Scientific Research Publishing

Journal of Mathematical Finance, 2026, 16(4), 226-246
https://www.scirp.org/journal/jmf
ISSN Online: 2162-2442
ISSN Print: 2162-2434

On the Importance of Model Selection in the Validation of Technical Strategies Using the Bootstrap Method

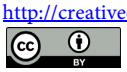
Kazuki Yahagi[1,2], Koichi Miyazaki[1]

[1]Graduate School of Business Administration, Saitama Gakuen University, Kawaguchi, Japan
[2]Sumitomo Mitsui Trust Bank, Limited, Tokyo, Japan
Email: 26DB0001@saigaku.ac.jp, k.miyazaki@saigaku.ac.jp

How to cite this paper: Yahagi, K. and Miyazaki, K. (2026) On the Importance of Model Selection in the Validation of Technical Strategies Using the Bootstrap Method. *Journal of Mathematical Finance*, **16**, 226-246.
https://doi.org/10.4236/jmf.2026.164011

Received: August 13, 2026
Accepted: November 21, 2026
Published: November 24, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



Open Access

Abstract

This study examines the importance of stock-return time-series model selection when bootstrap methods are used to validate technical trading strategies. Building on Brock, Lakonishok, and LeBaron, we compare eight models: Normal, AR(1), GARCH-M, EGARCH, two threshold models, and two hidden Markov switching models. Model adequacy is assessed using AIC and BIC, the Wasserstein distance between actual and bootstrapped return distributions, and Kendall's rank correlation between the resulting rankings. The empirical application uses daily Nikkei 225 data from January 2010 to December 2019 and six variable moving-average buy-signal strategies. The information-criterion and Wasserstein rankings are broadly consistent (Kendall's $\tau = 0.7143$), and the hidden Markov switching models show the strongest distributional reproducibility. Although the observed VMA strategies do not clearly outperform buy-and-hold, the best-fitting model indicates that an appropriately selected strategy may do so with a probability of 60% or more based on the model validation using in-sample data. The results show that pre-evaluating model fit and distributional reproducibility is essential for reliable bootstrap-based validation.

Keywords

Technical Trading Strategies, Bootstrap Method, Time-Series Model Selection, Wasserstein Distance, Hidden Markov Switching Model, Nikkei 225

1. Introduction

The efficient market hypothesis, typified by the random walk model, denies the

possibility of profits through technical strategies. In contrast, in futures and foreign exchange markets, many market participants widely employ technical strategies in practice and may recognize them as significant factors influencing short-term price fluctuations. Against the backdrop of this gap between theory and practice, technical strategies have long been studied from both theoretical and empirical perspectives, resulting in a wealth of prior research. An outstanding review paper on their profitability is Park and Irwin [1].

Park and Irwin [1] analyzed existing empirical studies and classified them into two categories—early studies and contemporary studies—based on differences in their validation procedures. They further point out that in contemporary research, many of the analytical shortcomings observed in early studies have been corrected through the use of conventional statistical tests, sophisticated bootstrap methods, or a combination of both. Although these advances in testing methods have provided evidence supporting the profitability of technical strategies, skepticism regarding their profitability remains deeply entrenched in the academic community. Given this situation, Timmermann and Granger ([2] p. 16) propose key research topics for future study. These include 1) the introduction of promising stock time-series models and their estimation methods for predicting stock returns, and 2) the introduction of statistical methods for selecting the optimal stock time-series model.

Brock, Lakonishok, and LeBaron [3] (hereinafter referred to as BLL) were the first to adopt a model-based bootstrap method to statistically verify the profitability of technical strategies. A key advantage of this method is that, by selecting an appropriate stock time-series model, it is possible to generate a large number of stock price paths that take into account the return characteristics of financial assets—such as leptokurticity (high kurtosis), autocorrelation, conditional heteroscedasticity, and time-varying distribution properties—and to statistically verify the profitability of technical strategies by applying them to these paths.

BLL adopted four models as stock time-series models: the random walk, the first-order autoregressive process (AR(1)), the generalized autoregressive conditional heteroscedasticity model with mean (GARCH-M), and the exponential GARCH model (EGARCH). Subsequent research following BLL includes studies by Kho [4], Mills [5], Kwon and Kish [6], Saacke [7], and Hambuckers and Heuchenne [8], all of which primarily employ autoregressive moving-average (ARMA)-type or generalized autoregressive conditional heteroscedasticity (GARCH)-type stock time-series models and follow the verification framework proposed by BLL. For this reason, BLL can be considered one of the most influential studies in the field of modern technical strategy research.

Building on this prior research, the objective of this study is to incorporate promising stock time-series models not adopted by BLL into the analysis, while also refining BLL's validation methodology and proposing additional steps to obtain more precise implications. Specifically, after evaluating the validity of the stock time-series models, we add a procedure to compare the returns obtained by applying the same technical strategy to actual stock price paths and to stock price

paths generated by each model.

In this study, to examine stock time-series models that can more appropriately represent the characteristics of actual stock returns, we analyze a total of eight models: the four models adopted by BLL, plus four new models with structures more complex than those of linear models. The newly added models are the normal-threshold model (abbreviated as TH-norm), based on the threshold model introduced by Tong [9], and the AR(1)-threshold model (abbreviated as TH-ar), as well as the normal-hidden Markov switching model (abbreviated as HMS-norm) and the AR(1)-hidden Markov switching model (abbreviated as HMS-ar), both based on the hidden Markov switching model introduced by Hamilton [10].

As mentioned earlier, when evaluating the validity of stock time-series models that were not examined by BLL, we analyze them from two perspectives: "To what extent do the characteristics of stock returns generated by the stock time-series model accurately represent the characteristics of actual stock returns?" and "To what extent are the two distributions similar?" We introduce the Akaike Information Criterion (AIC) and Schwarz Information Criterion (BIC) as evaluation criteria for the former, and the Wasserstein distance as an evaluation criterion for the latter. Using information criteria, we evaluate the extent to which each model appropriately captures the characteristics of actual stock returns, while taking model complexity into account. On the other hand, the Wasserstein distance is used to measure the distance between the distributions of stock returns generated from each model via a model-based bootstrap method and the distribution of actual stock returns, thereby evaluating the similarity between the two. Furthermore, the consistency between the model rankings based on the information criteria and the ranking based on the Wasserstein distance is verified using Kendall's rank correlation coefficient.

Based on these verifications, we finally apply the same technical strategy to the actual stock price path and the stock price paths generated by each model, and verify using in-sample data where the returns obtained from the actual stock price path fall within the distribution of returns obtained from each generated path.

The structure of this paper is as follows. Section 2 reviews BLL and clarifies the research questions regarding BLL addressed in this study. Section 3 introduces the eight stock time-series models and six technical strategies to be analyzed. In Section 4, we present the analytical framework after reviewing the bootstrap method, Wasserstein distance, and Kendall's rank correlation coefficient, which play important roles in the empirical analysis. In Section 5, we present the data and analytical settings used as examples in the empirical analysis, followed by the results and our discussion. The final section presents the conclusion.

2. Review of Previous Research (BLL), Research Questions, and Approach

2.1. Review of Previous Research (BLL)

BLL examines the effectiveness of a simple moving average strategy—a technical

strategy that defines a buy signal when the short-term moving average exceeds the long-term moving average by a specified bandwidth, and a sell signal when it falls below it—using daily data from the Dow Jones Industrial Average (hereinafter referred to as the Dow) from 1897 to 1986, based on statistical methods utilizing a stock time-series model.

First, they examined two types of strategies: a variable-length moving average strategy (abbreviated as VMA) that classifies each trading day as either a buy signal period or a sell signal period based on the relative positions of the short-term and long-term moving averages, and a fixed-length moving average strategy (abbreviated as FMA) that establishes a position on the trading day following the occurrence of a buy or sell signal and closes the position after holding it for 10 trading days. For each strategy, they employed 10 different combinations of short-term and long-term moving average periods and bandwidths.

They then conducted two-tailed tests for each strategy to determine whether the average daily returns during the buy signal periods and sell signal periods under the VMA strategy differed significantly from the average daily returns over the entire period. As a result, the null hypothesis that the averages were equal was rejected at the 5% significance level for 6 out of 10 strategies in the buy signal period and for all 10 strategies in the sell signal period. Similar results were obtained for the FMA, suggesting the effectiveness of moving average strategies.

Second, they proposed a framework for comparing the effectiveness of various technical strategies and conducted an empirical analysis. In the framework, they first introduced a stock time-series model and estimated the parameters of each model using Dow Jones Industrial Average return data. Next, using the bootstrap method (see Section 4.1), they generated 500 stock return series—that is, 500 stock price paths—from each model.

They applied the technical strategies to each of the 500 stock price paths generated from each model and calculated the mean and standard deviation of the resulting daily strategy returns, separately for buy signal periods and sell signal periods. Hereinafter, these are referred to as the "model mean" and "model standard deviation", respectively.

They also applied the same technical strategy to the actual Dow Jones Industrial Average stock price path and calculated the mean and standard deviation of the resulting daily strategy returns, separately for buy signal periods and sell signal periods. Hereinafter, these are referred to as the "actual mean" and "actual standard deviation," respectively.

Based on this, for each of the 500 stock price paths generated by each model, they calculated the proportion of paths where the model's mean exceeds the actual mean and the proportion of paths where the model's standard deviation exceeds the actual standard deviation, separately for buy signals and sell signals, to verify the effectiveness of each strategy.

For stock time-series models, they adopted four models: the random walk, AR(1), GARCH-M, and EGARCH. This model selection takes into account the

return characteristics of financial assets mentioned earlier. The results of the empirical analysis show that, compared to the random walk and AR(1) models, for the GARCH-M and EGARCH models, a certain proportion of stock price paths exhibited model averages that exceeded the actual averages for buy signals, and the proportion of stock price paths where the model standard deviation exceeded the actual standard deviation also increased. However, the model averages obtained by applying technical strategies often fell below the actual averages, and the model standard deviations also differed significantly from the actual standard deviations.

Based on the results of the above empirical analysis, BLL draws the following three implications.

1) The process of generating stock returns is more complex than that captured by the four models described above.

2) Technical strategies are likely to extract some hidden patterns.

3) Applying more sophisticated technical strategies may result in a greater discrepancy between actual stock price paths and those generated by the model in terms of strategy returns.

2.2. Issues with BLL and the Approach of This Study

What makes BLL's work compelling is that, unlike earlier studies that focused solely on actual stock price time-series data, it applies technical strategies not only to actual stock price paths but also to numerous stock price paths generated from stock time-series models, thereby comparing and evaluating their effectiveness. However, there appears to be a gap in BLL's analytical methodology that warrants further examination regarding the relationship between the validation of technical strategy effectiveness and the predictive power of stock time-series models. This issue is deeply related to two implications highlighted by BLL: "the process of generating stock returns is more complex than the four models described above" and "technical strategies are likely to extract some hidden patterns."

As discussed in the second part of Section 2.1, in BLL, technical strategies are applied to the 500 stock price paths generated from each stock time-series model, and strategies are compared by calculating the proportion of cases in which the mean and standard deviation of the resulting daily returns exceed the values obtained when the same strategy is applied to actual stock price paths. However, for this comparison to be valid, the stock time-series models must adequately represent the actual stock price process. Otherwise, it is impossible to distinguish whether the differences between the actual stock price paths and the generated paths stem from intrinsic hidden patterns extracted by the technical strategies or from the insufficient predictive power of the stock time-series models; as a result, the effectiveness of each technical strategy would be compared while the effects of both factors remain intermingled.

Based on the issues outlined above, this study adopts the following approach. First, to verify whether the stock return series generated by the stock time-series

models can adequately reproduce the characteristics of actual stock returns, using in-sample data, we analyze a total of eight models—the four models adopted by BLL plus four additional models (see Section 3.1 for details). The verification procedure is as follows.

1) For these eight models, the parameters of each model are estimated using the maximum likelihood method based on actual stock return time-series data.

2) We rank each model based on the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (BIC).

3) Using the bootstrap method, we measure the distance between the distribution of stock returns generated by each stock time-series model and the distribution of actual stock returns using the Wasserstein distance. We rank the models based on the Wasserstein distance: the smaller the distance, the more accurately the model's generated distribution of stock returns is deemed to replicate the distribution of actual stock returns.

4) We verify the consistency between the ranking of stock time-series models based on the information criteria and the ranking based on the Wasserstein distance using Kendall's rank correlation coefficient.

In this study, we use the information criteria and Wasserstein distance to select stock time-series models that exhibit a high degree of fit to the actual data and whose generated return distributions closely resemble the actual return distributions. By doing so, we aim to mitigate the impact of insufficient model reproducibility on the verification results and strengthen our analytical methodology to enable a more precise assessment of whether technical strategies capture the patterns inherent in stock return data.

3. Time-Series Models of Stock Returns and Technical Strategies

3.1. Time-Series Models of Stock Returns

BLL adopted the random walk (normal) model (abbreviated as "Normal"), the AR(1) model (abbreviated as "AR"), the mean GARCH model (abbreviated as "GARCH-M"), and the exponential GARCH model (abbreviated as "EGARCH"). We add the following four models to these: the normal-threshold model (abbreviated as TH-norm), based on the threshold model framework introduced by Tong [9], and the AR(1)-threshold model (abbreviated as TH-ar), as well as the normal-hidden Markov switching model (abbreviated as HMS-norm) and the AR(1)-hidden Markov switching model (abbreviated as HMS-ar), based on the hidden Markov switching model framework introduced by Hamilton [10], bringing the total number of models analyzed to eight.

Random Walk (Normal) Model (Normal)

$$r_t = \mu + \varepsilon_t, \varepsilon_t \sim W.N.\left(\sigma^2\right). \tag{1}$$

Here, ε_t is white noise with variance σ^2, and μ is the drift parameter.

AR(1) Model (AR)

$$r_t = c + \varphi r_{t-1} + \varepsilon_t, |\varphi| < 1, \varepsilon_t \sim W.N.\left(\sigma^2\right). \tag{2}$$

Here, ε_t is white noise with variance σ^2, and (c, φ) are parameters.

Mean GARCH Model (GARCH-M)

The GARCH-M model is a generalization of the conditional variance process in the autoregressive conditional heteroscedasticity-in-mean (ARCH-M) model introduced by Engle, Lilien, and Robins [11] to the GARCH framework proposed by Bollerslev [12].

$$r_t = a + \gamma h_t + b\varepsilon_{t-1} + \varepsilon_t, \tag{3}$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{4}$$

$$\varepsilon_t = h_t^{1/2} z_t, z_t \sim W.N.(1). \tag{5}$$

Here, z_t represents white noise with a variance of 1, and $(a, \gamma, b, \alpha_0, \alpha_1, \beta)$ are parameters. h_t is the conditional variance, which is expressed as a linear combination of the squared error term from the previous period and the conditional variance from the previous period. In other words, by introducing positive time-series correlation into h_t, the model captures the phenomenon where high (low) volatility is followed by high (low) volatility.

Exponential GARCH Model (EGARCH)

$$r_t = a + \gamma \mathrm{e}^{h_t} + b\varepsilon_{t-1} + \varepsilon_t, \tag{6}$$

$$h_t = \alpha_0 + g\left(z_{t-1}\right) + \beta h_{t-1}, \tag{7}$$

$$g\left(z_t\right) = \theta z_t + \omega\left(\left|z_t\right| - \left(2/\pi\right)^{1/2}\right), \tag{8}$$

$$\varepsilon_t = \mathrm{e}^{1/2^{h_t}} z_t, z_t \sim W.N.(1). \tag{9}$$

Here, z_t is white noise with a variance of 1, and $(a, \gamma, b, \alpha_0, \theta, \omega, \beta)$ are parameters. Similar to the GARCH model, it can capture the persistence of volatility; furthermore, since the logarithm of the conditional variance follows an autoregressive process, it can asymmetrically capture the influence of the previous period's standardized error on the current period's volatility depending on its sign. This captures the phenomenon noted by Black [13], in which volatility is higher following a negative return than following a positive return. For the formulation of the EGARCH model, see Nelson [14].

Normal-Threshold Model (TH-norm)

A threshold model is a model in which the state variable S_t is defined as a deterministic function of the observed variable r_{t-1}; the type of threshold model is determined by the combination of this function and the model specifications used within each regime.

The Normal-Threshold Model (TH-norm) is a model that uses the Random Walk (Normal) model as the model specification within each regime, with its parameters varying according to the regime $S_t = i$ $(i = 1, 2)$ at each time step t.

$$r_t = \mu^{(S_t)} + \varepsilon_t, \varepsilon_t \sim W.N.\left(\left(\sigma^{(S_t)}\right)^2\right), \tag{10}$$

$$S_t = \begin{cases} 1, & r_{t-1} < \tau, \\ 2, & \tau \le r_{t-1}. \end{cases} \tag{11}$$

Here, ε_t represents white noise with variance $\left(\sigma^{(S_t)}\right)^2$, $\mu^{(S_t)}$ is the drift parameter, and τ is the threshold. $\mu^{(S_t)}$ and $\sigma^{(S_t)}$ vary depending on the regime $S_t = i$ $(i = 1, 2)$.

AR(1)-Threshold Model (TH-ar)

The AR(1)-threshold model (TH-ar) adopts an AR model as the model specification within each regime, and its parameters depend on the regime $S_t = i$ $(i = 1, 2)$ at each time step t.

$$r_t = c^{(S_t)} + \varphi^{(S_t)} r_{t-1} + \varepsilon_t, \left|\varphi^{(S_t)}\right| < 1, \varepsilon_t \sim W.N.\left(\left(\sigma^{(S_t)}\right)^2\right), \tag{12}$$

$$S_t = \begin{cases} 1, & r_{t-1} < \tau, \\ 2, & \tau \le r_{t-1}. \end{cases} \tag{13}$$

Here, ε_t is white noise with variance $\left(\sigma^{(S_t)}\right)^2$, $\left(c^{(S_t)}, \varphi^{(S_t)}\right)$ are parameters, and τ is a threshold. $c^{(S_t)}, \varphi^{(S_t)}$ and $\sigma^{(S_t)}$ are determined according to the regime $S_t = i$ $(i = 1, 2)$.

Next, we introduce two stock time-series models that utilize the Hidden Markov Regime-Switching Model.

Hamilton [10] introduced the Hidden Markov Regime-Switching Model (hereinafter abbreviated as the HMS model) into time-series analysis in financial economics. In the HMS model, the state variable representing the regime at each time point is not observed; rather than being defined as a deterministic function of the observed variables, as in a threshold model, it transitions probabilistically according to a Markov chain. This makes it possible to capture regime shifts—such as those observed in business cycles—stochastically from observed data without directly observing the regimes themselves. In the HMS model, S_t represents the latent regime (state) at time t. The conditional probability density of the observed value r_t $(t = 1, \cdots, T)$, given the latent state S_t, is expressed as

$$f\left(r_t \mid S_t = k, \Theta_k\right). \tag{14}$$

Here, $f\left(r_t \mid S_t = k, \Theta_k\right)$ represents the conditional probability density of the observation r_t given the hidden state $S_t = k$, and Θ_k represents the model parameters at state $S_t = k$ $(k = 1, 2)$.

The initial probability

$$\pi_k = \Pr\left(S_1 = k\right), k = 1, 2 \tag{15}$$

represents the probability that the model starts from state k.

In a two-state HMS model ($K = 2$), two regimes with different characteristics—such as rising and falling price phases—are typically assumed. State transitions in a Markov chain are defined by the transition matrix $P = \left(p_{ij}\right)$, and in a two-state model, the transition matrix $P_{2\times 2}$ is given as follows:

$$P_{2\times 2} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \tag{16}$$

Here, the probability of transitioning from state i to state j is defined by the following equation.

$$p_{ij} = \Pr\left(S_t = j \mid S_{t-1} = i, S_{t-2} = k, \cdots, r_{t-1}, r_{t-2}, \cdots\right) = \Pr\left(S_t = j \mid S_{t-1} = i\right). \quad (17)$$

Normal-Hidden Markov Switching Model (HMS-norm)

The Normal-Hidden Markov Switching Model (HMS-norm) is a model that adopts a random walk (Normal) model for the model specification within each regime, with its parameters varying according to the latent regime $S_t = i \ (i = 1, 2)$.

$$r_t = \mu^{(S_t)} + \varepsilon_t, \varepsilon_t \sim W.N.\left(\left(\sigma^{(S_t)}\right)^2\right). \quad (18)$$

Here, ε_t represents white noise with variance $\left(\sigma^{(S_t)}\right)^2$, and $\mu^{(S_t)}$ is a parameter.

AR(1)-Hidden Markov Switching Model (HMS-ar)

The AR(1)-Hidden Markov Switching Model (HMS-ar) is a model that adopts the AR(1) model (AR) as the model specification within each regime, and whose parameters vary according to the latent regime $S_t = i \ (i = 1, 2)$.

$$r_t = c^{(S_t)} + \varphi^{(S_t)} r_{t-1} + \varepsilon_t, \left|\varphi^{(S_t)}\right| < 1, \varepsilon_t \sim W.N.\left(\left(\sigma^{(S_t)}\right)^2\right). \quad (19)$$

Here, ε_t is white noise with variance $\left(\sigma^{(S_t)}\right)^2$, and $\left(c^{(S_t)}, \varphi^{(S_t)}\right)$ are parameters.

3.2. Technical Strategies

As mentioned in Section 2, BLL adopts the VMA and FMA moving average strategies as the technical strategies under analysis. In this study, we conducted an empirical analysis focusing on the buy signals of the VMA strategy adopted by BLL, both to propose a method that complements the aspects overlooked by BLL and to compare our results with those of BLL. Furthermore, the reason for focusing on buy signals is to compare the returns during the buy signal period with the returns of a general buy-and-hold strategy (a strategy in which the entire period is treated as a long position period).

In the VMA strategy, the period during which the short-term moving average exceeds the long-term moving average is considered a period of an upward market trend and is designated as a buy signal period. Moving averages are used to smooth price series and identify market trends; however, when the short-term and long-term moving averages are close together, even slight price fluctuations can easily cause "false signals," where buy signals appear and disappear repeatedly within a short period. Therefore, to mitigate this effect, bands of a specified percentage are set above and below the long-term moving average, and only the periods during which the short-term moving average exceeds the upper band limit are designated as buy signal periods. In this case, the number of trading days that generate buy signals decreases compared to when no bands are set.

In this study, the short-term moving average period is set to 1 business day, and

the long-term moving average periods are set to 50, 150, and 200 business days. For each combination, we examine the results when no bands are set versus when bands are set at ±1% above and below the long-term moving average. Each strategy is denoted as (short-term moving average period, long-term moving average period, bandwidth), and we adopt a total of six VMA strategies: (1, 50, 0), (1, 50, 0.01), (1, 150, 0), (1, 150, 0.01), (1, 200, 0), and (1, 200, 0.01). Furthermore, signals determined based on the closing price on each trading day t are applied to the daily return on the following trading day $t+1$.

4. Statistical Methods, Analytical Framework, and Procedures

4.1. Bootstrap Method

Based on the stock time-series model, the procedure for generating multiple paths of stock returns using the bootstrap method is outlined in Steps 1 through 4.

Step 1 Convert the daily closing price data P_t into log returns r_t using the following equation.

$$r_t = \ln\left(P_t / P_{t-1}\right) \tag{20}$$

Step 2 Estimate the parameters of the stock time-series model introduced in Section 3.1 using the maximum likelihood method.

Step 3 Assuming that the error terms for each model follow a standard normal distribution, generate the required number of sequences of standard normal random numbers for validation (parametric bootstrap method).

Step 4 Based on the generated standard normal random numbers and the estimated parameters of each model, generate paths of log returns for stocks from each model and, if necessary, convert them to stock price paths using exponential transformation.

4.2. Wasserstein Distance

The Wasserstein distance is a measure of the distance between two probability distributions based on the concept of optimal transport. Specifically, it is defined as the "minimum transport cost" required to move from one probability distribution to another. Since it takes into account the distances between the values that a random variable can take, it is characterized by its ability to evaluate differences in the overall location and shape of distributions that cannot be fully captured by mean or variance alone. For the definition of the Wasserstein distance, see Vallender [15]; for the general theory of optimal transport, see Villani [16].

Definition

Let P and Q be probability distributions, and let their cumulative distribution functions (CDFs) be $F_P(x)$ and $F_Q(x)$, respectively. Then, the Wasserstein distance $W(P,Q)$ is defined as

$$W(P,Q) = \int_0^1 \left| F_P^{-1}(u) - F_Q^{-1}(u) \right| \mathrm{d}u. \tag{21}$$

Here, $F_P^{-1}(u)$ and $F_Q^{-1}(u)$ are the inverse CDFs (quantile functions) of

$F_P(x)$ and $F_Q(x)$, respectively. This is the integral, over all quantile levels, of the absolute difference between the quantiles of the two distributions at the same quantile level. Therefore, the Wasserstein distance captures the differences in the location and shape of the two distributions; the smaller the distance, the more similar the two distributions are, and when the distance is zero, the two distributions are identical.

4.3. Kendall's Rank Correlation Coefficient

In this study, we use Kendall's rank correlation coefficient, as proposed by Kendall [17], to verify the consistency between two ranking systems. This coefficient is calculated from n pairs of bivariate data

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n). \tag{22}$$

For any $i < j$, we focus on whether the order relations between x_i and x_j match those between y_i and y_j, or whether they are reversed. We then standardize the difference between the number of pairs where the order relations match and the number of pairs where they are reversed, such that the minimum value is -1 and the maximum value is 1, similar to the standard correlation coefficient.

Let $I(\cdot)$ be an indicator function—that is, a function that returns 1 if the condition in parentheses holds and 0 otherwise—and define τ as follows:

$$\tau = \frac{\sum_{i<j} I\left(\left(x_i - x_j\right)\left(y_i - y_j\right) > 0\right) - \sum_{i<j} I\left(\left(x_i - x_j\right)\left(y_i - y_j\right) < 0\right)}{\sqrt{\sum_{i<j} I\left(x_i \neq x_j\right) \times \sum_{i<j} I\left(y_i \neq y_j\right)}}. \tag{23}$$

Kendall's tau-b (τ_b) is a version of Kendall's rank correlation coefficient that adjusts the denominator for ties (equal ranks) in either ranking. In this paper, this coefficient is denoted by τ and is defined by Equation (23).

When testing the null hypothesis that there is no rank correlation in the population, it is more convenient to consider the numerator of τ:

$$K = \sum_{i<j} I\left(\left(x_i - x_j\right)\left(y_i - y_j\right) > 0\right) - \sum_{i<j} I\left(\left(x_i - x_j\right)\left(y_i - y_j\right) < 0\right). \tag{24}$$

In this study, when neither the observations $x_1, x_2, \cdots, x_n$ nor the observations $y_1, y_2, \cdots, y_n$ contain duplicate values, the probability distribution of K under the null hypothesis is approximated by a normal distribution with
expected value

$$E(K) = 0 \tag{25}$$

variance

$$V(K) = \frac{n(n-1)(2n+5)}{18}. \tag{26}$$

Under the one-sided alternative hypothesis that the rank correlation is positive, the one-sided p-value (significance probability) is given by

$$p = 1 - \Phi\left(\frac{K - E(K) - 1}{\sqrt{V(K)}}\right). \tag{27}$$

Φ is the cumulative distribution function of the standard normal distribution;

the subtraction of 1 from the numerator is a continuity correction.

4.4. Analytical Framework and Procedures

The analytical framework is based on BLL. However, when GARCH-M and EGARCH models are estimated using standard maximum likelihood estimation, the mean and variance of the generated stock returns may deviate from their observed values, potentially failing to adequately reproduce the risk-return characteristics of actual stock returns. In such cases, even if technical strategies are applied to the generated stock price paths, the results will be influenced by the model's insufficient replicability, leading to the problem that returns attributable to the technical strategies cannot be properly extracted. Therefore, in this study, when estimating GARCH-M and EGARCH, we use the maximum likelihood method under the constraint that the mean and standard deviation of stock returns during the analysis period match the actual values, and conduct the analysis according to the following procedures.

Here, the moment restrictions imposed on GARCH-M and EGARCH apply to the model-implied theoretical unconditional moments, rather than to the mean and variance calculated ex post from a finite simulated sample. Let $\bar{r}$, s, and s^2 denote the sample mean, sample standard deviation, and sample variance of the observed returns, respectively. For GARCH-M, the restrictions $E(r_t)=\bar{r}$ and $E(h_t)=s^2$ are imposed, and the following constrained conditional Gaussian log-likelihood is maximized.

$$\max_{a,\gamma,b,\alpha_0,\alpha_1,\beta} \ell_{\text{GARCH-M}}$$

$$\ell_{\text{GARCH-M}} = -\frac{1}{2}\sum_{t=1}^{n}\left[\log(2\pi)+\log h_t+\frac{\varepsilon_t^2}{h_t}\right]$$

For EGARCH, the restrictions $E(r_t)=\bar{r}$ and $E\left(\mathrm{e}^{h_t}\right)=s^2$ are imposed, and the following constrained conditional Gaussian log-likelihood is maximized.

$$\max_{a,\gamma,b,\alpha_0,\theta,\omega,\beta} \ell_{\text{EGARCH}}$$

$$\ell_{\text{EGARCH}} = -\frac{1}{2}\sum_{t=1}^{n}\left[\log(2\pi)+h_t+\frac{\varepsilon_t^2}{\mathrm{e}^{h_t}}\right]$$

where

$$\varepsilon_t = r_t - a - \gamma \mathrm{e}^{h_t} - b\varepsilon_{t-1}$$

This log-likelihood is maximized subject to $E(r_t)=\bar{r}, E\left(\mathrm{e}^{h_t}\right)=s^2$, and $|\beta|<1$. AIC and BIC are calculated using the maximized value of the constrained likelihood and the total number of parameters actually estimated under the moment restrictions for each model. In the simulations, standard normal random numbers corresponding to the error terms are generated; however, no ex post restandardization is applied to force the sample mean and sample standard deviation of each finite simulated series to match the observed values.

Procedure 1 For the eight stock time-series models described in Section 3.1, we estimate the parameters using ordinary least squares (OLS) or the maximum likelihood method under the aforementioned constraints and calculate the residuals. Next, based on information criteria (AIC and BIC), we evaluate the goodness of fit of each stock time-series model to actual stock returns and rank them.

Procedure 2 Using the estimated parameters and standard normal random numbers corresponding to the error terms of each model, we generate 10,000 stock return series from each stock time-series model. Furthermore, based on the Wasserstein distance measured between the distribution of actual stock returns and the distribution of stock returns generated from each stock time-series model, we evaluate the goodness of fit of each stock time-series model to the actual stock returns and rank the models from smallest to largest Wasserstein distance.

The bootstrap implementation for Procedure 2 is as follows. For each model, 10,000 paths are generated, and after a burn-in period of 500 observations, 2448 observations, corresponding to the length of the observed sample, are retained. For initialization, no initial value is required for the Normal model because it contains no recursive structure; the unconditional mean of the model is used for AR; $h_0 = s^2$ and $\varepsilon_0 = 0$ are used for GARCH-M; $h_0 = \log\left(s^2\right)$ and $z_0 = 0$ are used for EGARCH, where h_t denotes the logarithm of the conditional variance; and $r_0 = \bar{r}$ is used for the threshold models. For the HMS models, the initial state is set to state 1, and subsequent states are updated according to the estimated transition probabilities. For HMS-ar, the initial return is set to $r_0 = \bar{r}$. A different fixed seed is used for each of the eight models. The Wasserstein distance is calculated between the distribution obtained by pooling the 2448 observations from each of the 10,000 paths for a given model and the empirical distribution of the 2448 observed returns, using quantile differences at the midpoints of 2500 quantile intervals.

Procedure 3 We verify the consistency between the rankings based on AIC and BIC and the ranking based on the Wasserstein distance using Kendall's rank correlation coefficient.

Procedure 4 We apply each technical strategy to the actual stock price data and calculate the average daily return during the buy signal period. Next, we apply the same technical strategy to the 10,000 stock price paths generated by each stock time-series model and construct an empirical distribution based on the 10,000 average daily returns. We then calculate the proportion of points in that empirical distribution that exceed the average value obtained from the actual stock price data.

Furthermore, we examine the extent to which this proportion varies depending on the combination of the stock time-series model and the technical strategy.

5. Examples of Empirical Analysis

5.1. Data and Settings

As an example of empirical analysis, we use daily closing price data for the Nikkei Stock Average (hereinafter abbreviated as the Nikkei 225) for the 2010s (January 2010 through December 2019). The trend is shown in Figure 1.



Source: Compiled by the authors based on Yahoo Finance, "Nikkei 225 (^N225) Historical Data" [18].

Figure 1. Trends in Nikkei 225 closing prices during the 2010s.

We will now review the trends in the Nikkei 225 closing prices during the analysis period shown in Figure 1.

- From 2010 to 2012, against the backdrop of yen appreciation, the Great East Japan Earthquake, and the European debt crisis, the Nikkei 225's returns were sluggish, and volatility remained at high levels.
- From 2013 to 2015, against the backdrop of Abenomics, "unconventional monetary easing," and a weak yen, the Nikkei 225's returns rose sharply, while volatility remained at moderate to high levels.
- From 2016 to 2018, against the backdrop of Brexit and U.S.-China trade tensions, the Nikkei 225's returns fluctuated widely, and volatility remained at high levels.
- In 2019, against the backdrop of monetary easing and a global economic recovery, the Nikkei 225 posted high returns, while volatility remained at moderate levels.

Next, as the technical strategies used in the analysis, we will examine the six VMA strategies described in Section 3.2: (1, 50, 0), (1, 50, 0.01), (1, 150, 0), (1, 150, 0.01), (1, 200, 0), and (1, 200, 0.01).

5.2. Results and Discussion

5.2.1. Basic Statistics and Parameter Estimates for the Eight Stock Time-Series Models

Table 1 shows the basic statistics and parameter estimates for the eight stock time-series models. The daily returns of the Nikkei 225 during the 2010s had a mean of 0.033% and a standard deviation of 1.308%; when annualized based on 252 trading days per year, the mean is approximately 8.3% and the standard deviation is approximately 20.8%. Furthermore, the skewness S is −0.5729, the kurtosis K is 8.5309, and the first-order autocorrelation coefficient ρ is −0.0395. Skewness and kurtosis exceed the practical thresholds for non-normality $|S| \geq 0.5$ and $K \geq 4$, respectively, confirming a certain degree of negative skewness and considerably high kurtosis—that is, a heavy-tailed distribution. Furthermore, since stock returns are considered "uncorrelated" in practice if the absolute value of the first-order autocorrelation coefficient ρ is below a practical threshold ranging from 0.05 to 0.10, the daily returns of the Nikkei 225 during the analysis period can also be judged to be uncorrelated in practice. Based on the above, while the differences between the Normal and AR models are small among the eight stock time-series models, it is considered that there are corresponding differences in the ability to reproduce the distribution characteristics of daily returns among the conditional variance model, the threshold model, and the HMS model. Below, we examine the parameter estimation results for each stock time-series model based on Table 1.

Table 1. Basic statistics and parameter estimates for stock time-series models.

Panel A. Basic Statistics					
Number of Observations	Mean	Standard Deviation	Skewness S	Kurtosis K	First-Order Autocorrelation ρ
2448	0.03300%	1.30750%	−0.57290	8.53090	−0.03950

Panel B. Parameter Estimates				
Stock Time-Series Model	Estimates			
Normal	$\mu = 0.00033$	$\sigma = 0.01308$		
AR	$c = 0.00034$	$\sigma = 0.01307$	$\varphi = -0.03945$	
GARCH-M	$a = 0.00033$	$\gamma = 0.00000$	$b = 0.00000$	
	$\alpha_0 = 0.00001$	$\alpha_1 = 0.12272$	$\beta = 0.83762$	
EGARCH	$a = 0.00033$	$\gamma = 0.00000$	$b = 0.00000$	$\beta = 0.94054$
	$\alpha_0 = -0.52379$	$\theta = -0.11429$	$\omega = 0.21011$	
TH-norm	$\mu_1 = 0.00301$	$\sigma_1 = 0.02200$	$\tau = -0.01959$	
	$\mu_2 = 0.00016$	$\sigma_2 = 0.01227$		
TH-ar	$c_1 = -0.00306$	$\sigma_1 = 0.02288$	$\varphi_1 = -0.16519$	$\tau = -0.02687$
	$c_2 = 0.00029$	$\sigma_2 = 0.01281$	$\varphi_2 = -0.00030$	
HMS-norm	$\mu_1 = 0.00112$	$\sigma_1 = 0.00870$	$p_{11} = 0.97998$	
	$\mu_2 = -0.00116$	$\sigma_2 = 0.01864$	$p_{22} = 0.96228$	
HMS-ar	$c_1 = 0.00114$	$\sigma_1 = 0.00869$	$\varphi_1 = -0.00608$	$p_{11} = 0.97985$
	$c_2 = -0.00125$	$\sigma_2 = 0.01859$	$\varphi_2 = -0.06626$	$p_{22} = 0.96202$

Note: Maximum likelihood estimation was performed under the constraint that the mean and standard deviation of each model match those of the actual data.

For GARCH-M and EGARCH, as described in Section 4.4, we estimated the parameters using the maximum likelihood method under the constraint that the mean and standard deviation of the stock returns generated by the model match the actual values. The parameters of the mean equations for both models are $a = 0.00033$, $\gamma = 0.00000$, and $b = 0.00000$. Therefore, the mean equations of both models effectively consist only of constant terms, and their estimated values match the actual mean stock returns $\mu = 0.00033$ during the analysis period. This indicates that, among the constraints imposed during the maximum likelihood estimation shown in Section 4.4, the constraint regarding the mean is reflected in the estimation results. Furthermore, according to the TH-norm estimation results, when the return one business day prior falls below the threshold $\tau = -0.01959$, both the drift and volatility of the return on the following business day are relatively high; when it exceeds this threshold, both are relatively low. The HMS-norm distinguishes between two states: one characterized by large positive drift and low volatility, and another characterized by large negative drift and high volatility. Furthermore, while both states exhibit high persistence, the probability of remaining in the former state is slightly higher.

5.2.2. Ranking of Stock Return Time-Series Models Based on Information Criteria (AIC and BIC)

Table 2 shows the values of the information criteria (AIC and BIC) for the eight stock time-series models and the ranking of these models based on these criteria. As shown in Table 2, the rankings based on AIC and BIC were in complete agreement. As discussed in Section 5.2.1, the difference between the Normal and AR models is small. The goodness of fit based on the information criteria, listed in ascending order, is as follows: the single-state AR model, the Normal model, the threshold-switching models TH-ar and TH-norm, the conditional variance model GARCH-M, the HMS models HMS-ar and HMS-norm, and the conditional variance model EGARCH.

Table 2. Ranking of stock time-series models based on information criteria.

Model Classification	Model Name	AIC	Rank	BIC	Rank
Single-State Model	Normal	−14,287	7	−14,287	7
	AR	−14,283	8	−14,277	8
Conditional Variance Model	GARCH-M	−14,722	4	−14,699	4
	EGARCH	−14,799	1	−14,770	1
Threshold-Switching Model	TH-norm	−14,414	5	−14,390	5
	TH-ar	−14,380	6	−14,345	6
Hidden Markov Switching Model (HMS Model)	HMS-norm	−14,760	2	−14,737	2
	HMS-ar	−14,754	3	−14,719	3

5.2.3. Ranking of Stock Return Time-Series Models Based on Wasserstein Distance

Table 3 shows the Wasserstein distance values for the eight stock time-series models and the ranking of these models based on these values. As shown in Table 3, the Wasserstein distance increases in the following order: HMS-ar, HMS-norm, GARCH-M, EGARCH, TH-norm, TH-ar, Normal, and AR. Accordingly, their ability to reproduce the return distribution of the Nikkei 225 is ranked in the same order, from highest to lowest. As noted in Section 5.2.1, the difference between the Normal and AR models is small. Although the Wasserstein distances for the Normal and AR models are both 0.00173 when rounded to five decimal places, the values reported to seven decimal places are 0.0017303 and 0.0017309, respectively, with the Normal model having the slightly smaller value. Accordingly, the Normal model is ranked seventh and the AR model eighth. On the other hand, when viewed by model category, the ability to reproduce the distribution increases in the following order: single-state models, threshold-switching models, conditional variance models, and HMS models.

Table 3. Ranking of stock time-series models based on Wasserstein distance.

Model Category	Model Name	Wasserstein Distance	Rank
Single-State Model	Normal	0.0017303	7
	AR	0.0017309	8
Conditional Variance Model	GARCH-M	0.0010742	3
	EGARCH	0.0011686	4
Threshold-Switching Model	TH-norm	0.0014720	5
	TH-ar	0.0016369	6
Hidden Markov Switching Model (HMS Model)	HMS-norm	0.0006235	2
	HMS-ar	0.0006134	1

5.2.4. Consistency in the Ranking of Stock Time-Series Models Based on Information Criteria (AIC and BIC) and Wasserstein Distance

We used Kendall's rank correlation coefficient to verify the consistency between the rankings of stock time-series models based on information criteria (AIC and BIC) and the ranking based on the Wasserstein distance. The results showed that Kendall's rank correlation coefficient was 0.7143—a fairly high value—indicating that the two rankings were consistent, thereby supporting the reliability of the model selection results. The one-sided and two-sided p-values were 0.0071 and 0.0141, respectively.

To assess variability due to random-number generation, the 10,000 paths generated for each of the eight models were divided into 20 groups of 500 paths, and the Wasserstein distance from the observed-data distribution was calculated for each group. For each model, 20 distances were then drawn with replacement from the 20 group-level distances and averaged. The eight models were ranked on the basis of these mean distances, and Kendall's τ between this ranking and the

AIC/BIC ranking was calculated. This resampling procedure was repeated 10,000 times. HMS-ar ranked first in 97.1% of the replications and HMS-norm in 2.9%. The median Kendall's τ was 0.7143, and the 95% interval was [0.6429, 0.7143].

Figure 2 compares the actual Nikkei 225 return distribution with the return distributions generated using the bootstrap method from the Normal, EGARCH, and HMS-norm stock time-series models. As confirmed in Section 5.2.1, the actual return distribution exhibits extremely high kurtosis. Therefore, stock time-series models that can flexibly capture the characteristics of the actual Nikkei 225 return distribution—such as high kurtosis and fat tails—should be evaluated as having superior ability to reproduce the distribution and thus be selected. Indeed, Figure 2 confirms that the return distributions generated by the stock time-series models—in the order of Normal, EGARCH, and HMS-norm—closely approximate the actual return distribution.

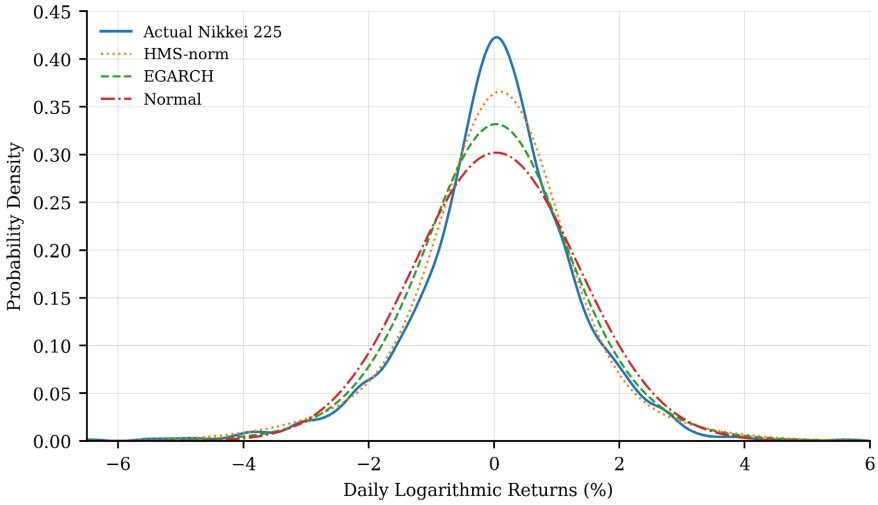


Figure 2. Comparison of Actual Nikkei 225 Returns and Model-Generated Return Distributions.

Based on the above, it has been demonstrated that by comprehensively evaluating the goodness of fit and distribution reproducibility relative to actual Nikkei 225 returns through Procedures 1 - 3 proposed in Section 4.4, it is possible to select a relatively superior model from among the eight stock time-series models examined in this study.

5.2.5. Implications Derived from the Validation of Technical Strategies Using the Bootstrap Method

Finally, following Procedure 4, we examine the implications derived from a bootstrap-based validation focused on the buy signal period of the variable-length moving average (VMA) strategy among the technical strategies.

Table 4 presents the results of applying the six VMA strategies (buy signals only) described in Section 5.1 to actual Nikkei 225 data and to stock price paths generated from each stock time-series model. The leftmost column of values shows the average daily returns obtained by applying each VMA to the actual Nik-

kei 225 data. The columns to the right show, from left to right, the results of the bootstrap method validation for the following stock time-series models: Normal, AR, GARCH-M, EGARCH, TH-norm, TH-ar, HMS-norm, and HMS-ar.

Specifically, 10,000 stock price paths were generated from each stock time-series model, and the same VMA strategy used on the actual Nikkei 225 data was applied to each of these stock price paths. Next, for each stock price path, we calculated the average daily return during the buy signal period and determined the proportion of the 10,000 averages obtained from the models and strategies that exceeded the average obtained by applying the strategy to the actual Nikkei 225 data.

Table 4. Percentage of cases where the model's average strategy return exceeds the Nikkei 225's average strategy return.

Strategy	Average Strategy Return	Percentage of Cases Where the Model's Average Strategy Return Exceeds the Nikkei 225's Average Strategy Return							
	Nikkei 225	Normal	AR	GARCH-M	EGARCH	TH-norm	TH-ar	HMS-norm	HMS-ar
(1, 50, 0)	0.0129%	70%	61%	70%	72%	64%	68%	86%	81%
(1, 50, 0.01)	0.0208%	61%	51%	61%	64%	54%	59%	79%	73%
(1, 150, 0)	0.0323%	49%	44%	49%	49%	45%	48%	63%	59%
(1, 150, 0.01)	0.0353%	46%	40%	45%	45%	41%	44%	60%	55%
(1, 200, 0)	0.0225%	59%	55%	60%	60%	56%	58%	71%	68%
(1, 200, 0.01)	0.0277%	54%	49%	53%	53%	49%	52%	66%	62%

First, examining the average daily returns during the VMA buy signal periods, Strategy (1, 150, 0) and Strategy (1, 150, 0.01) yielded returns of 0.032% and 0.035%, respectively, which were comparable to the average daily return of 0.033% for the buy-and-hold strategy (see the basic statistics in Table 1). On the other hand, the other four strategies all fell significantly below the average daily return of the buy-and-hold strategy. This suggests that, for the Nikkei 225 in the 2010s, unlike the results of the BLL empirical analysis targeting the long-term Dow Jones Industrial Average, we were unable to confirm the notable effectiveness of the VMA even in the in-sample data.

However, based on future research using out-of-sample data, if selection criteria for the "optimal model" regarding the Nikkei 225 return model are established, and the HMS-norm is selected from among those with high AIC/BIC rankings or Wasserstein rankings, it can be interpreted that for strategy (1, 150, 0.01), the probability that the average daily return during the VMA buy signal period exceeds the buy-and-hold strategy's average daily return of 0.033% is 60% or higher.

6. Summary and Future Research Directions

This study reviewed BLL—one of the most influential prior studies on the validation of technical strategies using the bootstrap method—and clarified the research objectives by identifying remaining challenges in its analytical methodology.

Building on this, we introduced a procedure that incorporates promising stock time-series models not used in BLL, pre-evaluates the predictive power of each model, and verifies the effectiveness of technical strategies based on that evaluation. Through this, we proposed a refinement to complement BLL's verification methodology and obtain more precise implications. Furthermore, through examples of empirical analysis, we confirmed how the proposed methodology functions and what implications can be derived from it.

While this study analyzed only in-sample data, we will use out-of-sample data to verify the validity of the selected stock time-series model and the effectiveness of the technical strategy.

In doing so, we would like to establish predefined rules for combining AIC/BIC and the Wasserstein distance in order to select the optimal model. Furthermore, regarding the verification of model validity, we would like to consider: 1) comparisons of autocorrelation in returns and the squares of returns, tail quantiles, and the persistence of volatility clusters; and 2) the introduction of a heavy-tailed innovation benchmark to verify whether the superiority of hidden Markov switching reflects regime dynamics rather than the constraints of Gaussian innovations.

Disclaimer

The content and opinions expressed in this paper do not represent the official views of the authors' affiliated institutions.

Acknowledgements

We deeply thank an anonymous reviewer for his/her valuable comments and suggestions to improve the initial version of this paper.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Park, C.H. and Irwin, S.H. (2004) The Profitability of Technical Analysis: A Review. AgMAS Project Research Report No. 2004-04, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign.

[2] Timmermann, A. and Granger, C.W.J. (2004) Efficient Market Hypothesis and Forecasting. *International Journal of Forecasting*, **20**, 15-27. https://doi.org/10.1016/s0169-2070(03)00012-8

[3] Brock, W., Lakonishok, J. and LeBaron, B. (1992) Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. *The Journal of Finance*, **47**, 1731-1764. https://doi.org/10.1111/j.1540-6261.1992.tb04681.x

[4] Kho, B.C. (1996) Time-Varying Risk Premia, Volatility, and Technical Trading Rule Profits: Evidence from Foreign Currency Futures Markets. *Journal of Financial Economics*, **41**, 249-290. https://doi.org/10.1016/0304-405x(95)00861-8

[5] Mills, T.C. (1997) Technical Analysis and the London Stock Exchange: Testing Trading Rules Using the FT30. *International Journal of Finance & Economics*, **2**, 319-331.

https://doi.org/10.1002/(sici)1099-1158(199710)2:4<319::aid-jfe53>3.0.co;2-6

[6] Kwon, K.Y. and Kish, R.J. (2002) Technical Trading Strategies and Return Predictability: NYSE. *Applied Financial Economics*, **12**, 639-653. https://doi.org/10.1080/09603100010016139

[7] Saacke, P. (2002) Technical Analysis and the Effectiveness of Central Bank Intervention. *Journal of International Money and Finance*, **21**, 459-479. https://doi.org/10.1016/s0261-5606(02)00009-8

[8] Hambuckers, J. and Heuchenne, C. (2016) Estimating the Out-of-Sample Predictive Ability of Trading Rules: A Robust Bootstrap Approach. *Journal of Forecasting*, **35**, 347-372. https://doi.org/10.1002/for.2380

[9] Tong, H. (1978) On a Threshold Model. In: Chen, C.H., Ed., *Pattern Recognition and Signal Processing*, Springer, 575-586. https://doi.org/10.1007/978-94-009-9941-1_24

[10] Hamilton, J.D. (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, **57**, 357-384. https://doi.org/10.2307/1912559

[11] Engle, R.F., Lilien, D.M. and Robins, R.P. (1987) Estimating Time Varying Risk Premia in the Term Structure: The Arch-M Model. *Econometrica*, **55**, 391-407. https://doi.org/10.2307/1913242

[12] Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, **31**, 307-327. https://doi.org/10.1016/0304-4076(86)90063-1

[13] Black, F. (1976) Studies of Stock Price Volatility Changes. *Proceedings of the* 1976 *Meetings of the Business and Economic Statistics Section*, Boston, 23-26 August 1976, 177-181.

[14] Nelson, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, **59**, 347-370. https://doi.org/10.2307/2938260

[15] Vallender, S.S. (1974) Calculation of the Wasserstein Distance between Probability Distributions on the Line. *Theory of Probability & Its Applications*, **18**, 784-786. https://doi.org/10.1137/1118101

[16] Villani, C. (2009) Optimal Transport: Old and New. Springer. https://doi.org/10.1007/978-3-540-71050-9

[17] Kendall, M.G. (1938) A New Measure of Rank Correlation. *Biometrika*, **30**, 81-93. https://doi.org/10.1093/biomet/30.1-2.81

[18] Yahoo Finance (n.d.) Nikkei 225 (^N225) Historical Data. https://finance.yahoo.com/quote/%5EN225/history/