Optimal Trend Following Rules in Two-State Regime-Switching Models

Valeriy Zakamulin* and Javier Giner†

This revision: November 4, 2023

Abstract

Academic research on trend-following investing has almost exclusively focused on testing various trading rules' profitability. However, all existing trend-following rules are essentially ad-hoc, lacking a solid theoretical justification for their optimality. This paper aims to address this gap in the literature. Specifically, we examine the optimal trend-following when the returns follow a two-state process, randomly switching between bull and bear markets. We show that if a Markov model governs the return process, it is optimal to follow the trend using the Exponential Moving Average rule. However, the Markov model is unrealistic because it does not represent the bull and bear market duration times correctly. It is more sensible to model the return process by a semi-Markov model where the state termination probability increases with age. Under this framework, the optimal trend-following rule resembles the Moving Average Convergence/Divergence rule. We confirm the validity of the semi-Markov model with an empirical study demonstrating that the theoretically optimal trading rule outperforms the popular 10-month Simple Moving Average and 12-month Momentum rules across a universe of international markets.

Key words: Markov model, semi-Markov model, bull-bear markets, optimal trend-following, moving averages

JEL classification: G11, G17

*School of Business and Law, University of Agder, Service Box 422, 4604 Kristiansand, Norway, E-mail: valeriz@uia.no

†Department of Economics, Accounting and Finance, University of La Laguna, Camino La Hornera s/n, 38071, Santa Cruz de Tenerife, Spain, E-mail: jginer@ull.edu.es

1 Introduction

Trend-following investing is popular for several reasons. One of the main reasons is that it is a relatively simple and straightforward investment strategy that can be easily understood and implemented by investors. Another reason is that trend-following has historically been shown to be profitable over the long term. By buying assets that are trending upward and selling assets that are trending downward, trend-following investors can potentially capture significant gains while minimizing losses. In addition, trend-following investing is appealing to some investors because it can help to diversify their portfolios and reduce risk. Finally, trend-following investing can be attractive to investors who are seeking a systematic approach to investing. By using objective rules and indicators to identify trends, investors can avoid emotional decision-making and potentially benefit from more consistent returns over time.

Many alternative trend-following rules are designed to identify the upward and downward trends of the financial markets. Most of these rules are based on moving averages of past prices. The most popular is the Simple Moving Average (SMA). Less commonly used types of moving averages are the Linear Moving Average (LMA) and Exponential Moving Average (EMA). Each moving average is computed using an averaging window of a particular size. Besides, a trend-following rule can be based on either a single moving average or a combination of moving averages. For example, a moving average "crossover" is a rule constructed using two moving averages: one with short window size and another with long window size. There are SMA, LMA, and EMA crossovers. A popular Moving Average Convergence/Divergence (MACD) rule is based on three EMAs. Last but not least, there is a Momentum (MOM) rule that can also be considered as a specific moving average rule.

In view of the aforesaid, investment professionals are overwhelmed by the variety of choices between different trend-following rules, types of moving averages, and averaging window sizes. From the academic point of view, all these trend-following rules are ad-hoc rules whose optimality has never been justified theoretically. The goal of this paper is to fill this gap in the literature. To achieve our goal, we need a model that can replicate the stylized facts about financial asset returns documented by the econometric literature. These facts include fat tails, negative skewness, volatility clustering, short-term momentum, and medium-term mean reversion (see, for example, Cont (2001)). A two-state regime switching model is a widely accepted model for stock returns that can reproduce these stylized facts (see Timmermann (2000), Frühwirth-Schnatter (2006), and Giner and Zakamulin (2023)). In this model, the returns follow a process that randomly switches between bull and bear states. The market states are identifiable and persistent. Consequently, this model justifies trend-following investing.

The main research question in our study is: What are the optimal trend-following rules when the returns follow a two-state regime-switching process? To the best knowledge of the authors, there is only a series of papers that answer this question using a continuous-time Markov Switching Model (MSM) and assuming the existence of trading costs (see Dai, Zhang, and Zhu (2010), Nguyen, Tie, and Zhang (2014a), Dai, Yang, Zhang, and Zhu (2016), and Tie and Zhang (2016)). A typical goal in these papers is to maximize the expected return of the trading strategy. The optimal trading strategy is represented by two time-dependent boundaries on the stock price. When the stock price is above (below) the upper (lower) boundary, a Buy (Sell) signal is generated. Unfortunately, the continuous-time optimization problem with transaction costs is not tractable analytically and finding the no-trading boundaries turns out to be an extremely difficult numerical task (Nguyen, Yin, and Zhang (2014b)).

Besides, a severe limitation of an MSM is that the state termination probability does not depend on the time already spent in that state. In other words, there is no duration dependence. By contrast, many empirical studies document that the stock market states exhibit positive duration dependence (see, among others, Cochran and Defina (1995), Ohn, Taylor, and Pagan (2004), Harman and Zuehlke (2007), and Zakamulin (2023)). A positive duration dependence means that the longer a bull (bear) market lasts, the higher its probability of ending. Because an MSM does not provide a correct representation of the bull and bear market duration times, the return process modeled by an MSM exhibits only short-term momentum. For the return process to exhibit both short-term momentum and subsequent medium-term mean reversion, the return process must be modeled by a Semi-Markov Switching Model (SMSM), where the stock market states exhibit positive duration dependence (see Giner and Zakamulin (2023)).

In contrast to the previous papers, we use a discrete-time model without transaction costs. We consider both a conventional MSM and an SMSM. The contributions of this paper are as follows. Our first contribution is to demonstrate that the problem of finding the optimal trend-following rule in an MSM is analytically tractable. We show that the EMA rule represents the optimal trend-following rule in the MSM and find the solution to the optimal window size (decay constant) in this rule.

Our second contribution is to examine the optimal trend-following rule in an SMSM where the state duration times exhibit positive duration dependence. This is a non-trivial task because an SMSM lacks analytical tractability. Besides, all numerical computations rely on using complicated recursive algorithms. Therefore, theoretical applications of an SMSM are extremely rare in finance.[1] Our analysis relies on an SMSM realized as an Expanded-State MSM (ESMSM) with a specific topology. This ESMSM allows some analytical tractability, and the numerical computations are of the same complexity as that of a conventional MSM. Our results show that the optimal trend-following rule in an SMSM is somewhat similar to the MACD rule.

Our third contribution lies in demonstrating that our theoretical ESMSM is in good agreement with the empirical data, and our theoretically optimal trend-following rule outperforms the popular trend-following rules used by investment professionals and academics. In our empirical study, we use data from a wide range of international stock markets. Through out-of-sample simulations, we provide clear evidence of the optimal rule's superiority over the popular 10-month SMA and the 12-month MOM rules. This outcome not only confirms the validity of our theoretical model but also advocates that our model is highly relevant in investment practice. The main explanation of the optimal trend-following rule's superior performance is as follows: Whereas the 10-month SMA and 12-month MOM rules exploit only the short-term return momentum, the optimal rule uses both the short-term momentum and subsequent mean reversion. Last but not least, our theoretical model offers an explanation for why the 10-month SMA rule typically performs better than the 12-month Momentum rule.

The rest of the paper is organized as follows. Subsequent Section 2 briefly reviews the main existing trend-following rules. Section 3 motivates the modeling of returns by a regime-switching process and advocates that the optimal trend-following rule is determined by the autoregressive (AR) coefficients of the return process. Section 4 derives the analytical solution to the AR coefficients in the MSM and demonstrates that following the trend using the EMA rule is optimal. Section 5 presents the numerical solutions to the AR coefficients in the ESMSM. Section 6 confirms that the return weights in the optimal trend-following rule are sufficiently

[1] In contrast, empirical applications of Hidden Semi-Markov Models (HSMM) is a growing field in empirical finance due to the availability of various estimation algorithms.

close to the AR coefficients of the return process. Section 7 demonstrates the validity of our ESMSM and the advantage of the theoretically optimal trend-following rule using real-world data. Finally, Section 8 concludes the paper.

2 Trend Following Rules: A Brief Review

2.1 Trend Following Rules Based on Past Prices

In most trend-following rules, the trading signal is generated depending on the value of either one or multiple moving averages of past prices. Formally, denote by $\{P_0, P_1, \dots, P_t\}$ a series of observations of the closing prices of a financial asset, where time t denotes the current time when the last closing price P_t is observed. A moving average of past n prices is computed as

$$MA_t(n, P) = \frac{\sum_{i=0}^{n-1} w_i P_{t-i}}{\sum_{i=0}^{n-1} w_i}, \tag{1}$$

where w_i is the weight of price P_{t-i} in the computation of a moving average. There are three basic types of moving averages: SMA, LMA, and EMA. The weights of the prices are given by $w_i = 1$ in $SMA_t(n, P)$, and $w_i = n - i$ in $LMA_t(n, P)$.

The EMA is computed as

$$EMA_t(n, P) = \sum_{i=0}^{\infty} (1 - \lambda)\lambda^i P_{t-i}, \quad \lambda = \frac{n-1}{n+1}, \tag{2}$$

where λ is a decay constant, $0 < \lambda < 1$. In contrast to all other types of moving averages, the EMA is computed using the averaging window of an infinite size. The parameter n in the EMA denotes the size of the averaging window in the SMA that has the same average lag time as the EMA (see Zakamulin (2017, Chapter 3)). This convention (to quote n instead of λ) is used to unify the notation for all types of moving averages.

In a trend-following rule based on one moving average, the last closing price P_t is compared with the value of the moving average $MA_t(n, P)$. A Buy (Sell) signal is generated when the last closing price is above (below) the moving average. Formally,

$$\text{Signal}_t^{\text{MA}} = \begin{cases} \text{Buy} & \text{if } P_t - MA_t(n,P) > 0, \\ \text{Sell} & \text{otherwise.} \end{cases} \tag{3}$$

The idea is that in an upward (downward) trending market, the moving average of past prices lags behind (leads) the current price.

In the Momentum (MOM) rule, the last observed price P_t is compared with the price n periods ago P_{t-n+1}. A Buy (Sell) signal is generated when the last price is above (below) the price n periods ago. The MOM rule can also be considered as a specific MA rule where all price weights are zero except the weight of the most distant price: $w_i = 0$ for $i \in [0, n-2]$ and $w_{n-1} = 1$.

In an MA crossover rule, the trading signal is computed based on the difference between a fast (short) and a slow (long) moving average. In particular, in this case, the signal is generated as

$$\text{Signal}_t^{\text{MAC}} = \begin{cases} \text{Buy} & \text{if } MA_t(s,P) - MA_t(l,P) > 0, \\ \text{Sell} & \text{otherwise,} \end{cases} \tag{4}$$

where s and l are the shorter and longer moving average sizes, respectively. A crossover rule replaces the last observed price with a short moving average of past prices. This replacement reduces the number of false signals (whipsaws). Typically, both moving averages are of the same type. Therefore, there are SMA crossovers, LMA crossovers, and EMA crossovers. Note that when $s = 1$, the MA crossover rule reduces to the corresponding MA rule.

The Moving Average Convergence/Divergence (MACD) rule is the most complicated and least understood trend-following rule based on moving averages of past prices. The trading signal of the MACD rule is computed as follows. First of all, one computes the MAC indicator using two EMAs

$$MAC_t(s,l) = EMA_t(s,P) - EMA_t(l,P).$$

Second, the signal is generated as

$$\text{Signal}_t^{\text{MACD}} = \begin{cases} \text{Buy} & \text{if } MAC_t(s,l) - EMA_t(n, MAC(s,l)) > 0, \\ \text{Sell} & \text{otherwise,} \end{cases} \tag{5}$$

where $EMA_t(n, MAC(s,l))$ denotes the results of the application of the EMA to the past values of $MAC(s,l)$. That is, the MACD rule commands buying the financial asset when the current value of $MAC(s,l)$ is greater than its value smoothed by an EMA.

2.2 Equivalent Formulation of Rules Using Past Returns

Acar (1998) and Lequeux (2005) were the first to demonstrate that the computation of the trading signal in the MA rules can be closely approximated using returns instead of prices. This idea was further extended by Beekhuizen and Hallerbach (2017) and Zakamulin (2017, Chapter 5) to cover other trend-following rules where the trading signal is computed using two or three moving averages of prices. Two key benefits are offered by formulating trading rules using returns instead of prices. First, this approach unifies the framework for trend-following rules, regardless of whether they are based on one or multiple moving averages. The trading signal for each rule can be expressed as a single moving average of past returns, simplifying the overall process. As a direct consequence, despite the many variations in rules and types of moving averages, the differences between them come down to the weighting function for returns. Second, the equivalent formulation allows us to model the return process using either a regime-switching model or the $ARMA(p,q)$ family of models and investigate the profitability and optimality of various trading rules.

Formally, the computation of the return-based trading indicator is given by

$$I_t(n) = \sum_{i=0}^{n-1} \theta_i r_{t-i}, \tag{6}$$

where n is the number of past return observations, $r_{t-i} = \frac{P_{t-i} - P_{t-i-1}}{P_{t-i-1}}$ is the time $t-i$ return, and θ_i is the weight of r_{t-i} in the computation of moving average. Note that if the trading rule uses n past returns, then the equivalent price-based trading indicator uses $n+1$ past prices.[2]

[2] The reader is reminded that the EMA uses an infinite size of the averaging window; the value of n is used to compute the decay constant.

The return weights represent an integrated version of the price weights due to the cumulative relation between prices and returns (for example, see Zakamulin (2017, Chapter 5)). Therefore, the weighting function for returns generally differs from that for prices. However, in the case of the EMA, both weighting functions are identical. In particular, the trading indicator of the EMA rule is computed as

$$I_t^{EMA}(n) = \sum_{i=0}^{\infty} (1-\lambda)\lambda^i r_{t-i}. \tag{7}$$

The trading signal of the return-based rule is generated as

$$\text{Signal}_t = \begin{cases} \text{Buy} & \text{if } I_t(n) > 0, \\ \text{Sell} & \text{otherwise.} \end{cases} \tag{8}$$

That is, the trading signal is generated based on the sign of the trading indicator. Note that the return weights in a trading indicator can be freely re-scaled without changing the trading signal. In particular, the trading indicator that uses the return weights $[a\,\theta_0, a\,\theta_1, \ldots, a\,\theta_{n-1}]$, where $a > 0$ is an arbitrary real number, generates the same trading signals as the trading indicator that uses the return weights $[\theta_0, \theta_1, \ldots, \theta_{n-1}]$. As an example, the trading indicator of the EMA rule can be computed[3] as

$$I_t^{EMA}(n) = \sum_{i=0}^{\infty} \lambda^i r_{t-i}. \tag{9}$$

For illustration, Figure 1 plots the shapes of the weighting functions for returns in the MOM, SMA, and EMA rules, the SMA and EMA crossover rules, and the MACD rule. These illustrations suggest that there are only four basic shapes of the return-weighting functions: (1) equal weighting of returns (as in the MOM rule), (2) overweighting the most recent returns (as in the SMA and EMA rules), (3) hump-shaped weighting function which underweights both the most recent and most distant returns (as in the SMA and EMA crossovers), and (4) a weighting function that has a damped waveform where the return weights alter sign (as in the MACD rule).

When a trading indicator uses monthly returns, the most popular trend-following rules are

[3]In this case, $a = (1-\lambda)^{-1}$.

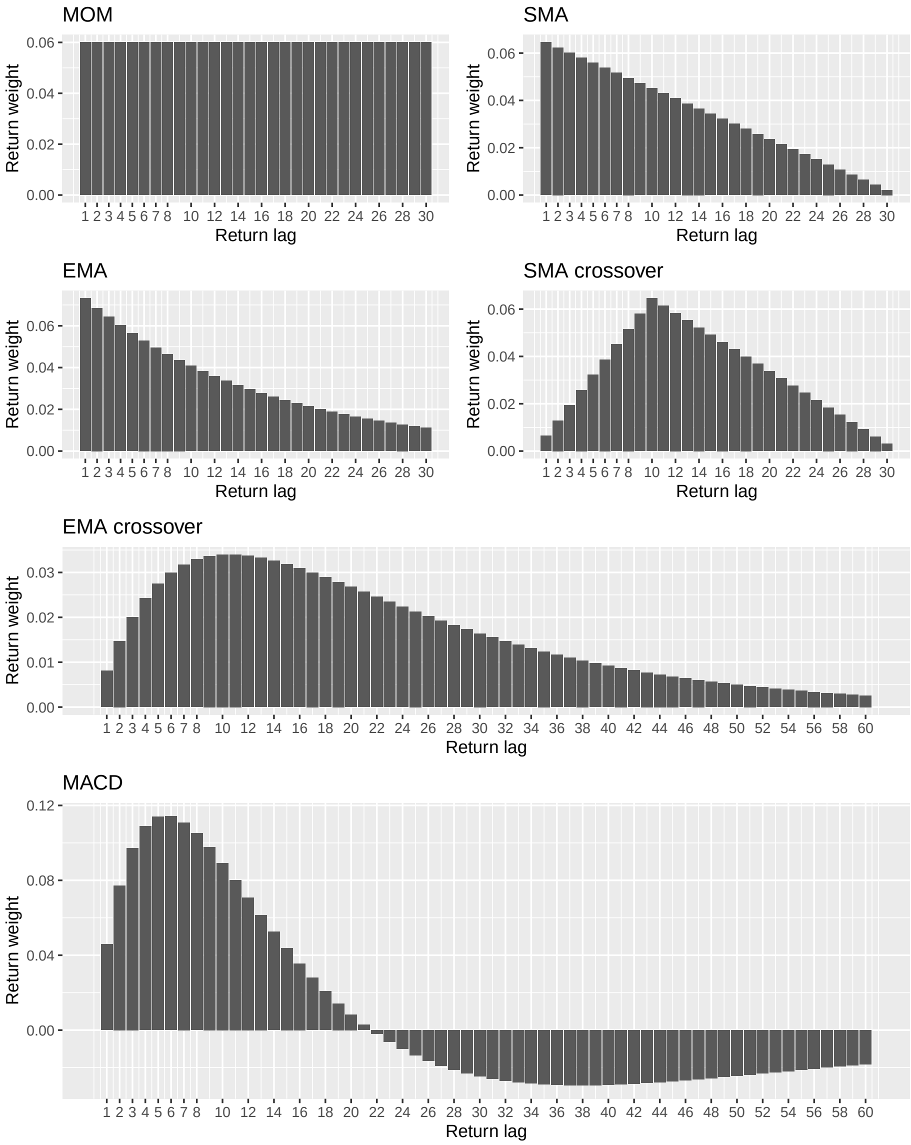


Figure 1: The shapes of the weighting functions for returns in various trend-following rules based on moving averages. In the MOM, SMA, and EMA rules, the size of the averaging window equals $n = 30$. In the SMA and EMA crossover rules, the size of the longer averaging window equals $l = 30$, while the size of the shorter averaging window equals $s = 10$ ($s = 15$) in the SMA (EMA) crossover rule. In the MACD rule, the window sizes are $s = 20$, $l = 50$, and $n = 10$. The return weights in the EMA rule are cut off at lag 30. The return weights in the EMA crossover and the MACD rules are cut off at lag 60.

the SMA(10) rule (Faber (2007), Gwilym, Clare, Seaton, and Thomas (2010), and Kilgallen (2012)) and the MOM(12) rule (Moskowitz, Ooi, and Pedersen (2012), Georgopoulou and Wang (2016), and Lim, Wang, and Yao (2018)). Note that these rules employ quite a different weighting of returns in a trading indicator. While the MOM(12) rule uses equal return weights, in the SMA(10) rule, the return weights linearly decrease with the return lag. However, it should be noted that the question of the optimality of either the MOM(12) or SMA(10) rule has never been raised in practice or theory. It is only obvious that no more than one of these rules is optimal.

3 Optimal Trend Following Rule: A Preliminary Discussion

3.1 Motivation

We know that, for the trend-following strategy to be beneficial, the returns must exhibit persistence in the short run. Put differently, the returns must display positive autocorrelation at short lags. Two examples of well-known return processes with short-term persistence are an ARMA-type process and a regime-switching model. As a matter of fact, Poskitt and Chung (1996) demonstrate that there is a one-to-one correspondence between a Markov switching model and an ARMA model. The ARMA-type process is easier to deal with than a regime-switching model. Therefore, we use the AR process to motivate the functional form of the return weights in the optimal trend-following rule.

Specifically, suppose that the market returns follow a p-order autoregressive process, $AR(p)$, defined by the following equation

$$r_t = \gamma + \sum_{i=1}^{p} \phi_i r_{t-i} + \varepsilon_t, \tag{10}$$

where γ is a constant, $\{\phi_1, \phi_2, \dots, \phi_p\}$ are the AR coefficients, and ε_t are i.i.d. random variables with zero mean and variance of σ_ε^2. It is well-known that if the returns follow an $AR(p)$ process, the best linear predictor has the same functional form as the $AR(p)$ process to be predicted, see Box, Jenkins, Reinsel, and Ljung (2016, Chapter 5). Specifically, the best linear predictor of the future return is given by

$$\hat{r}_{t+1} = \gamma + \sum_{i=0}^{p-1} \phi_{i+1} r_{t-i}, \tag{11}$$

where $\hat{r}_{t+1}$ is the predicted value at the future time $t+1$.

However, the aim of a trading indicator is not to predict the future return *per se* but to generate Buy and Sell trading signals. These trading signals must ensure the best risk-adjusted performance of the trend-following strategy. We consider the question of what return-weighting function maximizes the performance of the trend-following strategy later in this paper. As for now, a good educated guess is that the optimal trend-following indicator has the following form

$$I_t(p) = \sum_{i=0}^{p-1} (a\,\phi_{i+1})\, r_{t-i}, \tag{12}$$

where $a > 0$ is an arbitrary real number. That is, the return-weighting function of the trend-following indicator that maximizes the performance of the trend-following strategy represents a positive scaling transformation of the autoregressive coefficients of the underlying $AR(p)$ process, $\theta_i = a\,\phi_{i+1}$.

Consequently, from a theoretical point of view, the question of the optimal return weights in a trading indicator is rather trivial: the shape of the return-weighting function must resemble the shape of the AR coefficients of the return process. At first glance, all we need to do is estimate the AR coefficients of a real-life return process using linear regression. However, this straightforward approach cannot achieve the desired result. The fact is that the AR coefficients of a real-life return process are rather small and escape detection. Specifically, Zakamulin and Giner (2022) demonstrate that, even when one uses 150 years of monthly data, the values of the AR coefficients are comparable with the standard errors of their estimation. Their ballpark estimate is that we need about 650 years of monthly data to make the estimated AR coefficients statistically significant at the 5% level. Hence, the empirical estimation of AR coefficients is not a feasible approach. The only remaining avenue to gain insights into the return weights for the optimal trend-following rule is through theoretical analysis. In this paper, we employ a two-state regime-switching model to deduce the shape of the AR coefficients, with these states representing the bull and bear markets.

The empirical literature strongly supports using a regime-switching model for analyzing stock markets. Furthermore, given that such a model successfully captures the stylized facts associated with financial asset returns, it is likely to be capable of reproducing the AR coefficients. However, a fundamental question remains: why would the market adhere to such a process?

The existence of both bull and bear markets can be attributed to a combination of factors, with the primary drivers including economic conditions, bounded investor rationality, and investor sentiment. Notably, economic activity in each country experiences recurrent fluctuations known as business cycles. Understanding the underlying reasons for these cycles remains a prominent challenge in economic research, with numerous alternative business cycle theories available for exploration.[4]

In a rational expectations model, stock market movements should mirror the forward-looking behavior of investors who assess the future state of the economy. This synchronization between stock market fluctuations and business cycles is well-documented in the literature (see, for instance, Chauvet (1999) and Cruz, Nicolau, and Rodrigues (2021)).

Stock market cycles have a profound impact on investors' sentiment, and these changes in sentiment, as noted by Shiller (2000), are known to drive asset values away from their fundamentals. Additionally, investors' earnings forecasts play a crucial role in shaping stock market valuations. Due to the bounded rationality of investors, these forecasts often carry biases, as discussed by Bondt and Thaler (1990). These biases also lead to significant overvaluation in bull markets and undervaluation in bear markets (De Bondt and Thaler (1985) and De Bondt and Thaler (1989)). The resultant mispricing necessitates a correction phase during which stock prices return to their fundamental values, further reinforcing the dynamics of stock market cycles.

3.2 AR Coefficients in a Two-State Regime-Switching Model

We suppose that the return r_t is a discrete-time stochastic process that randomly switches between two states (regimes): A and B. We assume that state A is a bull state of the market, while state B is a bear state of the market. 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:

$$r_t = \begin{cases} \mu_A + \sigma_A z_t & \text{if } S_t = A, \\ \mu_B + \sigma_B z_t & \text{if } S_t = B, \end{cases} \tag{13}$$

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.

The conditional probabilities $p_{IJ}(k) = Prob(S_{t+k} = J|S_t = I)$ are called the multi-period

[4]For an extensive review, we refer the reader to Zarnowitz (1985) or Niemira and Klein (1994, Chapter 2)).

transition probabilities. In words, $p_{IJ}(k)$ is the probability that the process transits from state I to state J over k periods. The k-period transition probability distribution of the process can be represented by a 2×2 transition probability matrix $\mathbf{P}(k)$:

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

The steady-state (stationary or ergodic) probabilities are denoted by π_A and π_B and given by

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

The return autocorrelation function ρ_k is defined by (see Giner and Zakamulin (2023))

$$\rho_k = \frac{\pi_A \pi_B (\mu_A - \mu_B)^2 - (\mu_A - \mu_B)(\pi_A \, p_{AB}(k) \, \mu_A - \pi_B \, p_{BA}(k) \, \mu_B)}{\sigma^2}, \tag{14}$$

where

$$\sigma^2 = \pi_A \sigma_A^2 + \pi_B \sigma_B^2 + \pi_A \pi_B (\mu_A - \mu_B)^2.$$

It is essential to stress that the return autocorrelation function's dependence on k solely relies on the transition probabilities $p_{AB}(k)$ and $p_{BA}(k)$. The calculation of the k-period transition probabilities is significantly influenced by whether the regime-switching model is Markov or semi-Markov, and we will elaborate on this in the upcoming sections.

Once the $p \times 1$ vector $\boldsymbol{\rho}'_p = [\rho_1, \rho_2, \ldots, \rho_p]$ of the autocorrelation coefficients is determined, we can calculate the AR coefficients of the return process by solving the Yule-Walker equations

$$\mathbf{R}_{p,p} \boldsymbol{\phi}_p = \boldsymbol{\rho}_p, \tag{15}$$

where $\boldsymbol{\phi}'_p = [\phi_1, \phi_2, \ldots, \phi_p]$ is the $p \times 1$ vector that contains the AR coefficients and $\mathbf{R}_{p,p}$ is the

$p \times p$ matrix given by

$$\boldsymbol{R}_{p,p} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & 1 \end{bmatrix}. \tag{16}$$

The solution is given by

$$\boldsymbol{\phi}_p = \boldsymbol{R}_{p,p}^{-1}\boldsymbol{\rho}_p, \tag{17}$$

where $\boldsymbol{R}_{p,p}^{-1}$ is the inverse of matrix $\mathbf{R}_{p,p}$.

4 Conventional Markov Model

Panel A of Figure 2 shows the topology of a conventional two-state MSM. The one-period transition probability matrix in this model is given by

$$\mathbf{P} = \begin{pmatrix} p_{AA} & p_{AB} \\ p_{BA} & p_{BB} \end{pmatrix} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}. \tag{18}$$

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 k-period transition probability matrix is calculated as $\mathbf{P}(k) = \mathbf{P}^k$. The elements of the transition probability matrix $\mathbf{P}(k)$ are given by (see, for example, Hamilton (1994, Chapter 22))

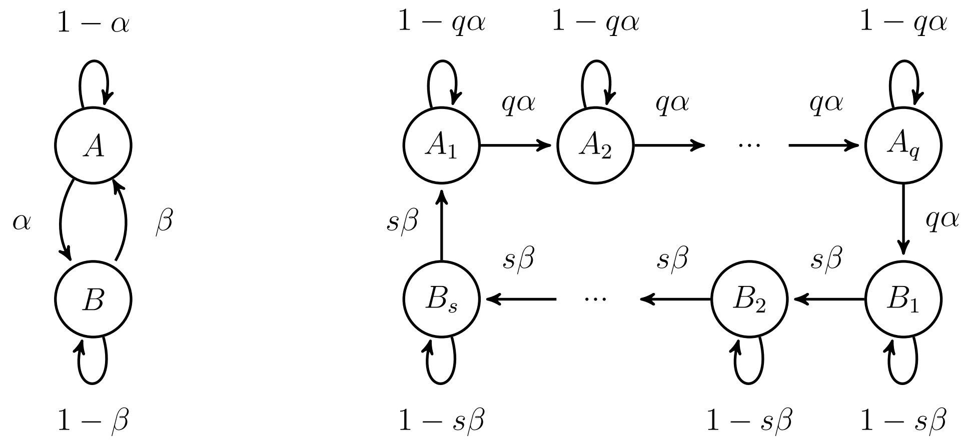
$$\mathbf{P}(k) = \begin{pmatrix} p_{AA}(k) & p_{AB}(k) \\ p_{BA}(k) & p_{BB}(k) \end{pmatrix} = \begin{pmatrix} \pi_A + \pi_B\delta^k & \pi_B - \pi_B\delta^k \\ \pi_A - \pi_A\delta^k & \pi_B + \pi_A\delta^k \end{pmatrix}, \tag{19}$$

where

$$\delta = 1 - \alpha - \beta, \tag{20}$$

and the steady-state probabilities are given by

$$\pi_A = \frac{\beta}{\alpha+\beta}, \quad \pi_B = \frac{\alpha}{\alpha+\beta}. \tag{21}$$



Panel A: Conventional Markov model — Panel B: Expanded-state Markov model

Figure 2: A conventional Markov model and an expanded-state Markov model. State A is a bull state of the market, while state B is a bear state of the market. In the expanded-state Markov model, macro-state (semi-Markovian state) A is represented by q sub-states (Markovian states), while macro-state B consists of s sub-states.

The solution for the lag-k return autocorrelation is obtained by inserting the solutions for $p_{AB}(k)$ and $p_{BA}(k)$ into equation (14). The final expression for the lag-k autocorrelation is given by

$$\rho_k = c\delta^k, \tag{22}$$

where the constant c is given by

$$c = \frac{\pi_A \pi_B (\mu_A - \mu_B)^2}{\sigma^2}. \tag{23}$$

Note that the autocorrelation is positive and exponentially decreases towards zero as k increases to infinity.

It is possible to solve analytically for the AR coefficients using the one-to-one correspondence between Markov models and ARMA models. Specifically, Poskitt and Chung (1996) prove that the process in an h-state MSM admits an $ARMA(h-1, h-1)$ representation with a homogeneous zero-mean white noise process. Consequently, the observations of the return process in a two-state MSM are indistinguishable from the observations of the return process that follows an ARMA(1,1) model. Put differently, the return process defined by equation (13) can alternatively be represented as an ARMA(1,1) process. In particular, the return process can be specified by

$$r_t = \omega + \varphi r_{t-1} + \varepsilon_t - \vartheta \varepsilon_{t-1}, \tag{24}$$

where ω, φ, and ϑ are some constants and ε_t is a homogeneous zero-mean white noise process. We assume that φ and ϑ satisfy the stationarity conditions.

It is well-known that any $ARMA(p,q)$ process admits an $AR(\infty)$ representation. The $AR(\infty)$ representation of the process specified by equation (24) is given by

$$r_t = \gamma + \sum_{i=1}^{\infty} (\varphi - \vartheta)\vartheta^{i-1} r_{t-i} + \varepsilon_t. \tag{25}$$

Note that equation (25) corresponds to equation (10) where the AR terms are given by

$$\phi_i = (\varphi - \vartheta)\vartheta^{i-1}. \tag{26}$$

This result allows us to conclude that the EMA indicator, with the decay constant $\lambda = \vartheta$, represents the optimal trend-following rule in a two-state MSM. Indeed, in this case, the optimal trend-following indicator is given by

$$I_t(n) = \sum_{i=0}^{\infty} a(\varphi - \vartheta)\vartheta^{i-1} r_{t-i}. \tag{27}$$

This trend-following indicator corresponds to the EMA trading indicator, specified by equation (9), when $a = (\varphi - \vartheta)^{-1}$ and where the value of n is given by $n = \frac{1+\vartheta}{1-\vartheta}$.

Our next goal is to find the analytical expressions for φ and ϑ. The lag-1 autocorrelation of the $ARMA(1,1)$ process is given by (see Box et al. (2016, Chapter 3))

$$\rho_1 = \frac{(\varphi - \vartheta)(1 - \varphi\vartheta)}{1 - 2\varphi\vartheta + \vartheta^2}. \tag{28}$$

Then for every ρ_k, $k > 1$, we have

$$\rho_k = \rho_1 \varphi^{k-1}. \tag{29}$$

The functional form of the lag-k return autocorrelation given by equation (29) is similar to that of the lag-k return autocorrelation specified by equation (22). This similarity is nothing other than a direct consequence of the duality between a two-state MSM process and an

ARMA(1,1) process. A comparison of (29) and (22) suggests that

$$\varphi = \delta = 1 - \alpha - \beta. \tag{30}$$

This expression is natural from the viewpoint that φ is interpreted as the persistence of the $ARMA(1,1)$ process. The smaller the values of α and β, the higher the value of φ, and the process is more likely to stay in the same state than to transit to another state.

Remains to find the analytical expression for ϑ. We find it by matching the expressions (28) and (22) for $k = 1$. This gives us the following equation:

$$\frac{(\varphi - \vartheta)(1 - \varphi\vartheta)}{1 - 2\varphi\vartheta + \vartheta^2} = c\varphi. \tag{31}$$

The result is a quadratic equation in ϑ that has two roots:

$$\vartheta_1 = d - \sqrt{d^2 - 1}, \quad \vartheta_2 = d + \sqrt{d^2 - 1},$$

where d is given by

$$d = \frac{1 + \varphi^2(1 - 2c)}{2\varphi(1 - c)}. \tag{32}$$

Since $d > 1$ (otherwise, both roots are complex numbers), we conclude that $\vartheta_2 > 1$ and, therefore, in this case, the $ARMA(1,1)$ process is non-invertible. Hence, the only satisfactory solution to equation (31) is provided by ϑ_1. Consequently, the analytical solution to ϑ is given by

$$\vartheta = d - \sqrt{d^2 - 1}. \tag{33}$$

5 Semi-Markov Model

The stylized facts about financial asset returns include fat tails, negative skewness, volatility clustering, short-term momentum, and medium-term mean reversion. A two-state conventional MSM reproduces most of these stylized facts except the mean reversion. The state duration times in a conventional MSM follow a memoryless geometric distribution. As a result, the probability of transitioning out of a state is independent of the time spent in that state, resulting in no duration dependence. However, numerous studies report that the stock market states show positive duration dependence (see, among others, Cochran and Defina (1995), Ohn et al. (2004), Harman and Zuehlke (2007), and Zakamulin (2023)). Positive duration dependence indicates that the probability of a bull or bear market ending increases as it lasts longer. Therefore, the traditional MSM cannot accurately represent the duration of bull and bear markets.

To incorporate duration dependence into a regime-switching model, the primary method is to use an SMSM instead of an MSM. Unlike the MSM, the SMSM allows for any probability distribution to govern the state duration time. However, the SMSM lacks analytical tractability and requires complex recursive algorithms for numerical computation. Recently, Giner and Zakamulin (2023) proposed an ESMSM that overcomes the disadvantages of the SMSM. The ESMSM has a specific topology[5] where the state duration times follow a negative binomial distribution, which displays positive duration dependence and simplifies to a geometric distribution under particular parameter constraints. This ESMSM has some analytical tractability and numerical computations with the same complexity as a conventional MSM. Giner and Zakamulin (2023) demonstrate that the ESMSM induces return momentum at short lags and reversal at subsequent lags.

The reader is referred to Giner and Zakamulin (2023) for all details regarding the topology of the ESMSM and the computations of transition probabilities between the bull and bear states. In this section, we briefly review the construction of the ESMSM and present the computed AR coefficients in this model.

Panel B of Figure 2 shows the general topology of the ESMSM with two macro (semi-Markovian) states, A and B. Macro-state A is represented by q sub-states (denoted by $A_1, A_2, \ldots, A_q$), while macro-state B consists of s sub-states (denoted by $B_1, B_2, \ldots, B_s$). This ESMSM extends the conventional two-state MSM depicted in Panel A of Figure 2. In this ESMSM, the state duration times follow a negative binomial distribution. For instance, each of the q sub-states of macro-state A is with self-transition, and the transition to the next macro-state is possible only from the last qth sub-state. It is assumed that the self-transition probability $p_{ii} = p_{A_i A_i}$ is the same in each sub-state $1, 2, \ldots, q$ of state A. Under this assumption, macro-state A duration time, d_A, follows a negative binomial distribution $d_A \sim NB(q)$. The probability

[5] In an ESMSM, several Markovian states (sub-states) represent one semi-Markovian state (macro-state).

mass function of the $NB(q)$ distribution is given by

$$f(n,q,p_{ii}) = Prob(d_A = n) = \binom{n-1}{n-q}(1-p_{ii})^q p_{ii}^{n-q}, \quad n \geq q. \tag{34}$$

The geometric distribution is a special case of the negative binomial distribution when $q = 1$. Consequently, when each macro-state consists of only one sub-state, $q = s = 1$, the ESMSM reduces to a conventional MSM depicted in Panel A of Figure 2.

A simple example to understand an ESMSM is when each macro state consists of only two sub-states, $q = s = 2$. Specifically, state A consists of two sub-states A_1 and A_2, while state B consists of two sub-states B_1 and B_2. In this case, the one-period transition probability matrix is given by:

$$\mathbf{P} = \begin{bmatrix} p_{A_1A_1} & p_{A_1A_2} & p_{A_1B_1} & p_{A_1B_2} \\ p_{A_2A_1} & p_{A_2A_2} & p_{A_2B_1} & p_{A_2B_2} \\ p_{B_1A_1} & p_{B_1A_2} & p_{B_1B_1} & p_{B_1B_2} \\ p_{B_2A_1} & p_{B_2A_2} & p_{B_2B_1} & p_{B_2B_2} \end{bmatrix} = \begin{bmatrix} 1-2\alpha & 2\alpha & 0 & 0 \\ 0 & 1-2\alpha & 2\alpha & 0 \\ 0 & 0 & 1-2\beta & 2\beta \\ 2\beta & 0 & 0 & 1-2\beta \end{bmatrix}. \tag{35}$$

For instance, if the process is in sub-state A_1, then over a single period, the process transits to sub-state A_2 with probability $p_{A_1A_2}$ or remains in sub-state A_1 with probability $p_{A_1A_1} = 1-p_{A_1A_2}$. If the process is in sub-state A_2, then over a single period, the process transits to sub-state B_1 with probability $p_{A_2B_1}$ or remains in sub-state A_2 with probability $p_{A_2A_2} = 1-p_{A_2B_1}$. Note that the self-transition probabilities of sub-states A_1 and A_2 (B_1 and B_2) are the same $p_{A_1A_1} = p_{A_2A_2}$ ($p_{B_1B_1} = p_{B_2B_2}$). As a result, the transition probabilities from one sub-state of macro-state A (B) to either another sub-state or another macro-state are the same $p_{A_1A_2} = p_{A_2B_1}$ ($p_{B_1B_2} = p_{B_2A_1}$).

In the ESMSM specified by the transition probability matrix in (35), the transition probabilities are computed as follows:

$$\begin{aligned} p_{AA} &= (p_{A_1A_1} + p_{A_1A_2} + p_{A_2A_1} + p_{A_2A_2})/2, \quad p_{BA} = (p_{B_1A_1} + p_{B_1A_2} + p_{B_2A_1} + p_{B_2A_2})/2, \\ p_{AB} &= (p_{A_1B_1} + p_{A_1B_2} + p_{A_2B_1} + p_{A_2B_2})/2, \quad p_{BB} = (p_{B_1B_1} + p_{B_1B_2} + p_{B_2B_1} + p_{B_2B_2})/2. \end{aligned} \tag{36}$$

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

$$\mathbf{P}(k) = \mathbf{P}^k = \begin{bmatrix} p_{A_1A_1}(k) & p_{A_1A_2}(k) & p_{A_1B_1}(k) & p_{A_1B_2}(k) \\ p_{A_2A_1}(k) & p_{A_2A_2}(k) & p_{A_2B_1}(k) & p_{A_2B_2}(k) \\ p_{B_1A_1}(k) & p_{B_1A_2}(k) & p_{B_1B_1}(k) & p_{B_1B_2}(k) \\ p_{B_2A_1}(k) & p_{B_2A_2}(k) & p_{B_2B_1}(k) & p_{B_2B_2}(k) \end{bmatrix}. \tag{37}$$

The k-period transition probabilities of macro-states A and B are computed similarly to (36). For example, the k-period self-transition probability of macro-state A is computed as $p_{AA}(k) = (p_{A_1A_1}(k) + p_{A_1A_2}(k) + p_{A_2A_1}(k) + p_{A_2A_2}(k))/2$.

Both the ESMSM and MSM have the same one-period transition probabilities for macro-states A and B. For example, $p_{AA} = 1 - \alpha$ in both the ESMSM and MSM. As a result, both the ESMSM and MSM have the same mean state duration times. Additionally, the ESMSM has the same macro-state stationary probabilities π_A and π_B. All these features provide simple comparability between the ESMSM and the corresponding MSM.

Now consider the general case where macro-state A is represented by q sub-states, while macro-state B consists of s sub-states. In this case, the one-period $(q+s) \times (q+s)$ 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 $q \times q$ sub-matrix, $\mathbf{P}_{AB}$ is the $q \times s$ sub-matrix, $\mathbf{P}_{BA}$ is the $s \times q$ sub-matrix, and $\mathbf{P}_{BB}$ is the $s \times s$ sub-matrix.[6] For instance, the self-transition probability p_{AA} (p_{BB}) of macro-state A (B) is computed by summing all elements of sub-matrix $\mathbf{P}_{AA}$ ($\mathbf{P}_{BB}$) and dividing the result by q (s). Then, the complementary probability p_{AB} (p_{BA}) can be calculated as $p_{AB} = 1 - p_{AA}$ ($p_{BA} = 1 - p_{BB}$). Regardless of the number of sub-states in each macro-state, the k-period transition probability matrix is computed in the usual manner as $\mathbf{P}(k) = \mathbf{P}^k$.

Giner and Zakamulin (2023) demonstrate that the ESMSM with two sub-states for each

[6]This model has a constraint that requires the fulfillment of the following conditions: $\alpha < 1/q$ and $\beta < 1/s$. In most cases, these conditions can be satisfied by real-world processes that have relatively small values for q and s.

macro-state is tractable analytically. Specifically, they derive analytical solutions to the k-period transition probabilities $p_{AB}(k)$ and $p_{BA}(k)$ needed to compute the lag-k return autocorrelation. Consequently, there is an analytical solution to the return weights in the optimal trend-following rule when the bull and bear state duration times are given by the $NB(2)$ distribution. In sum, the optimal trend-following rule in the ESMSM with no more than two sub-states for each market state admits an analytical solution. Regardless of the number of sub-states in each macro-state, the k-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.

Figure 3 plots the monthly AR coefficients in the ESMSM with q sub-states for each macro-state for various $q \in \{1, 2, 3, 4\}$. Regardless of the value of q, the model parameters are as follows. The annualized mean state returns are $\mu_A = 25\%$ and $\mu_B = -25\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 18\%$. The mean state A (bull market) duration time equals 28 months, and the mean state B (bear market) duration time equals 14 months. These model parameters closely resemble those estimated in the empirical section of this paper, as outlined below. The AR coefficients are computed using the following sequence of calculations. First, we compute numerically the k-period transition probabilities $p_{AB}(k)$ and $p_{BA}(k)$. Second, we compute the lag-k autocorrelation coefficients given by equation (14). Third and finally, we compute the AR coefficients[7] using equation (17).

The curves in Figure 3 suggest the following observations. In the conventional MSM, the state duration times follow the memoryless NB(1) distribution. In this case, the AR coefficients are always positive and exponentially decrease toward zero as the lag length increases. Consequently, in the MSM, the return process only exhibits short-term momentum. In the ESMSM with $q > 1$, the AR coefficients display a damped oscillating behavior around zero. That is, they periodically change the sign from positive to negative. Because oscillations decay rather fast, we mainly see positive autocorrelations at short lags and negative autocorrelations at longer lags.

To summarize, when positive duration dependence exists in the state duration times, the return process exhibits short-term momentum and subsequent mean reversion. The short-term

[7] In principle, the number of AR coefficients is infinite. However, the value of the AR coefficient quickly approaches zero with the lag length. In our numerical computations, we limit the total number of lags to 100 and show the first 30 AR coefficients.

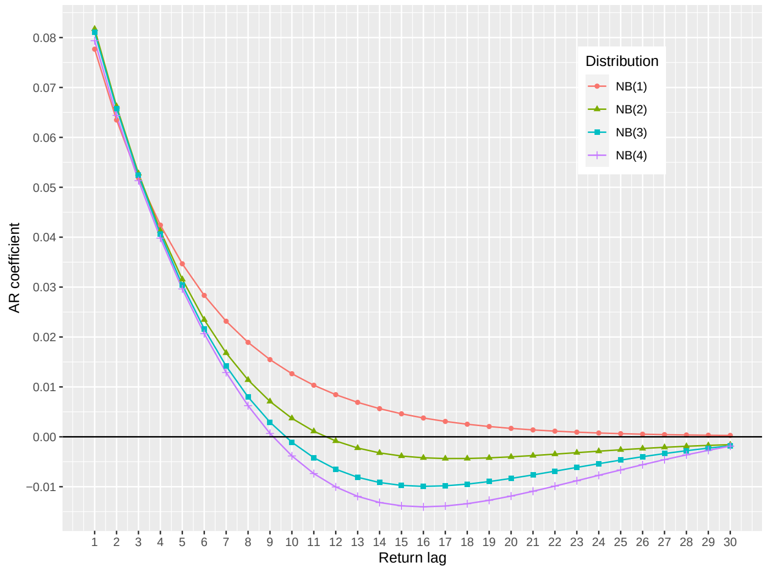


Figure 3: The monthly AR coefficients in an ESMSM with q sub-states for each macro-state for various $q \in \{1, 2, 3, 4\}$. The annualized mean state returns are $\mu_A = 25\%$ and $\mu_B = -25\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 18\%$. The mean state A duration time equals 28 months, whereas the mean state B duration time equals 14 months. The AR coefficients are cut off at lag 30.

momentum materializes as a positive return autocorrelation at short lags, while medium-term mean reversion materializes as a negative autocorrelation at longer lags. Our final observation is that the curves in Figure 3 clearly illustrate that the larger the q, the stronger the mean-reverting behavior. That is, the higher the degree of positive duration dependence, the more pronounced the mean-reverting behavior.

6 Return Weights in the Optimal Trend Following Rule

In this section, armed with the knowledge of the AR coefficients of the return process in the MSM and ESMSM, we examine the question of the optimal return weights in a trading indicator. Specifically, the question is as follows. Given the $p \times 1$ vector $\boldsymbol{\phi}_p' = [\phi_1, \phi_2, \ldots, \phi_p]$ of the AR coefficients, what is the $n \times 1$ vector $\boldsymbol{\theta}_n' = [\theta_0, \theta_1, \ldots, \theta_{n-1}]$ of the return weights in the optimal trading indicator?

We consider the long-only strategy that seeks to generate profits and limit losses by investing in the stocks only when prices trend upwards. The time t return to this trend-following strategy is given by

$$R_t = \begin{cases} r_t & \text{if } I_{t-1}(n) > 0, \\ r_f & \text{otherwise,} \end{cases} \tag{38}$$

where r_f denotes the risk-free rate of return. When the trading signal is Sell, the strategy requires selling the risky asset and investing in the risk-free instrument. Note that the time t return to the trading strategy is determined by the trading indicator's value at time $t-1$.

The optimal trend-following rule is constructed to ensure the best risk-adjusted performance of the trend-following strategy. The Sharpe ratio and CAPM alpha are the two most popular performance measures in finance. Denote by $S(R_t)$ and $\alpha(R_t)$ the Sharpe ratio and CAPM alpha, respectively, of the trend-following strategy. The Sharpe ratio is computed as

$$S(R_t) = \frac{E[R_t] - r_f}{\sqrt{Var[R_t]}},$$

where $E[R_t]$ and $Var[R_t]$ are the expected return and variance of the trend-following strategy. We assume that the market portfolio is the underlying asset in the trend-following strategy. In this case, the alpha is the intercept in the time-series regression of the trend-following strategy's excess returns on those of the passive market portfolio.

To find the optimal return weights in the trading indicator, we have to solve the following two optimization problems

$$\max_{\theta_n} \quad S(R_t) \text{ subject to } \theta_n' \mathbf{1}_n = 1, \tag{39}$$

$$\max_{\theta_n} \quad \alpha(R_t) \text{ subject to } \theta_n' \mathbf{1}_n = 1, \tag{40}$$

where $\mathbf{1}_n$ is the $n \times 1$ vector of ones. Note that a positive scaling of return weights $\boldsymbol{\theta}_n$ in a trading indicator produces an equivalent trading indicator (since the trading signal is generated based on the sign of the trading indicator's value). Therefore, there are infinite solutions to the optimal return weights without an additional constraint on the return weights. To find a unique solution, we suppose that the sum of all return weights equals one.

We remind the reader that, even though the original return process in our framework follows a regime-switching model, the return process can alternatively be viewed as an $AR(p)$ process that is specified by equation (10). Zakamulin and Giner (2022) provide analytical solutions to the Sharpe ratio and CAPM alpha of the trend-following strategy under the assumption that the innovations in the $AR(p)$ process for returns, ε_t, are normally distributed. These analytical solutions allow us to solve optimization problems (39) and (40) numerically using standard nonlinear optimization techniques.

When the ESMSM models the returns, the probability distribution of the innovations in the equivalent $AR(p)$ process for returns is largely unknown. However, to make the problem solvable, we assume that the innovations are normally distributed. In this case, the joint distribution of the returns r_t and the trading indicator $I_{t-1}(n)$ follows a bivariate normal distribution

$$\begin{bmatrix} r_t \\ I_{t-1}(n) \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mu \\ m \end{bmatrix}, \begin{bmatrix} \sigma^2 & \varrho\sigma v \\ \varrho\sigma v & v^2 \end{bmatrix} \right), \tag{41}$$

where μ and σ^2 are the mean and variance of r_t, m and v^2 are the mean and variance of $I_{t-1}(n)$, and ϱ is the correlation coefficient between r_t and $I_{t-1}(n)$.

The trend-following indicator is specified by equation (6). The mean of $I_{t-1}(n)$ equals the mean of $I_t(n)$ which is given by

$$m = E[I_t(n)] = \sum_{i=0}^{n-1} \theta_i E[r_{t-i}] = \mu \sum_{i=0}^{n-1} \theta_i. \tag{42}$$

The variance of $I_{t-1}(n)$ equals the variance of $I_t(n)$ which is given by

$$v^2 = Var(I_t(n)) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \theta_i \theta_j Cov(r_{t-i}, r_{t-j}) = \sigma^2 \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \theta_i \theta_j \rho_{|i-j|}. \tag{43}$$

To make the last step in the derivation, we use the knowledge that $Cov(r_{t-i}, r_{t-j}) = \rho_{|i-j|}\sigma^2$, where $\rho_{|i-j|}$ denotes the correlation between r_{t-i} and r_{t-j}. In matrix notation, the mean and variance of indicator $I_t(n)$ are given by

$$m = \boldsymbol{\theta}_n' \mathbf{1}_n \mu, \qquad v^2 = \boldsymbol{\theta}_n' \boldsymbol{R}_{n,n} \boldsymbol{\theta}_n \sigma^2, \tag{44}$$

where $\mathbf{1}_n$ is the $n \times 1$ vector of ones, $\mathbf{x}'$ denotes the transpose of $\mathbf{x}$, and matrix $\boldsymbol{R}_{n,n}$ is the

$n \times n$ matrix given by

$$\boldsymbol{R}_{n,n} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{bmatrix}, \tag{45}$$

where ρ_i is the autocorrelation of order i of the $AR(p)$ process for returns.

The correlation coefficient between r_t and $I_{t-1}(n)$ is computed as (see Zakamulin and Giner (2020))

$$\varrho = Cor(r_t, I_{t-1}(n)) = \frac{\theta_n' \boldsymbol{R}_{n,p} \boldsymbol{\phi}_p}{\sqrt{\theta_n' \boldsymbol{R}_{n,n} \theta_n}}, \tag{46}$$

where $\boldsymbol{R}_{n,p}$ is the $n \times p$ matrix given by

$$\boldsymbol{R}_{n,p} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & \rho_{|p-n|} \end{bmatrix}.$$

Zakamulin and Giner (2022) show that in this model, the analytical solutions for the mean returns, the variance of returns, and the CAPM alpha of the trend-following strategy are given by

$$E[R_t] = (\mu - r_f)\Phi(-d) + r_f + g, \tag{47}$$

$$Var[R_t] = (\mu^2 + \sigma^2)\Phi(-d) + g(2\mu + \sigma\varrho d) + r_f^2\Phi(d) - E[R_t]^2, \tag{48}$$

$$\alpha(R_t) = g\left(1 - \frac{(\mu - r_f)(\mu - r_f + \sigma\varrho d)}{\sigma^2}\right), \tag{49}$$

where

$$d = -\frac{m}{v}, \quad g = \sigma\varrho f(d),$$

and $f(.)$ and $\Phi(.)$ denote the probability density and the cumulative probability distribution function, respectively, of the standard normal random variable

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad \Phi(d) = \int_{-\infty}^{d} \varphi(z) dz.$$

In this context, Zakamulin and Giner (2020) provide another useful result. Specifically, they prove that the trading indicator with the return weights representing a positive scaling transformation of the AR coefficients of the return process

$$I_t(p) = \sum_{i=0}^{p-1} (a\, \phi_{i+1})\, r_{t-i}, \tag{50}$$

where $a > 0$ is an arbitrary real number, maximizes the correlation between the trading indicator and the future return, $Cor(r_t, I_{t-1}(n))$. This result is not surprising because it is congruent with our educated guess about the functional form of the optimal trend-following rule.

Besides, the trading indicator that maximizes the correlation between the trading indicator and the future return also maximizes the Sharpe ratio and CAPM alpha under the restrictive assumption that the mean market return is zero, $\mu = 0$. The proof of this result is quite illustrative because, in this case, assuming that $r_f = 0$ without loss of generality, the expressions for the mean return, the CAPM alpha, and the variance of returns of the trend-following strategy reduce to

$$E[R_t] = \alpha(R_t) = \frac{\sigma}{\sqrt{2\pi}} \varrho, \quad Var[R_t] = \frac{\sigma^2}{2} \left(1 - \frac{\varrho^2}{\pi}\right).$$

Therefore, assuming that $\varrho > 0$, the mean return and CAPM alpha increase when ϱ increases, while the variance decreases as ϱ increases. Consequently, the maximization of the correlation coefficient ensures the maximization of the Sharpe ratio and CAPM alpha of the trend-following strategy.

The result presented in the preceding paragraph suggests that when the mean market return is close to zero, $\mu \approx 0$, the trading indicator given by equation (50) is close to the optimal trading indicator. Our next goal is to examine the differences between the trading indicator specified by (50) and the optimal trading indicator when $\mu > 0$. To this end, we will numerically solve the optimization problems (39) and (40) using real-world model parameters similar to those estimated in the empirical section of this paper, as detailed below. The parameters are as follows. The annualized mean state returns are $\mu_A = 25\%$ and $\mu_B = -25\%$. The annualized standard deviations of state returns are $\sigma_A = \sigma_B = 18\%$. The annualized risk-free rate of return is 4%. The mean state A (bull market) duration time equals 28 months, and the mean state B (bear market) duration time equals 14 months.

We assume that the trend-following strategy is implemented on a monthly basis. The AR coefficients are computed assuming an equal number of sub-states in each market state. The computations are done for the conventional MSM and the ESMSM with 4 sub-states for each market state. A complication is that, in both the MSM and ESMSM, the AR process is of infinite order. That is, in principle, $n = p = \infty$. However, the AR coefficients decrease rather fast as the return lag increases. Therefore, when the trend-following strategy is implemented on a monthly basis, we can perform the calculations restricting both p and n to some sufficiently large number (for example, 100) and implement the trend-following strategy using 30 return lags.

Figure 4 shows the results of our numerical optimizations. In particular, the plots in this figure compare and contrast the shape of the AR coefficients of the return process and the shape of the return-weighting function in the optimal trend-following rules. Our main observation is that, regardless of the choice of the performance measure, there are only marginal differences between the AR coefficients and the return weights in the optimal trend-following rules. Therefore, we conclude that for practical purposes, the trend-following rule specified by equation (50) is sufficiently close to the optimal trend-following rule. Put differently, the shape of the return-weighting function in the optimal trend-following rule closely resembles the shape of the AR coefficients of the return process.

In summary, the optimal trend-following strategy in the conventional MSM is reasonably close to the EMA rule. In contrast, the optimal trend-following strategy in the ESMSM bears a close resemblance to the MACD rule. This resemblance becomes evident when visually comparing Figures 1 and 4. It is important to note that, in theory, the number of return lags in each trend-following strategy is infinite. However, in practical applications, this number must be finite. It should be sufficiently large to incorporate economically meaningful AR coefficients.

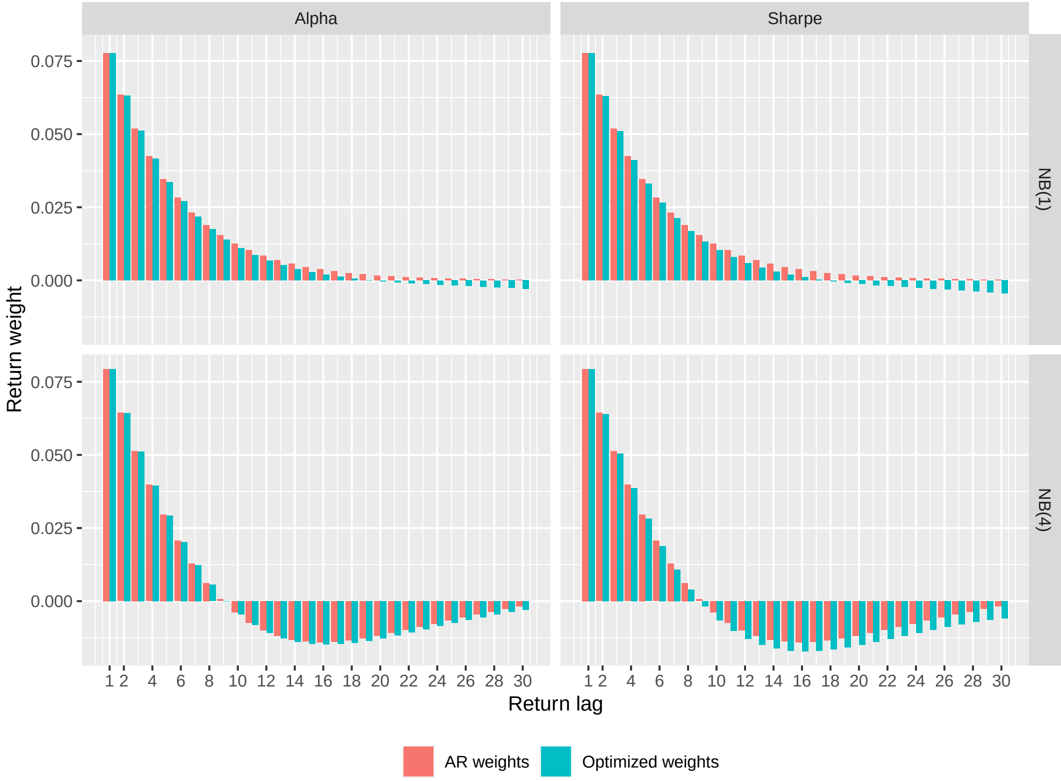


Figure 4: The shape of the AR coefficients of the return process and the shape of the return-weighting function in the optimal trend-following rules. The return weights are cut off at lag 30. The market returns follow either the conventional MSM (NB(1)) or the ESMSM with 4 sub-states for each market state (NB(4)). The objectives in the trend-following rules are to maximize either the Sharpe ratio (Sharpe) or the alpha in the CAPM (Alpha).

7 Empirical Application

The empirical literature on trend-following primarily focuses on the time series momentum strategy and begins by highlighting its profitability in various US financial markets since 1985. Subsequently, researchers have extended their analysis to encompass international equity markets and more extensive historical periods. For instance, in a study by Lim et al. (2018), data from 12 developed equity markets dating back to 1975 was used to illustrate the superiority of the time series momentum strategy. Additionally, Hurst, Ooi, and Pedersen (2017) examined the performance of trend-following investing across 11 international equity indices since 1903, finding consistent profitability over 110 years. Lempérière, Deremble, Seager, Potters, and Bouchaud (2014) collected data from seven developed equity markets dating back to 1800, documenting a stable time series momentum effect across different periods and equity markets.

This consistent long-term evidence suggests that trends are inherent features of international equity markets, and the superior performance of trend-following investing is not limited to a particular country or historical period. It is worth noting that the existing evidence primarily centers on investigating the time series momentum effect, which is based on the 12-month MOM strategy. No one has ever questioned the optimality of the 12-month MOM strategy or alternative strategies like the 10-month SMA in tracking equity market trends.

Our empirical study aims to test whether the theoretically optimal trend-following rule performs better than two commonly used rules: the 12-month MOM and the 10-month SMA rules. We test these strategies using monthly data from 16 international equity markets, including Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, and the USA. Table 1 lists the international markets used in our study and the start year in each market.

N	**Country**	**Start year**
1	Australia	1975
2	Austria	1990
3	Belgium	1975
4	Canada	1977
5	Denmark	1989
6	Finland	1988
7	France	1975
8	Germany	1975
9	Italy	1977
10	Japan	1980
11	Netherlands	1982
12	Norway	1979
13	Spain	1977
14	Sweden	1982
15	Switzerland	1975
16	USA	1875

Table 1: The list of the international markets used in our study and the start year in each of the markets.

In our sample, we collected data on market returns and risk-free rates of return for each country. All returns are denominated in the local currency. The start year for most markets in our sample ranges from 1975 to 1990, except for the US market, which has data dating back to January 1875. The data for all markets in our study extends up to December 2020.

We collected returns data for the US market from two different sources. Specifically, we obtained market returns from January 1926 to December 2020 from the Center for Research in Security Prices (CRSP). Additionally, returns from January 1875 to December 1925 were provided by William Schwert.[8] As a proxy for the risk-free rate of return in the USA, we used the Treasury-bill rate from January 1920 to December 2020. Given the absence of risk-free government debt instruments before the 1920s, we calculated the instrumented risk-free rate using the methodology recommended by Welch and Goyal (2008) based on Commercial Paper Rates for New York.

For all other countries, we gathered market return data from the online data library provided by Kenneth French.[9] Moreover, we sourced data on the risk-free rate of return for each of these countries from the Federal Reserve Economic Data (FRED) database.[10] In this context, we relied on the 3-month interbank rate as a proxy for the risk-free rate.

For each country in our study, we conducted simulations to assess the performance of the theoretically optimal trend-following strategy by employing the following out-of-sample approach. The return weights in the optimal trend-following rule are estimated using the US market data over the period from January 1875 to December 1974. Using this long-term historical period spanning a century allows us to estimate the theoretical AR coefficients of the return process that adheres to the ESMSM with a high degree of precision. Subsequently, we use these estimated return weights from the US market to simulate the performance of the theoretically optimal trend-following strategy for each country in our sample.

In the case of the US market, the out-of-sample period spans from January 1975 to December 2020. As the return weights for the optimal trend-following rule are derived exclusively from the in-sample US market data, for all other countries, the out-of-sample period covers the entire sample period. Our implicit assumption is that market trends not only prevail in all international equity markets but also exhibit shared and common characteristics.

To estimate the AR coefficients of the return process in the US market during the in-sample period, we begin by identifying the bull and bear states of the market. The turning points between the bull and bear markets are determined using a method proposed by Pagan and Sossounov (2003), which is widely accepted by researchers for this purpose. In essence, this method adapts the dating algorithm initially developed by Bry and Boschan (1971) with slight modifications. The algorithm was originally designed to identify US business cycle turning points using GDP data.

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

[9] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[10] https://fred.stlouisfed.org/

	Bull markets	**Bear markets**
Number of states	29	30
Minimum duration	4	5
Mean duration	23.56	16.07
Maximum duration	73	44
Mean return, %	25.11	−24.00
Standard deviation, %	16.27	17.86

Table 2: The summary statistics of the bull and bear states of the US market in the in-sample period from January 1875 to December 1974. Duration is measured in months. Mean returns and standard deviations are annualized and reported in percentages.

Table 2 provides an overview of the bull and bear market statistics. Notably, the mean return during a bull market stands at 25%, while during a bear market, it is −24%. The standard deviation of returns during bull and bear markets is 16% and 18%, respectively. The difference in mean returns between bull and bear markets is substantial, while the discrepancy in standard deviations is relatively small. On average, bull markets endure for approximately 24 months, whereas bear markets last around 16 months, resulting in a notably longer average duration for bull markets compared to bear markets.

Armed with the durations of the identified bull and bear states in hand, we proceed by fitting the $NB(q)$ distribution to the state duration data to determine the most suitable q value. In this context, the probability mass function $f(n, q, p)$ quantifies the likelihood of the qth success occurring in the nth Bernoulli trial, with the parameter p representing the probability of success in a single trial. Our approach relies on maximum likelihood estimation (MLE) to fit these distributions. The standard procedure involves identifying the pair of parameters (q, p) that maximizes the log-likelihood function.

However, a complication arises when dealing with the q parameter. Typically, q is considered a real number, but our model assumes it is an integer. To address this challenge, we suppose that q is a known value and proceed to find the maximum likelihood estimator for the p parameter only. We carry out this process sequentially for various integer values of q within the range $\{1, \ldots, 6\}$ and select the value of q that yields the maximum log-likelihood.

Table 3 provides the estimated values of p and the associated log-likelihood values resulting from the maximum likelihood estimation of p for different q values. The log-likelihood values strongly indicate that the negative binomial distribution with $q = 4$ maximizes the log-likelihood function for both bull and bear market states. As a result, we conclude that the

q	Bull markets		Bear markets	
	p	Log-likelihood	p	Log-likelihood
1	0.040	−121.76	0.056	−115.52
2	0.077	−115.42	0.106	−108.80
3	0.111	−113.59	0.152	−106.67
4	0.143	−113.30	0.192	−106.08
5	0.172	−113.73	0.229	−106.20
6	0.200	−114.52	0.263	−106.66

Table 3: The results of maximum likelihood estimations using the US market data for the in-sample period from January 1875 to December 1974. 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.

durations of both bull and bear states are best described by the $NB(4)$ distribution.

Given the estimated parameters of the ESMSM that governs the return process, we compute the theoretically implied AR coefficients. The procedure is described in Section 5. The shape of the returns weights of the theoretically optimal trend-following rule mirrors the shape of the AR coefficients. When simulating returns for the theoretically optimal trend-following strategy, we limit the number of return weights to 30.

In addition to the theoretically optimal trend-following strategy, we also simulate returns for the MOM(12) strategy and the SMA(10) strategy. After completing these simulations, we compute the Sharpe ratio of each competing trend-following strategy and the passive buy-and-hold benchmark. We then assess whether the trend-following strategy performs better than the buy-and-hold strategy by conducting a Sharpe ratio test, where the null hypothesis is as follows:

$$H_0 : S_{TF} = S_{BH} \text{ versus } H_A : S_{TF} > S_{BH},$$

where S_{TF} and S_{BH} represent the Sharpe ratios of the trend-following and buy-and-hold strategies, respectively. We perform this test using the Jobson and Korkie (1981) test, which is corrected by Memmel (2003). The test statistic used in the Jobson and Korkie (1981) test is given by:

$$z = \frac{S_{TF} - S_{BH}}{\sqrt{\frac{1}{T}\left[2(1-\rho) + \frac{1}{2}(S_{TF}^2 + S_{BH}^2 - 2\rho^2 S_{TF} S_{BH})\right]}},$$

where S_{TF}, S_{BH}, and ρ are the estimated Sharpe ratios and the correlation coefficient between the returns of the two strategies over a sample of T months. Under the null hypothesis, the test statistic z follows an asymptotic standard normal distribution.

	Sharpe ratio	P-value		Sharpe ratio	P-value
1. Australia			**9. Italy**		
B&H	0.30		B&H	0.26	
NB(4)	**0.42**	0.13	NB(4)	**0.50**	0.02
MOM(12)	0.34	0.29	MOM(12)	0.47	0.01
SMA(10)	0.33	0.37	SMA(10)	0.39	0.10
2. Austria			**10. Japan**		
B&H	0.31		B&H	0.23	
NB(4)	**0.54**	0.07	NB(4)	**0.50**	0.02
MOM(12)	0.40	0.27	MOM(12)	0.45	0.04
SMA(10)	0.50	0.11	SMA(10)	0.39	0.11
3. Belgium			**11. Netherlands**		
B&H	0.43		B&H	0.47	
NB(4)	**0.75**	0.01	NB(4)	0.63	0.12
MOM(12)	0.60	0.05	MOM(12)	**0.67**	0.04
SMA(10)	0.63	0.03	SMA(10)	0.66	0.06
4. Canada			**12. Norway**		
B&H	0.28		B&H	0.34	
NB(4)	**0.49**	0.05	NB(4)	**0.49**	0.12
MOM(12)	0.26	0.57	MOM(12)	0.45	0.15
SMA(10)	0.32	0.32	SMA(10)	0.44	0.18
5. Denmark			**13. Spain**		
B&H	0.44		B&H	0.35	
NB(4)	**0.92**	0.00	NB(4)	0.42	0.28
MOM(12)	0.56	0.17	MOM(12)	0.41	0.27
SMA(10)	0.77	0.01	SMA(10)	**0.48**	0.09
6. Finland			**14. Sweden**		
B&H	0.39		B&H	0.44	
NB(4)	0.58	0.10	NB(4)	**0.73**	0.03
MOM(12)	**0.74**	0.01	MOM(12)	0.61	0.08
SMA(10)	0.61	0.05	SMA(10)	0.71	0.02
7. France			**15. Switzerland**		
B&H	0.42		B&H	0.46	
NB(4)	0.61	0.06	NB(4)	**0.68**	0.03
MOM(12)	0.48	0.25	MOM(12)	0.55	0.16
SMA(10)	**0.63**	0.02	SMA(10)	0.56	0.16
8. Germany			**16. USA**		
B&H	0.34		B&H	0.44	
NB(4)	**0.59**	0.02	NB(4)	**0.52**	0.27
MOM(12)	0.42	0.21	MOM(12)	0.50	0.25
SMA(10)	0.45	0.14	SMA(10)	0.51	0.26

Table 4: The annualized out-of-sample Sharpe ratios for four strategies: Buy-and-Hold (B&H), MOM(12), SMA(10), and the theoretically optimal trend-following strategy (NB(4)). The p-values correspond to hypothesis tests, examining whether the Sharpe ratio of a trend-following strategy equals the Sharpe ratio of the passive strategy. Sharpe ratios highlighted in bold indicate the top-performing strategy in each market.

For all countries in our sample, Table 4 presents the out-of-sample Sharpe ratios for the buy-and-hold, MOM(12), SMA(10), and the theoretically optimal trend-following strategy. Additionally, the table presents p-values for hypothesis tests, which evaluate whether the Sharpe ratio of a trend-following strategy equals that of the passive strategy.

The Sharpe ratios in Table 4 suggest the following observations. First, in each market, trend-following strategies tend to outperform their respective buy-and-hold counterparts. Second, the theoretically optimal trend-following strategy seems to exhibit superior performance. Specifically, it boasts the highest Sharpe ratio in 12 out of 16 international markets. The MOM(12) strategy performs best in two markets, and the SMA(10) strategy excels in two others.

Regrettably, we cannot draw a scientifically sound conclusion solely based on the estimated Sharpe ratios of a trend-following strategy in comparison to its passive or active counterparts. The results presented in Table 4 currently indicate, for instance, that the Sharpe ratio of the optimal trend-following strategy is statistically significantly higher (at the 10% significance level) than the Sharpe ratio of the buy-and-hold strategy in 11 out of 16 international markets. However, this information alone does not suffice to conclusively establish that the optimal trend-following strategy outperforms the buy-and-hold strategy across all markets. To make such a conclusion, a joint test is necessary.

In our joint test, the null hypothesis posits that two alternative strategies yield similar performance across all international markets. Conversely, the alternative hypothesis suggests that one strategy outperforms its counterpart on average, across all international markets. We describe our joint test by explaining how to assess whether a trend-following strategy is superior to the buy-and-hold strategy across all international markets.

While conducting a joint test, it is essential to normalize the (out)performance statistic, as described by Hansen (2005). With this in mind, the difference in the Sharpe ratios $\Delta S = S_{TF} - S_{BH}$ is not a good performance statistic because international markets may differ substantially with respect to the probability distributions of the Sharpe ratio differences. In contrast, the z-statistic in the Sharpe ratio test represents a normalized performance measure.[11] In particular, assuming that the two Sharpe ratios are equal, the value of the z-statistic is zero.

Consequently, our joint test assumes that the true value of the z-statistic is zero in each international market. Formally, we assume that $z_i = 0$ for all i in the range $[1, 2, \dots, 16]$, where i represents a different country. The test statistic in our joint test quantifies the average

[11] Specifically, the z-statistic in the Sharpe ratio test represents the difference in two Sharpe ratios divided by the standard error of the estimated difference.

outperformance of a trend-following strategy across all international markets:

$$\bar{z} = \frac{1}{16}\sum_{i=1}^{16} z_i.$$

Given that the true value of each z_i is zero, the average outperformance is also expected to be zero. As a result, our null and alternative hypotheses for the joint test are as follows:

$$H_0 : \bar{z} = 0, \text{ versus } H_A : \bar{z} > 0.$$

Rejection of the null hypothesis in favor of the alternative hypothesis provides evidence that, on average across all markets, a trend-following strategy outperforms the buy-and-hold strategy.

The probability distribution of $\bar{z}$ under the null hypothesis is estimated using the stationary block-bootstrap method of Politis and Romano (1994). This method keeps the empirical time dependence in each return series. Besides, we bootstrap jointly all return series in order to keep the empirical correlations across the markets. We carry out $R = 1,000$ bootstrap trials in total.[12]

In each bootstrap trial, we calculate the z-statistic for all countries under the null hypothesis of a true value of zero for each statistic. Subsequently, we compute the average of all these z-statistics. Ultimately, the collection of average z-statistics from all trials forms the probability distribution of $\bar{z}$ under the null hypothesis. To determine the p-value of the test, we calculate the percentage of the simulated $\bar{z}$ values that are equal to or more extreme than the empirically estimated $\bar{z}$.

We systematically experimented with varying the block length within the stationary block-bootstrap method, ranging from 5 to 12. Remarkably, our findings revealed that the estimated p-values for the joint test exhibited very little sensitivity to the choice of block length. This robustness suggests that our results are reliable and not significantly affected by the specific block length used in the bootstrap method.

Table 5 presents the results of both individual and joint Sharpe ratio tests for each trend-

[12] In principle, if we assume that each z_i follows a standard normal distribution, the average z-statistic also follows a normal distribution with a mean of zero. However, to calculate its standard deviation, we need to know the correlation coefficients between different z_i. Since calculating these correlation coefficients analytically is extremely cumbersome, if not impossible, we turn to the bootstrap method. Another advantage of the bootstrap method is that it does not require any parametric assumptions.

	MOM(12) vs B&H		SMA(10) vs B&H		NB(4) vs B&H	
	z-statistic	p-value	z-statistic	p-value	z-statistic	p-value
Australia	0.55	0.29	0.34	0.37	1.13	0.13
Austria	0.61	0.27	1.24	0.11	**1.47**	0.07
Belgium	**1.66**	0.05	**1.82**	0.03	**2.51**	0.01
Canada	-0.18	0.57	0.47	0.32	**1.68**	0.05
Denmark	0.96	0.17	**2.44**	0.01	**3.10**	0.00
Finland	**2.41**	0.01	**1.61**	0.05	**1.27**	0.10
France	0.67	0.25	**2.01**	0.02	**1.55**	0.06
Germany	0.81	0.21	1.06	0.14	**2.09**	0.02
Italy	**2.20**	0.01	**1.29**	0.10	**2.07**	0.02
Japan	**1.80**	0.04	1.24	0.11	**2.05**	0.02
Netherlands	**1.73**	0.04	**1.54**	0.06	1.16	0.12
Norway	1.04	0.15	0.93	0.18	1.16	0.12
Spain	0.62	0.27	**1.32**	0.09	0.59	0.28
Sweden	**1.42**	0.08	**2.04**	0.02	**1.95**	0.03
Switzerland	0.99	0.16	1.01	0.16	**1.95**	0.03
USA	0.68	0.25	0.65	0.26	0.61	0.27
Average	**1.12**	0.05	**1.31**	0.04	**1.65**	0.01

Table 5: The results of both individual and joint Sharpe ratio tests to determine if each trend-following strategy outperforms the buy-and-hold strategy. In the individual Sharpe ratio test, the z-statistic is employed for each country, while the joint test of outperformance uses the average z-statistic. The corresponding p-value for each test is situated to the left of the respective z-statistic. Values that are statistically significant at the 10% level are indicated in bold.

following strategy. In the individual Sharpe ratio test, the z-statistic serves as the test statistic for each country, enabling us to determine whether a trend-following strategy outperforms its passive benchmark in that specific country. Meanwhile, the joint test of outperformance utilizes the average z-statistic as its test statistic, helping us ascertain whether, on average, a trend-following strategy outperforms the passive benchmark across all countries. The corresponding p-value for each test is reported to the left of its respective z-statistic.

The results presented in Table 5 show that, on average across all countries, each trend-following strategy outperforms its respective buy-and-hold counterpart. Notably, the theoretically optimal trend-following strategy achieves the highest average z-statistic, which is statistically significant at the 1% level. On the other hand, the MOM(12) strategy exhibits the smallest average z-statistic, but it remains statistically significant at the 5% level.

Based on the average z-statistics from the joint tests of outperformance, we can establish the following ranking for the competing trend-following strategies, from best to worst: the NB(4) strategy, the SMA(10) strategy, and finally the MOM(12) strategy. However, it is important to note that a conclusive scientific inference can only be drawn from the results of statistical tests. Therefore, we proceed to conduct the same individual and joint tests of outperformance to assess whether a specific trend-following strategy outperforms its competitor. The outcomes of these tests are presented in Table 6.

	NB(4) vs MOM(12)		NB(4) vs SMA(10)		SMA(10) vs MOM(12)	
	z-statistic	p-value	z-statistic	p-value	z-statistic	p-value
Australia	0.82	0.21	**1.36**	0.09	-0.19	0.58
Austria	0.99	0.16	0.52	0.30	0.81	0.21
Belgium	**1.32**	0.09	**1.27**	0.10	0.46	0.32
Canada	**2.04**	0.02	**1.73**	0.04	0.93	0.18
Denmark	**2.78**	0.00	**1.61**	0.05	**2.23**	0.01
Finland	-1.38	0.92	-0.39	0.65	-1.59	0.94
France	1.13	0.13	-0.29	0.61	**1.76**	0.04
Germany	**1.69**	0.05	**2.01**	0.02	0.40	0.34
Italy	0.32	0.37	**1.54**	0.06	-1.15	0.88
Japan	0.42	0.34	**1.29**	0.10	-0.58	0.72
Netherlands	-0.31	0.62	-0.28	0.61	-0.15	0.56
Norway	0.30	0.38	0.54	0.29	-0.10	0.54
Spain	0.10	0.46	-0.86	0.80	1.03	0.15
Sweden	0.95	0.17	0.15	0.44	1.11	0.13
Switzerland	**1.31**	0.09	**1.65**	0.05	0.14	0.45
USA	0.13	0.45	0.12	0.45	0.05	0.48
Average	**0.79**	0.06	**0.75**	0.03	0.32	0.26

Table 6: The results of both individual and joint Sharpe ratio tests to determine if a specific trend-following strategy outperforms its competitor. In the individual Sharpe ratio test, the z-statistic is employed for each country, while the joint test of outperformance uses the average z-statistic. The corresponding p-value for each test is situated to the left of the respective z-statistic. Values that are statistically significant at the 10% level are indicated in bold.

Our primary focus is to determine whether the theoretically optimal trend-following strategy outperforms both the MOM(12) and SMA(10) strategies. The findings presented in Table 6 affirm the superiority of the NB(4) strategy. For instance, when compared to the SMA(10) strategy, the NB(4) strategy exhibits a statistically significant Sharpe ratio in 8 out of 16 countries. Furthermore, the average z-statistic in the joint test is statistically significant at the 5% level. This outcome leads us to conclude that, on average across all countries, the NB(4) strategy outperforms the SMA(10) strategy.

Likewise, we conclude that the NB(4) strategy outperforms the MOM(12) strategy on average across all markets. However, it is worth noting that although the average z statistic in this test is the highest, it is statistically significant at the 10% level only. This can be explained by the fact that the probability distribution of this test statistic exhibits greater variability compared to the probability distribution of the same statistics in the joint test of the NB(4) strategy versus the SMA(10) strategy.

We also assess whether the SMA(10) strategy outperforms the MOM(12) strategy. It is important to note that the average z-statistic in the joint test of outperformance is positive, suggesting that there are indications that the SMA(10) strategy performs better than the MOM(12) strategy on average. However, the p-value of the test does not provide sufficient evidence to reject the null hypothesis of similar performance.

In summary, the results of our empirical study validate the soundness of our theoretical model and underscore the superiority of the theoretically optimal trend-following strategy. To conclude this section, we aim to delve further into the characteristics of the optimal trend-following rule and its distinctions and commonalities with the ad-hoc rules typically employed in practice.

First and foremost, it is worth noting that all existing ad-hoc rules, except the MACD rule, primarily capitalize on short-term momentum, which is prevalent in most financial markets. However, financial asset returns not only exhibit short-term momentum but also subsequent medium-term mean reversion. The MACD rule stands out as it seeks to harness both short-term momentum and medium-term mean reversion. Nevertheless, while the return weights in the MACD rule share similarities with those in the optimal rule, using the MACD rule in practice poses potential risks. First, the MACD rule is ad-hoc, with return weights resembling but differing from those in the theoretically optimal trend-following rule. Second, compared to the MOM(12) and SMA(10) rules, the MACD rule is over-parameterized, requiring three parameters, making it susceptible to backtest overfitting.[13]

Second but no less important, when we focus solely on short-term momentum, the return weights in the optimal trend-following rule decline as the return lag increases. This phenomenon clarifies why the SMA(10) rule often outperforms the MOM(12) rule in empirical tests. Neither the SMA(10) nor the MOM(12) rule can be deemed optimal, at least from a theoretical standpoint, but the return weights in the SMA(10) rule are closer to the theoretically optimal return weights compared to those in the MOM(12) rule. For a visual representation of this, please refer to Figure 5, which illustrates the return weights in the NB(4) and SMA(10) rules. In principle, the SMA(10) rule accurately captures the duration of short-term momentum, which typically lasts around 9 months before undergoing a reversal. The return weights in the SMA(10) rule exhibit a linear decrease, whereas the return weights in the optimal rule

[13]For a detailed discussion of backtest overfitting, we refer the reader to Bailey, Borwein, de Prado, and Zhu (2014). In brief, backtest overfitting refers to a situation in which a trading strategy optimized on the in-sample data performs poorly on the out-of-sample data.

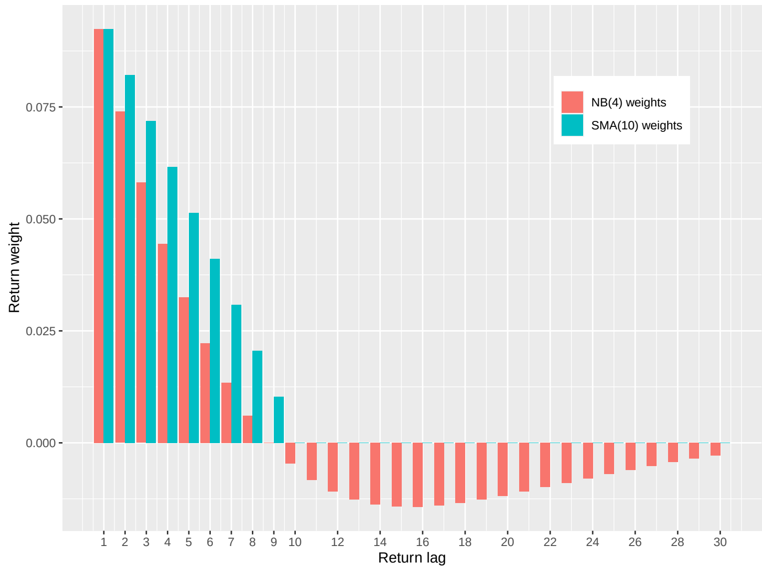


Figure 5: The return weights in the NB(4) and SMA(10) rules.

follow an exponential decline.

8 Conclusions

Despite the widespread popularity of trend-following investing, little is still known about the theoretically optimal trend-following rules. This paper fills this gap in the literature and examines the optimal trend-following in two-state discrete-time Markov and semi-Markov regime-switching models. We start with a conventional Markov model and find that, in this case, the optimal trend-following problem is analytically tractable. We demonstrate that following the trend using the EMA is optimal and provide the analytical solution to the optimal window size (decay constant) in the EMA.

The absence of duration dependence limits the validity of a conventional Markov model in financial applications. Therefore, we proceed to the case of a semi-Markov model where the distribution of the state duration times exhibits positive duration dependence. In this case, the optimal trend-following problem is generally not analytically tractable, but our choice of the Markov chain topology makes numerical computations quite simple. Our numerical results show that the optimal trend-following rule in a semi-Markov model is somewhat similar to the MACD rule.

In our empirical analysis, we utilize data from a diverse set of international stock markets.

Through out-of-sample simulations, we present compelling scientific evidence supporting the superiority of our theoretically optimal trading rule within the semi-Markov model over the widely used 10-month SMA and the 12-month MOM rules. These findings not only confirm the soundness of our theoretical model but also emphasize its practical relevance in the field of investment.

References

Acar, E. (1998). "Expected Returns of Directional Forecasters", In Acar, E. and Satchell, S. (Eds.), *Advanced Trading Rules*, pp. 51–76. Butterworth-Heinemann, Oxford, UK.

Bailey, D. H., Borwein, J. M., de Prado, M. L., and Zhu, Q. J. (2014). "Pseudomathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-Of-Sample Performance", *Notices of the American Mathematical Society*, *61*(5), 458–471.

Beekhuizen, P. and Hallerbach, W. G. (2017). "Uncovering Trend Rules", *Journal of Alternative Investments*, *20*(2), 28–38.

Bondt, W. F. M. D. and Thaler, R. H. (1990). "Do Security Analysts Overreact?", *American Economic Review*, *80*(2), 52–57.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2016). *Time Series Analysis: Forecasting and Control* (5 edition). John Wiley & Sons Inc., New Jersey.

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

Chauvet, M. (1999). "Stock Market Fluctuations and the Business Cycle", *Journal of Economic and Social Measurement*, *25*(3-4), 235–257.

Cochran, S. J. and Defina, R. H. (1995). "Duration Dependence in the US Stock Market Cycle: A Parametric Approach", *Applied Financial Economics*, *5*(5), 309–318.

Cont, R. (2001). "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues", *Quantitative Finance*, *1*(2), 223–236.

Cruz, J., Nicolau, J., and Rodrigues, P. (2021). "Structural Changes in the Duration of Bull Markets and Business Cycle Dynamics", *Asia-Pacific Financial Markets*, *28*, 333–352.

Dai, M., Zhang, Q., and Zhu, Q. J. (2010). "Trend Following Trading under a Regime Switching Model", *SIAM Journal on Financial Mathematics*, *1*(1), 780–810.

Dai, M., Yang, Z., Zhang, Q., and Zhu, Q. J. (2016). "Optimal Trend Following Trading Rules", *Mathematics of Operations Research*, *41*(2), 626–642.

De Bondt, W. F. M. and Thaler, R. (1985). "Does the Stock Market Overreact?", *Journal of Finance*, *40*(3), 793–805.

De Bondt, W. F. M. and Thaler, R. H. (1989). "Anomalies: A Mean-Reverting Walk Down Wall Street", *Journal of Economic Perspectives*, *3*(1), 189–202.

Faber, M. T. (2007). “A Quantitative Approach to Tactical Asset Allocation”, *Journal of Wealth Management*, *9*(4), 69–79.

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

Georgopoulou, A. and Wang, J. G. (2016). “The Trend Is Your Friend: Time-Series Momentum Strategies Across Equity and Commodity Markets”, *Review of Finance*, *21*(4), 1557–1592.

Giner, J. and Zakamulin, V. (2023). “A Regime-Switching Model of Stock Returns with Momentum and Mean Reversion”, *Economic Modelling*, *122*, 106237.

Gwilym, O., Clare, A., Seaton, J., and Thomas, S. (2010). “Price and Momentum as Robust Tactical Approaches to Global Equity Investing”, *Journal of Investing*, *19*(3), 80–91.

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

Hansen, P. R. (2005). “A Test for Superior Predictive Ability”, *Journal of Business & Economic Statistics*, *23*(4), 365–380.

Harman, Y. S. and Zuehlke, T. W. (2007). “Nonlinear Duration Dependence in Stock Market Cycles”, *Review of Financial Economics*, *16*(4), 350 − 362.

Hurst, B., Ooi, Y. H., and Pedersen, L. H. (2017). “A Century of Evidence on Trend-Following Investing”, *Journal of Portfolio Management*, *44*(1), 15–29.

Jobson, J. D. and Korkie, B. M. (1981). “Performance Hypothesis Testing with the Sharpe and Treynor Measures”, *Journal of Finance*, *36*(4), 889–908.

Kilgallen, T. (2012). “Testing the Simple Moving Average Across Commodities, Global Stock Indices, and Currencies”, *Journal of Wealth Management*, *15*(1), 82–100.

Lempérière, Y., Deremble, C., Seager, P., Potters, M., and Bouchaud, J.-P. (2014). “Two Centuries of Trend-Following”, *Journal of Investment Strategies*, *3*(3), 41–61.

Lequeux, P. (2005). “Optimal Allocation of Trend-Following Rules: An Application Case of Theoretical Results”, In Dunis, C. L., Laws, J., and Naim, P. (Eds.), *Applied Quantitative Methods for Trading and Investment*, pp. 333–347. Wiley-Blackwell, West Sussex, UK.

Lim, B. Y., Wang, J. G., and Yao, Y. (2018). “Time-Series Momentum in Nearly 100 Years of Stock Returns”, *Journal of Banking & Finance*, *97*, 283–296.

Memmel, C. (2003). “Performance Hypothesis Testing with the Sharpe Ratio”, *Finance Letters*, *1*, 21–23.

Moskowitz, T. J., Ooi, Y. H., and Pedersen, L. H. (2012). “Time Series Momentum”, *Journal of Financial Economics*, *104*(2), 228–250.

Nguyen, D., Tie, J., and Zhang, Q. (2014a). “An Optimal Trading Rule Under a Switchable Mean-Reversion Model”, *Journal of Optimization Theory and Applications*, *161*, 145–163.

Nguyen, D., Yin, G., and Zhang, Q. (2014b). "A Stochastic Approximation Approach for Trend-Following Trading", In Mamon, R. S. and Elliott, R. J. (Eds.), *Hidden Markov Models in Finance: Further Developments and Applications, Volume II*, pp. 167–184. Springer US, Boston, MA.

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

Ohn, J., Taylor, L. W., and Pagan, A. (2004). "Testing for Duration Dependence in Economic Cycles", *Econometrics Journal*, *7*(2), 528–549.

Pagan, A. R. and Sossounov, K. A. (2003). "A Simple Framework for Analysing Bull and Bear Markets", *Journal of Applied Econometrics*, *18*(1), 23–46.

Politis, D. and Romano, J. (1994). "The Stationary Bootstrap", *Journal of the American Statistical Association*, *89*(428), 1303–1313.

Poskitt, D. S. and Chung, S.-H. (1996). "Markov Chain Models, Time Series Analysis and Extreme Value Theory", *Advances in Applied Probability*, *28*(2), 405–425.

Shiller, R. J. (2000). *Irrational Exuberance*. Princeton University Press.

Tie, J. and Zhang, Q. (2016). "An Optimal Mean-Reversion Trading Rule Under a Markov Chain Model", *Mathematical Control & Related Fields*, *6*(3), 467–488.

Timmermann, A. (2000). "Moments of Markov Switching Models", *Journal of Econometrics*, *96*(1), 75–111.

Welch, I. and Goyal, A. (2008). "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction", *Review of Financial Studies*, *21*(4), 1455–1508.

Zakamulin, V. (2017). *Market Timing with Moving Averages: The Anatomy and Performance of Trading Rules*. Palgrave Macmillan.

Zakamulin, V. (2023). "Revisiting the Duration Dependence in the US Stock Market Cycles", *Applied Economics*, *55*(4), 357–368.

Zakamulin, V. and Giner, J. (2020). "Trend Following with Momentum Versus Moving Average: A Tale of Differences", *Quantitative Finance*, *20*(6), 985–1007.

Zakamulin, V. and Giner, J. (2022). "Time Series Momentum in the US Stock Market: Empirical Evidence and Theoretical Analysis", *International Review of Financial Analysis*, *82*, 102173.

Zarnowitz, V. (1985). "Recent Work on Business Cycles in Historical Perspective: A Review of Theories and Evidence", *Journal of Economic Literature*, *23*(2), 523–580.