Journal of
*Risk and Financial Management*

MDPI

*Article*

# Optimal and Non-Optimal MACD Parameter Values and Their Ranges for Stock-Index Futures: A Comparative Study of Nikkei, Dow Jones, and Nasdaq

**Byung-Kook Kang**

Department of Business Administration, Nanzan University, Nagoya 466-8673, Japan; kangbk@nanzan-u.ac.jp

**Abstract:** This study investigates the optimal and non-optimal parameter values of the MACD (Moving Average Convergence Divergence) technical analysis indicator for three major stock market index futures: the Nikkei 225, the Dow Jones, and the Nasdaq. Using a recently developed methodology, it reveals the existence of specific ranges of optimal and non-optimal values for each of the three parameters of the MACD indicator in these indices. Sample models employing the optimal parameter values in the three index futures generated significantly higher returns, outperforming both a non-technical buy-and-hold strategy and a random strategy that did not incorporate any market information. This discovery suggests that the three market indices may not be weak-form efficient. Therefore, this study contributes to the research on market efficiency by verifying inefficiency using a new approach. The highlight of this study is identifying that the ranges of optimal parameter values for the three indices are different from each other, but the optimal parameter value combinations for each of the three indices share a unique characteristic form. This issue and its finding have not been explored in the existing literature. Several interesting findings and valuable insights for market participants and researchers arise from this study. The new methodology is unique in finding optimal and non-optimal parameter values through the analysis of parameter sets used in well-performing and poorly performing sample models. Its validity and reliability have been confirmed by this study, making a useful contribution to the field of technical analysis research, particularly in parameter optimization insight.

**Keywords:** MACD; technical analysis; stock index futures; market efficiency; stock trading simulation; COVID-19

check for updates

**Citation:** Kang, Byung-Kook. 2023. Optimal and Non-Optimal MACD Parameter Values and Their Ranges for Stock-Index Futures: A Comparative Study of Nikkei, Dow Jones, and Nasdaq. *Journal of Risk and Financial Management* 16: 508. https://doi.org/10.3390/jrfm16120508

Academic Editor: Pankaj Topiwala

Received: 27 October 2023
Revised: 29 November 2023
Accepted: 2 December 2023
Published: 7 December 2023



## 1. Introduction

Profitable technical analysis is inconsistent with the efficient market hypothesis (EMH) of Fama (1970). Because the EMH argues that current market prices must reflect all available information, technical trading rules based on historical prices should not earn genuine profits. A substantial body of evidence supports this hypothesis. Nevertheless, both practical and intellectual doubts about the EMH remain, making it a hotly debated topic among academics and practitioners.

Within this context, many academic researchers have examined a very popular technical analysis indicator, the MACD, and tested its effectiveness or used it to gauge market efficiency. Many of these evaluations failed to confirm either the profitability of the technical approach or the existence of market efficiency based its traditional parameter settings of 12, 26, and 9 days. Several examples are described in the next section.

The MACD tool consists of three parameter values that define three periods (these are moving averages of historical prices. The first two parameters are for the calculation of the MACD series itself, constructing a line that shows the difference between short- and long-term exponential moving averages. The third parameter is for creating a signal line that depicts the exponential moving average of the MACD series. These three parameter values are written in the form MACD ($n_1$, $n_2$, $n_3$), for example, MACD (12, 26, 9). The tool can be interpreted as indicating to buy (sell) when the MACD line crosses up (down) through the signal line. This is called the "signal line crossover" trading rule.

The MACD indicator, like any technical tool, can generate false trading signals. This is because trade signals generated by the momentum indicator depend solely on its parameters. There is no means of identifying false trading signals in advance. Hence, adjusting the parameter values to a target investment is necessary to reduce false signal generation and improve performance. It is therefore natural to ask, "Why is the traditional parameter values (12, 26, 9) version the sole determinant of the profitability of all MACD models and the only valid test of market efficiency?" Changing parameter values will lead to different outcomes and conclusions.

Only a few researchers have paid attention to the importance of selecting optimal parameter values. Erić et al. (2009) examined various MACD models with different parameter values and identified the most profitable parameter values for the 48 companies listed on Belgrade Stock Exchange. They stated that "The application of the optimized MACD indicator of technical analysis to the financial market significantly contributes to the maximization of profitability on investments (p. 171)". However, the most profitable parameters for each company that were identified from in-sample tests (using data from June 2004 to May 2008) all resulted in losses in out-of-sample tests (using data from May 2008 to May 2009). Consequently, they concluded that "it is important to optimize the three parameters in time. (p. 185)" Borowski and Pruchnicka-Grabias (2019) also conducted a similar investigation into optimal parameter values for the 140 companies listed on the Warsaw Stock Exchange (using data from 2000 to 2018). They found that, while a few companies had identical optimal parameters, many had different ones, and concluded that "there is nothing like standard time lengths for moving averages (i.e., parameter values) for all kinds of investments. (p. 458)".

Kang (2021) expressed concerns about prior research that exclusively sought the most profitable parameter values for individual companies. He noted that the parameter values yielding the highest returns might not necessarily be the single best or optimal combination. He identified the MACD (4, 22, 3) model as one of the most optimal choices for Nikkei 225 index futures due to its balanced performance in terms of profit generation, loss avoidance, and the total number of transactions, even when other models offered higher returns. He also investigated the potential benefits of incorporating supplementary trading strategies to enhance profitability by reducing the number of false trading signals. His findings revealed that the number of models showing improved performance with supplementary strategies was significantly higher for models with optimized values than for models with non-optimized values. He concluded that "Prioritizing the optimization of the three parameter values should precede the search for supplementary trade strategies to avoid false trading signals. (p. 19)".

A common point of the research described above is examining many sample models with different parameter values. But Erić et al. (2009), along with Borowski and Pruchnicka-Grabias (2019), focused on finding the most profitable parameter value combination for stocks of individual companies, whereas Kang (2021) focused on the most optimal values for a stock market index. Investors in individual stocks may find the former approach and results interesting, but for investors interested in index trading, the latter may offer a new perspective on one's target market, especially for investors sympathetic with the EMH who tend to purchase index funds that track the market overall, that is, passive portfolio management.

Nevertheless, it is a questionable methodology to take the single model with the most profitable parameter values and discuss its optimality while overlooking other high-performance models with different parameter values but similar performance. In the same vein, if one compares a single model with the most profitable or the most optimal parameter values in one market to that in another market and argue that the difference between the two models' parameters reflects the two markets' features, it is not sufficient to reach general conclusions since it is just one example of comparison. Some examples of this are included in the following section. It is worth noting the work of Borowski and Pruchnicka-Grabias (2019) mentioned above and subsequent work by Kang (2022).

Borowski and Pruchnicka-Grabias (2019) extended their research to consider the consistency of the optimal parameter values for the 140 companies and investigated if each value of the optimal parameters was an even or an odd number. They concluded that "the transaction system was optimized mainly by short moving averages (p. 464)" and "the highest performances were obtained for 'odd-even-odd' combinations of the three parameters (p. 468)". While there is a question regarding the practical utility of this discovery for market participants, to the best knowledge of this author, this was the sole systematic endeavor to ascertain the features of optimal parameter value combinations for a given market.

On the other hand, Kang (2022) still had an issue with high-performing models, including the MACD (4, 22, 3) model, that were examined in his previous study, "What characteristics do those models' parameter value sets possess"? To address this question, he introduced a new methodology for discovering well-performing and poorly performing parameter value sets for the Nikkei market. In brief, he investigated 19,456 MACD models with different parameter values (using data from 2011 to 2021) and observed that "The most frequently used parameter values of the sample models in the top 100 and the bottom 100 ranked models in terms of profitability have distinctively different distributions from each other. (p. 4)" Based on this observation, he focused on the frequently used values that were equal to or greater than a certain minimum frequency ratio in each of the top/bottom 100 ranked models and considered those values as optimal/non-optimal parameters for the market. Using this approach, he identified specific ranges of optimal/non-optimal values and created several groups of hypothetical sample models with the analyzed optimal/non-optimal parameter values. Importantly, only 19 of the 992 hypothetical models were among the original top 100 models.

Nevertheless, the test outcomes indicated the following: (1) Most of the hypothetical optimal (non-optimal) models yielded positive (negative) returns, not only for in-sample tests but also for out-of-sample tests, suggesting the new methodology's effectiveness in finding optimal/non-optimal parameter values for the Japanese market. (2) The resulting combinations of the three parameter values ($n_1$, $n_2$, $n_3$), used in the highest performing group of sample models, exhibited a characteristic form: the second parameter value is larger than the other two parameters, i.e., $n_1 < n_2$ and $n_2 > n_3$.

However, as Kang's work aimed to demonstrate the effectiveness of the new approach, it became a bit too detailed by considering too many sample models, dividing them into "two primary groups" of optimal sample models and "four secondary groups" of suboptimal sample models. Furthermore, his research remains confined to the Japanese Nikkei futures market. Therefore, these aspects require further discussion:

1. If we change the minimum level of the frequency ratio used to identify optimal/non-optimal parameter values and redefine these parameters for the Japanese market, will the models with the newly identified optimal/non-optimal parameter values also perform well/poorly in the market?
2. If we apply the new methodology to different markets, such as the Dow Jones futures and the Nasdaq futures, will it also prove effective in finding the optimal/non-optimal parameters for the MACD trading system in these markets?
3. If we compare the resulting optimal/non-optimal parameter values for each of the three markets, including the Japanese market, how do these values for the three markets differ from each other? Additionally, can the characteristic form observed in the Japanese market also be found in the two U.S. markets, or will there be differences?

The first question posed above will give an opportunity to reaffirm the effectiveness of the new methodology presented by Kang (2022), and the second question will provide additional evidence to validate its effectiveness in different markets. The third question offers a country-by-country analysis that makes it possible to deepen one's perspective about the three markets, which has not been discussed in the existing literature. Investigating these three questions is the objective of this study, and it will also uncover other valuable insights from the study's results. The merit of the new methodology will be discussed in the final section.

## 2. Literature Review

Numerous studies have investigated the predictive power of technical analysis, but the literature on the MACD published in the last two decades can be categorized into four categories: (1) Research that examines the profitability of the MACD approach and measures of market efficiency with the traditional MACD (12, 26, 9) model; (2) Research that attempts to use other parameter value combinations that differ from the traditional (12, 26, 9) format; (3) Research that examines a larger number of parameter value combinations and selects the most profitable parameters for individual companies; and (4) Research that employs complex algorithms to find the best-performing optimal parameters or improves the performance of MACD applications.

Meissner et al. (2001) is an example of research in the first category. They examined the traditional MACD (12, 26, 9) model and found that it resulted in poor success rates for both DOW and NASDAQ stocks over the period from 1989 to 1999. They concluded that "the traditional MACD indicator can almost be regarded as a contra-indicator". Chen and Metghalchi (2012) tested the predictive power of 32 different models based on single, double, or triple-indicator combinations for the Brazilian stock index (BOVESPA) over the period from 1996 to 2011. None of their models, including the MACD (12, 26, 9) model, beat a simple buy-and-hold strategy. They concluded that "our results support strongly the weak form of market efficiency of the Brazilian stock market".

Rosillo et al. (2013) also reported that they obtained unsatisfactory results by applying the MACD (12, 26, 9) model to the companies listed on the Spanish Continuous Market from 1986 to 2009. Biondo et al. (2013) investigated whether the MACD (12, 26, 9) model, along with three other standard technical analysis tools, outperformed a random trading strategy (based on the uniform distribution) using 15 to 20 years of data for the FTSE-UK, FTSE-MIB, DAX, and S&P indices. They found that "the four conventional trading strategies perform no better than the random strategy on average" and concluded that "the random strategy is less risky than the considered standard trading strategies." Similarly, Nor and Wickremasinghe (2014) examined the MACD (12, 26, 9) model for the Australian All Ordinaries Index (XOA) using data from 1996 to 2014. They found that the traditional MACD model generally performed poorly, but another tested RSI (Relative Strength Index) model showed some potential. They concluded that "overall, the Australian stock market is not weak-form efficient".

Researchers have generally failed to find positive results for the MACD tool with its traditional parameter values. In other words, all the above studies prove that the conventional MACD (12, 26, 9) model does not work. If they had considered different parameter values, their conclusions might have been very different.

Hejase et al. (2017) examined the MACD (12, 26, 9) model to determine if the MACD tool could produce profits for Lebanese traders for six Lebanese banks and a company. They found unsatisfactory results from their examinations of data from 2004 to 2014. Interestingly, after filtering the empirical results, they applied three kinds of additional trading approaches to avoid false trading signals generated by the MACD. These approaches included filter rules such as buying (selling) whenever the stock price increased (decreased) by a given percentage. However, their conclusion was that "in the long run, MACD dynamic trades do not make sense, as none of the applied strategies outperformed the simple buy-and-hold strategy".

Of course, there is research that has found positive results using the MACD tool, but such examples are extremely rare. Research by Chong and Ng (2008) is the only example that can be discussed here. They examined the MACD (12, 26, 0) model and the RSI to determine if they are profitable for the FT30 index on the London Stock Exchange. Using monthly data for a 60-year period from 1935 to 1994, they found that the two technical trading rules generated higher returns than a buy-and-hold strategy. However, this result should be reconfirmed with new data from 2000 onwards. Regarding the actual use of technical analysis, the research of Menkhoff (2010) is notable. He pointed out with survey evidence from 692 fund managers in five countries that "The vast majority of fund managers use it to some extent . . . as a complement to fundamental analysis . . . (and) at shorter-term forecasting horizons. Up to horizons of weeks, it is more important than fundamental analysis in all countries (p. 2585)".

In the second category, Chong et al. (2014), a follow-up to Chong and Ng (2008), deserves attention. They re-investigated whether the MACD and the RSI generate excess returns for the stock markets in five OECD countries (Canada, Germany, Italy, Japan and the United States). They applied three conventional models, the MACD (12, 26, 0), MACD (12, 26, 9), and MACD (8, 17, 9) models, to market data gathered from more than 27 years, from 1976 to 2002, and reported interesting results: While the MACD (12, 26, 0) model had beaten the buy-and-hold strategy in the Italian market (Milan Comit General), the MACD (12, 26, 9) model did not perform in the same market. In addition, the MACD (8, 17, 9) model delivered significantly negative returns for the Italian and German markets (DAX30), and it had zero predictive power for the other markets. They concluded, "the three traditional MACD trading rules are not robust to the choice of market". However, if they had paid attention to the two opposite results for the MACD (12, 26, 0) and the MACD (12, 26, 9) models in the Italian market, they might have noticed that the profitability of the MACD approach depends on its parameter values.

An example of a relatively recent study is Montgomery et al. (2019). They conducted tests on the efficacy of technical trend-following rules, which included the MACD rule and several moving average rules, as well as the 52-week high strategy for 50 U.S. corporate bonds and their corresponding stocks. Their findings indicated that, "Over the 2002 to 2015 period, the technical rules and the 52-week strategy were unprofitable for both bonds and stocks". However, they used three types of MACD models, all with the traditional values of 12 and 26, including the MACD (12, 26, 0).

None of the previously mentioned research provided a clear rationale for using the traditional parameter values, except for ambiguous expressions such as, 'because it is most commonly used'.[1] On this point, Appel (1979), creator of the MACD indicator, does not seem irrelevant. He proposed to use the MACD (12, 26, 9) model for buying and the MACD (19, 39, 9) model for selling in the Dow Jones Industrial Average, Nasdaq Composite Index, and S&P 500, stating that "Different combinations are useful for buying and for selling" in Appel (2005). However, Appel also did not provide a satisfactory explanation for why he proposed the two specific models or how he determined these parameter combinations to be worthy of consideration.

Abbey and Doukas (2012) made an important assertion regarding this point. They examined whether technical analyses, including the MACD (12, 24, 0) model and three other well-known technical indicators, are profitable for individual currency traders. They used a database of 428 individuals from 2004 to 2009. However, they found that technical analysis was negatively associated with performance and concluded that this negative association occurred "because traders used well-known technical indicators to trade currencies". This statement shows the irrationality of using technical analysis tools and their parameter values solely because of their popularity.

Moving on to the third category, we consider Erić et al. (2009) and Borowski and Pruchnicka-Grabias (2019). As previously noted, both of these studies searched for the most profitable parameters among various combinations for individual companies and pointed out the necessity of optimizing the three parameter values. Adding a supplementary note to these two studies, they provided lists of the most profitable parameter values for individual companies (specifically 48 companies on the Belgrade Stock Exchange in the Republic of Serbia and 140 companies on the Warsaw Stock Exchange in Poland). Interestingly, Kang (2022) compared the sets of the most profitable parameter values (48 sets from the former study and 140 sets from the latter study) with the 288 sets of the best-performing parameter values found for the Japanese Nikkei 225 futures market and discovered that none of the former 48 sets (only 4 sets of the latter 140 sets) were identical to the 288 sets. Although that study was a simple comparative analysis that did not consider the difference between developing and developed markets, it shows a significant difference between the optimal parameter value sets for the markets of the two countries (Serbia vs. Japan and Poland vs. Japan). This suggests that exploring this potential in markets of other countries could provide unique insights tailed to each market's conditions, which is the goal of this study.

On the other hand, the research by Anghel (2015) is noteworthy in the sense that it was the first attempt to assess the information efficiency of global stock markets (75 countries with 1336 companies from 2001 to 2012) using the MACD indicator. He examined three parameter values that varied over a given range and maximized a target estimator function for the assessment. From the results of this assessment, he concluded that "traders using the MACD as a technical analysis investment method on the stock market could sometimes obtain abnormal cost- and risk-adjusted returns" and pointed out that "the world's stock markets present important inefficiencies".

The fourth category encompasses research approaches that utilize complex algorithms, such as genetic algorithms, or artificial intelligence (AI), like neural networks, as well as other methods to improve existing methodologies. Within this category, numerous applications for predicting stock prices exist, but a few examples related to the MACD include: (1) Bodas-Sagi et al. (2009), who demonstrated that the parameters of the MACD can be optimized using evolutionary algorithms; (2) Wiles and Enke (2015), who used genetic algorithms to optimize the three MACD parameter values for the soybean futures market; (3) Wang and Kim (2018), who introduced variable weights for calculating the exponential moving average of the MACD, deviating from the typical fixed weight; (4) Hashida and Tamura (2019), who presented a new MACD-histogram-based FCN (fully convolutional neural network) model using outcomes from the MACD (13, 26, 9) and MACD (3, 5, 4) models; and (5) Chou and Lin (2019), who presented an integrated fuzzy neural network that incorporates five technical indicators, including MACD.

As seen above, numerous studies have been conducted using the MACD. However, none of the literature considers the possibility of an optimal "range" for each of the three parameter values that would be suitable for a specific market. Acquiring such knowledge would help us understand the optimal parameter combinations for a given market and how they differ from those in other markets, ultimately enhancing our understanding of each market's characteristics.

Finally, let us mention a study that provides good guidelines for research using technical analysis tools. Park and Irwin (2007) conducted an extensive survey of the technical analysis literature published between 1960 and 2004. They pointed out the limitations of the early studies: no consideration of parameter optimization, no out-of-sample validation, no statistical tests of the significance of trading returns, and no consideration of data snooping biases. Viewed in this respect, only some of the literature reviewed above satisfy these requirements, but this study does.

## 3. Research Methodology

This section provides a summary of the research method, including data description, sample models tested, trading simulation rules, a brief explanation of the new methodology for finding optimal/non-optimal parameter values, and statistical tests of the significance of trading returns.

### 3.1. Data and Sample Models

This study utilizes the daily closing index values of Nikkei 225 futures (contracts near maturity), Dow Jones futures, and Nasdaq futures over the last 11 years (2011–2021). The historical data for Nikkei 225 futures (4 January 2011–30 December 2021) were obtained from an official data provider, JPX Data Cloud (http://db-ec.jpx.co.jp) operated by the Japan Exchange Group (JPX), and the data for Dow Jones futures and Nasdaq futures

(3 January 2011–31 December 2021) are from one of the top global financial platforms and news websites, Investing.com (http://www.investing.com). For reference, the Nikkei 225 comprises the leading 225 blue-chip companies listed on the Tokyo Exchange and serves as the primary JPY-denominated stock index futures.

We allocated the first 9 years of data (2011–2019) for in-sample testing purposes and the last 2 years of data (2020–2021) for out-of-sample testing. Additionally, we divided the initial 9 years into three fairly long sub-periods: 2011–2013, 2014–2016, and 2017–2019. This means that the calculation of profitability for each model is conducted separately within each sub-period to ensure the robustness of the sample test period selection. This approach is in line with Kang (2022), who followed the approach adopted by Chong and Ng (2008) to prevent data snooping (selection) bias.

As for the baseline models considered in this research, we employ the same set of 19,456 models used in Kang (2022). The set consists of models ranging from MACD (3, 5, 3) to MACD (20, 40, 40), with the three parameters ($n_1$, $n_2$, $n_3$) defined as follows: $n_1 = \{3, \cdots, 20\}$, $n_2 = \{5, \cdots, 40\}$, $n_3 = \{3, \cdots, 40\}$ in a one-day interval. We utilize this identical set of sample models for all three futures markets. The purpose of this is to identify which models, within a range of identical sample models, perform well and poorly in each market, and to ascertain how they differ from each other.

### 3.2. Trading Rules

In terms of trading simulations, this study employs the trading rule used in Kang (2022) to ensure consistency in the trading results across both research papers. The trading rule can be summarized as follows: (1) When a 'buy' (or 'sell') signal is generated based on the signal line crossover, a buy (or sell) order for 'one trading unit' is executed at the closing price (index value) on the next day. (2) Once a 'buy' (or 'sell') position is opened, any subsequent identical buy (or sell) signals are disregarded. However, if the first opposite trading signal, i.e., a 'sell' (or 'buy') signal, is generated, the buy (or sell) position is assumed to be closed out at the closing price reported on the next day. (3) Simultaneously, to implement the newly generated signal, a new 'sell' (or 'buy') order for one trading unit is assumed to be executed at the same closing price on the same day. In other words, when a position is closed, a reverse trade is automatically executed. (4) Consequently, only one position for 'one unit' can be open at a time, and all transactions need to be executed sequentially, one-by-one. Holding multiple positions is not allowed after a position is taken.

Note that this trading rule repeats a 'long trade' (to profit from increases in the index values) and a 'short trade' (to profit from decreases in the index values) in sequence. Therefore, the performance of all sample models can be distinguished by long trades and short trades, which is why this study chose index futures for its tests.

Transaction fees are not taken into account in this study, similarly to the previous study. Specifically, the current round-trip commission for one large contract unit in the Nikkei 225 futures market is negligible in comparison to the high leverage of 1000 times the index value. It amounts to approximately 0.33 percent of the positive return equivalent to JPY 100,000 when the index moves up by just 10 ticks (with each tick size being JPY 10). This approach is then applied to the other two U.S. futures markets to ensure consistency in the trading results across all three markets. Hence, the return of each transaction is computed using index values according to the following equation:

$$r = P_{sell} - P_{buy} \tag{1}$$

where $P_{buy}$ ($P_{sell}$) signifies the closing index value on the day when a buying (selling) transaction is carried out. A positive value of $r$ indicates a profit, while a negative value of $r$ indicates a loss, regardless of whether it a long or short trade. However, log returns are used for conducting statistical tests to determine the significance of trading returns. For reference, this study employs the following formula for calculating the Exponential Moving Average (EMA) of the MACD series: $EMA_t = [(2/(n+1)(P_t - EMA_{t-1})] + EMA_{t-1}$, where $n$ represents the number of periods for the EMA, and $P_t$ indicates the closing day price (index value) at time $t$.

### 3.3. The Rule of the New Methodology for Identifying Optimal/Non-Optimal Parameter Values

The new methodology presented in Kang (2022) is divided into three parts: examining a large number of sample models; finding the range of optimal/non-optimal parameter values; and creating new sample models with those parameter values. To summarize and redefine this approach:

- Step 1: Examining the profitability of a large number of sample models.
- Step 2: Focusing on a certain level of top/bottom models. Note that this study examines the profitability of the 19,456 sample models using the first nine years of data (2011–2019) for each of the three index markets, and focuses on the top/bottom 100 ranked models in each market as target models.
- Step 3: Checking the frequency distributions of the three parameter values.
- Step 4: Identifying good/poor performing parameter values. For the top (bottom) 100 ranked models, examine the frequency distributions of the three parameter values ($n_1$, $n_2$, $n_3$) used in the models. Then, determine the values for which the frequency ratio is greater than or equal to a specific percent, representing a given 'minimum cutoff point' in the sense that these values can be considered as optimal (non-optimal) parameter values that perform well (poorly). However, if a predetermined value appears in both the top and bottom 100 models, exclude it in the sense that we cannot determine whether it is an optimal or non-optimal value.
- Step 5: Defining the range of optimal/non-optimal parameter values.
- Step 6: Creating new sample models using the predetermined parameter values. For the remaining identified parameter values of the top and bottom 100 models, focus on the values which are predominantly distributed around the most frequently observed value. If these values appear consecutively, classify them as a group, considering them as elements belonging to a range of optimal or non-optimal parameter values. If this rule results in multiple groups, differentiate them as 'A', 'B', and so on. If an identified value does not fit into a predetermined group, place it within the nearest group to avoid creating numerous subsets with small elements. Moving on to the next step, create new sample models with the optimal or non-optimal parameter values within the predefined ranges.

Note that "the minimum cutoff point" mentioned in the 4th step above does not have a strict criterion. However, it should be considered that setting a higher (lower) cutoff point will result in considering fewer (more) parameter values later when creating new models. In consideration of this point, this study adopts a minimum cutoff point of "5%", which makes it possible to create a single group of new sample models for each of the three markets. This facilitates comparisons of their performance in each market.[2]

The question regarding this methodology is: Do the new sample models created using the derived optimal (non-optimal) parameter values from the above rules perform well (poorly)? If these models align with our expectations, we can consider their parameter values as indicative of good-performing optimal (poor-performing non-optimal) ones. This concept forms the basis of this rule-based methodology.

### 3.4. Statistical Tests for Assessing the Robustness of Simulation Returns

For the new sample models with predetermined optimal parameter values from the new methodology, we will conduct tests to determine if these models outperform two benchmark trading strategies: the classic non-technical 'buy-and-hold' investment strategy and the purely 'random' strategy based on a uniform distribution which was adopted by Biondo et al. (2013).

The first benchmark strategy does not consider market fluctuations or technical indicators. Therefore, testing against this benchmark strategy helps confirm whether the MACD technical analysis approach (i.e., the new sample models created using the derived optimal parameter values) can outperform the non-technical strategy. The second benchmark strategy, which follows a purely random approach without utilizing any market information or pre-established trading mechanisms, deserves consideration in the sense that the random strategy is a more objective and appropriate benchmark for testing market efficiency. The specific details of these two benchmark strategies and the statistical tests will be provided in Section 5.6.

## 4. Preliminaries for Creating New Sample Models

This section demonstrates the process of creating new sample models by applying the rule described in the preceding section to the three markets, covering Step 3 to Step 6.

### *4.1. Distribution of the Three Parameter Values in the Top/Bottom 100 Ranked Models*

Figure 1 shows the frequency distributions of the three parameter values in the models for the Nikkei 225 (left), Dow Jones (middle) and Nasdaq (right) futures markets. The three vertical histograms in panel [a] show the frequencies of the three parameter values observed in the top/bottom (blue/red) 100 models out of the 19,456 sample models. Panel [b] and panel [c] show the frequencies observed in the top/bottom 500 and 1000 models, respectively.


**Figure 1.** Frequencies of the three parameter values ($n_1$, $n_2$, $n_3$) in the top/bottom 100, 500, and 1000 models.

Observing the histograms for the Nikkei depicted on the left side of panel [a], [b], and [c], it becomes immediately apparent that the most frequently used parameter values in the top/bottom groups of models are significantly distant from each other and exhibit minimal overlap, even with an increase in sample size. As the sample sizes grow, the variability of each sampling distribution decreases, providing a clearer depiction of the range of collectively distributed values with higher frequency ratios. This pattern remains consistent and retains its prototype without experiencing significant changes, at least for the top/bottom 1000 models. Based on this observation, this study chooses to focus on the three parameter values used in the top/bottom 100 ranked models.

The same observations apply to the Dow Jones and the Nasdaq futures. Surprisingly, the Nasdaq shows a clear distinction in the most frequently used parameter values between the top and bottom group of models. These unexpected findings suggest the existence of a boundary between the optimal and non-optimal ranges of parameter values for each of the three parameters.

### *4.2. Defining the Ranges of Optimal and Non-Optimal Parameter Values*

The next step is to investigate parameter values with a frequency ratio equal to or higher than the 5% minimum cutoff point. This information can be found in the numeric values (frequency ratios) shown on the histogram bars in Figure 1. For brevity, only the results are presented below.

We defined the range of "optimal" parameter values for the three parameters as $A_{n1}$, $A_{n2}$, $A_{n3}$, respectively, by following the procedure in step 5 described in Section 3.3. The three ranges for each of the three markets are as follows:

- $A_{n1}$: $\{3, \cdots, 8\}$, $A_{n2}$: $\{8, \cdots, 10\}$, $A_{n3}$: $\{3, \cdots, 8\}$ for the Nikkei.
- $A_{n1}$: $\{4, \cdots, 8\}$, $A_{n2}$: $\{18, \cdots, 21\}$, $A_{n3}$: $\{6, \cdots, 9\}$ for the Dow Jones.
- $A_{n1}$: $\{6, \cdots, 11\}$, $A_{n2}$: $\{10, \cdots, 20\}$, $A_{n3}$: $\{6, \cdots, 16\}$ for the Nasdaq.

where the subscript $n_1$, $n_2$, $n_3$ of $A_{n1}$, $A_{n2}$, $A_{n3}$ denotes the order of the three parameters ($n_1$, $n_2$, $n_3$) respectively. Similarly, we defined the range of "non-optimal" parameter values for the three parameters as $W_{n1}$, $W_{n2}$, $W_{n3}$, respectively. The three ranges for each of the three markets are as follows:

- $W_{n1}$: $\{9, \cdots, 18\}$, $W_{n2}$: $\{29, \cdots, 35\}$, $W_{n3}$: $\{9, \cdots, 12\}$ for the Nikkei.
- $W_{n1}$: $\{18, \cdots, 20\}$, $W_{n2}$: $\{34, \cdots, 38\}$, $W_{n3}$: $\{18, \cdots, 22\}$ for the Dow Jones.
- $W_{n1}$: $\{17, \cdots, 20\}$, $W_{n2}$: $\{34, \cdots, 40\}$, $W_{n3}$: $\{33, \cdots, 40\}$ for the Nasdaq.

The reason for considering these "non-optimal" parameter values together with the "optimal" parameter values is to examine the optimality of the former and the non-optimality of the latter. If models with optimal/non-optimal parameter values in the ranges specified above exhibit significantly different and opposing performance, with the former performing well but the latter underperforming, it establishes the non-optimality of the latter as supporting evidence for the optimality of the former.

### *4.3. Creating New Sample Models*

Using the ranges presented in the preceding section, we derive two sets of parameter value combinations, each comprising three ranges corresponding to the three parameters ($n_1$, $n_2$, $n_3$): one set labeled as $A_{n1} - A_{n2} - A_{n3}$ and the other labeled as $W_{n1} - W_{n2} - W_{n3}$. Then, for each of these two combinations, we can create new sample models with the predetermined optimal/non-optimal parameter values.

First, let us examine the number of new sample models that can be created from the combination $A_{n1} - A_{n2} - A_{n3}$.

- In the case of the Nikkei, the number of elements within each range presented above is $n(A_{n1}) = 6$, $n(A_{n2}) = 3$, and $n(A_{n3}) = 6$. Accordingly, from the combination $A_{n1} - A_{n2} - A_{n3}$, a total of 108 (=6 × 3 × 6) unique sample models can be created. However, this combination generates 6 (=1 × 1 × 6) irrational models where the values of $n_1$

and $n_2$ are equal. Therefore, we only consider the remaining 102 (=108 − 6) models as the sample models for this combination.

- In the Dow Jones, we obtain 80 (=5 × 4 × 4) unique sample models from $n(\mathrm{A_{n1}})$ = 5, $n(\mathrm{A_{n2}})$ = 4, and $n(\mathrm{A_{n3}})$ = 4. No irrational models are included.
- For the Nasdaq, we have a large number of 792 (=6 × 11 × 12) sample models from $n(\mathrm{A_{n1}})$ = 6, $n(\mathrm{A_{n2}})$ = 11, and $n(\mathrm{A_{n3}})$ = 12. However, including all of these models would lead to an imbalance when compared to the similar models from the preceding two markets. This disparity arises from applying the same criteria (the "5%" minimum cutoff point) to all three markets. To address this issue, we decided to increase the cutoff point for this market from 5% to 7% and accordingly redefine the range of optimal parameter values as follows: $\mathrm{A_{n1}}$: {6, · · · , 11}, $\mathrm{A_{n2}}$: {12, · · · , 20}, $\mathrm{A_{n3}}$: {7, · · · , 10}. As a result, we obtain 216 (=6 × 9 × 4) sample models from these new ranges, with $n(\mathrm{A_{n1}})$ = 6, $n(\mathrm{A_{n2}})$ = 9, and $n(\mathrm{A_{n3}})$ = 4.

Next, let us consider the new sample models that can be created from the combination $\mathrm{W_{n1}} - \mathrm{W_{n2}} - \mathrm{W_{n3}}$.

- For the Nikkei, we can create 280 (=10 × 7 × 4) unique sample models from $n(\mathrm{W_{n1}})$ = 10, $n(\mathrm{W_{n2}})$ = 7, and $n(\mathrm{W_{n3}})$ = 4.
- For the Dow Jones, we obtain 75 (=3 × 5 × 5) unique sample models from $n(\mathrm{W_{n1}})$ = 3, $n(\mathrm{W_{n2}})$ =5, and $n(\mathrm{W_{n3}})$ =5.
- For the Nasdaq, we apply the 7% cutoff point. This led to a redefinition of the range of non-optimal parameter values for this market as follows: $\mathrm{W_{n1}}$: {17,· · · ,20}, $\mathrm{W_{n2}}$: {36, · · · , 40}, $\mathrm{W_{n3}}$: {35, · · · , 40}. As a result, we have 120 (=4 × 5 × 6) unique sample models from $n(\mathrm{W_{n1}})$ = 4, $n(\mathrm{W_{n2}})$ = 5, and $n(\mathrm{W_{n3}})$ = 6.

To sum up, the number of new sample models created above is:

- For the Nikkei: $n(\mathrm{A_{n1}} - \mathrm{A_{n2}} - \mathrm{A_{n3}})$ = 102, $n(\mathrm{W_{n1}} - \mathrm{W_{n2}} - \mathrm{W_{n3}})$ = 280.
- For the Dow Jones: $n(\mathrm{A_{n1}} - \mathrm{A_{n2}} - \mathrm{A_{n3}})$ = 80, $n(\mathrm{W_{n1}} - \mathrm{W_{n2}} - \mathrm{W_{n3}})$ = 75.
- For the Nasdaq: $n(\mathrm{A_{n1}} - \mathrm{A_{n2}} - \mathrm{A_{n3}})$ = 216, $n(\mathrm{W_{n1}} - \mathrm{W_{n2}} - \mathrm{W_{n3}})$ = 120.

Note that these models are "hypothetical" models with the "predefined" optimal/non-optimal parameter values at the current stage. Nonetheless, if these sample models exhibit good/poor performances, it can be inferred that their performance is attributed to their optimal/non-optimal parameter values because their performance depends solely on their three parameter value settings.

One thing to add here is whether the $\mathrm{A_{n1}} - \mathrm{A_{n2}} - \mathrm{A_{n3}}$ group includes the original models belonging to the "top" 100 models. We confirmed that: (1) in case of the Nikkei, none of the original models are included in the new set of 102 sample models; (2) for the Dow Jones, the same holds true for the new set of 80 sample models; however, (3) for the Nasdaq, 28 original models from the top 100 are included in the new set of 216 sample models.

Regarding the $\mathrm{W_{n1}} - \mathrm{W_{n2}} - \mathrm{W_{n3}}$ group, the following number of original models from the "bottom" 100 models are included: (1) 44 of the new 280 models for the Nikkei; (2) 10 of the new 75 models for the Dow Jones; and (3) 60 of the new 120 models for the Nasdaq.

As evident from the above, the inclusion of the original models from the top/bottom 100 models in the set of new sample models varies in each market. However, we do not exclude the original models in advance; instead, we examine the performance of all the hypothetical sample models in the next section. This approach is taken because our objective is to identify the range of well-performing optimal and poorly performing non-optimal parameter values for each of the three markets. If the out-of-sample tests yield results that are similar to their original performance during the in-sample tests, then their parameter values should be considered as optimal/non-optimal parameter values.

## 5. Empirical Results

We now assess whether the hypothetical sample models indeed perform well (poorly). Before moving on to the empirical results, let us take a moment to observe the movements of the three indices over the past 11 years from Figure 2.

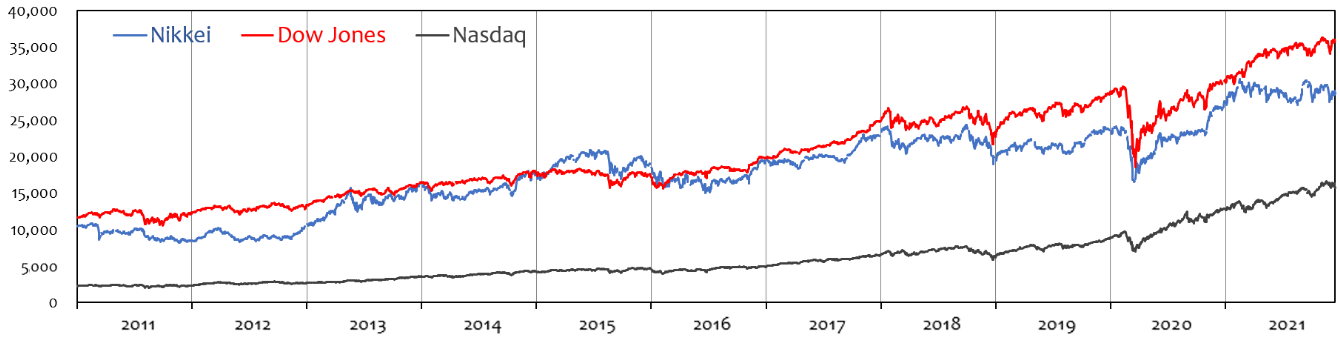


**Figure 2.** Index values of the Nikkei 225, Dow Jones, and Nasdaq futures (2011–2021). Note: This figure shows the historical changes in the three indices: Nikkei 225 (blue), Dow Jones (red), and Nasdaq (black). The dataset for each index covers 2692, 2896, and 2877 trading days, respectively.

The first point to note is that the three index values exhibited a consistent upwards trend over the past 11 years, although the pace of their upward trajectory varied. The second point to note is the turbulent roller coaster ride observed in 2020, reflecting the impact of the global COVID-19 pandemic and the various stimulus packages implemented by governments throughout the year. It will be interesting to assess how the sample models in this study performed during the pandemic period (2020–2021), especially in the initial outbreak year of 2020. These two factors—long-term upward trends with variability and the pandemic crisis situation—will play a significant role in assessing the performance of the sample models.

### *5.1. Results of In-Sample Tests*

Table 1 presents summary statistics for the returns of the sample models belonging to the two groups, $A_{n1} - A_{n2} - A_{n3}$ and $W_{n1} - W_{n2} - W_{n3}$. Note that the returns of the two groups are presented as raw returns calculated from index values to facilitate a comparison of their performances in the three markets.

From Table 1, we observe the following key performance patterns: (1) The sample models belonging to the $A_{n1} - A_{n2} - A_{n3}$ group for the Nikkei yield a positive mean return of '4244.5'. (2) Conversely, the sample models in the $W_{n1} - W_{n2} - W_{n3}$ group for the same market exhibit a large negative mean return of '−10,928.7'. Moreover, their returns range from '−2570' to '−16,120', all of which are negative. (3) Similar observations can be made for the sample models in the Dow Jones and the Nasdaq.

**Table 1.** Summary statistics of in-sample test results (2011–2019).

| | n | Max | Min | Mean | Median | Std. E. | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| [a] Nikkei | | | | | | | | |
| $A_{n1} - A_{n2} - A_{n3}$ | 102 | 8130 | −120 | 4244.5 | 4160 | 191.9 | −0.0302 | −0.6706 |
| $W_{n1} - W_{n2} - W_{n3}$ | 280 | −2570 | −16,120 | −10,928.7 | −12,065 | 199.9 | 0.6188 | −0.6099 |
| [b] Dow Jones | | | | | | | | |
| $A_{n1} - A_{n2} - A_{n3}$ | 80 | 5376 | −1151 | 2512.8 | 2270 | 142.9 | −0.1579 | −0.1423 |
| $W_{n1} - W_{n2} - W_{n3}$ | 75 | −1061 | −9497 | −7167.0 | −7789 | 217.1 | 1.2838 | 1.2754 |
| [c] Nasdaq | | | | | | | | |
| $A_{n1} - A_{n2} - A_{n3}$ | 216 | 2011 | −1485 | 591.6 | 976 | 65.4 | −0.4697 | −1.1993 |
| $W_{n1} - W_{n2} - W_{n3}$ | 120 | −4017 | −5196 | −4668.5 | −4701 | 20.3 | 0.4286 | 0.2199 |

Note: The "n" column represents the number of sample models in each group of $A_{n1} - A_{n2} - A_{n3}$ and $W_{n1} - W_{n2} - W_{n3}$. The "Std. E." column shows the standard error.

All these results demonstrate a significant difference in mean return between the two groups that is consistently observed across all three index markets. Therefore, for the out-of-sample tests, we can anticipate good performances from the models within the $A_{n1} - A_{n2} - A_{n3}$ group, but not from the models within the $W_{n1} - W_{n2} - W_{n3}$ group.

### 5.2. Results of Out-of-Sample Tests

Table 2 shows several noteworthy points, including: (1) The sample models in the $A_{n1} - A_{n2} - A_{n3}$ group for the Nikkei exhibit a mean return of '6301.6', and all of their returns are positive without exception. (2) Conversely, the sample models in the $W_{n1} - W_{n2} - W_{n3}$ group for the same market show a negative mean return of '−1494.5'. (3) Similar observations can be made for the sample models in the Dow Jones and the Nasdaq.

**Table 2.** Summary statistics of out-of-sample test results (2020–2021).

| | n | Max | Min | Mean | Median | Std. E | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| [a] Nikkei | | | | | | | | |
| $A_{n1} - A_{n2} - A_{n3}$ | 102 | 11,430 | 100 | 6301.6 | 6615 | 320.5 | −0.1273 | −1.2131 |
| $W_{n1} - W_{n2} - W_{n3}$ | 280 | 2940 | −4210 | −1494.5 | −1410 | 73.6 | 0.4984 | 1.4623 |
| [b] Dow Jones | | | | | | | | |
| $A_{n1} - A_{n2} - A_{n3}$ | 80 | 5162 | −3607 | 2441.8 | 3133 | 272.5 | −1.1720 | 0.4307 |
| $W_{n1} - W_{n2} - W_{n3}$ | 75 | −7905 | −10677 | −8875.1 | −8740 | 75.0 | −0.9226 | 0.3854 |
| [c] Nasdaq | | | | | | | | |
| $A_{n1} - A_{n2} - A_{n3}$ | 216 | 4738 | −709 | 3185.4 | 3217 | 46.3 | −1.6560 | 7.1227 |
| $W_{n1} - W_{n2} - W_{n3}$ | 120 | 533 | −3038 | −1676.6 | −2052 | 94.1 | 0.6208 | −0.7794 |

All these results indicate that the sample models within the $A_{n1} - A_{n2} - A_{n3}$ group continue to perform well, while the sample models in the $W_{n1} - W_{n2} - W_{n3}$ group still experience poor performance, consistent with the results from the in-sample tests.

In light of these findings, what we should focus on here are: (1) The performance of all the tested models in the two sample tests is solely determined by their three parameter values. In simpler terms, the good (poor) performance of the sample models in the $A_{n1} - A_{n2} - A_{n3}$ ($W_{n1} - W_{n2} - W_{n3}$) group is directly linked to their optimal (non-optimal) parameter value settings. (2) As a result, the poor performance of the sample models in the $W_{n1} - W_{n2} - W_{n3}$ group serves as strong evidence supporting the optimality of the parameter values used in the models of the $A_{n1} - A_{n2} - A_{n3}$ group.

### 5.3. A Brief Conclusion and an Emerging Issue

Based on the results shown in Tables 1 and 2, we can draw the following conclusions:

- The parameter values utilized in the models of the $A_{n1} - A_{n2} - A_{n3}$ ($W_{n1} - W_{n2} - W_{n3}$) group demonstrated their effectiveness (ineffectiveness) as well-performing optimal (poorly performing non-optimal) values.
- The optimality (non-optimality) of the parameter values used in the sample models has been confirmed and verified, not only in the Japanese market but also in the two U.S. index futures markets.
- These findings collectively indicate that the new methodology possesses significant validity in identifying well-performing models with optimal parameter values and detecting poorly performing models with non-optimal parameter values.

These concluding remarks answer the first two questions stated at the end of the first section. However, a new issue arises when we scrutinize the data in the "Min" column in both Tables 1 and 2. In the case of the Nikkei, the minimum return of a sample model belonging to the $A_{n1} - A_{n2} - A_{n3}$ group is '−120' in Table 1 and '100' in Table 2. These values indicate the possibility that a certain number of sample models within this group initially had negative returns for the in-sample tests but later turned positive in the out-of-sample tests. As a result, such models contributed to the overall positive mean return of the group in the out-of-sample tests. Considering the negative minimum returns such as '−1151' and '−3607' for the Dow Jones, and '−1485' and '−709' for the Nasdaq, recorded in the "Min" column of Tables 1 and 2 for the $A_{n1} - A_{n2} - A_{n3}$ group, we need to investigate how many models within the group had negative returns in either the in-sample tests, out-of-sample tests, or both.

The same reasoning applies to the returns of the sample models in the $W_{n1} - W_{n2} - W_{n3}$ group. However, further investigation into these models is unnecessary, as their poor performance has already been verified. Therefore, we will focus on examining the number of models within the $A_{n1} - A_{n2} - A_{n3}$ group that achieved a 'positive' return not only for the in-sample tests but also for the out-of-sample tests, as well as the other models that experienced negative returns for either the in-sample tests, the out-of-sample tests, or both.

Table 3 shows that the $A_{n1} - A_{n2} - A_{n3}$ group for the Nikkei has 101 'P-P' models and only one 'N-P' model. Therefore, even if we exclude the 'N-P' model and focus solely on the 'P-P' models, it does not significantly affect the validity of the three key conclusions stated earlier. However, in the Dow Jones, there are 66 'P-P' models and 14 (=11 + 1 + 2) models of other types. In the Nasdaq, there are 142 'P-P' models and 74 (=72 + 2) models of other types. These results indicate the need to distinguish and analyze the performance of 'P-P' type models from the other types of models. However, for the sake of brevity, we will first provide a brief overview of our findings from examining the question, "What distinguishes 'P-P' type models from the rest?" Next, we will shift our focus to the 'P-P' type models and explore the question, "What features do the 'P-P' type models have?".

- In the case of the Dow Jones, when comparing the three parameter values ($n_1$, $n_2$, $n_3$) used in the 66 'P-P' models with those of the other 14 models, we identified three distinct features: (1) Almost all (92.9%) of the 14 models concentrated their first parameter values in the shorter range {4, 5}, whereas a significant number (71.2%) of the 66 'P-P' models had longer values {6, 7, 8}; (2) The average difference between $n_1$ and $n_2$ parameters in the 14 models was generally wider, averaging 14.7, compared to 11.9 in the 'P-P' models; and (3) The average value of parameter $n_3$ in the 14 models was shorter than that of the 'P-P' models, i.e., 6.64 vs. 7.68. In summary, these findings indicate that the 14 models were driven by trading signals generated by the MACD line "with a shorter $n_1$ and a longer interval between $n_1$ and $n_2$ values", as well as the Signal line derived from averaging the "relatively shorter period of the MACD series".
- A similar observation was also made for the sample models for the Nasdaq. Therefore, we can conclude that a large number of models, excluding the 'P-P' type models in the two U.S. markets, have reacted to the relatively short-term recent market movements, suggesting that employing such short parameter values or parameter value setting is not suitable for achieving consistent positive returns in the two U.S. markets.
- Regarding the ratio of the 'P-P' models (65.7%) in the Nasdaq, which is lower than the ratios in the preceding markets (99.0% for the Nikkei and 82.5% for the Dow Jones), the following findings were obtained: (1) This difference is attributed to the wider range of their second parameter values in the Nasdaq compared to the other two markets, i.e., $\{12, \cdots, 20\}$ vs. $\{8, \cdots, 10\}$ and $\{18, \cdots, 21\}$; (2) An increase in the cut-off point for identifying frequently used parameter values as optimal ones results in a higher ratio of 'P-P' models and a better mean return. For instance, raising the point from the current 7% for this market to 8% results in a narrower and higher range of values for $n_2$, changing it from $\{12, \cdots, 20\}$ to $\{15, \cdots, 20\}$. As a result, the current ratio of 65.7% (=142/216) and the total mean return of '4,347.8′ (which will be presented in Table 4) for the 'P-P' models over the entire test period goes up to 76.7% (=46/60) and '4518.9′, respectively. This suggests that narrowing the range of the second parameter values to those more frequently used, higher values can potentially lead to the discovery of better-performing models in the Nasdaq.

**Table 3.** Classification by performance of the sample models within $A_{n1} - A_{n2} - A_{n3}$ group (2011–2021).

| | P-P | | P-N | | N-P | | N-N | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | n | Ratio | n | Ratio | n | Ratio | n | Ratio | n | Ratio |
| Nikkei | 101 | (99.0%) | 0 | (0.0%) | 1 | (1.0%) | 0 | (0.0%) | 102 | (100.0%) |
| Dow Jones | 66 | (82.5%) | 11 | (13.8%) | 1 | (1.3%) | 2 | (2.5%) | 80 | (100.0%) |
| Nasdaq | 142 | (65.7%) | 0 | (0.0%) | 72 | (33.3%) | 2 | (0.9%) | 216 | (100.0%) |

Note: The "P-P" (or "N-N") column shows the number of models with consistent positive (or negative) returns in both the in-sample and the out-of-sample tests. The "P-N" (or "N-P") column indicates the number of models with positive (or negative) returns in the in-sample tests but negative (or positive) returns in the out-of-sample tests.

**Table 4.** Profitability analysis of long and short trades for P-P models.

| | In-Sample Test (2011–2019) | | | Out-of-Sample Test (2020–2021) | | | Whole Test Period | | |
|---|---|---|---|---|---|---|---|---|---|
| | Long | Short | Subtotal | Long | Short | Subtotal | Long | Short | Total |
| [a] Nikkei | | | | | | | | | |
| Mean return | 9491.1 | −5203.4 | 4287.7 | 5580.7 | 691.9 | 6272.6 | 15,071.8 | −4511.5 | 10,560.3 |
| Average number of trades (%) | 189.6 (50%) | 189.3 (50%) | 378.9 (100%) | 39.2 (50%) | 39.2 (50%) | 78.4 (100%) | 228.8 (50%) | 228.5 (50%) | 457.3 (100%) |
| [b] Dow Jones | | | | | | | | | |
| Mean return | 9432.1 | −6671.8 | 2760.3 | 5135.7 | −1771.4 | 3364.3 | 14,567.9 | −8443.3 | 6124.6 |
| Average number of trades (%) | 125.6 (50%) | 125.2 (50%) | 250.8 (100%) | 28.4 (50%) | 28.8 (50%) | 57.2 (100%) | 154.0 (50%) | 154.0 (50%) | 308.0 (100%) |
| [c] Nasdaq | | | | | | | | | |
| Mean return | 3571.9 | −2349.3 | 1222.6 | 5096.1 | −1970.9 | 3125.2 | 8668.0 | −4320.3 | 4347.8 |
| Average number of trades (%) | 112.5 (50%) | 112.8 (50%) | 225.3 (100%) | 24.5 (50%) | 24.9 (50%) | 49.3 (100%) | 136.9 (50%) | 137.7 (50%) | 274.6 (100%) |

These findings have many practical implications for investors in the three markets. However, our focus will be on assessing how well the 'P-P' models performed in each of the three index markets and what characteristics the 'P-P' models have.

### *5.4. Performance of P-P Models*

See the right-most "Total" column in Table 4, which shows that the 101 'P-P' models for the Nikkei produced a total mean return of '10,560.3' over the two sample test periods. Similarly, the 66 'P-P' models for the Dow Jones and the 142 'P-P' models for the Nasdaq yielded total mean returns of '6124.6'and '4347.8', respectively. These results represent a 0.1%, 31.6%, and 15.3% increase in mean returns compared to the overall mean returns generated by all four types of models including 'P-P' models—'10,546.1' (=4244.5 + 6301.6) for the Nikkei, '4654.6' (=2512.8 + 2441.8) for the Dow Jones, and '3777.0' (=591.6 + 3185.4) for the Nasdaq—which were presented in Tables 1 and 2. This implies that restricting our focus to the 'P-P' models alone has merit.

One thing to recall here is that the trading rule applied to all the sample models in this study is to alternate between a 'long' and a 'short' trade in sequence. As a result, the total number of transactions executed by the 'P-P' models comprises nearly equal proportions of long and short trades, as Table 4 shows. This prompts the need to assess the profitability of the 'P-P' models separately for long trades and short trades.

Focusing on the results for the Nikkei, we can observe that, on average over the entire testing period, each of the 'P-P' models lost nearly one-third (29.9%) of their earnings (15,071.8) from long trades due to losses (−4511.5) incurred in short trades. These findings indicate that the profitability of the 'P-P' models in the Japanese market is significantly higher for long trades compared to short trades. This is a novel aspect that was not discussed in the previous study by Kang (2022).

Similar observations can also be made for the two U.S. markets. The 66 'P-P' models for the Dow Jones lost more than half (58.0%) of their earnings (14,567.9) from long trades due to their losses (–8443.3) from short trades. Likewise, the 142 'P-P' models for the

Nasdaq lost almost half (49.8%) of their earnings (8668.0) from long trades due to their losses (−4320.3) from short trades.

Here, a natural question arises: "Why do the 'P-P' models perform significantly better in long trades?" One may argue that, in markets with a clear long-term uptrend, as seen in Figure 2, taking a long position tends to be more profitable than taking a short position. However, a better understanding of the performance of the 'P-P' models requires further investigation.

### 5.5. Market Momentum and the Performance of the P-P Models

We now examine the profitability of the 'P-P' models in relation to the market momentum of the three index markets (Table 5). The row labeled "*Up*" ("*Down*") represents the total amount of increase (decrease) in each of the three index values per day when the index value on the previous day went up (down) on the following day. Based on these results, we introduce a new simple momentum index, denoted as $M$, which measures the overall annual market momentum. It is defined as the ratio of "*Up*" to "*Down*":

$$M = Up/Down \tag{2}$$

This index can be interpreted as follows: (1) If $M > 1$, it signifies a stronger bullish trend in the market compared to the bearish trend, suggesting a robust rising uptrend that is more significant than the downtrend during the corresponding year. Therefore, it indicates that long trades are more favorable than short trades. (2) If $M < 1$, it indicates a stronger bearish trend, suggesting short trades are more favorable than long trades.

**Table 5.** Profitability of 'P-P' models in relation to market momentum for each index market.

| | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [a] Nikkei | | | | | | | | | | | | |
| *Up* | 10,550 | 10,080 | 24,750 | 18,630 | 23,460 | 25,400 | 16,220 | 23,430 | 19,400 | 33,050 | 34,100 | 21,576.4 |
| *Down* | 12,490 | 8090 | 18,860 | 17,590 | 21,820 | 25,310 | 12,560 | 26,340 | 15,600 | 29,220 | 32,700 | 19,931.8 |
| *M* | 0.84 | 1.25 | 1.31 | 1.06 | 1.08 | 1.00 | 1.29 | 0.89 | 1.24 | 1.13 | 1.04 | 1.10 |
| Long | −1338 | 1571 | 3019 | 1133 | 1664 | 241 | 852 | −14 | 2363 | 2723 | 2858 | 1370.2 |
| Short | 619 | −290 | −2490 | −1619 | 293 | −226 | −2969 | 2355 | −877 | −577 | 1269 | −410.1 |
| L&S | −719 | 1280 | 529 | −486 | 1957 | 15 | −2117 | 2341 | 1487 | 2146 | 4127 | 960.0 |
| [b] Dow Jones | | | | | | | | | | | | |
| *Up* | 14,457 | 9893 | 10,995 | 11,168 | 15,033 | 13,295 | 10,636 | 24,481 | 21,765 | 46,012 | 28,310 | 18,731 |
| *Down* | 13,892 | 9016 | 7528 | 9910 | 15,444 | 10,916 | 5621 | 25,948 | 16,525 | 44,023 | 22,581 | 16,491 |
| *M* | 1.04 | 1.10 | 1.46 | 1.13 | 0.97 | 1.22 | 1.89 | 0.94 | 1.32 | 1.05 | 1.25 | 1.22 |
| Long | −916 | −1042 | 2992 | 1462 | −388 | 1284 | 2664 | −777 | 4154 | 5414 | −278 | 1324.4 |
| Short | −908 | −1882 | −126 | −534 | −15 | −1128 | −1719 | 875 | −1234 | 4860 | −6632 | −767.6 |
| L&S | −1824 | −2925 | 2866 | 927 | −403 | 156 | 945 | 98 | 2920 | 10274 | −6910 | 556.8 |
| [c] Nasdaq | | | | | | | | | | | | |
| *Up* | 3097 | 2662 | 2740 | 3532 | 4686 | 4129 | 4138 | 8810 | 8662 | 20,873 | 17,719 | 7368 |
| *Down* | 3073 | 2282 | 1811 | 2883 | 4331 | 3853 | 2593 | 8886 | 6243 | 16740 | 14284 | 6089 |
| *M* | 1.01 | 1.17 | 1.51 | 1.23 | 1.08 | 1.07 | 1.60 | 0.99 | 1.39 | 1.25 | 1.24 | 1.23 |
| Long | −14 | 76 | 273 | 565 | 428 | 361 | 490 | −68 | 1460 | 1772 | 3324 | 788.0 |
| Short | 8 | −274 | −662 | −107 | 54 | 67 | −754 | 11 | −693 | −1447 | −524 | −392.8 |
| L&S | −5 | −198 | −389 | 458 | 482 | 429 | −264 | −56 | 767 | 326 | 2799 | 395.3 |

Note: The row labeled "Long", "Short", and "L&S" represents the average annual returns for 'long', 'short', and both 'long and short' trades executed by the 'P-P' models. The returns of these three types of trades are rounded to integers.

Let us examine the results for the Nikkei in Table 5. The market momentum ($M$) for the two years 2011 and 2018 is less than '1' (0.84 and 0.89), while for the other 9 years, it exceeds '1'. This implies that short trades are more profitable in those 2 years, whereas long trades are more favorable in the other years. This can be verified from the "Long" and "Short" rows, where the 101 'P-P' models yielded higher returns from short trades than from long trades in 2011 and 2018 (i.e., $-1388 < 619$, $-14 < 2355$). Conversely, in the other years, the situation was reversed. These outcomes indicate that the 'P-P' models performed well in short trades when $M < 1$, and in long trades when $M > 1$. These findings collectively demonstrate the effectiveness of the 101 'P-P' models as trend-following trading systems to effectively capture market trends. This aspect holds significant importance as it supports the profitability of the 'P-P' models and the optimality of their parameter value settings.

In the case of the Dow Jones, the years with market momentum less than '1' are 2015 and 2018 (0.97 and 0.94). During these 2 years, the 66 'P-P' models for this market exhibited smaller losses and higher returns for short trades compared to long trades (i.e., $-388 < -15$, $-777 < 875$). In the other years, the situation was reversed. Similarly, in the Nasdaq, only the year 2018 has market momentum less than '1' (0.99). Thus, the 142 'P-P' models for this market achieved greater returns from short trades than from long trades during that year (i.e., $-68 < 11$). These outcomes confirm that the 'P-P' models for the two U.S. markets also serve as effective trend-following systems under varying market conditions.

Next, let us see how the 'P-P' models performed during the 2020 pandemic. Looking at the "*Up*" and "*Down*" rows for the year, the total amounts of increase and decrease in the three index values are much higher compared to any previous year. This reflects the extraordinary fluctuations caused by the pandemic during that year.

Delving into the performance of the 'P-P' models for the Nikkei in the pandemic year, the 101 'P-P' models achieved an average positive return of '2723' from 'long' trades, while incurring an average negative return of '−577' from 'short' trades. However, their total mean return from both 'long and short' trades remained positive at '2146'. This demonstrates the robust performance of the 'P-P' models during the crisis as they were able to offset losses from 'short' trades. Shifting to the Dow Jones, the 66 'P-P' models exhibited positive returns for both long trades (5414) and short trades (4860). As a result, their total mean return from both trade types reached a surprisingly high '10,274'. Lastly, in the case of the Nasdaq, the 142 'P-P' models generated a positive mean return from long trades (1772) and a negative mean return from short trades (−1447), akin to the Nikkei. Nevertheless, the combined total mean return from both long and short trades remained positive at '326'. These results demonstrate that the 'P-P' models for each of the two U.S. markets also performed well even during the pandemic crisis.

Based on these findings, we have made another interesting discovery. When we calculated the total average return of 'long' trades for the last 2 years and compared it with that of the first 9 years, we obtained: '2790.5 >1054.6' for the Nikkei, '2568.0 > 1048.1' for the Dow Jones, and '2548.0 > 396.8' for the Nasdaq. These results indicate that the 'P-P' models for each of the three markets were more profitable during the "abnormal" crisis period of 2020–2021 compared to the "normal" period of 2011–2019. Interestingly, this pattern also holds true for both 'long and short' trades, with '3136.5 > 476.3' for the Nikkei, '1682.0 > 306.7' for the Dow Jones, and '1562.5 > 136.0' for the Nasdaq. One possible interpretation may be that the excess volatility incurred by the COVID-19 pandemic has contributed to increasing the effectiveness of technical analysis, i.e., the performance of the 'P-P' models.

### *5.6. Two Test Results for Robustness*

One remaining issue is whether the 'P-P' models outperform the classic 'buy-and-hold' strategy. However, we need to consider that, in markets with a long-term uptrend, long trades tend to be more profitable than short trades. This means that directly comparing the performance of the 'P-P' models (which involve both 'long' and 'short' trades) with that of the buy-and-hold strategy (which assumes a 'long' trade only over an extended period) would not be equitable. To address this concern, we focused on comparing the returns from long trades achieved by the 'P-P' models with those of a monthly-based buy-and-hold strategy that follows the simple rule to purchase a futures contract at the start of each month and sell it at the end of the month. The reason for considering this monthly-based benchmark strategy was to ensure a fair comparison with the 'P-P' models, aligning the number of trades as closely as possible with the 'long' trades executed by the 'P-P' models.[3]

However, conducting a statistical significance test for annual mean returns between the 'buy-and-hold' strategy and the 'P-P' models is not suitable as a test of the difference in means. This is because we can obtain only a single sample of annual return from the former, which aggregates 12 monthly returns, while we have multiple individual samples from the latter. Therefore, we examined each annual return of the former and focused on how many 'P-P' models outperformed the buy-and-hold strategy in individual years.

When focusing on the boldface numbers in Table 6[b], the following observation can be made: (1) In the case of the Nikkei, the 101 'P-P' models in 2011, 2018, and 2021 perfectly or almost perfectly outperformed the buy-and-hold strategy not only for 'long' trades but also for the combined 'long and short' trades; (2) For the Dow Jones, the 66 'P-P' models in 2013, 2015, 2018, and 2020 exhibited such superior performance, as mentioned above; (3) For the Nasdaq, the 142 'P-P' models in 2015, 2016, 2018, and 2021 showed similar results; and (4) Outside of the years shown in boldface, many 'P-P' models outperformed the benchmark strategies in various years (e.g., 58 models in 2014 for the Nikkei).

To summarize these results, while there is no 'P-P' model that consistently generated significant returns over the entire test period, there are numerous 'P-P' models that outperformed the buy-and-hold strategy in individual or consecutive years.

**Table 6.** Performance comparison between buy-and-hold strategy and P-P models.

| | | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [a] Annual returns of the monthly buy-and-hold strategy | | | | | | | | | | | | | |
| Nikkei | | −2470 | 2310 | 5580 | 990 | 2190 | 1100 | 1510 | −3260 | 3370 | 4470 | −960 | 1348.2 |
| Dow Jones | | 1409 | 825 | 2675 | 1906 | −836 | 2187 | 4132 | −1378 | 5541 | 1264 | 4486 | 2019.1 |
| Nasdaq | | 180 | 308 | 728 | 759 | 178 | 183 | 1449 | −225 | 2578 | 3405 | 2748 | 1117.4 |
| [b] Numbers of P-P models with higher returns than the buy-and-hold strategy | | | | | | | | | | | | | |
| Nikkei | Long | **101** | 0 | 0 | 58 | 13 | 25 | 13 | **101** | 2 | 1 | **101** | 37.7 |
| | L&S | **101** | 0 | 0 | 3 | 50 | 31 | 0 | **101** | 1 | 8 | **99** | 21.8 |
| Dow | Long | 0 | 0 | **57** | 0 | **57** | 0 | 0 | **66** | 1 | **66** | 0 | 22.5 |
| | L&S | 0 | 0 | **47** | 0 | **43** | 0 | 0 | **66** | 1 | **66** | 0 | 20.3 |
| Nasdaq | Long | 13 | 0 | 0 | 1 | **136** | **136** | 0 | **121** | 0 | 0 | **142** | 49.9 |
| | L&S | 18 | 0 | 0 | 5 | **123** | **121** | 0 | **104** | 0 | 0 | **73** | 40.4 |

Note: Panel [a] presents the annual returns of the monthly buy-and-hold strategy in the three markets. In panel [b], the row labeled "Long" ("L&S") indicates the number of 'P-P' models for which annual returns from 'long' (and 'long and short') trades exceed the annual returns of the buy-and-hold strategy.

The next issue is whether the 'P-P' models outperform the 'random' trading strategy mentioned in Section 3.4. In order to assess this, we employed a monthly-based random strategy, identical to the monthly buy-and-hold strategy, as follows: (1) It initiates a buy order for one trading unit on a randomly chosen day within a month and closes the position on another randomly chosen day within the same month. (2) The two days for opening and closing a long position are determined by generating two random numbers from a uniform distribution.

Table 7 presents the empirical results for the random strategy in the same format as Table 6 to ensure comparability. The results in panel [a] display the annual average returns of the random strategy for the three markets that were obtained through 1000 simulations of this strategy. Examining the right-most "Mean" column in panel [a], it becomes immediately evident that the profitability of the random strategy is significantly lower compared to that of the 'P-P' models from long trades shown in Table 5 (i.e., '750.5 < 1370.2' for the Nikkei, '694.8 < 1324.4' for the Dow Jones, and '502.3 < 788.0' for the Nasdaq). Focusing on the boldface numbers presented in panel [b], we can make the same observation as before: while there is no 'P-P' model that consistently generated significant returns over the entire test period, numerous 'P-P' models outperformed the random strategy in individual years or consecutive years.

**Table 7.** Performance comparison between random trading strategy and P-P models.

| | | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [a] Annual returns of the monthly buy-and-hold strategy | | | | | | | | | | | | | |
| Nikkei | | −1137 | 1205 | 2337 | 320 | 1183 | 1117 | 698 | −1275 | 1961 | 1103 | 744 | 750.5 |
| Dow Jones | | 244 | 298 | 1223 | 720 | −36 | 1033 | 1559 | −887 | 2616 | −922 | 1795 | 694.8 |
| Nasdaq | | −27 | 120 | 231 | 327 | 238 | 246 | 547 | −217 | 1077 | 995 | 1459 | 502.3 |
| [b] Numbers of P-P models with higher returns than the buy-and-hold strategy | | | | | | | | | | | | | |
| Nikkei | Long | 44 | **95** | **96** | **101** | 74 | 25 | 57 | **91** | 76 | **93** | **94** | 76.9 |
| | L&S | 49 | 63 | 2 | 13 | 70 | 31 | 1 | **99** | 27 | **77** | **90** | 47.5 |
| Dow | Long | 0 | 0 | **66** | **66** | 11 | **65** | **66** | 40 | **66** | **66** | 0 | 40.5 |
| | L&S | 0 | 0 | **66** | **50** | 25 | 0 | 11 | **64** | 37 | **66** | 0 | 29.0 |
| Nasdaq | Long | 45 | 40 | 117 | **142** | **136** | **123** | 58 | **121** | **142** | **142** | **142** | 109.8 |
| | L&S | 44 | 0 | 0 | **98** | **103** | **113** | 0 | **104** | 16 | 0 | **142** | 142.0 |

In addition to the results presented above, we conducted year-by-year statistical significance tests for the performance of the 'P-P' models and the random strategy. Regarding the performance of the random strategy, we utilized the above-mentioned 1000 simulations outcomes which had converged to provide sufficiently reliable expected values for its annual mean returns. Specially, we compared the annual log returns of the 'P-P' models in long trades ($n_1$ = 101 for the Nikkei, $n_1$ = 66 for the Dow Jones, and $n_1$ = 142 for the Nasdaq) with the annual log returns of the random strategy ($n_2$ = 1000) for each individual year by employing a standard $t$-test to the two groups. The null hypothesis tested was the equality between the mean return of the 'P-P' models ($\overline{x}_1$) and that of the random strategy ($\overline{x}_2$), represented as $H_o$: $\overline{x}_1 = \overline{x}_2$, against the alternative hypothesis $H_1$: $\overline{x}_1 > \overline{x}_2$.

Table 8 shows that the null hypothesis is rejected at the 1% or 5% significance level for 9 years in the Nikkei, 6 years in the Dow Jones, and 7 years in the Nasdaq. This provides robust evidence of a highly significant difference in the annual mean return performance between the 'P-P' models and the random strategy in these specific years. The boldface results in panel [b] of the preceding Table 7 corroborate these findings, indicating that almost all of the 'P-P' models outperformed the random strategy for long trades in those years.

Considering these results, we can conclude that, overall, the 'P-P' models possess high potential to outperform the random strategy. This finding contradicts the conclusion reached by Biondo et al. (2013) that "the random strategy is less risky than the considered standard trading strategies," which was based on their discovery that "standard trading strategies (including the traditional MACD (12, 26, 9) model), may occasionally be successful within small temporal windows, but on a large temporal scale perform no better on average than the purely random strategy". This suggests that employing different parameter values in their models, rather than the conventional ones, might have led to different conclusions. This is another intriguing aspect that emerges from this study.

**Table 8.** $t$-test outputs for P-P models and random strategy.

| | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| [a] Nikkei | | | | | | | | | | | |
| $\overline{x}_1$ (P-P) | −0.1525 | 0.1692 | 0.2263 | 0.0687 | 0.0912 | 0.0087 | 0.0447 | −0.0042 | 0.1105 | 0.1229 | 0.0980 |
| $\overline{x}_2$ (Random) | −0.1210 | 0.1293 | 0.1716 | 0.0199 | 0.0637 | 0.0636 | 0.0328 | −0.0601 | 0.0904 | 0.0411 | 0.0268 |
| $t$-statistic | −2.944 | 4.796 ** | 4.964 ** | 5.840 ** | 3.181 ** | −4.902 | 2.304 * | 6.085 ** | 3.549 ** | 5.597 ** | 9.007 ** |
| [b] Dow Jones | | | | | | | | | | | |
| $\overline{x}_1$ (P-P) | −0.0858 | −0.0819 | 0.2018 | 0.0872 | −0.0234 | 0.0710 | 0.1176 | −0.0316 | 0.1619 | 0.2116 | −0.0092 |
| $\overline{x}_2$ (Random) | 0.0212 | 0.0238 | 0.0814 | 0.0427 | −0.0029 | 0.0555 | 0.0702 | −0.0380 | 0.1002 | −0.0383 | 0.0542 |
| $t$-statistic | −10.887 | −16.058 | 20.488 ** | 6.802 ** | −2.744 | 2.207 * | 11.481 ** | 0.623 | 10.517 ** | 13.034 ** | −10.196 |
| [c] Nasdaq | | | | | | | | | | | |
| $\overline{x}_1$ (P-P) | −0.0068 | 0.0366 | 0.0821 | 0.1427 | 0.0965 | 0.0788 | 0.0858 | −0.0064 | 0.2015 | 0.1972 | 0.2286 |
| $\overline{x}_2$ (Random) | −0.0118 | 0.0486 | 0.0739 | 0.0838 | 0.0534 | 0.0519 | 0.0965 | −0.0329 | 0.1415 | 0.0905 | 0.1009 |
| $t$-statistic | 0.590 | −1.954 | 1.585 ^ | 9.974 ** | 6.382 ** | 4.628 ** | −2.775 | 3.313 ** | 10.950 ** | 8.706 ** | 21.015 ** |

Note: **, *, ^ denote statistical significance at the 1%, 5%, and 10% level, respectively, for one-tailed tests.

### 5.7. Market Returns and the Performance of the P-P Models

For the performance of the 'P-P' models in the preceding section, we reassess their performance against the market returns of each market, as Figure 3 illustrates.

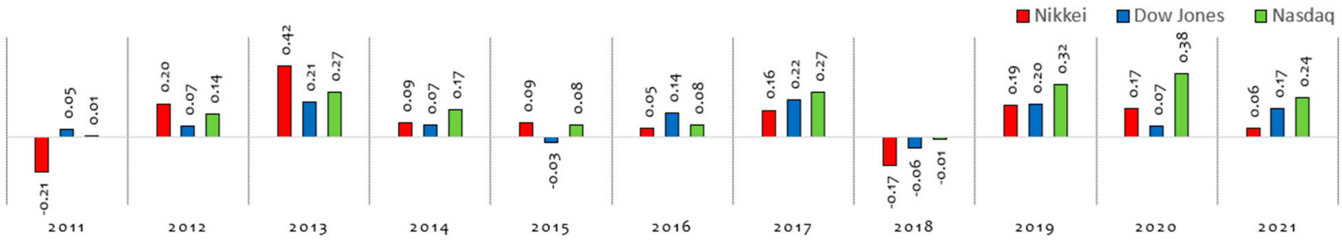


**Figure 3.** Annual market returns of Nikkei, Dow Jones, and Nasdaq. Note: The annual market returns for the three indices are obtained from the natural logarithm of the ratio between the year-end index value and the year-beginning index value.

Let us first examine the differences in each market as seen from market returns. In 2011, the Nikkei had a negative return (−0.21), while the Dow Jones and the Nasdaq both yielded positive returns. In 2012, the Nikkei showed a higher return (0.20) compared to the two U.S. markets. In 2015, the Dow Jones was the only market with a negative return (−0.03), but the situation reversed in 2016 when it recorded the highest return (0.14). In 2018, all three markets experienced negative returns, although the extent of decline varied.[4] These differences in market returns are the reasons we selected these markets to assess the effectiveness of the new methodology.

Calculating the Pearson's correlation coefficient between the annual market returns of the Nikkei and the Dow Jones (Nasdaq) shown in Figure 3, we obtain values of 0.617 (0.729). The coefficient between the market returns of the Dow Jones and the Nasdaq is 0.658. These values indicate a moderate positive correlation among the annual market returns of the three markets, suggesting a degree of relevance in these markets, but not an overly strong one.

Now, let us compare the performance of 'P-P' models to the total average market return of each market. Table 9 shows that the 'P-P' models for the Nikkei achieved a return of 0.0712 on average over the entire test period, which represents approximately 73.6% of the market return. In the Dow Jones, the 'P-P' models' profitability accounts for about 55.5% of the market return, and in the Nasdaq, it accounts for approximately 58.9%. Overall, the 'P-P' models generate returns that, on average, amount to approximately two-thirds or more of the market returns across all three indices.

**Table 9.** Total average annual log returns of P-P Models, buy-and-hold strategy, and random strategy (2011–2021).

| | **Nikkei** | **Dow Jones** | **Nasdaq** |
|---|---|---|---|
| Market return | 0.0967 | 0.1014 | 0.1755 |
| P-P models | 0.0712 (73.6%) | 0.0563 (55.5%) | 0.1033 (58.9%) |
| Buy-and-hold strategy | 0.0801 (82.8%) | 0.0984 (97.0%) | 0.1628 (92.8%) |
| Random strategy | 0.0416 (43.0%) | 0.0336 (33.1%) | 0.0633 (36.1%) |

Note: The "Market return" in the first row is calculated from the annual log returns presented in Figure 3. The total average annual log returns of "P-P models" and "Random strategy" are from the annual log returns in Table 8. The "Buy-and-hold strategy" in the third row represents the total average annual log return, which is based on the annual log returns in Figure 4.

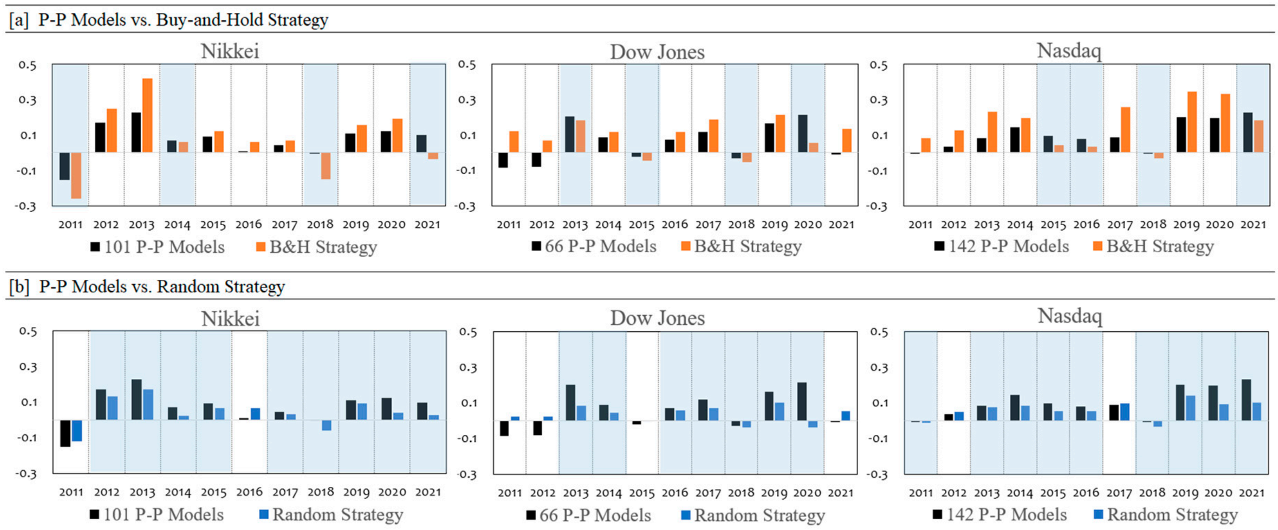


**Figure 4.** Performance comparison: P-P models vs. Benchmark strategies.

It is worth revisiting the results in Table 9. Historically, stock markets have demonstrated an average annual return of around 10%, without adjusting for inflation. For example, the S&P 500 has consistently maintained an average yearly return of approximately 10%, despite occasional fluctuations such as the −2.97% decline in 2015 and the over 20% increase in 2021. Considering these historical norms, the profitability of the 'P-P' models, as shown in Table 9, is noteworthy.

Figure 4 provides a comprehensive comparison of the performance of the 'P-P' models. The blue bands in panels [a] and [b] indicate the years when the 'P-P' models outperformed the buy-and-hold strategy and the random strategy, resulting in higher gains or reduced losses. Note that the returns of the 'buy-and-hold' strategy depicted in panel [a] are annual log returns that were separately calculated using its monthly log returns for this comparison, and the returns of the 'P-P' models and the 'random' strategy depicted in this figure match the log returns presented in Table 8.

### *5.8. Characteristics of Optimal Parameter Values of the P-P Models*

Let us now compare the optimal parameter values for each of the three markets. The dots within the blue shaded areas of Figure 5 represent the ranges of optimal parameter values used in the 'P-P' models for each market. From this figure, the following observations can be made:

- The 'P-P' models for each market have different ranges of three optimal parameter values ($n_1$, $n_2$, $n_3$).
- However, their value combinations all share a common characteristic form: the second parameter value is longer than the other two parameters, i.e., $n_1 < n_2$ and $n_2 > n_3$.
- The second parameter values ($n_2$) used in the 'P-P' models for the two U.S. markets, especially for the Dow Jones, are longer than those for the Nikkei. They fall within the range of $\{18, \cdots, 21\}$ for the Dow Jones, $\{12, \cdots, 20\}$ for the Nasdaq, and $\{8, \cdots, 10\}$ for the Nikkei. Meanwhile, the range of the first parameter values ($n_1$) for the three markets shows no significant difference.

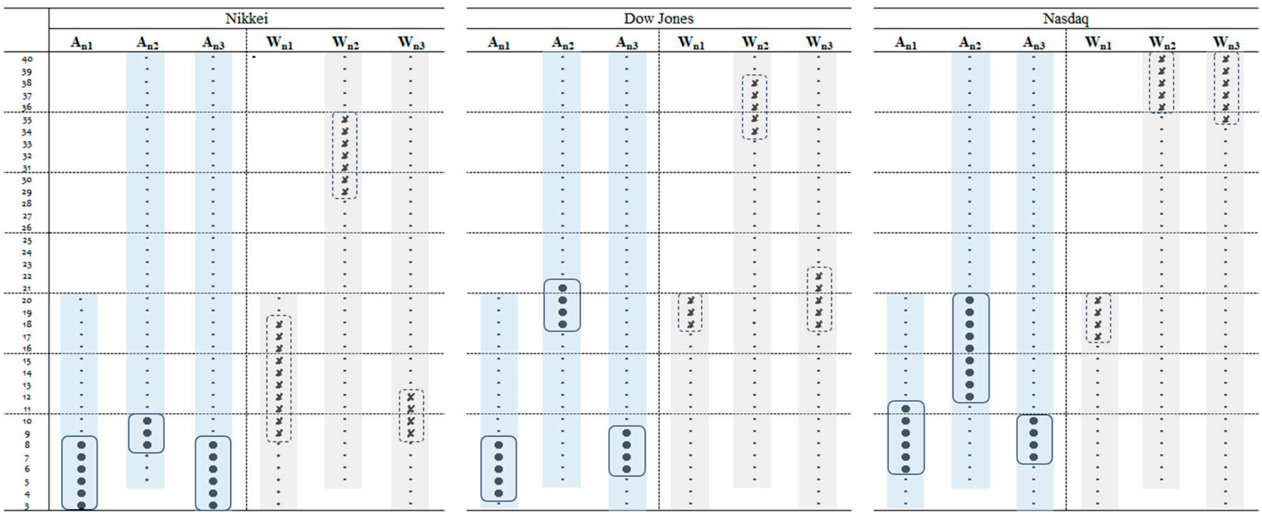


**Figure 5.** Optimal and non-optimal parameter value ranges in Nikkei, Dow Jones, and Nasdaq. Note: The optimal parameter ranges are represented by dots in each rectangle, corresponding to the parameter values used in the 'P-P' models of the $A_{n1} - A_{n2} - A_{n3}$ group. In contrast, non-optimal parameter ranges are depicted by crosses in each rectangle, indicating the parameter values used in the sample models of the $W_{n1} - W_{n2} - W_{n3}$ group. Blue bands and gray bands indicate the whole parameter value ranges examined in this study.

The first two findings, related to the different ranges of optimal parameter values and their common characteristic form in their parameter value combinations, seem to reflect differences among the three markets and their global synchronization. The second finding is the most interesting result of this study, suggesting that adopting this characteristic form may lead to favorable outcomes across all three markets. Additionally, the third finding highlights that the intervals between the first two parameters ($n_1$ and $n_2$), which makes up the MACD line, are relatively wider in 'P-P' models for the two U.S. markets compared to those in the Japanese market. To better understand this discrepancy, we examined the 'mean market momentum' values in Table 5 and the 'total average market returns' in Table 9 for the three markets. Notably, the Japanese market exhibits lower values compared to the two U.S. markets, with $1.10 < 1.22 < 1.23$ for market momentum and $0.0967 < 0.1014 < 0.1755$ for market returns, respectively. Considering these market distinctions, it can be inferred that a wider interval between the first two optimal parameters could be advantageous in markets with higher momentum and growth potential as they contribute to reducing the MACD's sensitivity to minor market or price fluctuations.

Consider the crosses within the gray shaded areas of Figure 5, which represent the ranges of "non-optimal" parameter values in the three markets. Looking at the Nikkei and the Dow Jones, it is immediately evident that these ranges exhibit the same characteristic form as the optimal parameter values, with $n_1 < n_2$ and $n_2 > n_3$. This implies that, even if we construct a model following the same pattern as the optimal parameter combinations but use values within these non-optimal ranges, negative outcomes may result; this was demonstrated earlier in this study. This highlights the importance of analyzing both optimal and non-optimal parameter value ranges.

## 6. Concluding Remarks

This study presented the first country-by-country analysis of the optimal and non-optimal parameter value ranges of the MACD technical indicator in three major index futures: the Nikkei 225 in Japan, and the Dow Jones and the Nasdaq in the United States. Evidence from all three indices confirmed the strong performance of models using optimal parameter values determined by the new methodology, referred to as 'P-P' models in this paper. Specifically: (1) 'P-P' models, totaling 101 models for the Nikkei, 66 models for the

Dow Jones, and 142 models for the Nasdaq, consistently produced positive returns in both in-sample and out-of-sample tests. (2) Conversely, nearly all sample models employing 'non-optimal' parameter values delivered notably negative returns in both in-sample and out-of-sample tests across all three markets.

Therefore, we can conclude: (1) There is considerable validity in the new methodology for finding good-performing models with "optimal" parameter values and identifying poor-performing models with "non-optimal" parameter values. (2) Its validity is verified not only for the Japanese market but also for the two U.S. markets.

Several extended analyses provided more interesting findings on the performance of the 'P-P' models:

- Throughout the entire test period from 2011 to 2021, the 'P-P' models for each of the three markets experienced a substantial reduction in their 'long' trade earnings due to losses in 'short' trades. This reduction amounted to approximately 29.9% in the Nikkei, 58.0% in the Dow Jones, and 49.8% in the Nasdaq. However, this result is not surprising or unnatural, considering that long trades tend to be more favorable than short trades in markets with long-term uptrends as commonly observed in these markets.
- During the pandemic crisis in 2020, 'P-P' models in all three markets showed notably higher mean returns in 'long' trades compared to 'short' trades. When comparing the total average return for the last 2 years with that of the first 9 years, the 'P-P' models in all three markets were more profitable during the crisis period of 2020–2021 compared to the normal period of 2011–2019.
- The extended analysis combined with market momentum ($M$) revealed that the 'P-P' models for each of the three markets serve as effective trend-following trading systems to capture market trends under varying market conditions for the last 11 years.

In summary, these results demonstrate the effectiveness of the 'P-P' models in each of the three markets throughout the entire test period, including crisis periods. However, the true significance of this study lies in uncovering "the existence of specific ranges of optimal and non-optimal values" for each of the three parameters of the MACD indicator in the three major indices. This discovery goes beyond merely validating the efficacy of the new methodology and deserves attention, as it has not been addressed in the existing literature. It holds significant implications for both market participants and researchers.

For traders, the identified optimal parameter ranges offer helpful guidance to customize their MACD models to align the frequency of signal generation with their preferred trading style and objectives, whether they prefer short-term or long-term investments, by selectively choosing parameter values from the specified ranges in Figure 5. See the Appendix A for straightforward recommendations and practical guidelines for individual traders. For market researchers, this study explores distinctions among the three major markets. Researchers can gain insights by examining the 'P-P' model performances under varying market momentum and market return conditions in each market (Section 5.7) and by examining the differences and commonalities in the optimal parameter value ranges (Section 5.8) to enhance their understanding of each market's unique characteristics. Additionally, exploring numerous models with highly performing parameter values within the extensive pool of the 'P-P' models provides a means of making detailed discussions about the effectiveness of the MACD technical system or the existence of market efficiency. This goes beyond the limitations of previous research that often focused on testing only a few models with traditional parameter values or a single model with the most profitable parameters. Discussions based on a large number of highly performing models are more robust and reliable in addressing the aforementioned issues.

The highlight of this study is identifying the difference and the characteristic of the optimal parameter values for the three indices:

- The ranges of optimal parameter values for the three indices are different from each other.

- Despite the differences, the combinations of the three optimal parameter values share a common characteristic form, where the second parameter ($n_2$) has a longer length value than the other two parameters ($n_1$ and $n_3$).

Considering these discoveries, if we conduct further investigations into this possibility for financial markets in other countries or diverse financial assets (e.g., currencies and commodities) in future research, it will offer insightful perspectives that are unique to each market. This, in turn, would contribute to a deeper understanding of the dynamics of global financial markets. To accomplish this, future research could require a more comprehensive and detailed analysis, considering various market dynamics, such as economic conditions, geopolitical events, and sector-specific factors, that could influence the differences and commonalities of the optimal parameter values for each financial market.

Regarding the market efficiency issue of the three markets, it is crucial to consider the key finding that the 'P-P' models delivered significantly higher returns over multiple years, outperforming both the non-technical buy-and-hold strategy and the random strategy, neither of which use any market information. Specifically: (1) Almost all or all of the 'P-P' models for each market outperformed the 'buy-and-hold' strategy and/or the 'random' strategy in various individual or consecutive years. (2) In particular, the 'P-P' models achieved significantly higher annual mean returns than the 'random' strategy for 9 years in the Nikkei, 6 years in the Dow Jones, and 7 years in the Nasdaq, all with high statistical significance.

These results provide strong evidence that the 'P-P' models, employing the simple MACD technical analysis tool, exhibited robust performance over the last 11-year test period in the three index markets. Consequently, these findings raise questions about the efficiency of the Efficient Market Hypothesis (EMH) and suggest that the three markets may not follow weak-form efficiency as they do not fully incorporate all publicly available and relevant information. This implication is consistent with the conclusions drawn from the research of Anghel (2015) where the information efficiency of stock markets was assessed for 1336 companies in 75 countries. His market efficiency ranking revealed that Japan and the United States were among the countries where abnormal profits could be obtained by employing MACD technical trading rules. This further supports the notion that the three index markets examined may not efficiently incorporate past market data and other relevant information into their pricing mechanisms.

**Funding:** Nanzan University provided funding for the APC.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. The historical data for Nikkei 225 futures can be obtained from JPX Data Cloud (http://db-ec.jpx.co.jp), and the data for Dow Jones futures and Nasdaq futures from Investing.com (http://www.investing.com).

**Acknowledgments:** The author would like to acknowledge Marc Bremer for his proofreading and valuable comments and Masayuki Arakawa for reviewing the empirical results in this paper. The author also expresses gratitude for the valuable comments from multiple reviewers.

**Conflicts of Interest:** The author declares no conflict of interest.

## Appendix A. Additional Recommendations and Guidelines for Traders

1. Look at the optimal and non-optimal parameter ranges in Figure 5 and discern which values are optimal and which are not for the three major markets.
   - Suggestion: If the model parameters you are currently using differ from the optimal values in Figure 5, consider modifying them.
2. Select the best parameter value combination that aligns with your preferred trading style and objectives. Note that using short length values increases the frequency of signal generation, and vice versa.
3. Consider which trade between long and short could be favorable under the market momentum of your target market.

- Suggestion: Refer to the empirical results in this study, especially Table 5 for valuable information and insights into the market momentum of the three major index markets and the profitability of long and short trades.

4. For more detailed and prudent adjustments of parameter values, consider utilizing the following two approaches:
   - Sensitivity Analysis: Evaluate the MACD's sensitivity to parameter changes to assess how small adjustments affect the signal generations.
   - Cross-Validation: Apply the parameters to different time frames (i.e., short-, medium-, and/or long-term) to ensure consistency and effectiveness across these periods.
5. However, optimizing the parameters of the MACD model and revising existing trading strategies should be done constantly to reduce false signal generations and enhance its overall performance. This need arises from the inherent lagging nature, a weakness in the MACD. For refinement in such instances, consider employing the new methodology demonstrated in this study as an effective option.
   - Suggestion: Refer to the example of analysis results when raising the minimum cut-off point, provided at the end of Section 5.3. It serves as a helpful guide for understanding the potential impact of such adjustments.

## Notes

[1] On this point, Kang (2022) stated, "It has no logic other than examining the performance of a single parameter combination based on an unreliable assumption that it is the most commonly used. Therefore, it only makes sense if the single examination result is a counterexample that suffices to refute an assertion or a hypothesis such as the EMH. (p. 2)".

[2] In Kang (2022), the minimum cutoff point for identifying the "primary group of optimal" parameter values was set at 4 percent. Parameter values falling within the range of 2 or 3 percent were categorized as the "secondary group of sub-optimal" values. This led to the generation of too many groups of sample models to consider, as mentioned in the first section.

[3] Looking the average annual number of long trades of the 'P-P' models, as shown on the right side of Table 4, we observe a value of '228.8' for the Nikkei. This corresponds to an average of approximately 1.73 (=228.8/(11 years $\times$ 12 months) long trades per month. Similarly, for the Dow Jones and the Nasdaq, we have approximately 1.17 and 1.04 long trades per month, respectively.

[4] The same observations can be made for the "market momentum" values presented in Table 5. This suggests that the market momentum index defined in this study effectively corresponds to the "market return" values in Figure 3.

## References

Abbey, Boris, and John Doukas. 2012. Is technical analysis profitable for individual currency traders? *Journal of Portfolio Management* 39: 142–50. [CrossRef]

Anghel, Gabriel Dan. 2015. Stock market efficiency and the MACD: Evidence from countries around the world. *Procedia Economics and Finance* 32: 1414–31. [CrossRef]

Appel, Gerald. 1979. *The Moving Average Convergence Divergence Trading Method*. Great Neck: Signalert Corp.

Appel, Gerald. 2005. *Technical Analysis: Power Tools for Active Investors*. Upper Saddle River: Financial Times Press.

Biondo, Alessio, Pluchino Alessandro, Rapisarda Andrea, and Helbing Dirk. 2013. Are random trading strategies more successful than technical ones? *PLoS ONE* 8: e68344. [CrossRef] [PubMed]

Bodas-Sagi, Diego, Fernandez Pablo, Hidalgo Ignacio, Soltero Francisco, and Risco-Martin Jose. 2009. Multiobjective optimization of technical market indicators. In Paper presented at GECCO '09: Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, Montreal, QC, Canada, July 8–12.

Borowski, Krzysztof, and Izabela Pruchnicka-Grabias. 2019. Optimal lengths of moving averages for the MACD oscillator for companies listed on the Warsaw Stock Exchange. *Bank i Kredyt (National Bank of Poland)* 50: 457–78. Available online: https://bankikredyt.nbp.pl/content/2019/05/BIK_05_2019_02.pdf (accessed on 10 August 2023).

Chen, Chien-Ping, and Massoud Metghalchi. 2012. Weak-form market efficiency: Evidence from the Brazilian stock market. *International Journal of Economics and Finance* 4: 22–32. [CrossRef]

Chong, Terence Tai-Leung, and Wing-Kam Ng. 2008. Technical analysis and the London stock exchange: Testing the MACD and RSI rules using the FT30. *Applied Economics Letters* 15: 1111–14. [CrossRef]

Chong, Terence Tai-Leung, Wing-Kam Ng, and Venus Khim-Sen Liew. 2014. Revisiting the performance of MACD and RSI oscillators. *Journal of Risk and Financial Management* 7: 1. [CrossRef]

Chou, Chien-Chang, and Keng-Shou Lin. 2019. A fuzzy neural network combined with technical indicators and its application to Baltic Dry Index forecasting. *Journal of Marine Engineering & Technology* 18: 82–91. [CrossRef]

Erić, Dejan, Andjelic Goran, and Redžepagić Srdjan. 2009. Application of MACD and RVI indicators as functions of investment strategy optimization on the financial market. *Journal of Economics and Business* 27: 171–96. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2267753 (accessed on 10 August 2023).

Fama, Eugene. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25: 383–417. [CrossRef]

Hashida, Shuichi, and Keiichi Tamura. 2019. MACD-histogram-based fully convolutional neural networks for classifying time series. Paper presented at 2019 6th International conference on Control, Decision and Information Technologies (CoDIT), Paris, France, April 23–26.

Hejase, Ale, Srour Ruba, Hejase Hussin, and Joumana Younis. 2017. Technical analysis: Exploring MACD in the Lebanese stock market. *Journal of Research in Business, Economics and Management* 8: 1493–502. Available online: https://scitecresearch.com/journals/index.php/jrbem/article/view/1077 (accessed on 10 August 2023).

Kang, Byung-Kook. 2021. Improving MACD technical analysis by optimizing parameters and modifying trading rules: Evidence from the Japanese Nikkei 225 futures market. *Journal of Risk and Financial Management* 14: 37. [CrossRef]

Kang, Byung-Kook. 2022. Finding optimal parameter values for the MACD indicator: Evidence from the Japanese Nikkei 225 futures market using a new methodology. *International Journal of Financial Research* 13: 1–24. [CrossRef]

Meissner, Gunter, Albin Alex, and Kai Nolte. 2001. A refined MACD indicator–Evidence against the random walk hypothesis? *ABAC Journal* 21: 1–17. Available online: https://www.semanticscholar.org/paper/A-Refined-MACD-Indicator-%25E2%2580%2593-Evidence-against-the-Meissner-Alex/53b278167b6a2d53caf37bd790555112eedcd0e8 (accessed on 10 August 2023).

Menkhoff, Lukas. 2010. The use of technical analysis by fund managers: International evidence. *Journal of Banking & Finance* 34: 2573–86. [CrossRef]

Montgomery, Willam, Raza Ahmad, and Ülkü Numan. 2019. Tests of technical trading rules and the 52-week high strategy in the corporate bond market. *Global Finance Journal* 40: 85–103. [CrossRef]

Nor, Safwan Mohd, and Guneratne Wickremasinghe. 2014. The profitability of MACD and RSI trading rules in the Australian stock market. *Investment Management and Financial Innovations* 11: 194–99. Available online: https://www.researchgate.net/publication/272238241_The_profitability_of_MACD_and_RSI_trading_rules_in_the_Australian_stock_market (accessed on 10 August 2023).

Park, Cheol-Ho, and Scott Irwin. 2007. What do we know about the profitability of technical analysis? *Journal of Economic Surveys* 21: 786–826. [CrossRef]

Rosillo, Rafael, David de la Fuente, and Jose Brugos. 2013. Technical analysis and the Spanish stock exchange: Testing the RSI, MACD, momentum and stochastic rules using Spanish market companies. *Applied Economics* 45: 1541–50. [CrossRef]

Wang, Jian, and Junseok Kim. 2018. Predicting Stock Price Trends using MACD Optimized by Historical Volatility. *Mathematical Problems in Engineering* 2018: 9280590. [CrossRef]

Wiles, Phoebe, and David Enke. 2015. Optimizing MACD parameters via genetic algorithms for soybean futures. *Procedia Computer Science* 61: 85–91. [CrossRef]