_{0:n}\right] \text{ for } \ell \le n\,, \tag{6.3a}$$

$$\breve{N}^{i\cdot}_{n} = \mathbf{E}_\theta\left[\sum_{\ell=1}^{n}\langle X_{\ell-1}, e_i\rangle \middle| Y_{0:n} = y_{0:n}\right] \tag{6.3b}$$

and

$$\breve{T}_n^{ij} = \mathbf{E}_\theta\left[\sum_{\ell=1}^{n}\log\langle\mathfrak{B}'_{\ell-1}(X_{\ell-1})X_{\ell-1},X_\ell\rangle\langle X_{\ell-1},e_i\rangle\langle X_\ell,e_j\rangle\middle| Y_{0:n}=y_{0:n}\right]. \quad (6.3c)$$

The following terms will be estimated using recursive estimation for:

$$\mathbf{E}_\theta\left[\langle X_\ell,e_j\rangle\, X_k \middle| Y_{0:k}=y_{0:k}\right], \quad (6.4)$$

$$\mathbf{E}_\theta\left[\sum_{\ell=1}^{k}\langle X_{\ell-1},e_i\rangle\, X_k\middle| Y_{0:k}=y_{0:k}\right] \quad (6.5)$$

and

$$\mathbf{E}_\theta\left[\sum_{\ell=1}^{k}\log\langle\mathfrak{B}'_{\ell-1}(X_{\ell-1})X_{\ell-1},X_\ell\rangle\langle X_{\ell-1},e_i\rangle\,\langle X_\ell,e_j\rangle\, X_k\middle| Y_{0:k}=y_{0:k}\right]. \quad (6.6)$$

After these expressions are obtained with $k=n$, one takes the inner product of these expressions with $\mathbf{1}=(1,...,1)^\top\in\mathbb{R}^n$ to obtain expressions (6.3a)-(6.3c).

Lemma 6.1. *We have the expression* $q_k=\overline{\mathbf{E}}[\overline{\Lambda}_k X_k|\mathcal{Y}_k]=\sum_{d=1}^{k+1}q_k(d)$ *where* $q_k(d)=\overline{\mathbf{E}}\left[\overline{\Lambda}_k\, X_k\, I(h_k=d)\middle|\,\mathcal{Y}_k\right]$. *Here h_k is an abbreviation for $h_k(X_k)$. Then for $k\geq 1$ and $d=1$,*
$q_k(1)=\sum_{j=1}^{N}\sum_{d'=1}^{k}A(d')e_j\,\frac{p_j(d')}{F_j(d')}\,\Gamma_j(Y_k,Y_{k-1})\,\langle q_{k-1}(d'),e_j\rangle$ *and for $d>1$,*
$q_k(d)=\sum_{j=1}^{N}\frac{F_j(d)}{F_j(d-1)}\,\Gamma_j(Y_k,Y_{k-1})\,e_j\langle q_{k-1}(d-1),e_j\rangle$ *with initial value,*

$$q_0(d)=\begin{cases}0 & \text{if } d>1\,,\\ \pi & \text{if } d=1\,.\end{cases}$$

In above formula: $\Gamma_j(Y_k,Y_{k-1})=\Gamma(e_j,Y_k,Y_{k-1})$ *for* $j=1,...,N$. *We also have* $\overline{\mathbf{E}}\,[\overline{\Lambda}_k|\mathcal{Y}_k]=\langle q_k,\,\mathbf{1}\rangle$.

Proof. Knowing that $\overline{\mathbf{E}}\,[M_k\,|\,\mathcal{Y}_k\vee\mathcal{F}_{k-1}]=0$ then
$q_k(d)=\overline{\mathbf{E}}\left[\overline{\Lambda}_{k-1}\,\overline{\mathbf{E}}\left[\overline{\lambda}_k\,\mathfrak{B}_{k-1}(X_{k-1})X_{k-1}\,I(h_k=d)\middle|\,\mathcal{Y}_k\vee\mathcal{F}_{k-1}\right]\middle|\,\mathcal{Y}_k\right]$. Now for $d=1$,

$$=\sum_{d'=1}^{k}\mathfrak{B}_{k-1}(X_{k-1})X_{k-1}\,I(h_{k-1}=d')\,\Gamma(X_{k-1},Y_k,Y_{k-1}),$$

and hence $q_k(1)=\sum_{j=1}^{N}\sum_{d'=1}^{k}A(d')e_j\,\frac{p_j(d')}{F_j(d')}\,\Gamma_j(Y_k,Y_{k-1})\,\langle q_{k-1}(d'),e_j\rangle$. For $d>1$

$$=\mathfrak{B}_{k-1}(X_{k-1})X_{k-1}\,I(h_{k-1}=d-1)\,\Gamma(X_{k-1},Y_k,Y_{k-1})\,,$$

and hence $q_k(d)=\sum_{j=1}^{N}\frac{F_j(d)}{F_j(d-1)}\,\Gamma_j(Y_k,Y_{k-1})\,e_j\langle q_{k-1}(d-1),e_j\rangle$. Now for the initial value $q_0(d)=\begin{cases}0 & \text{if } d>1\,,\\ \pi & \text{if } d=1\,.\end{cases}$ This follows because
$q_0(d)=\overline{\mathbf{E}}[\overline{\Lambda}_0\,X_0\,I(h_0=d)|\mathcal{Y}_0]$ is zero if $d>1$ and
$q_0(1)=\overline{\mathbf{E}}\,[\overline{\Lambda}_0\,X_0\,|\mathcal{Y}_0]=\sum_{j=1}^{N}\pi(j)\,e_j=\pi$. This proves the lemma. □

For developing recurrences relationships, we used un-normalized quantities such as $\widehat{q}_k$, but in numerical implementation we use normalized quantities for, instance $\widehat{v}_k$ which we now discuss. The normalized, nonlinear recurrences must be used because of their numerical stability.

For numerically implementation, we use $v_k = \mathbf{E}_\theta\left[X_k|\mathcal{Y}_k\right]$, $\widehat{v}_k = \mathbf{E}_\theta\left[X_k|Y_{0:k} = y_{0:k}\right]$ and $\widehat{v}_k(d) = \mathbf{E}_\theta\left[X_k I(h_k = d)|\, Y_{0:k} = y_{0:k}\right]$.

Lemma 6.2. *We have* $\widehat{v}_k = \widehat{v}_k(1) + \sum_{d=2}^{k+1}\widehat{v}_k(d)$ *where for* $d = 1$,

$$\widehat{v}_k(1) = \frac{\displaystyle\sum_{j=1}^{N}\sum_{d'=1}^{k} A(d')e_j\,\frac{p_j(d')}{F_j(d')}\,\Gamma_j(y_k, y_{k-1})\,\langle\widehat{v}_{k-1}(d'), e_j\rangle}{\displaystyle\sum_{j=1}^{N}\Gamma_j(y_k, y_{y-1})\ \langle\widehat{v}_{k-1}, e_j\rangle}$$ *and for* $d > 1$,

$$\widehat{v}_k(d) = \frac{\displaystyle\sum_{j=1}^{N}\frac{F_j(d)}{F_j(d-1)}\,\Gamma_j(y_k, y_{k-1})\,e_j\langle\widehat{v}_{k-1}(d-1), e_j\rangle}{\displaystyle\sum_{j=1}^{N}\Gamma_j(y_k, y_{y-1})\ \langle\widehat{v}_{k-1}, e_j\rangle}$$, *with initial values* $\widehat{v}_0 = \pi$ *and*

$$\widehat{v}_0(d) = \begin{cases} 0 & \text{if } d > 1\,, \\ \pi & \text{if } d = 1\,. \end{cases}$$

Proof. We start from

$$\widehat{v}_k = \mathbf{E}_\theta\left[X_k|Y_{0:k} = y_{0:k}\right] = \frac{\widehat{q}_k(1)}{\langle\widehat{q}_k, \mathbf{1}\rangle} + \frac{\displaystyle\sum_{d=2}^{k+1}\widehat{q}_k(d)}{\langle\widehat{q}_k, \mathbf{1}\rangle}\,. \tag{6.7}$$

Now we have to compute the denominator $\langle\widehat{q}_k, \mathbf{1}\rangle$
$\langle q_k, \mathbf{1}\rangle = \overline{\mathbf{E}}\left[\overline{\Lambda}_k|\mathcal{Y}_k\right] = \sum_{j=1}^{N}\Gamma_j(Y_k, Y_{k-1})\,\langle q_{k-1}, e_j\rangle$, which immediately implies $\langle\widehat{q}_k, \mathbf{1}\rangle = \sum_{j=1}^{N}\Gamma_j(y_k, y_{k-1})\,\langle\widehat{q}_{k-1}, e_j\rangle$. Now
$\frac{\langle\widehat{q}_k, \mathbf{1}\rangle}{\langle\widehat{q}_{k-1}, \mathbf{1}\rangle} = \sum_{j=1}^{N}\Gamma_j(y_k, y_{k-1})\,\left\langle\frac{\widehat{q}_{k-1}}{\langle\widehat{q}_{k-1}, \mathbf{1}\rangle}, e_j\right\rangle$. Then

$$\langle\widehat{q}_k, \mathbf{1}\rangle = \sum_{j=1}^{N}\Gamma_j(y_k, y_{k-1})\ \langle\widehat{v}_{k-1}, e_j\rangle\ \langle\widehat{q}_{k-1}, \mathbf{1}\rangle\,. \tag{6.8}$$

Now returning to equation (6.7) and dividing it by the new dominator which is equation (6.8) we obtain for $d \geq 1$, $\widehat{v}_k(d) = \frac{\widehat{q}_k(d)}{\langle\widehat{q}_k, \mathbf{1}\rangle}$. Then by Lemma 6.1 and equation (6.8)

$$\widehat{v}_k(1) = \frac{\displaystyle\sum_{j=1}^{N}\sum_{d'=1}^{k} A(d')e_j\,\frac{p_j(d')}{F_j(d')}\,\Gamma_j(y_k, y_{k-1})\,\langle\widehat{v}_{k-1}(d'), e_j\rangle}{\displaystyle\sum_{j=1}^{N}\Gamma_j(y_k, y_{k-1})\ \langle\widehat{v}_{k-1}, e_j\rangle}\,.$$

The formula for $\widehat{v}_k(d)$ for $d > 1$ is obtained in the same way. For the initial conditions $\widehat{v}_0(d) = \frac{\widehat{q}_0(d)}{\langle\widehat{q}_0, \mathbf{1}\rangle} = 0$ if $d > 1$ and $\widehat{v}_0(1) = \frac{\widehat{q}_0(1)}{\langle\widehat{q}_0, \mathbf{1}\rangle} = \frac{\pi}{\langle\pi, \mathbf{1}\rangle} = \pi$ from which $\widehat{v}_0 = \pi$ follows. The lemma is proved. □

Smoother Estimates: For equation (6.2) with $k \geq \ell$, we compute

$$\mathbf{E}_\theta \left[\langle X_\ell, e_j \rangle X_k \middle| Y_{0:k} = y_{0:k} \right] = \frac{\overline{\mathbf{E}} \left[\overline{\Lambda}_k \langle X_\ell, e_j \rangle X_k \middle| Y_{0:k} = y_{0:k} \right]}{\overline{\mathbf{E}} \left[\overline{\Lambda}_k \middle| Y_{0:k} = y_{0:k} \right]} .$$

Firstly we compute the numerator in (6.2). We use similar arguments to Lemma 6.1 and write

$Z^j_{\ell,k} = \overline{\mathbf{E}} \left[\overline{\Lambda}_k \langle X_\ell, e_j \rangle X_k \middle| \mathcal{Y}_k \right]$ and $\widehat{Z}^j_{\ell,k} = \overline{\mathbf{E}} \left[\overline{\Lambda}_k \langle X_\ell, e_j \rangle X_k \middle| Y_{0:k} = y_{0:k} \right]$.

Lemma 6.3. *Write* $\widehat{Z}^j_{\ell,k} = \sum_{d=1}^{k+1} \widehat{Z}^j_{\ell,k}(d)$ *. For* $\ell \leq k$*, we have recurrence relationships for* $\widehat{Z}^j_{\ell,k}(d) = \overline{\mathbf{E}} \left[\overline{\Lambda}_k \langle X_\ell, e_j \rangle I(h_k = d) \middle| Y_{0:k} = y_{0:k} \right]$ *. For* $d = 1$*,*
$\widehat{Z}^j_{\ell,k}(1) = \sum_{i=1}^N \sum_{d'=1}^k A_i(d') \frac{p_i(d')}{F_i(d')} \Gamma_i(y_k, y_{k-1}) \langle \widehat{Z}^j_{\ell,k-1}(d'), e_i \rangle$ *and for* $d > 1$*,*
$\widehat{Z}^j_{\ell,k}(d) = \sum_{i=1}^N \frac{F_i(d)}{F_i(d-1)} \Gamma_i(y_k, y_{k-1}) \langle \widehat{Z}^j_{\ell,k-1}(d-1), e_i \rangle e_i$ *, the initial values are*
$\widehat{Z}^j_{\ell,\ell}(d) = \hat{q}^j_\ell(d)$ *.*

Proof. We shall refer the recurrences for $\widehat{Z}^j_{\ell,k}(d)$ from those for $Z^j_{\ell,k}(d)$, in the usual way. We have for $k > \ell$,

$$\begin{aligned} Z^j_{\ell,k}(d) =& \overline{\mathbf{E}} \left[\overline{\Lambda}_k \langle X_\ell, e_j \rangle X_k I(h_k = d) \middle| \mathcal{Y}_k \right] \\ =& \overline{\mathbf{E}} \left[\overline{\Lambda}_{k-1} \langle X_\ell, e_j \rangle \overline{\mathbf{E}} \left[\overline{\lambda}_k \mathfrak{B}_{k-1}(X_{k-1}) X_{k-1} I(h_k = d) \middle| \mathcal{Y}_k \vee \mathcal{F}_{k-1} \right] \middle| \mathcal{Y}_k \right] . \end{aligned} \tag{6.9}$$

For $d = 1$,

$$\begin{aligned} &\overline{\mathbf{E}} \left[\overline{\lambda}_k \mathfrak{B}_{k-1}(X_{k-1}) X_{k-1} I(h_k = d) \middle| \mathcal{Y}_k \vee \mathcal{F}_{k-1} \right] \\ &= \sum_{d'=1}^{k} \mathfrak{B}_{k-1}(X_{k-1}) X_{k-1} I(h_{k-1} = d') \Gamma(X_{k-1}, Y_k, Y_{k-1}). \end{aligned}$$

Then writing $A_i(d')$ for the i–th column of A,

$$Z^j_{\ell,k}(1) = \sum_{i=1}^N \sum_{d'=1}^k A_i(d') \frac{p_i(d')}{F_i(d')} \Gamma_i(Y_k, Y_{k-1}) \langle Z^j_{\ell,k-1}(d'), e_i \rangle.$$

For $d > 1$ and $k > \ell$,

$$\begin{aligned} &\overline{\mathbf{E}} \left[\overline{\lambda}_k \mathfrak{B}_{k-1}(X_{k-1}) X_{k-1} I(h_k = d) \middle| \mathcal{Y}_k \vee \mathcal{F}_{k-1} \right] \\ &= \mathfrak{B}_{k-1}(X_{k-1}) X_{k-1} I(h_{k-1} = d-1) \Gamma(X_{k-1}, Y_k, Y_{k-1}). \end{aligned}$$

Hence $Z^j_{\ell,k}(d) = \sum_{i=1}^N \frac{F_i(d)}{F_i(d-1)} \Gamma_i(Y_k, Y_{k-1}) \langle Z^j_{\ell,k-1}(d-1), e_i \rangle e_i$. We now conclude that for $k > \ell$, and for $d = 1$

$Z^j_{\ell,k}(1) = \sum_{i=1}^N \sum_{d'=1}^k A_i(d') \frac{p_i(d')}{F_i(d')} \Gamma_i(Y_k, Y_{k-1}) \langle Z^j_{\ell,k-1}(d'), e_i \rangle$, for $d > 1$,

$Z^j_{\ell,k}(d) = \sum_{i=1}^N \frac{F_i(d)}{F_i(d-1)} \Gamma_i(Y_k, Y_{k-1}) \langle Z^j_{\ell,k-1}(d-1), e_i \rangle e_i$ and $Z^j_{\ell,\ell}(d) = q^j_\ell(d)$.

Then for $d = 1$, $\widehat{Z}^j_{\ell,k}(1) = \sum_{i=1}^N \sum_{d'=1}^{k+1} A_i(d') \frac{p_i(d')}{F_i(d')} \Gamma_i(y_k, y_{k-1}) \langle \widehat{Z}^j_{\ell,k-1}(d'), e_i \rangle$,

and for $d > 1$, $\widehat{Z}^j_{\ell,k}(d) = \sum_{i=1}^N \frac{F_i(d)}{F_i(d-1)} \Gamma_i(y_k, y_{k-1}) \langle \widehat{Z}^j_{\ell,k-1}(d-1), e_i \rangle e_i$ and initially, $\widehat{Z}^j_{\ell,\ell}(d) = \widehat{q}^j_\ell(d)$. The lemma is proved. □

We compute the denominator $\overline{\mathbf{E}}\left[\overline{\Lambda}_k \middle| Y_{0:k} = y_{0:k}\right]$ in two ways as $\sum_{v=1}^{N} \langle \widehat{Z}^v_{\ell,k}, \mathbf{1}\rangle = \langle \widehat{q}_k, \mathbf{1}\rangle$ and so $\mathbf{E}_\theta\left[\langle X_\ell, e_j\rangle \middle| Y_{0:n} = y_{0:n}\right] = \frac{\langle \widehat{Z}^j_{\ell,n}, \mathbf{1}\rangle}{\sum_{v=1}^{N}\left\langle \widehat{Z}^v_{\ell n}, \mathbf{1}\right\rangle} = \frac{\langle \widehat{Z}^j_{\ell n}, \mathbf{1}\rangle}{\langle \widehat{q}_n, \mathbf{1}\rangle}$.

Numerical calculations will use recursion for the normalized quantities. We define $\widehat{W}^j_{\ell,k} = \mathbf{E}_\theta\left[\langle X_\ell, e_j\rangle X_k \middle| Y_{0:k} = y_{0,k}\right] = \frac{\overline{\mathbf{E}}\left[\overline{\Lambda}_k \langle X_\ell, e_j\rangle X_k \middle| Y_{0:k} = y_{0,k}\right]}{\overline{\mathbf{E}}\left[\overline{\Lambda}_k \middle| Y_{0:k} = y_{0,k}\right]}$

$$= \frac{\widehat{Z}^j_{\ell,k}}{\langle \widehat{q}_k, \mathbf{1}\rangle} = \frac{\sum_{d=1}^{k+1} \widehat{Z}^j_{\ell,k}(d)}{\langle \widehat{q}_k, \mathbf{1}\rangle}.$$

We shall write $\widehat{W}^j_{\ell,k}(d) = \mathbf{E}_\theta\left[\langle X_\ell, e_j\rangle X_k\, I(h_k = d) \middle| Y_{0:k} = y_{0,k}\right]$ for which we now provide recursions.

Lemma 6.4. *We let* $\widehat{W}^j_{\ell,k} = \widehat{W}^j_{\ell,k}(1) + \sum_{d=2}^{k+1} \widehat{W}^j_{\ell,k}(d)$, *where for* $d = 1$

$$\widehat{W}^j_{\ell,k}(1) = \frac{\sum_{i=1}^{N}\sum_{d'=1}^{k} A_i(d')\,\frac{p_i(d')}{F_i(d')}\,\Gamma_i(y_k, y_{k-1})\,\langle \widehat{W}^j_{\ell,k-1}(d'), e_i\rangle}{\sum_{i=1}^{N}\Gamma_i(y_k, y_{y-1})\ \langle \widehat{v}_{k-1}, e_i\rangle} \quad \textit{and for } d > 1$$

$$\widehat{W}^j_{\ell,k}(d) = \frac{\sum_{i=1}^{N}\frac{F_i(d)}{F_i(d-1)}\,\Gamma_i(y_k, y_{k-1})\,\langle \widehat{W}^j_{\ell,k-1}(d-1), e_i\rangle\, e_i}{\sum_{i=1}^{N}\Gamma_i(y_k, y_{y-1})\ \langle \widehat{v}_{k-1}, e_i\rangle} \quad \textit{with initial value}$$

$\widehat{W}^j_{\ell,\ell}(d) = \langle \widehat{v}_\ell(d), e_j\rangle\ e_j$. *Furthermore,* $\breve{Z}^j_{\ell,n} = \left\langle \widehat{W}^j_{\ell,n}, \mathbf{1}\right\rangle$.

Proof. We use results from Lemma 6.3 and equation (6.8) for $k > \ell$. If $d = 1$,

$$\widehat{W}^j_{\ell,k}(1) = \frac{\widehat{Z}^j_{\ell,k}(1)}{\langle \hat{q}^j_k, \mathbf{1}\rangle} = \frac{\sum_{i=1}^{N}\sum_{d'=1}^{k} A_i(d')\,\frac{p_i(d')}{F_i(d')}\,\Gamma_i(y_k, y_{k-1})\,\langle \widehat{W}^j_{\ell,k-1}(d'), e_i\rangle}{\sum_{i=1}^{N}\Gamma_i(y_k, y_{y-1})\ \langle \widehat{v}_{k-1}, e_i\rangle}.$$

If $d > 1$,

$$\widehat{W}^j_{\ell,k}(d) = \frac{\widehat{Z}^j_{\ell,k}(d)}{\langle \hat{q}^j_k, \mathbf{1}\rangle} = \frac{\sum_{i=1}^{N}\frac{F_i(d)}{F_i(d-1)}\,\Gamma_i(y_k, y_{k-1})\langle \widehat{W}^j_{\ell,k-1}(d-1), e_i\rangle\, e_i}{\sum_{i=1}^{N}\Gamma_i(y_k, y_{y-1})\ \langle \widehat{v}_{k-1}, e_i\rangle}.$$

Initial values for $k = \ell$ are $\widehat{W}^j_{\ell,\ell}(d) = \langle \widehat{v}_\ell(d), e_j\rangle\, e_j$.

Also $\breve{Z}^j_{\ell,n} = \mathbf{E}_\theta\left[\langle X_\ell, e_j\rangle \middle| Y_{0:n} = y_{0,n}\right] = \left\langle \widehat{W}^j_{\ell,n}, \mathbf{1}\right\rangle$. The lemma is proved. □

Computing State Occupation estimates

To obtain equation (6.5), we note that $\sum_{\ell=1}^{n}\langle X_{\ell-1}, e_i\rangle$ gives the number of times $X_0, ..., X_{n-1}$ occupy the state e_i. Then
$\breve{N}_n^{i\cdot} = \mathbf{E}_\theta\left[\sum_{\ell=1}^{n}\langle X_{\ell-1}, e_i\rangle \middle| Y_{0:n} = y_{0:n}\right] = \left\langle \sum_{\ell=1}^{n}\widehat{W}_{\ell-1,n}^{i}, \mathbf{1}\right\rangle$. This is because
$\mathbf{E}_\theta\left[\langle X_{\ell-1}, e_i\rangle \middle| Y_{0:n} = y_{0:n}\right] = \langle \mathbf{E}_\theta\left[\langle X_{\ell-1}, e_i\rangle X_n \middle| Y_{0:n} = y_{0:n}\right], \mathbf{1}\rangle$.

State transition the Smoother Estimates

Finally, to compute expression in (6.3c)
$\breve{T}_n^{ij} = \mathbf{E}_\theta\left[\sum_{\ell=1}^{n}\log\langle \mathfrak{B}'_{\ell-1}(X_{\ell-1})X_{\ell-1}, X_\ell\rangle\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle \middle| Y_{0:n} = y_{0:n}\right]$, we adapt the method from [17]. We obtain $\breve{T}_n^{ij}$ from J_n^{ij}, where
$J_k^{ij} = \mathbf{E}_\theta\left[\sum_{\ell=1}^{k}\log\langle \mathfrak{B}'_{\ell-1}(X_{\ell-1})X_{\ell-1}, X_\ell\rangle\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle \middle| \mathcal{Y}_k\right]$.

This is the same as

$$\mathbf{E}_\theta\left[\sum_{\ell=1}^{k}\sum_{d=1}^{\ell}\log\langle \mathfrak{B}'_{\ell-1}(X_{\ell-1})X_{\ell-1}, X_\ell\rangle\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle I(h_{\ell-1} = d)\middle| \mathcal{Y}_k\right] =$$

$$\sum_{\ell=1}^{k}\sum_{d=1}^{\ell}\mathbf{E}_\theta\left[\log\left\{A'_{ji}(d)\frac{p'_i(d)}{F'_i(d)} + \delta_{ij}\frac{F'_i(d+1)}{F'_i(d)}\right\}\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle I(h_{\ell-1} = d)\middle| \mathcal{Y}_k\right].$$

We define $\zeta_{\ell-1}^{k}(j,i,d) = \mathbf{E}_\theta\left[\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle I(h_{\ell-1} = d)\middle| \mathcal{Y}_k\right]$. As $A_{ii}(d) = 0$,

$$\text{we obtain } J_k^{ij} = \begin{cases} \displaystyle\sum_{\ell=1}^{k}\sum_{d=1}^{\ell}\log\left\{A'_{ji}(d)\frac{p'_i(d)}{F'_i(d)}\right\}\zeta_{\ell-1}^{k}(j,i,d) & \text{if } \quad i \neq j \\ \displaystyle\sum_{\ell=1}^{k}\sum_{d=1}^{\ell}\log\left\{\frac{F'_i(d+1)}{F'_i(d)}\right\}\zeta_{\ell-1}^{k}(i,i,d) & \text{if } \quad i = j \end{cases}$$

$$= \begin{cases} \displaystyle\sum_{\ell=1}^{k}\sum_{d=1}^{\ell}\log A'_{ji}(d)\zeta_{\ell-1}^{k}(j,i,d) & \\ \displaystyle\quad + \sum_{\ell=1}^{k}\sum_{d=1}^{\ell}\{\log p'_i(d) - \log F'_i(d)\}\,\zeta_{\ell-1}^{k}(j,i,d) & \text{if } \quad i \neq j \\ \displaystyle\sum_{\ell=1}^{k}\sum_{d=1}^{\ell}\{\log F'_i(d+1) - \log F'_i(d)\}\,\zeta_{\ell-1}^{k}(i,i,d) & \text{if } \quad i = j\,. \end{cases}$$

Now as $\zeta_{\ell-1}^{k}(j,i,d)$

$$= \mathbf{E}_\theta\left[\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle I(h_{\ell-1} = d)\middle| \mathcal{Y}_k\right] = \frac{\overline{\mathbf{E}}\left[\overline{\Lambda}_k\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle I(h_{\ell-1} = d)\middle| \mathcal{Y}_k\right]}{\langle q_k, \mathbf{1}\rangle}.$$

In order to obtain $\zeta_{\ell-1}^{k}(j,i,d)$ for $k > \ell$, we compute the numerator. Define
$H_{\ell-1}^{ijd}(k,r,\delta) = \overline{\mathbf{E}}\left[\overline{\Lambda}_k\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle I(h_{\ell-1} = d)\langle X_k, e_r\rangle I(h_k = \delta)\middle| \mathcal{Y}_k\right]$ and then

$$\zeta_{\ell-1}^{k}(j,i,d) = \frac{\displaystyle\sum_{r=1}^{N}\sum_{\delta=1}^{k+1}H_{\ell-1}^{ijd}(k,r,\delta)}{\langle q_k, \mathbf{1}\rangle}\,. \tag{6.10}$$

We then set $\zeta_{\ell-1}(j,i,d) = \zeta^n_{\ell-1}(j,i,d)$ evaluated at $Y_{0:n} = y_{0:n}$. We obtain a recurrence relationship for $k \to H^{ijd}_\ell(k+1,r,\delta)$. For this, suppose $k+1 \geq \ell+1$, then

$$H^{ijd}_\ell(k+1,r,\delta) = \overline{\mathbf{E}}\left[\overline{\Lambda}_{k+1}\langle X_\ell, e_i\rangle I(h_\ell = d)\langle X_{\ell+1}, e_j\rangle\langle X_{k+1}, e_r\rangle I(h_{k+1}=\delta)\middle|\, \mathcal{Y}_{k+1}\right] \tag{6.11}$$

which is equal to

$$\begin{aligned}&\overline{\mathbf{E}}\left[\overline{\mathbf{E}}\left[\overline{\Lambda}_{k+1}\langle X_\ell, e_i\rangle I(h_\ell = d)\langle X_{\ell+1}, e_j\rangle\langle X_{k+1}, e_r\rangle I(h_{k+1}=\delta)\middle|\, \mathcal{Y}_{k+1}\vee\mathcal{F}_k\right]\middle|\, \mathcal{Y}_{k+1}\right] = \\ &\overline{\mathbf{E}}\left[\overline{\Lambda}_k\langle X_\ell, e_i\rangle\, I(h_\ell = d)\langle X_{\ell+1}, e_j\rangle \overline{\mathbf{E}}\left[\overline{\lambda}_{k+1}\langle X_{k+1}, e_r\rangle I(h_{k+1}=\delta)\middle|\, \mathcal{Y}_{k+1}\vee\mathcal{F}_k\right]\middle|\, \mathcal{Y}_{k+1}\right].\end{aligned} \tag{6.12}$$

Lemma 6.5. *We have* $\overline{\mathbf{E}}\left[\overline{\lambda}_{k+1}\langle X_{k+1}, e_r\rangle I(h_{k+1}=\delta)\middle|\, \mathcal{Y}_{k+1}\vee\mathcal{F}_k\right]$ *equal to*

$$\Gamma_r(Y_k, Y_{k+1})\,\frac{F_r(\delta)}{F_r(\delta-1)}\,\langle X_k, e_r\rangle I(h_k = \delta-1) \tag{6.13}$$

if $\delta > 1$ *and for* $\delta = 1$*, it is equal to*

$$\sum_{\delta'=1}^{k+1}\sum_{s\neq r=1}^{N} A_{rs}(\delta')\frac{p_s(\delta')}{F_s(\delta')}\,\Gamma_s(Y_k, Y_{k+1})\,\langle X_k, e_s\rangle\, I(h_k = \delta')\,. \tag{6.14}$$

Proof. We have

$$\begin{aligned}&\overline{\mathbf{E}}\left[\overline{\lambda}_{k+1}\langle X_{k+1}, e_r\rangle I(h_{k+1}=\delta)\middle|\, \mathcal{Y}_{k+1}\vee\mathcal{F}_k\right]\\ =&N\,\langle\mathfrak{B}_k(X_k)X_k, e_r\rangle\,\frac{1}{|\det(D(X_k))|}\,\frac{\psi\,(b_{k+1})}{\psi\,(Y_{k+1})}\,\overline{\mathbf{E}}\left[\,\langle X_{k+1}, e_r\rangle I(h_{k+1}=\delta)\right|\mathcal{F}_k\right]\,.\end{aligned} \tag{6.15}$$

Now $\overline{\mathbf{E}}\left[\,\langle X_{k+1}, e_r\rangle I(h_{k+1}=\delta)\right|\mathcal{F}_k\right] = \sum_{s=1}^N \overline{\mathbf{E}}\left[\,\langle X_k, e_s\rangle\langle X_{k+1}, e_r\rangle I(h_{k+1}=\delta)\right|\mathcal{F}_k\right]$ equals

$$\langle X_k, e_r\rangle I(h_k = \delta - 1)\,\frac{1}{N}\,, \tag{6.16}$$

if $\delta > 1$, and equals

$$\sum_{r\neq s=1}^{N}\langle X_k, e_s\rangle\,\frac{1}{N} \tag{6.17}$$

if $\delta = 1$.

We substitute equations (6.16) and (6.17) into (6.15) separately and respectively, then for $\delta > 1$,

$$N\,\langle\mathfrak{B}_k(X_k)X_k, e_r\rangle\,\frac{1}{|\det(D(X_k))|}\,\frac{\psi\,(b_{k+1})}{\psi\,(Y_{k+1})}\,\langle X_k, e_r\rangle I(h_k = \delta-1)\,\frac{1}{N}$$

$$= \Gamma_r(Y_k, Y_{k+1})\,\tfrac{F_r(\delta)}{F_r(\delta-1)}\,\langle X_k, e_r\rangle I(h_k = \delta-1)\,,$$

and for $\delta = 1$, $\quad N\,\langle\mathfrak{B}_k(X_k)X_k, e_r\rangle\,\frac{1}{|\det(D(X_k))|}\,\frac{\psi(b_{k+1})}{\psi(Y_{k+1})}\,\sum_{s\neq r=1}^N\langle X_k, e_s\rangle\,\frac{1}{N}$

$= \sum_{\delta'=1}^{k+1}\sum_{s\neq r=1}^N A_{rs}(\delta')\frac{p_s(\delta')}{F_s(\delta')}\,\Gamma_s(Y_k, Y_{k+1})\,\langle X_k, e_s\rangle\, I(h_k = \delta')\,.$

The lemma is proved. ∎

Now we substitute equations (6.13) and (6.14) into (6.12) separately and respectively, for $\delta > 1$,

$\overline{\mathbf{E}}\left[\overline{\Lambda}_k\langle X_\ell, e_i\rangle\langle X_{\ell+1}, e_j\rangle\, I(h_\ell = d)N\Gamma_r(Y_k, Y_{k+1})\frac{F_r(\delta)}{F_r(\delta-1)}\langle X_k, e_r\rangle I(h_k = \delta-1)\Big|\,\mathcal{Y}_{k+1}\right]$
$= \Gamma_r(Y_k, Y_{k+1})\frac{F_r(\delta)}{F_r(\delta-1)}\overline{\mathbf{E}}\left[\overline{\Lambda}_k\langle X_\ell, e_i\rangle\langle X_{\ell+1}, e_j\rangle I(h_\ell = d)\langle X_k, e_r\rangle I(h_k = \delta-1)\big|\,\mathcal{Y}_k\right]$,

and for $\delta = 1$,

$$\overline{\mathbf{E}}\Bigg[\overline{\Lambda}_k\langle X_\ell, e_i\rangle\langle X_{\ell+1}, e_j\rangle I(h_\ell = d)\sum_{\delta'=1}^{k+1}\sum_{s\neq r=1}^{N} A_{rs}(\delta')\frac{p_s(\delta')}{F_s(\delta')}\,\Gamma_s(Y_k, Y_{k+1})\,\langle X_k, e_s\rangle$$
$$I(h_k = \delta')|\,\mathcal{Y}_{k+1}\Bigg]$$
$$= \sum_{\delta'=1}^{k+1}\sum_{s\neq r=1}^{N} A_{rs}(\delta')\frac{p_s(\delta')}{F_s(\delta')}\Gamma_s(Y_k, Y_{k+1})\overline{\mathbf{E}}\left[\overline{\Lambda}_k\langle X_\ell, e_i\rangle\langle X_{\ell+1}, e_j\rangle I(h_\ell = d)\langle X_k, e_s\rangle\right.$$
$$\left. I(h_k = \delta')|\,\mathcal{Y}_k\right].$$

Now we recall the definition (6.11)

$H_\ell^{ijd}(k+1, r, \delta) = \overline{\mathbf{E}}\left[\overline{\Lambda}_{k+1}\langle X_\ell, e_i\rangle\,\langle X_{\ell+1}, e_j\rangle I(h_\ell = d)\langle X_{k+1}, e_r\rangle I(h_{k+1} = \delta)|\,\mathcal{Y}_{k+1}\right]$

and can express it in terms of $H_\ell^{ijd}(k, s, \delta-1)$ with $k > \ell$. For $\delta > 1$,

$H_\ell^{ijd}(k+1, r, \delta) = \Gamma_r(Y_k, Y_{k+1})\,\frac{F_r(\delta)}{F_r(\delta-1)}\,H_\ell^{ijd}(k, r, \delta-1)$, and if $\delta = 1$,
$H_\ell^{ijd}(k+1, r, \delta) = \sum_{\delta'=1}^{k}\sum_{s=1}^{N} A_{rs}(\delta')\frac{p_s(\delta')}{F_s(\delta')}\,\Gamma_s(Y_k, Y_{k+1})\,H_\ell^{ijd}(k, s, \delta')$. Then the corresponding estimators for $\delta > 1$

$\widehat{H}_\ell^{ijd}(k+1, r, \delta) = \Gamma_r(y_k, y_{k+1})\,\frac{F_r(\delta)}{F_r(\delta-1)}\,\widehat{H}_\ell^{ijd}(k, r, \delta-1)$ and if $\delta = 1$

$\widehat{H}_\ell^{ijd}(k+1, r, \delta) = \sum_{\delta'=1}^{k}\sum_{s=1}^{N} A_{rs}(\delta')\frac{p_s(\delta')}{F_s(\delta')}\,\Gamma_s(y_k, y_{k+1})\,\widehat{H}_\ell^{ijd}(k, s, \delta')$. In order to compute the initial value $H_\ell^{ijd}(\ell+1, r, \delta-1)$ where $\ell+1 = k$, we use the following lemma.

Lemma 6.6. *The expression* $\widehat{H}_\ell^{ijd}(\ell+1, r, \delta-1)$
$= \overline{\mathbf{E}}\left[\overline{\Lambda}_{\ell+1}\langle X_\ell, e_i\rangle I(h_\ell = d)\langle X_{\ell+1}, e_j\rangle\langle X_{\ell+1}, e_r\rangle I(h_{\ell+1} = \delta)\big|\,Y_{0:\ell} = y_{0:\ell}\right]$
is equal to zero if $r \neq j$ or $j \neq i$. For other choices where ($r = j = i$) and $d = \delta - 1$, it equals for $\delta > 1$ $\Gamma_j(y_\ell, y_{\ell+1})\,\frac{F_j(\delta)}{F_j(\delta-1)}\,\langle\widehat{q}_\ell(\delta-1), e_j\rangle$ and for $\delta = 1$, it equals

$$A_{ji}(d)\frac{p_i(d)}{F_i(d)}\,\Gamma_i(y_\ell, y_{\ell+1})\,\langle\widehat{q}_\ell(d), e_i\,\rangle\,. \tag{6.18}$$

Proof. The expression in equation (6.18) is clearly zero if $r \neq j$. We shall obtain $\widehat{H}_\ell^{ijd}(\ell+1, r, \delta-1)$ from those for $H_\ell^{ijd}(\ell+1, r, \delta-1)$. If $r = j$, we have

$$H_\ell^{ijd}(\ell+1, r, \delta-1) =$$
$$\overline{\mathbf{E}}\left[\overline{\Lambda}_\ell\langle X_\ell, e_i\rangle\, I(h_\ell = d)\,\overline{\mathbf{E}}\left[\overline{\lambda}_{\ell+1}\langle X_{\ell+1}, e_j\rangle I(h_{\ell+1} = \delta)\big|\,\mathcal{Y}_{\ell+1}\vee\mathcal{F}_\ell\right]\big|\,\mathcal{Y}_{\ell+1}\right]. \tag{6.19}$$

Now by Lemma 6.5, we have a formula for $\overline{\mathbf{E}}\left[\overline{\lambda}_{\ell+1}\langle X_{\ell+1}, e_j\rangle I(h_{\ell+1} = \delta)\big|\,\mathcal{Y}_{\ell+1}\vee\mathcal{F}_\ell\right]$ in (6.13) where $\delta > 1$ and (6.14) where $\delta = 1$. Substitute them into equation

(6.19) separately and respectively; if $\delta > 1$ it is

$$\overline{\mathbf{E}}\left[\overline{\Lambda}_\ell \langle X_\ell, e_i\rangle\, I(h_\ell = d)\,\Gamma_j(Y_\ell, Y_{\ell+1})\,\frac{F_j(\delta)}{F_j(\delta-1)}\,\langle X_\ell, e_j\rangle I(h_\ell = \delta - 1)\,\Big|\,\mathcal{Y}_{\ell+1}\right]$$
$$= \Gamma_j(Y_\ell, Y_{\ell+1})\,\frac{F_j(\delta)}{F_j(\delta-1)}\,\overline{\mathbf{E}}\left[\overline{\Lambda}_\ell \langle X_\ell, e_i\rangle\, I(h_\ell = d)\,\langle X_\ell, e_j\rangle I(h_\ell = \delta - 1)\,\big|\,\mathcal{Y}_\ell\right]$$

which is zero if $j \neq i$ so if $j = i$ for $d = \delta - 1$ it is,

$$= \Gamma_j(Y_\ell, Y_{\ell+1})\,\frac{F_j(\delta)}{F_j(\delta-1)}\,\langle q_\ell(\delta-1), e_j\rangle\,.$$

So for $\delta > 1$, $\widehat{H}_\ell^{ijd}(\ell+1, r, \delta-1) = \Gamma_j(y_\ell, y_{\ell+1})\,\frac{F_j(\delta)}{F_j(\delta-1)}\,\langle \widehat{q}_\ell(\delta-1), e_j\rangle$ if $r = j = i$ and zero otherwise. For $\delta = 1$

$$\overline{\mathbf{E}}\left[\overline{\Lambda}_\ell \langle X_\ell, e_i\rangle\, I(h_\ell = d)\sum_{\delta'=1}^{\ell+1}\sum_{s\neq j=1}^{N} A_{js}(\delta')\,\frac{p_s(\delta')}{F_s(\delta')}\,\Gamma_s(Y_\ell, Y_{\ell+1})\,\langle X_\ell, e_s\rangle\, I(h_\ell = \delta')\,\Bigg|\,\mathcal{Y}_{\ell+1}\right]$$
$$= \sum_{d'=1}^{\ell+1}\sum_{s\neq j=1}^{N} A_{js}(\delta')\,\frac{p_s(\delta')}{F_s(\delta')}\,\Gamma_s(Y_\ell, Y_{\ell+1})\,\overline{\mathbf{E}}\left[\overline{\Lambda}_\ell\,\langle X_\ell, e_i\rangle I(h_\ell = d)\,\langle X_\ell, e_s\rangle\, I(h_\ell = \delta')\,\big|\,\mathcal{Y}_\ell\right]$$

if $s = i$, $\delta' = d$ and $i \neq j$ otherwise zero. So

$$= A_{ji}(d)\,\frac{p_i(d)}{F_i(d)}\,\Gamma_i(Y_\ell, Y_{\ell+1})\,\langle q_\ell(d), e_i\rangle$$

using equation (6.9). It follows that for $\delta = 1$
$\widehat{H}_\ell^{ijd}(\ell+1, i, \delta) = A_{ji}(d)\,\frac{p_i(d)}{F_i(d)}\,\Gamma_i(y_\ell, y_{\ell+1})\,\langle \widehat{q}_\ell(d), e_i\rangle$. This proves the lemma. □

That means we can compute the expression in (6.10):

$$\zeta_{\ell-1}^n(j, i, d) = \frac{\displaystyle\sum_{r=1}^{N}\sum_{\delta=1}^{n+1} H_{\ell-1}^{ijd}(n, r, \delta)}{\langle q_n, \mathbf{1}\rangle}\,. \tag{6.20}$$

We shall derive some recursion for normalized quantities and their estimators. We make the following definition. For $k > \ell + 1$, let

$\mathcal{H}_\ell^{ijd}(k, r, \delta) = \mathbf{E}_\theta[\,\langle X_\ell, e_i\rangle\langle X_{\ell+1}, e_j\rangle I(h_\ell = d)\langle X_k, e_r\rangle I(h_k = \delta)|\,\mathcal{Y}_k] = \frac{H_\ell^{ijd}(k,r,\delta)}{\langle q_k, \mathbf{1}\rangle}$ and the corresponding estimators $\widehat{\mathcal{H}}_\ell^{ijd}(k, r, \delta)$
$= \mathbf{E}_\theta[\,\langle X_\ell, e_i\rangle\langle X_{\ell+1}, e_j\rangle I(h_\ell = d)\langle X_k, e_r\rangle I(h_k = \delta)|\,Y_{0:k} = y_{0:k}] = \frac{\widehat{H}_\ell^{ijd}(k,r,\delta)}{\langle \widehat{q}_k, \mathbf{1}\rangle}$. Also
$U_{\ell,k}^{ijd} = \zeta_{\ell-1}^k(i, j, d) = \mathbf{E}_\theta\left[\,\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle I(h_{\ell-1} = d)|\,\mathcal{Y}_k\right] = \displaystyle\sum_{r=1}^{N}\sum_{\delta=1}^{k+1}\mathcal{H}_{\ell-1}^{ijd}(k, r, \delta)$

and $\widehat{U}_{\ell,k}^{ijd} = \mathbf{E}_\theta\left[\,\langle X_{\ell-1}, e_i\rangle\langle X_\ell, e_j\rangle I(h_{\ell-1} = d)|\,Y_{0:k} = y_{0:k}\right] = \displaystyle\sum_{r=1}^{N}\sum_{\delta=1}^{k+1}\widehat{\mathcal{H}}_{\ell-1}^{ijd}(k, r, \delta)\,.$

Further, $\zeta_{\ell-1}(i, j, d) = \widehat{U}_{\ell,n}^{ijd}$. We now provide recursion and initial values.

Lemma 6.7. *For $k > \ell + 1$, let*

$$\widehat{\mathcal{H}}_\ell^{ijd}(k, r, \delta) = \frac{\widehat{H}_\ell^{ijd}(k, r, \delta)}{\langle \widehat{q}_k, \mathbf{1} \rangle},$$

where for $\delta = 1$,

$$\widehat{\mathcal{H}}_\ell^{ijd}(k, r, 1) = \frac{\sum_{\delta'=1}^{k} \sum_{s=1}^{N} A_{rs}(\delta') \frac{p_s(\delta')}{F_s(\delta')} \Gamma_s(y_\ell, y_{\ell+1}) \widehat{\mathcal{H}}_\ell^{ijd}(k-1, s, 1)}{\sum_{\nu=1}^{N} \Gamma_\nu(y_k, y_{k-1}) \langle \widehat{v}_{k-1}, e_\nu \rangle},$$

and for $\delta > 1$,

$$\widehat{\mathcal{H}}_\ell^{ijd}(k, r, \delta) = \frac{\Gamma_j(y_\ell, y_{\ell+1}) \frac{F_j(\delta)}{F_j(\delta-1)} \widehat{\mathcal{H}}_\ell^{ijd}(k-1, r, \delta-1)}{\sum_{\nu=1}^{N} \Gamma_\nu(y_k, y_{k-1}) \langle \widehat{v}_{k-1}, e_\nu \rangle}.$$

Proof. We have, using equation (6.8),

$$\mathcal{H}_\ell^{ijd}(k, r, \delta) = \frac{H_\ell^{ijd}(k, r, \delta)}{\langle q_k, \mathbf{1} \rangle} = \frac{H_\ell^{ijd}(k, r, \delta)}{\sum_{\nu=1}^{N} \Gamma_\nu(Y_k, Y_{k-1}) \langle v_{k-1}, e_\nu \rangle \langle q_{k-1}, \mathbf{1} \rangle}.$$

For $\delta = 1$, $\mathcal{H}_\ell^{ijd}(k, r, 1) = \dfrac{\sum_{\delta'=1}^{k} \sum_{s=1}^{N} A_{rs}(\delta') \frac{p_s(\delta')}{F_s(\delta')} \Gamma_s(Y_\ell, Y_{\ell+1}) \mathcal{H}_\ell^{ijd}(k-1, s, 1)}{\sum_{\nu=1}^{N} \Gamma_\nu(Y_k, Y_{k-1}) \langle v_{k-1}, e_\nu \rangle}$,

and for $\delta > 1$, $\mathcal{H}_\ell^{ijd}(k, r, \delta) = \dfrac{\Gamma_j(Y_\ell, Y_{\ell+1}) \frac{F_j(\delta)}{F_j(\delta-1)} \mathcal{H}_\ell^{ijd}(k-1, r, \delta-1)}{\sum_{\nu=1}^{N} \Gamma_\nu(Y_k, Y_{k-1}) \langle v_{k-1}, e_\nu \rangle}$. Then it follows that

for $\delta = 1$, $\widehat{\mathcal{H}}_\ell^{ijd}(k, r, 1) = \dfrac{\sum_{\delta'=1}^{k} \sum_{s=1}^{N} A_{rs}(\delta') \frac{p_s(\delta')}{F_s(\delta')} \Gamma_s(y_\ell, y_{\ell+1}) \widehat{\mathcal{H}}_\ell^{ijd}(k-1, s, 1)}{\sum_{\nu=1}^{N} \Gamma_\nu(y_k, y_{k-1}) \langle \widehat{v}_{k-1}, e_\nu \rangle}$,

and for $\delta > 1$, $\widehat{\mathcal{H}}_\ell^{ijd}(k, r, \delta) = \dfrac{\Gamma_j(y_\ell, y_{\ell+1}) \frac{F_j(\delta)}{F_j(\delta-1)} \widehat{\mathcal{H}}_\ell^{ijd}(k-1, r, \delta-1)}{\sum_{\nu=1}^{N} \Gamma_\nu(y_k, y_{k-1}) \langle \widehat{v}_{k-1}, e_\nu \rangle}$.

The lemma is proved. □

Initial values for the recursion are provided in the next Lemma.

We define, for $k = \ell + 1$,

$$\mathcal{H}_\ell^{ijd}(\ell+1, r, \delta) = \mathbf{E}_\theta \left[\langle X_\ell, e_i \rangle \langle X_{\ell+1}, e_j \rangle I(h_\ell = d) \langle X_{\ell+1}, e_r \rangle I(h_{\ell+1} = \delta) \middle| \mathcal{Y}_{\ell+1} \right]$$

$$= \frac{H_\ell^{ijd}(\ell+1,r,\delta)}{\langle q_{\ell+1},\mathbf{1}\rangle}$$

and the corresponding estimators

$$\widehat{\mathcal{H}}_\ell^{ijd}(\ell+1,r,\delta) = \mathbf{E}_\theta\left[\langle X_\ell,e_i\rangle\langle X_{\ell+1},e_j\rangle I(h_\ell=d)\langle X_{\ell+1},e_r\rangle I(h_{\ell+1}=\delta)|\,Y_{0:\ell+1}=y_{0:\ell+1}\right]$$
$$= \frac{\widehat{H}_\ell^{ijd}(\ell+1,r,\delta)}{\langle \widehat{q}_{\ell+1},\mathbf{1}\rangle}.$$

Lemma 6.8. *We have, for $k=\ell+1$ and $\delta=1$,*

$$\widehat{\mathcal{H}}_\ell^{ij}(\ell+1,r,1) = \frac{A_{ji}(d)\frac{p_i(d)}{F_i(d)}\,\Gamma_i(y_\ell,y_{\ell+1})\langle v_\ell(d),e_i\rangle}{\sum_{\nu=1}^{N}\Gamma_\nu(y_\ell,y_{\ell+1})\langle\widehat{v}_\ell,e_\nu\rangle}$$

and for $\delta>1$, $\widehat{\mathcal{H}}_\ell^{ij}(\ell+1,r,\delta) = \dfrac{\Gamma_j(y_\ell,y_{\ell+1})\,\frac{F_j(\delta)}{F_j(\delta-1)}\,\langle v_\ell(\delta-1),e_j\rangle}{\sum_{\nu=1}^{N}\Gamma_\nu(y_\ell,y_{\ell+1})\langle\widehat{v}_\ell,e_\nu\rangle}.$

Proof. We find $\widehat{\mathcal{H}}_\ell^{ijd}(\ell+1,r,\delta)$ in term of $\mathcal{H}_\ell^{ijd}(\ell+1,r,\delta)$, so

$$\mathcal{H}_\ell^{ijd}(\ell+1,r,\delta) = \frac{H_\ell^{ijd}(\ell+1,r,\delta)}{\langle q_{\ell+1},\mathbf{1}\rangle} = \frac{H_\ell^{ijd}(\ell+1,r,\delta)}{\sum_{\nu=1}^{N}\Gamma_\nu(Y_\ell,Y_{\ell+1})\langle v_\ell,e_\nu\rangle\langle q_\ell,\mathbf{1}\rangle}.$$

For $k=\ell+1$, $\mathcal{H}_\ell^{ijd}(\ell+1,r,\delta)=0$ unless $\delta'=d$, $r=i$ and $i\neq j$, then by using results of Lemma 6.6 for $\delta=1$,

$$\mathcal{H}_\ell^{ijd}(\ell+1,r,1) = \frac{A_{ji}(d)\frac{p_i(d)}{F_i(d)}\,\Gamma_i(Y_\ell,Y_{\ell+1})\langle v_\ell(d),e_i\rangle}{\sum_{\nu=1}^{N}\Gamma_\nu(Y_\ell,Y_{\ell+1})\langle v_\ell,e_\nu\rangle},$$

and $\mathcal{H}_\ell^{ijd}(\ell+1,r,\delta)=0$ unless $d=\delta-1$ and $r=i=j$, for $\delta>1$,

$$\mathcal{H}_\ell^{ij}(\ell+1,r,\delta) = \frac{\Gamma_j(Y_\ell,Y_{\ell+1})\,\frac{F_j(\delta)}{F_j(\delta-1)}\,\langle v_\ell(\delta-1),e_j\rangle}{\sum_{\nu=1}^{N}\Gamma_\nu(Y_\ell,Y_{\ell+1})\langle v_\ell,e_\nu\rangle}.$$

The corresponding estimators for $\delta=1$,

$$\widehat{\mathcal{H}}_\ell^{ij}(\ell+1,r,1) = \frac{A_{ji}(d)\frac{p_i(d)}{F_i(d)}\,\Gamma_i(y_\ell,y_{\ell+1})\langle v_\ell(d),e_i\rangle}{\sum_{\nu=1}^{N}\Gamma_\nu(y_\ell,y_{\ell+1})\langle\widehat{v}_\ell,e_\nu\rangle}$$

and for $\delta>1$, $\widehat{\mathcal{H}}_\ell^{ij}(\ell+1,r,\delta) = \dfrac{\Gamma_j(y_\ell,y_{\ell+1})\,\frac{F_j(\delta)}{F_j(\delta-1)}\,\langle v_\ell(\delta-1),e_j\rangle}{\sum_{\nu=1}^{N}\Gamma_\nu(y_\ell,y_{\ell+1})\langle\widehat{v}_\ell,e_\nu\rangle}.$

The lemma is proved. □

For the maximization part, we use the results which we have calculated in the Expectation step and then perform the maximization step. We compute $\hat{\theta}' = \arg\max_{\theta'} Q(\theta', \theta)$ under various constraints on θ'. We then maximize with respect to $\theta' = (\pi', q', B', C', D')$. The updates for $A_{ji}(d)$ and $p_i(d)$ can be made separately.

7. The Numerical Results

In order to make a simulation, we selected a set of parameters for a HSMM and generated a series of observations $\{y_\ell : \ell = 0, ..., n\}$. We let $N = 2$ and then we choose the $p_i(d) = 0$ if $d \geq 4$. We choose $\pi(1) = 0.5, \pi(2) = 0.5$, $A(d) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ for all $d \geq 1$ which is the only choice when $N = 2$. The values of $(B(e_i), C(e_i), D(e_i))$ are chosen as $(B(e_1) = 1.5, B(e_2) = 0.9, C(e_1) = 0.4, C(e_2) = 0.9, D(e_1) = 0.4, D(e_2) = 0.4)$. Also the values of $p_1(1) = 0.4, p_1(2) = 0.3, p_1(3) = 0.2, p_1(4) = 0.1, p_2(1) = 0.6, p_2(2) = 0.4, p_2(3) = 0, p_2(4) = 0$. We simulate as follows:

Let u_1 be a draw from $\mathcal{U}(0,1)$ a uniformly distributed random variable on $(0,1)$ and let $X_0 = \begin{cases} e_1 & \text{if } u_1 < \pi(1)\,, \\ e_2 & \text{else}\,. \end{cases}$ If $X_0 = e_i$. We make another draws for $u_2 \sim \mathcal{U}(0,1)$ and

$$d = \begin{cases} 1 & \text{if } \; u_2 < p_i(1)\,, \\ 2 & \text{if } \; p_i(1) \leq u_2 < p_i(1) + p_i(2)\,, \\ 3 & \text{if } \; p_i(1) + p_i(2) \leq u_2 < p_i(1) + p_i(2) + p_i(3)\,, \\ 4 & \text{otherwise}\,. \end{cases}$$

Then $X_\ell = e_i$ for $\ell = 0, 1, ..., d-1$. If $d > 1$, then $X_1 = e_i$. We assume $W_1 \sim \mathcal{N}(0,1)$ and let $y_1 = B(e_i) + C(e_i)y_0 + D(e_i)W_1$ where $y_0 = B(X_0)/(1 - C(X_0))$, say. Also $y_2 = B(e_i) + C(e_i)y_1 + D(e_i)W_2$ where W_2 is another draw from $\mathcal{N}(0,1)$ and so on for $0 \leq \ell \leq d$.

To determine a simulation for X_d we use $X_d = e_j$ with probability $A_{ji}(d)$ for $j = 1, ..., N$. We then repeat the argument above with $\ell = d$, the new starting time. We generated 1000 observations by this procedure. Using these simulated observations we proceeded to estimate the parameters of the model using the EM algorithm of this chapter with 50 iterations. The starting values of the parameters are given in a table for iteration zero. The out puts are given in tables (1) and (2). The plots of the process $\{X_\ell\}$ and $\{y_\ell\}$ are given in figures (1) and (2).

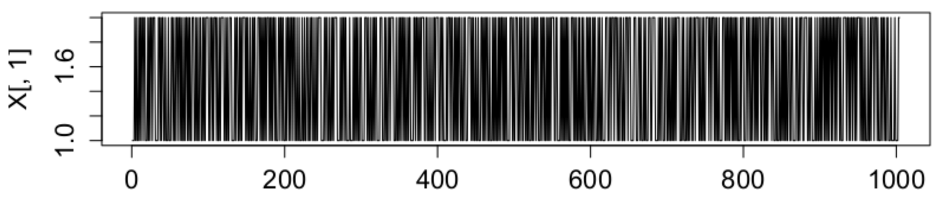

FIGURE 1. The plot of hidden state space values X

Table 1: Fitting VAR(1) with Two-state Hidden semi-Markov model on 1000 observations simulated data

Iterations	$B[1]$	$B[2]$	$C[1]$	$C[2]$	$D[1]$	$D[2]$	$\pi[1]$	$\pi[2]$
0	2.00	0.50	0.30	1.50	0.30	0.70	0.5	0.5
1	1.56	-1.38	0.61	1.43	0.79	2.11	0.5	0.5
2	1.55	-1.03	0.60	1.34	0.59	1.08	0.5	0.5
3	1.52	-0.81	0.59	1.26	0.59	0.90	0.5	0.5
25	1.17	0.88	0.54	0.71	0.59	0.61	0.5	0.5
45	1.10	1.06	0.41	0.69	0.60	0.60	0.5	0.5
50	1.10	1.06	0.40	0.69	0.60	0.60	0.5	0.5

Table 2: Fitting VAR(1) with Two-state Hidden semi-Markov model on 1000 observations length data series for sojourn time $p_i(d)$

Iterations	$p_1(1)$	$p_1(2)$	$p_1(3)$	$p_1(4)$	$p_2(1)$	$p_2(2)$	$p_2(3)$	$p_2(4)$
0	0.4000	0.3000	0.2000	0.1000	0.5000	0.2000	0.2000	0.1000
1	0.8107	0.1534	0.0290	0.0069	0.5661	0.2456	0.1066	0.0817
2	0.7286	0.1977	0.0537	0.0200	0.6144	0.2369	0.0913	0.0574
3	0.6851	0.2157	0.0679	0.0313	0.6492	0.2277	0.0799	0.0432
25	0.6691	0.2214	0.0733	0.0362	0.6643	0.2230	0.0749	0.0378
45	0.4110	0.3200	0.1720	0.0970	0.5043	0.4220	0.0359	0.0378
50	0.4110	0.3200	0.1720	0.0970	0.5043	0.4220	0.0359	0.0378

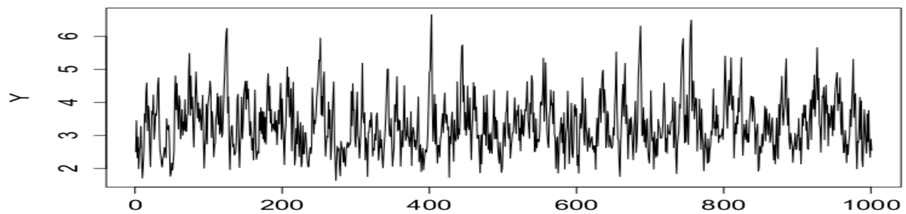

Figure 2. The plot of simulated Y series which includes 1000 observations

Several authors indicate that an initial guess (see for example [20])for the EM algorithm can be important, and we see this. Also a long data set of observations may be needed to well estimate the semi-Markov chain parameters (see for example [3]).

8. Conclusion

In this article, we provided an analytical study of regime switching $VAR(1)$ time series models where the regimes change according to hidden semi-Markov chains and we assumed that the transition matrix does depend on sojourn time. The formulation of the semi-Markov chain generalizes the one used by Ferguson, by allowing the transition in state probabilities to be duration dependent. We also estimate empirical distributions for the sojourn times of the hidden semi-Markov chains. Using a simple ($N = 2$ and $m = 4$) choice of parameters, we simulated observations, and using these simulations we applied the estimation algorithms to re-estimate the model.

Acknowledgment. The results of this paper were obtained during my Ph.D. study at the University of South Australia.

References

1. Barbu, V. & Limnios, N. :Semi-Markov chains and hidden semi-Markov models toward applications: their Use in Reliability and DNA analysis. (Springer,2008)
2. Bulla, J. :Application of hidden Markov models and hidden semi-Markov models to financial Time series. (Georg-August-Universität Göttingen,2006).
3. Bulla, J. & Bulla, I.:Stylized facts of financial Time series and hidden semi-Markov models. *Computational Statistics And Data Analysis.* **51** pp. 2192-2209 (2006)
4. Çinlar, E. :Introduction to Stochastic Processes. (Prentice-Hall.Inc,1975)
5. Elliott, R., Limnios, N. & Swishchuk, A. :Filtering hidden semi-Markov chains. *Statistics And Probability Letters.* **83**, 2007-2014 (2013)
6. Ferguson, J. :Variable duration models for speech. *Proceedings Of The Symposium On The Applications Of Hidden Markov Models To Text And Speech*
7. Hamilton, J.:Time series analysis. (Princeton University Press,1994)
8. Krolzig, H.-M.: Predicting Markov-switching vector autoregressive processes . (Nuffield College,2000)
9. Krolzig, H. :Markov-switching vector autoregression: modelling, Statistical Inference, and application to business cycle analysis. (Springer,1997)

10. Lütkepohl, H.: New introduction to multiple Time series analysis. (Springer,2007)
11. Luetkepohl, H.:Vector autoregressive models. (2011)
12. Malefaki, S., Trevezas, S. & Limnios, N.:An EM and a stochastic version of the EM algorithm for nonparametric hidden semi-Markov models. *Communication In Statistics-Simulation And Computation.* **39**, 240-261 (2010)
13. Pertsinidou, C., Tsaklidis, G., Papadimitriou, E. & Limnios, N. :Application of hidden semi-Markov models for the seismic hazard assessment of the North and South Aegean Sea, Greece. *Applied Statistics.* pp. 1-22 (2016)
14. Setiawaty, B. :Consistent estimation of the Order for hidden Markov models. (Univerity of Adelaide, South Australia,1999)
15. Shiryaev, A. :Probability. (Springerlink,1996)
16. Shumway, R. & Stoffer, D. :Time series analysis and its applications. (Springer,2011)
17. Van Der Hoek, J. & Elliott, R. : Introduction to Hidden Semi-Markov Models . (Cambridge University Press,2018)
18. Xie, Y., Tang, S. & Huang, X. :An efficient algorithm for parameterizing HsMM with Gaussian and Gamma distributions. *Information Processing Letters.* **112** pp. 732-737 (2012)
19. Yu, S.:Hidden semi-Markov models. *Artificial Intelligence.* **174** pp. 215-243 (2010)
20. Zucchini, W. & MacDonald, I. :Hidden Markov Model for Time Series: An Introduction using R. (CRS Press,2009)

Kurdstan Abdullah: Department of Statistics and Informatics, University of Sulaimani, Sulaymaniyah, Iraq
Email address: Kurdstan.abdullah@univsul.edu.iq

John van der Hoek: School of Mathematics and Statistics, University of South Australia, Adelaide, Australia
Email address: John.vanderHoek@unisa.edu.au