Economic Modelling 122 (2023) 106237

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

A regime-switching model of stock returns with momentum and mean reversion☆

Javier Giner [a], Valeriy Zakamulin [b,*]

[a] Department of Economics, Accounting and Finance, University of La Laguna, Camino La Hornera s/n, 38071, Santa Cruz de Tenerife, Spain
[b] School of Business and Law, University of Agder, Service Box 422, 4604 Kristiansand, Norway

ARTICLE INFO

JEL classification:
C32
C51
G10

Keywords:
Time-series momentum
Mean reversion
Bull and bear markets
Duration dependence
Semi-Markov model
Return autocorrelation function

ABSTRACT

A vast body of empirical literature documents the existence of short-term momentum and medium-term mean reversion in various financial markets. However, few theoretical models explain these two common phenomena. A Markov model, wherein the return process randomly switches between bull and bear states, can reproduce many stylized facts of financial asset returns, excluding the mean reversion. An important limitation of the Markov model is that the state termination probability does not depend on age. We develop a semi-Markov model wherein, following the empirical evidence, the state termination probability increases with age. We demonstrate that this model induces short-term return momentum and subsequent reversal. We calibrate our model to real-world data and show that the empirical results agree with our theoretical model.

1. Introduction

Numerous empirical studies have documented momentum and mean reversion, two all-pervading phenomena in financial markets. Momentum denotes a stock price tendency to continue moving in the same direction over a short run. For instance, if stock returns have been high in the recent past, they will likely remain high in the immediate future. The concept of mean reversion refers to a stock price tendency to revert to a trend path in the medium run. For example, if stock returns have been unusually high (low) in the past, they are likely to be unusually low (high) in the foreseeable future.

Momentum and mean reversion come in two flavors: cross-sectional and time-series. While the cross-sectional momentum (discovered by Jegadeesh and Titman (1993)) and mean reversion (first reported by De Bondt and Thaler (1985)) focus on the relative performance of stocks in the cross-section, the time-series momentum and mean reversion focus exclusively on a financial asset's performance. This study deals only with the time-series momentum and mean reversion. In particular, we examine the time-series dependence in the returns on a single stock or a stock market index. In this case, short-term momentum and medium-term mean reversion materialize as a positive return autocorrelation at short lags and a negative autocorrelation at longer lags.

Several earlier studies have examined the momentum and mean reversion in stock prices (Summers, 1986; Fama and French, 1988; Lo and MacKinlay, 1988; Poterba and Summers, 1988; Jegadeesh, 1991). For example, Poterba and Summers (1988) document that stock returns exhibit a positive autocorrelation over periods shorter than one year and a negative autocorrelation over longer periods. Fama and French (1988) find a negative autocorrelation in returns aggregated over three to five years. Later studies spotlight how one can exploit the momentum and mean reversion to beat the market. For example, Moskowitz et al. (2012) document a strategy of buying stocks with positive returns in the past 12 months and selling stocks with negative returns, which delivers superior performance in various financial markets. Subsequently, similar findings are reported by Georgopoulou and Wang (2016), Hurst et al. (2017), Lim et al. (2018), and many others. Balvers et al. (2000) present evidence of momentum and mean reversion and show how mean reversion can be exploited to generate abnormal returns; Balvers and Wu (2006) and Balvers et al. (2012) report similar results.

The empirical literature on momentum and mean reversion is vast and rapidly expanding. In contrast, there is a significant shortage of theoretical models that explain the presence of short-term momentum and subsequent mean reversion. Almost exclusively, these models are equilibrium models (Hong and Stein, 1999; Barberis and Shleifer, 2003;

☆ The authors are grateful to Angus C. Chu, two anonymous referees, and the participants at the 10th International Conference on Mathematical and Statistical Methods for Actuarial Sciences and Finance for many insightful comments and suggestions.

* Corresponding author.
E-mail addresses: jginer@ull.edu.es (J. Giner), valeriz@uia.no (V. Zakamulin).

https://doi.org/10.1016/j.econmod.2023.106237
Received 20 October 2022; Received in revised form 15 February 2023; Accepted 15 February 2023
Available online 17 February 2023

He and Li, 2015) that assume the existence of several types of traders in a financial market: rational traders, noisy (irrational) traders, momentum traders (trend-followers), and contrarian traders. Calibrating empirical data with these elaborate and complicated theoretical models is challenging or impossible. Furthermore, these models are difficult to solve analytically; therefore, the researchers must resort to numerical solutions.

Whereas it is challenging to explain the momentum and mean reversion phenomena, time-series modeling of these phenomena is relatively trivial using the autoregressive-moving average family of models. The simplest modeling is through the use of the second-order autoregressive process (Box et al., 2016 (Chapter 3); Khil and Lee, 2002). Under a particular set of parameters in this process, the return autocorrelation function represents a damped cosine wave. Another approach to modeling momentum and reversal is to add a cyclical component to a continuous return process (Stein and Pozharny, 2022).

This paper's novel approach uses a two-state regime-switching model of stock returns and demonstrates that under realistic conditions, the return process in this model exhibits momentum and subsequent reversion. In particular, this paper's model rests upon a simple, plausible, easy-to-understand assumption that the stock market switches between bull and bear states. Our model lies between the two extremes: the equilibrium models that explain the presence of momentum and mean reversion and the time-series models that take the existence of momentum and reversal as granted.

Modeling financial returns by switching between bull and bear states is not new. On the contrary, Markov switching models (MSM) of stock returns are well-known. These models can reproduce many stylized facts of financial asset returns, including heavy tails, negative skewness, volatility clustering, and short-term momentum (Timmermann, 2000; Frühwirth-Schnatter, 2006 (Chapter 10)). However, these models cannot reproduce mean reversion.

A severe limitation of an MSM is that a memoryless geometric distribution governs the state duration times. As a result, the state termination probability does not depend on the time already spent in that state. That is, there is no state duration dependence in an MSM. In contrast, many empirical studies document that the stock market states exhibit positive duration dependence (Cochran and Defina, 1995; Ohn et al., 2004; Harman and Zuehlke, 2007; Zakamulin, 2023). Positive duration dependence means that the longer a bull (bear) market lasts, the higher its probability of ending. Consequently, an MSM does not correctly represent the bull and bear market durations.

The primary approach to incorporate the duration dependence in a regime-switching model is to replace an MSM with a semi-Markov switching model (SMSM). An SMSM generalizes the MSM by allowing the state duration time to follow any probability distribution; however, a serious disadvantage of an SMSM is the lack of analytical tractability. Furthermore, all numerical computations rely on using complicated recursive algorithms. For example, D'Amico and Petroni (2012) derive, among other things, the solution for the volatility autocorrelation function in an SMSM for stock returns. This solution assumes the knowledge of the first passage time distribution and involves two recursive equations that must be solved numerically.

Alternatively, an SMSM can be realized as an expanded-state MSM (ESMSM) where several Markovian states represent one semi-Markovian state. Our choice is an ESMSM with a specific topology where a negative binomial distribution governs the state duration times. A negative binomial distribution exhibits positive duration dependence and reduces to a geometric distribution under particular parameter constraints. Compared to an original SMSM formulation (D'Amico and Petroni, 2012), an ESMSM formulation lacks flexibility but presents two crucial advantages. First, an ESMSM provides some degree of analytical tractability. Second, this formulation enables us to apply all well-established methods available for Markov models.

Our paper's most important theoretical results are as follows. First, we propose a conceptual construction of an ESMSM where the return process randomly switches between bull and bear states. Second, we determine the form of the solution to the return autocorrelation function in the general case. Third, we offer the analytical solution to the return autocorrelation function for the simplest case where two Markovian states represent each semi-Markovian state. To the best of our knowledge, this is the first time an SMSM has been solved analytically for the return autocorrelation function. This analytical solution significantly contributes to our fundamental understanding of the conditions shaping the return autocorrelation function.

In particular, we demonstrate that the return autocorrelation function in our ESMSM exhibits short-term momentum and medium-term mean reversion. Under realistic model parameters, the shape of the autocorrelation function represents a damped cosine wave that decays relatively quickly. Qualitatively, the shape of the return autocorrelation function remains the same in the general case, where many Markovian states represent each semi-Markovian state. In the general case, the return autocorrelation function can be computed using simple numerical methods.

One of the crucial advantages of our model is the ease of calibration to empirical data. We demonstrate the applicability of our theoretical model by calibrating it to the monthly returns on the Standard and Poor's Composite index and the Dow Jones Industrial Average index. We show that the model-implied autocorrelation function fits reasonably well with the empirically estimated autocorrelation function. In particular, our model correctly captures the duration of the short-term momentum that lasts about 12 months and subsequently reverses.

The rest of the paper is organized as follows. Subsequent Section 2 describes how the return autocorrelation function is computed in a two-state regime-switching model. Section 3 presents a conventional MSM and the return autocorrelation function in this model for completeness and comparability. Section 4 explains the construction of our ESMSM, offers the analytical solution to the return autocorrelation function for the simplest case, and demonstrates that the return autocorrelation function exhibits short-term momentum and medium-term mean reversion. Section 5 calibrates our model to empirical data and illustrates the goodness of fit, and finally, Section 6 concludes.

2. Return autocorrelation in a regime-switching model

Denote by X_t the period-t log return on a financial asset. We assume that X_t is a discrete-time stochastic process that randomly switches between two states (regimes): A and B. Formally, the state space of the process is $S_t \in \{A, B\}$. The return distribution depends on the state S_t in the following manner:

$$X_t = \begin{cases} \mu_A + \sigma_A z_t & \text{if } S_t = A, \\ \mu_B + \sigma_B z_t & \text{if } S_t = B, \end{cases}$$

where μ_A and σ_A are the mean and standard deviation of returns in state A, μ_B and σ_B are the mean and standard deviation of returns in state B, and z_t is an identically and independently distributed over time random variable with zero mean and unit variance.

Throughout the paper, we assume that state A is a bull state of the market, while state B is a bear state. A bull market is typically a high-return low-volatility state, whereas a bear market is a low-return high-volatility state.

The conditional probabilities $p_{IJ}(n) = Prob(S_{t+n} = J | S_t = I)$ are called multi-period transition probabilities. That is, $p_{IJ}(n)$ is the probability that the process transits from state I to state J over n periods. The n-period transition probability distribution of the process can be represented by a 2×2 transition probability matrix $\mathbf{P}(n)$:

$$\mathbf{P}(n) = \begin{pmatrix} p_{AA}(n) & p_{AB}(n) \\ p_{BA}(n) & p_{BB}(n) \end{pmatrix}.$$

$\boldsymbol{\pi} = [\pi_A, \pi_B]$ represents the vector of the steady-state (stationary or ergodic) probabilities. Specifically,

$$\pi_A = Prob(S_t = A), \quad \pi_B = Prob(S_t = B).$$

The return autocorrelation function ρ_n is defined by Timmermann (2000); Frühwirth-Schnatter (2006, Chapter 10))

$$\rho_n = \frac{E[X_t X_{t+n}] - \mu^2}{\sigma^2},$$

where

$$\mu = E[X_t] = \pi_A \mu_A + \pi_B \mu_B,$$
$$\sigma^2 = Var[X_t] = \pi_A \sigma_A^2 + \pi_B \sigma_B^2 + \pi_A \pi_B (\mu_A - \mu_B)^2,$$

$$E[X_t X_{t+n}] = \pi_A \mu_A \left(p_{AA}(n) \mu_A + p_{AB}(n) \mu_B \right) + \pi_B \mu_B \left(p_{BA}(n) \mu_A + p_{BB}(n) \mu_B \right).$$

The expression for the lag-n autocorrelation can be rewritten in the following form:

$$\rho_n = \frac{\pi_A \pi_B (\mu_A - \mu_B)^2 - (\mu_A - \mu_B)(\pi_A p_{AB}(n) \mu_A - \pi_B p_{BA}(n) \mu_B)}{\sigma^2}. \tag{1}$$

Notably, the return autocorrelation function depends on n only through transition probabilities $p_{AB}(n)$ and $p_{BA}(n)$. The computation of the n-period transition probabilities depends largely on whether the regime-switching model is a Markov or a semi-Markov model. We discuss the computation of the transition probabilities in the subsequent sections.

Typically, the return autocorrelation at any lag is very weak and escapes detection in empirical studies because it is usually statistically insignificant (Zakamulin and Giner, 2022). Therefore, reliable detection of return autocorrelation is only possible using returns aggregated over multiple periods. This idea was proposed by Fama and French (1988), who suggest using the first-order autocorrelation of k-period returns:

$$AC1(k) = Cor(X_{t+k,t+1}, X_{t,t-k+1}), \tag{2}$$

where

$$X_{t+k,t+1} = \sum_{i=1}^{k} X_{t+i}, \quad X_{t,t-k+1} = \sum_{i=1}^{k} X_{t-k+i}.$$

Proposition 1. *The first-order autocorrelation of k-period returns is given by*

$$AC1(k) = \frac{\mathbf{1}' \boldsymbol{U} \mathbf{1}}{\mathbf{1}' \boldsymbol{R} \mathbf{1}}, \tag{3}$$

where $\mathbf{1}$ is the $k \times 1$ vector of ones and $\boldsymbol{R}$ and $\boldsymbol{U}$ are the $k \times k$ matrices given by

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix},$$

$$\boldsymbol{U} = \begin{bmatrix} \rho_k & \rho_{k+1} & \rho_{k+2} & \cdots & \rho_{2k-1} \\ \rho_{k-1} & \rho_k & \rho_{k+1} & \cdots & \rho_{2k-2} \\ \rho_{k-2} & \rho_{k-1} & \rho_k & \cdots & \rho_{2k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1 & \rho_2 & \rho_3 & \cdots & \rho_k \end{bmatrix}, \tag{4}$$

where ρ_i is the lag i autocorrelation of X_t.

The proof is given in Zakamulin and Giner (2022).

Note that the return autocorrelation function ρ_n fully determines the first-order autocorrelation of k-period returns. $AC1(k)$ equals the sum of all elements of the matrix $\boldsymbol{U}$ divided by the sum of all elements of the matrix $\boldsymbol{R}$.

3. Return autocorrelation in a Markov model

In an MSM, the return process satisfies the "Markov property" (memorylessness):

$$Prob(S_{t+1} | S_t, \ldots, S_0) = Prob(S_{t+1} | S_t). \tag{5}$$

In words, the conditional probability distribution of the state process at a future time, $t+1$, depends only upon the present state at time t and not upon the past states at times $t-i$, for all $i \geq 1$.

Assume the following one-period transition probability matrix:

$$\mathbf{P} = \begin{pmatrix} p_{AA} & p_{AB} \\ p_{BA} & p_{BB} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}. \tag{6}$$

For instance, if the process is in state A, then over a single period, the process transits to state B with probability p_{AB} or remains in state A with probability $p_{AA} = 1 - p_{AB}$. The probability p_{AA} is called the self-transition probability of state A. Fig. 1 illustrates an MSM specified by the transition probability matrix in (6).

The state duration time is the random time of staying in a particular state. In an MSM, the state I duration time d_I follows the geometric distribution

$$Prob(d_I = n) = p_{II}^{n-1}(1 - p_{II}), \quad n \geq 1,$$

where p_{II} is the self-transition probability of state I, d_I is the duration time of state I, and n is the number of periods. In a Markov process, there is no duration dependence in the sense that

$$Prob(d_I > s + n | d_I \geq s) = Prob(d_I > n) \quad \forall \quad n, s \geq 1.$$

The intuition behind this property is as follows. If the process is in state I at time t, then the remaining state duration time does not depend on the time already spent in that state.

When the state duration times are geometrically distributed, the mean state duration times are given by

$$E[d_A] = \frac{1}{1 - p_{AA}} = \frac{1}{\alpha}, \quad E[d_B] = \frac{1}{1 - p_{BB}} = \frac{1}{\beta}, \tag{7}$$

where $E[d_A]$ and $E[d_B]$ denote the mean state A and B duration times, respectively.

Assuming that the transition probability matrix $\mathbf{P}$ is the same after each period, the n-period transition probability matrix can be computed as $\mathbf{P}(n) = \mathbf{P}^n$. The elements of the transition probability matrix $\mathbf{P}(n)$ are given by Hamilton (1994, Chapter 22)

$$\mathbf{P}(n) = \begin{pmatrix} p_{AA}(n) & p_{AB}(n) \\ p_{BA}(n) & p_{BB}(n) \end{pmatrix} = \begin{pmatrix} \pi_A + \pi_B(1-\alpha-\beta)^n & \pi_B - \pi_B(1-\alpha-\beta)^n \\ \pi_A - \pi_A(1-\alpha-\beta)^n & \pi_B + \pi_A(1-\alpha-\beta)^n \end{pmatrix}. \tag{8}$$

Note that in the limit when $n \to \infty$, the matrix $\mathbf{P}(n)$ converges to a matrix that contains the stationary probabilities in each row. Assuming that $\alpha + \beta < 1$, the transition probability $p_{AA}(n)$ ($p_{BB}(n)$) monotonically decreases from 1 (when $n = 0$) to π_A (π_B). In contrast, the transition probability $p_{AB}(n)$ ($p_{BA}(n)$) monotonically increases from 0 to π_B (π_A). The stationary probabilities satisfy the following condition $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$. This condition can be used to find the expression for the stationary probabilities

$$\pi_A = \frac{\beta}{\alpha + \beta}, \quad \pi_B = \frac{\alpha}{\alpha + \beta}. \tag{9}$$

In an MSM, the return autocorrelation function, given by Eq. (1), reduces to

$$\rho_n = \frac{\pi_A \pi_B (\mu_A - \mu_B)^2}{\sigma^2} (1 - \alpha - \beta)^n. \tag{10}$$

It is essential to note that the lag-n autocorrelation is always nonnegative, $\rho_n \geq 0$, given that $\alpha + \beta < 1$. Additionally, the autocorrelation is strictly positive if $\mu_B \neq \mu_A$. The autocorrelation exponentially decreases toward zero as n increases. Consequently, in an MSM, the return process exhibits a short-term momentum.

Fig. 2 illustrates the return autocorrelations in an MSM for monthly returns. Specifically, the red line with points plots the month-k return autocorrelation, and the blue line with points plots the first-order autocorrelation of k-month returns. The annualized mean state returns are $\mu_A = 20\%$ and $\mu_B = -30\%$. The annualized standard deviations

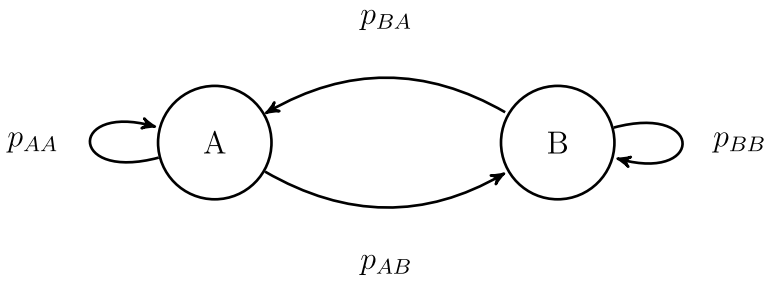

Fig. 1. A two-state Markov switching model. p_{IJ} denotes the conditional probability that the process transits from state I to state J over a single period.

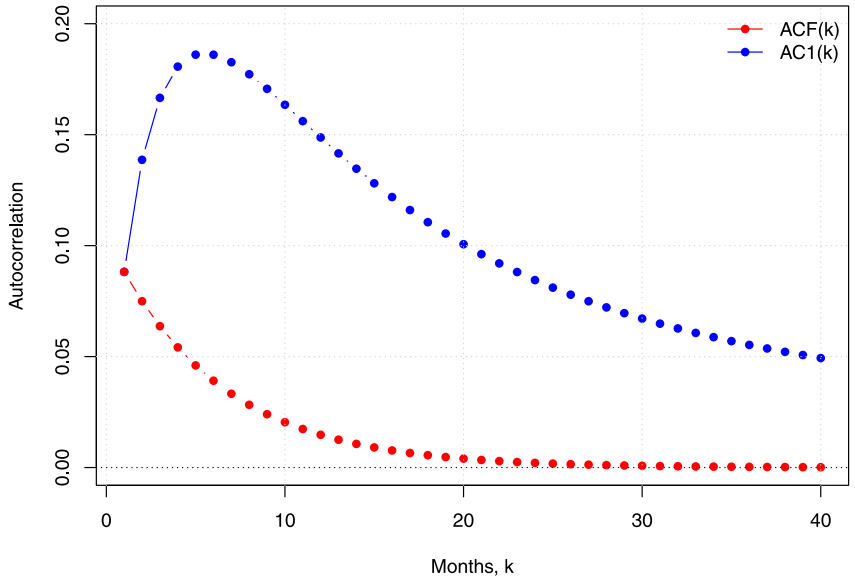

Fig. 2. The return autocorrelations in a Markov switching model. $ACF(k)$ denotes the month-k return autocorrelation ρ_k. $AC1(k)$ denotes the first-order autocorrelation of k-month returns. The annualized mean state returns are $\mu_A = 20\%$ and $\mu_B = -30\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 20\%$. The mean state A duration time equals 20 months, whereas the mean state B duration time equals 10 months.

of state returns are $\sigma_A = \sigma_B = 20\%$. The mean state A (bull market) duration time equals 20 months, whereas the mean state B (bear market) duration time equals 10 months. Note that the month-k return autocorrelation exponentially decreases toward zero as k increases. In contrast, the first-order autocorrelation of k-month returns quickly increases before gradually decreasing toward zero after reaching the maximum. It is worth emphasizing that the first-order autocorrelation of k-month returns is substantially larger than the month-k return autocorrelation for $k > 1$.

4. Return autocorrelation in a semi-Markov model

4.1. Preliminaries

A notable limitation of a conventional Markov model is that the state duration times are geometrically distributed. Consequently, the probability of state change does not depend on the time passed since the entry into the state. This behavior is not a reasonable representation of many real-world processes. For example, most empirical studies document that the stock market cycles exhibit duration dependence (Cochran and Defina, 1995; Maheu and McCurdy, 2000; Lunde and Timmermann, 2004; Ohn et al., 2004; Harman and Zuehlke, 2007; Zakamulin, 2023). The researchers often find positive duration dependence in the bull and bear market states. Positive duration dependence means that the longer a state lasts, the higher its probability of ending. The main approach to incorporate the duration dependence in a regime-switching model is to replace an MSM with an SMSM.

An SMSM generalizes an MSM by allowing the state duration time to follow any probability distribution. Contrary to a Markov model, an SMSM does not have the Markov property at each time t. The Markov property is satisfied only when the process changes the state. When a process enters state I, it determines the next state J according to the transition probability p_{IJ}; however, after J has been selected – but before making the transition – the process holds in state I for a random time d_{IJ}. Generally, a two-state SMSM is specified by a 2×2 transition probability matrix and 4 probability mass functions that determine the distribution of the state duration times when the process transits from state I to state J. A significant advantage of an SMSM is that it is very flexible and can incorporate any duration distribution; however, a serious disadvantage is that an SMSM does not provide analytical solutions to the state transition probabilities. Moreover, the state transition probabilities must always be computed using complicated recursive numerical algorithms (Howard, 1971; Barbu and Limnios, 2008).

Ferguson (1980) was the first to note that an SMSM can be realized as an expanded-state MSM (ESMSM). Russell and Cook (1987), Johnson (2005), Guèdon (2005), and Langrock and Zucchini (2011, Chapter 12) present overviews of various approaches to constructing an ESMSM. In an ESMSM, each state I of the process is represented by q substates in the conventional Markov model: $\{i_1, i_2, \ldots, i_q\}$. The state process is in macro-state I, $S_t = I$, provided the state process is in the set of q substates[1] $S_t \in \{i_1, i_2, \ldots, i_q\}$. Compared with an

[1] Note that in our definitions, a macro-state is a semi-Markovian state, while a substate is a Markovian state. In an ESMSM, a semi-Markovian state consists of several Markovian states.

authentic SMSM, an ESMSM is not flexible regarding the distribution of the state duration times; however, an immense advantage of using the ESMSM formulation is that this approach enables one to apply all well-established methods available for Markov models. For example, instead of implementing a complicated recursive numerical algorithm, one can compute the n-period transition probability matrix using matrix multiplication (power).

In an ESMSM, the state duration distributions depend on the chosen topology. We use an ESMSM with a specific topology where the state duration times follow a negative binomial distribution. In this topology, each q substate of a macro-state has self-transition, and the transition to the next macro-state is possible only from the last qth substate. We assume that the self-transition probability p_{ii} is the same in each substate $1, 2, \dots, q$ of state I. Under this assumption, the macro-state I duration time follows a negative binomial distribution $d_I \sim NB(q)$ (Johnson, 2005; Guèdon, 2005; Zhu et al., 2006; Tejedor et al., 2015). The probability mass function of the $NB(q)$ distribution is given by

$$f(n, q, p_{ii}) = Prob(d_I = n) = \binom{n-1}{n-q}(1-p_{ii})^q p_{ii}^{n-q}, \quad n \geq q. \tag{11}$$

The transition probability $1 - p_{ii}$ can be interpreted as the probability of success on one Bernoulli trial; thus, the negative binomial distribution gives the probability that the qth success will occur in the nth Bernoulli trial. The expected number of trials to obtain q successes (or equivalently, the macro-state I mean duration time) equals $E[d_I] = q/(1 - p_{ii})$. The geometric distribution is a special case of the negative binomial distribution when $q = 1$. Consequently, when $q = 1$ for all states, an ESMSM reduces to a conventional MSM.

The hazard function is a very convenient function for duration analysis,

$$h(n) = \frac{f(n)}{1 - F(n)},$$

where $f(n)$ is the probability mass function of the state durations and $F(n)$ is the corresponding cumulative distribution function. The hazard function provides a conditional failure rate. Specifically, the hazard function is a probability that the market state ends at time n under the condition that the market state lasts until time n. When the hazard function is constant, there is no duration dependence. Provided the absence of duration dependence, at any time, the probability that a market state ends is independent of how long the state has lasted. Positive duration dependence exists if a hazard function is an increasing function of time. In this case, the longer a market state lasts, the higher the probability it ends.

Fig. 3 illustrates various shapes of the negative binomial distribution $NB(q)$ for $q \in \{1, 2, 3, 4\}$ and the shapes of the corresponding hazard functions. For all q, the mean state duration time is always 20. Only for $q = 1$ the distribution of the state duration is memoryless. For $q > 1$, the state duration distribution exhibits positive duration dependence. In particular, the probability that a state terminates increases as it ages. Numerous researchers report that the negative binomial distribution describes many empirical phenomena much better than the geometric distribution (Levinson, 1986; Burshtein, 1996; Johnson, 2005).

4.2. Topology of ESMSM and functional form of the solution

We turn to the presentation of the topology of an ESMSM where the state duration times follow a negative binomial distribution. We begin with the simplest case, depicted in Fig. 4, where each macro-state A and B is represented by two substates. Specifically, macro-state A consists of substates 1 and 2, while macro-state B consists of substates 3 and 4. This ESMSM extends the conventional two-state MSM depicted in Fig. 1. In this ESMSM, the duration times for states A and B follow the $NB(2)$ distribution.

Before proceeding further, we must clarify our notation for a general state transition probability. The states written in capital letters I and J denote macro-states (or semi-Markovian states), while the states in lower case letters i and j or integer numbers denote substates (or Markovian states); therefore, p_{IJ} denotes the transition probability between two macro-states, whereas p_{ij} (or, for instance, p_{12}) denotes the transition probability between two substates.

The one-period transition probability matrix for the ESMSM in Fig. 4 is given by

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} = \begin{bmatrix} 1-2\alpha & 2\alpha & 0 & 0 \\ 0 & 1-2\alpha & 2\alpha & 0 \\ 0 & 0 & 1-2\beta & 2\beta \\ 2\beta & 0 & 0 & 1-2\beta \end{bmatrix}. \tag{12}$$

Each element p_{ij} of the transition probability matrix is defined in the usual manner: $p_{ij} = Prob(S_{t+1} = j | S_t = i)$. Note that the self-transition probabilities of substates 1 and 2 (3 and 4) are the same: $p_{11} = p_{22}$ ($p_{33} = p_{44}$). As a result, the transition probabilities from one substate of macro-state A (B) to either another substate or another macro-state are the same $p_{12} = p_{23}$ ($p_{34} = p_{41}$).

Provided that $\alpha < 1/2$ and $\beta < 1/2$, the matrix $\mathbf{P}$ is a stochastic matrix whose entries are nonnegative and whose rows all sum to 1. We constructed our ESMSM to reproduce the mean state duration times of the conventional two-state MSM in Fig. 1. For example, the mean state A duration time in the ESMSM is $2/(1 - p_{11}) = 2/(2\alpha) = 1/\alpha$, which is the same as the mean state A duration time in the conventional MSM. As a result, in our ESMSM, the one-period transition probabilities p_{IJ} are the same as in the corresponding traditional MSM (see below). Additionally, our ESMSM has the same stationary probabilities π_A and π_B. These features provide simple comparability between the ESMSM and the corresponding MSM.

In the ESMSM specified by the transition probability matrix in (12), the self-transition probability of macro-state A is computed as follows. If we know that the process is in macro-state A, then the process is equally likely[2] to be either in substate 1 or 2. If the process is in substate 1, then the probability of remaining in macro-state A is $p_{11} + p_{12}$. If the process is in substate 2, then the probability of remaining in macro-state A is $p_{21} + p_{22}$. Consequently, the probability p_{AA} is computed as $(p_{11} + p_{12})/2 + (p_{21} + p_{22})/2$. All other transition probabilities are computed in the same manner:

$$\begin{aligned} p_{AA} &= (p_{11} + p_{12} + p_{21} + p_{22})/2, \quad & p_{BA} &= (p_{31} + p_{32} + p_{41} + p_{42})/2, \\ p_{AB} &= (p_{13} + p_{14} + p_{23} + p_{24})/2, \quad & p_{BB} &= (p_{33} + p_{34} + p_{43} + p_{44})/2. \end{aligned} \tag{13}$$

It is easy to check that the ESMSM and MSM have the same one-period transition probabilities for states A and B. For example, $p_{AA} = 1 - \alpha$ in the ESMSM and MSM; however, the multi-period transition probabilities differ.

The n-period transition probability matrix in the ESMSM is given by

$$\mathbf{P}(n) = \mathbf{P}^n = \begin{bmatrix} p_{11}(n) & p_{12}(n) & p_{13}(n) & p_{14}(n) \\ p_{21}(n) & p_{22}(n) & p_{23}(n) & p_{24}(n) \\ p_{31}(n) & p_{32}(n) & p_{33}(n) & p_{34}(n) \\ p_{41}(n) & p_{42}(n) & p_{43}(n) & p_{44}(n) \end{bmatrix}. \tag{14}$$

The n-period transition probabilities of macro-states A and B are computed similarly to (13). For example, the n-period self-transition probability of state A is computed as $p_{AA}(n) = (p_{11}(n) + p_{12}(n) + p_{21}(n) + p_{22}(n))/2$.

We next consider the general case where macro-state A is represented by g substates, while macro-state B consists of p substates. In the general case, the state A duration time follows the $NB(g)$ distribution, while the state B duration time follows the $NB(p)$ distribution. In this

[2] When we observe that the process is in macro-state A, nothing distinguishes the substates in A from each other.

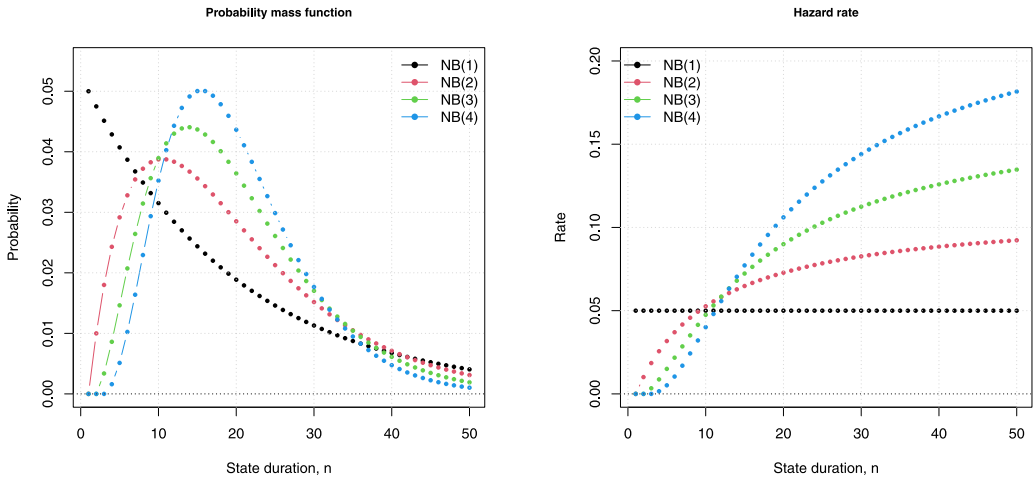

Fig. 3. The left panel shows the probability mass function of state durations for the negative binomial distribution $NB(q)$ for various q. The right panel shows the hazard rate for the negative binomial distribution $NB(q)$ for various q. For all q, the mean state duration time is always 20.

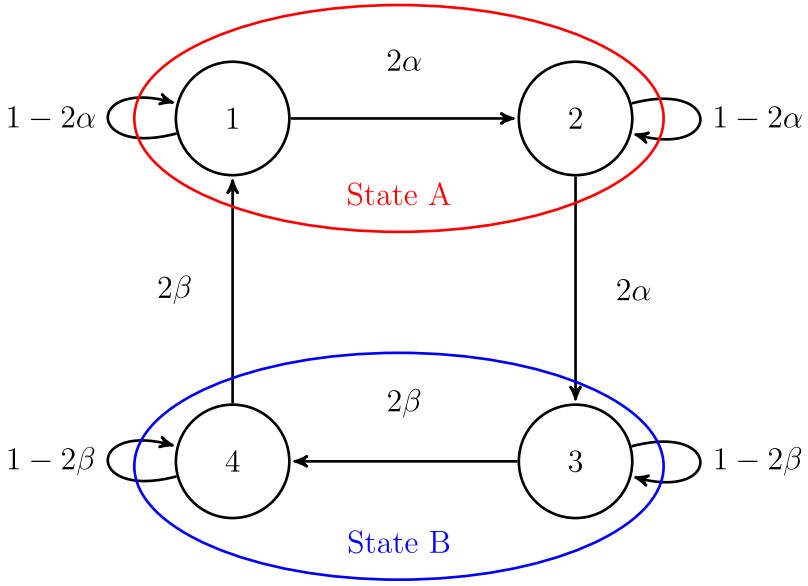

Fig. 4. An ESMSM with two substates for each macro-state A and B. Specifically, macro-state A consists of substates 1 and 2, while macro-state B is represented by substates 3 and 4.

case, the one-period $(g+p)\times(g+p)$ transition probability matrix $\mathbf{P}$ is given by the following partitioned matrix

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{P}_{AA} & \mathbf{P}_{AB} \\ \hline \mathbf{P}_{BA} & \mathbf{P}_{BB} \end{array}\right],$$

where $\mathbf{P}_{AA}$ is the $g\times g$ submatrix, $\mathbf{P}_{AB}$ is the $g\times p$ submatrix, $\mathbf{P}_{BA}$ is the $p\times g$ submatrix, and $\mathbf{P}_{BB}$ is the $p\times p$ submatrix.[3] For instance, the self-transition probability p_{AA} (p_{BB}) of macro-state A (B) is computed by summing all elements of submatrix $\mathbf{P}_{AA}$ ($\mathbf{P}_{BB}$) and dividing the result by g (p). Then, the complementary probability p_{AB} (p_{BA}) can be calculated as $p_{AB} = 1 - p_{AA}$ ($p_{BA} = 1 - p_{BB}$).

The n-period transition probability matrix is computed in the usual manner as $\mathbf{P}(n) = \mathbf{P}^n$. We next determine the functional form of the analytical solution for the elements of $\mathbf{P}^n$. Assuming that the matrix $\mathbf{P}$ is diagonalizable over the field of complex numbers,[4] one can determine the analytical solution through the diagonalization of $\mathbf{P}$. This method aims to find a diagonal matrix $\mathbf{D}$ that allows us to express $\mathbf{P}$ as $\mathbf{P} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{-1}$. Then, the nth power of $\mathbf{P}$ can be computed as $\mathbf{P}^n = \mathbf{Q}\mathbf{D}^n\mathbf{Q}^{-1}$.

The diagonalization procedure consists of the following steps. First, we find the eigenvalues λ_i, $i \in \{1, 2, \ldots, g+p\}$, of $\mathbf{P}$. The eigenvalues are the values of λ that satisfy the equation $|\mathbf{P} - \lambda\mathbf{I}| = 0$. Second, we determine the eigenvectors corresponding to each eigenvalue. The eigenvector $\mathbf{v}_i$ is found by solving the equation $(\mathbf{P} - \lambda_i\mathbf{I}) = \mathbf{v}_i$. The diagonal matrix $\mathbf{D}$ contains the eigenvalues along the main diagonal, $\mathbf{D} = \text{diag}(\lambda_1, \ldots, \lambda_{g+p})$; all off-diagonal elements of the matrix $\mathbf{D}$ equal zero. The matrix $\mathbf{Q}$ is composed of eigenvectors $\mathbf{Q} = [\mathbf{v}_1, \ldots, \mathbf{v}_{g+p}]$. Third, we find the inverse matrix $\mathbf{Q}^{-1}$, and finally, we perform the

[3] A limitation of our model is that the following conditions must be satisfied: $\alpha < 1/g$ and $\beta < 1/p$. Real-world processes typically meet these conditions when g and p are relatively small.

[4] The matrix $\mathbf{P}$ defined by (12) is diagonalizable because it has four distinct eigenvalues, see the proof of Proposition 2 in the subsequent section.

matrix multiplication

$$\mathbf{P}(n) = \mathbf{Q}\begin{bmatrix} \lambda_1^n & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_{g+p}^n \end{bmatrix}\mathbf{Q}^{-1}. \tag{15}$$

The general case is not analytically tractable; however, the computational method (15) for the elements of the n-period transition probability matrix allows us to deduce the functional form of the solution to the state transition probabilities. For example, the solution to the n-period transition probabilities $p_{AB}(n)$ and $p_{BA}(n)$ (used to compute the autocorrelation function given by Eq. (1)) has the following functional form

$$p_{IJ}(n) = c_{1,IJ}\lambda_1^n + c_{2,IJ}\lambda_2^n + \cdots + c_{g+p,IJ}\lambda_{g+p}^n,$$

where $c_{s,IJ}$, $s \in \{1, 2, \ldots, g+p\}$, are some functions of the one-period state transition probabilities in the corresponding conventional Markov model, $c_{s,IJ} = c_{s,IJ}(\alpha, \beta)$, λ_1 is the largest eigenvalue, λ_2 is the second-largest eigenvalue, and λ_{g+p} is the smallest eigenvalue. Hence, the first conclusion is that a sum of exponential functions represents the functional form of the solution.

As a rule, the largest eigenvalue of a stochastic matrix is 1, that is, $\lambda_1 = 1$. All other eigenvalues are in absolute value smaller than 1; therefore, each exponential function λ_s^n, $s > 1$, approaches zero as n increases. We know that in the limit as n increases, the n-period transition probability $p_{AB}(n)$ ($p_{BA}(n)$) approaches the stationary probability π_B (π_A). Consequently, the second conclusion is that $c_{1,IJ} = \pi_J$; thus, the above equation can be rewritten as

$$p_{IJ}(n) = \pi_J + c_{2,IJ}\lambda_2^n + \cdots + c_{p+q,IJ}\lambda_{g+p}^n.$$

A transition probability matrix may have complex eigenvalues. These eigenvalues always occur in complex conjugate pairs; hence, the transition probability $p_{IJ}(n)$ approaches π_J in an oscillating manner. In the case all eigenvalues are real, the transition probability $p_{IJ}(n)$ approaches π_J in a nonoscillating manner. Thus, the third conclusion is that the transition probability $p_{IJ}(n)$ can approach the stationary probability in two fundamental manners: either oscillating or nonoscillating.

4.3. Analytical solutions when a macro-state has two substates

In some simple cases, the diagonalization method allows one to derive analytical solutions to the n-period state transition probabilities. The two-state conventional Markov model is the simplest case where the second-largest eigenvalue is $\lambda_2 = 1 - \alpha - \beta$ and $c_{2,IJ} = -\pi_J$, see Eq. (8). In the ESMSM, the matrix $\mathbf{P}$ is a sparse matrix that contains many zero elements; therefore, when two substates represent each macro-state, it is challenging but possible to derive the analytical solutions to the n-period state transition probabilities.

Proposition 2. *The solutions to the n-period state transition probabilities of macro-states A and B, with two substates for each macro-state, are given by*

$$p_{AB}(n) = \pi_B - \frac{1}{4\beta}\psi(n), \quad p_{BA}(n) = \pi_A - \frac{1}{4\alpha}\psi(n), \tag{16}$$

where function $\psi(n)$ is given by

$$\psi(n) = \frac{(\delta + C)^2}{4C}\lambda_3^n - \frac{(\delta - C)^2}{4C}\lambda_4^n - \frac{(\alpha - \beta)^2}{\delta}(1 - 2\delta)^n, \tag{17}$$

π_A and π_B are the stationary probabilities given by Eq. (9), $\delta = \alpha + \beta$, $\lambda_3 = 1 - \delta - C$, $\lambda_4 = 1 - \delta + C$, and $C = \sqrt{\alpha^2 + \beta^2 - 6\alpha\beta}$ assuming that $C \neq 0$.

Appendix A presents the proof.

Therefore, in the ESMSM presented in Fig. 4, the solution for the lag-n autocorrelation yields

$$\rho_n = \frac{(\mu_A - \mu_B)^2}{4\sigma^2(\alpha + \beta)}\psi(n). \tag{18}$$

Function $\psi(n)$ determines the functional form of the lag-n autocorrelation in the ESMSM. This function represents the sum of three exponential functions, where the first two are functions of C. Note that C can be either a nonzero real number, zero, or a complex number depending on the sign and value of $\alpha^2 + \beta^2 - 6\alpha\beta$. In particular,

$$C \text{ is } \begin{cases} \text{a complex number} & \text{if } (3 - \sqrt{8})\beta < \alpha < (3 + \sqrt{8})\beta, \\ 0 & \text{if } \alpha = (3 \pm \sqrt{8})\beta, \\ \text{a real number} & \text{if } \alpha < (3 - \sqrt{8})\beta \text{ or } \alpha > (3 + \sqrt{8})\beta. \end{cases} \tag{19}$$

We can easily deduce that C is a real number when the mean duration of one state is approximately more than six times greater than the mean duration of the other state. We do not observe such a notable difference between the mean durations of bull and bear markets. Consequently, in the context of the stock market cycles, we expect that C is a complex number. In this case, λ_3 and λ_4 is a complex conjugate pair, and the following proposition provides the analytical solution to function $\psi(n)$.

Proposition 3. *If C is a complex number, then function $\psi(n)$ given by Eq. (17) can be rewritten as follows:*

$$\psi(n) = R\lambda^n \cos(n\varphi + \theta) - \frac{(\alpha - \beta)^2}{\delta}(1 - 2\delta)^n, \tag{20}$$

where

$$\lambda = \sqrt{1 - 2\delta + 8\alpha\beta}, \quad \varphi = \arctan\left(\frac{\sqrt{6\alpha\beta - \alpha^2 - \beta^2}}{1 - \delta}\right), \tag{21}$$

$$R = \sqrt{\delta^2 + \frac{(\alpha - \beta)^4}{6\alpha\beta - \alpha^2 - \beta^2}}, \quad \theta = \arctan\left(\frac{(\alpha - \beta)^2}{\delta\sqrt{6\alpha\beta - \alpha^2 - \beta^2}}\right). \tag{22}$$

The proof is given in Appendix A.

Consequently, if C is a complex number, the expression for the n-period state transition probabilities represents the difference between the two components. The first component is an exponentially damped cosine wave with a phase shift, while the second is exponential decay. Therefore, ρ_n approaches zero, oscillating as n increases. To gain further insight into the behavior of the lag-n autocorrelation, let us assume that $\alpha = \beta$. In this case, the expression for ρ_n can be simplified to[5]

$$\rho_n = \frac{(\mu_A - \mu_B)^2}{4\sigma^2}\lambda^n \cos(n\varphi). \tag{23}$$

Under this simplified assumption, an exponentially damped cosine wave without a phase shift represents the shape of the lag-n autocorrelation. In particular, ρ_n periodically changes sign beginning from a positive one.[6] Typically, because the cosine wave decays rather fast, the full oscillating behavior is hard to notice; however, a positive autocorrelation appears over the short run and a subsequent negative autocorrelation over the medium run. The return process exhibits short-term momentum and medium-term mean reversion.

Consider the case where C is a nonzero real number. In that case, λ_3 and λ_4 are also real numbers. Function $\psi(n)$ represents the sum of three real-valued exponential functions. As in the preceding case, the lag-n autocorrelation approaches zero as n increases, though in a nonoscillating manner. Finally, consider the last case where $C \to 0$. The subsequent proposition (the Appendix presents the proof) provides the solution to the n-period state transition probabilities.

[5] If $\alpha = \beta$, then $R = \alpha + \beta$ and $\theta = 0$. Furthermore, the second term on the right-hand side of Eq. (20) disappears. Finally, $\pi_A = \pi_B = 0.5$.

[6] This function crosses zero each time when $n\varphi = k\pi$ radians, where k is a positive integer value.

Proposition 4. *As $C \to 0$, function $\psi(n)$ given by Eq. (17) converges to*

$$\lim_{C\to 0} \psi(n) = \left(\delta - \frac{(\alpha-\beta)^2}{1-\delta} n\right)(1-\delta)^n - \frac{(\alpha-\beta)^2}{\delta}(1-2\delta)^n. \tag{24}$$

Consequently, in this case, the return autocorrelation ρ_n also decreases toward zero in a nonoscillating manner as n increases.

We finish this section by presenting some illustrations in Fig. 5, which shows three examples of the return autocorrelation functions in an ESMSM with two substates for each macro-state. Specifically, the top panel plots the month-k return autocorrelation function, while the bottom panel displays the first-order autocorrelation function of k-month returns. The annualized mean state returns in all plots are $\mu_A = 20\%$ and $\mu_B = -30\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 20\%$, and the one-period transition probability from a bear to a bull state of the market is $\beta = 0.1$. The one-period transition probability from a bull to a bear state of the market takes three alternative values: $\alpha \in \{0.01, (3-\sqrt{8})\beta, 0.05\}$. In the case $\alpha = 0.05$ ($\alpha = 0.01$), the month-k return autocorrelation approaches zero in an oscillating (nonoscillating) manner. The case $\alpha = (3-\sqrt{8})\beta$ is the border between the oscillatory and nonoscillatory behavior.

In all three cases, the month-k return autocorrelation crosses zero at least once. In particular, in all cases, the autocorrelation function changes sign from positive to negative no less than once. Regarding the shape of the first-order autocorrelation of k-month returns, qualitatively, it remains the same in all three cases. Specifically, it quickly increases before gradually decreasing below zero after reaching the maximum. These examples show that the return process in our ESMSM exhibits short-term momentum and subsequent medium-term mean reversion.

4.4. Numerical solutions for the general case

If more than two substates represent one macro-state in an ESMSM, then the n-period transition probabilities can be computed using matrix multiplication routines available in many mathematical software programs. All that is needed is to define the one-period transition probability matrix in an ESMSM. For example, in an ESMSM where three substates represent either macro-state, the one-period transition probability matrix is given by

$$\mathbf{P} = \begin{bmatrix} 1-3\alpha & 3\alpha & 0 & 0 & 0 & 0 \\ 0 & 1-3\alpha & 3\alpha & 0 & 0 & 0 \\ 0 & 0 & 1-3\alpha & 3\alpha & 0 & 0 \\ 0 & 0 & 0 & 1-3\beta & 3\beta & 0 \\ 0 & 0 & 0 & 0 & 1-3\beta & 3\beta \\ 3\beta & 0 & 0 & 0 & 0 & 1-3\beta \end{bmatrix}. \tag{25}$$

Under our convention, the mean state duration times in an ESMSM are the same as in the conventional MSM. Therefore, for example, the transition probability from one substate to another substate of macro-state A equals 3α. This choice ensures that the mean state A duration time equals $1/\alpha$; however, whereas the state duration times in a conventional MSM follow the geometric distribution, the state duration times in an ESMSM specified by the transition probability matrix in (25) are governed by the $NB(3)$ distribution.

Our numerical experiments reveal that, under realistic model parameters, the solution to the return autocorrelation function ρ_n in an ESMSM with more than two substates for each macro-state is qualitatively similar to that where an ESMSM has two substates for each macro-state. For illustration, Fig. 6 shows the return autocorrelations in an ESMSM with three substates for each macro-state. As in the preceding section, this ESMSM assumes monthly returns. In the figure, the red line with points plots the month-k return autocorrelation, whereas the blue line with points plots the first-order autocorrelation of k-month returns. Excluding the number of substates for each macro-state, the other model parameters are the same as those in Fig. 2. Specifically, the annualized mean state returns are $\mu_A = 20\%$ and $\mu_B = -30\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 20\%$. The mean state A (bull market) duration time equals 20 months, and the mean state B (bear market) duration time equals 10 months.

It is instructive to compare the shapes of the autocorrelation functions in the conventional MSM depicted in Fig. 2 and those in the ESMSM presented in Fig. 6. The month-k return autocorrelation exponentially decreases toward zero in the MSM; however, the month-k return autocorrelation exhibits a damped oscillating behavior around zero in the ESMSM. While the first-order autocorrelation of k-month returns is always positive in the MSM, the first-order autocorrelation of k-month returns is initially positive; subsequently, its sign changes to negative in the ESMSM. Again, it is worth noting that the first-order autocorrelation of k-month returns is notably larger in absolute value than the month-k return autocorrelation for $k > 1$.

4.5. Duration dependence and mean reversion

In concluding the main theoretical section, we want to elaborate on the role of duration dependence in the presence of mean-reverting behavior. First, it should be emphasized that a semi-Markov model can reproduce the medium-term mean reversion under positive duration dependence. In other words, the state termination probability must increase with the state age. Most empirical studies document positive duration dependence in bull and bear market states; thus, this condition is satisfied. Why does positive duration dependence induce medium-term mean reversion? Diebold and Rudebusch (1990) present a good general discussion of the effect of positive duration dependence on the properties of economic cycles. The negative binomial distribution can provide a more direct answer to this question.

As motivated by Fig. 3, the larger the value of q in the $NB(q)$ distribution, the lower (higher) the probability of state termination if the state is young (old). Therefore, the larger the q is, the higher concentration of the probability mass around the mean value is. Put differently, the larger the q, the lesser the uncertainty in the state duration time. We can demonstrate this reduction of uncertainty using the following mathematical arguments. Under our construction, the mean macro-state I duration time does not depend on q and equals to $q/(q\,p_{IJ}) = 1/p_{IJ}$, where p_{IJ} is the probability of transiting from macro-state I to macro-state J over one period. However, the variance of the state duration time equals $q(1-q\,p_{IJ})/(q\,p_{IJ})^2 = 1/\left(q\,p_{IJ}^2\right) - 1/p_{IJ}$. Consequently, as q increases, the mean macro-state I duration time remains the same; however, the variance of the macro-state I duration time decreases. As a result, as q increases, the probability distribution of the state duration increasingly concentrates around the mean. As first observed by Diebold and Rudebusch (1990), positive duration dependence induces some regularity of economic cycles. In our case, as q increases, the stock market states interchange with higher regularity that materializes in negative medium-term autocorrelations. The medium-term mean reversion manifests some regularity in the stock market cycles. To stress the random nature of this regularity, Diebold and Rudebusch (1990) introduce the notion of a "stochastic weak form of periodicity" in economic cycles.

Additionally, the higher the degree of positive duration dependence, the more regular the stock market cycles and the more pronounced the mean-reverting behavior. This relationship is depicted in Fig. 7, which plots the month-k return autocorrelation in the ESMSM where each macro-state is represented by $q \in \{1, 2, 3, 4\}$ substates. For all q, the mean state A (bull market) duration time is 20, while the mean state B (bear market) duration time is 10. For each q, the state A and B duration times follow $NB(q)$ distribution. The curves in the figure clearly illustrate that the larger the value of q, the stronger the mean-reverting behavior.

The intuition behind the strengthening of mean reversion due to higher regularity in state changes can be reinforced if we consider what happens with the return autocorrelation function in the ESMSM

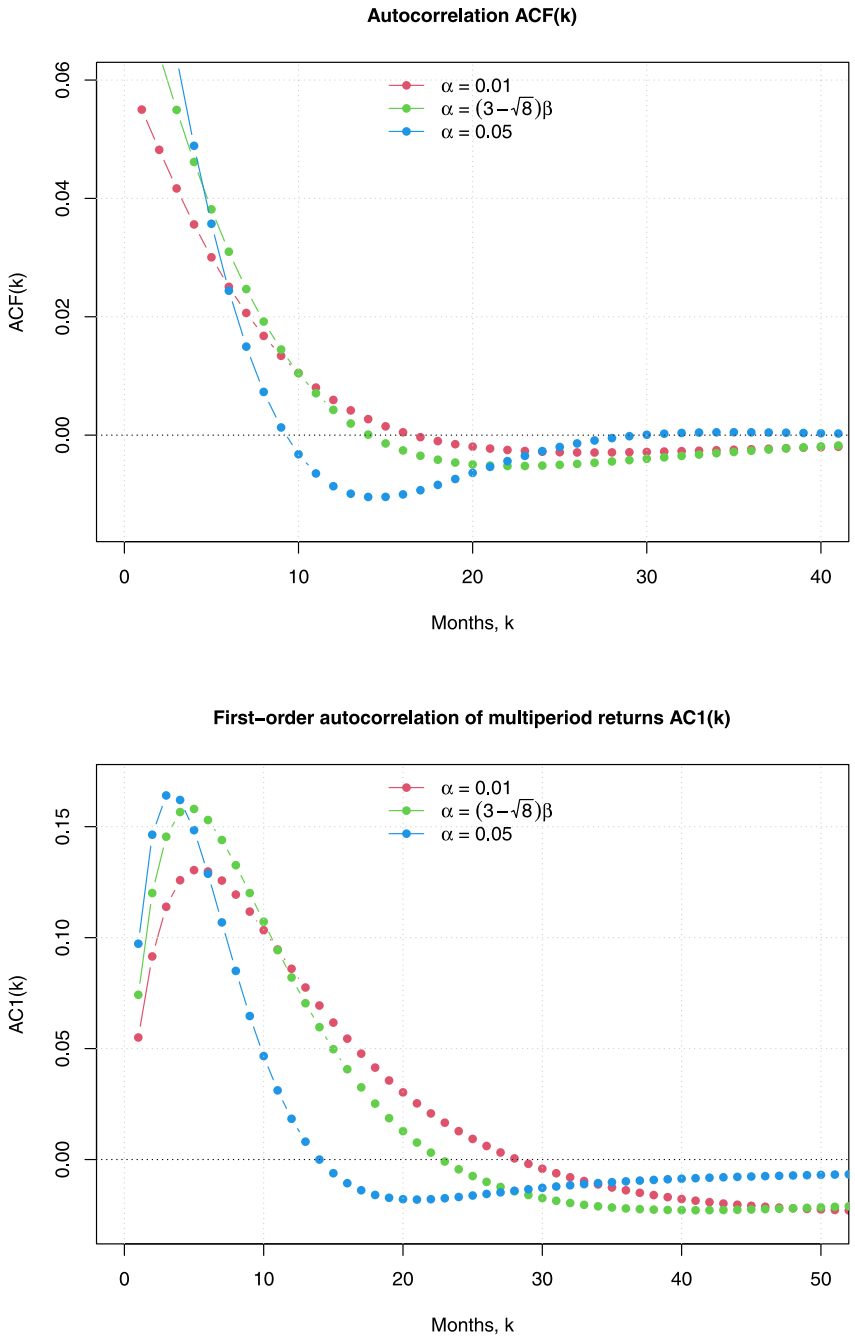

Fig. 5. The return autocorrelations in an ESMSM with two substates for each macro-state. $ACF(k)$ denotes the month-k return autocorrelation ρ_k. $AC1(k)$ denotes the first-order autocorrelation of k-month returns. The annualized mean state returns are $\mu_A = 20\%$ and $\mu_B = -30\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 20\%$. The one-period transition probability $\beta = 0.1$. The one-period transition probability α takes three alternative values $\{0.01, (3-\sqrt{8})\beta, 0.05\}$.

with two substates for each macro-state when the state duration times become certain. Specifically, consider the case where

$$\alpha \to 1/2 \text{ and } \beta \to 1/2. \tag{26}$$

In this limiting case, the variance of the macro-state I duration time is zero, $\alpha = \beta$; therefore, the return autocorrelation function is given by Eq. (23). Furthermore, it is easy to check that under conditions (26), we obtain $\lambda = 1$ and $\varphi = \pi/2$. Therefore, the return autocorrelation function reduces to

$$\rho_n = \frac{(\mu_A - \mu_B)^2}{4\sigma^2} \cos(n\pi/2). \tag{27}$$

The conclusion is that, with deterministic state duration times, the shape of the lag-n autocorrelation is represented by a cosine function without a phase shift and damping (if we extend n to real numbers). Put differently, when the variance of the state duration time approaches zero, the market states start to interchange with perfect regularity.

The general observations can be summarized as follows. In an ESMSM with certain state duration times, the shape of the return autocorrelation function is represented by a cosine wave. When we introduce some uncertainty in the state duration times, the return autocorrelation function displays a damped oscillatory behavior. A further increase in the uncertainty of the state duration times increases the decay (damping) rate. Eventually, the oscillatory behavior disappears if the uncertainty in the state duration times increases beyond some critical level.

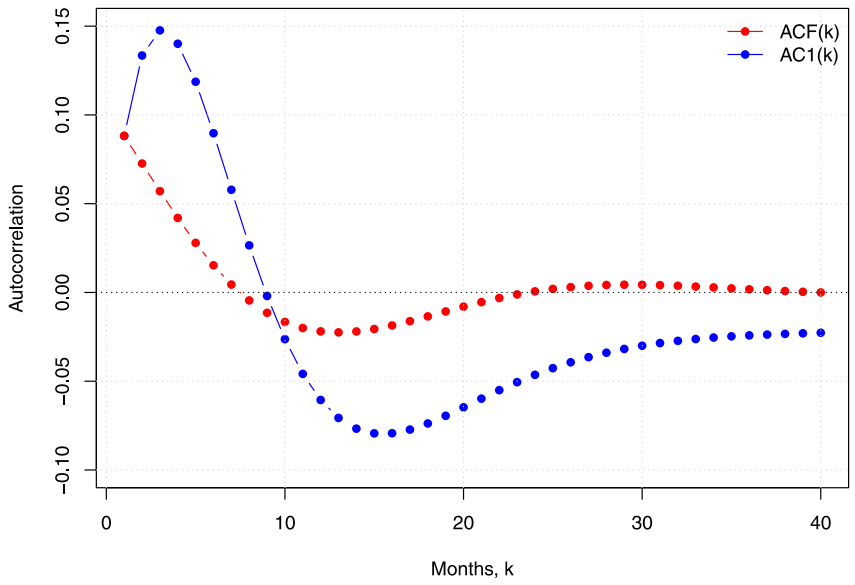

Fig. 6. Return autocorrelations in an expanded-state Markov switching model with three substates for each macro-state. $ACF(k)$ denotes the month-k return autocorrelation ρ_k. $AC1(k)$ denotes the first-order autocorrelation of k-months returns. The annualized mean state returns are $\mu_A = 20\%$ and $\mu_B = -30\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 20\%$. The mean state A duration time equals 20 months, whereas the mean state B duration time equals 10 months.

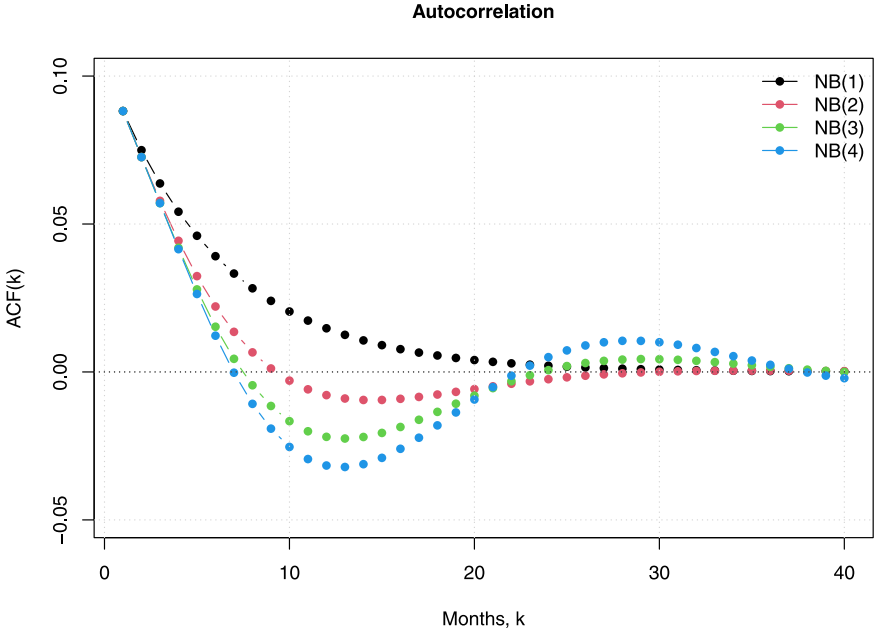

Fig. 7. The return autocorrelation ρ_n in an expanded-state Markov switching models with q substates for each macro-state for various $q \in \{1, 2, 3, 4\}$. The annualized mean state returns are $\mu_A = 20\%$ and $\mu_B = -30\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 20\%$. The mean state A duration time equals 20 months, whereas the mean state B duration time equals 10 months.

The stock market cycles exhibit positive duration dependence, which is also documented in the business cycle expansions and contractions (Diebold and Rudebusch, 1990; Kim and Nelson, 1998; Zuehlke, 2003; Layton and Smith, 2007). Poghosyan (2018) finds positive duration dependence in the credit cycles; therefore, positive duration dependence seems to be a universal phenomenon in all economic cycles.

The relationship between the stock market cycles and business cycles is not clear-cut; however, in a rational expectation model, the stock market movements should reflect the forward-looking behavior of stock market investors who assess the future state of the economy. Therefore, if we assume that the stock market cycles are related to the business cycles and the business cycles exhibit positive duration dependence, then positive duration dependence in the stock market cycles appears naturally. Consequently, the short-term momentum and subsequent mean reversion can be explained without assuming that the investors are not rational.[7]

[7] To date, the momentum and mean reversion in the stock market returns are explained by behavioral theories that challenge the assumption of strict rationality. In particular, these theories presume that investors have several cognitive and emotional biases. The behavioral explanation for short-term momentum and subsequent medium-term mean reversion rests upon the assumptions that investors underreact to news in the short run and overreact in the medium run.

Table 1
The summary statistics of the bull and bear market states.

	S&P Composite		Dow Jones	
Statistic	Bull	Bear	Bull	Bear
Number of states	34	33	38	37
Mean duration	29.03	14.52	27.11	13.00
Stationary probability, %	66.67	33.33	67.59	32.41
Transition probability, %	3.44	6.89	3.69	7.69
Mean return, %	23.00	−27.33	23.92	−28.87
Standard deviation, %	15.51	18.47	15.95	19.08

Notes: Duration is measured in months. Mean returns and standard deviations are annualized and reported in percentages. The transition probability is the probability of transiting to another state over a month.

In summation, our economic explanation for positive duration dependence in the stock market cycles (and, hence, the mean reversion phenomenon) rests upon the view that the stock market cycles are related to the business cycles, and the latter cycles exhibit positive duration dependence. Why do the business cycles show positive duration dependence? There are dozens of alternative business cycle theories, see Niemira and Klein (1994, Chapter 2) for a review. Some of these theories produce a relatively regular repetition of these cycles. This regularity is nothing other than the manifestation of positive duration dependence.

5. Empirical application

5.1. Data and descriptive statistics of bull and bear markets

Our empirical application uses the data on two famous stock market indices: the Standard and Poor's (S&P) Composite index and the Dow Jones Industrial Average index. All data are released monthly and represent capital gain returns. Our sample period begins in January 1897 and ends in December 2020 (124 full years), providing 1488 monthly observations. The data on the S&P Composite index is collected from two sources. In particular, the index returns over the period from January 1897 to December 1925 are provided by William Schwert.[8] The index returns for this period are constructed using a collection of early stock market indices for the US. Schwert (1990) describes the construction methodology in detail. From January 1926 to February 1957, the index returns are the S&P 90 stock market index returns. Beginning in March 1957, the index returns are the S&P 500 stock market index returns. The index returns from January 1926 to December 2020 are from the Center for Research in Security Prices. The data on the Dow Jones index over the total sample period are provided by S&P Dow Jones Indices LLC.[9]

Using the capital gain returns, we reconstruct each stock index value. The bull and bear market turning points are identified using the method proposed by Pagan and Sossounov (2003), which seems to be the most widely accepted method among researchers for such purposes; some notable examples are Gonzalez et al. (2005), Kaminsky and Schmukler (2007), and Claessens et al. (2012). In brief, with minor modifications, this method adopts the dating algorithm developed by Bry and Boschan (1971) to identify the US business cycle turning points using the gross domestic product data. This algorithm is a pattern recognition algorithm based on a set of rules. First, the algorithm finds peaks and troughs in a data series. Second, the algorithm performs several censoring operations to ensure that a complete stock cycle lasts at least 16 months and a market state lasts at least 5 months unless a rise or fall in the stock price exceeds 20%.

For each stock market index, Table 1 presents the summary statistics of the bull and bear markets. Even though some differences exist between the descriptive statistics for each market index, they share many similarities. The mean return is equal to 23% (−28%) in a bull (bear) market, while the standard deviation of returns amounts to 16% (19%) in a bull (bear) market. The difference between the mean returns in the bull and bear markets is substantial. In contrast, the difference between the standard deviation of returns in the bull and bear states is negligible.

The variable of primary interest is the discrepancy between the mean durations of the bull and bear market states. The average bull (bear) market duration is approximately 28 (14) months. Consequently, the average bull market duration is twice as long as the average bear market duration. The values of the stationary state probabilities confirm this observation. In particular, the probability that the US stock market is in the bull (bear) state amounts to 67% (33%). Similarly, the one-period transition probability from a bear to a bull state equals roughly 7%, about twice as large as the transition probability from a bull to a bear state (3.50%). Given these results, we expect the return autocorrelation function to exhibit a damped oscillatory behavior.

Table 2
The results of maximum likelihood estimations.

	Bull markets		Bear markets	
q	p	Log-likelihood	p	Log-likelihood
Panel A: S&P Composite index				
1	0.033	−140.33	0.064	−114.98
2	0.064	−133.38	0.121	−108.82
3	0.094	−131.45	0.171	−107.21
4	0.121	−131.27	0.216	−107.05
5	0.147	−131.89	0.256	−107.50
6	0.171	−132.94	0.292	−108.23
Panel B: Dow Jones index				
1	0.036	−150.26	0.071	−122.93
2	0.070	−142.83	0.132	−117.23
3	0.102	−140.90	0.185	−116.36
4	0.131	−140.88	0.233	−116.95
5	0.159	−141.74	0.275	−118.11
6	0.185	−143.05	0.313	−119.50

Notes: p is the probability of success in one Bernoulli trial in the $NB(q)$ distribution. Log-likelihood is the value of the maximum likelihood estimation of p for various $q \in \{1, \dots, 6\}$ in the $NB(q)$ distribution.

5.2. Fitting statistical distributions to bull and bear duration data

In our semi-Markov model, the state duration times follow the negative binomial $NB(q)$ distribution. Eq. (11) provides the probability mass function of the $NB(q)$ distribution. The probability mass function $f(n, q, p)$ describes the probability that the qth success occurs in the nth Bernoulli trial (the parameter p is the probability of success in a single trial). Recall that the $NB(1)$ distribution is equivalent to the geometric distribution. This section fits the $NB(q)$ distribution to the state duration data to determine which q fits the data best.

We rely on the maximum likelihood estimation (MLE) method to fit the distributions. The standard procedure is to find the pair of parameters (q, p) that maximizes the log-likelihood function. A complication is that q typically extends to real numbers; however, our model assumes that q is an integer. To address this problem, we assume that q is known and find the maximum likelihood estimator for p only. This is done sequentially for various integer values $q \in \{1, \dots, 6\}$, and we select the value of q, which maximizes the log-likelihood.

For each stock market index and market state, Table 2 reports the estimated p and the log-likelihood values of the maximum likelihood estimation of p for various $q \in \{1, \dots, 6\}$. The log-likelihood values advocate that the negative binomial distribution with $q = 4$ maximizes the log-likelihood function for virtually all stock market indices and market states. Therefore, for uniformity, we assume that the state duration times are governed by the $NB(4)$ distribution.

[8] https://www.billschwert.com/gws_data.htm
[9] https://www.spglobal.com/en/

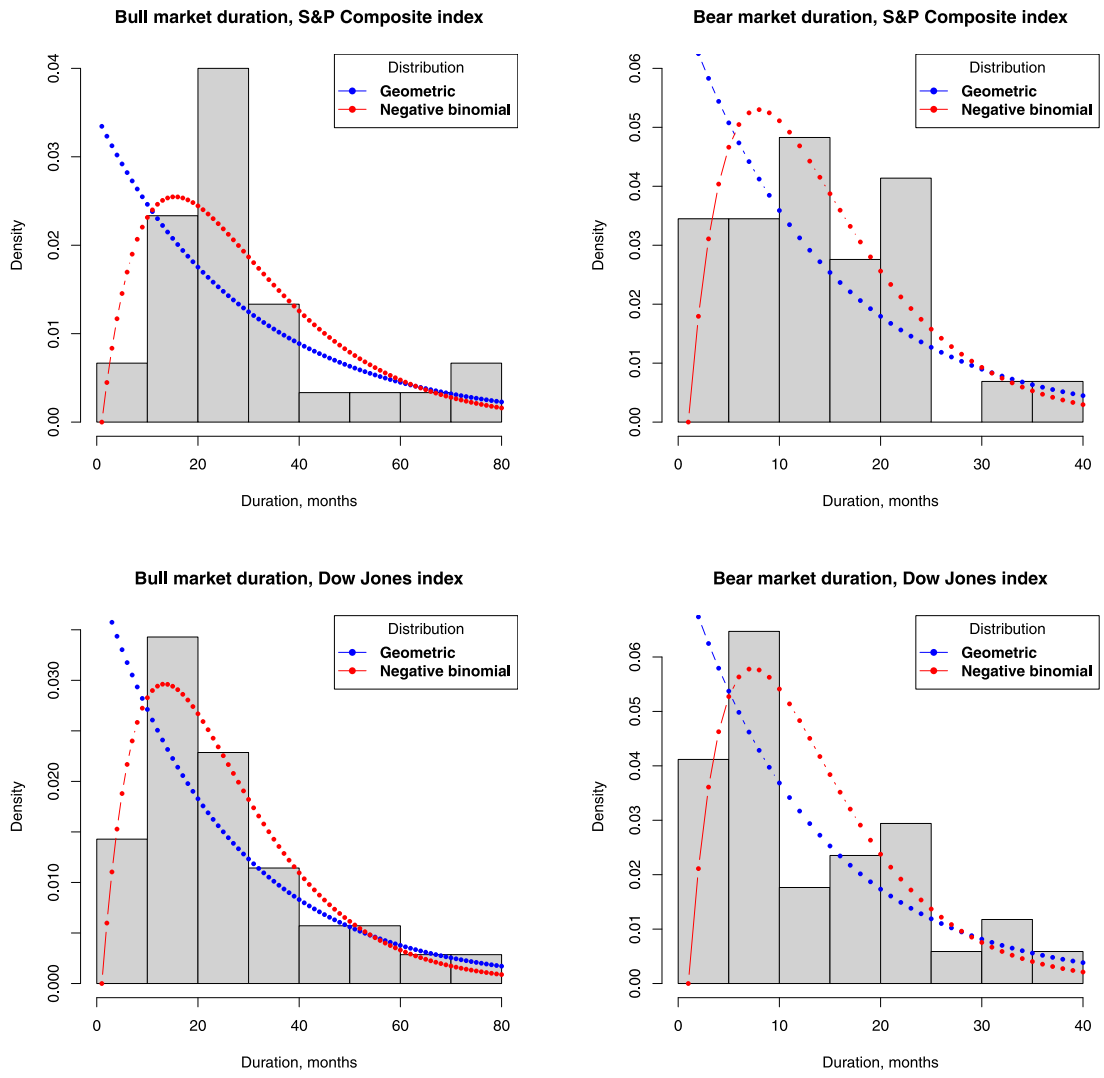

Fig. 8. The histograms of the bull and bear market durations Notes: The lines with blue points plot the fitted geometric distribution, while the lines with red points plot the fitted $NB(4)$ distribution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The top panels in Fig. 8 plot the histograms of the bull and bear market durations for the S&P Composite index. The bottom panels in Fig. 8 plot the histograms of the bull and bear market durations for the Dow Jones index. In each panel, the lines with blue points plot the fitted geometric distribution, while the red lines with points plot the fitted $NB(4)$ distribution. The curves in these panels indicate that the negative binomial distribution fits the state duration data substantially better than the geometric distribution.

5.3. Model calibration and results

This section estimated the empirical lag-n return autocorrelation ρ_n and the first-order autocorrelation of k-period returns $AC1(k)$. Subsequently, using the fitted model parameters, we compute the model-implied ρ_n and $AC1(k)$ in our ESMSM and the conventional MSM. Finally, we compare and contrast the empirical autocorrelations with the model-implied autocorrelations.

All autocorrelations in our study are estimated using a highly robust covariance (and correlation) estimation method suggested by Rousseeuw (1984) and further developed by Rousseeuw (1985). The covariance is estimated using the minimum covariance determinant (MCD) method, which is highly resistant to outliers.[10] The problem is that the exact MCD method is extremely time-consuming. In our study, we rely on the FAST-MCD method developed by Rousseeuw and Driessen (1999).

We estimate the first-order autocorrelation of k-period returns for $k \in \{1, \ldots, 30\}$ months. The fundamental problem with these estimations is that we have only a relatively small number of nonoverlapping intervals of length 30 months; therefore, to increase the number of observations of k-month returns, we follow Fama and French (1988) and employ overlapping intervals of k months.[11]

After estimating ρ_n and $AC1(k)$, we test the following null hypotheses to ensure that the estimated autocorrelations are statistically

[10] In contrast, Fama and French (1988) estimate the first-order autocorrelation of multi-period returns using a standard ordinary least squares regression; however, a few outliers in the return data may highly influence the covariance and correlation estimation by biasing the estimates away from values representative of most of the sample.

[11] Estimates obtained using overlapping blocks of data are biased in short samples (Fama and French, 1988; Kim et al., 1991; Nelson and Kim, 1993). However, our sample is not short because it contains 1488 monthly observations. Our extensive simulation experiments confirm that the bias in estimating the first-order autocorrelation of k-period returns is negligibly small.

significantly different from zero:

$$H_0 : \rho_n = 0 \text{ versus } H_A : \rho_n \neq 0,$$

$$H_0 : AC1(k) = 0 \text{ versus } H_A : AC1(k) \neq 0.$$

Under each null hypothesis, all autocorrelations are zeros. These null hypotheses are equivalent to a presumption that the returns are independent and identically distributed. We employ the randomization method to conduct the test of the null hypotheses. In essence, randomization consists of reshuffling the data and then recalculating the test statistics for each reshuffling to estimate its distribution under the null hypothesis.

Specifically, we randomize the return series 1000 times, each time obtaining a new estimate for ρ_n^* and $AC1(k)^*$.[12] Then, for example, the collection of all estimates for $AC1(k)^*$ constitutes the probability distribution of $AC1(k)$ under the null hypothesis. Using this probability distribution, we compute the 90% confidence interval for $AC1(k)$ under the null hypothesis. In this case, if the estimated value of $AC1(k)$ lies outside the 90% confidence interval, this value is statistically significantly different from zero at the 5% level in a one-tailed test.

The ESMSM and the corresponding MSM are calibrated to empirical data using the following methodology. Our procedure aims to ensure that in the ESMSM and corresponding MSM, the mean state duration times and stationary state probabilities are the same. In the geometric distribution, the mean state duration times are given by Eq. (7). Therefore, in the MSM, the one-period transition probability from state A (bull market) to state B (bear market) equals $\alpha = 1/E[d_A]$, where $E[d_A]$ is the mean state A duration time. Similarly, the one-period transition probability from state B to state A is given by $\beta = 1/E[d_B]$, where $E[d_B]$ is the mean state B duration time. Table 1 reports the estimated state duration times for each market index. The one-period transition probabilities computed in the manner described above differ only marginally from those reported in Table 2. Note that under our theoretical construction, in the ESMSM with four substates for each macro-state, the one-period transition probability from one substate of macro-state A (B) to another substate equals 4α (4β).

The top panels in Fig. 9 plot the results of estimations and calibrations for the S&P Composite index, while the bottom panels show the results of estimations and calibrations for the Dow Jones index. The left panels in Fig. 9 plot the lag-n autocorrelation of log returns, ρ_n, while the right panels plot the first-order autocorrelation of k-period log returns, $AC1(k)$. The black lines with points show the empirically estimated autocorrelations. The shaded areas indicate the 90% confidence interval for the estimated autocorrelations under the null hypothesis of independent and identically distributed (i.i.d.) returns. The blue lines with points depict the autocorrelations implied by the fitted conventional Markov model. The red lines with points show the autocorrelations implied by the fitted semi-Markov model, where four Markovian states represent one semi-Markovian state.

First, our results present convincing evidence of short-term momentum and medium-term mean reversion in the returns on the two stock market indices. This evidence is mainly obtained by comparing the empirically estimated first-order autocorrelation of k-period returns with the boundaries of the 90% confidence interval under the null hypothesis of i.i.d. returns. The evidence is stronger in the returns on the S&P Composite index. In particular, for this index, the estimated values of $AC1(k)$ are statistically significantly above zero over 3 to 9 months and below zero over 14 to 18 months. Using the returns on the Dow Jones index, the values of $AC1(k)$ are statistically significantly positive (negative) over the periods from 4 to 6 months (from 15 to 17 months).

[12] Asterisk is used to indicate that each estimate is calculated on a randomized sample.

Most of the estimated lag-n autocorrelations for both stock market indices lie inside the 90% confidence interval. For both indices, the lag-5 (lag-22) autocorrelation is statistically significantly above (below) zero at the 5% level. Additionally, for the S&P Composite (Dow Jones) index, the lag-11 (lag-18) autocorrelation is statistically significantly above (below) zero. Therefore, the evidence of short-term momentum and medium-term mean reversion is weaker, judging by the estimated lag-n autocorrelation values. Furthermore, because of the limited number of statistically significant values, another problem arises in drawing inferences from the lag-n autocorrelations. Specifically, due to the multiple-testing issue, some of the estimated lag-n autocorrelations can be statistically significant due to luck or chance.

Second but no less crucial, our results present convincing evidence that the semi-Markov model can explain the shape of the empirically estimated $AC1(k)$ function much better than the conventional Markov model. Specifically, the fitted conventional Markov model implies only a short-term momentum, which should be strong and cause statistically significant values of $AC1(k)$ over periods from 1 to 20 months. In contrast, the fitted semi-Markov model predicts a short-term momentum, which should generate statistically significant values of $AC1(k)$ over periods from 1 to 8 months. Subsequently, over periods longer than 11–12 months, the fitted semi-Markov model forecasts negative values of $AC1(k)$ that should not be statistically significant.

Purely qualitatively, the shape of the semi-Markov model-implied $AC1(k)$ and the shape of the empirically estimated $AC1(k)$ look similar. The semi-Markov model correctly captures the duration of the short-term momentum that lasts about 10–12 months and subsequently reverses. Quantitatively though, the model-implied momentum is stronger than the estimated momentum. The difference is especially noticeable over periods from 1 to 5 months. In contrast, the model-implied mean reversion is weaker than the estimated mean reversion. Here the difference is noticeable over periods from 14 to 17 months.

5.4. Explaining the gap between theory and empirical evidence

The results reported in the preceding section demonstrate that the fitted semi-Markov model generates the shape of the first-order autocorrelation function of multi-period returns, $AC1(k)$, that is qualitatively similar to the empirically estimated shape; however, the match between the empirical and model-implied $AC1(k)$ is far from perfect. Specifically, the short-term empirical momentum is weaker than the model-implied momentum, while the empirical medium-term mean reversion is stronger than the model-implied mean reversion. This section discusses some potential explanations for the observed mismatch and the limitations of our semi-Markov model.

Universally, any discrepancy between the model predictions and empirical data stems from model misspecification. One apparent problem with our model is the potential for more than two regimes in the return process in real markets. In particular, although the researchers often assume only two states in the stock market, several studies extend the number of market states. For example, Dias et al. (2015) and Liu and Wang (2017) employ a three-state regime-switching model. Maheu et al. (2012) and Jiang and Fang (2015) operate with a four-state regime-switching model. Finally, De Angelis and Paas (2013) estimate a seven-state model. The presence of more than two regimes in the stock market can potentially account for the mismatch between our semi-Markov model's predictions and the empirical estimates.

For example, the four-state conventional MSM considered in Maheu et al. (2012) assumes the presence of a shorter-term reversal effect in the bull and bear states of the market. Specifically, the model supposes that bull markets contain short periods of bull market corrections, whereas bear markets have short periods of bear market rallies.

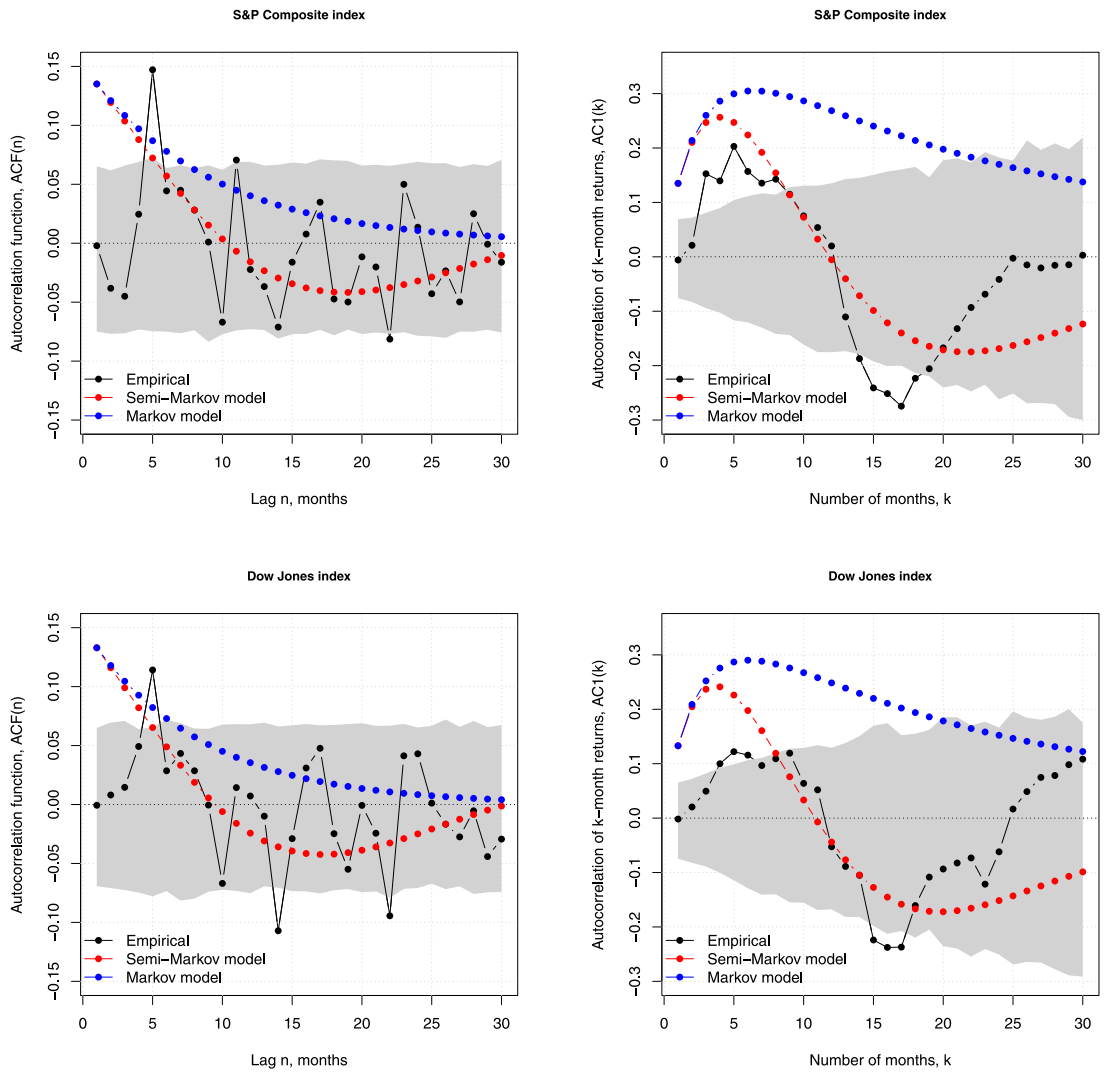

Fig. 9. The results of estimations and calibrations Notes: The top panels in Fig. 9 plot the results of estimations and calibrations for the S&P Composite index, while the bottom panels show the results of estimations and calibrations for the Dow Jones index. The left panels plot the lag-n autocorrelation of log returns ($ACF(n) \equiv \rho_n$). The right panels plot the first-order autocorrelation of k-period log returns ($AC1(k)$). The black lines with points show the empirically estimated autocorrelations. The shaded areas indicate the 90% confidence interval for the estimated autocorrelation under the null hypothesis of i.i.d. returns. The blue lines with points depict the autocorrelations implied by the fitted conventional Markov model. The red lines with points depict the autocorrelations implied by the fitted semi-Markov model, where four Markovian states represent one semi-Markovian state. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This model is motivated by the empirical literature that documents a one-month return reversal in the US and many international stock markets (Rosenberg et al., 1985; Jegadeesh, 1990; Lehmann, 1990). Our simulation experiments verify that this shorter-term return reversal significantly reduces the autocorrelations at first lags.

The existence of a positive feedback effect in the stock market can explain that the empirical medium-term mean reversion is stronger than the model-implied one; a significant price increase tends to be followed by a significant price drop and vice versa (Zarowin, 1989; Renshaw, 1995; Hirschey, 2003). In our semi-Markov model, the mean state return is constant. In real markets, on the other hand, the mean state return can be different. Furthermore, if, for example, the bull state return has been extremely high, then there is a clear tendency for the subsequent bear market return to be extremely low.[13] It is reasonable to suppose that this positive feedback effect makes the medium-term mean reversion more powerful.

[13] This effect corresponds to common investment wisdom, which says that "an excess in one direction leads to an excess in the opposite direction".

6. Conclusions

We present a semi-Markov model where the return process randomly switches between bull and bear states. Our semi-Markov model is realized as an expanded-state Markov model where several Markovian states represent one semi-Markovian state. A negative binomial distribution governs the state duration times in our model; this distribution exhibits positive state duration dependence reported in many empirical studies. We offer the analytical solution to the return autocorrelation function for the simplest case, where two Markovian states represent each semi-Markovian state. In the general case, the return autocorrelation function can be computed using simple numerical methods. We demonstrate that our model's return process induces short-term momentum and medium-term mean reversion. Under realistic model parameters, the shape of the autocorrelation function represents a damped cosine wave that decays relatively quickly.

Positive autocorrelations at shorter lags occur because the return process is more likely to remain in the same state than to switch to another. The intuition behind the appearance of negative autocorrelations at longer lags is as follows. Provided the absence of duration

dependence, the switching between the two states is entirely irregular; in this case, the return process exhibits only short-term momentum. When both states exhibit positive duration dependence, some regularity in the state changes emerges. As a result, the return process shows short-term momentum and medium-term mean reversion.

Our model is easy to fit into empirical data. We calibrate our model to monthly returns on the S&P Composite and Dow Jones indices. We demonstrate that the fit is reasonably good; in particular, our model correctly captures the duration of short-term momentum that lasts about 12 months and subsequently reverses. Our model represents a parsimonious, simple-to-compute, and easy-to-calibrate regime-switching model for stock returns. This model explains short-term momentum and medium-term mean reversion documented by numerous empirical studies.

Finally, our model has some limitations, as it is designed to explain the 12-month momentum and subsequent mean reversion documented in various empirical studies that use monthly data. Therefore, our model cannot explain the stylized facts of the intraweek and intraday returns documented by studies that use daily and intraday returns. Similarly, our model cannot explain the long-term properties of stock market returns, for example, the secular mean reversion documented by Zakamulin (2017).

CRediT authorship contribution statement

Javier Giner: Conception and design of study, Analysis and/or interpretation of data, Writing – review & editing. **Valeriy Zakamulin:** Conception and design of study, Analysis and/or interpretation of data, Writing – review & editing.

Data availability

The data/code are available as Supplementary Material.

Acknowledgment

All authors approved the version of the manuscript to be published.

Appendix A

Proof of Proposition 2

The detailed proof of this proposition is very lengthy. Below, we present only a sketch of the proof. Full details of the proof are available from the authors upon request.

First, we find the eigenvalues λ_i, $i \in \{1,\dots,4\}$, of $\mathbf{P}$ by solving the equation $|\mathbf{P}-\lambda\mathbf{I}| = 0$. This gives us the following four eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 1-2\delta$, $\lambda_3 = 1-\delta-C$, and $\lambda_4 = 1-\delta+C$. The diagonal matrix $\mathbf{D}$ contains the eigenvalues $\{\lambda_1,\dots,\lambda_4\}$ along the main diagonal.

Second, we find the eigenvectors corresponding to each eigenvalue by solving the equation $(\mathbf{P}-\lambda_i\mathbf{I}) = \mathbf{v}_i$ for each λ_i. These eigenvectors constitute the columns in matrix $\mathbf{Q}$. After some extremely tedious but straightforward computations, we obtain the following matrices $\mathbf{Q}$ and $\mathbf{Q}^{-1}$:

$$\mathbf{Q} = \begin{bmatrix} 1 & -\frac{\alpha}{\beta} & \frac{\beta-\alpha-C}{2\beta} & \frac{\beta-\alpha+C}{2\beta} \\ 1 & 1 & -1 & -1 \\ 1 & -\frac{\beta}{\alpha} & \frac{\beta-\alpha+C}{2\alpha} & \frac{\beta-\alpha-C}{2\alpha} \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{Q}^{-1} = \begin{bmatrix} \frac{\beta}{2\delta} & \frac{\beta}{2\delta} & \frac{\alpha}{2\delta} & \frac{\alpha}{2\delta} \\ -\frac{\beta}{2\delta} & \frac{\alpha}{2\delta} & -\frac{\alpha}{2\delta} & \frac{\beta}{2\delta} \\ -\frac{\beta}{2C} & \frac{d}{4C} & \frac{\alpha}{2C} & \frac{c}{4C} \\ \frac{\beta}{2C} & -\frac{c}{4C} & -\frac{\alpha}{2C} & -\frac{d}{4C} \end{bmatrix},$$

where $c = \beta-\alpha+C$ and $d = \beta-\alpha-C$ are two constants introduced to shorten the expressions.

Finally, we derive the expressions for the elements of matrix $\mathbf{P}(n) = \mathbf{P}^n = \mathbf{Q}\mathbf{D}^n\mathbf{Q}^{-1}$. Using four 2×2 submatrices, matrix $\mathbf{P}(n)$ can be written as

$$\mathbf{P}(n) = \begin{bmatrix} \mathbf{P}_{AA}(n) & \mathbf{P}_{AB}(n) \\ \mathbf{P}_{BA}(n) & \mathbf{P}_{BB}(n) \end{bmatrix}.$$

To compute the transition probabilities, we must know the elements of submatrices $\mathbf{P}_{AB}(n)$ and $\mathbf{P}_{BA}(n)$. The submatrix $\mathbf{P}_{AB}(n)$ is

$$\mathbf{P}_{AB}(n) = \begin{bmatrix} p_{13}(n) & p_{14}(n) \\ p_{23}(n) & p_{24}(n) \end{bmatrix} = \begin{bmatrix} \pi_3 + \pi_3\frac{\alpha}{\beta}\lambda_2^n + \frac{\alpha d}{4\beta C}\lambda_3^n - \frac{\alpha c}{4\beta C}\lambda_4^n & \pi_3 - \pi_1\frac{\alpha}{\beta}\lambda_2^n + \frac{cd}{8\beta C}\lambda_3^n - \frac{cd}{8\beta C}\lambda_4^n \\ \pi_3 - \pi_3\lambda_2^n - \frac{\alpha}{2C}\lambda_3^n + \frac{\alpha}{2C}\lambda_4^n & \pi_3 + \pi_1\lambda_2^n - \frac{c}{4C}\lambda_3^n + \frac{d}{4C}\lambda_4^n \end{bmatrix},$$

where $\pi_1 = \frac{\beta}{2\delta} = \frac{\pi_A}{2}$ and $\pi_3 = \frac{\alpha}{2\delta} = \frac{\pi_B}{2}$ are the stationary probabilities of substates 1 and 3. The probability $p_{AB}(n)$ is computed as one-half of the sum of these four probabilities (see Eq. (13)):

$$p_{AB}(n) = \pi_B + \frac{(\alpha-\beta)^2}{4\beta\delta}\lambda_2^n - \frac{(\delta+C)^2}{16\beta C}\lambda_3^n + \frac{(\delta-C)^2}{16\beta C}\lambda_4^n.$$

The submatrix $\mathbf{P}_{BA}(n)$ is

$$\mathbf{P}_{BA}(n) = \begin{bmatrix} p_{31}(n) & p_{32}(n) \\ p_{41}(n) & p_{42}(n) \end{bmatrix} = \begin{bmatrix} \pi_1 + \pi_1\frac{\beta}{\alpha}\lambda_2^n - \frac{\beta c}{4\alpha C}\lambda_3^n + \frac{\beta d}{4\alpha C}\lambda_4^n & \pi_1 - \pi_3\frac{\beta}{\alpha}\lambda_2^n + \frac{cd}{8\alpha C}\lambda_3^n - \frac{cd}{8\alpha C}\lambda_4^n \\ \pi_1 - \pi_1\lambda_2^n - \frac{\beta}{2C}\lambda_3^n + \frac{\beta}{2C}\lambda_4^n & \pi_1 + \pi_3\lambda_2^n + \frac{d}{4C}\lambda_3^n - \frac{c}{4C}\lambda_4^n \end{bmatrix}.$$

Probability $p_{BA}(n)$ is computed as one-half of the sum of these four probabilities:

$$p_{BA}(n) = \pi_A + \frac{(\alpha-\beta)^2}{4\alpha\delta}\lambda_2^n - \frac{(\delta+C)^2}{16\alpha C}\lambda_3^n + \frac{(\delta-C)^2}{16\alpha C}\lambda_4^n.$$

A useful property

Property 1. *Suppose that $C_3 = u - iv$ and $C_4 = u + iv$ are a complex conjugate pair. Additionally, suppose that $\lambda_3 = x - iy$ and $\lambda_4 = x + iy$ are another complex conjugate pair. Then the following result holds:*

$$C_3\lambda_3^n + C_4\lambda_4^n = 2\lambda^n R\cos(n\varphi+\theta), \tag{28}$$

where $\lambda = \sqrt{x^2+y^2}$, $\varphi = \arctan(y/x)$, $R = \sqrt{u^2+v^2}$, and $\theta = \arctan(v/u)$.

Proof. Using De Moivre's formula, we obtain

$$\lambda_3^n = (x-iy)^n = \lambda^n e^{-in\varphi}, \quad \lambda_4^n = (x+iy)^n = \lambda^n e^{in\varphi}.$$

Euler's formula implies that

$$2\cos(n\varphi) = e^{in\varphi} + e^{-in\varphi}, \quad 2i\sin(n\varphi) = e^{in\varphi} - e^{-in\varphi}.$$

Therefore,

$$\begin{aligned} C_3\lambda_3^n + C_4\lambda_4^n &= (u-iv)(x-iy)^n + (u+iv)(x+iy)^n \\ &= \lambda^n\left((u-iv)e^{-in\varphi} + (u+iv)e^{in\varphi}\right) \\ &= 2\lambda^n\left(u\cos(n\varphi) - v\sin(n\varphi)\right). \end{aligned} \tag{29}$$

Finally, a linear combination of cosine and sine waves is equivalent to a single cosine wave with a phase shift and rescaled amplitude:

$$u\cos(n\varphi) - v\sin(n\varphi) = R\left(\frac{u}{R}\cos(n\varphi) - \frac{v}{R}\sin(n\varphi)\right) = R\cos(n\varphi+\theta).$$

Proof of Proposition 3

If C is a complex number, then $C = i|C|$ where $|C| = \sqrt{|\alpha^2+\beta^2-6\alpha\beta|}$. In this case, the eigenvalues λ_3 and λ_4 can be written in the following form:

$$\lambda_3 = 1-\delta-i|C|, \quad \lambda_4 = 1-\delta+i|C|. \tag{30}$$

Apparently, λ_3 and λ_4 are complex conjugate numbers. Consider now the coefficients in front of λ_3 and λ_4 in Eq. (17).

$$C_3 = \frac{(\delta + C)^2}{4C} = \frac{\delta}{2} - i\frac{(\alpha-\beta)^2}{2|C|}, \quad C_4 = -\frac{(\delta - C)^2}{4C} = \frac{\delta}{2} + i\frac{(\alpha-\beta)^2}{2|C|}. \quad (31)$$

Consequently, C_3 and C_4 are also complex conjugate numbers. The final result follows from Property 1.

Proof of Proposition 4

When $C \to 0$, the expression for $\psi(n)$ gives rise to an indeterminate $0/0$ form. We evaluate this indeterminate form using l'Hôpital's rule. We consider the situation when C is a complex number approaching zero.

The expressions for λ_3 and λ_4 are given by Eqs. (30), whereas the coefficients in front of λ_3 and λ_4 in Eq. (17) are given by Eqs. (31). Using Eq. (29), the expression for $\psi(n)$ is given by

$$\psi(n) = \lambda^n \left(\delta \cos(n\varphi(C)) - \frac{(\alpha-\beta)^2}{|C|} \sin(n\varphi(C)) \right) - \frac{(\alpha-\beta)^2}{\delta}(1-2\delta)^n, \quad (32)$$

where

$$\lambda = \sqrt{(1-\delta)^2 + |C|^2}, \quad \varphi(C) = \arctan\left(\frac{|C|}{1-\delta}\right),$$

and notation $\varphi(C)$ emphasizes that φ is a function of C.

As $C \to 0$, $\lambda \to 1-\delta$, $\varphi(C) \to 0$, and, hence, $\cos(n\varphi(C)) \to 1$ while $\sin(n\varphi(C)) \to 0$. One term in (32) has an indeterminate $0/0$ form, in particular, $\sin(n\varphi(C))/|C|$. The application of l'Hôpital's rule gives us (to shorten the notation, we replace $|C|$ by c)

$$\lim_{c\to 0} \frac{\sin(n\varphi(c))}{c} \overset{\text{l'H}}{=} \lim_{c\to 0} \frac{n\cos(n\varphi(c))\varphi'(c)}{c'} = \lim_{c\to 0} n\cos(n\varphi(c)) \frac{1-\delta}{(1-\delta)^2 + c^2} = \frac{n}{1-\delta}.$$

Consequently,

$$\lim_{C\to 0} \psi(n) = \left(\delta - \frac{(\alpha-\beta)^2}{1-\delta} n\right)(1-\delta)^n - \frac{(\alpha-\beta)^2}{\delta}(1-2\delta)^n.$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.econmod.2023.106237.

References

Balvers, R.J., Hu, O., Huang, D., 2012. Transitory market states and the joint occurrence of momentum and mean reversion. J. Financial Res. 35 (4), 471–495.

Balvers, R.J., Wu, Y., 2006. Momentum and mean reversion across national equity markets. J. Empir. Financ. 13 (1), 24–48.

Balvers, R., Wu, Y., Gilliland, E., 2000. Mean reversion across national stock markets and parametric contrarian investment strategies. J. Finance 55 (2), 745–772.

Barberis, N., Shleifer, A., 2003. Style investing. J. Financ. Econ. 68 (2), 161–199.

Barbu, V.S., Limnios, N., 2008. Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications: their Use in Reliability and DNA Analysis. Springer-Verlag, New York.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2016. Time Series Analysis: Forecasting and Control, fifth ed. John Wiley & Sons Inc., New Jersey.

Bry, G., Boschan, C., 1971. Cyclical Analysis of Time Series: Selected Procedures and Computer Programs. NBER.

Burshtein, D., 1996. Robust parametric modeling of durations in hidden Markov models. IEEE Trans. Speech Audio Process. 4 (3), 240–242.

Claessens, S., Kose, M.A., Terrones, M.E., 2012. How do business and financial cycles interact? J. Int. Econ. 87 (1), 178–190.

Cochran, S.J., Defina, R.H., 1995. Duration dependence in the US stock market cycle: A parametric approach. Appl. Financial Econ. 5 (5), 309–318.

D'Amico, G., Petroni, F., 2012. A semi-Markov model for price returns. Phys. A 391 (20), 4867–4876.

De Angelis, L., Paas, L.J., 2013. A dynamic analysis of stock markets using a hidden Markov model. J. Appl. Stat. 40 (8), 1682–1700.

De Bondt, W.F.M., Thaler, R., 1985. Does the stock market overreact? J. Finance 40 (3), 793–805.

Dias, J.G., Vermunt, J.K., Ramos, S., 2015. Clustering financial time series: New insights from an extended hidden Markov model. European J. Oper. Res. 243 (3), 852–864.

Diebold, F.X., Rudebusch, G.D., 1990. A nonparametric investigation of duration dependence in the American business cycle. J. Polit. Econ. 98 (3), 596–616.

Fama, E.F., French, K.R., 1988. Permanent and temporary components of stock prices. J. Polit. Econ. 96 (2), 246–273.

Ferguson, J.D., 1980. Variable duration models for speech. In: Ferguson, J.D. (Ed.), Proceedings of the Symposium on the Application of Hidden Markov Models to Text and Speech. Princeton, New Jersey, pp. 143–179.

Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer, New York.

Georgopoulou, A., Wang, J.G., 2016. The trend is your friend: Time-series momentum strategies across equity and commodity markets. Rev. Finance 21 (4), 1557–1592.

Gonzalez, L., Powell, J.G., Shi, J., Wilson, A., 2005. Two centuries of bull and bear market cycles. Int. Rev. Econ. Finance 14 (4), 469–486.

Guèdon, Y., 2005. Hidden hybrid Markov/semi-Markov chains. Comput. Statist. Data Anal. 49 (3), 663–688.

Hamilton, J.D., 1994. Time Series Analysis. Princeton, New Jersey.

Harman, Y.S., Zuehlke, T.W., 2007. Nonlinear duration dependence in stock market cycles. Rev. Financ. Econ. 16 (4), 350–362.

He, X.-Z., Li, K., 2015. Profitability of time series momentum. J. Bank. Financ. 53, 140–157.

Hirschey, M., 2003. Extreme return reversal in the stock market. J. Portf. Manag. 29 (3), 78–90.

Hong, H., Stein, J.C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. J. Finance 54 (6), 2143–2184.

Howard, R.A., 1971. Dynamic Probabilistic Systems, Volume II: Semi-Markov and Decision Processes. John Wiley & Sons, Inc., New York.

Hurst, B., Ooi, Y.H., Pedersen, L.H., 2017. A century of evidence on trend-following investing. J. Portf. Manag. 44 (1), 15–29.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. J. Finance 45 (3), 881–898.

Jegadeesh, N., 1991. Seasonality in stock price mean reversion: Evidence from the U.S. and the U.K.. J. Finance 46 (4), 1427–1444.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. J. Finance 48 (1), 65–91.

Jiang, Y., Fang, X., 2015. Bull, bear or any other states in US stock market? Econ. Model. 44, 54–58.

Johnson, M.T., 2005. Capacity and complexity of HMM duration modeling techniques. IEEE Signal Process. Lett. 12 (5), 407–410.

Kaminsky, G.L., Schmukler, S., 2007. Short-run pain, long-run gain: Financial liberalization and stock market cycles. Rev. Finance 12, 253–292.

Khil, J., Lee, B.-S., 2002. A time-series model of stock returns with a positive short-term correlation and a negative long-term correlation. Rev. Quant. Financ. Account. 18 (4), 381–404.

Kim, C.-J., Nelson, C.R., 1998. Business cycle turning points, A new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. Rev. Econ. Stat. 80 (2), 188–201.

Kim, M.J., Nelson, C.R., Startz, R., 1991. Mean reversion in stock prices? A reappraisal of the empirical evidence. Rev. Econom. Stud. 58 (3), 515–528.

Langrock, R., Zucchini, W., 2011. Hidden Markov models with arbitrary state dwell-time distributions. Comput. Statist. Data Anal. 55 (1), 715–724.

Layton, A.P., Smith, D.R., 2007. Business cycle dynamics with duration dependence and leading indicators. J. Macroecon. 29 (4), 855–875.

Lehmann, B.N., 1990. Fads, martingales, and market efficiency. Q. J. Econ. 105 (1), 1–28.

Levinson, S., 1986. Continuously variable duration hidden Markov models for speech analysis. In: ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 11, pp. 1241–1244.

Lim, B.Y., Wang, J.G., Yao, Y., 2018. Time-series momentum in nearly 100 years of stock returns. J. Bank. Financ. 97, 283–296.

Liu, Z., Wang, S., 2017. Decoding Chinese stock market returns: Three-state hidden semi-Markov model. Pac.-Basin Finance J. 44, 127–149.

Lo, A.W., MacKinlay, A.G., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. Rev. Financ. Stud. 1 (1), 41–66.

Lunde, A., Timmermann, A., 2004. Duration dependence in stock prices: An analysis of bull and bear markets. J. Bus. Econom. Statist. 22 (3), 253–273.

Maheu, J.M., McCurdy, T.H., 2000. Identifying bull and bear markets in stock returns. J. Bus. Econom. Statist. 18 (1), 100–112.

Maheu, J.M., McCurdy, T.H., Song, Y., 2012. Components of bull and bear markets: Bull corrections and bear rallies. J. Bus. Econom. Statist. 30 (3), 391–403.

Moskowitz, T.J., Ooi, Y.H., Pedersen, L.H., 2012. Time series momentum. J. Financ. Econ. 104 (2), 228–250.

Nelson, C.R., Kim, M.J., 1993. Predictable stock returns: The role of small sample bias. J. Finance 48 (2), 641–661.

Niemira, M.P., Klein, P.A., 1994. Forecasting Financial and Economic Cycles. John Wiley & Sons Inc., New York.

Ohn, J., Taylor, L.W., Pagan, A., 2004. Testing for duration dependence in economic cycles. Econom. J. 7 (2), 528–549.

Pagan, A.R., Sossounov, K.A., 2003. A simple framework for analysing bull and bear markets. J. Appl. Econometrics 18 (1), 23–46.
Poghosyan, T., 2018. How do financial cycles affect public debt cycles? Empir. Econ. 54, 425–460.
Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices: Evidence and implications. J. Financ. Econ. 22, 27–59.
Renshaw, E., 1995. Is the stock market more stable than it used to be? Financ. Anal. J. 51 (6), 81–88.
Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. J. Portf. Manag. 11 (3), 9–16.
Rousseeuw, P.J., 1984. Least median of squares regression. J. Amer. Statist. Assoc. 79 (388), 871–880.
Rousseeuw, P.J., 1985. Multivariate estimation with high breakdown point. In: Grossmann, W., Pflug, G., Vincze, I., Wertz, W. (Eds.), Mathematical Statistics and Applications. Vol. B, Reidel Publishing Company, Dordrecht, Nederland, pp. 283–297.
Rousseeuw, P.J., Driessen, K.V., 1999. A fast algorithm for the minimum covariance determinant estimator. Technometrics 41 (3), 212–223.
Russell, M., Cook, A., 1987. Experimental evaluation of duration modelling techniques for automatic speech recognition. In: ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 12, pp. 2376–2379.
Schwert, G.W., 1990. Indexes of U.S. stock prices from 1802 to 1987. J. Bus. 63 (3), 399–426.
Stein, H.J., Pozharny, J., 2022. Modeling momentum and reversals. Risks 10 (10).
Summers, L.H., 1986. Does the stock market rationally reflect fundamental values? J. Finance 41 (3), 591–601.
Tejedor, A., Gómez, J., Pacheco, A., 2015. The negative binomial distribution as a renewal model for the recurrence of large earthquakes. Pure Appl. Geophys. 172, 23–31.
Timmermann, A., 2000. Moments of Markov switching models. J. Econometrics 96 (1), 75–111.
Zakamulin, V., 2017. Secular mean reversion and long-run predictability of the stock market. Bull. Econ. Res. 69 (4), E66–E93.
Zakamulin, V., 2023. Revisiting the duration dependence in the US stock market cycles. Appl. Econ. 55 (4), 357–368.
Zakamulin, V., Giner, J., 2022. Time series momentum in the US stock market: Empirical evidence and theoretical analysis. Int. Rev. Financ. Anal. 82, 102173.
Zarowin, P., 1989. Does the stock market overreact to corporate earnings information? J. Finance 44 (5), 1385–1399.
Zhu, H., Wang, J., Yang, Z., Song, Y., 2006. A method to design standard HMMs with desired length distribution for biological sequence analysis. In: Bücher, P., Moret, B.M.E. (Eds.), Algorithms in Bioinformatics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 24–31.
Zuehlke, T.W., 2003. Business cycle duration dependence reconsidered. J. Bus. Econom. Statist. 21 (4), 564–569.