Temporal aggregation of Markov switching models

Wai-Sum Chan\textsuperscript{a,}\,*, Li-Xin Zhang\textsuperscript{b}, Siu Hung Cheung\textsuperscript{c}

\textsuperscript{a}Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong, PR China
\textsuperscript{b}Department of Mathematics, Zhejiang University, Hangzhou 310028, PR China
\textsuperscript{c}Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong, PR China

Abstract

In this paper we investigate the effects of temporal aggregation of a class of Markov switching models known as MSN models. The growing popularity of the MSN models can be attributed to their inherit flexibility characteristics, allowing for heteroscedascity, asymmetry and excess kurtosis. The distributions of the process described by the basic MSN model and the model of the corresponding temporal aggregate data are derived. They belong to a general class of mixture normal distributions. The limiting behaviour of the aggregated MSN model, as the order of aggregation tends to infinity, is studied. We provide explicit formulae for the volatility, autocovariance, skewness and kurtosis of the aggregated processes. A numerical illustration of the results is given using S&P500 total return data.

\textit{JEL Classification}: C10

\textit{Keywords}: Autocovariance function; Characteristic function; High-order moments; Markov switching; Mixing sequence

1. Introduction

In econometric time-series analysis, research on the effects of temporal aggregation has important practical implications. The derivation of the mathematical link between the basic model and the model of the corresponding aggregate data reveals important information on parameter estimation, forecasting and hypothesis testing (Granger and Siklos, 1995; Teles and Wei, 2000; Haug, 2002; Souza and Smith, 2004). In many economic and financial applications, the problem of temporal aggregation arises when data are observed at a lower frequency than the data generation frequency of the underlying model. The resulting observed data, which is referred to as the aggregate series, contain less information, and may lead to a distorted view of the true model, leading to potential errors in decision making. Therefore, the study of the effects of temporal aggregation is important for making proper decisions that are based on aggregate data (Man, 2004).

Let \( \{r_t, t = 1, 2, \ldots\} \) be the basic series. Its \( m \)-component non-overlapping temporal aggregated series, which is denoted by \( \{R_T\}, T \geq 1 \), is given by

\[
R_T = \sum_{t=m(T-1)+1}^{mT} r_t.
\]
Temporal aggregation has been widely studied in linear time-series models. For the class of ARIMA models, Tiao (1972) and Wei and Stram (1988) investigate the limiting behaviour of the aggregated models as the order of aggregation \( m \) tends to infinity. For the finite order of aggregation, Stram and Wei (1986) study the autocorrelation structure of the aggregated ARIMA model, and hence derive its corresponding model orders. For more discussions on temporal aggregation of ARIMA models, see Wei (1990, Chapter 16).

In comparison to the abundant research on linear models, little is known about the effects of temporal aggregation on non-linear time-series models. Among the different types of non-linear models, the temporal aggregation of GARCH-type models has been explored by Drost and Numan (1993), and Meddahi and Renault (2004). Souza and Smith (2004) examine the aggregation effects of a stationary non-linear Auto-Regressive Fractionally Integrated Moving-Average (ARFIMA) process.

Recently, Markov switching models have gained popularity in the modelling of economic variables, such as foreign exchange rates (Dewachter, 2001); equity returns and interest rates (Hamilton and Raj, 2002); and inflation rates (Chopin and Pelgrin, 2004). These models allow for heteroscedascity, asymmetry and excess kurtosis. To better grasp the characteristics of this class of models, in particular serial correlation, asymmetry and tail behaviour, Timmermann (2000) gives a detailed account of the moments of three types of Markov switching models. In this paper, we consider the renowned class of mean-variance Markov switching normal (MSN) processes. A \( K \)-state MSN model is defined as

\[
\begin{align*}
\epsilon_t & \sim N(\mu_1, \sigma_1^2), \quad \text{if } S_t = 1, \\
\vdots & \vdots \\
\epsilon_t^{(K)} & \sim N(\mu_K, \sigma_K^2), \quad \text{if } S_t = K,
\end{align*}
\]

where \( S_t = 1, 2, \ldots, K \) denotes the unobservable state indicator, which follows an ergodic \( K \)-state Markov process; \( \epsilon_t^{(k)} \), for \( k = 1, \ldots, K \), are independent standard Gaussian white noise processes; \( \mu_{S_t} \) and \( \sigma_{S_t} \) are the state-dependent mean and standard deviation respectively. Let the transition matrix of the Markov process be denoted by the \( (K \times K) \) matrix \( P \) with elements \( (p_{ij}) \) and the corresponding steady-state probabilities be given by the \( (K \times 1) \) vector \( \pi = (\pi_1, \cdots, \pi_K)' \). That is

\[
P' \pi = \pi.
\]

For example, in the two-regime case where \( K = 2 \), we have

\[
P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},
\]

where \( 0 \leq p_{ij} \leq 1 \) for \( i = 1, 2 \), and \( j = 1, 2; \) \( (p_{11} + p_{12}) = (p_{21} + p_{22}) = 1; \)

\[
\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} \frac{p_{21}}{p_{11} + p_{21}} \\ \frac{p_{12}}{p_{12} + p_{22}} \end{pmatrix}.
\]

Thus far there is no study on the moment structure of the temporal aggregate series which is generated from a basic MSN model. In this paper we attempt to derive the mathematical link between the basic MSN process \( (r_t) \) and the model of the corresponding aggregate data \( (R_T) \). The rest of this paper is organised as follows. Section 2 provides the distributions of \( r_t \) and \( R_T \), while Section 3 studies the moments and autocovariance functions of the aggregate MSN model. Section 4 illustrates the results by a real example, and finally some concluding remarks are given in Section 5.
2. The distributions of \( r_t \) and \( R_T \)

2.1. The general case

A statistical distribution is uniquely specified by its characteristic function (Papoulis, 1984, p.155). In this section we derive the characteristic functions of \( r_t \) and \( R_T \).

Let \( \psi_k(s) \) be the characteristic function of the normal random variable \( e_t^{(k)} \sim N(\mu_k, \sigma_k^2) \) for \( k = 1, \ldots, K \), i.e.,

\[
\psi_k(s) = E[e^{is e_t^{(k)}}] = e^{i\mu_k s - \sigma_k^2 s^2/2}.
\] (5)

We first derive the characteristic function of \( r = (r_1, r_2, \cdots, r_n)' \), where \( r_t \) follows model (2), a \( K \)-State MSN process. Let

\[
\psi(s) = \mathcal{P} \times \text{Diag}(\psi_1(s), \cdots, \psi_K(s)),
\]

where \( \text{Diag}(V) \) is a matrix function which returns a diagonal square matrix with elements in the vector \( V \) as the corresponding diagonal elements. Define \( X_t = (X_{t,1}, \cdots, X_{t,K})' \) with \( X_{t,k} = I\{S_t = k\} \), where

\[
I\{S_t = k\} = \begin{cases} 
1 & \text{if } S_t = k, \\
0 & \text{if } S_t \neq k.
\end{cases}
\]

It is obvious that

\[
E[X_t'|S_t-1] = X_{t-1} \mathcal{P}.
\]

Let \( \mathcal{F}_t \) be the information set up to time \( t \). For a given \( s = (s_1, \cdots, s_n) \), we define \( Q_0 = 0 \) and \( Q_l = \sum_{t=1}^l s_t r_t \) for \( l = 1, \ldots, n \). Then we have

\[
E\left[ e^{iQ_n X_n'} | \mathcal{F}_{n-1} \right] = e^{iQ_n} E\left[ (X_{n,1} e^{i s_n r_n}, \cdots, X_{n,K} e^{i s_n r_n}) | \mathcal{F}_{n-1} \right] = e^{iQ_n} E\left[ (X_{n,1} e^{i s_n r_n}, \cdots, X_{n,K} e^{i s_n r_n}) | S_n | \mathcal{F}_{n-1} \right] = e^{iQ_n} E\left[ (X_{n,1} \psi_1(s_n), \cdots, X_{n,K} \psi_K(s_n)) | \mathcal{F}_{n-1} \right] = e^{iQ_n} E\left[ X_n'| \mathcal{F}_{n-1} \right] \times \text{Diag}(\psi_1(s_n), \cdots, \psi_K(s_n)) = e^{iQ_n} X_n'| \mathcal{F}_{n-1} \psi(s_n).
\]

It follows recursively that

\[
E\left[ e^{iQ_n X_n'} \right] = E\left[ e^{iQ_{n-1} X_n'} \right] \psi(s_n) = \cdots = e^{iQ_1 X_0'} \psi(s_1) \psi(s_1) \cdots \psi(s_{n-1}) \psi(s_n) = \pi' \psi(s_1) \cdots \psi(s_n).
\]

Therefore, the characteristic function of \( r = (r_1, r_2, \cdots, r_n)' \) is

\[
E[e^{is r}] = E\left[ e^{iQ_n} X_n' \right] = E\left[ e^{iQ_n} X_n' \right] 1 = \pi' \psi(s_1) \cdots \psi(s_n) 1,
\] (6)

where \( 1 = (1, \cdots, 1)' \). In particular, the characteristic function of \( r_t \) is

\[
\Psi_{r_t}(s) = \pi' \psi(s) 1.
\] (7)
Next, we derive the characteristic function of \( R = (R_1, R_2, \cdots, R_N)' \). Analogous to the above derivation of the characteristic function for \( r \), we obtain

\[
E[e^{is'R}] = \pi' [\psi(s_1)]^m \cdots [\psi(s_N)]^m 1,
\]

where \( s = (s_1, \cdots, s_N)' \). In particular, the characteristic function of \( R_T \) is

\[
\Psi_{R_T}(s) = \pi' [\psi(s)]^m 1 = \sum \beta_{t_1t_2\ldots t_K} [\psi_1(s)]^{t_1} \cdots [\psi_K(s)]^{t_K},
\]

where the summation is taken over all possible non-negative integer combinations of \( \{l_1, \ldots, l_K\} \) with the constraint \( \{l_1 + l_2 + \cdots + l_K = m\} \). The \( \beta \) coefficients satisfy

\[
\beta_{t_1t_2\ldots t_K} \geq 0 \quad \text{and} \quad \sum \beta_{t_1t_2\ldots t_K} = 1.
\]

It should be noted that \( [\psi_1(s)]^{t_1} \cdots [\psi_K(s)]^{t_K} \) in the equation (9) is a characteristic function of a normal distribution with mean \((l_1\mu_1 + \cdots + l_K\mu_K)\) and variance \((l_1\sigma_1^2 + \cdots + l_K\sigma_K^2)\).

Hence, the distribution of \( R_T \) is a multiple-point mixture of normal distributions. The mixing probabilities \( \beta_{t_1\ldots t_K} \) can be derived in the following way.

Notice that the characteristic function of \( R_T \) can be re-written as

\[
\Psi_{R_T}(s) = \pi' [P \times \text{Diag}(\psi_1(s), \cdots, \psi_K(s))]^m 1.
\]

Let

\[
P_k = P \times \text{Diag}(0, \cdots, 1, \cdots, 0)
\]

be the matrix which only keeps values in the \( k \)-th column of \( P \). Then, for any \((1 \times K)\) dimension real vector \((x_1, \cdots, x_K)\), we have

\[
[P \times \text{Diag}(x_1, \cdots, x_K)]^m = [x_1P_1 + \cdots + x_KP_K]^m = \sum \beta_{t_1\ldots t_K}x_1^{t_1} \cdots x_K^{t_K}.
\]

Hence,

\[
\beta_{t_1\ldots t_K} = \pi' \left[ \sum \frac{P_{1i} \cdots P_{Kj} \cdots P_{m}}{(#(i,j)=l_k ; k=1,\cdots,K)} \right] 1 \triangleq \pi' Q_{t_1\ldots t_K} 1,
\]

where the summation is taken over all possible sequence combinations of the product of \( m \) \( P_i \)'s (consisting of \( l_1P_1 \)'s, \( l_2P_2 \)'s, \ldots, \( l_KP_K \)'s with \( \sum l_i = m \)) and \( Q_{t_1\ldots t_K} = (q_{t_1\ldots t_K}) \). There is an interesting interpretation of \( Q_{t_1\ldots t_K} \). Let \( E_{t_1\ldots t_K} \) be the event that, from step 1 to step \( m \), the Markov Chain visits the State-\( k \) \( l_k \) times, for \( k = 1, 2, \ldots, K \), and \( l_1 + \cdots + l_K = m \). It can be seen that \( q_{t_1\ldots t_K} = \Pr(E_{t_1\ldots t_K}, S_m = j | S_0 = i) \) is the transition probability from \( S_0 = i \) to \( S_m = j \) through the path on \( E_{t_1\ldots t_K} \). Hence,

\[
\beta_{t_1\ldots t_K} = \Pr(E_{t_1\ldots t_K}).
\]

The distribution of \( R_T \) is a mixture of normal distributions consisting of \( N(l_1\mu_1 + \cdots + l_K\mu_K, l_1\sigma_1^2 + \cdots + l_K\sigma_K^2) \) with mixing probability \( \beta_{t_1\ldots t_K} \). Finally, the \( \beta \) coefficients can be obtained via the equation (10) or (11).
The joint distribution of $R_T$ and $R_{T+\ell}$ can be derived from the general characteristic function of $R$ in (8). For $\ell \geq 1$, the characteristic function of $(R_T, R_{T+\ell})$ is

$$
\pi' \left( [P \times \text{Diag}(\psi_1(s_1), \ldots, \psi_K(s_1))]^m \right) \frac{\mathcal{P}^{m(\ell-1)}}{\mathcal{P}^{m}} \left( [P \times \text{Diag}(\psi_1(s_2), \ldots, \psi_K(s_2))]^m \right).$$

(12)

The joint distribution of $(R_T, R_{T+\ell})$ is a mixture of two-dimensional normal distributions consisting of

$$
N(l_1\mu_1 + \cdots + l_K\mu_K, l_1\sigma_1^2 + \cdots + l_K\sigma_K^2) \otimes N(i_1\mu_1 + \cdots + i_K\mu_K, i_1\sigma_1^2 + \cdots + i_K\sigma_K^2)
$$

with mixing probability

$$
\{\pi' Q_{l,1}, \ldots, Q_{l,K} \mathcal{P}^{m(\ell-1)} Q_{i,1}, \ldots, Q_{i,K} \},
$$

(13)

$k = 1, \ldots, K$, and

$$
\sum_{k=1}^{K} l_k = \sum_{k=1}^{K} i_k = m.
$$

In fact, the mixing probability is the probability of the event that, from step 1 to step $m\ell$, the Markov chain visits each State-$k l_k$ times in the first $m$ steps, and visits each State-$k i_k$ times in the last $m$ steps, for $k = 1, \ldots, K$.

2.2. The case of $K = 2$

Users of MSN processes have found, in general, that two or three regimes (that is, $K = 2$ or $K = 3$) are often sufficient for modelling most economic variables. In particular, the two-regime model appears to be empirically popular and sufficient for most share return data (Timmermann, 2000; and Hardy, 2003, p.31). In this section we illustrate the application of the general results in Section 2.1 to derive distribution functions for the series $R_T$ which is aggregated temporally from the basic two-regime ($K = 2$) MSN process $r_t$. The following proposition provides the distribution formulae.

**Proposition 1.** Suppose that $r_t$ follows the two-regime stationary MSN process (2), (3) started from its steady state characterised by the set of unconditional probabilities (4). Let $R_T$ be the $m$-component non-overlapping temporal aggregated series (1). Then

(a) The distribution function $F_R(\cdot)$ of $R_T$ is a $(m + 1)$-point mixture of normal distributions, $Z_{l,m-1} \sim N(l_1\mu_1 + (m-l)\mu_2, l_1\sigma_1^2 + (m-l)\sigma_2^2)$ with mixing probability $\beta_{l,m-1}$, for $l = 0, \ldots, m$. That is

$$
F_R(x) = \sum_{l=0}^{m} \beta_{l,m-1} \Phi_{l,m-1}(x).
$$

with $\Phi_{l,m-1}(\cdot)$ is the distribution function of the $Z_{l,m-1}$ normal random variable,

$$
\beta_{l,m-1} = \Pr(E_{l,m-1})
$$

where $E_{l,m-1}$ is the event that, from step 1 to step $m$, the Markov chain visits “State-1” $l$ times and visits “State-2” $(m - l)$ times.
(b) For $\ell \geq 1$, the joint distribution function of $(R_T, R_{T+\ell})$ is

$$F_T(x, y) = F_R(x)F_R(y) + \pi_1\pi_2\phi^{m(\ell-1)+1}G_R(x)G_R(y),$$

where $\phi = 1 - p_{12} - p_{21}$,

$$G_R(y) = \sum_{l=0}^{m} \alpha_{l,m-l}\Phi_{l,m-l}(y), \quad \phi = (1 - p_{12} - p_{21})$$

and

$$\alpha_{l,m-l} = \Pr(E_{l-1,(m-1)-(l-1)}|S_0 = 1) - \Pr(E_{l,(m-1)-l}|S_0 = 2).$$

A proof of Proposition 1 is given in the appendix. Results for the case of $K = 3$ can be obtained similarly. Determination of the $\beta$ and $\alpha$ coefficients in the above proposition depends on the order of aggregation $m$. We illustrate the technical details of computing these coefficients in the following two examples.

**Example 1.** We consider the case of $m = 3$. The distribution of $R_T$ is

$$F_R(x) = \sum_{l=0}^{3} \beta_{l,3-l}\Phi_{l,3-l}(x),$$

which is a four-point mixture of normal distributions consisting of $N(3\mu_1, 3\sigma_1^2)$ with probability $\beta_{3,0}$; $N(2\mu_1 + \mu_2, 2\sigma_1^2 + \sigma_2^2)$ with probability $\beta_{2,1}$; $N(\mu_1 + 2\mu_2, \sigma_1^2 + 2\sigma_2^2)$ with probability $\beta_{1,2}$; and $N(3\mu_2, 3\sigma_2^2)$ with probability $\beta_{0,3}$; where

$$\begin{align*}
\beta_{3,0} &= \Pr(111) = \pi_1p_{11}^2, \\
\beta_{2,1} &= \Pr(112) + \Pr(121) + \Pr(211) = \pi_1p_{11}p_{12} + \pi_1p_{12}p_{21} + \pi_2p_{21}p_{11}, \\
\beta_{1,2} &= \Pr(221) + \Pr(212) + \Pr(122) = \pi_2p_{22}p_{21} + \pi_2p_{21}p_{12} + \pi_1p_{12}p_{22}, \\
\beta_{0,3} &= \Pr(222) = \pi_2p_{22}^2.
\end{align*}$$

Here $\Pr(uvw)$ denotes $\Pr(S_1 = u, S_2 = v, S_3 = w)$. Furthermore, the joint distribution function of $(R_T, R_{T+\ell})$ is

$$F_T(x, y) = F_R(x)F_R(y) + \phi^{3(\ell-1)+1}G_R(x)G_R(y),$$

where

$$G_R(y) = \sum_{l=0}^{3} \alpha_{l,3-l}\Phi_{l,3-l}(y) \quad \phi = (1 - p_{12} - p_{21})$$

with

$$\begin{align*}
\alpha_{3,0} &= \Pr(111|S_1 = 1) - \Pr(111|S_1 = 2) = p_{11}^2, \\
\alpha_{2,1} &= [\Pr(112|S_1 = 1) + \Pr(121|S_1 = 1) + \Pr(211|S_1 = 1)] \\
&\quad - [\Pr(112|S_1 = 2) + \Pr(121|S_1 = 2) + \Pr(211|S_1 = 2)] = p_{11}p_{12} + p_{12}p_{21} - p_{21}p_{11}, \\
\alpha_{1,2} &= [\Pr(221|S_1 = 1) + \Pr(212|S_1 = 1) + \Pr(122|S_1 = 1)] \\
&\quad - [\Pr(221|S_1 = 2) + \Pr(212|S_1 = 2) + \Pr(122|S_1 = 2)] = p_{12}p_{22} - p_{22}p_{21} - p_{21}p_{12}, \\
\alpha_{0,3} &= \Pr(222|S_1 = 1) - \Pr(222|S_1 = 2) = -p_{22}^2.
\end{align*}$$
Example 2. We consider the case of $m = 1$. This is a special case where $R_T = r_t$. The distribution function of $r_t$ is

$$F_r(x) = \pi_1 \Phi_1(x) + \pi_2 \Phi_2(x).$$

Notice that

$$\alpha_{1,0} = \Pr(1|S_1 = 1) - \Pr(1|S_1 = 2) = 1,$$
$$\alpha_{0,1} = \Pr(2|S_1 = 1) - \Pr(2|S_1 = 2) = -1.$$

The distribution joint function of $(r_t, r_{t+\ell})$, for $\ell \geq 1$, is

$$F_r(x)F_r(y) + \phi^\ell \pi_1 \pi_2 [\Phi_1(x) - \Phi_2(x)][\Phi_1(y) - \Phi_2(y)],$$

where $\Phi_k(x)$ is the distribution function of $\epsilon_t^{(k)} \sim N(\mu_k, \sigma_k^2)$, $k = 1, 2$.

2.3. The limiting case of $m \to \infty$

In this section we study the asymptotic distribution of $R_T$. In the case of the two-regime MSN model where $K = 2$, we have the following proposition.

Proposition 2. Suppose that $r_t$ follows the two-regime stationary MSN process (2), (3) started from its steady state characterised by the set of unconditional probabilities (4). Let $R_T$ be the $m$-component non-overlapping temporal aggregated series (1). Then

$$m^{-1/2} \sum_{t=1}^{m} (r_t - \mu) \overset{D}{\to} N(0, \sigma^2), \text{ as } m \to \infty,$$

where

$$\mu = \pi_1 \mu_1 + \pi_2 \mu_2,$$
$$\sigma^2 = \pi_1 \sigma_1^2 + \pi_2 \sigma_2^2 + \frac{p_{11} + p_{22}}{p_{12} + p_{21}} \pi_1 \pi_2 (\mu_1 - \mu_2)^2.$$

Furthermore, there is a standard Brownian motion $W(t)$ such that for any constant $\tau > 0$,

$$\sum_{t=1}^{m} r_t - m\mu = \sigma W(m) + o(m^{1/4}(\log m)^{9/4+\tau}) \text{ a.s. as } m \to \infty. \quad (14)$$

Hence, when $m \to \infty$, $\{r_t\}$ is approximately independent and

$$R_1, \cdots, R_T \text{ are approximately i.i.d. } N(m\mu, m\sigma^2) \text{ random variables.} \quad (15)$$

A proof of Proposition 2 is given in the appendix. In fact, we can also prove that (15) is true for the general case of $K \geq 2$. This proposition has an interesting interpretation. We can conclude from (15) and (14) that if the switching time interval is very short, or the time
horizon is very long, the Markov switching process can be approximated by an independent model.

3. Mean, variance, autocovariance, skewness and kurtosis

The characteristic function can be used to generate raw moments of a random variable (vector). Econometricians are often particularly interested in the mean, variance, autocovariance, skewness and kurtosis of their data. In this section we provide these moments more explicitly for the temporal aggregates $R_T$. Without loss of generality, we derive these moment formulae for the case of $K = 2$. The results are summarised in the following proposition.

**Proposition 3.** Suppose that $r_t$ follows the two-regime stationary MSN process (2), (3) started from its steady state characterised by the set of unconditional probabilities (4). Let $R_T$ be the $m$-component non-overlapping temporal aggregated series (1). Then the unconditional mean $(\mu_R)$, variance $(\sigma_R^2)$, lag-$\ell$ autocovariance $(\gamma_R(\ell))$, coefficient of skewness $(\sqrt{b_1})$ and coefficient of excess kurtosis $(b_2)$ of $R_T$ are given by

$$
\mu_R \equiv \mathbb{E}(R_T) = m(\pi_1\mu_1 + \pi_2\mu_2),
$$

$$
\sigma_R^2 \equiv \mathbb{E}[(R_T - \mu_R)^2] = m(\pi_1\sigma_1^2 + \pi_2\sigma_2^2) + (\mu_1 - \mu_2)^2\pi_1\pi_2A_2,
$$

$$
\gamma_R(\ell) \equiv \mathbb{E}[(R_T - \mu_R)(R_{T+\ell} - \mu_R)] = \pi_1\pi_2\phi^{m(\ell-1)+1}(\mu_1 - \mu_2)^2(1 - \phi^m)^2/(1 - \phi)^2, \quad \ell \geq 1,
$$

$$
\sqrt{b_1} \equiv \frac{\mathbb{E}[(R_T - \mu_R)^3]}{(\sigma_R^2)^{3/2}} = \pi_1\pi_2(\mu_1 - \mu_2)^3A_3 + 3(\sigma_1^2 - \sigma_2^2)(\mu_1 - \mu_2)A_2/\sigma_R^4,
$$

$$
b_2 \equiv \frac{\mathbb{E}[(R_T - \mu_R)^4]}{(\sigma_R^2)^2} - 3 = \pi_1\pi_2(\mu_1 - \mu_2)^4A_4 + 6(\mu_1 - \mu_2)^2(\sigma_1^2 - \sigma_2^2)A_3 + 3(\sigma_1^2 - \sigma_2^2)^2A_2/\sigma_R^4,
$$
where
\[
\phi = (1 - p_{12} - p_{21}),
\]
\[
A_2 = \frac{1 + \phi - 2\phi(1 - \phi^m)}{1 - \phi},
\]
\[
A_3 = m(\pi_2 - \pi_1) \left[ 1 + \frac{6\phi(1 + \phi^m)}{(1 - \phi)} \right] - (\pi_2 - \pi_1) \frac{6\phi(1 - \phi^m)}{(1 - \phi)^3}(1 + \phi),
\]
\[
A_4 = m^2(\pi_2 - \pi_1)^2 \frac{12\phi^{m+1}}{(1 - \phi)^2} + m \frac{1 + \phi}{(1 - \phi)^3} \left[ 1 + 10\phi + \phi^2 + 24\phi^{m+1} - 6\pi_1\pi_2(1 + 8\phi + \phi^2 + 20\phi^{m+1}) \right]
- \frac{2\phi(1 - \phi^m)}{(1 - \phi)^4} \left[ 7 + 22\phi + 7\phi^2 - 6\pi_1\pi_2(6 + 17\phi + 6\phi^2 + \phi^{m+1}) \right].
\]

A proof of Proposition 3 is given in the appendix. It is useful to study the formulae in the proposition through a numerical example. We consider a two-regime MSN process with \( \mu_1 = 1, \mu_2 = -3, \sigma^2_1 = 4 \) and \( \sigma^2_2 = 16 \). This set of parameters is close to those of the fitted MSN models for monthly excess return series in three major stock markets (Timmermann, 2000, p.80). Fig. 1 provides surface plots of the coefficient of skewness \( \sqrt{b_1} \) upon different order of aggregation \( m \) versus the probabilities of staying in the two states \( (p_{11}, p_{22}) \) which are varied over the grid \([0.01, 0.99]\). To investigate the impact of the order of aggregation on the skewness of the aggregate series, we consider \( m = 1, 3, 6, 12, 30, \infty \). The case of \( m \to \infty \) is approximated by setting \( m = 100,000 \). Fig. 2 shows the surface plots for the coefficient of excess kurtosis \( b_2 \).

When \( m = 1 \), the expressions in Proposition 3 reduce to those in Corollary 1 of Timmermann (2000). Hence, Fig. 1(a) and Fig. 2(a) match exactly to Fig. 2 in Timmermann (2000) as expected. When \( m > 1 \) with moderate size, temporal aggregation magnifies both the skewness and excess kurtosis. High, negative values of both skewness and excess kurtosis are obtained when both \( p_{11} \) and \( p_{22} \) are close to 0.99. On the other hand, a large positive skewness is obtained when \( p_{22} = 0.99 \) and \( p_{11} \) is at around 0.90; a large positive excess kurtosis is obtained when \( p_{11} = 0.99 \) and \( p_{22} \) is at around 0.90. Finally, when \( m \to \infty \), both \( \sqrt{b_1} \) and \( b_2 \) tend to zero for any combinations of \((p_{11}, p_{22})\) as expected from the result in Proposition 2.
Fig. 1. Skewness of a $m$-order aggregated two-regime MSN process with $\mu_1 = 1$, $\mu_2 = -3$, $\sigma_1^2 = 4$ and $\sigma_2^2 = 16$. 
Fig. 2. Excess kurtosis of a $m$-order aggregated two-regime MSN process with $\mu_1 = 1$, $\mu_2 = -3$, $\sigma_1^2 = 4$ and $\sigma_2^2 = 16$. 
Table 1), many insurance companies are now being asked to employ stochastic models to measure solvency risk created by their products with equity-linked guarantees. In the United States, the Life Capital Adequacy Subcommittee (LCAS) of the American Academy of Actuaries (AAA) issued the C-3 Phase II Risk-Based Capital (RBC) proposal in September 2004. The full report of the Subcommittee (LCAS, 2004) is published by AAA and is available from the Academy’s website. Monthly S&P500 total return index from December 1953 to December 2002 is chosen as the baseline series. A stochastic model should be fitted to the baseline data and the implied equity return distributions over three different holding periods (1 year, 5 years and 10 years) are used to determine capital requirements.

Table 1
Performance of G7 stock markets, December 2000 to December 2003

<table>
<thead>
<tr>
<th>Country</th>
<th>Market Index</th>
<th>Dec 2000</th>
<th>Dec 2003</th>
<th>Change(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>TSE 300 Stock Index</td>
<td>8934</td>
<td>8293</td>
<td>-7.17%</td>
</tr>
<tr>
<td>France</td>
<td>CAC 40 Stock Index</td>
<td>5926</td>
<td>3597</td>
<td>-39.30%</td>
</tr>
<tr>
<td>Germany</td>
<td>DAX Stock Index</td>
<td>6433</td>
<td>3965</td>
<td>-38.36%</td>
</tr>
<tr>
<td>Italy</td>
<td>BCI Stock Index</td>
<td>1939</td>
<td>1271</td>
<td>-34.45%</td>
</tr>
<tr>
<td>Japan</td>
<td>Nikkei 225 Stock Index</td>
<td>15096</td>
<td>12801</td>
<td>-15.20%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>FTSE 100 Stock Index</td>
<td>6222</td>
<td>4510</td>
<td>-27.52%</td>
</tr>
<tr>
<td>United States</td>
<td>S&amp;P 500 Stock Index</td>
<td>1320</td>
<td>1112</td>
<td>-15.76%</td>
</tr>
</tbody>
</table>

Let \( S_t \) be the monthly S&P500 total return index value at \( t \), for \( t = 0, 1, 2, \ldots, n \). Define

\[ r_t = \log \left( \frac{S_t}{S_{t-1}} \right) \]  

(16)

as the log return for the \( t \)-th month. The log return series for a \( m \)-month non-overlapping holding period can be constructed by

\[ R_T = \log \left( \frac{S_{mT}}{S_{m(T-1)}} \right) \]

(17)

\[ = \log \left[ \frac{S_{m(T-1)+1}}{S_{m(T-1)}} \right] \left( \frac{S_{m(T-1)+2}}{S_{m(T-1)+1}} \right) \cdots \left( \frac{S_{mT}}{S_{m(T-1)+m-1}} \right) \]

(18)

\[ = \log \left( \frac{S_{m(T-1)+1}}{S_{m(T-1)}} \right) + \log \left( \frac{S_{m(T-1)+2}}{S_{m(T-1)+1}} \right) + \cdots + \log \left( \frac{S_{mT}}{S_{m(T-1)+m-1}} \right) \]

(19)

\[ = \sum_{t=m(T-1)+1}^{mT} \left( r_t \right). \]

(20)

for \( T = 1, 2, \ldots, M \), and we assume that \( M = \left[ n/m \right] \) is an integer. The accumulation factor for the period \([m(T - 1), mT]\) is given by

\[ A_T = \left( \frac{S_{mT}}{S_{m(T-1)}} \right) \]

(21)

\[ = \exp(R_T). \]

(22)
Even though regulators do not prescribe a mandatory class of stochastic models for fitting the baseline data, the LCAS employs a two-regime MSN process because it captures many of the dynamics which have been observed in the data (LCAS, 2004, p.15). It is anticipated that the use of two-regime MSN models is highly persuasive in the United States. However, the baseline series (600 monthly data from 1953 to 2002) are too sparse to provide sufficient empirical information for the 1-year, 5-year and 10-year guarantee horizons. There are only 50 non-overlapping annual observations, 10 five-year observations and 5 ten-year observations that can be constructed from the baseline series.

The equation (20) is exactly the same as the equation (1). Therefore, results in this paper can be applied to study the characteristics of $R_T$ for 1-year, 5-year and 10-year holding periods (i.e., $m = 12, 60$ and 120 respectively). Fig. 3 shows monthly total returns on the S&P 500 index, together with estimated volatility, computed using a 12-month moving standard deviation of the log returns. Maximum likelihood estimates of the two-regime MSN model for this baseline series ($r_t$, from January 1954 to December 2002) are obtained:

$$\mu_1 = 0.01282, \quad \mu_2 = -0.00983,$$
$$\sigma_1^2 = 0.00121, \quad \sigma_2^2 = 0.00406,$$
$$p_{12} = 0.03377, \quad p_{21} = 0.15412.$$

Using formulae developed in the previous two sections, we are able to compute characteristics of the aggregate series $R_T$ implied by the fitted baseline MSN model. The results are summarised in Table 2.
Fig. 3. Monthly total returns and annual volatility, S&P 500

Table 2
Two-regime MSN model fitted to S&P500 log returns (monthly returns, 1953-2002). This table reports the summary characteristics implied by the fitted model. For comparative purpose, sample characteristics computed from the monthly and annual observed series are also shown.

<table>
<thead>
<tr>
<th></th>
<th>Baseline (monthly)</th>
<th>1-year</th>
<th>5-year</th>
<th>10-year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>Sample Size</td>
<td>600</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.00876</td>
<td>0.00875</td>
<td>0.10508</td>
<td>0.10499</td>
</tr>
<tr>
<td>Variance</td>
<td>0.00180</td>
<td>0.00180</td>
<td>0.02669</td>
<td>0.02624</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.57209</td>
<td>-0.38769</td>
<td>-0.43816</td>
<td>-0.64312</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>2.34161</td>
<td>1.36182</td>
<td>-0.40688</td>
<td>1.14273</td>
</tr>
<tr>
<td>Autocorrelations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lag</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.017</td>
<td>0.034</td>
<td>0.010</td>
<td>0.056</td>
</tr>
<tr>
<td>2</td>
<td>-0.029</td>
<td>0.028</td>
<td>-0.173</td>
<td>0.005</td>
</tr>
<tr>
<td>3</td>
<td>0.032</td>
<td>0.023</td>
<td>0.042</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>0.014</td>
<td>0.018</td>
<td>0.170</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>0.095</td>
<td>0.015</td>
<td>-0.135</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>-0.041</td>
<td>0.012</td>
<td>-0.128</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>-0.010</td>
<td>0.010</td>
<td>0.026</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>-0.019</td>
<td>0.008</td>
<td>0.043</td>
<td>0.000</td>
</tr>
<tr>
<td>9</td>
<td>0.013</td>
<td>0.006</td>
<td>0.069</td>
<td>0.000</td>
</tr>
<tr>
<td>10</td>
<td>0.033</td>
<td>0.005</td>
<td>-0.047</td>
<td>0.000</td>
</tr>
<tr>
<td>11</td>
<td>-0.008</td>
<td>0.004</td>
<td>-0.107</td>
<td>0.000</td>
</tr>
<tr>
<td>12</td>
<td>0.048</td>
<td>0.003</td>
<td>0.020</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Note: The sample sizes are too small for computing the sample statistics for the 5-year and 10-year periods.
5. Concluding remarks

The class of Markov switching models has been widely employed in analysing economic data, as many economic theories are naturally expressed in terms of regimes and the transition from one regime to another is often described by exogenous stochastic processes. In applications Markov switching models have been specified for data at different frequencies. For examples, Turner et al. (1989) and Timmermann (2000) employ Markov switching models for analysing monthly stock return data, while Sarno and Valente (2000) and Chesnay and Jondeau (2001) study Markov regime-switching behaviour of weekly stock return series. In general, we say that a model is closed under temporal aggregation if the model keeps the same structure, with possibly different parameter values, for any data frequency. In this paper, we have shown that Markov switching models, in general, are not closed upon temporal aggregation. Even for the simple two-regime MSN case, the model is closed only for the following three trivial situations: (i) $p_{12} = 0$ or $p_{21} = 0$; (ii) $\phi = 0$; or (iii) $p_{11} = p_{22} = 0$.

Perez-Quiros and Timmermann (2001) propose a two-state Markov mixture model with normal and Student-$t$ distributions. This class of models can be extended to the mixture of Gaussian and any heavy-tailed distributions (Rachev, 2003) that can capture outliers in the data. It might be able to generalise the results in this paper to these classes of Markov mixture models. For $K = 2$, we can try to replace the second characteristic function in (5) by the characteristic function of the heavy-tailed distribution. However, characteristic functions of heavy-tailed distributions are often very complicated (see, e.g., Ortobelli et al., 2003). Explicit expressions for the moment structure of temporal aggregates as in Section 3 of this paper might not be easy to obtain. Research in this direction is in process.
Appendix

Proof of Proposition 1. Part (a) of the proposition can be easily proved from the equations (9) and (11). As for Part (b) of the proposition, we denote \( \phi = 1 - p_{12} - p_{21} \). It should be noted that the \( s \)-step transition matrix of the Markov process is

\[
\mathcal{P}^s \triangleq (p_{ij}^{(s)}) = (\pi_1, \pi_2)'(1, 1) + \phi^s(\pi_2, -\pi_1)'(1, -1).
\]

(A.1)

As shown in Section 2.1, the joint distribution of \((R_T, R_{T+\ell})\) is a mixture of two-dimensional normal distributions with mixing probability (13). Now, we look into the mixing probability,

\[
\pi' \mathcal{Q}_{i,m-i} \mathcal{P}^{m(\ell-1)} \mathcal{Q}_{j,m-j} 1
\]

\[
= \pi' \mathcal{Q}_{i,m-i} 1 \cdot \pi' \mathcal{Q}_{i,m-j} 1 + \phi^{m(\ell-1)} \pi' \mathcal{Q}_{i,m-i} (\pi_2, -\pi_1)'(1, -1) \mathcal{Q}_{j,m-j} 1
\]

\[
= \Pr(E_{i,m-i}) \Pr(E_{j,m-j}) + \phi^{m(\ell-1)} \pi' \mathcal{Q}_{i,m-i} (\pi_2, -\pi_1)'(1, -1) \mathcal{Q}_{j,m-j} 1.
\]

For the two-regime case where \( K = 2 \), the Markov chain is time reversible, and the event \( E_{i,m-i} \) is also time reversible. It follows that

\[
\Pr(E_{i,m-i}, S_m = i) = \Pr(E_{i,m-i}|S_m = i) \Pr(S_m = i) = \Pr(E_{i,m-i}|S_1 = i) \pi_i.
\]

Notice that

\[
\pi' \mathcal{Q}_{i,m-i}(\pi_2, -\pi_1)' = \pi_2 \Pr(E_{i,m-i}, S_m = 1) - \pi_1 \Pr(E_{i,m-i}, S_m = 2)
\]

\[
= \pi_1 \pi_2 [\Pr(E_{i,m-i}|S_1 = 1) - \Pr(E_{i,m-i}|S_1 = 2)]
\]

\[
= \pi_1 \pi_2 [\Pr(E_{i-1,(m-1)-(i-1)}|S_0 = 1) - \Pr(E_{i,(m-1)-i}|S_0 = 2)]
\]

(A.2)

and

\[
(1, -1) \mathcal{Q}_{j,m-j} 1 = \Pr(E_{j,m-j}|S_0 = 1) - \Pr(E_{j,m-j}|S_0 = 2)
\]

\[
= [p_{11} \Pr(E_{j,m-j}|S_1 = 1) + p_{12} \Pr(E_{j,m-j}|S_1 = 2)]
\]

\[
- [p_{21} \Pr(E_{j,m-j}|S_1 = 1) + p_{22} \Pr(E_{j,m-j}|S_1 = 2)]
\]

\[
= \phi [\Pr(E_{j,m-j}|S_1 = 1) - \Pr(E_{j,m-j}|S_1 = 2)].
\]

(A.3)

Combined (A.2) and (A.3), we have

\[
\pi' \mathcal{Q}_{i,m-i} \mathcal{P}^{m(\ell-1)} \mathcal{Q}_{j,m-j} 1 = \Pr(E_{i,m-i}) \Pr(E_{j,m-j})
\]

\[
+ \phi^{m(\ell-1)+1} \pi_1 \pi_2 [\Pr(E_{i,m-i}|S_1 = 1) - \Pr(E_{i,m-i}|S_1 = 2)]
\]

\[
\times [\Pr(E_{j,m-j}|S_1 = 1) - \Pr(E_{j,m-j}|S_1 = 2)].
\]

(A.4)

Finally, the joint distribution of \((R_T, R_{T+\ell})\) is

\[
F_\ell(x, y) = \sum_{i,j=0}^m \pi' \mathcal{Q}_{i,m-i} \mathcal{P}^{m(\ell-1)} \mathcal{Q}_{j,m-j} 1 \Phi_{i,m-j}(x) \Phi_{j,m-j}(y)
\]

\[
= F_R(x) F_R(y) + \phi^{m(\ell-1)+1} \pi_1 \pi_2 G_R(x) G_R(y)
\]

(A.5)
where

\[ F_R(x) = \sum_{l=0}^{m} \Pr(E_{l,m-l}) \Phi_{l,m-l}(y) = \sum_{l=0}^{m} \beta_{l,m-l} \Phi_{l,m-l}(y), \]

\[ G_R(y) = \sum_{l=0}^{m} [\Pr(E_{l,m-l}|S_1 = 1) - \Pr(E_{l,m-l}|S_1 = 2)] \Phi_{l,m-l}(y) \]

\[ = \sum_{l=0}^{m} [\Pr(E_{l-1,(m-1)-(l-1)}|S_0 = 1) - \Pr(E_{l,(m-1)-(l-1)}|S_0 = 2)] \Phi_{l,m-l}(y) \]

\[ \triangleq \sum_{l=0}^{m} \alpha_{l,m-l} \Phi_{l,m-l}(y). \]

\[ \square \]

**Proof of Proposition 2.** In order to prove the proposition, we need the following definitions for a mixing sequence and a lemma.

**Definitions:**

Let \( \{X_n; n = 0, \pm 1, \pm 2, \ldots \} \) be a sequence of random variables defined on \((\Omega, \mathcal{F}, \Pr)\). For \( n \leq m \), denote \( \mathcal{F}_n = \sigma(X_i; n \leq i \leq m) \). The sequence \( \{X_n; n = 0, \pm 1, \pm 2, \ldots \} \) is said to be \( \varphi \)-mixing if

\[ \varphi(\nu) \triangleq \sup_{X \in \mathcal{F}_n, Y \in \mathcal{F}_{n+\nu}} \text{Corr}\{X, Y\} \to 0, \text{ as } \nu \to \infty, \]

where the supremum is taken over all \( \mathcal{F}_n \)-measurable functions \( X \) and \( \mathcal{F}_{n+\nu} \)-measurable functions \( Y \) and all \( n \). The sequence is said to be \( \varphi \)-mixing if

\[ \varphi(\nu) \triangleq \sup_{A \in \mathcal{F}_n, B \in \mathcal{F}_{n+\nu}} |\Pr(B|A) - \Pr(B)| \to 0, \text{ as } \nu \to \infty, \]

where the supremum is taken over all \( \mathcal{F}_n \)-measurable events \( A \) and \( \mathcal{F}_{n+\nu} \)-measurable events \( B \) and all \( n \). The sequence is said to be double-side \( \varphi \)-mixing if

\[ \overline{\varphi}(\nu) \triangleq \sup_{A \in \mathcal{F}_n, B \in \mathcal{F}_{n+\nu}} \max\{|\Pr(B|A) - \Pr(B)|, |\Pr(A|B) - \Pr(A)|\} \to 0, \text{ as } \nu \to \infty. \]

**Lemma A.1.** For the two-regime MSN process where \( K = 2 \), the sequence \( r_t \) is a \( \varphi \)-mixing sequence with mixing coefficients \( \varphi(\nu) \leq |1 - p_{12} - p_{21}|^{\nu-1} \), and also a double-side \( \varphi \)-mixing sequence with mixing coefficients \( \overline{\varphi}(\nu) \leq |1 - p_{12} - p_{21}|^{\nu-1} \).

**Proof of Lemma A.1.** Denote \( \phi = (1 - p_{12} - p_{21}) \). Let \( \xi = (r_1, \ldots, r_n)' \), \( \eta_\nu = (r_{\nu+n}, \ldots, r_{\nu+n+m-1})' \). By (6), the joint characteristic function of \( \xi \) and \( \eta_\nu \) is

\[ \pi' \psi(t_1) \cdots \psi(t_n) \phi^{\nu-1} \psi(s_1) \cdots \psi(s_m) 1 \]

\[ = \pi' \psi(t_1) \cdots \psi(t_n) 1 \cdot \pi' \psi(s_1) \cdots \psi(s_m) 1 \]

\[ + \phi^{\nu-1} \pi' \psi(t_1) \cdots \psi(t_n)(\pi_2, -\pi_1)' \cdot (1, -1) \psi(s_1) \cdots \psi(s_m) 1. \]

It follows that the joint distribution function of \( \xi \) and \( \eta_\nu \) is

\[ F_\nu(x, y) = F(x)G(y) + \phi^{\nu-1} H_1(x)H_2(y), \quad x = (x_1, \ldots, x_n)', y = (y_1, \ldots, y_m)'. \]
Here, $F$ and $G$ are the distribution functions of $\xi$ and $\eta_\nu$, respectively; $H_1$ and $H_2$ are the inverse Fourier transform of

$$\pi' \psi(t_1) \cdots \psi(t_n)(\pi_2, -\pi_1)' \text{ and } (1, -1) \psi(s_1) \cdots \psi(s_m) 1,$$

respectively.

It should be noted that $H_1$ and $H_2$ are both linear combinations of multi-dimensional normal distributions and they do not depend on $\nu$.

Letting $\nu = 1$ yields

$$H_1(x) H_2(y) = F_1(x, y) - F(x) G(y).$$

Hence

$$F_\nu(x, y) - F(x) G(y) = \phi^{\nu - 1} [F_1(x, y) - F(x) G(y)]. \quad (A.6)$$

Now for any measurable function $f(x)$ and $f(y)$, according to (A.6) we have

$$\text{Cov}\{f(\xi), g(\eta_\nu)\} = \phi^{\nu - 1} \text{Cov}\{f(\xi), g(\eta_1)\}. \quad (A.7)$$

Notice that

$$|\text{Cov}\{f(\xi), g(\eta_1)\}| \leq \sqrt{\text{Var}\{f(\xi)\} \text{Var}\{g(\eta_1)\}} = \sqrt{\text{Var}\{f(\xi)\} \text{Var}\{g(\eta_\nu)\}}.$$

We have

$$|\text{Corr}\{f(\xi), g(\eta_\nu)\}| \leq |\phi|^{\nu - 1}.$$

By the arbitrariness of $f, g, n, m$ and the definition of the $\varphi$-mixing, we conclude that

$$\varphi(\nu) \leq |\phi|^{\nu - 1}.$$

Let $A$ and $B$ be Borel sets in $R^n$ and $R^m$, respectively. According to (A.7), we also get

$$|\text{Pr}(\xi \in A, \eta_\nu \in B) - \text{Pr}(\xi \in A) \text{Pr}(\eta_\nu \in B)|$$

$$= |\phi|^{\nu - 1} |\text{Pr}(\xi \in A, \eta_1 \in B) - \text{Pr}(\xi \in A) \text{Pr}(\eta_1 \in B)|$$

$$\leq |\phi|^{\nu - 1} |\{|\text{Pr}(\xi \in A, \eta_1 \in B)\} \setminus \{|\text{Pr}(\xi \in A) \text{Pr}(\eta_1 \in B)\}|$$

$$\leq |\phi|^{\nu - 1} \text{Pr}(\xi \in A).$$

That is

$$|\text{Pr}(\eta_\nu \in B | \xi \in A) - \text{Pr}(\eta_\nu \in B)| \leq |\phi|^{\nu - 1}.$$

Similarly,

$$|\text{Pr}(\xi \in A | \eta_\nu \in B) - \text{Pr}(\xi \in A)| \leq |\phi|^{\nu - 1}.$$

So, $\varphi(\nu) \leq |\phi|^{\nu - 1}$. This completes the proof of Lemma A.1. \qed

Lemma A.1 infers that the sequence $R_T$ is a $\varphi$-mixing sequence with mixing coefficients $\varphi(\nu) \leq |1 - p_{12} - p_{21}|^{(\nu - 1)}$, and also a double-side $\varphi$-mixing sequence with mixing coefficients $\varphi(\nu) \leq |1 - p_{12} - p_{21}|^{(\nu - 1)}$. It should be noted that the $\varphi$ and $\varphi$ mixing coefficients are decaying in an exponential rate. It implies that when $m$ is large, the dependence between $R_T$’s is often very weak. For example, when $m = 12$ and $(p_{11}, p_{22}) = (0.8, 0.7), \varphi(\nu)$ and $\varphi(\nu)$ are both less than 0.00024140625$^{\nu - 1}$. 

18
Now, since \( \{ r_1 \} \) is a \( \varphi \)-mixing sequence with \( \varphi(\nu) \leq |1 - p_{12} - p_{21}|^{\nu-1} \), (14) follows directly from the strong invariance principle for \( \varphi \)-mixing random variables (c.f., Corollary 9.1.1 of Lin and Lu (1997)). According to (14),

\[
\frac{R_i - m\mu}{\sqrt{m\sigma}} = \frac{W(im) - W((i-1)m)}{\sqrt{m}} + o(1), \text{ a.s., } i = 1, \ldots, T.
\]

Notice that \( \frac{W(im) - W((i-1)m)}{\sqrt{m}}, i = 1, \ldots, T \), are i.i.d. \( N(0,1) \) random variables. The proof of (15) is completed. This completes the proof of Proposition 2.

**Proof of Proposition 3.** The mean \((\mu_R)\) can be obtained directly from expectation. Next, we consider the lag-\( \ell \) autocovariance \((\gamma_R(\ell))\). Suppose that \( K = 2 \) and \( \ell \geq 1 \), by (A.5), we have

\[
\gamma_R(\ell) = \text{Cov}\{R_T, R_{T+\ell}\} = \phi^{m(\ell-1)+1} \pi_1 \pi_2 \left( \int x dG_R(x) \right)^2.
\]

Notice that

\[
\sum_{l=0}^{m} l \text{Pr}(E_{l,m-l} | S_1 = u) = \mathbb{E} [\# \{ S_j = 1 \} | S_1 = u]
\]

\[
= \sum_{j=1}^{m} \text{Pr}(S_j = 1 | S_1 = u) = \sum_{j=1}^{m} p_{u1}^{(j-1)}
\]

and

\[
\sum_{l=0}^{m} (m - l) \text{Pr}(E_{l,m-l} | S_1 = u) = \mathbb{E} [\# \{ S_j = 2 \} | S_1 = u]
\]

\[
= \sum_{j=1}^{m} \text{Pr}(S_j = 2 | S_1 = u) = \sum_{j=1}^{m} p_{u2}^{(j-1)}.
\]

Then, by (A.1) we have

\[
\int x dG_R(x) = \sum_{l=0}^{m} [\text{Pr}(E_{l,m-l} | S_1 = 1) - \text{Pr}(E_{l,m-l} | S_1 = 2)] (l\mu_1 + (m - l)\mu_2)
\]

\[
= \sum_{j=1}^{m} (p_{1j}^{(j-1)} - p_{2j}^{(j-1)}) \mu_1 + \sum_{j=1}^{m} (p_{2j}^{(j-1)} - p_{22}^{(j-1)}) \mu_2
\]

\[
= \sum_{j=1}^{m} \phi^{j-1} (\mu_1 - \mu_2) = \frac{1 - \phi^{m}}{1 - \phi} (\mu_1 - \mu_2).
\]

Hence,

\[
\gamma_R(\ell) = \phi^{m(\ell-1)+1} \pi_1 \pi_2 (\mu_1 - \mu_2)^2 (1 - \phi^{m})^2 / (1 - \phi)^2, \quad \ell \geq 1.
\]

Finally, we consider the variance \((\sigma_R^2)\), coefficient of skewness \((\sqrt{\vartheta_1})\) and coefficient of
excess kurtosis \((b_2)\) of \(R_T\). Let \(N_k = \#\{j : S_j = k, 1 \leq j \leq m\}, k = 1, \ldots, K\). Then

\[
E[(R_T - E R_T)^n] = \sum_{j=0}^{n} \binom{n}{j} c_j \sum l_1 l_2 \ldots l_K (l_1 \sigma_1^2 + \cdots + l_K \sigma_K^2)^{j/2} (l_1 (\mu_1 - \mu) + \cdots + l_K (\mu_K - \mu))^{n-j}
\]

\[
= \sum_{j=0}^{n} \binom{n}{j} c_j E \left[ (N_1 \sigma_1^2 + \cdots + N_K \sigma_K^2)^{j/2} (N_1 (\mu_1 - \mu) + \cdots + N_K (\mu_K - \mu))^{n-j} \right],
\]

where \(c_j\) is the \(j\)-th moment of a standard normal random variable. We consider the case of \(K = 2\).

\[
E[(R_T - E R_T)^n] = \sum_{j=0}^{n} \binom{n}{j} c_j E \left[ (N_1 \sigma_1^2 + N_2 \sigma_2^2)^{j/2} (N_1 (\mu_1 - \mu) + N_2 (\mu_2 - \mu))^{n-j} \right]
\]

\[
= \sum_{j=0}^{n} \binom{n}{j} c_j E \left[ (N_1 (\sigma_1^2 - \sigma_2^2) + m \sigma_2^2)^{j/2} (N_1 - m \pi_1)^{n-j} (\mu_1 - \mu_2)^{n-j} \right].
\]

In particular,

\[
\text{Var}\{R_T\} = E[(R_T - E R_T)^2] = \sum_{l=0}^{m} \beta_{l,n-l} (l - m \pi_1)^2 (\mu_1 - \mu_2)^2 + \sum_{l=0}^{m} \beta_{l,n-l} (l (\sigma_1^2 - \sigma_2^2) + m \sigma_2^2)
\]

\[
= m (\pi_1 \sigma_1^2 + \pi_2 \sigma_2^2) + (\mu_1 - \mu_2)^2 \text{Var}\{N_1\}, \tag{A.8}
\]

\[
E[(R_T - E R_T)^3] = \sum_{l=0}^{m} \beta_{l,n-l} (l - m \pi_1)^3 (\mu_1 - \mu_2)^3
\]

\[
+ 3 \sum_{l=0}^{m} \beta_{l,n-l} (l (\sigma_1^2 - \sigma_2^2) + m \sigma_2^2)(l - m \pi_1) (\mu_1 - \mu_2)
\]

\[
= E[(N_1 - m \pi_1)^3] (\mu_1 - \mu_2)^3 + 3 (\sigma_1^2 - \sigma_2^2) (\mu_1 - \mu_2) \text{Var}\{N_1\} \tag{A.9}
\]

and

\[
E[(R_T - E R_T)^4] = (\mu_1 - \mu_2)^4 E[(N_1 - m \pi_1)^4] + 6 (\mu_1 - \mu_2)^2 (\sigma_1^2 - \sigma_2^2) E[(N_1 - m \pi_1)^3] + 6 (\mu_1 - \mu_2)^2 (\sigma_1^2 + \sigma_2^2) \text{Var}\{N_1\}
\]

\[
+ 3 \text{Var}\{N_1\} (\sigma_1^2 - \sigma_2^2)^2 + 3 (m (\pi_1 \sigma_1^2 + \pi_2 \sigma_2^2))^2.
\]
So,
\[
E[(R_T - ER_T)^4] - 3[Var\{R_T\}]^2
= (\mu_1 - \mu_2)^4 \left[ E[(N_1 - m\pi_1)^4] - 3[Var\{N_1\}]^2 \right] \\
+ 6(\mu_1 - \mu_2)^2(\sigma_1^2 - \sigma_2^2)E[(N_1 - m\pi_1)^3] + 3Var\{N_1\}(\sigma_1^2 - \sigma_2^2)^2. \quad (A.10)
\]

Next, we need to compute the values of $Var\{N_1\}, E[(N_1 - EN_1)^3]$ and $E[(N_1 - EN_1)^4]$. Let $I_j = I\{S_j = 1\}$. It is easily seen that
\[
EN_1 = E[\sum_j I_j] = m\pi_1
\]
and
\[
Var\{N_1\} = E[N_1^2] - (EN_1)^2 = E[\sum I_i I_j + \sum_j I_j] - (m\pi_1)^2 \\
= m\pi_1 + 2 \sum_{j>i} p_{11}^{(j-i)} \pi_1 - (m\pi_1)^2 \\
= m\pi_1 + 2 \sum_{j>i} (\pi_1 + \phi^{j-i}\pi_2)\pi_1 - (m\pi_1)^2 \\
= \frac{1 + \phi}{1 - \phi} \frac{m\pi_1\pi_2}{\pi_1\pi_2} - \frac{2\phi(1 - \phi^m)}{(1 - \phi)^2} \pi_1\pi_2. \quad (A.11)
\]

For the third moment, we have
\[
EN_1^3 = E \left[ \sum_j I_j^3 + 3 \sum_{i\neq j} I_i^2 I_j + \sum_{i\neq j \neq k} I_i I_j I_k \right] \\
= E \left[ \sum_j I_j + 3 \sum_{i\neq j} I_i I_j + \sum_{i\neq j \neq k} I_i I_j I_k \right] \\
= EN_1 + 3(EN_1^2 - EN_1) + 6E \left[ \sum_{k>j>i} I_i I_j I_k \right].
\]
Notice that by (A.1),
\[
E \left[ \sum_{k>j>i} I_i I_j I_k \right] = \sum_{k>j>i} p_{11}^{(k-j)} p_{11}^{(j-i)} \pi_1 = \sum_{k>j>i} (\pi_1 + \phi^{k-j}\pi_2) (\pi_1 + \phi^{j-i}\pi_2) \pi_1 \\
= \frac{1}{6} (m^3 - 3m^2 + 2m)\pi_1^3 + \frac{1}{6} \phi \left[ m^2 + \frac{2(1 - \phi^m)}{(1 - \phi)^2} - m \right] \\
+ \frac{\phi}{(1 - \phi)^2} \left[ m\phi + m\phi^m + \frac{2\phi(\phi^m - 1)}{1 - \phi} \right] \pi_1^2 \pi_1.
\]
Finally, for the third central moment of \( N_1 \), we have

\[
\mathbb{E}[(N_1 - \mathbb{E}N_1)^3] = \mathbb{E}[N_1^3] - 3\mathbb{E}[N_1^2]\mathbb{E}N_1 + 3(\mathbb{E}N_1)^2\mathbb{E}N_1 - (\mathbb{E}N_1)^3
\]

\[
= 3\text{Var}\{N_1\} + 3(\mathbb{E}N_1)^2 - 2\mathbb{E}N_1 - 3\text{Var}\{N_1\}\mathbb{E}N_1 - (\mathbb{E}N_1)^3 + 6\mathbb{E}\left[ \sum_{k>j>i} I_i I_j I_k \right]
\]

\[
= 3 \left[ \frac{m}{1 - \phi} \pi_1 \pi_2 - \frac{2\phi(1 - \phi^m)}{(1 - \phi)^2} \pi_1 \pi_2 \right]
\]

\[
+ 3(m\pi_1)^2 - 2m\pi_1 - 3m\pi_1 \left[ \frac{m}{1 - \phi} \pi_1 \pi_2 - \frac{2\phi(1 - \phi^m)}{(1 - \phi)^2} \pi_1 \pi_2 \right]
\]

\[
+ (-3m^2 + 2m)\pi_1 + 6\pi_1^2 \pi_2 \left[ \frac{\phi}{1 - \phi} \left[ m^2 + \frac{2(1 - \phi^m)}{(1 - \phi)^2} - m \frac{3 - \phi}{1 - \phi} \right] \right.
\]

\[
+ \left. 6 \frac{\phi}{(1 - \phi)^2} \left[ m\phi + m\phi^m + \frac{2\phi(\phi^m - 1)}{1 - \phi} \right] \pi_2^2 \pi_1 \right]
\]

\[
= m(\pi_2 - \pi_1)\pi_1 \pi_2 \left[ 1 + \frac{6\phi(1 + \phi^m)}{(1 - \phi)^2} \right] + \pi_1 \pi_2 \frac{6\phi(1 - \phi^m)}{(1 - \phi)^3} (\pi_1 - \pi_2)(1 + \phi). \quad (A.12)
\]

For the fourth moment, we have

\[
\mathbb{E}N_1^4 = \mathbb{E}\left[ \sum_j I_j^4 + 4 \sum_{i \neq j} I_i I_j^3 + 3 \sum_{i \neq j} I_i^2 I_j^2 + 6 \sum_{i \neq j \neq k} I_i I_j I_k + \sum_{i \neq j \neq k \neq l} I_i I_j I_k I_l \right]
\]

\[
= \mathbb{E}\left[ \sum_j I_j + 7 \sum_{i \neq j} I_i I_j + 6 \sum_{i \neq j \neq k} I_i I_j I_k + \sum_{i \neq j \neq k \neq l} I_i I_j I_k I_l \right]
\]

\[
= \mathbb{E}N_1 + 7[\mathbb{E}N_1^2 - \mathbb{E}N_1] + 6[\mathbb{E}N_1^3 - 3\mathbb{E}N_1^2 + 2\mathbb{E}N_1] + \mathbb{E}\left[ \sum_{i \neq j \neq k \neq l} I_i I_j I_k I_l \right]
\]

\[
= 6\mathbb{E}N_1^3 - 11\mathbb{E}N_1^2 + 6\mathbb{E}N_1 + 24\mathbb{E}\left[ \sum_{l>k>j>i} I_i I_j I_k I_l \right]
\]

and

\[
\mathbb{E}\left[ \sum_{l>k>j>i} I_i I_j I_k I_l \right] = \sum_{l>k>j>i} \left( p_{11}^{(1-k)} p_{11}^{(k-j)} p_{11}^{(j-i)} \pi_1 \right)
\]

\[
= \sum_{l>k>j>i} (\pi_1 + \phi^{l-k}\pi_2) (\pi_1 + \phi^{k-j}\pi_2) (\pi_1 + \phi^{j-i}\pi_2) \pi_1.
\]
Furthermore,

\[ E(N_1 - EN_1)^4 - 3(\text{Var}\{N_1\})^2 = E[N_1^4] - 4E[N_1^3]EN_1 + 6E[N_1^2](EN_1)^2 - 3(EN_1)^4 - 3(\text{Var}\{N_1\})^2 \\
= (6 - 4EN_1)E[N_1^3] - 11EN_1^2 + 6EN_1 + 6E[N_1^2](EN_1)^2 - 3(EN_1)^4 \\
+ 24E \left[ \sum_{L>k>j>i} I_i I_j I_k I_l \right] - 3(\text{Var}\{N_1\})^2 \\
= (6 - 4EN_1) \left[ E[(N_1 - EN_1)^3] + (EN_1)^3 \right] + 6\text{Var}\{N_1\} \left[ 3EN_1 - (EN_1)^2 \right] \\
- 11[\text{Var}\{N_1\} + (EN_1)^2] + 6EN_1 + 3(EN_1)^4 \\
+ 24E \left[ \sum_{L>k>j>i} I_i I_j I_k I_l \right] - 3(\text{Var}\{N_1\})^2 \\
= m^2 \frac{12\phi^{m+1}}{(1 - \phi)^2 \pi_1 \pi_2 (\pi_2 - \pi_1)^2} \\
+ m \frac{1 + \phi}{(1 - \phi)^3} \pi_1 \pi_2 \left[ 1 + 10\phi + \phi^2 + 24\phi^{m+1} - 6\pi_1 \pi_2 (1 + 8\phi + \phi^2 + 20\phi^{m+1}) \right] \\
- \frac{2\phi(1 - \phi^m)}{(1 - \phi)^4} \pi_1 \pi_2 \left[ 7 + 22\phi + 7\phi^2 - 6\pi_1 \pi_2 (6 + 17\phi + 6\phi^2 + \phi^{m+1}) \right]. \\
\text{(A.13)}

The tedious expansion from (A.13) to (A.14) can be easily performed by many symbolic algebra software such as Maple, Matlab and Mathematica.

Finally, the proof is completed by substituting (A.11), (A.12) and (A.14) into (A.8), (A.9) and (A.10), respectively. \(\square\)
References


