Finding Optimal Parameter Values for the MACD Indicator: Evidence From the Japanese Nikkei 225 Futures Market Using a New Methodology

Byung-Kook Kang[1]

[1] Department of Business Administration, Nanzan University, Nagoya, Japan

Correspondence: Byung-Kook Kang, Department of Business Administration, Nanzan University, Aichi 466-8673, Japan. Tel: 82-52-832-3111. E-mail: kangbk@nanzan-u.ac.jp

Received: June 15, 2022 Accepted: July 20, 2022 Online Published: July 28, 2022

doi:10.5430/ijfr.v13n3p1 URL: https://doi.org/10.5430/ijfr.v13n3p1

Abstract

This paper explores: (1) what parameter values are most often used to optimize the Moving Average Convergence Divergence (MACD) trading system for the Japanese Nikkei 225 futures market; and, (2) the characteristics of good-performing models with the optimized parameter values. To accomplish this purpose, this paper presents a new methodology to find the three optimal parameter values of the MACD trading system; this approach systematically examines specific ranges of optimal parameter values. Evidence from the Japanese futures market demonstrates the validity of this new methodological approach. From this, we find that for the Japanese market the technical trading system is most often optimized by three parameter values within three specific ranges over the last 11 years (2011–2021). These optimal value combinations have a unique characteristic form. These findings give insightful and broader perspectives about the market. This issue, methodology and the results have not been discussed in the existing literature. This paper also considers how the models with optimal parameter values performed during the pandemic period (2020–2021).

Keywords: MACD, technical analysis, trading simulation, Japanese Nikkei 225 futures, efficient market hypothesis, COVID-19

1. Introduction

The MACD (Moving Average Convergence Divergence) is one of the most popular technical analysis indicators which generates buying and selling trade signals based on analysis of historical stock price data. Technical analysts believe that technical analysis is useful to determine trends and to predict future stock prices. But it is true that the use of technical analysis is often regarded with skepticism, especially from the academic viewpoint, as it is difficult to explain why analyzing past market data can help investors predict future prices. It seems that this skepticism is widespread since Fama (1970) claimed that analysis of past prices cannot deliver profitable forecasts in informationally efficient markets. His efficient market hypothesis (EMH) states that current market prices reflect all past market information even in the case of a weak-form efficient market, so technical analysis based on past market data cannot work. However, it is also true that while academics point to a large body of evidence that supports the hypothesis, a substantial amount of dissension remains. As to the voluminous amount of research supporting the hypothesis, there is no point in reviewing all of it here. Regarding the latter research that rejects the hypothesis, the work of Anghel (2015) is notable as a recently published study. He assessed the state of information efficiency in the stock markets of 75 countries (1268 companies) over the period from 2001 to 2012 by using two MACD-based trading rules. He reported that “Weak form efficiency can be discarded for 34 of the 75 studied markets, when applying Appel’s MACD as an investment technique (p. 1429)” and concluded that “The world’s stock markets present important inefficiencies (p. 1414).” As for the use of technical analysis, Menkhoff (2010) 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).”

As we have seen above, technical analysis is clearly at odds with the weak-form of market efficiency hypothesis. Within this context, many academic researchers have evaluated the effectiveness of the MACD approach or used it to measure market efficiency, or both. However, quite a few of these evaluations fail to verify the effectiveness of the technical approach nor market efficiency based on results obtained with the traditional parameter settings of 12, 26 and 9 days. Section 2 describes some of the literature that reported negative results using the traditional parameter values. In the author's opinion, the skepticism mentioned above can be linked to these negative results. The points to discuss below concern:

- Why is using optimal parameter values necessary and important?
- How best to find the optimal parameter values for a market?
- Why is the most profitable parameter value combination the best optimal?
- What characteristics do high-performing parameter values have?
- How to find the characteristics that optimize the trading system for a specific market?

Let us start with a brief explanation of the indicator for the first question stated above. The MACD consists of three parameter values to define three time-periods (moving averages of historical prices): two for the calculation of the MACD series—which composes a line to illustrate the difference between short- and long- term exponential moving averages, the remaining value is for the calculation of the signal series—which is a line to capture the exponential moving average of the MACD series. In terms of signal generation, it is common to interpret it as follows: buy (sell) when the MACD line crosses up (down) through the signal line—which is called the "signal line crossover" trading rule. Three parameter values are usually represented in the form MACD (n_1, n_2, n_3), e.g., MACD (12,26,9).

As we can see from this, all signals generated by the MACD momentum indicator depend on the three parameter values. Taking this into consideration, a simple question arises for the research that has utilized the traditional parameter values and reported negative results: "Why persist in using the traditional (12,26,9) parameter settings?" In other words, "Why is the model with traditional parameter values the sole determinant of the profitability of all MACD models and the only valid test of market efficiency?" If they had considered other values, they may have reached different conclusions.

One may here ask a question: "What is the rationale for using the traditional values and why consider changing to new values?" Using the traditional parameter values is just the momentum of tradition. Accordingly, the above-mentioned research that has used the traditional parameter values did not provide any rationale for the values, except for vague words such as, 'because it is most commonly used'. It has no logic other than examining the performance of a single parameter combination based on an unreliable assumption that it is 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.

On the other hand, considering new values is essential in using the MACD indicator because, as has already been mentioned, "all signals generated by the MACD indicator depend on the three parameter values." This can be also explained in the following way. The MACD indicator can generate false trade signals since it is a lagging indicator as are other trend following technical analysis tools. But there is no perfect way to predict and eliminate all false signals. One possible approach is changing the three parameter values to control directly the frequency of signal generation. Another approach is using additional trading strategy criteria to modify indirectly the trade signals generated by the trading rule. As these approaches illustrate, adjusting the three parameter values for each market is crucial to avoid false trade signals and improve the performance of the MACD trading system.

On this point, Kang (2021) demonstrated that the application of the traditional MACD (12,26,9) model to Japan's Nikkei 225 futures market produced surprisingly negative returns using 9 years of data (2011–2019). In contrast to this poor result, his research verified that applying the MACD trading rule to the Japanese market can achieve significant positive returns for investors if optimal values are used for the three parameters of the momentum indicator—which suggests that the Japanese market is not weak-form efficient in the sense that futures prices do not reflect all public information. He also tested whether adopting supplemental trading strategies makes it possible to improve profitability by reducing false trade signals. From a simulation using a larger number of sample models with different parameter values, he found that the number of models with improved performance resulting from the supplemental strategies is far greater for models with optimized parameter values than for models with non-optimized values. He therefore concluded that the three parameter values of the MACD tool should be optimized for each market and this should take precedence over finding additional strategies to reduce false trade signals.

Based on this discussion, the author of this research emphasizes that the three parameters of the MACD model delivering the best result possible for a chosen asset or fitting a market should be utilized when we test the effectiveness of the MACD approach or use it to measure market efficiency. Regarding the former—testing the effectiveness of the MACD approach, Park and Irwin (2007) said that “Optimizing trading rules is important because actual traders are likely to choose the best-performing rules in advance (p. 792).” Regarding the latter—testing market efficiency using the MACD, Anghel (2015) said that “The ultimate criterion in determining market efficiency is always practical, so the true market nature, in the sense of information efficiency, cannot be determined until all practical methods have been tested and validated or invalidated on the market (p. 1415).”

Then, the next issue is the second question stated above: “How to find optimal parameter values for a market?” This leads to optimization being employed to select the most successful set of parameter values using past trading data; and then to confirm its validity in trading simulations on out-of-sample data for robustness. Regarding this point, Park and Irwin (2007) pointed out that, “Since there is not a structural form of a technical trading system that pre-specifies parameters, technical trading studies inevitably tend to search over a larger number of parameters (p. 1791).” However, as we will see in the next section, much of the existing literature on MACD indicator usage is biased toward the examination of the profitability of one or two (at best three) models with traditional parameter values such as (12,26,9), (12,26,0) and/or (8,17,9)—although the third model is not so popular. Much of this literature confirms that the MACD trading rule does not produce good results with the three traditional value settings. Among these, only a few papers looked for the optimal parameter values. Two of these are Erić, Andjelic and Redźepagić (2009) and Borowski and Pruchnicka-Grabias (2019). The common approach of these two papers was to examine the profitability of a large number of sample models with different parameter values and to select the most profitable ones. They reported the most profitable parameter value combination for each company listed on the stock market they tested.

However, one may have an issue with this research approach that seeks only the most profitable parameter values: “Why is the most profitable parameter value combination the best optimal?” The set of three parameter values with the highest return is not necessarily the single best, optimal one. For example, if the model with the highest return for a company has a high ratio of unprofitable trades based on the past trading data, we cannot expect that it actually delivers a sustainable trading profit. Another set of parameter values with less risk of having experienced unprofitable trades in the past should be substituted for the most profitable parameter set, even though its profitability in the past is less than the most profitable one. Besides this, other combinations of parameter values potentially can be better for individual traders when the resulting number of trade signals (i.e., the sensitivity of the signal generation), profit acquisition and/or loss avoidance matches one’s investment style and goal. This means that profitability is not the sole determinant of the performance of all MACD models but just a major factor in their evaluation. Therefore, the evaluation criteria for performance in selecting parameters needs to be improved to consider more practical and broadened perspectives. This is not only for individual traders, since it suggests that just as the optimal parameter combination for individual investor can be different from each other, the most optimized parameter values for each market might be different.

On this point, Kang (2021) evaluated the performance of the 19,456 MACD sample models not only based on their returns but also in terms of the rate of profitable trades, the rate of unprofitable trades and the number of total trades from the perspectives of actual traders. As thus, he determined the MACD (4,22,3) model is an optimal model on the Japanese futures market that is balanced in terms of profit acquisition and loss avoidance and the number of transactions—although there were other sample models with higher returns than the MACD (4,22,3) model. It was a new approach to determine optimal values for a market that has not been considered in the existing literature.

But he still had an issue with the other high-performing models that were almost comparable to the most profitable one: "How to explain why those sample models with different parameter values had high performance?” This is a significant question because it is hard to answer—by simply comparing a model with most profitable parameter values in one market to that in another market—that the difference between the two models’ parameter value combinations reflects the two markets’ characteristics. However, for reasons to be described later, he could not answer the question about the other high-performing models. This is the point of this paper.

Before we go on to that, let’s explain briefly why he could not address the issue. Suppose that we evaluate the profitability of a large number of MACD models with different parameter values and focus our attention on the parameter values in a top-ranked group of models. At that time, we often encounter a situation where the three parameter values of the models in the top group have no consistency and relevance to each other because they are just sorted in terms of their profitability. This fact makes it difficult to find any feature that explains the high-performing sets of the three parameter values. We can try an alternative to sorting models by other criteria at that stage but in most cases, it brings us back to the issue. The ambiguous information of the high-performing models (parameter sets) raises these questions:

- Why are those various models with different parameter values able to perform well in the absence of any consistency?
- For the particular market tested, what characteristics do those high-performing models have?

These questions suggest that: if we can find high-performing parameter value sets with a specific feature or regularity for a certain financial market, it may give us insightful new perspectives about the market. For investors who seek a good-performing model, there may be practical insights. For researchers who are interested in market characteristics, there could be an opportunity to deepen one's perspective about the market. However, little attention has been given to a methodological approach to find answers to the questions as stated above.

From these viewpoints, this research presents a new methodology to find good-performing and well-fitted parameter value sets for a market and clarifies what characteristics those parameter value sets have. To accomplish this purpose, this study revisits the performance of the 19,456 MACD sample models that were developed in Kang (2021) for the Japanese futures market. This research re-examines the way to optimize the MACD trading system for the market by using the new methodology.

One of the most interesting findings is that the most frequently used parameter values of the best- and the worst-performing group of the sample models—in the top 100 and the bottom 100 ranked models in profitability—have distinctively different distributions from each other and almost do not overlap even when the sample size increases. Focusing on this finding, we defined some specific ranges of frequently used values in the top (bottom) 100 models as newly analyzed optimal (non-optimal) parameter values. We then construct several groups of new hypothetical sample models with the analyzed optimal (non-optimal) parameter values. Therefore, this paper examines whether the new hypothetical optimal (non-optimal) sample models perform well (poorly). Two of the most remarkable results are:

(1) Almost all of the new hypothetical optimal (non-optimal) sample models achieved positive (negative) returns not only for in-sample tests (2011–2019) but also for out-of-sample tests (2020–2021). Nevertheless, the original models belonging to the top (bottom) 100 group are not or rarely included among the hypothetical sample models.

(2) Among the several groups of the hypothetical optimal sample models, a certain group of sample models for which parameters consist of the values in the pre-specified range— n_1: $\{3,\cdots,8\}$, n_2: $\{18,\cdots,25\}$, n_3: $\{3,\cdots,8\}$—outperform the other groups for both in- and out-of- sample tests.

The first finding demonstrates that the new methodology used in this study has considerable effectiveness in finding optimal parameter values of the MACD trading system in a market—which gives an answer the question: "How to find good-performing and well-fitted parameter values for a market?" The second finding confirms that for the Japanese futures market the trading system is most often optimized by the three parameter values in the three specified ranges as stated above for the last 11 years at least—which answers the question: "What parameter values are most often used to optimize the trading system on the tested market?" Now, look again at the pre-specified range stated in the second item above. We can then see from it that the best combination of the three parameter values (n_1, n_2, n_3) using the values in the three curly brackets makes a characteristic form like a 'top hat' in that the second parameter value n_2 has a longer length than the other two parameters n_1 and n_3.

This paper is not just to present only the most profitable parameter values as other conventional research have done but to present three specific ranges of the three optimal parameter values for the tested market—which makes it possible to select various values and create a model to fit one's trading style and goals. This point will be taken up again in the last section with an explanation for the 'top hat' shaped parameter combinations stated above.

The next section is a brief review of previous research. The third section explains the methodology and data. Section 4 shows how the optimal and non-optimal parameter values are found and determined, and how the new optimal and non-optimal sample models are created. Section 5 presents empirical results. Section 6 provides concluding remarks on the implications of this study.

2. Literature Review

The existing literature on MACD indicator use can be classified into two categories. The first one tests the effectiveness of the MACD approach and measures market efficiency by using the traditional MACD (12,26,9) model or including other technical analysis tools. The other one searches for the most profitable parameter values of the MACD for each of the companies listed on a financial market.

In the first category, Meissner, Alex and Nolte (2001) is the first notable research in the last two decades. They tested whether the traditional MACD (12,26,9) model is profitable. But they found that it results in a poor success rate of about 32 percent for both the DOW 30 stocks and NASDAQ-100 stocks over a 10-year period (1989–1999). They thus concluded that the traditional MACD indicator can almost be regarded as a contra-indicator. Armour, Lofton, Oyenekan and Metghalchi (2010), cited by Anghel (2015), also tested the MACD (12,26,9) trading rule, including a simple moving average rule, on 20 years of data for the Irish Stock Market Main Index and found that the MACD rule did not outperform a benchmark buy-and-hold strategy. Chen and Metghalchi (2012) attempted to test the predictive power of 32 models of single, double or triple-indicator combinations based on the most popular six technical indicators for the Brazilian stock index (BOVESPA) over a 5-year period (1996–2011). They found that none of the trading models, including the MACD (12,26,9) model, can beat a buy-and-hold strategy. So, they concluded that the Brazilian stock index was weak-form efficient. Abbey and Doukas (2012) tested whether technical analysis is profitable for individual currency traders by using a MACD (12,24,0) model and three other well-known technical analysis indicators for a proprietary database of 428 individual currency traders over the period from March 2004 to September 2009. Their results were that technical analysis is negatively associated with performance. They concluded that this result arose because individual currency traders used well-known technical indicators to trade currencies, which implies that such currency traders suffer from reduced performance. Rosillo, Fuente and Brugos (2013) also examined the profitability of four popular technical analysis tools, including the traditional MACD (12,26,9) model, for the companies of the Spanish Continuous Market from 1986 to 2009 and they reported that the total net benefits generated by applying the MACD model were an unsatisfactory 2.48 percent. Du Plessis (2013) examined the MACD (12,26,9) model for the South African stock market index (FTSE/JSE Top 40) using 10 years (2001–2010) of data and found that the traditional MACD model was less effective than a buy-and-hold strategy. He thus stated that the MACD was not an effective investment strategy using the default parameter settings. Biondo, Pluchino, Rapisarda and Helbing (2013) investigated whether a random trading strategy (which is based on the uniform distribution) outperforms four other standard technical analysis tools, including the MACD (12,26,9) model. From results using 15–20 years of data for the stock indexes, FTSE-UK, FTSE-MIB, DAX and S&P500, they found that the four standard trading strategies perform on average not better than a random strategy. So, they concluded that the random strategy is less risky than the considered standard trading strategies. Nor and Wickremasinghe (2014) also examined the profitability of the MACD (12,26,9) model for the Australian All Ordinaries Index (XOA) by using data over 1996 to mid-2014. They found that the traditional MACD model generally performs poorly in the market but the RSI (Relative Strength Index) model showed some profit potential. So, they concluded that the Australian stock market is not weak-form efficient overall.

As we have seen above, there are many research papers which proved that the MACD trading rule does not produce good results by using the traditional three value settings. Of course, it is not that there is no research that finds a positive result for the MACD approach, although such examples are extremely rare. Chong and Ng (2008) examined the MACD (12,26,0) model and the RSI to see if these technical trading rules are profitable. Using 60 years of monthly data (1935–1994) for the London Stock Exchange FT30 index, they found that the two trading tools can generate higher returns than a buy-and-hold strategy in the market. Chong, Ng and Liew (2014) examined again whether the MACD and the RSI can generally generate excess returns for the stock markets of five other OECD countries (Italy, Canada, Germany, United States and Japan). They applied three traditional MACD (12,26,0), MACD (12,26,9) and MACD (8,17,9) models to market data over 27 years (1976–2002) and reported several interesting findings: while the MACD (12,26,0) model outperformed the buy-and-hold strategy in the Italian market (Milan Comit General), the MACD (12,26,9) model did not show profit potential in the same market; in addition, the returns of the MACD (8,17,9) model were significantly negative in the Italian and the German market (DAX30) while it had no predictive power for the other markets. So, they concluded that the three traditional MACD trading rules are not robust to the choice of market.

In turn, Hejase, Srour, Hejase and Younis (2017) also applied the traditional MACD (12,26,9) model to the stock prices of six Lebanese banks and a real estate company to see if the MACD tool is able to deliver high profits to Lebanese stock traders. Interestingly, they tested three trading strategies different from the typical 'signal line crossover' trading rule of the MACD: such as to execute a 'buy (sell)' transaction only when three 'buy (sell)' signals are generated on three consecutive days and so on. The reason why they introduced different trading strategies was to avoid false trade signals after filtering the empirical results over 11 years of data (2004–2014). Nevertheless, they found that all three non-conventional strategies did not outperform a buy-and-hold strategy. So, they concluded that in the long run, MACD dynamic trades do not make sense.

Again, many researchers have failed to obtain satisfactory results for the MACD indicator by using its traditional parameter settings, and have drawn negative conclusions for it. Yet, these researchers did not explore the performance of other different parameter value settings.

On the other hand, in the second category, there were two groups of researchers who pointed out the necessity of optimizing the three parameter values. Erić et al. (2009) considered a large number of MACD models with different parameter values staring from the shortest (2,3,2) through the longest (29,30,20) and identified the most profitable parameter values for the 48 companies listed on the market of the Republic of Serbia, Belgrade Stock Exchange. However, the most profitable parameter values for each company that are identified from the in-sample test (using data from June 2004 to May 2008) all turned out to have negative profitability in the out-of-sample tests (using data from May 2008 to May 2009). So, they pointed out that it is important to optimize the three parameters over time. Borowski and Pruchnicka-Grabias (2019) also investigated optimal parameter values of the MACD indicator for 140 companies listed on the Warsaw Stock Exchange using data over the period of 2000–2018. They found that some companies had identical optimal parameters but many were different. So, they stated that there is nothing like standard time lengths for moving averages (i.e., parameter values) for all kinds investments for which the transaction system based on the MACD tool is applied. However, as mentioned in the preceding section, the two papers mentioned above were to trying to find the most profitable parameter values for individual companies.

As can be seen from the above, much of the literature in the first category did not consider the need to use optimal parameter values in testing the effectiveness of the MACD approach or in evaluating market efficiency. On the other hand, the studies in the second category pointed out the importance of the optimization of the three parameter values but focused on the most profitable parameter values for each company while failing to consider other good-performing parameter value combinations which may lead to a better understanding of the parameters that are optimal for entire markets. In the author's opinion, all of these things are a consequence of ignoring the need for a new methodology to find good-performing and well-fitted parameter value sets for an entire market. This is the goal of this research which makes importantly different from previous studies and its main contribution to this area of research. (See Section 3.2 for an outline of the new methodology and Section 6 for the results and more on its original contribution.)

In the context of broader research in related areas such as artificial intelligence, research by Wiles and Enke (2015) can be cited. They adopted a genetic algorithm to optimize the three parameter values of the MACD for soybean futures on the Chicago Mercantile Exchange (CME). But the computational best-performing per round trip parameter values (6.866 days, 33.812 days, 4.575 days) obtained their 'genetic algorithm MACD crossover heuristics' were not practical for daily investors since they generate at most only a single trade signal per month for the test period of the optimized parameters (December 2013 to August 2014). Outside this research, to the best knowledge of this author, none of the existing literature has explored a new methodology to find the three optimal parameter values, nor to answer the question stated in the preceding section: "What parameter values are most often used to optimize the trading system for the market?" or "What characteristics do those high-performing parameter values have?"

Finally, let's here add a brief note about the traditional parameter value settings, which were explained in Kang (2021, p.1). The (12,26,9) format is neither a formal standard nor a combination recommended by Appel (1979) who developed the MACD. It is said that Appel originally suggested two different settings on a daily chart: (8,17,9) for buy signals and (12,25,9) for sell signals. But Murphy (1999, p.253) later discussed these two different setting values of Appel and added: "Most traders, however, utilize the default values of 12, 26, and 9 in all instances."

3. Research Methodology

As mentioned before, this study starts with the models ranked in the top 100 and the bottom 100 among the 19,456 models with different parameter values which were examined in Kang (2021). This section provides a data description, information about the 19,456 sample models, the trading simulation method to be used in this study, and a brief explanation of the new methodology mentioned in the first section.

3.1 Sample Data and Trading Rules

This study uses the daily closing index values of Nikkei 225 futures contracts near maturity over 11 years (4 January 2011 to 30 December 2021) which were obtained from an official data provider, JPX Data Cloud (http://db-ec.jpx.co.jp). The Nikkei 225 includes the top 225 blue-chip companies listed on the Osaka Exchange of the Japan Exchange Group (JPX); it is the primary yen-denominated stock index future. We allot the first 9 years of data (2011–2019)—which was used in the previous research by Kang (2021)—for in-sample tests and the last 2 years of data (2020–2021) —which is newly added—for out-of-sample tests.

Regarding the data used for the in-sample test, note that Kang (2021) divided the sample data over the 9 years

(2011–2019) into three fairly long sub-periods: 2011–2013, 2014–2016 and 2017–2019 following the approach of Chong and Ng (2008) in order to avoid data snooping (selection) bias. This means that the calculation of profitability for every model is carried out in each sub-period separately for robustness in the choice of sample period.

Now, the 19,456 sample models examined by Kang (2021) start from the MACD (3,5,3) model and end with the MACD (20,40,40) model where the three parameters (n_1, n_2, n_3) are taken into consideration over a given range as follows: n_1={3,⋯,20}, n_2={5,⋯,40}, n_3={3,⋯,40} at an interval of one day. This study revisits these sample models in order to examine what characteristic parameter value combinations have good performance and fit well in the Japanese market.

In terms of trading simulations, this study applies the same trading rule adopted in Kang (2021) for consistency in the trading results of the two research papers. To summarize: (1) When a 'buy (or sell)' signal is generated according to 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) After having opened a 'buy (or sell)' position, all subsequent identical buy (or sell) signals are ignored. However, when the first opposite trading signal from the opening position, i.e., '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) At the same time, in order 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 of the same day. That is, 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 have to be executed one-by-one sequentially. This means that holding multiple positions is not permitted after a position is taken.

For reference, one contract unit on the Nikkei 225 futures for a large contract is 1000 times the value of the Nikkei 225 index value and its tick size is 10 Japanese yen (JPY). Therefore, one tick up for a one-unit contract leads to a positive return of JPY 10,000 (=10×1000). On the other hand, the current round-trip commission for one large contract unit is so small—approximately 0.33 percent of the positive return of JPY 100,000 corresponding to when the index moves just 10 ticks up—that it can be ignored, compared to the high leverage at 1000 times the index value. Thus, this tiny commission will not be taken into consideration in this study.

3.2 The Outline of the New Methodology

In the 1st place, we investigate the frequency distributions of the parameter values of the top/bottom 100 models and compare how different they are. The reason for considering these models that are ranked in the two opposite groups is that their ranks in profitability are solely determined by the difference in the three parameter values used so that we can consider these models as a representative group of models for which parameter value settings are optimal and non-optimal. (The result will be shown in the following subsections 4.1–4.2.)

In the 2nd place, we differentiate the most frequently used values in the top 100 models from those values in the bottom 100 models and then define a few ranges of parameter values which are collectively distributed around the most frequently used value in the top/bottom 100 models. This work will be conducted in a consistent and systematic way by applying a pre-established selection rule. Note that we refer to the parameter values belonging to the pre-defined ranges in the top (bottom) 100 models as "analyzed optimal (non-optimal) parameter values" and the ranges as "preset optimal (non-optimal) ranges". (The results will be shown in the following subsection 4.3.)

In the 3rd place, we create some groups of sample models for which the parameters consist of the analyzed optimal (non-optimal) values belonging to the pre-defined ranges and refer to the models as new "hypothetical optimal (non-optimal) sample models" with the analyzed optimal (non-optimal) parameter values. Several hundred sample models are created in this process. (The results will be shown in the following subsections 4.4–4.6.)

In the last place, we examine the profitability of the hypothetical optimal (non-optimal) sample models by conducting in-sample tests to confirm the optimality (non-optimality) of their parameter values and then confirm the results again in out-of-sample tests. Further discussion will concern which group of models have the best optimal parameter values. (All of this will be shown in Section 5.)

4. Preliminary Analysis by Applying the New Methodology

This section deals with the three issues described in the preceding section: from the first issue—investigating the frequency distributions of the three parameter values used in the top/bottom 100 models; to the second issue—determining ranges of optimal (non-optimal) parameter values; and to the third issue—creating new hypothetical optimal (non-optimal) sample models.

4.1 Frequency Distributions of the Three Parameter Values in the Top/Bottom 100 Sample Models

Figure 1 illustrates the frequency ratios of the three parameter values (n_1, n_2, n_3) observed for the top 100 and the bottom 100 models among the 19,456 sample models. We can see at once from the figure that the most frequently used parameter values of the two opposite groups have distinctively different distributions from each other. Below we will look more closely at the frequency distributions of the three parameter values one by one. Note that we will use the word 'length' instead of a 'parameter value' in following discussion for convenience.

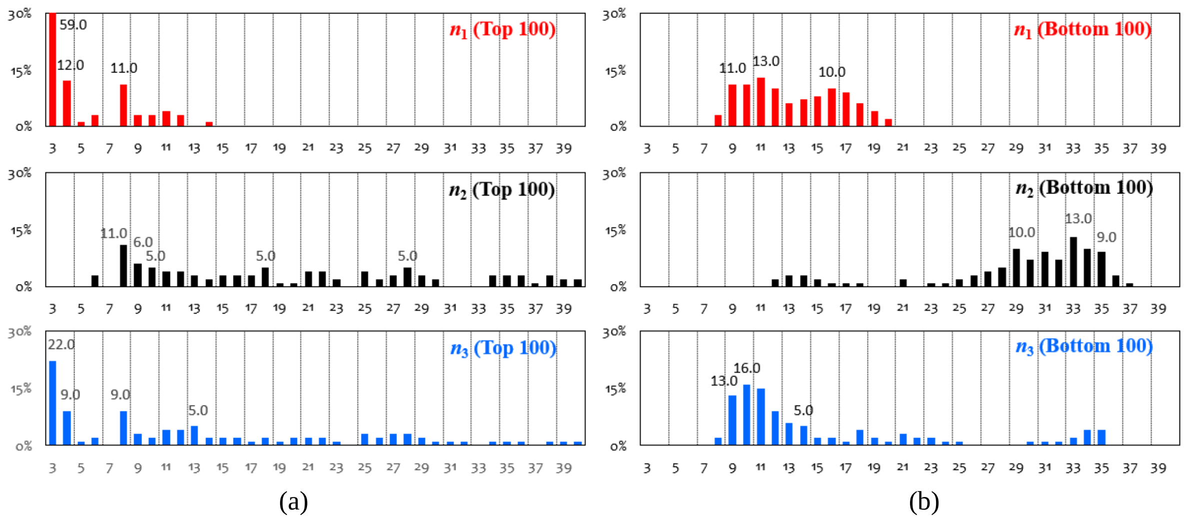


Figure 1. (a) Frequencies of the three parameter values (n_1, n_2, n_3) in the top 100 models. (b) Frequencies of the three parameter values (n_1, n_2, n_3) in the bottom 100 models

4.1.1 The First Parameter (n_1)

See the histogram in the top left of Figure 1. The most frequently used lengths in the "top" 100 models are 3-day (59%), 4-day (12%) and 8-day (11%). Three notable points are:

- The 3-day length dominates the set with the largest share at 59 percent.
- The three most frequently used lengths (3-, 4- and 8-day) are all distributed ahead of the 10-day length.
- No length longer than 14-day is observed.

On the other hand (see the histogram in the top right of Figure 1), the most frequently used lengths in the "bottom" 100 models are 9- (11%), 10- (11%), 11- (13%), 12- (10%) and 16-day (10%) while any length less than the 8-day length is not observed. This is in marked contrast to the features of the moving average lengths adopted in the "top" 100 models. That is:

- The most frequently used five lengths in the "bottom" 100 models are distributed in the range of 9–16, where lengths with none or small shares in the "top" 100 models are recorded.
- On the contrary, lengths with no or a small share in the 100 "bottom" models are placed in the range of 3–8, but the most frequently used lengths with very high shares in the "top" 100 models are distributed around there.

That is to say, the two ranges of the most frequently used moving average lengths in the "bottom" 100 models (i.e., 9–16) and the "top" 100 models (i.e., 3–8) do not overlap each other.

4.1.2 The Second Parameter (n_2)

See the histogram in the middle left of Figure 1. The most frequently used lengths in the "top" 100 models are 8-day (11%), 9-day (6%) and 10-day (5%). Local extrema with a share of 5 percent are found on an occasional basis at 18- (5%) and 28-day (5%). Two notable points about this distribution are:

- It has a wide range of moving average lengths with relatively small shares.
- Next to the 8-day length, no length with a share exceeding 10 percent is observed.

On the other hand (see the histogram in the middle right of Figure 1), the most frequently used lengths in the "bottom" 100 models are collectively distributed in the range of 29 (10%)–35 (9%) with 33-day (13%) as a peak. As before, this is in marked contrast to the features of the moving average lengths adopted in the "top" 100 models. That is:

- The most frequently used lengths for the "bottom" 100 models are collectively distributed in the range of 29–35, but the lengths with no or small shares in the "top" 100 models are placed there.
- On the contrary, lengths with no shares for the "bottom" 100 models are observed in the range of 5–11, but the most frequently used lengths with relatively large shares in the "top" 100 models are distributed there.

These two points indicate that the two ranges of the most frequently used moving average lengths in the "bottom" 100 models (i.e., 29–35) and the "top" 100 models (i.e., 5–11) do not overlap each other.

4.1.3 The Third Parameter (n_3)

See the histogram in the bottom left of Figure 1. We can then see that 3-day (22%), 4-day (9%) and 8-day (9%) lengths dominate the set. What has to be noticed is that the third parameter has both features of the first parameter (n_1) and the second parameter (n_2) as described above. That is:

- The most frequently used lengths (3-, 4- and 8-day) are same as was in the case for the first parameter values and it is also the same case that those lengths are all distributed ahead of the 10-day length—which is a similar feature to that of the first parameter (n_1).
- Almost every length is distributed across the whole range—which resembles a feature of the second parameter (n_2).

On the other hand (see the histogram in the bottom right of Figure 1), the most frequently used lengths of moving averages are distributed collectively in the range of 9 (13%)–14 (5%) with 10-day (16%) as a peak. Almost the same observation as before can be made:

- The most frequently used lengths for the "bottom" 100 models are collectively distributed in the range of 9–14, but the lengths with small shares in the "top" 100 models are placed there.
- On the contrary, lengths with no or a small share in the "bottom" 100 models are placed in the range of 3–8, but the most frequently used lengths with very high shares in the "top" 100 models are distributed around there.

In conclusion, the two ranges of the most frequently used moving average lengths in the "bottom" 100 models (i.e., 9–14) and the "top" 100 models (i.e., 3–8) do not overlap each other.

4.2 Frequency Distributions of the Three Parameter Values in Extended Cases of Sample Models

Let us now turn to examine whether the features of the three parameter values found above hold true if we expand the number of sample models to the top/bottom 500 and 1000. Look at Figure 2 to Figure 4. We can then see that the features discovered in the preceding subsections for each of the three parameter values become clearer since the variability of each sampling distribution decreases as the sample sizes increase. But to save space, we will focus on the most notable findings about the extended cases of the top ranked groups of sample models.

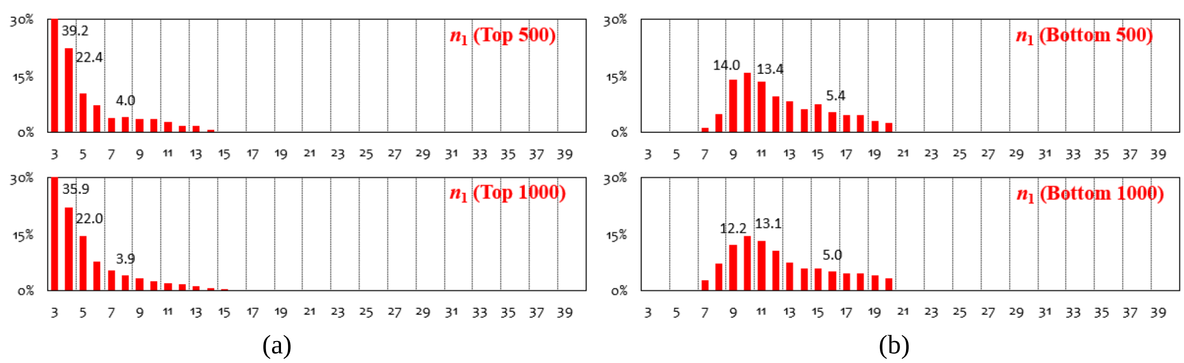
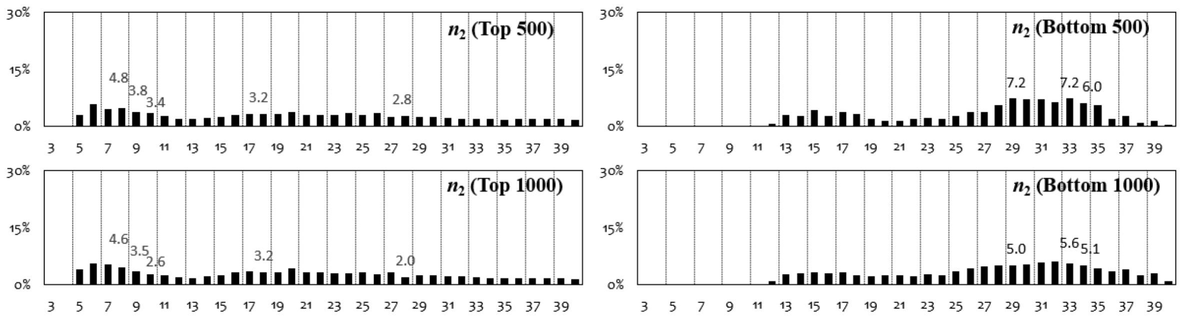


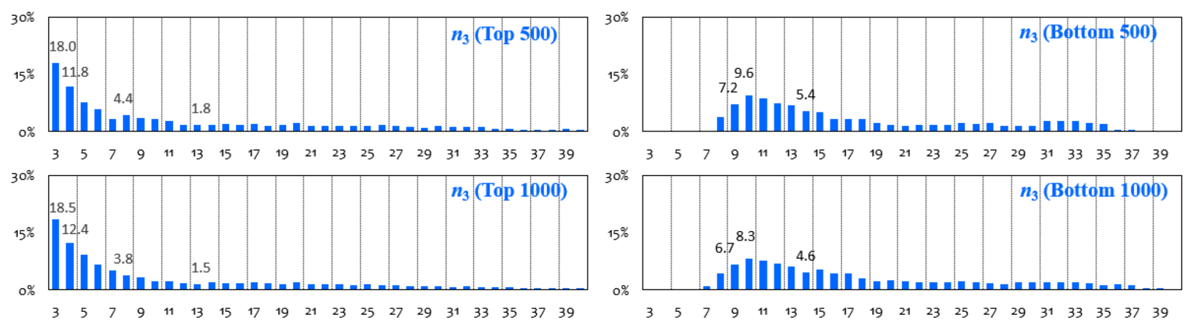
Figure 2. (a) Frequencies of the first parameter values n_1 in the top 500 and 1000 models. (b) Frequencies of the first parameter values n_1 in the bottom 500 and 1000 models

See again Figure 2(a) for the first parameter (n_1). It confirms that the three items described before (subsection 4.1.1) almost follow the prototype even though we increase the sample size up to 500 and 1000. Leaving the details of minor changes aside, note that the histogram makes a curve sloping downward from left to right as the sample sizes increase which looks like a ‘reverse J-shaped distribution’. This suggests that: *the optimality of the first parameter* n_1 *in the best-performing models gets worse as the length of the moving average becomes longer and vice versa*: *the shorter the length of the moving average, the better*.



(a) (b)

Figure 3. (a) Frequencies of the second parameter values n_2 in the top 500 and 1000 models. (b) Frequencies of the second parameter values n_2 in the bottom 500 and 1000 models



(a) (b)

Figure 4. (a) Frequencies of the third parameter values n_3 in the top 500 and 1000 models. (b) Frequencies of the third parameter values n_3 in the bottom 500 and 1000 models

See Figure 3(a) for the second parameter (n_2). Almost the same observations as the two items described before (subsection 4.1.2) can be made for the top 500 and 1000 models. Keeping this point in mind and paying attention to the values with relatively higher shares, we can say that: *the second parameter* n_2 *has a wide range of moving average lengths but, it is most often optimized by the two clusters of lengths at approximately 6–10 or around 18–28 as the next best*.

See Figure 4(a) for the third parameter (n_3). The frequency distribution of this parameter’s values also makes a clear ‘reverse J-shaped distribution’ with ‘a long tail on the right’ as the sample sizes increase. This reflects the two features of this parameter that have been mentioned before (subsection 4.1.3). Focusing attention to the lengths with higher shares, this suggests that: *the third parameter* n_3 *is most often optimized by short moving average lengths of less than 10 days*.

To summarize what we have found above: (1) For each of the three parameters, the two ranges of the most frequently used moving average lengths in the best-performing and the worst-performing models are far from each other and almost do not overlap. (2) This distinctive difference found for the top/bottom 100 models becomes clearer, keeping its prototype (without significant change) as the sample size increases, at least for the top/bottom 1000 models. Taking these findings into consideration, we can say that: *there exists a boundary for optimal and non-optimal ranges of moving average lengths for each of the three parameters*.

4.3 A Rule for Finding Optimal Parameter Values

Based on the finding stated at the end of the preceding section, we next attempt to find a boundary to distinguish optimal and non-optimal moving average lengths for each of the three parameters. To accomplish this, we set the frequency distributions of the moving average lengths in the top/bottom 100 models as testing targets and apply the following five-step rule.

- **Step 1**: Identifying good-performing moving average lengths

For each of the three parameters (n_1, n_2, n_3) of the "top" 100 models, identify moving average lengths for which the share is greater than or equal to a certain percent (e.g., 4 percent)—in the sense that those time spans of moving averages deserve consideration as optimal parameter values. The "4 percent" is an example of the "minimum cutoff point" adopted for this study. (See the column titled "Best" of the three parameters in Table 1 in appendix. Bold figures with underline indicate those lengths and note (a) in the following paragraph after Step 5.)

- **Step 2**: Filtering out poor-performing moving average lengths

For each of the three parameters of the "bottom" 100 models, do the same thing as indicated in the first step. If a length in the bottom 100 models is also identified as one in the top 100 models, put a cross "x" next to it—in the sense that we exclude it. (See the cross in the column titled "P/F" in Table 1, which implies 'pass or fail'.)

- **Step 3**: Grouping of good-performing moving average lengths

For the remaining identified moving average lengths, when they are placed in consecutive order, classify them together into a group (if this rule yields multiple groups, differentiate them as A or B and so on)—in the sense that those consecutive lengths should be considered as elements belonging to a range of optimal parameter values. (See the column "P/F" of the first parameter n_1 in Table 1 for example. The two moving average lengths 3-day and 4-day are classified into the same group A_{n1}.)

- **Step 4**: Adding separated good-performing lengths to the nearest predetermined group

If an identified moving average length is not in a predetermined group, include it in the nearest group to avoid creating too many subsets with small elements. (See the column "P/F" of the parameter n_1 in Table 1 for example. The 8-day length is added for the A_{n1} group.)

- **Step 5**: Identifying supplementary groups

For the remaining identified moving average lengths in the "top" 100 models, if some of them have a small share of 2 or 3 percent (less than the minimum cutoff point of 4 percent) but they are not included in the opposite "bottom" 100 models, designate them as a "secondary group" of good-performing moving average lengths. (See that "C_{n2}" and "B_{n3}" for the parameter values n_2 and n_3 are added in and note (b) in the following paragraph after Step 5.)

Note that: (a) In this study, the "4 percent" works as a "boundary" to distinguish moving average lengths that are mainly used in the best-performing models and those used in the worst-performing models; and, (b) The reason for considering these secondary groups is to see how the models with these groups' moving average lengths perform.

4.4 Defining Ranges of Optimal Parameter Values

As described in the preceding subsection, we define the ranges of analyzed optimal moving average lengths for the three parameters as follows:

- A_{n1}: {3,⋯,8} for the first parameter n_1
- A_{n2}: {8,⋯,12}, B_{n2}: {18,⋯,25}, C_{n2}: {38,⋯,40} for the second parameter n_2
- A_{n3}: {3,⋯,8}, B_{n3}: {26,⋯,29} for the third parameter n_3

where the subscript n_1, n_2, n_3 of A_{n1}, A_{n2}, A_{n3} and so on indicates the order of the three parameters (n_1, n_2, n_3) respectively.

4.5 Creating New Sample Models With the Analyzed Optimal Parameter Values

Using the pre-specified ranges presented above, we obtain six possible combinations where each combination consists of three ranges corresponding to the three parameters (n_1, n_2, n_3): $A_{n1}–A_{n2}–A_{n3}$, $A_{n1}–B_{n2}–A_{n3}$, $A_{n1}–C_{n2}–A_{n3}$, $A_{n1}–A_{n2}–B_{n3}$, $A_{n1}–B_{n2}–B_{n3}$, and $A_{n1}–C_{n2}–B_{n3}$. Note that the last four combinations include the secondary group of good-performing moving averages lengths "C_{n2}" or "B_{n3}", or both mentioned in Step 5 in the preceding subsection. We distinguish these from the others as follows:

- Primary combination of three optimal parameter values: $A_{n1}–A_{n2}–A_{n3}$, $A_{n1}–B_{n2}–A_{n3}$

- Secondary combination including a suboptimal parameter value: $A_{n1}-C_{n2}-A_{n3}$, $A_{n1}-A_{n2}-B_{n3}$, $A_{n1}-B_{n2}-B_{n3}$ and $A_{n1}-C_{n2}-B_{n3}$

For this reason, we will refer to the sample models to be constructed below from the former as "primary groups of optimal sample models" and those constructed from the latter as "secondary groups of suboptimal sample models" in the next section. Also, we refer to them as "optimal sample models" when referring to these two groups together.

Now, the number of elements belonging to each range presented above is: $n(A_{n1}) = 6$, $n(A_{n2}) = 5$, $n(B_{n2}) = 8$, $n(C_{n2}) = 3$, $n(A_{n3}) = 6$ and $n(B_{n3}) = 4$. Accordingly, from the first combination "$A_{n1}-A_{n2}-A_{n3}$" we can create 180 (=6×5×6) unique sample models in total, However, this combination produces 6 (=1×1×6) irrational models where the value of n_1 is equal to that of n_2. We thus consider the 174 (=180−6) models as the sample models of this combination. For the same reason, from the fourth combination "$A_{n1}-A_{n2}-B_{n3}$" we have 116 (=6×5×4−4) sample models. The other combinations include no irrational models as mentioned above. Therefore, from the second combination "$A_{n1}-B_{n2}-A_{n3}$" we obtain 288 (=6×8×6) unique sample models in total; from the third "$A_{n1}-C_{n2}-A_{n3}$", 108 (=6×3×6); from the fifth "$A_{n1}-B_{n2}-B_{n3}$", 192 (=6×8×4); and, from the sixth "$A_{n1}-C_{n2}-B_{n3}$", 72 (=6×3×4), respectively. To sum up, we have 950 (=174+116+288+108+192+72) unique sample models to consider in this study.

Note that among the 950 models, 19 models are the original ones that were included in the top 100 models. This implies that all the top 100 models are not included in the new 950 models. In addition, these models are "hypothetical" sample models with the "analyzed" optimal/suboptimal parameter values at the present stage. Nevertheless, if these sample models have good performances, we can consider their parameter values to be "good-performing optimal ones".

4.6 Creating New Sample Models With the Analyzed Non-optimal Parameter Values

We have also defined ranges of non-optimal moving averages for the three parameters by applying the same rule stated in Section 4.3, i.e., grouping such moving average lengths that are used with a share greater than or equal to 4 percent in the bottom 100 models but rarely used in the top 100 models (See Table 1 in appendix). The reason for this is to examine: "If we make models with the non-optimal parameter values, will all of them be poor-performing?" If these models match our expectations, we will gain helpful insights about both optimal and non-optimal parameter values. Thus, the following ranges of analyzed non-optimal moving average lengths are defined.

- W_{n1}: {13,⋯,19} for the first parameter n_1
- W_{n2}: {31,⋯,33} for the second parameter n_2
- W_{n3}: {34, 35} for the third parameter n_3

Using the preset ranges presented above, we obtain

- The worst combination of non-optimal three parameter values: $W_{n1}-W_{n2}-W_{n3}$.

We can create from this combination 42 (=7×3×2) unique sample models. We will refer to these hypothetical sample models with the analyzed non-optimal parameter values as "non-optimal sample models" or "the worst group of sample models with non-optimal parameter values" hereafter. Note that none of the new 42 models are not included in the bottom 100 models. Nevertheless, if we find that these 42 models do not perform well, we can say that it is due to their "non-optimal parameter ones".

5. Empirical Results

We examine whether the hypothetical sample models with the newly analyzed optimal (non-optimal) parameter values are genuinely good (poor) performing models by conducting in-sample tests using data (2011–2019) for the Japanese futures market. After confirming these results, we reexamine their performance in out-of-sample tests using the most recent data (2020–2021).

5.1 In-sample Test Results

We focus on the following two points: first, "Do the primary groups of sample models with the optimal parameter values have higher returns than the secondary groups of sample models with the suboptimal parameter values?" to see how the difference of optimal and suboptimal parameter values affects their performance; and second, "Which group of sample models are best-performing?"

5.1.1 An Overview of the Results of the In-sample Tests

Table 2 summarizes the characteristics of the returns for the sample models belonging to the six groups discussed in the previous section. The profitability of the six groups is shown here by their raw returns to facilitate comparison.

Table 2. Summary statistics of the in-sample tests (2011–2019)

		n	Max	Min	Mean	Median	S.E	Skewness	Kurtosis
Primary Groups	$A_{n1}-A_{n2}-A_{n3}$	174	8130	-750	4252.1	4160	152.8	-0.1929	-0.4869
	$A_{n1}-B_{n2}-A_{n3}$	288	13930	2560	9554.4	9875	133.9	-0.6870	0.1035
Secondary Groups	$A_{n1}-C_{n2}-A_{n3}$	108	15140	-2340	7691.6	9110	420.5	-0.2634	-1.1508
	$A_{n1}-A_{n2}-B_{n3}$	116	15240	-4720	3933.1	3280	444.2	0.6132	0.0139
	$A_{n1}-B_{n2}-B_{n3}$	192	1640	-13530	-6873.9	-7620	314.0	0.3088	-1.2119
	$A_{n1}-C_{n2}-B_{n3}$	72	4140	-9630	-5256.0	-6255	424.9	1.3021	0.8530

Note: The first column 'n' indicates the number of sample models belonging to each group and the column 'S.E' indicates the standard error.

The points to be observed are:

(1) The 1st group ($A_{n1}-A_{n2}-A_{n3}$) and the 2nd group ($A_{n1}-B_{n2}-A_{n3}$) of sample models have good performance with high mean returns of 4252.1 and 9554.4 respectively (the total average of the two mean returns is 6903.3).

(2) The 3rd group ($A_{n1}-C_{n2}-A_{n3}$) and the 4th group ($A_{n1}-A_{n2}-B_{n3}$) have also good performance with mean returns of 7691.6 and 3933.1 respectively. But the total average (5812.4) of the two mean returns does not exceed that (6903.3) of the 1st and the 2nd groups.

(3) The 5th ($A_{n1}-B_{n2}-B_{n3}$) and the 6th ($A_{n1}-C_{n2}-B_{n3}$) groups deliver negative mean returns of '-6873.9' and '-5256.0' respectively—opposite to what was expected.

(4) The 2nd group ($A_{n1}-B_{n2}-A_{n3}$) has the best high performance and the most stable profitability in the sense that its mean return (9554.4) is the highest and its standard error (133.9) is the smallest; more noteworthy is that its returns are all positive from 2560 to 13930.

To summarize, the sample models with the analyzed optimal/suboptimal parameter values have good performance overall, excepted for the 5th and the 6th groups. Before going into further discussion, it is interesting to note here that the traditional MACD (12,26,9) model earns a surprisingly large 'minus 4,180' for the same test period while the MACD (4,22,3) model—presented by Kang (2021) as a model with optimal parameter values—earns '9,420'. Another interesting point is that this study makes it possible to find that the MACD (4,22,3) model is a member of the good-performing 2nd group ($A_{n1}-B_{n2}-A_{n3}$) defined in this study.

5.1.2 The Frequency Distribution of Returns for the In-sample Tests

Let us look at the results seen above from a different angle. Table 3 displays the frequency distribution of returns produced by the sample models belonging to the six groups.

Table 3. Frequencies of returns produced by the six groups for the in-sample tests (2011–2019)

	$A_{n1}-A_{n2}-A_{n3}$		$A_{n1}-B_{n2}-A_{n3}$		$A_{n1}-C_{n2}-A_{n3}$		$A_{n1}-A_{n2}-B_{n3}$		$A_{n1}-B_{n2}-B_{n3}$		$A_{n1}-C_{n2}-B_{n3}$	
	n	Ratio	n	Ratio	n	Ratio	n	Ratio	n	Ratio	n	Ratio
Less than 0	5	(2.9%)	0	(0.0%)	3	(2.8%)	16	(13.8%)	179	(93.2%)	61	(84.7%)
0 – 2,500	23	(13.2%)	0	(0.0%)	9	(8.3%)	32	(27.6%)	13	(6.8%)	6	(8.3%)
2,500 – 5,000	81	(46.6%)	14	(4.9%)	25	(23.1%)	30	(25.9%)	0	(0.0%)	5	(6.9%)
5,000 – 7,500	58	(33.3%)	37	(12.8%)	11	(10.2%)	17	(14.7%)	0	(0.0%)	0	(0.0%)
7,500 – 10,000	7	(4.0%)	97	(33.7%)	13	(12.0%)	5	(4.3%)	0	(0.0%)	0	(0.0%)
10,000 – 12,500	0	(0.0%)	118	(41.0%)	32	(29.6%)	5	(4.3%)	0	(0.0%)	0	(0.0%)
12,500 – 15,000	0	(0.0%)	22	(7.6%)	14	(13.0%)	10	(8.6%)	0	(0.0%)	0	(0.0%)
15,000 – 17,500	0	(0.0%)	0	(0.0%)	1	(0.9%)	1	(0.9%)	0	(0.0%)	0	(0.0%)
Total	174	(100.0%)	288	(100.0%)	108	(100.0%)	116	(100.0%)	192	(100.0%)	72	(100.0%)

The points to be confirmed are:

(1) In the case of the 1st group ($A_{n1}-A_{n2}-A_{n3}$), 97.1 (=100–2.9) percent of the sample models have positive returns that are concentrated in the range of below 10,000. In the case of the 2nd group ($A_{n1}-B_{n2}-A_{n3}$), 100 percent of the sample models achieve positive returns and the returns of 48.6 (=41.0+7.6) percent of the models are concentrated in the high range of 10,000 to 15,000.

(2) In the case of the 3rd group ($A_{n1}-C_{n2}-A_{n3}$), 97.2 (=100–2.8) percent and for the 4th group ($A_{n1}-A_{n2}-B_{n3}$), 86.2 (=100–13.8) percent of the sample models have positive returns. These two groups of sample models have also high levels of returns recorded in the range of more than 10,000, i.e., 43.5 (=29.6+13.0+0.9) percent and 13.8 (=4.3+8.6+0.9) percent, respectively. But, the total average of the positive return ratios (98.6= (97.1+100) ÷2) of the 1st and the 2nd groups is higher than that (91.7= (97.2+86.2) ÷2) of the 3rd and the 4th groups.

(3) In the case of the 5th ($A_{n1}-B_{n2}-B_{n3}$) and the 6th ($A_{n1}-C_{n2}-B_{n3}$) groups, only 6.8 and 8.3 percent of the models have positive returns, respectively.

(4) In terms of the ratio of high levels of returns that are greater than 10,000, the 2nd group is the best. That is, the 2nd group has the most numerous high-performing models.

Judging from the results shown in Table 3 and the mean returns shown in Table 2, we can expect good performance from sample models belonging to the 1st and the 2nd groups, including even the 3rd and the 4th groups, but not from models belonging to the 5th and the 6th groups.

This means that there is no need for further discussion of the 5th and the 6th groups of models. This is because their overwhelmingly poor performance has been verified over the nine years and thus their relative disadvantages in profitability are not to be denied. Moreover, as long as we cannot expect good performance in the future from these two secondary groups of models, even though they might bring some good outcomes in the future, we shall consider the results as obtained by chance. For this reason, we exclude the 5th and the 6th groups of sample models from further discussion.

5.1.3 The Performance of the Worst-performing Group for the In-sample Tests

Let us confirm the performance of the hypothetical worst group of sample models with the analyzed non-optimal parameter values. As can be seen from Table 4, all of the sample models have negative returns without exception. This result confirms that parameter values adopted in the sample models of this group—which were derived by using the new methodology—are actually non-optimal values. Nevertheless, we will examine again this group's performance in the out-of-sample tests.

Table 4. Summary statistics of the worst-performing group (2011–2019)

	n	Max	Min	Mean	Median	S.E	Skewness	Kurtosis
$W_{n1}-W_{n2}-W_{n3}$	42	-2850	-6640	-4800.7	-4865	140.3	0.0896	-0.1081

5.1.4 Hypotheses

Taking all findings described in the previous subsections into consideration, several reasonable hypotheses are suggested for the out-of-sample tests.

- **Hypothesis 1**: Models belonging to the 1st and the 2nd groups (primary groups) perform well due to their optimal parameter value settings.
- **Hypothesis 2**: Models belonging to the 3rd and the 4th groups (secondary groups) also perform well, but their performances do not exceed those of the 1st and the 2nd groups of models on average because they have one or two suboptimal parameter values.
- **Hypothesis 3**: The 2nd group has the largest number and the highest ratio of high-performing models.
- **Hypothesis 4**: Models belonging to the worst-performing group do not perform well due to their non-optimal parameter value settings.

For Hypothesis 1, see points (1) in section 5.1.1 and (1) in section 5.1.2; for Hypothesis 2, see (2) in section 5.1.1 and (2) in section 5.1.2. As for Hypothesis 3, see points (4) in section 5.1.1 and (4) in section 5.1.2; and, for Hypothesis 4, see Table 4 in section 5.1.3.

5.2 Out-of-sample Test Results

Before going into the results of the out-of-sample tests, let us begin with a simple observation on the changes in the trends in the Nikkei 225 futures values. Figure 5 shows how the index changes from 4 January 2011 to 30 December 2021 (which covers 2,692 trading days).

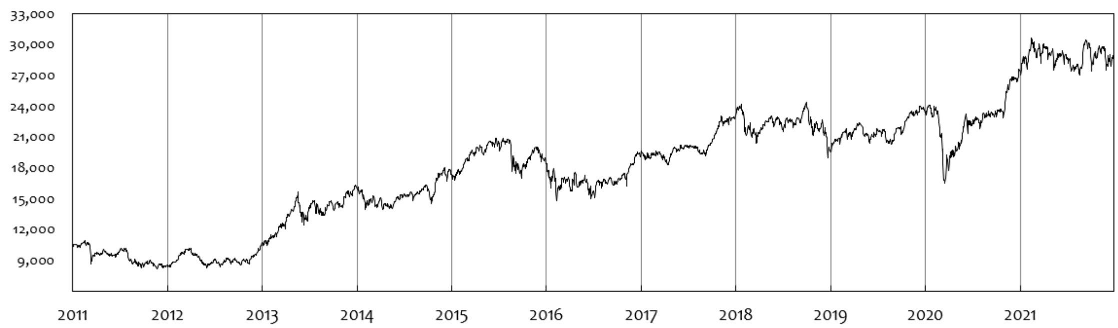


Figure 5. Index values of Nikkei 225 futures over the last 11 years

As we can see from the figure, the index value goes up suddenly from 2013 and then manifests a long-term increasing trend to 2019. The latter period is quite different from the relatively stable levels of the period before 2013 (in fact, the stable flat trend to the end of 2012 goes back to 2009). As for these changes in the trend, see Kang (2021) where explanatory notes are provided, which are based on expansionary monetary policies implemented by the Federal Reserve Board of the United States and the Bank of Japan.

As for the wild roller-coaster ride in 2020, it seems to be a reflection of the impact of the ongoing COVID-19 pandemic on financial markets around the world and the various stimulus packages implemented by governments. The World Health Organization declared a “Public Health Emergency of International Concern” on 30 January 2020, and a pandemic on 11 March 2020. In the year since, the extraordinary fall started to bounce back after mid-March as governments began responding with record stimulus packages to support their economies, and by early June regained the market losses. As for the rapid increasing trend propelled from the beginning of November, it seems to have been caused by news of vaccines including Pfizer’s release of their trial results for a candidate vaccine which was reported on 9, November.

In 2021, the index was driven up and down in response to the developments of the COVID-19 pandemic as well as political changes in Japan. It will be interesting to see how the sample models defined in this study perform in the pandemic period (2020–2021). Note that the models were fitted to data during a “normal” period, yet will be tested in the pandemic period.

5.2.1 An Overview of the Results of the Out-of-sample Tests

Table 5 summarizes the descriptive characteristics for the returns of the sample models which belong to the four groups. As we can see from the table, the four groups have all positive mean returns.

Table 5. Summary statistics for the out-of-sample tests (2020–2021)

	n	Max	Min	Mean	Median	S.E	Skewness	Kurtosis
$A_{n1}-A_{n2}-A_{n3}$	174	12230	20	5920.7	5920	240.0	0.0495	-1.0175
$A_{n1}-B_{n2}-A_{n3}$	288	11630	-3800	4930.7	5450	174.8	-0.4627	-0.2936
$A_{n1}-C_{n2}-A_{n3}$	108	4480	-3880	910.6	1340	186.7	-0.6710	-0.3513
$A_{n1}-A_{n2}-B_{n3}$	116	5420	-1980	1877.4	1610	131.8	0.0053	0.0912

The points to be observed are:

(1) Models belonging to the 1st group ($A_{n1}-A_{n2}-A_{n3}$) and the 2nd group ($A_{n1}-B_{n2}-A_{n3}$) have good performance

with high mean returns of 5920.7 and 4930.7 respectively (the total average of these two groups is 5425.7)—which is the same as was suggested from the results for the in-sample tests.

(2) Models belonging to the 3rd group ($A_{n1}-C_{n2}-A_{n3}$) and the 4th group ($A_{n1}-A_{n2}-B_{n3}$) also perform well with mean returns of 910.6 and 1877.6 respectively. But the total average (1394.1) of these two groups does not exceed that (5425.7) of the 1st and the 2nd groups of models—as was confirmed by the in-sample tests.

(3) Models belonging to the 2nd group have good performance with a relatively high mean return of 4930.7 but the mean return is not the highest among the four groups—which is inconsistent with the result of the in-sample tests.

The first and the second items stated above suggest that "Hypothesis 1" and "Hypothesis 2" established in the preceding section are all verified. But the third item stated above suggests that "Hypothesis 3" is not valid in the sense that the mean return is not the highest. But this is not a direct denial of "Hypothesis 3" since we have not yet found enough evidence to confirm how many high-performing models each group has. We will see the results in the next subsection.

For reference, let us report the results of the two groups—the 5th and the 6th groups that have been excluded from further discussion in the previous subsection—produced for the out-of-sample tests. Their mean returns were '224.8' and 'minus 4267.4' respectively, which were both unambiguously low.

5.2.2 The Frequency Distribution of Returns for the Out-of-sample Tests

Table 6 displays the frequency distribution of returns produced by the sample models belonging to the four groups. From these results, we can see how many models of each group had positive returns and how many high-performing models each group contains. We will focus on the latter issue to save space.

Table 6. Frequencies of returns achieved by the four groups for the out-of-sample tests (2020–2021)

	Primary Groups				Secondary Groups				Total	
	$A_{n1}-A_{n2}-A_{n3}$		$A_{n1}-B_{n2}-A_{n3}$		$A_{n1}-C_{n2}-A_{n3}$		$A_{n1}-A_{n2}-B_{n3}$			
Return range	n	ratio	n	ratio	n	ratio	n	ratio	n	ratio
less than 0	0	(0.0%)	18	(6.3%)	27	(25.0%)	5	(4.3%)	50	(7.3%)
0 − 2,500	28	(16.1%)	41	(14.2%)	58	(53.7%)	71	(61.2%)	198	(28.9%)
2,500 − 5,000	48	(27.6%)	66	(22.9%)	23	(21.3%)	39	(33.6%)	176	(25.7%)
5,000 − 7,500	36	(20.7%)	106	(36.8%)	0	(0.0%)	1	(0.9%)	143	(20.8%)
7,500 − 10,000	38	(21.8%)	51	(17.7%)	0	(0.0%)	0	(0.0%)	89	(13.0%)
10,000 − 12,500	24	(13.8%)	6	(2.1%)	0	(0.0%)	0	(0.0%)	30	(4.4%)
Total	174	(100.0%)	288	(100.0%)	108	(100.0%)	116	(100.0%)	686	(100.0%)

The first point to notice is that 56.3 (=20.7+21.8+13.8) percent of the models belonging to the 1st group ($A_{n1}-A_{n2}-A_{n3}$) have high levels of returns recorded in the range of more than 5,000 while 56.6 (=36.8+17.7+2.1) percent of the models belonging to the 2nd group ($A_{n1}-B_{n2}-A_{n3}$) have similar high levels of returns. From these results, we can say with fair certainty that "Hypothesis 1" is supported.

On the other hand, none of the models belonging to the 3rd ($A_{n1}-C_{n2}-A_{n3}$) and the 4th ($A_{n1}-A_{n2}-B_{n3}$) groups yield higher returns of more than 5,000, except just one case for the 4th group. One possible explanation for this difference is that these results were caused by the difference of parameter values that were used in "the 1st and the 2nd groups" and "the 3rd and the 4th groups". Considering this and recalling that the mean returns of the two primary groups are both far greater than those of the other two secondary groups, we can say that "Hypothesis 2" is essentially valid.

On the remaining issue about "Hypothesis 3", we have already seen from Table 5 that the 1st group has a higher mean return (5920.7) than that (4930.7) of the 2nd group. As long as we focus only on this result, it can be said that the 1st group has better performance than the 2nd group on average. But, just because of this we cannot simply conclude that the 1st group has an advantage in profitability relative to the 2nd group, at least at this stage. Moreover, if we pay attention to the above-mentioned ratios of the high level of returns more than 5,000, there is almost no difference between the ratios (56.3% vs. 56.6%) of the two groups. Hence, we need to look into the performance of all sample models in more detail.

One thing to note in relation to this point is that our interest is not in models with lower returns than their average return but in models with higher returns than average. This is the reason why we focused on the high level of returns in the range above 5,000, i.e., which is near to the total average (5425.7) of the two mean returns of the 1st and the 2nd groups as shown in the preceding subsection.

Another thing that we must not overlook—and is yet to be confirmed—is that every group may include models that have positive returns for the in-sample tests but experience negative returns for the out-of-sample tests, and vice versa. If such models are included in a group, their returns play a role to affect the mean return of the group for both the in- and out-of- sample tests. Therefore, excluding them will lead to a more accurate evaluation of the good-performing models and helps to address Hypothesis 3. For this reason, we extend our discussion to the following two questions: "How many sample models with *positive* returns for the in-sample tests were also able to achieve *positive* returns for the out-of-sample tests?" and "How good were the returns for these models over the whole period consisting of the two sample tests?"

5.2.3 Models With Positive Returns for Both the In-sample and the Out-of-sample Tests

In Table 7, see the row titled 'P to P' ('N to N') which indicates the number of models where returns for the in-sample tests are positive (negative) and returns for the out-of-sample tests are also positive (negative).

Table 7. Changes in the performance for the in-sample tests to the out-of-sample tests (2011–2021)

	A_{n1}−A_{n2}−A_{n3}		A_{n1}−B_{n2}−A_{n3}		A_{n1}−C_{n2}−A_{n3}		A_{n1}−A_{n2}−B_{n3}		Subtotal (a)		Subtotal (b)		Total	
	n	Ratio	n	Ratio	n	Ratio	n	Ratio	n	Ratio	n	Ratio	n	Ratio
N to N	0	(0.0%)	0	(0.0%)	0	(0.0%)	3	(2.6%)	0	(0.0%)	3	(1.3%)	3	(0.4%)
N to P	5	(2.9%)	0	(0.0%)	3	(2.8%)	13	(11.2%)	5	(1.1%)	16	(7.1%)	21	(3.1%)
P to N	0	(0.0%)	18	(6.3%)	27	(25.0%)	2	(1.7%)	18	(3.9%)	29	(12.9%)	47	(6.9%)
P to P	169	(97.1%)	270	(93.8%)	78	(72.2%)	98	(84.5%)	439	(95.0%)	176	(78.6%)	615	(89.7%)
Total	174	(100.0%)	288	(100.0%)	108	(100.0%)	116	(100.0%)	462	(100.0%)	224	(100.0%)	686	(100.0%)

Note: The column 'Subtotal(a)' indicates the sum of results for the 1st and the 2nd groups shown in the left side; the 'Subtotal(b)' indicates the sum of results for the 3rd and the 4th groups shown in the middle.

First of all, look at the number of 'P to P' models in the 1st group (A_{n1}−A_{n2}−A_{n3}) and the 2nd group (A_{n1}−B_{n2}−A_{n3}). Note that 169 (97.1%) of the models belonging to the 1st group and 270 (93.8%) of the models belonging to the 2nd group are 'P to P' models. The two ratios (97.1% vs. 93.8%) are almost the same, so it is still not possible to determine which of these two groups performs better and have more good-performing models. But, one thing is certain: almost all models belonging to these two groups achieved positive returns not only for the in-sample tests but also for the out-of-sample tests. Based on this finding, we move to the next issue: how good are the returns for these 'P to P' models for the whole period?

Before moving to that, let us point out here something about the five 'N to P' models in the 1st group and the eighteen 'P to N' models in the 2nd group. The former (latter) had negative (positive) returns for the in-sample tests but changed to have positive (negative) returns for the out-of-sample tests. In either case, these types of models do not merit further discussion.

Another point concerns the columns "Subtotal (a)" and "Subtotal (b)" in Table 7. We can see that the first two groups have 439 (95.0%) 'P to P' models while the last two groups have 176 (78.6%). This result suggests that the former has far more good-performing models than the latter. Second, look at the column "Total" on the right-hand-side and note that 615 (89.7%) of the models have turned out to be the 'P to P' models. This confirms that there is considerable validity to our methodology in finding good-performing models.

5.2.4 The Frequency Distribution of Returns Achieved by the 'P to P' Models

Table 8 shows more detailed information to confirm the profitability of the 'P to P' models belonging to the four groups.

Table 8. Frequencies of returns achieved by the 'P to P' models belonging to the four groups over the whole period which includes both in- and out-of-sample tests (2011–2021)

	$A_{n1}-A_{n2}-A_{n3}$		$A_{n1}-B_{n2}-A_{n3}$		$A_{n1}-C_{n2}-A_{n3}$		$A_{n1}-A_{n2}-B_{n3}$		Subtotal (a)		Subtotal (b)	
Return range	n	Ratio	n	Ratio	n	Ratio	n	Ratio	n	Ratio	n	Ratio
0 – 5,000	20	(11.8%)	0	(0.0%)	21	(26.9%)	32	(32.7%)	20	(4.6%)	53	(30.1%)
5,000 – 10,000	55	(32.5%)	25	(9.3%)	33	(42.3%)	50	(51.0%)	80	(18.2%)	83	(47.2%)
10,000 – 15,000	62	(36.7%)	113	(41.9%)	22	(28.2%)	11	(11.2%)	175	(39.9%)	33	(18.8%)
15,000 – 20,000	32	(18.9%)	120	(44.4%)	2	(2.6%)	5	(5.1%)	152	(34.6%)	7	(4.0%)
20,000 – 25,000	0	(0.0%)	12	(4.4%)	0	(0.0%)	0	(0.0%)	12	(2.7%)	0	(0.0%)
Total	169	(100.0%)	270	(100.0%)	78	(100.0%)	98	(100.0%)	439	(100.0%)	176	(100.0%)

In the case of the 1st group ($A_{n1}-A_{n2}-A_{n3}$), 18.9 (=18.9+0.0) percent of the 'P to P' models belonging to the group have higher returns of more than 15,000. In the case of the 2nd group ($A_{n1}-B_{n2}-A_{n3}$), 48.8 (=44.4+4.4) percent of the 'P to P' models belonging to the group have similar high levels of returns; moreover, the number of the high-performing 'P to P' models included in the 2nd group (132=120+12) is the highest of all groups. In addition to this, note the row titled "Total mean return" below in Table 9. It indicates the average return that was achieved by a 'P to P' model in each group over the whole period for the in- and out-of- sample tests. The mean return (14867.4) of the 2nd group is greater than that (10352.5) of the 1st group; furthermore, it is the highest of all groups. Based on these two results, we can say that "Hypothesis 3" is supported.

Turn back to Table 8 and look at the columns "Subtotal (a)" and "Subtotal (b)." The first two groups have 37.3 (=34.6+2.7) percent of the 'P to P' models with higher returns than 15,000 as stated above, while the ratio of the last two groups is just 4.0 (=4.0+0.0) percent. This implies that the former has far more good-performing models than the latter. From this, we can confirm again the validity of "Hypothesis 1" and "Hypothesis 2."

Table 9. The performance of the 'P to P' models belonging to the four groups over the whole period which includes both in- and out-of- sample tests (2011–2021)

	$A_{n1}-A_{n2}-A_{n3}$	$A_{n1}-B_{n2}-A_{n3}$	$A_{n1}-C_{n2}-A_{n3}$	$A_{n1}-A_{n2}-B_{n3}$	Total Average
Total mean return	10352.5	14867.4	8295.3	6955.0	11532.3
Annual mean return	941.1	1351.6	754.1	632.3	1048.4
Total average number of transactions	450.6	385.2	324.0	279.5	378.5
Annual average number of transactions	41.0	35.0	29.5	25.4	34.4

To summarize the important points that we have confirmed so far: (1) every model's performance examined for the in- and out-of- sample tests depends solely on its three parameter value settings; (2) the parameter value settings of the 'P to P' models made it possible to accomplish good and high performances as we have seen above; and, (3) their performance has been verified for the long period of the two sample tests which cover the last 11 years.

For these reasons, we conclude that *all the 'P to P' models have good-performing 'optimal' parameter values*. In particular, the parameter value combinations of the 'P to P' models belonging to the 2nd group are most optimal in the sense that their returns for the in- and the out-of- sample tests turned out to be all positive; and the ratio (48.8%) of the high-preforming 'P to P' models is the highest of all groups.

See Table 9 and note that the total mean return (14867.4) of the 2nd group is almost same level as the level of returns (15,000) that we considered on so far; and both are higher than the total average (11532.3) achieved by all the 'P to P' models over the whole period. Now, the annual mean return '1351.6' of the 2nd group in Table 9 implies that every 'P to P' model of the 2nd group has earned 1,351,600 Japanese yen (=1351.6×1000) a year on average—for the past 11 years by trading repeatedly only one unit at a time. It is approximately 13,030 US dollars based on the average exchange rate of ¥/$=103.73 for the whole period. At this point, one may ask: "Which model has the best optimal parameter values?" or "Which parameter value combinations are best optimal?" This question is taken up later.

5.2.5 The Performance of the Worst-performing Group for the Out-of-sample Tests

One remaining issue is to verify Hypothesis 4: "Do the sample models belonging to the worst-performing group perform better in the out-of-sample tests?" Table 10 shows that all of the models belonging to this group still had negative returns—as was observed for the in-sample tests. Their mean return 'minus 9035.5' is a clear contrast to those of the four groups with all positive mean returns (see again Table 5). This result confirms that "Hypothesis 4" is supported.

Table 10. Summary statistics of the worst-performing group (2020–2021)

	n	Max	Min	Mean	Median	S.E	Skewness	Kurtosis
$W_{n1}–W_{n2}–W_{n3}$	42	-5180	-–10620	-9035.5	-9460	167.2	1.0650	2.1111

Although now obvious, it is appropriate to say that we cannot expect good performance from this group of models. We therefore conclude that *good performance comes from good parameter value settings, and vice versa*—which is strong evidence to support the main assertion of this study.

5.2.6 The Best-performing Parameter Value Combinations

To discuss "which model is the best?" is beyond the scope of this brief paper and would undoubtedly obscure the main point of our discussion. Instead, we propose to answer the following three questions. It will give helpful information to traders who are looking for an optimal model.

- Which group has the most numerous high-performing models?
- What parameter values are most often used in the high-performing models?
- What characteristics do the high-performing models have?

For the first question, we can answer: *it is the 2nd group*. The reasons for this can be summarized as follows:

(1) The 2nd group of models have the highest mean return (9554.4) for the in-sample tests. In addition, a large number of models (140, 48.6%) belonging to this group have higher returns than 10,000 for the in-sample tests; this number of high-performing models is the highest of all groups. (See Table 2 and Table 3.)

(2) Almost the same observations as stated above can be made for the performance of the 2nd group of models for the out-of-sample tests. The mean return (4930.7) of this group of models for the out-of-sample tests and the ratio (56.6%) of high-performing models with higher returns than 5,000 are as good as those (5920.7, 56.3%) of the 1st group, which makes a big difference from the other two groups. (See Table 5 and Table 6.)

(3) In more detail, almost all (270, 93.8%) of the 2nd group of models turned out to be the 'P to P' models that produce positive returns not only for the in-sample tests but also for the out-of-sample tests. In addition, almost half (132, 48.8%) of the 'P to P' models belonging to this group have higher levels of returns than 15,000 over the whole period of the two sample tests. This number and ratio of high-performing models are overwhelmingly greater than those of the other groups. (See Table 7 and Table 8.)

For the second question, we investigated the frequencies of the three parameter values being applied to the above-mentioned 132 high-performing 'P to P' models and found that there is no particular range of values that are used intensively. That is, all of the three parameter values that have been pre-defined as optimal parameter values of the sample models of the 2nd group—n_1: {3,···,8}, n_2: {18,···,25}, n_3: {3,···,8}—are used without exception. This implies that every combination of the three parameter values derived from the above three brackets are all optimal ones—although there was a difference in whether their resulting returns are higher than 15,000 or not. This alone may be enough to explain why the 2nd group of sample models achieved such good performance as the three items stated in the preceding paragraph.

As for the third question, look again at the values that were presented in the three brackets above. We can then see that the resulting combination of the three parameter values (n_1, n_2, n_3) make a characteristic form like a 'top hat' in that the second parameter value n_2 has a longer length than the other two parameters n_1 and n_3— i.e., "$n_1 < n_2$ and $n_2 > n_3$." This finding admits an interpretation that this form of the three parameter values is an optimal fit for the Japanese futures market for the last 11 years.

5.2.7 A Comparison of the Performances of the ‘P to P’ Models With the Buy-and-hold Strategy

Let us now consider whether the above-mentioned ‘P to P’ models in the 2nd group can statistically outperform the classic benchmark buy-and-hold trading strategy. For the statistical significance test, log returns of monthly and ten-day buy-and-hold strategies are compared with log returns produced by the ‘P to P’ models (Note 1). Yet, among the 270 ‘P to P’ models belonging to the 2nd group, there was no model that consistently generated statistically significant returns over the whole test period from 2011 to 2021—although some had significant returns over shorter periods. As for the reason for this, it could be that the buy-and-hold strategy has the potential to deliver better performance than technical trading rules during long-term clear uptrends in the Japanese market as we have seen in Figure 1.

Table 11 shows that the ‘P to P’ models compare favorably to the two buy-and-hold strategies. Look at the row titled “Average”. It demonstrates that for five out of the 11 years, the annual mean return of the 270 ‘P to P’ models is higher than the monthly and/or ten-day buy-and-hold strategies. Also, the total annual mean return (1352) of the 270 ‘P to P’ models over the 11 years is almost the same or higher than the two buy-and-hold strategies (1348 and 1221). Another noteworthy point is that the median (1734) of the annual average retunes of the 270 ‘P to P’ models is higher than those (1510 and 1130) of the two buy-and-hold strategies. This implies, if we judge from the median, as the total performance over the 11 years, that more than half of the 270 ‘P to P’ models perform better than the two buy-and-hold strategies.

Table 11. The returns of the two buy-and-hold strategies and the ‘P to P’ models of the 2nd group

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Mean	Median	Total
Panel A: Returns of the monthly and ten-day buy-and-hold strategies														
Monthly	-2470	2310	5580	990	2190	1100	1510	-3260	3370	4470	-960	1348	1510	14830
Ten-day	-1460	1960	6300	2140	-350	-910	1440	-1450	810	3820	1130	1221	1130	13430
Panel B: Returns of the 270 ‘P to P’ models belonging to the 2nd group														
Average	**378**	1756	1898	-580	**2337**	–1668	618	**1734**	**3057**	3605	**1732**	1352	1734	14867
Panel C: Examples of good performing and highly effective models in the year of 2020														
(5,18,7)	-220	**1970**	1730	-70	**2740**	-1600	410	**2580**	**1480**	**7000**	**3150**	1743	1730	19170
(7,25,3)	**580**	1890	1230	-–960	**3100**	**500**	1290	**1620**	**3300**	**6590**	**2920**	2005	1620	22060
(4,22,6)	**320**	1930	2490	-720	**2680**	**-620**	790	**1580**	**2180**	**6280**	**2540**	1768	1930	19450

Note: Bold font represents trading returns whereby the trading returns of the ‘P to P’ models exceeded the returns from the monthly and/or ten-day buy-and-hold strategy.

A further point to note is “How do the ‘P to P’ models in the 2nd group perform in the pandemic period (2020–2021)?” Recall the situation described in Section 5.2 about the pandemic period and look at the results shown in the two columns “2020” and “2021” in Table 11. We can then see that the total mean return (5337=3605+1732) of the 270 ‘P to P’ models for the two years is higher than those (3510=4470–960, 4950=3820+1130) of the two buy-and-hold strategies. This implies that many of the ‘P to P’ models were more profitable during the pandemic period. Let us focus on the returns in the year 2020 when the market was fluctuating like a roller-coaster. Look at the “Panel C" in the bottom of the table. It shows a handful of good performing and highly effective models—especially in 2020 (Note 2). Their average returns (7000, 6590, 6280) in this year are much higher than those (4470, 3820) of the two buy-and-hold strategies. Note that these three example models also have relatively good performance during the normal period (2011–2019) before COVID-19. This shows that the 270 ‘P to P’ sample models include many models that perform well not only during the “normal” period but also during “abnormal” times of crisis. Regarding these models with high profitability in the year 2020, it may be that excess volatility triggered by the COVID-19 pandemic contributed to the profitability of technical analysis.

Regarding this point, the research of Lento and Gradojevic (2022) is relevant. They explored the profitability of technical trading rules (including the traditional moving average cross-over rule) around the COVID-19 pandemic market crash (from January to May 2020) for five asset markets (bitcoin, gold, oil, and so on). They reported that “Overall, the analysis reveals that many trading rules could generate positive profits on the observed data before transaction costs. However, most of these profits were not robust enough to persist through transaction costs and statistical significance testing (p.7)".

6. Summary and Concluding Remarks

Finding optimal parameter values is the most crucial matter for traders who use the MACD tool because the buying and selling signals generated from it depend on its three parameter settings. The same thing can be said for academic researchers who test the effectiveness of the MACD approach or use it to measure market efficiency. This is because optimal parameter values should be used to get the best result possible for their chosen assets or markets as well as to justify their research. However, much of the existing MACD literature is biased toward examination of the profitability of models with traditional parameter values. Overall, this previous research reported negative results for the technical trading approach and failed to reject market efficiency. In this literature, only a few papers looked for the optimal parameter values from a large number of sample models with different parameter values. Two of these are Erić et al. (2009) and Borowski and Pruchnicka-Grabias (2019). The common purpose of these two papers was to find the most profitable parameter combination for each company listed on each stock market they tested.

It would be useful for traders to know the most profitable parameter combinations for individual companies and how different they are from each other. But if we know the answers to the questions "what parameter values are most optimal to use in a market?" and "what characteristics do the optimal parameter values have?", we would gain insight and broader perspectives about markets. Each trader's strategy could be optimized and suited to each market. However, little attention has been given to the method of how best to select parameter values.

This paper applied a new methodology to find optimal parameter values of the MACD for the Japanese Nikkei 225 futures market and found answers to the above questions:

- The most often optimized MACD transaction system on the market has three parameter values in the following ranges— n_1: {3,⋯,8}, n_2: {18,⋯,25}, and n_3: {3,⋯,8}.
- The resulting combinations of the three parameter values from these ranges make forms like a 'top hat' in the sense that the second parameter (n_2) is longer than the other two, i.e., "$n_1 < n_2$ and $n_2 > n_3$".

This is the main result of this paper. Its evidence is summarized in Section 5.2.6. From the above results, traders participating in the market can get useful insights as to how to create a model matching one's trading style and goals. For example, if one prefers short-term (long-term) investment, he/she can select three values from the three specified ranges presented in the three brackets above and create a model where the parameter interval lengths are relatively short (long) to match one's preferred frequency of signal generation. In addition to this and noting the finding presented in Table 9—the best-performing models with the parameter values stated above have an average number of annual transactions of 35.0 (about 3 transactions a month), one can optimize one's trading strategy with a broader perspective on the market. That is, the most often used optimal parameter value combinations stated above are suitable for weekly or ten-day swing trades. In other words, the best optimal trading interval for the MACD is neither a short period like a few days nor a long period like a month.

This study also performed a comparative analysis of the performance of the sample models with the hypothetical optimal and non-optimal parameter values. The results were quite surprising—none of models with the non-optimal parameter values had positive returns for both the in- and out-of- sample tests which cover the last 11 years (see Table 4 and Table 10) while almost all of the models with the optimal parameter values performed well. This salient result confirmed that *good performance comes from good parameter value settings, and vice versa*. In this respect, it is interesting to note here again the specified ranges of the three non-optimal parameter values— n_1: {13,⋯,19}, n_2: {31,⋯,33}, and n_3: {34,35}. The resulting combinations of the three parameter values from the three specified ranges make a form like 'two cascading falls' in that the difference between the first and the second parameters is greater than the difference between the second and the third parameters, i.e., "$(n_2 - n_1) > (n_3 - n_2)$ where "$n_1 < n_2 < n_3$". We found that the worst-performing models with the parameter values stated above have 9.4 (less than 1 transaction a month) transactions on average a year. This suggests that these combinations with relatively longer values than the optimal parameter values stated above do not perform well for the market.

Several other suggestions and insights can be obtained from the results of this study. What is most significant is that: *for a major stock market (e.g., the Japanese Nikkei 225 futures index), a simple technical analysis tool (i.e., the MACD) has been most often optimized—for a long time (at least for the last 11 years)—by the three parameter values in the three specified ranges which are presented in this paper*. This finding deserves more than passing notice since it might leave room for argument about the EMH. As recently suggested by Kang (2021), the Japanese market is not weak-form efficient in the sense that the market does not incorporate all public information. This suggestion is consistent with the work of Anghel (2015) who assessed the state of information efficiency of the stock markets of

75 countries. According to his ranking of relative market efficiency, Japan and the United States belonged to a group of 34 countries where it was possible to obtain abnormal profits by using one of two MACD trading rules.

Inefficiency means the existence of opportunities to exploit profits. If traders exploit new information such as the findings in this study to optimize their trading strategies, their collective behavior and expectations may affect the dynamics of the market until the possibility of profit disappears. However, as Biondo et al. (2013) wrote, since Fama “traders and financial analysts continuously seek to expand their information set to gain the opportunity to choose the best strategy (p.3).” New information will continue to be discovered and exploited. In connection with this point, there is one thing to note about the above-mentioned research.

Borowski and Pruchnicka-Grabias (2019) investigated optimal parameter values of MACD models for 140 companies listed on the Warsaw Stock Exchange and concluded that “the transaction system was optimized mainly by short moving averages (p.464)” and “the moving averages that optimize the tested transaction system are most often a few sessions long (p.468)” with an attached list of the most profitable parameter values for each of the companies in their appendix. Interestingly, they extended their study to consider the regularity of the optimal parameter values and whether each value of the best-performing three parameters was an even or an odd number. From the results of this analysis, they reported “the highest rates of returns were obtained for the “odd-even-odd” combinations of the moving averages (p.468).” This is the only attempt to determine the characteristics of the best-performing models, to the best knowledge of this author.

In contrast to these results with no explicit mention of the lengths of “short moving averages”, this study specifies the ranges of three optimal parameter values and describes the characteristic form of the three parameter values—with its practical implications as described above. In these respects, this paper contributes to the existing literature.

There is no point to add to the oversimplified discovery of the “odd-even-odd” combinations. But Borowski and Prucinicka-Grabia’s list provides an opportunity to compare the Polish market and the Japanese market. Comparing the 140 sets of the most profitable parameter values in their list with the 288 combinations of the best-performing three parameter sets in this study shows that there were only 4 identical sets. This confirms the very interesting point that the optimal parameter value settings for the two countries’ financial markets are different from each other (Note 3). Therefore, if we exploit this possibility for financial markets in other countries in future study, it provides more insightful perspectives that are unique for each market, which help to better understand the dynamics of global financial markets.

This research is limited because its target is the Japanese futures market. Accordingly, the feedback effects of global correlations between different financial markets have to be considered when exploring the validity of the new methodological approach presented in this research. Explicit modeling of global linkages between financial markets might improve the development of MACD models.

Acknowledgements

The author would like to acknowledge Prof. Marc Bremer for his proofreading and valuable comments. The author also thanks Masayuki Arakawa for his detailed checking of the empirical results of this paper. The author appreciates valuable suggestions from the anonymous reviewers.

References

Abbey, B. S., & Doukas J. A. (2012). Is technical analysis profitable for individual currency traders?. *Journal of Portfolio Management*, *39*, 142-150. https://doi.org/10.3905/jpm.2012.39.1.142

Anghel, G. D. (2015). Stock market efficiency and the MACD: Evidence from countries around the world. *Procedia Economics and Finance*, *32*, 1414-1431. https://doi.org/10.1016/S2212-5671(15)01518-X

Appel, G. (1979). *The moving average convergence divergence trading method*. Great Neck, NY: Signalert Corp.

Armour, J., Lofton, M., Oyenekan, O., & Metghalchi, M. (2010). Efficient market hypothesis and technical analysis: The Irish stock index. In *International handbook of academic research and teaching 71*. Heidelberg, Berlin: Springer.

Biondo, A. E., Pluchino A., Rapisarda, A., & Helbing, D. (2013). Are random trading strategies more successful than technical ones?. *PLoS ONE*, *8*(7), e68344, 1-13. https://doi.org/10.1371/journal.pone.0068344

Borowski, K., & Pruchnicka-Grabias, I. (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*(5), 457-478. Retrieved from https://bankikredyt.nbp.pl/content/2019/05/BIK_05_2019_02.pdf

Chen C-P., & Metghalchi, M. (2012). Weak-form market efficiency: Evidence from the Brazilian stock market. *International Journal of Economics and Finance*, *4*(7), 22-32. http://dx.doi.org/10.5539/ijef.v4n7p22

Chong, T. T., & Ng, W-K. (2008). Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30. *Applied Economics Letters*, *15*(14), 1111-1114. https://doi.org/10.1080/13504850600993598

Chong, T. T., Ng, W-K., & Liew, V. K. (2014). Revisiting the performance of MACD and RSI oscillators. *Journal of Risk and Financial Management*, *7*(1), 1-12. https://doi.org/10.3390/jrfm7010001

Du Plessis, A. W. (2013). The effectiveness of a technical analysis strategy versus a buy-and-hold strategy on the FTSE/JSE Top 40 index shares of the JSE Ltd: The case of the moving average convergence divergence indicator. *Ph.D. dissertation*, University of Johannesburg, Johannesburg, South Africa. Retrieved from http://hdl.handle.net/10210/8408

Erić, D., Andjelic, G. B., & Redžepagić, S. (2009). Application of MACD and RVI indicators as functions of investment strategy optimization on the financial market. *Journal of Economics and Business*, *27*(1), 171-196. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2267753

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383-417. https://doi.org/10.2307/2325486

Hejase, A. J., Srour R. M., Hejase, H, J., & Younis, J. (2017). Technical analysis: Exploring MACD in the Lebanese stock market. *Journal of Research in Business, Economics and Management*, *8*(4), 1493-1502. Retrieved from https://scitecresearch.com/journals/index.php/jrbem/article/view/1077

Kang, B-K. (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. https://doi.org/10.3390/jrfm14010037

Lento C., & Gradojevic, N. (2022). The profitability of technical analysis during the COVID-19 market meltdown. *Journal of Risk and Financial Management, 15*, 192. https://doi.org/10.3390/jrfm15050192

Meissner, G., Alex, A., & Nolte, K. (2001). A refined MACD indicator–Evidence against the random walk hypothesis?. *ABAC Journal*, *21*(1), 10-17. Retrieved from http://www.assumptionjournal.au.edu/index.php/abacjournal/article/view/732/650

Menkhoff, L. (2010). The use of technical analysis by fund managers: International evidence. *Journal of Banking & Finance*, *34*(11), 2573-2586. https://doi.org/10.1016/j.jbankfin.2010.04.014

Murphy, J. (1999). *Technical analysis of the financial markets*. New York: Prentice Hall Press.

Nor, S. M., & Wickremasinghe, G. (2014). The profitability of MACD and RSI trading rules in the Australian stock market. *Investment Management and Financial Innovations*, *11*(4), 194-199. Retrieved from https://www.researchgate.net/publication/272238241_The_profitability_of_MACD_and_RSI_trading_rules_in_the_Australian_stock_market

Park, C-H., & Irwin, S. H. (2007). What do we know about the profitability of technical analysis?. *Journal of Economic Surveys*, *21*(4), 786-826. https://doi.org/10.1111/j.1467-6419.2007.00519.x

Rosillo, R., Fuente D., & Brugos, J. (2013). Technical analysis and the Spanish stock exchange: Testing the RSI, MACD, momentum and stochastic rules using Spanish market companies. *Applied Economics*, *45*(12), 1541-1550. https://doi.org/10.1080/00036846.2011.631894

Wiles, P., & Enke, D. (2015). Optimizing MACD parameters via genetic algorithms for soybean futures. *Procedia Computer Science*, *61*, 85-91. https://doi.org/10.1016/j.procs.2015.09.157

Notes

Note 1. The ten-day buy-and-hold strategy is considered so as to compare it fairly to the ‘P to P’ models in the sense of matching the number of transactions to that of the ‘P to P’ models. Note that the average number of transactions is approximately 3 trades a month for both.

Note 2. We observed that 132 (142) models among the 270 ‘P to P’ models have higher returns than that of the monthly (ten-day) buy-and-hold strategy in the year of 2020.

Note 3. Erić et al. (2009) also provided a list of the most profitable parameter values for the 48 companies listed on the Belgrade Stock Exchange. But none of these were identical to the 288 combinations of this research.

Appendix

Table 1. Selected ranges of optimal, sub-optimal and non-optimal parameter values based on the frequencies of the three parameter values (n_1, n_2, n_3) used in the top 100 and the bottom 100 models

Length	Parameter n_1			Parameter n_2			Parameter n_3		
	Best	P/F	Worst	Best	P/F	Worst	Best	P/F	Worst
3	**59.0**	A_{n1}	0.0	–		–	**22.0**	A_{n3}	0.0
4	**12.0**	A_{n1}	0.0	–		–	**9.0**	A_{n3}	0.0
5	1.0		0.0	0.0		0.0	1.0		0.0
6	3.0		0.0	3.0		0.0	2.0		0.0
7	0.0		0.0	0.0		0.0	0.0		0.0
8	**11.0**	A_{n1}	3.0	**11.0**	A_{n2}	0.0	**9.0**	A_{n3}	2.0
9	3.0	x	**11.0**	**6.0**	A_{n2}	0.0	3.0	x	**13.0**
10	3.0	x	**11.0**	**5.0**	A_{n2}	0.0	2.0	x	**16.0**
11	**4.0**	x	**13.0**	**4.0**	A_{n2}	0.0	**4.0**	x	**15.0**
12	3.0	x	**10.0**	**4.0**	A_{n2}	2.0	**4.0**	x	**9.0**
13	0.0	x	**6.0**	3.0		3.0	**5.0**	x	**6.0**
14	1.0	x	**7.0**	2.0		3.0	2.0	x	**5.0**
15	0.0	x	**8.0**	3.0		2.0	2.0		2.0
16	0.0	x	**10.0**	3.0		1.0	2.0		2.0
17	0.0	x	**9.0**	3.0		1.0	1.0		1.0
18	0.0	x	**6.0**	**5.0**	B_{n2}	1.0	2.0	x	**4.0**
19	0.0	x	**4.0**	1.0		0.0	1.0		2.0
20	0.0	x	2.0	1.0		0.0	2.0		1.0
21				**4.0**	B_{n2}	2.0	2.0		3.0
22				**4.0**	B_{n2}	0.0	2.0		2.0
23				2.0		1.0	1.0		2.0
24				0.0		1.0	0.0		1.0
25				**4.0**	B_{n2}	2.0	3.0		1.0
26				2.0		3.0	2.0	B_{n3}	0.0
27				3.0	x	**4.0**	3.0	B_{n3}	0.0
28				**5.0**	x	**5.0**	3.0	B_{n3}	0.0
29				3.0	x	**10.0**	2.0	B_{n3}	0.0
30				2.0	x	**7.0**	1.0		1.0
31				0.0	x	**9.0**	1.0		1.0
32				0.0	x	**7.0**	1.0		1.0
33				0.0	x	**13.0**	0.0		2.0
34				3.0	x	**10.0**	1.0	x	**4.0**
35				3.0	x	**9.0**	1.0	x	**4.0**
36				3.0		3.0	1.0		0.0
37				1.0		1.0	0.0		0.0
38				3.0	C_{n2}	0.0	1.0		0.0
39				2.0	C_{n2}	0.0	1.0		0.0
40				2.0	C_{n2}	0.0	1.0		0.0

Note: Blue (Red) boxes indicate ranges of optimal (sub-optimal) parameter values; and, gray boxes indicate ranges of non-optimal parameter values.