A Comprehensive Look at the Empirical Performance of Moving Average Trading Strategies*

Valeriy Zakamulin†

This revision: December 11, 2015

Abstract

Despite the enormous current interest in market timing and a series of publications in academic journals, there is still lack of comprehensive research on the evaluation of the profitability of trading rules using methods that are free from the data-snooping bias. In this paper we utilize the longest historical dataset that spans 155 years and extend previous studies on the performance of moving average trading rules in a number of important ways. Among other things, we investigate whether overweighting the recent prices improves the performance of timing rules; whether there is a single optimal lookback period in each trading rule; and how accurately the trading rules identify the bullish and bearish stock market trends. In our study we, for the first time, use both the rolling- and expanding-window estimation scheme in the out-of-sample tests; study the performance of trading rules across bull and bear markets; and perform numerous robustness checks and tests for regime shifts in the stock market dynamics. Our main results can be summarized as follows: There is strong evidence that the stock market dynamics are changing over time. We find no statistically significant evidence that market timing strategies outperformed the market in the second half of our sample. Neither the shape of the weighting function nor the type of the out-of-sample estimation scheme allows a trader to improve the performance of timing rules. All market timing rules generate many false signals during both bullish and bearish stock market trends, yet these rules tend to outperform the market in bear states.

Key words: technical analysis, market timing, moving averages, regime switching, bull and bear markets, out-of-sample testing

JEL classification: G11, G17.

*The author is grateful to Steen Koekebakker, Vadim Linetsky, Pete Nikolai, and the participants of the 3rd Economics & Finance Conference (Rome, Italy) for their insightful comments on the earlier drafts of this paper. The usual disclaimer applies.

†a.k.a. Valeri Zakamouline, School of Business and Law, University of Agder, Service Box 422, 4604 Kristiansand, Norway, Tel.: (+47) 38 14 10 39, E-mail: `Valeri.Zakamouline@uia.no`

1 Introduction

Technical analysis represents a methodology of forecasting the future price movements through the study of past price data and uncovering some recurrent regularities, or patterns, in price dynamics. One of the fundamental principles of technical analysis is that prices move in trends. Analysts firmly believe that these trends can be identified in a timely manner to generate profits and limit losses. Market timing is an active trading strategy that implements this idea in practice. Specifically, this strategy is based on switching between the market and the cash depending on whether the prices trend upward or downward. A moving average is one of the oldest and most popular tools used in technical analysis for detecting a trend.

A moving average of prices is calculated using a fixed size data window that is rolled through time. The length of this window of data, also called the lookback period or averaging period, is the time interval over which the moving average is computed. Moving average strategies are simple to understand because the trading signals are computed using primitive rules and the results of market timing can be easily visualized in a compelling manner. Yet, despite a long history, modern market timing with moving averages still remains art rather than science. This is because there exists several popular trading rules, many types of moving averages, and lots of possible choices for the length of the lookback period. One of the unresolved controversies in market timing is over which combination of trading rule, moving average weighting scheme, and the length of the lookback period produces the best performance.

Whereas technical analysis has been extensively used by traders for over a century and the majority of active traders strongly believe in market timing, academics had long been skeptical about the usefulness of technical analysis. The academics' attitude towards the technical analysis was altered by a series of papers published in prominent academic journals,[1] because the findings in these papers tend to suggest that one should not dismiss the value of technical analysis. Recently we have witnessed a constantly increasing interest in technical analysis from both the practitioners and academics alike (Park and Irwin (2007)). This interest developed because over the decade of 2000s many technical trading rules outperformed the market by a large margin. Nowadays many academics seem to have gone from one extreme to another and

[1] Examples are Brock, Lakonishok, and LeBaron (1992), Brown, Goetzmann, and Kumar (1998), Neely, Weller, and Dittmar (1997), Lo, Mamaysky, and Wang (2000), Okunev and White (2003), and Moskowitz, Ooi, and Pedersen (2012).

believe that one can beat the market by using some technical trading rules. Numerous papers published in academic journals[2] after the global financial crisis of 2007-08 left their readers with the impression that "market timing works".

However, despite the enormous current interest in market timing and a series of publications in academic journals, there is still a great shortage of comprehensive research on the evaluation of the profitability of technical trading rules using methods that are free from the data-snooping bias. Our goal in this paper is to fill this gap in the literature and critically re-examine the empirical performance of moving average trading rules. Moreover, in this paper we extend previous studies in a number of important ways.

First, in this paper we utilize a historical dataset that spans 155 years and perform the longest out-of-sample test of a few market timing rules over the period of 145 years. It is worth mentioning that, to the best knowledge of the author, there are only two papers to date in which the researchers implement out-of-sample tests of profitability of some trading rules in the stock market. Specifically, in the studies by Sullivan, Timmermann, and White (1999) and Zakamulin (2014) the length of the out-of-sample period was 10 and 84 years respectively. Presumable, using a much longer history can provide us with much richer information about the performance of market timing rules.

Second, in previous studies the researchers usually selected a set of so-called "most popular combinations" of trading rules with moving average weighting schemes (examples are Brock et al. (1992), Brown et al. (1998), Sullivan et al. (1999), Okunev and White (2003)) without any analysis of commonalities and differences between miscellaneous choices for trading rules and moving average weighting schemes. However, in a recent study by Zakamulin (2015) the author uncovers the anatomy of market timing rules with moving averages and demonstrates that the computation of every technical trading indicator can equivalently be interpreted as the computation of a weighted moving average of price changes. Consequently, the only real difference, between diverse market timing rules coupled with various types of moving averages, lies in the weighting scheme used to compute the moving average of price changes. Motivated by this result, in sharp contrast to previous studies we select a set of rules with clearly distinct shapes of the moving average weighting scheme. This approach to selecting the set of rules

[2] Examples are Faber (2007), Gwilym, Clare, Seaton, and Thomas (2010), Kilgallen (2012), Moskowitz et al. (2012), Clare, Seaton, Smith, and Thomas (2013), Pätäri and Vilska (2014), and Glabadanidis (2015).

allows us to test the common belief among the traders that in the computation of a moving average one has to overweight the most recent prices because they contain more relevant information on the future direction of the price than earlier prices (see Murphy (1999), Chapter 9, and Kirkpatrick and Dahlquist (2010), Chapter 14).

Third, there is still a major controversy among technical traders about the optimal length of the lookback period in each trading rule. Analysis of the popular advice reveals that traders believe that in each trading rule there exits some certain, time-invariant, length of the lookback period that produces the best performance, but they strongly disagree on its specific length. Even for the most popular simple moving average trading rule, the popular advice on the length of the lookback period varies from 10 to 200 days (Kirkpatrick and Dahlquist (2010), Chapter 14). In previous studies the researchers used several choices for the lookback period in each tested rule. In our study we, for the first time, analyze the time variations in the length of the optimal lookback period for each trading rule.

Forth, since we find evidence the that length of the optimal lookback period is not stable over time, in our out-of-sample tests we, for the first time, use both the rolling- and expanding-window estimation scheme to dynamically determine the length of the optimal lookback period. This allows us to find out which estimation scheme produces the best performance of a market timing strategy. It is worth noting that the expanding-window estimation scheme is used when the parameter of estimation is supposed to be constant, whereas the rolling-window estimation scheme is used when parameter instability is suspected.

Our fifth extension of the previous studies is motivated by the following idea. Starting with Charles H. Dow, technical analysts believe that there are two types of primary trends in the stock market: Bullish and Bearish. A Bull (Bear) market is defined as a period of generally rising (declining) prices. Since a market timing rule is supposed to detect a trend in the stock market, the Buy-Sell signals generated by a trading rule must coincide with the Bull-Bear market states. In our study we, for the first time, measure the similarity between the Buy-Sell trading signals and the Bull-Bear stock market states. In addition, we evaluate the performance of market timing strategies not only over the periods that span a series of interchanging Bull and Bear markets, but also over the Bull and Bear markets separately.

Despite the fact that over a very long horizon (which is beyond the investment horizon of most individual investors) an active timing strategy seems to outperform the passive strategy, the outperformance generated by an active strategy is highly uneven over time. Therefore, as argued by Zakamulin (2014), the traditional performance measurement, which consists in reporting a single number for performance, is very misleading for investors with short- to medium-term horizons. To give a broader and clearer picture of market timing performance, we provide a detailed descriptive statistics of market timing performance over 5-year investment horizons.

Our sixth extension of the previous studies consists in the following. Even though analyzing the longest possible history of the empirical performance of technical trading rules allows us to better understand its properties, it is dangerous to assume that the observed performance over a very long run can be used as a reliable estimate of the expected performance in the near future. This is because not only the performance of a market timing strategy is highly uneven over a short run, but also the medium- to long-run performance of a market timing strategy may change over time. Such a change might occur as a result of changing dynamics of the stock market. In order to check whether we can use the observed performance over a very long run as a reliable estimate of expected future performance, we split the whole sample period into two sub-periods and perform a series of robustness tests and tests for regime shifts. Specifically, we test for structural breaks in the growth rate of the stock market index and for regime shifts in the return distribution of the stock market prices and the dynamics of the Bull-Bear stock market cycles. In addition, we test the stability of the length of the mean optimal lookback period in market timing rules over time. Last but not least, we compare the performances of trading rules over the two sub-periods and check whether we have evidence that the market timing strategies outperform the market in both sub-periods.

The rest of the paper is organized as follows. Section 2 presents the set of tested moving average rules whereas Section 3 presents our data and tests for regime shifts between subsample periods. In Section 4 we determine the turning points between the Bull and Bear markets, present the descriptive statistics on the Bull and Bear markets, and test for the regime shift in the dynamics and parameters of Bull and Bear markets. Section 5 presents our empirical research design, whereas Section 6 present the results of our empirical tests. In Section 7 we discuss the empirical results. The concluding remarks to our study are given in Section 8.

2 Market Timing Rules and Moving Average Weighting Schemes

A moving average of prices is calculated using a fixed size data "window" that is rolled through time. The length of this window of data, also called the lookback period or averaging period, is the time interval over which the moving average is computed. Denote by P_t the period t closing price of a stock market index. Furthermore, denote by $MA_t(k)$ the general weighted moving average at period-end t with k lagged prices. The general weighted moving average is computed using the following formula:

$$MA_t(k) = \frac{w_t P_t + w_{t-1} P_{t-1} + w_{t-2} P_{t-2} + \ldots + w_{t-k} P_{t-k}}{w_t + w_{t-1} + w_{t-2} + \ldots + w_{t-k}} = \frac{\sum_{j=0}^{k} w_{t-j} P_{t-j}}{\sum_{j=0}^{k} w_{t-j}}, \tag{1}$$

where w_{t-j} is the weight of price P_{t-j} in the computation of the weighted moving average. The most commonly used types of moving averages are: the Simple Moving Average (SMA), the Linear (or linearly weighted) Moving Average (LMA), and the Exponential Moving Average (EMA). A less commonly used type of moving average is the Reverse Exponential Moving Average (REMA). These moving averages at month-end t are computed as

$$\begin{aligned} SMA_t(k) &= \frac{1}{k+1} \sum_{j=0}^{k} P_{t-j}, \quad LMA_t(k) = \frac{\sum_{j=0}^{k} (k-j+1) P_{t-j}}{\sum_{j=0}^{k} (k-j+1)}, \\ EMA_t(k) &= \frac{\sum_{j=0}^{k} \lambda^j P_{t-j}}{\sum_{j=0}^{k} \lambda^j}, \quad REMA_t(k) = \frac{\sum_{j=0}^{k} \lambda^{k-j} P_{t-j}}{\sum_{j=0}^{k} \lambda^{k-j}}, \end{aligned} \tag{2}$$

where $0 < \lambda \leq 1$ is a decay factor.

The most popular trading rules used for timing the market are: the Momentum rule (MOM), the Price-minus-Moving-Average rule (P-MA), the Moving-Average-Change-of-Direction rule (ΔMA), and the Double-Crossover Method (DCM). The technical trading indicators in these rules are computed as

$$\begin{aligned} \text{Momentum rule:} \quad & \text{Indicator}_t^{\text{MOM}(k)} = P_t - P_{t-k}, \\ \text{Price-minus-Moving-Average rule:} \quad & \text{Indicator}_t^{\text{P-MA}(k)} = P_t - MA_t(k), \\ \text{Moving-Average-Change-of-Direction rule:} \quad & \text{Indicator}_t^{\Delta\text{MA}(k)} = MA_t(k) - MA_{t-1}(k), \\ \text{Double-Crossover Method:} \quad & \text{Indicator}_t^{\text{DCM}(s,k)} = MA_t(s) - MA_t(k), \end{aligned}$$

where $s < k$ defines the size of a shorter window. In all these market timing rules, the Buy signal is generated when the value of a technical trading indicator is positive. Otherwise, the Sell signal is generated.

Zakamulin (2015) demonstrates that despite being computed seemingly differently at the first sight, all technical trading indicators presented above are computed in the same general manner. In particular, the computation of every technical trading indicator can equivalently be interpreted as the computation of the weighted moving average of price changes. Specifically, every technical trading indicator can be equivalently computed using the following formula:

$$\text{Indicator}_t = \frac{\sum_{i=1}^{k} y_{t-i} \Delta P_{t-i}}{\sum_{i=1}^{k} y_{t-i}}, \tag{3}$$

where $\Delta P_{t-i} = P_{t-i+1} - P_{t-i}$ denotes the price change over the period from $t-i$ to $t-i+1$, and y_{t-i} is the weight of ΔP_{t-i} in the computation of the moving average of price changes. The weights y_{t-i} are computed using the weights $\{w_t, w_{t-1}, \dots, w_{t-k}\}$ that specify how the moving average of prices is computed.

The result given by formula (3) suggests that the only real difference, between diverse market timing rules coupled with various types of moving averages, lies in the weighting scheme $\{y_{t-1}, y_{t-2}, \dots, y_{t-k}\}$ used to compute the moving average of price changes. Even though there are various trading rules based on moving averages of prices and various types of moving averages, there are basically only three types of the shape of weighting scheme that are used in practice: equal weighting of price changes, underweighting the most old price changes, and underweighting both the most recent and the most old price changes. In order to generate the most typical shapes of the weighting function, we will employ four types of the weighting scheme $\{y_{t-1}, y_{t-2}, \dots, y_{t-k}\}$.

The equal weighting of price changes is used in the MOM rule and the Simple-Moving-Average-Change-of-Direction rule (ΔSMA) (Zakamulin (2015)). The value of the trading indicator in these rules is computed as:

$$\text{Indicator}_t^{\text{MOM}(k)} = \text{Indicator}_t^{\Delta\text{SMA}(k-1)} = \frac{1}{k} \sum_{i=1}^{k} \Delta P_{t-i}. \tag{4}$$

In the Price-minus-Simple-Moving-Average rule (P-SMA) and the Linear-Moving-Average-Change-of-Direction rule (ΔLMA), the value of the trading indicator is computed using the linearly weighted moving average of price changes (see Zakamulin (2015)):

$$\text{Indicator}_t^{\text{P-SMA}(k)} = \text{Indicator}_t^{\Delta\text{LMA}(k-1)} = \frac{\sum_{i=1}^{k}(k-i+1)\Delta P_{t-i}}{\sum_{i=1}^{k}(k-i+1)}. \tag{5}$$

In the linearly weighted moving average the weights decrease in arithmetic progression. In particular, in this weighting scheme the latest price change has weight k, the second latest $k-1$, etc. down to one. A disadvantage of the linearly weighted moving average is that the weighting scheme is too rigid. In contrast, by varying the value of λ in the Price-minus-Reverse-Exponential-Moving-Average (P-REMA), one is able to adjust the weighting to give greater or lesser weight to the most recent price. The value of the trading indicator in this case is computed as (Zakamulin (2015)):

$$\text{Indicator}_t^{\text{P-REMA}(k)} = \frac{\sum_{i=1}^{k}\left(1-\lambda^{k-i+1}\right)\Delta P_{t-i}}{\sum_{i=1}^{k}\left(1-\lambda^{k-i+1}\right)}. \tag{6}$$

When $\lambda = 0$, this weighting scheme reduces to the simple moving average of price changes (as in the MOM rule). When $\lambda \to 1$, this weighting scheme reduces to the linear moving average of price changes (Zakamulin (2015)). Thus, by varying the value of λ, one is able to adjust the weighting to give greater or lesser weights to the most recent price changes.

The weighting scheme in the DCM has a hump-shaped form and underweights both the most recent and the most old price changes. Specifically, in this weighting scheme the largest weight is given to the $(s+1)$-th price change. Most often, to compute the value of the DCM, one uses the exponential moving average in both the short and long windows. The value of the trading indicator for the DCM in this case is computed as (Zakamulin (2015)):

$$\text{Indicator}_t^{\text{DCM}(s,k)} = \frac{\sum_{i=1}^{k}\left(\lambda^{i}-\lambda^{k+1}\right)\Delta P_{t-i}}{1-\lambda^{k+1}} - \frac{\sum_{i=1}^{s}\left(\lambda^{i}-\lambda^{s+1}\right)\Delta P_{t-i}}{1-\lambda^{s+1}}.$$

The value of the technical indicator for the P-REMA rule depends on the length of the lookback period k and the value of the decay factor λ which determines the degree of over-weighting the most recent price changes. In order to avoid over-optimization in out-of-sample tests, we perform the optimization with respect to k only; the value of λ is held constant through time. In this rule we use $\lambda = 0.8$ which provides the degree of overweighting the most recent price changes somewhere in between the degrees provided by the equally weighted and linearly weighted schemes. The value of the technical indicator for the DCM depends on the lengths of shorter and longer averages. Among traders, one of the most popular combination is to use 50-day and 200-day averages in this timing rule. Therefore we fix the length of the shorter average to be $s = 2$ (months), whereas the length of the longer average $k > 2$ is determined by the dynamic optimization procedure.[3] In this rule we also use $\lambda = 0.8$ in both shorter and longer exponential moving averages. Figure 1 illustrates the four weighting schemes used for the computations of technical trading indicators in our study.

[Insert Figure 1 about here]

3 Data

3.1 Data Sources and Data Construction

In our empirical study we use the capital appreciation and total returns (denoted by CAP and MKT respectively) on the Standard and Poor's Composite stock price index, as well as the risk-free rate of return (denoted by RF) proxied by the Treasury Bill rate. Our sample period begins in January 1860 and ends in December 2014 (155 full years), giving a total of 1860 monthly observations. The data on the S&P Composite index comes from two sources. The returns for the period January 1860 to December 1925 are provided by William Schwert.[4] The returns for the period January 1926 to December 2014 are computed from the closing monthly priced of the S&P Composite index and corresponding dividend data provided by Amit Goyal.[5] The Treasury Bill rate for the period January 1920 to December 2014 is also provided by Amit Goyal. Because there was no risk-free short-term debt prior to the 1920s, we estimate it in the same manner as in Welch and Goyal (2008) using the monthly data for the Commercial Paper Rates for New York. These data are available for the period January 1857 to December 1971 from the National Bureau of Economic Research (NBER) Macrohistory database.[6] First, we

[3]Since our data comes at the monthly frequency, the choice of $s = 2$ is equivalent to using 42-days window in the shorter average.

[4]`http://schwert.ssb.rochester.edu/data.htm`

[5]`http://www.hec.unil.ch/agoyal/`

[6]`http://research.stlouisfed.org/fred2/series/M13002US35620M156NNBR`

run a regression

$$\text{Treasury-bill rate}_t = \alpha + \beta \times \text{Commercial Paper Rate}_t + e_t$$

over the period from January 1920 to December 1971. The estimated regression coefficients are $\alpha = -0.00039$ and $\beta = 0.9156$; the goodness of fit, as measured by the regression R-square, amounts to 95.7%. Then the values of the Treasury Bill rate over the period January 1860 to December 1919 are obtained using the regression above with the estimated coefficients for the period 1920 to 1971.

3.2 Descriptive Statistics and Tests

Table 1 summarizes the descriptive statistics for the data used in our study. Since the main goal of our empirical study is to estimate the out-of-sample performance of a few distinct technical trading rules and investigate how robust their performance is in sub-samples of data, the descriptive statistics are reported for the total out-of-sample period from January 1870 to December 2014 as well as for the first and second halves of the total out-of-sample period (from January 1870 to December 1942 and from January 1942 to December 2014 respectively). The total out-of-sample period spans 145 years, the first and the second halves span 72 and 73 years respectively. The split point between the halves is chosen to have the same numbers of Bull and Bear markets in each half (see the subsequent section).

The results of the Shapiro-Wilk test reject the normality in all data series over the total period as well as over each subperiod. It is worth noting that the first half of the out-of-sample period was much more turbulent than the second one. In particular, the mean stock returns during the first half were substantially lower than those during the second half, whereas the volatility, as well as the kurtosis, were considerably higher. In addition, during the first half of the out-of-sample period the stock return series exhibited a statistically significant positive autocorrelation. Over the second half, on the other hand, the autocorrelation in stock returns was neither economically nor statistically significant.

[Insert Table 1 about here]

To find out whether the means and standard deviations of the return series are the same in both sub-periods, we test the following null-hypotheses:

Equality of mean returns: $H_0^1 : \mu^1_{CAP} = \mu^2_{CAP}, \quad H_0^2 : \mu^1_{MKT} = \mu^2_{MKT}, \quad H_0^3 : \mu^1_{RF} = \mu^2_{RF},$

Equality of standard deviations: $H_0^4 : \sigma^1_{CAP} = \sigma^2_{CAP}, \quad H_0^5 : \sigma^1_{MKT} = \sigma^2_{MKT}, \quad H_0^6 : \sigma^1_{RF} = \sigma^2_{RF},$

where, for example, μ^1_{CAP} and μ^2_{CAP} denote the mean capital appreciation return during the first and the second sub-period respectively, and σ^1_{CAP} and σ^2_{CAP} denote the standard deviation of the capital appreciation return during the first and the second sub-period respectively. To test the hypotheses 1-3, we perform a standard two-sample t-test for equal means. To test the hypotheses 4-6, we perform a standard two-sample F-test for equal variances.

[Insert Table 2 about here]

Table 2 reports the results of the hypothesis tests. These results suggest that we have strong statistical evidence that the standard deviations of all return series have changed over time. In addition, we have evidence that the mean capital appreciation return and the mean risk-free rate of return have changed over time (at the 6% and 1% significance levels respectively). Yet, we cannot reject the hypothesis that the mean total market return has been stable over time.

3.3 A Structural Break Analysis

The results of the previous sub-section advocate that there are economically and statistically significant differences in the mean capital appreciation return across the two sub-samples of data. In this sub-section we perform a structural break analysis. Our goal is twofold. The first goal is to verify that there is a major break in the growth rate of the stock price index. The second goal is to find the date of the breakpoint.

Our null hypothesis is that the period t log return on the (not adjusted) stock price index, r_t, is normally distributed with constant mean μ and variance σ^2. More formally, $r_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Under this hypothesis the log of the stock price index at time t is given by the following linear model

$$\log\left(I_t\right) = \log\left(I_0\right) + \mu\, t + \varepsilon_t, \tag{7}$$

where I_0 is the index value at time 0 and $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2 t\right)$. Our alternative hypothesis is that the mean log return on the stock index varies over time. To test the null hypothesis, there are many formal tests (see Zeileis, Kleiber, Krämer, and Hornik (2003) and references therein). Unfortunately, the error term in regression (7) does not satisfy the standard i.i.d. assumptions (because ε_t exhibits heteroskedasticity and autocorrelation) and therefore these tests are not applicable in our case.

Our simplified alternative hypothesis is that the mean log return at time t^* changes from μ to $\mu + \delta$. Under the alternative hypothesis the log of the stock price index at time t is given by the following segmented model

$$\log(I_t) = \log(I_0) + \mu\, t + \delta\,(t - t^*)^+ + \varepsilon_t, \tag{8}$$

where $(t - t^*)^+$ denotes the positive part of the difference $(t - t^*)$. In this case the natural test of the null hypothesis is

$$H_0 : \delta = 0.$$

We find the breakpoint t^* using the methodology presented in Muggeo (2003). Both the alternative models are estimated using the total sample period 1860-2014. The results of the estimation of the two alternative models are reported in Table 3. The p-values of the estimated coefficients are computed using the heteroskedasticity and autocorrelation consistent standard errors. Figure 2 plots the log of the stock price index versus the fitted segmented model.

[Insert Table 3 about here]

Apparently, the we can reject the null hypothesis of constant log mean return at less than the 1% significance level. The segmented model has a higher R-squared (98% versus 90% for the linear model) and double as low the residual standard deviation (27% versus 62% for the linear model). The estimated date of the breakpoint is May 1944; this date roughly coincides with the split point between our two sub-samples of data. The 95% confidence interval for the breakpoint date is from May 1943 to May 1945. Under the assumption of a constant mean log returns, over the total sample period (that spans 155 years) the estimated mean log return amounts to approximately 4% in annualized terms. However, this assumption proofs to be wrong and a more detailed examination of the growth rate of the log of the stock price index suggests that around year 1944 (84 years from the start of the sample) there was a major break in the growth rate. Specifically, prior to 1944 the estimated mean log return was about 2%, thereafter about 7% in annualized terms.

[Insert Figure 2 about here]

4 Bull and Bear Market Cycles

4.1 The Dating Algorithm and the Results

It is an old tradition to describe cycles in stock prices as Bull and Bear markets. Yet, since there is no generally accepted formal definition of Bull and Bear markets in the finance literature, there is no single preferred method to identify the state of the stock market. Specifically, the literature offers two fundamentally different types of methods to detect the turning points between the phases of the stock market: non-parametric and parametric methods. The non-parametric methods are based on rules whereas the parametric ones are based on models. Among the rule-based methods, the most popular ones[7] adopt, with slight modifications, the formal dating methods used to identify turning points in the business cycle (Bry and Boschan (1971)). This type of a particular dating algorithm is based on a complex set of rules. In contrast, Lunde and Timmermann (2004) propose a very simple dating rule based on imposing a minimum on the price change since the last peak or trough. Among the parametric methods, the most prevailing one[8] is based on using a regime switching model (pioneered by Hamilton (1989)) with two or more laten states. However, a significant drawback of this approach is that a detected turning point usually does not coincide with a historical peak or trough in stock prices.

In our study, to detect the turning points between the Bull and Bear markets, we employ the dating algorithm proposed by Pagan and Sossounov (2003). This dating algorithm consists of two main steps: determination of initial turning points in raw data and censoring operations. In order to determine the initial turning points, first of all one uses a window of length $\tau_{\text{window}} = 8$ months on either side of the date and identifies a peak (trough) as a point higher (lower)

[7] Examples are Pagan and Sossounov (2003) and Gonzalez, Powell, Shi, and Wilson (2005).
[8] Examples are Maheu and McCurdy (2000) and Maheu, McCurdy, and Song (2009).

than other points in the window. Second, one enforces the alternation of turning points by selecting highest of multiple peaks and lowest of multiple troughs. Censoring operations require: eliminating peaks and troughs in the first and last $\tau_{\text{censor}} = 6$ months; eliminating cycles that last less than $\tau_{\text{cycle}} = 16$ months; and eliminating the phases that last less than $\tau_{\text{phase}} = 4$ months (unless the absolute price change in a month exceeds $\theta = 20\%$).[9]

[Insert Figure 3 about here]

Panels A and B in Figure 3 plot the natural log of the monthly Standard and Poor's Composite stock price index over the two sub-periods: 1870-1942 and 1942-2014. Shaded areas in the figure indicate the Bear market phases. Table 4 reports the descriptive statistics of Bull and Bear markets for the whole period and the two sub-periods. Over the total period, there were 41 Bull markets and 40 Bear markets. Each of the two sub-periods contains 21 Bull markets and 20 Bear markets.

[Insert Table 4 about here]

Over the whole period, the average length of a Bull market is close to 29 months, whereas the average Bear market length is close to 15 months. It is clear that Bull markets tend to be longer than the Bear markets and the durations of phases agree quite closely with those reported by Pagan and Sossounov (2003) and Gonzalez et al. (2005). The average Bull market duration exceeds the average Bear market duration by a factor of 1.9. The comparison of the lengths of the two stock market phases in the first and the second sub-periods suggests that over time the Bull markets tend to be longer while the Bear markets tend to be shorter. Whereas for the first sub-period the ratio of the average Bull market length to the average Bear market length amounts to 1.4, for the second sub-period this ratio amounts to 2.7. In other words, this ratio has almost doubled over time. On average, the stock index price increases by 65% during a Bull market and decreases by 24% during a Bear market. Our results suggest that over time the average amplitude of Bull markets tends to increase whereas the average amplitude of Bear markets tends to decrease.

[9] Gonzalez et al. (2005) use the same algorithm with $\tau_{\text{window}} = 6$, $\tau_{\text{cycle}} = 15$, and $\tau_{\text{phase}} = 5$. Despite the differences, the Bull and Bear markets in the study by Gonzalez et al. (2005) largely coincide with the Bull and Bear markets in the study by Pagan and Sossounov (2003).

All the Bull markets exhibit positive mean return while all the Bear markets have negative mean return. We do not observe any economically significant time-variation in the value of mean returns during the two stock market phases across subsample periods. In addition, somewhat surprisingly,[10] the return volatility during Bull and Bear markets is economically insignificantly different (a similar result is reported in Gonzalez et al. (2005)). This finding implies that Bull markets differ from Bear markets mainly in terms of mean returns, not in terms of standard deviation of returns.

4.2 Testing for a Structural Break

Our results, together with those obtained previously by Pagan and Sossounov (2003) and Gonzalez et al. (2005), advocate that the properties of cycles in stock prices has significantly changed over time. Yet so far we do not have any scientific evidence of the presence of structural breaks in the parameters and dynamics of Bull-Bear cycles. Since the presence of structural breaks is of crucial importance for the ability of a market timing strategy to outperform the passive market strategy, in the rest of this section we analyze whether there are statistically significant changes in the distribution parameters of Bull and Bear markets over time. Similarly to Maheu and McCurdy (2000), we use a two-state first-order Markov switching model for monthly returns as a parametric model that describes the dynamics of the stock market.

Mover formally, consider a two-state Markov switching model for returns where S_t denotes the latent state variable at time t. The state variable can take one of two possible values: 0 (denotes the Bear market state) and 1 (denotes the Bull market state). This Markov switching model for returns in sub-period $m \in \{1, 2\}$ can be written as

$$\begin{aligned} r_t^m | S_t &\sim N\left(\mu_{S_t}^m, (\sigma_{S_t}^m)^2\right), \\ p_{ij}^m &= P^m(S_t = j | S_{t-1} = i), \end{aligned}$$

where $i, j \in \{0, 1\}$. This model assumes that the stock market returns at time t of sub-period m are normally distributed with mean μ_0^m and standard deviation σ_0^m if the market is in state 0. Otherwise, in state 1, the stock market returns are normally distributed with mean μ_1^m and standard deviation σ_1^m. p_{ij}^m is the probability of transition from state i to state j in sub-period

[10]It is customary to assume that a Bear market is the low-return high-volatility state, whereas a Bull market is the high-return low-volatility state.

m. The transition probability matrix in sub-period m is given by

$$P^m = \begin{bmatrix} p_{00}^m & p_{01}^m \\ p_{10}^m & p_{11}^m \end{bmatrix}.$$

To find out whether the parameters of the Bull and Bear markets are the same in both sub-periods, we test the following null-hypotheses:

$$\text{Equality of mean returns: } H_0^1 : \mu_0^1 = \mu_0^2, \quad H_0^2 : \mu_1^1 = \mu_1^2,$$

$$\text{Equality of standard deviations: } H_0^3 : \sigma_0^1 = \sigma_0^2, \quad H_0^4 : \sigma_1^1 = \sigma_1^2,$$

$$\text{Equality of probability transition matrices: } H_0^5 : P^1 = P^2.$$

To test hypotheses 1-2, we perform a standard two-sample t-test for equal means. To test hypotheses 3-4, we perform a standard two-sample F-test for equal variances.

We test the equalities of the two transition probability matrices by performing element-by-element tests of the stability of each entry p_{ij}^m. To estimate the transition probability p_{ij}^m and standard errors of estimation of p_{ij}^m we use a bootstrap estimation approach proposed by Kulperger and Rao (1989). The bootstrap approach follows these steps: First, using the original data sequence of Bull and Bear markets, we estimate the transition probability matrix by employing the maximum likelihood estimator. Second, we generate 100 bootstrap samples of the data sequences following the conditional distributions of states estimated from the original one. Third, we apply maximum likelihood estimation on each bootstrapped data sequence. Forth, the estimated transition probability is computed as the average of all maximum likelihood estimators. Finally, after computing the average, we compute the standard deviation of our estimator and corresponding standard error of estimation. The hypothesis $H_0^{5_q} : p_{ij}^1 = p_{ij}^2$, $q \in \{1, 2, 3, 4\}$, is tested assuming that errors are normally distributed.

[Insert Table 5 about here]

Table 5 reports the estimated transition probabilities of the Markov switching model for the stock market states over two the historical sub-periods. The comparison of the values of transition probabilities over the two historical sub-periods also advocates that the duration of the Bear (Bull) markets has decreased (increased) over time. Specifically, $p_{01}^1 = 0.057$ whereas $p_{01}^2 = 0.096$. This says that during the first sub-period the transition probability form Bear to Bull market has been 5.7%, whereas over the second sub-period the transition probability form Bear to Bull market has been 9.6%. That is, the transition probability from the Bear state to the Bull state has substantially increased over time (almost doubled). As a consequence, the average length of Bear markets has become shorter over time. Similarly, $p_{10}^1 = 0.040$ whereas $p_{10}^2 = 0.031$. This says that during the first sub-period the transition probability form Bull to Bear market has been 4.0%, whereas over the second sub-period the transition probability form Bull to Bear market has been 3.1%. As a result, the average duration of Bull markets has become longer over time.

Table 6 reports the results of the hypothesis tests. These results suggest that we have strong statistical evidence that all the transition probabilities between the states of the stock market have changed over time (that is, we can reject the equality of the transition probability matrices over the two sub-periods) and that the volatility of the states have changed over time as well. Yet, we cannot reject the hypothesis that the mean stock market returns during the states have been stable over time. Anyway, the results of our statistical tests have important implications for the performance of the market timing strategies. There is clear scientific evidence that the duration of the Bull markets has increased over time, whereas the duration of the Bear markets has decreased over time. As a result, as compared with the first sub-period, over the second sub-period the stock market has been much more often in the Bull state than in the Bear state. Since the superior performance of a market timing strategy can appear only as the result of switching to cash during the Bear market states, it is logical to deduce that we should observe a substantial deterioration of the performance of market timing strategies (relative to that of the market) over the second sub-period.

[Insert Table 6 about here]

5 Empirical Research Design

5.1 Returns to the Market Timing Strategy

In our empirical study the main goal is to measure the out-of-sample performance of four clearly distinct market timing rules based on moving averages. The computation of the trading indicators for these rules is described in Section 2. The weighting schemes for these rules are illustrated in Figure 1. In each market timing rule the generation of a trading signal is a two-step process. At the first step, one computes the value of a technical trading indicator using the last closing price and k lagged prices

$$\text{Indicator}_t^{TR(k)} = Eq(P_t, P_{t-1}, \ldots, P_{t-k}),$$

where TR denotes the timing rule and $Eq(\cdot)$ is the equation that specifies how the technical trading indicator is computed. At the second step, using a specific function one translates the value of the technical indicator into the trading signal. In all market timing rules considered in this paper the Buy signal is generated when the value of a technical trading indicator is positive. Otherwise, the Sell signal is generated. Thus, the generation of a trading signal can be interpreted as an application of the following (mathematical) *indicator function* to the value of the technical indicator

$$\delta_{t+1|t} = \mathbf{1}_+ \left(\text{Indicator}_t^{TR(k)} \right),$$

where the indicator function $\mathbf{1}_+(\cdot)$ is defined by

$$\mathbf{1}_+(x) = \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases}$$

and $\delta_{t+1|t} \in \{0, 1\}$ is a trading signal for month $t+1$ (0 means Sell and 1 means Buy) generated at the end of month t.

In order to assess the real-life performance of a market timing rule, we need to account for the fact that rebalancing an active portfolio incurs transaction costs. Transaction costs in capital markets consist of the following three primary components: half-size of the quoted bid-ask spread, brokerage fees (commissions), and market impact costs. In addition, there are various taxes, delay costs, opportunity costs, etc. (see, for example, Freyre-Sanders, Guobuzaite, and Byrne (2004)). In our study we consider the average bid-ask half-spread as the only determinant of the one-way transaction costs, and we neglect all other components of transaction costs. Berkowitz, Logue, and Noser (1988), Chan and Lakonishok (1993), and Knez and Ready (1996) estimate the average one-way transaction costs for institutional investors to be in the range of 0.23% to 0.25%. Therefore in our study, we assume that the one-way transaction costs in the stock market amount to 0.25%. Denoting by γ the one-way transaction costs, the return to the market timing strategy over month t is given by

$$r_t = \begin{cases} r_{Mt} & \text{if } (\delta_{t|t-1} = \text{Buy}) \text{ and } (\delta_{t-1|t-2} = \text{Buy}), \\ r_{Mt} - \gamma & \text{if } (\delta_{t|t-1} = \text{Buy}) \text{ and } (\delta_{t-1|t-2} = \text{Sell}), \\ r_{ft} & \text{if } (\delta_{t|t-1} = \text{Sell}) \text{ and } (\delta_{t-1|t-2} = \text{Sell}), \\ r_{ft} - \gamma & \text{if } (\delta_{t|t-1} = \text{Sell}) \text{ and } (\delta_{t-1|t-2} = \text{Buy}), \end{cases} \tag{9}$$

where r_{Mt} and r_{ft} are the month t returns on the stock market (including dividends) and the risk-free asset respectively.

5.2 Out-of-Sample Testing of Trading Rules

To simulate the returns to the market timing strategy that are given by (9), for each market timing rule we need to compute the value of the technical indicator which provides us with Buy and Sell signals. It is crucial to observe that in order to compute the value of the technical indicator we need to specify the length of the lookback period k. One approach to the choice of k is to use the full historical data sample, simulate the returns to the market timing strategy for different k, and pick up the value of k which produces the best performance. Yet this approach is termed as "data-mining" and the performance of the best trading rule in a back test (that is, "in-sample" performance) generally severely over-estimates the real-life performance.

It is widely believed that the out-of-sample performance of a trading strategy provides a much more reliable estimate of it's real-life performance than the in-sample performance (see Sullivan et al. (1999), White (2000), and Aronson (2006)). The out-of-sample performance measurement method is based on simulating the real-life trading where a trader has to make a choice of what length of the lookback period k to use given the information about the past performances of the market timing strategy for different values of k. Specifically, the out-of-sample testing procedure begins with splitting the full historical data sample $[1, T]$ into the initial in-sample subset $[1, p]$ and out-of-sample subset $[p+1, T]$, where T is the last observation in the full sample and p denotes the splitting point. Then the best rule discovered in the mined data (in-sample) is evaluated on the out-of-sample data.

The out-of-sample performance can be evaluated with either a rolling- or expanding-window estimation scheme. A common belief among traders is that, regardless of the choice of historical period, the same specific value of k is optimal for using in a given technical indicator. For example, the majority of traders believe that in the P-SMA trading rule the optimal value of k equals to 10. If the length of the optimal lookback period is constant through time, then it is natural to use the expanding-window estimation scheme to determine the value of k. Out-of-sample simulation of a market timing strategy using an expanding-window estimation of k is performed as follows. The in-sample period of $[1, t]$, $t \in [p, T-1]$, is used to complete the procedure of selecting the best trading rule given some optimization criterion $O(r_1, r_2, \ldots, r_t)$ defined over the returns to the market timing strategy up to month t.[11] Formally, in our study the choice of the optimal k_t^* is given by

$$k_t^* = \arg \max_{k \in [k^{\min}, k^{\max}]} O(r_1, r_2, \ldots, r_t),$$

where $k^{\min}$ and $k^{\max}$ are the minimum and maximum values for k. Subsequently, the trading signal for month $t+1$ is determined using the lookback period of length k_t^*. Then the in-sample period is expanded by one month, and the best trading rule selection procedure is performed once again using the new in-sample period of $[1, t+1]$ to determine the trading signal for month $t+2$. This procedure is repeated, by pushing the endpoint of the in-sample period ahead by one month with each iteration of this process, until the trading signal for the last month T is determined.

The rolling-window estimation scheme is used when parameter instability is suspected. That is, when the length of the optimal lookback period varies through time. In the rolling-window estimation scheme the choice of k is done using the most recent n observations. In

[11] We follow closely the methodology employed by Lukac, Brorsen, and Irwin (1988), Lukac and Brorsen (1990), and Zakamulin (2014) among others. Note that this methodology has a dynamic aspect, in which the trading rule is being modified over time as the market evolves.

this case the choice of the optimal k_t^* is given by

$$k_t^* = \arg \max_{k \in [k^{\min}, k^{\max}]} O(r_{t-n+1}, r_{t-n+2}, \ldots, r_t).$$

We set the value of $k^{\min}$ to be the minimum possible length (measured in the number of lagged prices) of the lookback period for a given trading rule. For the Double Crossover Method $k^{\min} = 3$, for all other rules $k^{\min} = 1$. To select the appropriate value for $k^{\max}$, we studied the most popular recommendations of technical analysts for the choice of the optimal lookback period. In practice, the recommended value for the length k virtually never exceeds 12 months. To be on the safe side, in our empirical study we set $k^{\max} = 24$.

Despite many advantages of the out-of-sample performance measurement method, it has one unresolved deficiency that may seriously corrupt the estimation of the real-life performance of a market timing strategy. The primary concern is that no guidance exists on how to choose the split point between the in-sample and out-of-sample subsets. One possible approach is to choose the initial in-sample segment with a minimum length and use the remaining part of the sample for the out-of-sample test (see Marcellino, Stock, and Watson (2006) and Pesaran, Pick, and Timmermann (2011)). Another potential approach is to do the opposite and reserve a small fraction of the sample for the out-of-sample period (as in Sullivan et al. (1999)). Alternatively, the split point can be selected to lie somewhere in the middle of the sample. In any case, according to conventional wisdom, the out-of-sample performance of a trading strategy provides an unbiased estimate of its real-life performance.

Yet recently, the conventional wisdom about the unbiased nature of traditional out-of-sample testing has been challenged. In the context of out-of-sample forecast evaluation, Rossi and Inoue (2012) and Hansen and Timmermann (2013) report that the results of out-of-sample forecast tests depend significantly on how the sample split point is determined. Zakamulin (2014) also demonstrates that the out-of-sample performance of market timing strategies depends critically on the choice of a split point. The primary argument (put forward in the paper by Zakamulin (2014)), for why the choice of split point sometimes dramatically affects the out-of-sample performance of the market timing strategy, lies in the fact that the performance of market timing strategies is highly non-uniform. Generally, a market timing strategy under-performs the passive strategy during Bull markets and shows a superior performance during Bear markets. According to our results reported in Section 4, the mean Bull market duration exceeds the mean Bear market duration by a factor of 1.4-2.7, depending on the historical sub-period. As a result, one has to expect that over short-term horizons a market timing strategy under-performs the market most of the time, but occasionally it delivers an extraordinary outperformance. In addition, a Bull market might last for more than 6 years whereas a subsequent Bear market might last only a few months. As a result, a rather short out-of-sample period may contain basically only Bull markets and the results of out-of-sample testing may lead to an erroneous conclusion that market timing does not work at all (as in the tests performed by Sullivan et al. (1999)). Therefore, as argued in Zakamulin (2014), in the out-of-sample testing one has to choose the initial in-sample segment to have a minimum length. Another potential reason that may seriously distort the estimate for the real-life performance of market timing is the presence of structural breaks in the dynamics and parameters of Bull-Bear market cycles.

Motivated by the discussion above, we estimate the out-of-sample performance of market timing strategies over the total historical sample and two sub-samples, and choose the length of the initial in-sample period to be 10 years. Specifically, for the total sample that spans the period from January 1860 to December 2014, the first 10 years are reserved as the initial in-sample period, and, consequently, the total out-of-sample period is from January 1870 to December 2014 which spans 145 years. In addition, we measure the out-of-sample performance in the two sub-samples of data: January 1860-December 1942 and January 1932-December 2014. In each sub-sample, the first 10 years are again used as the initial in-sample period. As the result, the two out-of-sample sub-periods are: January 1870-December 1942 and January 1942-December 2014. Both of these sub-periods contain the same number of Bull and Bear phases and we know that the stock market dynamics is different in these sub-periods. Therefore we expect different performance of market timings strategies in these sub-periods.

5.3 Choice of Performance Measure

There is an uncertainty about what optimization criterion to use in the determination of the best trading rule using the past data. To limit the choice of optimization criteria, we consider an investor who decides whether to follow the passive buy-and-hold strategy or to follow the active market timing strategy. Since the two strategies are supposed to be mutually exclusive, it is natural to employ a reward-to-total-risk performance measure as the optimization criterion. That is, our investor chooses the value of k which maximizes some portfolio performance measure in a back test, that is, using the past (in-sample) data.

The most widely recognized reward-to-risk measure is the Sharpe ratio. Thus, the Sharpe ratio represents the natural optimization criterion to find the best trading rule. The Sharpe ratio uses the mean excess returns as a measure of reward, and the standard deviation of excess returns as a measure of risk. Specifically, the Sharpe ratio of trading strategy i with excess returns $r^e_{it} = r_{it} - r_{ft}$ is computed as (according to Sharpe (1994))

$$SR_i = \frac{\mu(r^e_i)}{\sigma(r^e_i)},$$

where $\mu(r^e_i)$ and $\sigma(r^e_i)$ denote the mean and standard deviation of r^e_{it} respectively.

For the Sharpe ratio of each market timing strategy we report the p-value of testing the null hypothesis that it is *equal to* the Sharpe ratio of the market portfolio (denoted by SR_M). For this purpose we apply the Jobson and Korkie (1981) test with the Memmel (2003) correction. Specifically, given SR_i, SR_M, and ρ as the estimated Sharpe ratios and correlation coefficient over a sample of size T, the test of the null hypothesis: $H_0 : SR_i = SR_M$ is obtained via the test statistic

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

which is asymptotically distributed as a standard normal.

As with any reward-to-risk ratio, the use of the Sharpe ratio has some inconveniences. In particular, its value is difficult to interpret, and to decide whether the timing strategy outperforms the market, one also needs to compute the Sharpe ratio of the market portfolio and compare one to the other. To facilitate performance measurement with the Sharpe ratio, we closely follow the method presented by Modigliani and Modigliani (1997) and employ in addition (to the Sharpe ratio) the M^2 measure (Modigliani-Modigliani measure or Modigliani-squared measure). The idea is to mix the active portfolio with a position in the risk-free asset so that the adjusted portfolio has the same (average) risk as the passive market. The returns to the adjusted portfolio are

$$r^*_{it} = a(r_{it} - r_{ft}) + r_{ft},$$

where a is the proportion invested in the active portfolio. When $a > 1$ ($a < 1$), it means that the adjusted portfolio represents a levered (unlevered) version of the original portfolio. The value of a that equates the risk of the adjusted portfolio with the risk of the market portfolio is

$$a = \frac{\sigma(r_M^e)}{\sigma(r_i^e)}.$$

In a similar manner to Bodie, Kane, and Marcus (2007), we compute the M^2 measure as the difference between the return to the adjusted portfolio and the return to the market portfolio. As the result, the expression for M^2 measure is given by

$$M_i^2 = \mu(r_i^{*e}) - \mu(r_M^e) = (SR_i - SR_M) \times \sigma(r_M^e).$$

Note that this M^2 measure produces the same ranking of risky portfolios as the Sharpe ratio, but it has the significant advantage of being in units of percent return, which makes it dramatically more intuitive to interpret. Specifically, this measure tells us by how much, in basis points, portfolio i outperformed (if $M_i^2 > 0$) or underperformed (if $M_i^2 < 0$) the market portfolio on a risk-adjusted basis.

Because the Sharpe ratio is often criticized on the grounds that the standard deviation appears to be an inadequate measure of risk, as a robustness test, we also used the Sortino ratio (due to Sortino and Price (1994)) and a few other popular reward-to-risk ratios as the optimization criterion instead of the Sharpe ratio. The results of these tests showed that regardless of the reward-to-risk ratio used, the comparative performance of the active market timing strategy and the passive market strategy remains virtually the same.

5.4 Measuring the Similarity Between the Bull-Bear Markets and Buy-Sell Trading Signals

How good are the market timing rules in detecting the Bull and Bear stock market phases? To find this out, we measure the similarity between the Bull-Bear market states (given by S_t) and Buy-Sell trading signals (given by δ_t) using the Simple Matching Coefficient (SMC). In addition, using the Jaccard Similarity Coefficient (JSC, due to Jaccard (1901)), we measure the similarity between the Bull market states and the Buy trading signals and the similarity between the Bear market states and the Sell trading signals. The computation of these similarity coefficients is described below.

We remind the reader that both S_t and δ_t, $t \in [1, T]$, can either be 0 or 1. First of all, we compute the following quantities:

$$\begin{aligned}
M_{00} &= \text{the number of instances where } S_t = 0 \text{ and } \delta_t = 0, \\
M_{01} &= \text{the number of instances where } S_t = 0 \text{ and } \delta_t = 1, \\
M_{10} &= \text{the number of instances where } S_t = 1 \text{ and } \delta_t = 0, \\
M_{11} &= \text{the number of instances where } S_t = 1 \text{ and } \delta_t = 1.
\end{aligned}$$

In our study t denotes a month's number. Therefore M_{00} and M_{11} can be interpreted as the number of months with correct Sell and Buy signals respectively. In contrast, M_{01} and M_{10} can be interpreted as the number of months with false Buy and Sell signals respectively. For any t, each instance must fall into one of these four categories, meaning that

$$M_{00} + M_{01} + M_{10} + M_{11} = T.$$

The Simple Matching Coefficient is computed as the number of months with correct Buy and Sell trading signals divided by the total number of months

$$SMC = \frac{M_{00} + M_{11}}{M_{00} + M_{01} + M_{10} + M_{11}}.$$

The Jaccard Similarity Coefficient between the Bull market states and the Buy trading signals is computed as the number of months with correct Buy signals divided by the aggregate number of months with correct Buy signals and all incorrect trading signals

$$JSC(\text{Bull, Buy}) = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}.$$

The computation of the Jaccard Similarity Coefficient between the Bear market states and the Sell trading signals goes along the similar lines

$$JSC(\text{Bear, Sell}) = \frac{M_{00}}{M_{00} + M_{01} + M_{10}}.$$

The value of each similarity coefficient is constrained to lie within the range $[0, 1]$. For example,

$$0 \leq SMC \leq 1,$$

where the case $SMC = 1$ indicates a perfect match between the Bull-Bear markets states and the Buy-Sell trading signals. Therefore the closer a similarity coefficient to unity, the better a market timing rule identifies the stock market phases.

A passive buy-and-hold strategy can be considered as a strategy that generates $\delta_t = 1$ for all $t \in [1, T]$. Therefore for the passive strategy $M_{00} = 0$ and consequently

$$JSC(\text{Bull, Buy}) = SMC, \quad JSC(\text{Bear, Sell}) = 0.$$

6 Empirical Results

6.1 Time-Variations in the Length of the Optimal Lookback Period

As it was mentioned earlier, in the literature on market timing one usually supposes that there is some specific length of the lookback period, k, which is optimal for using in a given technical indicator. Yet there is a major controversy among technical analysts about the optimal value of k. For instance, for the P-SMA rule the recommended value of k varies from 10 to 200 days (see Brock et al. (1992), Sullivan et al. (1999), Okunev and White (2003), and Kirkpatrick and Dahlquist (2010), Chapter 14). This common belief, that the optimal lookback period is constant, justifies the use of the expanding-window estimation scheme in the out-of-sample simulation of the trading strategy. And that is why the rolling-window estimation scheme is practically never used. The goal of this section is to check whether this common belief is fallacious or not.

In order to find out how stable the length of the optimal lookback period is for each trading rule, we use a long rolling window of n months. For each specific window $[t, t+n]$, for $t \in [1, T-n]$, we find the value of k which maximizes the in-sample performance of a trading rule. Specifically, the optimal k_t^* is given by

$$k_t^* = \arg \max_{k \in [k^{\min}, k^{\max}]} SR(r_t, r_{t+1}, \ldots, r_{t+n}),$$

where $SR(\cdot)$ is the Sharpe ratio computed using the returns $r_t, r_{t+1}, \ldots, r_{t+n}$ to a trading strategy under investigation. Then we report the descriptive statistics of k_t^* for each trading rule used in our empirical study.

We need to choose a suitable period length, n, that covers a series of alternating Bull and Bear markets. Our choice is $n = 120$ (10 years) and is motivated by the results reported in Section 4. In particular, in that section we studied the durations of Bull and Bear markets in our sample and found that the mean durations of the Bull and Bear markets are 29.3 and 15.5 months respectively, with the longest Bull and Bear market durations of about 6.5 and 3.5 years respectively. Therefore with the lookback period of 10 years we are guaranteed to cover at least one stock market cycle.

[Insert Figure 4 about here]

The results of our investigation are visualized in Figure 4 and the descriptive statistics of the optimal lookback period (which is the number of lagged monthly prices) for different technical trading rules are reported in Table 7. Specifically, Table 7 reports the descriptive statistics of the lengths of the optimal lookback periods for our total historical sample and the two sub-samples. We remind the reader that our total sample is from January 1860 to December 2014 and covers the period of 155 years (1860 months). With a 10-year window, it includes 1741 different values for the optimal k_t^*, where the first value is for the period from January 1860 to December 1869, the second value is for the period from February 1860 to January 1870, etc.

[Insert Table 7 about here]

Apparently, the results suggest that for each technical trading rule there is no single optimal lookback period. On the contrary, the results indicate that there are substantial time-variations in the length of the optimal lookback period. For example, for the most popular P-SMA rule the optimal lookback period varies from 1 to 23 months. Nevertheless, over the total historical sample, the mean value of the optimal lookback period for this rule amounts to 9.7 months which is very close to the most often recommended value of 10 months (200 days). In contrast, in our study the mean value of the optimal lookback period for the MOM rule is 8.1 months which is substantially lower than the most often recommended value of 12 months (Moskowitz et al. (2012)). The comparison of the mean lengths of the optimal lookback periods for the first sub-sample versus those for the second sub-sample reveals that for the majority of trading rules the mean length of the optimal lookback period has decreased notably. This result probably comes as no surprise because in Section 4 we found evidence of the presence of a structural break in the parameters and dynamics of Bull-Bear cycles. Specifically, we found that the stock market has been much less volatile in the second half of our sample, with longer Bull markets and shorter Bear markets as compared with those in the first half. We conjecture that the length of the optimal lookback period depends on the volatility of the stock market: when the volatility is low (high) it is optimal to decrease (increase) the length of the lookback period used to compute the moving average.[12]

We perform a formal statistical test of the equality of the mean lengths of the optimal lookback periods over the two historical sub-samples. For this purpose, we assume that the optimal lookback period over the horizon of 10 years follows the auto-regressive process of order one:

$$k_t = c + \varphi\, k_{t-1} + \varepsilon_t.$$

The coefficients c and φ of this process can be estimated using OLS. The mean value and variance of k_t can be computed as

$$E[k_t] = \frac{c}{1-\varphi}, \quad Var[k_t] = \frac{Var[\varepsilon_t]}{1-\varphi^2}.$$

After the computation of the mean values and variances of k_t over the two historical sub-periods, we perform a standard t-test of the null-hypothesis $H_0 : E[k_t^1] = E[k_t^2]$, where $E[k_t^1]$ and $E[k_t^2]$ are the mean values of the optimal lookback period over the first and the second sub-sample respectively. The results of this test are reported in Table 8 and allows us to reject the null hypothesis for 3 out of 4 trading rules. Specifically, only for the DCM rule we cannot reject the hypothesis on the stability of the mean length of the optimal lookback period.

[Insert Table 8 about here]

[12] A similar idea is presented by Kaufman (1995) who proposed using an Adaptive Moving Average in trading rules. This adaptive moving average automatically adjusts the length of the lookback period to the changes in stock market volatility.

The results reported in this section have two important implications. First of all, these results challenge the common belief on the constancy of the length of the optimal lookback period. As an immediate consequence, these results advocate that the trading strategy, simulated with the rolling-window estimation of the optimal lookback period, might produce better performance than that with the expanding-window estimation. Secondly, these results propose a simple explanation for the existing big diversity of the popular recommendations concerning the choice of the optimal lookback period. Specifically, different recommendations for the length of the lookback period might appear as the results of finding the best trading rule in the back-test using different historical periods.

6.2 Results of Performance Measurement in Out-of-Sample Tests

We perform the simulation of the returns to the market timing rules using both expanding- and rolling-window estimation schemes to dynamically determine the lookback period length that gives the best performance in a back test. In the rolling-window estimation scheme the length of the window is chosen to be 10 years. However, it is worth noting that, in principle, the performance of the market timing rule implemented with a rolling-window estimation scheme depends on the length of the rolling window. As a matter of fact, we tested different lengths of the rolling window (in the interval $n \in [2, 20]$ years) and our experiments showed that the performance of a market timing strategy varies insignificantly as long as the length of the rolling window exceeds 5 years. That is, our experiments indicated that decreasing the length of a rolling window to a period shorter than 5 years usually substantially deteriorates the performance of a market timing strategy.[13]

6.2.1 Performance over Bull and Bear Markets

We start presenting the results on the performance of trading strategies over Bull and Bear markets separately. For each trading rule we simulate real-life technical trading using both the rolling- and expanding-window estimation schemes and compute the descriptive statistics and performances of the passive market strategy and each active trading strategy over Bull and Bear markets. Tables 9 and 10 report the descriptive statistics and performances. Specifically,

[13]It is easy to understand why the length of the rolling window cannot be less than 4-5 years. In case the length of the rolling window is rather short, it may contain only a single stock market phase (Bull or Bear). As a result, the trading rule cannot be optimized to detect the changes between the phases of the stock market.

these tables report the mean returns, standard deviations, and Sharpe ratios of the passive strategy and each active trading strategy over Bull and Bear markets. The results are reported for the whole sample period and the two sub-periods. In addition, in these tables we report the p-values of testing the following three hypotheses $H_0 : \mu_i = \mu_M$, $H_0 : \sigma_i = \sigma_M$, and $H_0 : SR_i = SR_M$ where μ_i, σ_i, and SR_i are the mean returns, volatility, and the Sharpe ratio of strategy i; μ_M, σ_M, and SR_M are the mean returns, volatility, and the Sharpe ratio of the passive market portfolio.

[Insert Table 9 about here]

[Insert Table 10 about here]

The results presented in these two tables can be summarised as follows. Regardless of the choice of the estimation scheme and historical sub-period, over either Bull or Bear markets the mean returns and standard deviations of virtually all market timing strategies are statistically significantly different from those of the passive market strategy. Specifically, over both Bull and Bear markets the standard deviation of returns of market timing strategies is less than the standard deviation of returns of the market portfolio. Over the whole sample period, as compared with the volatility of the market portfolio, the average volatility of the market timing strategies is reduced by 27% (44%) over the Bull (Bear) markets. Over the Bull (Bear) markets the mean returns of market timing strategies are below (above) the mean returns of the market portfolio. Even though the Sharpe ratios of market timing strategies are virtually never statistically significantly different from those of the passive market strategy, our results suggest that, on average, the market timing strategies underperform the passive strategy in Bull states and outperform the passive strategy in Bear states of the market. It is worth noting that in Bear states the mean returns of market timing strategies are negative. That is, on average, technical traders also lose money in Bear markets; yet their losses are less than those of the market portfolio.

6.2.2 Similarity between the Bull-Bear states and Buy-Sell trading signals

In order to explain the results presented in Tables 9 and 10 and gain additional insights into the nature of market timing strategies, Tables 11 and 12 report the number of trading signals generated by each active trading rule and different similarity coefficients. The results on the number of generated trading signals show that this number is from 2 to 3 times higher than the number of stock market phases. For example, over the total sample period there were 41 Bull markets. Over the same period, the popular P-SMA rule generated 123 (98) Buy signals when implemented with the rolling (expanding) window estimation scheme. Apparently, market timing strategies generate many false signals. In addition, these results suggest that a trading rule implemented with a rolling-window estimation scheme generates more trading signals than the same trading rule implemented with an expanding-window estimation scheme. Regardless of the choice of estimation scheme, the number of trading signals is highest for the MOM rule and lowest for the DCM rule.

[Insert Table 11 about here]

[Insert Table 12 about here]

When it comes to the similarity coefficients, over the whole sample period the average (among all trading rules) JSC(Bull,Buy) amounts to 64% and the average JSC(Bear,Sell) is 42%. These numbers suggest that the market timing rules identify more accurately the Bull market states than the Bear market states. Since these numbers are substantially below 100%, we can again conclude that the trading rules generate many false signals. During Bull market states, false Sell signals deteriorate the performance of market timing strategies relative to the performance of the passive market portfolio. In contrast, during Bear market states, correct Sell signals improve the performance of market timing strategies relative to the performance of the passive market portfolio. The negative mean returns to the market timing strategies during Bear market states can be explained by the fact that the average JSC(Bear,Sell) is rather low. As compared with the first sub-period, during the second sub-period the average JSC(Bull,Buy) is higher whereas the average JSC(Bear,Sell) is lower. This observation suggests that the accuracy of detecting the stock market state is proportional to the average state duration. Specifically, the longer the average duration of a stock market state, the better the accuracy of detecting the state.

The results on the simple matching coefficients, reported in Table 12, suggest that the overall accuracy of trading rules in detecting the stock market states varies insignificantly between different rules and is only marginally better than the overall accuracy of the passive market strategy (assuming that the latter always generates a Buy trading signal). Over the whole sample period, the accuracy of the passive market strategy was 74%,[14] whereas the percentage of correct trading signals generated by different trading rules was only from 1% to 3% higher than that of the passive market strategy. As compared with the first sub-period, during the second sub-period the average percentage of correct trading signals generated by trading rules is higher, whereas the improvement in the accuracy, relative to that of the passive market strategy, is lower.

6.2.3 Overall Performance

In this subsection we present the results on the overall performance of trading strategies, that is, the performance over Bull and Bear markets simultaneously. Table 13 reports the descriptive statistics and performances of the active trading rules and the passive market portfolio as well. Specifically, this table reports the means, standard deviations, skewness, minimum, and maximum of monthly returns. In addition, this table reports the Sharpe ratio of each strategy. For each market timing strategy we test the hypothesis that its Sharpe ratio is equal to the Sharpe ratio of the passive market strategy. Specifically, we test $H_0 : SR_i = SR_M$. The descriptive statistics is reported for the whole out-of-sample period and for the two sub-periods.

[Insert Table 13 about here]

Judging by the Sharpe ratios, every market timing strategy outperforms the passive market strategy on the risk-adjusted basis. This observation applies equally to the performances over the whole period and the two sub-periods. Over the whole period, only two trading strategies exhibit performances that are statistically significantly different (at the 5% level) from that of

[14]This number also tells us that over the total sample period the market was in Bull state 74% of time.

the passive market portfolio. These strategies are based on the MOM rule and the P-REMA rule. Both the trading rules outperform the market when they are simulated with the rolling-window estimation scheme. Over the first sub-period, only the performance of the MOM rule is statistically significantly different from that of the passive market strategy. Even though the Sharpe ratios of the other rules are from 25% to 70% higher than that of the passive market strategy, we cannot reject the hypotheses (at conventional statistical levels) that they are equal to the Sharpe ratio of the market portfolio. Despite the fact that the two sub-periods have the same number of Bull and Bear markets, in the second sub-period the stock market has been much more often in the Bull state. Therefore as expected, over the second sub-period the market timing strategies outperformed the passive strategy to a much lesser extent. Specifically, over this sub-period the performance of market timing strategies is neither economically nor statistically significantly different from that of the passive market strategy. Here the Sharpe ratios of all the timing rules are only from 7% to 14% higher than that of the passive market strategy.

Our results on the performance of trading rules over Bull and Bear markets indicated that the market timing strategies underperform the market in Bull states and outperform the market in Bear states. Since over a sequence of interchanging Bull and Bear markets the active trading strategies tend to outperform the passive market portfolio, our results suggest that on average the out-performance in Bear states is greater than under-performance in Bull states.

We remind the reader that our trading rules are chosen to have four distinct shapes of the weighting scheme used for the computations of moving averages. Specifically, the MOM rule employs equal weighting of price changes, whereas both the P-REMA rule and P-SMA rule overweight the most recent price changes. The P-SMA rule employs the linear weighting of price changes and overweights the most recent price changes to a greater degree than the P-REMA rule. The weighting scheme in the DCM rule underweights both the most recent and the most old price changes. The comparison of the performances of different trading rules suggests that, contrary to the common belief, neither over-weighting nor under-weighting the recent price changes improves the performance of a market timing strategy based on moving averages. In particular, over the second sub-period, all trading rules generated virtually the same performance. Over the first sub-period, we find that the MOM rule produced the best performance in out-of-sample tests.

Our results on the length of the optimal lookback period revealed that it varies substantially over time and motivated us to consider the rolling-window estimation scheme as a potentially better alternative to the commonly used expanding-window estimation scheme. However, the comparison of the performances of trading rules implemented with two alternative estimation schemes suggests that the choice of the estimation scheme has only marginal impact on the performance. Only for the MOM rule over the first sub-period there is an economically significant difference in the Sharpe ratios of the trading strategies implemented with rolling- and expanding window estimation schemes. Yet even this difference is not statistically significant.[15] Therefore our results do not allow us to conclude which estimation scheme should be preferred in practice.

The comparison of the descriptive statistics of the returns to market timing strategies reveals the following. All market timing strategies are virtually equally risky (judging by the values of the standard deviation of monthly returns). We observe a significant risk reduction as compared to the riskiness of the passive market portfolio. However, the reduction of risk is not surprising because virtually in any market timing strategy about 1/3 of the time the money are held in cash. The mean returns to market timing strategies are also below the mean returns to the market portfolio. Yet for the majority of timing rules the decrease in mean (excess) returns is lesser than the decrease in risk. This property improves the risk-adjusted performance of a market timing strategy as compared with that of the passive market portfolio. The comparison of the maximum and minimum monthly returns to the market portfolio and those to the market timing strategies reveals an interesting observation. Specifically, whereas the minimum monthly returns are virtually the same, the passive market portfolio has as a rule much higher maximum monthly returns. This observation suggests that the majority of market timing rules lets the "big downward mover" months pass through, but plainly misses the "big upward mover" months. This is because the market timing rules detect the change in the stock market phase with some delay (which on average amounts to 3-4 months). The existence of a delay in the identification of a stock market state has another important implication for the performance of market timing rules during Bull markets. In particular, the first 6 months of Bull markets exhibit significantly higher returns than do the remaining months of Bull markets (this is reported by Maheu and McCurdy (2000) and Gonzalez et al. (2005)). Therefore technical

[15] In order to save the space, the result of this test is not reported.

traders usually miss high returns generated at the beginning of Bull markets. In contrast, there is no significant difference in returns between the first 6 months and the remaining months of Bear markets (Gonzalez et al. (2005)).

6.2.4 Performance Over Short- to Medium-Term Horizons

In this subsection we provide an alternative presentation of the performance of market timing strategies. This alternative presentation is suggested by Zakamulin (2014) and motivated as follows. The traditional performance measurement uses a single number (for example, the value of a Sharpe ratio) for performance that is estimated usually over a very long out-of-sample period (which is beyond the investment horizon of most individual investors). Such a number is very misleading for investors with short- and medium-term investment horizons. This is because a single number for performance creates a wrong impression that performance is time-invariant, whereas in reality it varies dramatically over time. Not only the relative performance of market timing rules varies over the stock market phases, but also there is a big variation in the outperformance during the Bear stock markets. Therefore, it is impossible to provide an accurate picture of market timing performance without taking into account the time-varying nature of performance. With this fact in mind, instead of providing a single number for the performance of a market timing strategy over a very long historical period, we measure the performance over shorter 5-year disjoint periods, and then provide the descriptive statistics of the historical performance over these 5-year periods.

Because our primer interest is to find out whether a market timing strategy can beat the market, we always need to compare the performance of a market timing strategy with that of the market. To simplify the performance comparison in this case, we employ the Modigliani-Modigliani measure. To illustrate the fact that the market timing performance is very uneven over time, Figure 5 plots the annualized M^2 performance measure computed over disjoint intervals of 5 years. The first period is from January 1870 to December 1874, the second period is from January 1875 to December 1879, etc. Then we plot the value of M^2 measure versus the historical period. The plots in this figure clearly indicate that the superior outperformance of market timing was generated mainly over relatively few particular historical episodes. Specifically, they are the severe Bear markets of the decades of 1870s, 1900s, 1930s, 1970s, and finally 2000s. Many market timing rules consistently under-performed the market over the course of several decades. For example, the most popular P-SMA rule under-performed the market over the period from early 1930s to late 1960s, and then over the period from late 1970s to the beginning of 2000s.

[Insert Figure 5 about here]

The alternative presentation of the performance of the market timing rules is reported in Table 14. In particular, this table reports the descriptive statistics of the annualized M^2 performance measure over the investment horizon of 5 years. First of all, the table reports the mean value of M^2, which reflects the average performance of a market timing strategy over a 5-year horizon. In addition to the mean, the table reports the standard deviation, which reflects the variability of M^2, as well as the minimum and maximum values, which define the range of possible values for M^2. The table also reports the median value of M^2. The median divides the range of M^2 in the middle and has 50% of the data below it. Thus, the probability that the performance of a market timing strategy over a 5-year horizon will be below the median equals 50%. Finally, this table reports the probability that a market timing strategy outperforms the passive strategy over a 5-year investment horizon.

[Insert Table 14 about here]

First, we interpret the descriptive statistics for the performance of market timing strategies over the whole sample period. Over this period, the majority of timing strategies exhibit a positive mean value for M^2 over a 5-year horizon. In other words, the majority of market timing rules outperform the market on average. Yet only the MOM rule has a positive median value for M^2. This means that only the MOM rule outperforms the market more than 50% of time over a 5-year horizon. For all other trading rules the probability of outperformance is less than 50%. For all trading rules the median performance is below the mean performance. When the mean is larger than the median, the probability distribution of M^2 is positively skewed. This suggests that on average the out-performance is greater than under-performance. Such an observation is confirmed by comparing the minimum and the maximum values for M^2. Specifically, on average the maximum values are substantially larger than the absolute values of the minima. The variability of the performance of the MOM rule is the largest one. However, this is because the MOM rule showed an extraordinary good outperformance over the period January 1930 to December 1934. As a matter of fact, the value of the performance measure for this particular historical period, in statistical terms, should be considered as an outlier because this value is very distant from the other values for the performance measure.

The comparison of the descriptive statistics for the performance of market timing strategies over the two sub-period reveals that the performance has deteriorated over time. Specifically, the mean and median values of M^2 for all trading rules are substantially lower for the second sub-period as compared with the first sub-period. This result is in agreement with the result on the performance measurement presented in the preceding subsection. Whereas over the first sub-period the outperformance probability is equal or above 50% for the majority of market timing rules, over the second sub-period the outperformance probability is below 50% for all the rules. It is interesting to observe the following. Over the second sub-period the Sharpe ratios of all the trading rules are virtually equal (see Panel C in Table 13) which may suggest that the choice of the trading rule is irrelevant. However, the mean values of M^2 and probabilities of outperformance (see Panel C in Table 14) reveal that according to these statistics the P-REMA rule should be preferred by medium-term investors. In particular, the P-REMA rule delivers notable higher values of M^2 and outperformance probability than the rest of the rules.

7 Discussion

In our study we utilized the longest historical dataset and extended previous studies on the performance of moving average trading rules in a number of ways. Yet while long history provides us with rich information about the past performance of market timing rules, the availability of long-term data is both a blessing and a curse. This is because in order to use the observed performance over a very long-term as a reliable estimate of the expected performance in the future, we need to make sure that the stock market dynamics both in the distant and near past were the same. However, the results from our numerous robustness tests revealed evidence of regime shifts in the stock market dynamics. All but one robustness tests were performed by splitting the whole very long-term sample into two virtually equal long-term sub-samples (with the same number of Bull and Bear market phases in each sub-sample) and comparing the results from the two sub-samples. Under the assumption of no regime shifts, we expect to obtain similar results in both sub-samples. Yet we found that the results for the first sub-sample were in sharp contrast with those in the second sub-sample. We also performed a formal test for the presence of a single structural break in the growth rate of the stock price index. We found statistically significant evidence of a structural break that occurred around 1944, practically in the middle of our total out-of-sample period. Specifically, starting from around 1944 the growth rate of the stock price index has more than doubled.

Most importantly, in our study we found evidence that the average Bull (Bear) market duration has increased (decreased) over time. As compared with the first sub-sample, over the second sub-sample the ratio of the average Bull market length to the average Bear market length has almost doubled. Since the benefits of moving average trading strategies come from timely identification of Bear market states and moving to cash, it is only logical to conclude that the potential advantage of market timing strategies over the passive strategy has diminished dramatically over time. Using our total historical sample we did find evidence that moving average rules significantly outperformed the market. However, this evidence comes mainly from the superior performance of market timing rules over the first half of our total sample. In contrast, over the second half of our total sample, even though both halves have exactly the same number of Bull and Bear market phases, we did not find statistically significant evidence of outperformance.

The evidence of regime shifts in the properties of cycles in stock prices is of paramount importance when it comes to the market timing issue. This evidence has two important implications. First, the fact that the stock market dynamics is subject to changes advocates that the observed past performance of trading rules is a poor estimate of the rules' expected future performance. Second, as applied to our question of investigation, the observed performance of trading rules over the first half of our sample has little or no relevance to the rules' expected performance in the future. Under the assumption that the stock market dynamics is changing very slowly over time, the observed performance of trading rules in the second half of our sample provides a more reliable estimate of the rules' expected future performance.

In our tests we selected the set of market timing rules with a few distinct shapes of the moving average weighting function. This allowed us to see whether the empirical performance of moving average rules depends on the weighting scheme. The common belief among the traders is that a moving average weighting scheme that overweights the most recent prices performs better than the moving average weighting scheme with equal weighting of prices. As the result, the P-SMA rule is much more popular among traders than the MOM rule. The other most popular rule among the traders is the DCM rule whose weighting function underweights both the most recent and the most distant prices. Many traders believe that this rule performs the best because it smoothes the noise in the most recent prices. However, we found no support for these common beliefs. On the contrary, over the first half of the total sample our results suggested that equal weighting of prices (as in the MOM rule) was the most optimal weighting scheme to use in market timing. In addition, over this half the DCM rule showed the worst performance. Over the second half of our sample, on the other hand, there was virtually no difference in the performances of trading rules with various weighting schemes.

We did not find the support for another common belief that there is some specific optimal lookback period in each trading rule. On the contrary, our results revealed that there are substantial time variations in the length of the optimal lookback period for each trading rule. For the majority of rules we found evidence that over the second half of our sample the mean length of the optimal lookback period was shorter than over the first half. The evidence of time variations in the length of the optimal lookback period motivated us to use, in out-of-sample tests, the rolling-window estimation scheme as a potentially better method to find the optimal lookback period than the commonly used expanding-window estimation scheme. However, we did not find statistically significant evidence that some specific estimation scheme produces better performance than the other one. Yet we found indications that in the MOM and P-REMA rules the rolling-window estimation scheme produces marginally better performance than the expanding-window estimation scheme. For the other two rules, P-SMA and DCM, the evidence was inconclusive. One potential disadvantage of the rolling-window estimation scheme is that it produces a larger number of trading signals than the expanding-window estimation scheme. Therefore the expanding-window estimation scheme might be advantageous when the transaction costs are large.

In our study we examined the performance of market timing rules over Bull and Bear markets separately. In addition, we studied the similarity between the Bull-Bear market states and the Buy-Sell trading signals generated by different rules. We found that all trading rules generate lots of false signals during both Bull and Bear markets. As a result, over Bull markets the performance of market timing rules was worse than that of the passive strategy. Despite the presence of false Buy signals during Bear markets, the market timing rules outperformed the market in Bear states. We found that the overall accuracy of trading rules in detecting the stock market states varies insignificantly between different rules and is only marginally better than the overall accuracy of the passive market strategy (assuming that the latter always generates a Buy trading signal). Over the whole sample, the overall accuracy (percentage of correctly classified stock market states) of market timing rules was about 75%. In other words, the market timings rules produced correct trading signals about 3/4 of time.

We also studied the performance of market timing rules over short- to medium-term horizons. We found that the outperformance generated by market timing rules is highly uneven over time. Our visual inspection of the historical performance over shorter horizons suggested that, over the whole sample, the superior outperformance of market timing rules was generated mainly over relatively few particular historical episodes: the severe Bear markets of the decades of 1870s, 1900s, 1930s, 1970s, and 2000s. Our analysis revealed that the median outperformance for the majority of rules was positive during the first half of our sample, whereas over the second half the median outperformance was negative for all the rules. If we use the historical outperformance during the second half of the sample as a reliable estimate of the expected future outperformance, this result tells us that the probability that a market timing rule will outperform the market over a short- to medium-term horizon is below 50%.

8 Concluding Remarks

In this paper we used the longest historical dataset, comprehensively re-examined the empirical performance of a few distinct market timing rules, and extended the previous studies in a number of ways. Our main results are as follows. We discovered strong evidence that the stock market dynamics are changing over time. Specifically, our findings revealed that over the second half of our sample the stock market was less volatile, the stock prices grew with a rate that was more than double as much as that over the first half, and the ratio of the average Bull market length to the average Bear market length was almost double as much as that over the fist half. We found evidence that over the total sample the moving average strategies outperformed the market. However, over the second half of our total sample, even though both halves were chosen to have exactly the same number of Bull and Bear market phases, we did not find statistically significant evidence of outperformance. Contrary to the common belief, our results indicated that there is no single optimal lookback period in each trading rule, as well as we found no support for the common belief that over-weighting the recent prices allows one to improve the performance of a market timing rule. Whereas we found some indications that over very long-term horizons the market timing strategy tends to outperform the market, over more realistic short- to medium-term horizons the market timing strategy is more likely to underperform the market than to outperform.

Thereby our findings cast doubts that market timing strategies can consistently beat the market. Therefore our findings are in sharp contrast with the findings reported in the majority of previous studies where the authors document that "market timing works". However, it is important to emphasize that our findings do not indicate that previous studies were implemented with some errors. In fact, we can easily reconcile our findings with prior studies. Already Zakamulin (2014) pointed to the following features of the market timing strategies: the outperformance delivered by market timing strategies is highly uneven over time; most of the outperformance is generated mainly over relatively few particular historical episodes; and, as the immediate consequence from these two features, the outcome of both in- and out-of-sample tests of profitability depends crucially on the choice of the historical sample period. If one chooses the sample period to be, for instance, either 1900-2010, 1970-2010, or 1990-2010, and simulates, for example, the SMA(10) strategy, then one comes to conclusion that market timing works. Yet strictly speaking, such a result tells us that a market timing strategy outperformed the market in some particular historical period in the past. The question of paramount importance is whether such a result represents a typical performance of the SMA(10) strategy, and whether the performance in this specific historical period can be used as a realiable estimate of the expected future performance.

Based on the findings revealed by our study, we can argue that the most relevant sample period for an empirical study on the profitability of trading rules is the whole post World War II period, as we found that prior to this period the stock market dynamics were significantly different. By starting the sample period from 1970, one excludes from the study a long period of 25 years where market timing strategies underperformed the market. Finally, by choosing a period that starts not long before the Dot-Com bubble crush and ends not long after the Global Financial Crisis, one captures the most successful period for market timing strategies where all of them delivered extraordinary good outperformance.

Concluding this paper, we would like to mention that the results of our empirical study, as the results of every empirical study, are, in principle, data-set specific and data-frequency specific. The data-frequency issue seems to be the least of these two concerns. In particular, Clare et al. (2013) find that there are no advantages in trading daily rather than monthly. That is, the performance of market timing rules virtually does not depend on the choice of the data frequency. In contrast, Zakamulin (2014) documents that the advantages of market timing rules depend on the choice of the underlying passive index. Specifically, the advantages of market timing are more apparent when the passive index is the S&P 500 and are less obvious when the passive index is the Dow Jones Industrial Average.

References

Aronson, D. (2006). *Evidence-Based Technical Analysis: Applying the Scientific Method and Statistical Inference to Trading Signals*. John Wiley & Sons, Ltd.

Berkowitz, S. A., Logue, D. E., and Noser, E. A. (1988). "The Total Costs of Transactions on the NYSE", *Journal of Finance*, *43*(1), 97–112.

Bodie, Z., Kane, A., and Marcus, A. J. (2007). *Investments*. McGraw Hill.

Brock, W., Lakonishok, J., and LeBaron, B. (1992). "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns", *Journal of Finance*, *47*(5), 1731–1764.

Brown, S. J., Goetzmann, W. N., and Kumar, A. (1998). "The Dow Theory: William Peter Hamilton's Track Record Reconsidered", *Journal of Finance*, *53*(4), 1311–1333.

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

Chan, L. K. C. and Lakonishok, J. (1993). "Institutional Trades and Intraday Stock Price Behavior", *Journal of Financial Economics*, *33*(2), 173–199.

Clare, A., Seaton, J., Smith, P. N., and Thomas, S. (2013). "Breaking Into the Blackbox: Trend Following, Stop losses and the Frequency of Trading - The Case of the S&P500", *Journal of Asset Management*, *14*(3), 182–194.

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

Freyre-Sanders, A., Guobuzaite, R., and Byrne, K. (2004). "A Review of Trading Cost Model: Reducing Transaction Costs", *Journal of Investing*, *13*(3), 93–116.

Glabadanidis, P. (2015). "Market Timing With Moving Averages", *International Review of Finance*, *15*(3), 387–425.

Gonzalez, L., Powell, J. G., Shi, J., and Wilson, A. (2005). "Two Centuries of Bull and Bear Market Cycles", *International Review of Economics and Finance*, *14*(4), 469 – 486.

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

Hamilton, J. D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, *57*(2), 357–384.

Hansen, P. R. and Timmermann, A. (2013). "Choice of Sample Split in Out-of-Sample Forecast Evaluation", Working paper, European University Institute, Stanford University and CREATES.

Jaccard, P. (1901). "Etude comparative de la distribution florale dans une portion des Alpes et des Jura", *Bulletin de la Societe Vaudoise des Sciences Naturelles*, *37*, 547–579.

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

Kaufman, P. J. (1995). *Smarter Trading: Improving Performance in Changing Markets.* McGraw-Hill, Inc.

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

Kirkpatrick, C. D. and Dahlquist, J. (2010). *Technical Analysis: The Complete Resource for Financial Market Technicians.* FT Press; 2nd edition.

Knez, P. J. and Ready, M. J. (1996). "Estimating the Profits from Trading Strategies", *Review of Financial Studies*, *9*(4), 1121–1163.

Kulperger, R. J. and Rao, B. L. S. P. (1989). "Bootstrapping a Finite State Markov Chain", *Sankhya: Indian Journal of Statistics, Series A (1961-2002)*, *51*(2), 178–191.

Lo, A. W., Mamaysky, H., and Wang, J. (2000). "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation", *Journal of Finance*, *55*(4), 1705–1770.

Lukac, L. P. and Brorsen, B. W. (1990). "A Comprehensive Test of Futures Market Disequilibrium", *Financial Review*, *25*(4), 593–622.

Lukac, L. P., Brorsen, B. W., and Irwin, S. H. (1988). "A Test of Futures Market Disequilibrium Using Twelve Different Technical Trading Systems", *Applied Economics*, *20*(5), 623–639.

Lunde, A. and Timmermann, A. (2004). "Duration Dependence in Stock Prices: An Analysis of Bull and Bear Markets", *Journal of Business and Economic Statistics*, *22*(3), 253–273.

Maheu, J. M. and McCurdy, T. H. (2000). "Identifying Bull and Bear Markets in Stock Returns", *Journal of Business and Economic Statistics*, *18*(1), 100–112.

Maheu, J. M., McCurdy, T. H., and Song, Y. (2009). "Extracting Bull and Bear Markets From Stock Returns", Working paper, University of Toronto.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series", *Journal of Econometrics*, *135*(1-2), 499–526.

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

Modigliani, F. and Modigliani, L. (1997). "Risk-Adjusted Performance", *Journal of Portfolio Management*, *23*(2), 45–54.

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

Muggeo, V. M. R. (2003). "Estimating Regression Models With Unknown Break-Points", *Statistics in Medicine*, *22*(19), 3055–3071.

Murphy, J. J. (1999). *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications.* New York Institute of Finance.

Neely, C., Weller, P., and Dittmar, R. (1997). "Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach", *Journal of Financial and Quantitative Analysis*, *32*, 405–426.

Okunev, J. and White, D. (2003). "Do Momentum-Based Strategies Still Work in Foreign Currency Markets?", *Journal of Financial and Quantitative Analysis*, *38*(2), 425–447.

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

Park, C.-H. and Irwin, S. H. (2007). "What Do We Know About the Profitability of Technical Analysis?", *Journal of Economic Surveys*, *21*(4), 786–826.

Pätäri, E. and Vilska, M. (2014). "Performance of Moving Average Trading Strategies over Varying Stock Market Conditions: the Finnish Evidence", *Applied Economics*, *46*(24), 2851–2872.

Pesaran, M. H., Pick, A., and Timmermann, A. (2011). "Variable Selection, Estimation and Inference for Multi-Period Forecasting Problems", *Journal of Econometrics*, *164*(1), 173–187.

Rossi, B. and Inoue, A. (2012). "Out-of-Sample Forecast Tests Robust to the Choice of Window Size", *Journal of Business and Economic Statistics*, *30*(3), 432–453.

Sharpe, W. F. (1994). "The Sharpe Ratio", *Journal of Portfolio Management*, *21*(1), 49–58.

Sortino, F. A. and Price, L. N. (1994). "Performance Measurement in a Downside Risk Framework", *Journal of Investing*, *3*(3), 59 – 65.

Sullivan, R., Timmermann, A., and White, H. (1999). "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap", *Journal of Finance*, *54*(5), 1647–1691.

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

White, H. (2000). "A Reality Check for Data Snooping", *Econometrica*, *68*(5), 1097–1126.

Zakamulin, V. (2014). "The Real-Life Performance of Market Timing with Moving Average and Time-Series Momentum Rules", *Journal of Asset Management*, *15*(4), 261–278.

Zakamulin, V. (2015). "Anatomy of Market Timing with Moving Averages", Working paper, University of Agder.

Zeileis, A., Kleiber, C., Krämer, W., and Hornik, K. (2003). "Testing and Dating of Structural Changes in Practice", *Computational Statistics and Data Analysis*, *44*(1-2), 109–123.

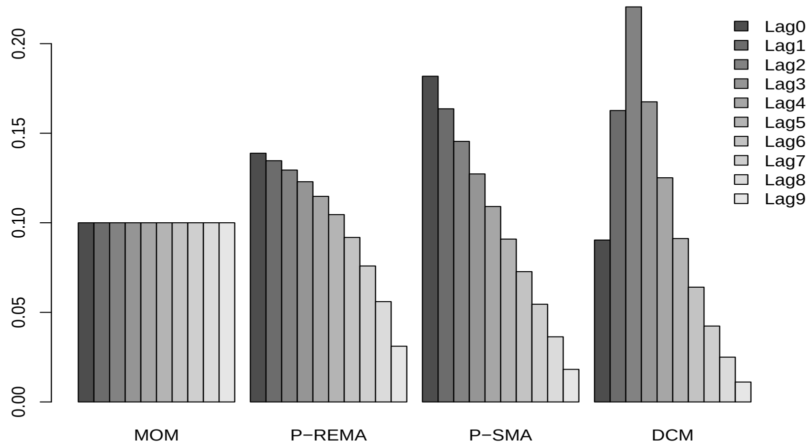

Figure 1: Weights of monthly price changes used for the computations of the technical trading indicators with $k = 10$. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$ and $s = 2$). **Lag**$(i-1)$ denotes the weight of the lag ΔP_{t-i}, where Lag0 denotes the most recent price change ΔP_{t-1} and Lag9 denotes the most oldest price change ΔP_{t-10}.

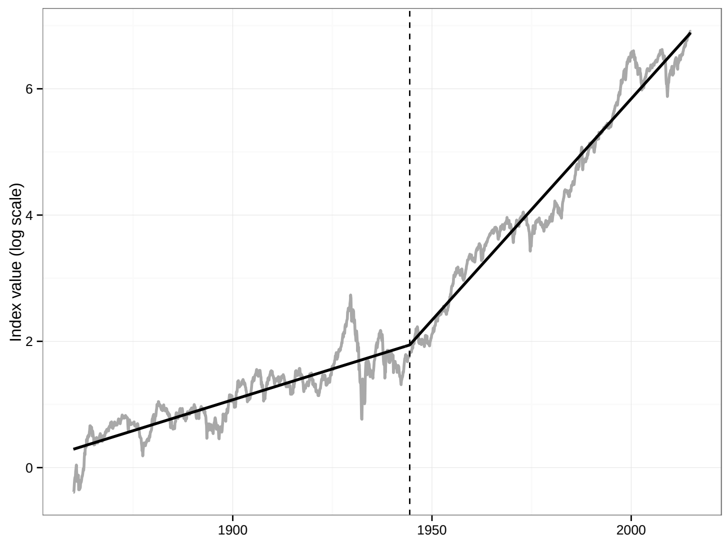

Figure 2: The log of the stock price index over 1860-2014 (gray line) versus the fitted segmented model (black line) given by $\log(I_t) = \log(I_0) + \mu\, t + \delta\, (t - t^*)^+ + \varepsilon_t$, where t^* is the breakpoint date, μ is the growth rate before the breakpoint, and $\mu + \delta$ is the growth rate after the breakpoint. The estimated date of the breakpoint is May 1944.

Panel A : Period 1870-1942

Panel B : Period 1942-2014

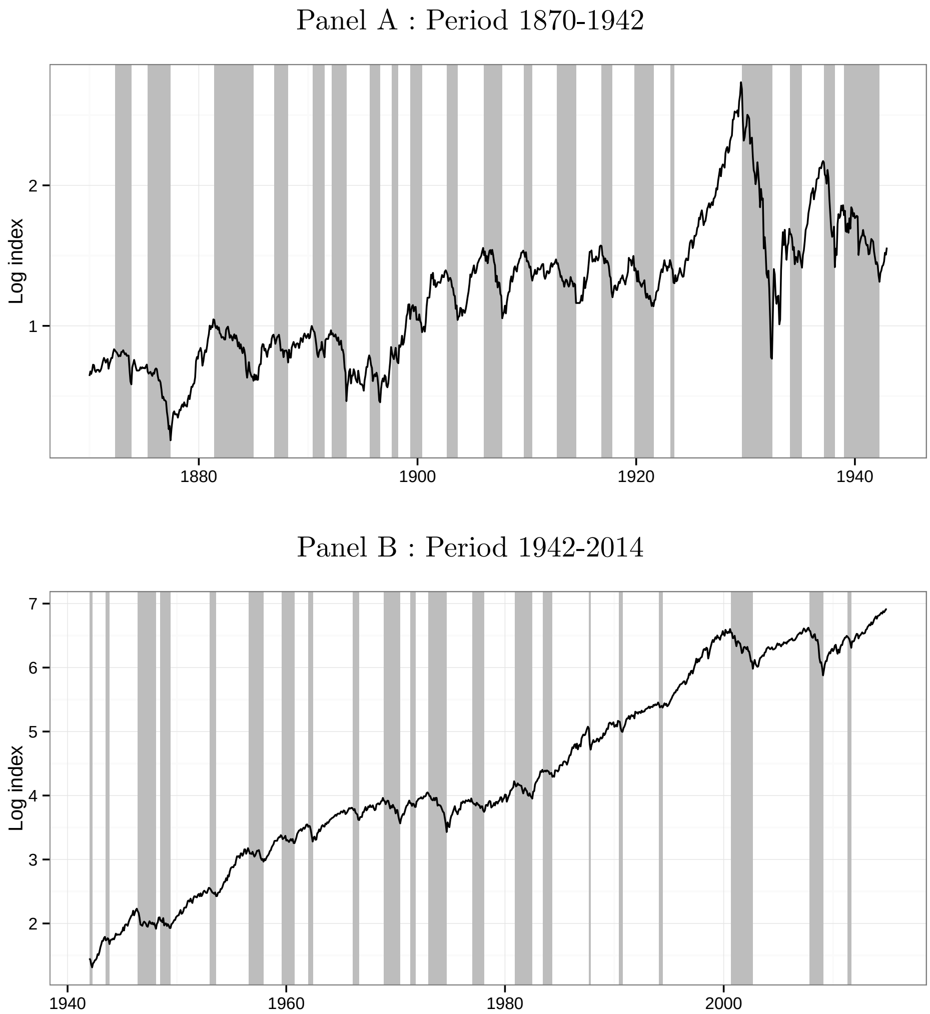

Figure 3: Bull and Bear markets over the two historical sub-periods: 1870-1942 and 1942-2014. Shaded areas indicate the Bear market phases.

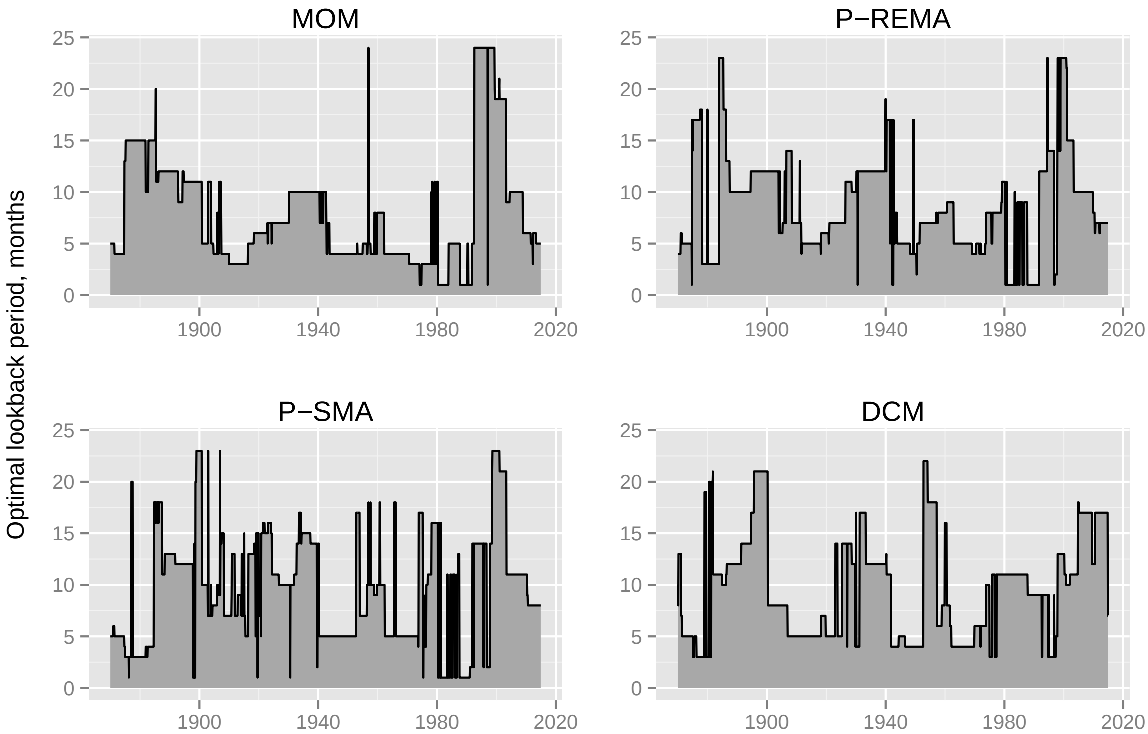

Figure 4: The optimal lookback period, measured in months, for different technical trading rules over a rolling window of 10 years. The first reported value for the optimal lookback period in the graphs is for the period from January 1860 to December 1869. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$).

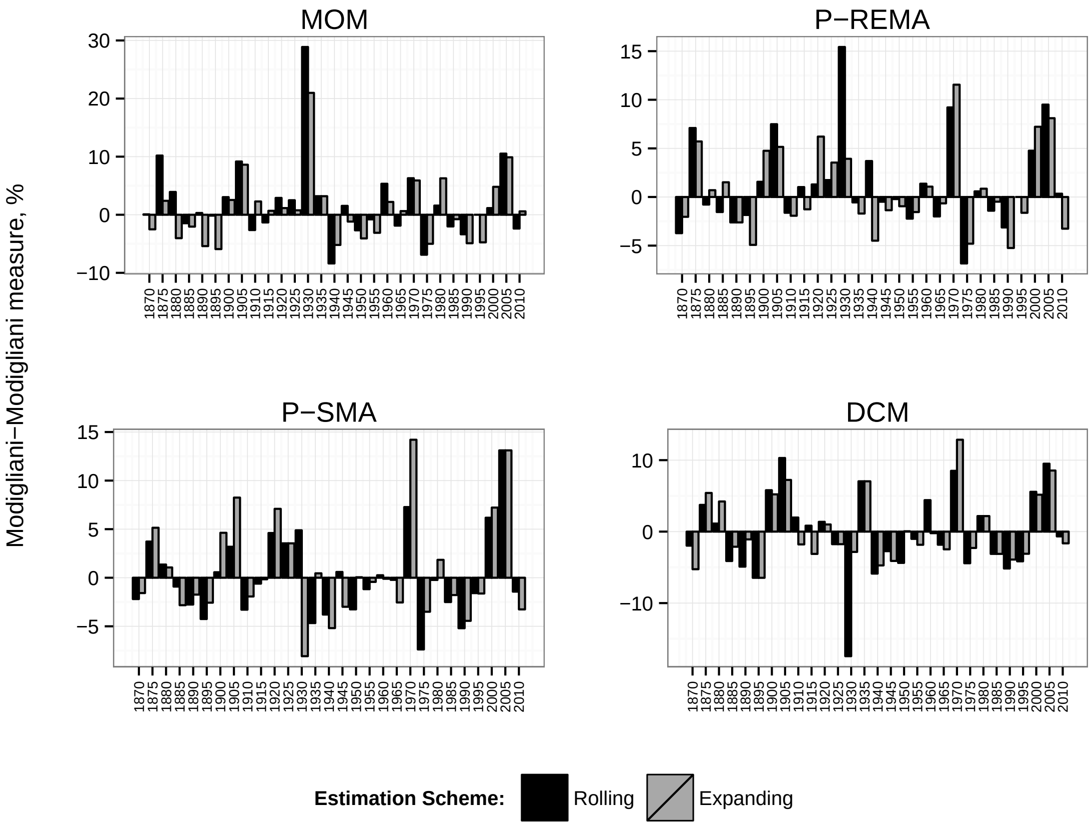

Figure 5: Annualized M^2 performance measure computed over disjoint intervals of 5-years. For each trading rule and estimation scheme we simulate real-life technical trading and compute the out-of-sample performance of market timing strategies over 5-year periods. The first period is from January 1870 to December 1874, the second period is from January 1875 to December 1879, etc. Then we plot the value of M^2 measure versus the historical period. The values of M^2 are reported in percents. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$).

Statistics	1870-2014			1870-1942			1942-2014		
	CAP	**MKT**	**RF**	**CAP**	**MKT**	**RF**	**CAP**	**MKT**	**RF**
Mean, %	0.49	0.85	0.31	0.27	0.70	0.29	0.71	1.01	0.33
Std. deviation, %	4.99	4.99	0.21	5.69	5.70	0.15	4.15	4.16	0.26
Skewness	0.23	0.28	1.03	0.52	0.57	0.19	-0.43	-0.43	0.94
Kurtosis	8.64	8.88	2.48	9.48	9.72	2.03	1.55	1.57	1.06
Shapiro-Wilk	**0.93**	**0.93**	**0.93**	**0.90**	**0.90**	**0.93**	**0.98**	**0.98**	**0.93**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AC1	**0.08**	**0.08**	**0.98**	**0.10**	**0.10**	**0.96**	0.03	0.04	**0.99**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.31)	(0.28)	(0.00)

Table 1: Descriptive statistics of data used in the study. **CAP**, **MKT**, and **RF** denote the capital appreciation return, the total market return, and the risk-free rate of return respectively. Means and standard deviation are annualized and reported in percents. **Shapiro-Wilk** denotes the value of the test statistics in the Shapiro-Wilk normality test. The p-values of the normality test are reported in brackets below the test statistics. $\mathbf{AC_1}$ denotes the first-order autocorrelation. For each AC_1 we test the hypothesis $H_0 : \text{AC}_1 = 0$. The p-values are reported in brackets below the values of autocorrelation. Bold text indicates values that are statistically significant at the 5% level.

Hypothesis	p-value
$H_0^1 : \mu^1_{CAP} = \mu^2_{CAP}$	0.06
$H_0^2 : \mu^1_{MKT} = \mu^2_{MKT}$	0.19
$H_0^3 : \mu^1_{RF} = \mu^2_{RF}$	0.00
$H_0^4 : \sigma^1_{CAP} = \sigma^2_{CAP}$	0.00
$H_0^5 : \sigma^1_{MKT} = \sigma^2_{MKT}$	0.00
$H_0^6 : \sigma^1_{RF} = \sigma^2_{RF}$	0.00

Table 2: Results of the hypothesis testing on the stability of the means and standard deviations of **CAP**, **MKT**, and **RF** which denote the capital appreciation return, the total market return, and the risk-free rate of return respectively. For example, the null hypothesis $H_0^1 : \mu^1_{CAP} = \mu^2_{CAP}$ is that the mean capital appreciation return during the first sub-period, μ^1_{CAP}, is equal to the mean capital appreciation return during the second sub-period, μ^2_{CAP}. Similarly, the null hypothesis $H_0^4 : \sigma^1_{CAP} = \sigma^2_{CAP}$ is that the volatility of the capital appreciation return during the first sub-period, σ^1_{CAP}, is equal to the volatility of the capital appreciation return during the second sub-period, σ^2_{CAP}.

	Linear model	Segmented model
(Intercept)	-5.96e-01	2.89e-01
	(0.00)	(0.00)
μ	3.45e-03	1.63e-03
	(0.00)	(0.00)
δ		4.21e-03
		(0.00)
R^2	0.90	0.98
adj. R^2	0.90	0.98
Residual st. dev.	0.62	0.27

Table 3: Results of the estimation of the two alternative models using the total sample period 1860-2014. The linear model is given by $\log(I_t) = \log(I_0) + \mu t + \varepsilon_t$. The segmented model is given by $\log(I_t) = \log(I_0) + \mu t + \delta (t - t^*)^+ + \varepsilon_t$. The p-values of the estimated coefficients are given in brackets. These p-values are computed using the heteroskedasticity and autocorrelation consistent standard errors. The estimated breakpoint date is May 1944.

Statistics	**1870-2014**		**1870-1942**		**1942-2014**	
	Bull	**Bear**	**Bull**	**Bear**	**Bull**	**Bear**
Number of phases	41	40	21	20	21	20
Minimum duration	6	3	9	5	6	3
Average duration	29.3	15.5	25.7	18.9	32.0	12.2
Median duration	28.0	14.0	25.0	15.0	32.0	11.5
Maximum duration	76	44	75	44	76	25
Average amplitude, %	65.19	-24.44	60.74	-29.05	66.90	-19.84
Average cum. return, %	75.12	-22.64	72.32	-25.82	75.27	-19.46
Mean monthly return, %	2.00	-1.46	2.15	-1.40	1.88	-1.54
Standard deviation, %	4.65	4.86	5.51	5.29	3.81	4.09

Table 4: Descriptive statistics of Bull and Bear markets. Duration is measured in the number of months. Amplitudes are defined as % changes in the stock index prices (not adjusted for dividends). Cumulative returns, mean monthly return and the standard deviations are computed using the total return (adjusted for dividends).

	1870-1942		**1942-2014**	
	Bear	Bull	Bear	Bull
Bear	0.943	0.057	0.904	0.096
Bull	0.040	0.960	0.031	0.969

Table 5: The estimated transition probabilities of the two-states Markov switching model for the stock market returns over two historical sub-periods: 1870-1942 and 1942-2014. The transition probabilities between the states $p_{ij} = P(S_t = j|S_{t-1} = i)$, where S_t denotes the latent state variable at time t.

Hypothesis	p-value
$H_0^1 : \mu_0^1 = \mu_0^2$	0.72
$H_0^2 : \mu_1^1 = \mu_1^2$	0.34
$H_0^3 : \sigma_0^1 = \sigma_0^2$	0.00
$H_0^4 : \sigma_1^1 = \sigma_1^2$	0.00
$H_0^{5_1} : p_{00}^1 = p_{00}^2$	0.00
$H_0^{5_2} : p_{01}^1 = p_{01}^2$	0.00
$H_0^{5_3} : p_{10}^1 = p_{10}^2$	0.00
$H_0^{5_4} : p_{11}^1 = p_{11}^2$	0.00

Table 6: Results of the hypothesis testing on the stability of the parameters of the two-states Markov switching model for the stock market returns r_t over the two sub-periods. The model for returns is given by $r_t^m | S_t \sim N\left(\mu_{S_t}^m, \left(\sigma_{S_t}^m\right)^2\right)$, where S_t denotes the latent state variable at time t (0-Bear state, 1-Bull state) and $m \in \{1, 2\}$ denotes the number of a sub-period. This model assumes that the stock market returns at time t of sub-period m are normally distributed with mean μ_0^m and standard deviation σ_0^m if the market is in state 0. Otherwise, in state 1, the stock market returns are normally distributed with mean μ_1^m and standard deviation σ_1^m. The transition probabilities between the states $p_{ij}^m = P^m(S_t = j | S_{t-1} = i)$, $i, j \in \{0, 1\}$, denote the probability of transition from state i to state j in sub-period m. For example, p_{01}^1 denotes the transition probability from state 0 to state 1 over the first sub-period.

	MOM	P-REMA	P-SMA	DCM
Panel A : Period 1870-2014				
Mean	8.1	8.8	9.7	9.7
Median	6	8	10	10
Minimum	1	1	1	3
Maximum	24	23	23	22
Panel B : Period 1870-1942				
Mean	8.7	9.6	10.3	9.6
Median	10	10	10	8
Minimum	3	1	1	3
Maximum	20	23	23	21
Panel C : Period 1942-2014				
Mean	7.4	7.8	9.0	9.8
Median	4	7	8	10
Minimum	1	1	1	3
Maximum	24	23	23	22

Table 7: Descriptive statistics of the optimal lookback period (the number of lagged monthly prices) for different technical trading rules over a rolling window of 10 years. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$).

	MOM	P-REMA	P-SMA	DCM
Mean optimal lookback period over 1870-1942	8.15	9.52	10.30	9.61
Mean optimal lookback period over 1942-2014	7.42	7.80	9.09	9.89
P-value of testing the equlity of means	0.01	0.00	0.00	0.26

Table 8: Mean optimal lookback periods over each of the two historical sub-samples (1870-1942 and 1942-2014) and the results of testing the null hypothesis of the equality of means. **MOM** denotes the Momentum rule. The optimal lookback period k_t is assumed to follow the process $k_t = c + \varphi\, k_{t-1} + \varepsilon_t$. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$).

	Bull markets					Bear markets				
	MKT	MOM	P-REMA	P-SMA	DCM	MKT	MOM	P-REMA	P-SMA	DCM
Panel A : Period 1870-2014										
Mean return	2.00	**1.56**	**1.49**	**1.45**	**1.46**	-1.46	**-0.59**	**-0.57**	**-0.64**	**-0.69**
		(0.01)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Std. deviation	4.65	**3.50**	**3.29**	**3.28**	**3.28**	4.86	**2.65**	**2.54**	**2.69**	**2.78**
		(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Sharpe ratio	1.27	1.24	1.25	1.21	1.22	-1.28	-1.21	-1.24	-1.26	-1.27
		(0.75)	(0.86)	(0.51)	(0.60)		(0.67)	(0.81)	(0.90)	(0.98)
Panel B : Period 1870-1942										
Mean return	2.15	1.64	**1.44**	**1.43**	**1.43**	-1.40	**-0.50**	**-0.44**	**-0.57**	**-0.63**
		(0.08)	(0.01)	(0.01)	(0.01)		(0.00)	(0.00)	(0.01)	(0.02)
Std. deviation	5.51	**3.89**	**3.44**	**3.44**	**3.44**	5.29	**2.65**	**2.38**	**2.76**	**2.90**
		(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Sharpe ratio	1.17	1.20	1.16	1.15	1.16	-1.11	-1.03	-1.06	-1.09	-1.11
		(0.83)	(0.94)	(0.87)	(0.92)		(0.70)	(0.82)	(0.90)	(0.99)
Panel C : Period 1942-2014										
Mean return	1.90	**1.49**	1.52	**1.48**	**1.48**	-1.55	**-0.74**	**-0.79**	**-0.76**	**-0.76**
		(0.03)	(0.05)	(0.03)	(0.03)		(0.01)	(0.02)	(0.01)	(0.01)
Std. deviation	3.80	**3.14**	**3.16**	**3.14**	**3.14**	4.08	**2.62**	**2.75**	**2.58**	**2.57**
		(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Sharpe ratio	1.44	1.29	1.32	1.28	1.28	-1.64	-1.48	-1.48	-1.53	-1.56
		(0.11)	(0.22)	(0.09)	(0.10)		(0.48)	(0.47)	(0.64)	(0.71)

Table 9: Descriptive statistics and performances of the trading strategies over the Bull and Bear markets. For each trading rule we simulate real-life technical trading using the **rolling-window** estimation scheme and compute the descriptive statistics of the passive market strategy and each active trading strategy. The descriptive statistics is for monthly returns, means and standard deviation are reported in percentages. In brackets we report the p-values of the following three hypothesis tests $H_0 : \mu_i = \mu_M$, $H_0 : \sigma_i = \sigma_M$, and $H_0 : SR_i = SR_M$ where μ_i, σ_i, and SR_i are the mean returns, volatility, and the Sharpe ratio of strategy i; μ_M, σ_M, and SR_M are the mean returns, volatility, and the Sharpe ratio of the market. The bold text indicates values that are statistically significant at the 5% level. **MKT** denotes the passive market strategy. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$).

	Bull markets					Bear markets				
	MKT	MOM	P-REMA	P-SMA	DCM	MKT	MOM	P-REMA	P-SMA	DCM
Panel A : Period 1870-2014										
Mean return	2.00	**1.49**	**1.46**	**1.47**	**1.45**	-1.46	**-0.63**	**-0.56**	**-0.55**	**-0.62**
		(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Std. deviation	4.65	**3.44**	**3.29**	**3.36**	**3.29**	4.86	**2.79**	**2.64**	**2.68**	**2.79**
		(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Sharpe ratio	1.27	1.19	1.21	1.19	1.20	-1.28	-1.19	-1.17	-1.15	-1.19
		(0.34)	(0.56)	(0.40)	(0.47)		(0.57)	(0.51)	(0.42)	(0.56)
Panel B : Period 1870-1942										
Mean return	2.15	**1.51**	**1.46**	**1.42**	**1.42**	-1.40	**-0.61**	**-0.52**	**-0.49**	**-0.59**
		(0.03)	(0.02)	(0.01)	(0.01)		(0.01)	(0.00)	(0.00)	(0.01)
Std. deviation	5.51	**3.80**	**3.46**	**3.59**	**3.44**	5.29	**2.83**	**2.63**	**2.72**	**2.87**
		(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Sharpe ratio	1.17	1.12	1.17	1.10	1.14	-1.11	-1.10	-1.07	-1.01	-1.07
		(0.68)	(0.99)	(0.59)	(0.82)		(0.97)	(0.85)	(0.59)	(0.82)
Panel C : Period 1942-2014										
Mean return	1.90	**1.44**	**1.40**	**1.40**	**1.49**	-1.55	**-0.67**	**-0.65**	**-0.67**	**-0.75**
		(0.02)	(0.01)	(0.01)	(0.04)		(0.01)	(0.01)	(0.01)	(0.01)
Std. deviation	3.80	**3.11**	**3.05**	**3.05**	**3.17**	4.08	**2.61**	**2.56**	**2.52**	**2.54**
		(0.00)	(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Sharpe ratio	1.44	**1.25**	**1.23**	**1.23**	1.28	-1.64	-1.42	-1.43	-1.46	-1.54
		(0.05)	(0.04)	(0.04)	(0.09)		(0.33)	(0.36)	(0.46)	(0.68)

Table 10: Descriptive statistics and performances of the trading strategies over the Bull and Bear markets. For each trading rule we simulate real-life technical trading using the **expanding-window** estimation scheme and compute the descriptive statistics of the passive market strategy and each active trading strategy. The descriptive statistics is for monthly returns, means and standard deviation are reported in percentages. In brackets we report the p-values of the following three hypothesis tests $H_0 : \mu_i = \mu_M$, $H_0 : \sigma_i = \sigma_M$, and $H_0 : SR_i = SR_M$ where μ_i, σ_i, and SR_i are the mean returns, volatility, and the Sharpe ratio of strategy i; μ_M, σ_M, and SR_M are the mean returns, volatility, and the Sharpe ratio of the market. The bold text indicates values that are statistically significant at the 5% level. **MKT** denotes the passive market strategy. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$).

Period	Statistics	MOM		P-REMA		P-SMA		DCM	
		Buy	**Sell**	**Buy**	**Sell**	**Buy**	**Sell**	**Buy**	**Sell**
Panel A : Rolling-window estimation scheme									
1870-2014	Number of signals	121	121	108	108	123	123	101	101
	Jaccard Similarity Coef.	0.63	0.42	0.64	0.43	0.64	0.43	0.64	0.43
1870-1942	Number of signals	58	58	58	58	65	65	54	54
	Jaccard Similarity Coef.	0.55	0.45	0.54	0.46	0.54	0.44	0.56	0.46
1942-2014	Number of signals	64	64	51	51	59	59	48	48
	Jaccard Similarity Coef.	0.70	0.37	0.73	0.39	0.73	0.41	0.71	0.39
Panel B : Expanding-window estimation scheme									
1870-2014	Number of signals	119	119	91	91	98	98	83	83
	Jaccard Similarity Coef.	0.62	0.41	0.64	0.43	0.66	0.45	0.64	0.44
1870-1942	Number of signals	64	64	55	55	58	58	47	47
	Jaccard Similarity Coef.	0.52	0.42	0.55	0.46	0.57	0.48	0.56	0.47
1942-2014	Number of signals	56	56	37	37	41	41	37	37
	Jaccard Similarity Coef.	0.71	0.40	0.71	0.38	0.73	0.41	0.71	0.39

Table 11: Number of trading signals and Jaccard similarity coefficients. For each trading rule and estimation scheme we simulate real-life technical trading and compute the number of trading signals generated by each active trading strategy. In addition, we compute the Jaccard similarity coefficients between the Bull market states and Buy signals (JSC(Bull,Buy) in columns **Buy**) and between the Bear market states and Sell signals (JSC(Bear,Sell) in columns **Sell**). **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$).

Period	MKT	MOM	P-REMA	P-SMA	DCM
Panel A : Rolling-window estimation scheme					
1870-2014	0.67	0.71	0.72	0.71	0.72
1870-1942	0.59	0.67	0.67	0.66	0.68
1942-2014	0.74	0.74	0.77	0.77	0.75
Panel B : Expanding-window estimation scheme					
1870-2014	0.67	0.70	0.72	0.73	0.72
1870-1942	0.59	0.64	0.68	0.69	0.68
1942-2014	0.74	0.76	0.76	0.77	0.76

Table 12: Simple matching coefficients between the Bull-Bear market states and generated Buy-Sell trading signals. **MKT** denotes the passive market strategy. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$). It is assumed that the passive market strategy generates a Buy signal for all months.

	MKT	MOM		P-REMA		P-SMA		DCM	
		Rol	**Exp**	**Rol**	**Exp**	**Rol**	**Exp**	**Rol**	**Exp**
Panel A : Period 1870-2014									
Mean	0.85	0.84	0.78	0.80	0.79	0.75	0.79	0.74	0.76
Std. dev.	4.99	3.40	3.39	3.21	3.23	3.25	3.29	3.28	3.28
Skewness	0.28	0.79	0.68	-0.14	-0.45	-0.40	-0.50	-0.54	-0.51
Minimum	-29.43	-23.51	-23.51	-21.54	-23.51	-23.51	-23.51	-23.51	-23.76
Maximum	42.91	42.66	42.66	16.09	16.09	16.09	16.09	13.46	16.09
Sharpe ratio	0.37	**0.54**	0.48	**0.53**	0.51	0.47	0.51	0.46	0.47
		(0.01)	(0.11)	(0.03)	(0.06)	(0.18)	(0.06)	(0.24)	(0.16)
Panel B : Period 1870-1942									
Mean	0.70	0.77	0.65	0.68	0.65	0.61	0.64	0.59	0.60
Std. dev.	5.70	3.60	3.59	3.19	3.29	3.32	3.39	3.38	3.37
Skewness	0.57	1.63	1.49	0.24	-0.45	-0.33	-0.54	-0.62	-0.55
Minimum	-29.43	-23.51	-23.51	-19.66	-23.51	-23.51	-23.51	-23.51	-23.76
Maximum	42.91	42.66	42.66	16.09	16.09	16.09	16.09	13.33	16.09
Sharpe ratio	0.25	**0.46**	0.35	0.42	0.39	0.34	0.36	0.31	0.32
		(0.04)	(0.34)	(0.12)	(0.22)	(0.41)	(0.30)	(0.57)	(0.52)
Panel C : Period 1942-2014									
Mean	1.01	0.91	0.89	0.93	0.87	0.90	0.87	0.90	0.91
Std. dev.	4.16	3.17	3.13	3.22	3.06	3.16	3.06	3.16	3.17
Skewness	-0.43	-0.42	-0.43	-0.49	-0.57	-0.46	-0.57	-0.42	-0.35
Minimum	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54	-21.54
Maximum	16.78	13.21	13.21	13.46	12.17	13.46	12.17	13.46	13.46
Sharpe ratio	0.56	0.63	0.62	0.64	0.60	0.62	0.60	0.62	0.63
		(0.40)	(0.49)	(0.34)	(0.61)	(0.48)	(0.65)	(0.47)	(0.40)

Table 13: Descriptive statistics and performances of the trading strategies. For each trading rule and estimation scheme we simulate real-life technical trading and compute the descriptive statistics of the passive market strategy and each active trading strategy. The descriptive statistics is for monthly returns, means and standard deviation are reported in percentages. The Sharpe ratios are annualized; the p-values of testing the null hypothesis $H_0 : SR_i = SR_M$, where SR_i is the Sharpe ratio of trading strategy i and SR_M is the Sharpe ratio of the market, are reported in brackets. The bold text indicates values that are statistically significant at the 5% level. **MKT** denotes the passive market strategy. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$). **Rol** and **Exp** denote the rolling- and expanding-window estimation schemes respectively.

	MOM		P-REMA		P-SMA		DCM	
	Rol	**Exp**	**Rol**	**Exp**	**Rol**	**Exp**	**Rol**	**Exp**
Panel A : Period 1870-2014								
Minimum	-8.38	-5.91	-6.83	-5.25	-7.38	-8.08	-17.43	-6.46
Median	0.31	0.63	0.00	-0.66	-0.61	-0.43	-1.01	-1.76
Mean	1.95	0.83	1.24	0.74	0.13	0.75	-0.26	0.24
Maximum	28.88	20.98	15.43	11.55	13.11	14.21	10.30	12.86
Std. Deviation	6.83	5.80	4.70	4.32	4.35	5.25	5.85	4.76
Outperf. Prob.	0.55	0.55	0.48	0.45	0.41	0.41	0.45	0.38
Panel B : Period 1870-1939								
Minimum	-2.63	-5.91	-3.72	-4.93	-4.67	-8.08	-17.43	-6.46
Median	2.70	0.96	0.23	1.11	-0.02	0.15	0.98	-1.42
Mean	4.19	1.62	1.64	1.22	0.23	0.81	-0.32	0.40
Maximum	28.88	20.98	15.43	6.21	4.89	8.25	10.30	7.23
Std. Deviation	8.02	6.78	5.15	3.66	3.37	4.47	6.81	4.59
Outperf. Prob.	0.71	0.64	0.50	0.57	0.50	0.50	0.57	0.43
Panel C : Period 1940-2014								
Minimum	-8.38	-11.26	-6.83	-7.68	-7.38	-8.58	-5.87	-6.90
Median	-0.80	-1.88	0.00	-0.23	-1.18	-1.09	-1.84	-1.84
Mean	-0.13	-1.05	0.88	1.30	0.04	0.06	-0.21	0.46
Maximum	10.52	9.24	9.50	15.95	13.11	8.11	9.50	18.16
Std. Deviation	4.89	5.63	4.39	7.50	5.23	4.79	5.04	6.52
Outperf. Prob.	0.40	0.27	0.47	0.47	0.33	0.47	0.33	0.33

Table 14: Descriptive statistics of the annualized M^2 performance measure which tells us how much a market timing strategy outperforms the passive strategy on a risk adjusted basis. For each trading rule and estimation scheme we simulate real-life technical trading and compute the descriptive statistics of M^2 measure over 5-year non-overlapping periods. The values of all descriptive statistics are reported in percents. **Outperf. Prob.** denotes the historical probability of outperformance. **MOM** denotes the Momentum rule. **P-REMA** denotes the Price-Minus-Reverse-Exponential-Moving-Average rule (with $\lambda = 0.8$). **P-SMA** denotes the Price-Minus-Simple-Moving-Average rule. **DCM** denotes the Double Crossover Method (based on using two exponential moving averages with $\lambda = 0.8$). **Rol** and **Exp** denote the rolling- and expanding-window estimation schemes respectively.