Investing in the S&P 500 index: Can anything beat the buy-and-hold strategy?*

Hubert Dichtl[#]

dichtl research & consulting GmbH
Am Bahnhof 7, 65812 Bad Soden am Taunus, Germany

First Version: July 2015
This Version: February 2018

Abstract

Determining whether investment strategies exist that provide higher (risk-adjusted) returns than buying and holding the S&P 500 stock market index is not only highly relevant for finance theory, but also for the asset management industry. This study conducts a comprehensive test using realistic investment strategies based on monthly seasonalities, technical indicators, and fundamental factors (over 4,100 strategies in total). To assess statistical significance, we use Hansen's (2005) data-snooping-resistant SPA test. The results show that only investment strategies trying to exploit underreaction and overreaction effects with technical indicators dominate the buy-and-hold strategy in some simulation setups. These investment strategies are clearly superior to the strategies based on seasonalities and fundamental factors. Given that underreaction and overreaction effects are mainly justified with cognitive biases, our results support the economic relevance of behavioral finance insights.

JEL Classification: G11, G12, G14

Keywords: S&P 500 index, investment strategies, monthly seasonalities, fundamental factors, technical indicators, market anomalies, behavioral finance, superior predictive ability test

---

* I am grateful to Wolfgang Drobetz for helpful comments and suggestions.
# Corresponding author: Tel.: +49-6196-972-99-00; E-mail: dichtl@dichtl-rc.de.
Address: dichtl research & consulting GmbH, Am Bahnhof 7, 65812 Bad Soden am Taunus, Germany.

1. Introduction

Determining whether investment strategies exist that can provide higher (risk-adjusted) returns than simply buying and holding the S&P 500 is not only highly relevant for finance theory, but also for the asset management industry. Investment strategies that attempt to outperform a buy-and-hold strategy with publicly available information can be roughly divided into three approaches: strategies relying on fundamental factors, seasonalities, or technical indicators.

One possible way trying to outperform a passive buy-and-hold strategy is the application of investment strategies based on economic predictors. These models usually use fundamental and macroeconomic factors such as short-term interest rates (Fama and Schwert, 1977), credit spreads (Keim and Stambaugh, 1986), the term structure (Campbell, 1987), and the dividend yield (Fama and French, 1988) as predictors. Several studies find that the power of fundamental prediction approaches can be substantially improved when they are applied in combination with more sophisticated forecast techniques, for example, principal component analysis (e.g., Ludvigson and Ng, 2007; Rapach and Zhou, 2013) or forecast combination techniques (e.g., Rapach et al., 2010; Rapach and Zhou, 2013).

Beside these investment strategies that are founded in economic fundamentals, there is also empirical evidence that profitable seasonal anomalies exist. For example, Andrade et al. (2013), Haug and Hirschey (2011), and Jacobsen and Zhang (2014) all provide evidence of superior trading strategies based on calendar effects. The explanations of such effects are distinct. While one strand of literature explains the seasonality patterns with more rational arguments such as vacations (e.g., Jacobsen and Zhang, 2014; Zhang, 2014) or the tax-loss selling hypothesis (e.g., D'Mello et al., 2003), another strand of literature justifies these phenomenon with weather- or temperature-induced mood (e.g., Kamstra et al., 2003; Cao and Wei, 2005; Goetzmann et al., 2015).

In a widely cited study, Moskowitz et al. (2012) document significant time series momentum effects in stock, bond, currency, and commodity markets. They demonstrate that deploying this pricing anomaly with a trading strategy based on a simple momentum indicator leads to a higher performance compared to the buy-and-hold strategy. In contrast to investment strategies utilizing fundamental prediction models, the success of trading strategies based on technical indicators (such as the momentum or moving average indicator) can be predominantly explained with behavioral biases.[1] Investment strategies trying to exploit time series momentum can only be successful when markets move in trends. Hurst et al. (2013) demonstrate that trends result from underreaction and overreaction effects of market prices that can be attributed to behavioral phenomena such as anchoring (e.g., Tversky and Kahneman, 1974), the disposition effect (e.g., Shefrin and Statman, 1985; Frazzini, 2006), representativeness and conservatism (Barberis et al., 1998), or the overconfidence and self-attribution effect (Daniel et al., 1998).

Despite the justification of the investment strategies described above with either economic explanations or cognitive biases, current insights from the academic finance research call for a comprehensive and integrated test if it is really possible to outperform the S&P 500 index statistically and economically significant. Firstly, and most important, Harvey emphasizes in his AFA Presidential Address 2017 „The Scientific Outlook in Financial Economics“ that many of the published results will fail to hold up. He argues that there is an incentive to produce "significant" results due to the competition for top journal space in combination with unreported tests, the lack of adjustment for multiple tests, as well as direct or indirect p-hacking (Harvey, 2017). Secondly, McLean and Pontiff (2016) provide evidence that investors learn about mispricing from academic publications, with the effect that the post-publication returns of trading strategies based on

[1] Beside simple momentum indicators, moving average indicators are also often applied for implementing time series momentum-strategies (e.g., Zhu and Zhou, 2009; Neely et al, 2014; Marshall et al., 2017). See also section 3.

anomalies are significantly lower compared to the returns reported in the original studies. Given their results, it is not clear whether the documented superiority of some trading strategies still holds up when they are implemented with more recent data.

Against this background, this study tries to identify investment strategies that are superior to the simple buy-and-hold-strategy of the S&P 500 index. We simulate investment strategies based on seasonal patterns, technical indicators, as well as fundamental forecasting models – over 4,100 tested strategies in total – under realistic conditions (e.g., consideration of transaction costs), and compare the results with a buy-and-hold strategy.[2] In this way, our study also comprises a systematic peer-group comparison of "rationally-motivated" fundamental strategies, that are economically grounded with more "behavioral-motivated" technical analysis strategies trying to exploit pricing anomalies.[3]

In order to account for the data-snooping bias problem (Lo and MacKinlay, 1990; Harvey, 2017), we apply Hansen's (2005) test for superior predictive ability (the SPA test). This multiple test-framework enables a systematic comparison of a population of competing investment strategies in terms of economic evaluation criteria (i.e., absolute or risk-adjusted profit).[4]

Our results can be summarized as follows: For the seasonality-based investment strategies, there are time periods when the well-known "Sell in May and Go Away" or "September swoon" strategies dominate buying and holding the S&P 500, but these results are neither statistically

---

[2] Despite this large number of tested models, yet we are limited in the capacity. While we still test all possible monthly seasonal allocation models, within the group of fundamental forecast models and technical indicator models we try to cover the (in our view) most popular strategies. Therefore, our study doesn't assert a claim for considering "all available" models.

[3] As already mentioned, until today it is not clear if seasonal patterns can be better explained with rational issues (e.g., vacations) or weather- and temperature-induced mood.

[4] In a highly respected study, Harvey et al. (2016) emphasizes the data-snooping problematic in hundreds of papers attempting to explain the cross-section of expected returns and recommend larger critical t-values in econometric models (e.g., linear regression models). However, as we compare trading strategies with economic evaluation criteria (i.e., mean return and Sharpe ratio) and don't evaluate the statistical significance of regression parameters, we decided to apply Hansen's (2005) SPA test.

significant nor robust over different time periods. The same result also holds for the fundamental-based investment strategies.

We observe the best results for technical indicator strategies, where statistically significant higher risk-adjusted excess returns compared to the buy-and-hold strategy are obtained in the time period from 2000 to 2014.[5] When substituting for the S&P 500 buy-and-hold benchmark with a risk-reduced 50:50 constant mix strategy, the risk-adjusted outperformance still persists over this "crisis period".[6] However, when we go beyond the mean-variance framework and examine higher moments of the return distribution, our analysis confirms the findings of existing studies. In particular, moving average strategies have a positive effect on skewness and lead to remarkably lower maximum drawdowns than a buy-and-hold strategy. Furthermore, we find that the moving average strategy provides a statistically significant alpha return during turbulent stock market cycles.

All in all, our results show that investment strategies that attempt to exploit underreaction and overreaction effects in the S&P 500 index with technical indicators clearly dominate fundamental and seasonality investment approaches. Given that underreaction and overreaction effects are mainly justified with cognitive biases, our results support the economic relevance of behavioral finance insights.

The remainder of this paper is organized as follows. Section 2 provides a literature review of various recent studies that have challenged the buy-and-hold strategy as superior investment approach. The implemented investment strategies are described in section 3, while section 4

[5] In our longer backtest period ranging from 1966 to 2014 we observe a risk-adjusted outperformance against the buy-and-hold-strategy at a significance level around 10 percent. Given that Hansen's SPA test is a very conservative test (e.g., Hsu et al., 2014), this superiority could also be judged as significant. See the discussion in section 6.

[6] This investment period is suffered from two serious crises: the dot.com bubble from 2000-2002 and the 2008/2009 global financial markets crises (Hurst et al., 2017).

outlines the data-snooping-resistant hypothesis test that is used to verify the results in terms of statistical significance. Section 5 presents the empirical results, and section 6 provides further insights and economic explanations. Section 7 concludes, and gives an outlook for further research.

2. Literature review

In order to verify the outperformance potential of active investment strategies compared to buying and holding the S&P 500, it is important to evaluate three different strands of the literature: Calendar anomalies, technical analysis, and fundamental analysis.

Calendar anomalies

A calendar anomaly refers to the phenomenon whereby systematically higher returns are observed in some months than in others.[7] For example, the well-known January effect posits higher returns in January (e.g., Haug and Hirschey, 2006), while the "Halloween" effect posits systematically lower returns from May to October than from November to April.

In their renowned study, Bouman and Jacobsen (2002) document systematically higher returns for the "Sell in May and Go Away" strategy that invests in the stock market from November to April and then switches to cash from May to October. Of eighteen tested stock markets, Bouman and Jacobsen (2002) found only two cases (the Hong Kong and South African stock markets) where the Halloween strategy provided lower returns than a buy-and-hold. Logically, because of the 100% cash market allocation from May to October, the Halloween strategy in all cases exhibits lower volatility than the buy-and-hold strategy. Some recent research has cast doubt on whether

[7] Calendar anomalies bounded on daily data frequencies also exist (e.g., the weekend effect and the turn-of-the-month effect). However, due to the monthly availability of many macroeconomic and fundamental data, the analysis is conducted using monthly data.

the outperformance potential of this simple calendar-based strategy still exists (e.g., Lucey and Zhao, 2008; Jones and Lundstrum, 2009; Dichtl and Drobetz, 2014, 2015), but some other recent studies claim the opposite (Swinkels and van Vliet, 2012; Andrade et al., 2013; Jacobsen and Zhang, 2014).

Another oft-cited calendar anomaly is the "September swoon," which documents systematically negative stock market returns during September. Haug and Hirschey (2011) emphasize that the magnitude of this effect should be large enough for a promising trading strategy. Due to the diverging results of existing studies, we consider calendar-based investment strategies here.

Technical analysis

Another approach trying to beat the passive buy-and-hold strategy is the possibility of implementing promising investment strategies that are based on historical prices and volume data (technical analysis). A variety of technical indicators exist (e.g., Sullivan et al., 1999). But two of them – time series momentum and the moving average indicator – seem especially promising when implemented with a monthly frequency time series.[8]

Moskowitz et al. (2012) find a persistent and significant time series momentum in the fifty-eight liquid instruments for equity index, currency, commodity, and bond markets. They document that the strongest relationship exists between a security's next month excess return and the lagged twelve-month return. This relationship can be exploited in simple trading strategies to generate a significant outperformance. The time series momentum evaluates the difference between the actual

[8] Various studies analyzing technical indicators use daily data (e.g., Brock et al., 1992; Sullivan et al., 1999; Bajgrowicz and Scaillet, 2012; Marshall et al., 2017). Because this analysis is conducted with monthly data (due to the availability of economic data for the fundamental models), these studies are not discussed in detail here.

price and the n-month lagged price, while a simple moving average strategy uses the difference between the actual price and the moving average calculated from the actual and last n prices.[9]

Various studies have provided evidence that an investment strategy based on moving averages applied with monthly time series data can generate an outperformance (for examples, see Faber, 2007, 2009; Kilgallen, 2012; and Clare et al., 2016). This strategy has been successfully applied to stock and bond markets, commodities, currencies, etc. Moreover, Glabadanidis (2014) documents a clear dominance of the moving average trading strategy over buying and holding U.S. REIT indices.

In addition to historical prices, volume data play a significant role in technical analysis. Blume et al. (1994) and Lo et al. (2000) demonstrate that volume data provide further information that cannot be deduced from the price statistic alone. Thus, many studies that consider technical indicators like time series momentum and moving average also take volume data into account (e.g., Neely et al., 2014). Because of this positive evidence, technical indicators (based on historical prices and volume data) are also considered in this analysis.

Fundamental analysis

Another strand of the literature develops predictive regression models with fundamental and/or macroeconomic data and implements investment strategies on their forecasts.

The empirical evidence regarding the potential to outperform the passive buy-and-hold strategy with such an investment approach is unclear. Some recent studies emphasize that fundamental models not even beat the historical mean of (excess) returns (e.g., Welch and Goyal, 2008); others find evidence that fundamental data can provide recoverable forecast information.

[9] An alternative to the moving average strategy is to replace the actual price with a short moving average of recent prices, and then calculate the price difference between the short and long moving averages (e.g., Neely et al., 2014).

For example, Campbell and Thompson (2008) argue that the forecast power of fundamental predictive regressions can be substantially improved when weak restrictions on the signs of coefficients and return forecasts are imposed. Rapach et al. (2010) and Rapach and Zhou (2013) show that a combination of fundamental and macroeconomic forecasts leads to better results than those obtained by Welch and Goyal (2008). Ludvigson and Ng (2007), Rapach and Zhou (2013), and Neely et al. (2014) provide evidence that predictive regressions with diffusion indices (principal components) derived from fundamental and macroeconomic variables can significantly improve forecast power. Furthermore, following Hammerschmid and Lohre (2018), the incorporation of the regime factors leads to a better prediction of the equity risk premium with fundamental variables.[10]

To summarize, it seems to be fruitful in this study to consider investment strategies based on monthly seasonality effects, technical indicators, and fundamental-based predictive regression models.

3. Implemented investment strategies

This section describes the active investment strategies that are implemented based on the three different strategy types noted above.

3.1 Investment strategies based on monthly calendar effects

As section 2 discussed, numerous studies have analyzed the January effect, the Halloween effect, and the September swoon. This study follows Dichtl and Drobetz (2014), and implements a more comprehensive approach, as follows. In each month, one can either be invested 100% in the

[10] Other pertinent research along these lines includes Dangl and Halling (2012), who demonstrate that out-of-sample predictability can be enhanced by taking the time variation of coefficients into account.

stock market, or 100% in cash. In this vein, we obtain 2^{12} = 4,096 different monthly allocation models (labeled as SEA0 to SEA4095). For each seasonal trading strategy, the same monthly allocation is applied in each year during the sample period.

For example, while the SEA0 model is invested in the stock market for each of the twelve months (the buy-and-hold benchmark), the SEA9 model is suitable for exploiting the September swoon (cash market investment in September and stock market investment in all other months). The SEA2482 model is also of particular interest because it implements the Halloween strategy (for a full model list, see Appendix A1).[11]

Furthermore, while the SEA0 model is permanently invested in the stock market, the SEA4095 model invests in the cash market in all twelve months. In contrast to all other monthly allocation models, these two models do not increase transaction costs.

3.2 Investment strategies based on technical indicators

For the technical indicator-based investment strategies, the moving average indicator, the momentum indicator, and a volume indicator are used (e.g., Neely et al., 2014). Exhibit 1 shows that these popular technical indicators are also implemented in many other studies, with mostly positive results. The moving average trading rule is defined as:

(1a) $$S_{i,t} = \begin{cases} 1 & if \;\; MA_{s,t} \geq MA_{l,t} \\ 0 & if \;\; MA_{s,t} < MA_{l,t} \end{cases} \qquad with \; s < l$$

where:

(1b) $$MA_{j,t} = (1/j) \sum_{i=0}^{j-1} P_{t-i} \qquad for \; j = s, l$$

---

[11] The Halloween strategy is one of the monthly allocation strategies that invests for exactly six months in the stock market and six months in the cash market. To summarize, we have $\binom{12}{6}$ = 924 different strategies that invest this way.

The moving average indicator is based on the moving averages on the S&P 500 stock price index (P_t), as defined in Equation (1b). According to Equation (1a), we invest in the S&P 500 index (including all dividends) if the short moving average $\left(MA_{s,t}\right)$ is above the long moving average $\left(MA_{l,t}\right)$. Otherwise, an allocation in the cash market is taken $\left(S_{i,t} = 1 \text{ or } S_{i,t} = 0\right)$.[12] Exhibit 1 gives an overview of the various studies that implement moving average strategies and their specific parameterizations.

[12] In the context of short term trading models based on technical indicators (e.g., daily data frequency), the cash position is often substituted with a short sale (e.g., Brock et al., 1992; Sullivan et al., 1999). However, if moving average strategies are implemented with monthly data (like in our case), a cash allocation seems to be the prefered choice (e.g., Faber, 2007, 2009; Kilgallen, 2012; Clare et al., 2016). Due to the unlimited loss potential of a short position, a holding period of one month bears – compared to a daily trading interval – an extraordinary high risk potential. Furthermore, Barberis and Thaler (2003) emphasize that short selling is not allowed for a large fraction of money managers (especially pension funds and mutual fund managers).

Exhibit 1: Alternative parameterization of technical trading strategies

Study	Parameterization
	Panel A: Momentum Strategies
Moskowitz et al. (2012)	1, 3, 6, 9, 12, 24, 36, 48 months
Neely et al. (2014)	9, 12 months
Baetje and Menkhoff (2016)	6,12 months
Hammerschmid and Lohre (2018)	1, 3, 6, 9, 12 months
	Panel B: Moving Average Strategies
Faber (2007, 2009)	Short MA: 1 month
	Long MA: 10 months (robustness check: 6, 8, 12, 14 months)
Kilgallen (2012)	Short MA: 1 month
	Long MA: 11 months (robustness check: 7, 9, 13 months)
Glabadanidis (2014)	Short MA: 1 month
	Long MA: 24 months (robustness check: 6, 12, 36, 48, 60 months)
Neely et al. (2014)	Short MA: 1, 2, 3 months
	Long MA: 9, 12 months
Baetje and Menkhoff (2016)	See Neely et al. (2014)
Hammerschmid and Lohre (2018)	See Neely et al. (2014)
Clare et al. (2016)	Short MA: 1 month
	Long MA: 6, 8, 10, 12 months
	Panel C: Volume-based Strategies
Neely et al. (2014)	Short MA: 1, 2, 3 months
	Long MA: 9, 12 months
Baetje and Menkhoff (2016)	See Neely et al. (2014)
Hammerschmid and Lohre (2018)	See Neely et al. (2014)

Notes: This table provides an overview of empirical studies that apply momentum strategies, moving average strategies, and volume-based strategies (column 1). The parameterizations of the technical indicator models are in column 2.

Following some of the studies listed in Exhibit 1, the short index for the moving average is set to $s = 1{,}2{,}3$ and the long index to $l = 9{,}12$, which results in six moving average strategies. We also set $s = 1$ and $l = 10, 24, 36, 48$ to cover some other popular parameterizations. We thus have ten different moving average strategies, labeled as MA*s*-*l* (e.g., MA1-10).

The time series momentum indicator is calculated as the difference between the actual price (P_t) and the *m*-month lagged price (P_{t-m}):

(2) $$S_{i,t} = \begin{cases} 1 & if \;\; P_t \geq P_{t-m} \\ 0 & if \;\; P_t < P_{t-m} \end{cases}$$

In this way, we invest in the S&P 500 index if we have a positive time series momentum $\left(S_{i,t} = 1\right)$. Otherwise, an allocation in the cash market is taken $\left(S_{i,t} = 0\right)$.[13] Following Moskowitz et al. (2012) we set $m = 1, 3, 6, 9, 12, 24, 36, 48$. The resulting eight momentum strategies are labeled as MOMm (e.g., MOM9).

In order to use the information contained in the volume data, we follow Granville's (1963) method, and implement a technical indicator based on the "on-balance" volume, which is defined as:

(3a) $$OBV_t = \sum_{k=1}^{t} VOL_k \, D_k \qquad with \; D_k = \begin{cases} 1 & if \;\; P_k \geq P_{k-1} \\ -1 & if \;\; P_k < P_{k-1} \end{cases} .$$

The trading signals for the OBV indicator thus result from a comparison of a short and a long moving average calculated on the OBV_t values, as follows:

[13] While Moskowitz et al. (2012) implement the momentum strategy with a short position, we follow Zakamulin (2014) and Marshall et al. (2017) and substitute the short position with an allocation in cash. Thus, we implement the momentum strategies in the same way as our seasonality strategies, the moving average strategies, as well as our fundamental-based predictive regression models. Marshall et al. (2017) additionally show the performance effects when the cash position is substituted with short sales in momentum and moving average strategies.

$$(3b) \quad S_{i,t} = \begin{cases} 1 & if \ MA_{s,t}^{OBV} \geq MA_{l,t}^{OBV} \\ 0 & if \ MA_{s,t}^{OBV} < MA_{l,t}^{OBV} \end{cases} \qquad with\ s < l$$

Along with some of the studies listed in Exhibit 1, the short index for the moving average is set to $s = 1,2,3$ and the long-index to $l = 9,12$, resulting in six volume-based strategies labeled as VOL*s*-*l* (e.g., VOL1-9).

3.3 Investment strategies based on fundamental factors

For the fundamental-based trading strategies, we choose the conventional framework, and estimate the following bivariate regression model (e.g., Welch and Goyal, 2008):

$$(4a) \quad r_{t+1}^{exc} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{i,t+1} \qquad (4b) \quad \hat{r}_{t+1}^{exc} = \hat{\alpha}_i + \hat{\beta}_i x_{i,t}$$

In Equation (4a), r_{t+1}^{exc} represents the log return on the S&P 500 index (including all dividends) in excess of the log risk-free rate from period *t* to *t*+1. $x_{i,t}$ is a predictor variable, and α_i and β_i are regression parameters that can be estimated using the OLS method. $\varepsilon_{i,t+1}$ denotes the regression residuum. Once the regression parameters are estimated, they can be used together with an observed value of the predictor variable to forecast the excess return (Equation 4b). As predictor variables, Welch and Goyal's (2008) updated data are used:

1. DP: Dividend price ratio (log), calculated as the log of a twelve-month moving sum of dividends paid on the S&P 500 index minus the log of stock prices (S&P 500 index).
2. DY: Dividend yield (log), calculated as the log of a twelve-month moving sum of dividends paid on the S&P 500 index minus the log of lagged stock prices (S&P 500 index).
3. EP: Earnings-price ratio (log), calculated as the log of a twelve-month moving sum of earnings on the S&P 500 index minus the log of stock prices (S&P 500 index).

4. DE: Dividend payout ratio (log), calculated as the log of a twelve-month moving sum of dividends minus the log of a twelve-month moving sum of earnings.
5. RVOL: Volatility of the equity risk premium, based on a twelve-month moving standard deviation estimator (Mele, 2007).
6. BM: Book-to-market value ratio for the Dow Jones Industrial Average.
7. NTIS: Net equity expansion, the ratio of a twelve-month moving sum of net equity issues by New York Stock Exchange (NYSE)-listed stocks to total year-end market capitalization of NYSE stocks.
8. TBL: Interest rate on a three-month Treasury bill.
9. LTY: Long-term government bond yield.
10. LTR: Return on long-term government bonds.
11. TMS: Term spread, computed as the long-term yield minus the Treasury bill rate.
12. DFY: Default yield spread, computed as the difference between Moody's BAA- and AAA-rated corporate bond yields.
13. DFR: Long-term corporate bond return minus long-term government bond return.
14. INFL: Calculated from the CPI for all urban consumers, with a publication lag of one month.

Rapach et al. (2010) and Rapach and Zhou (2013) show that the forecast accuracy of simple bivariate predictive regression models can be enhanced when their forecasts are combined. Apart from several more sophisticated forecast combination methods, the simple average of all single forecasts performs surprisingly well (Rapach and Zhou, 2013). However, one of the shortcomings of the simple bivariate predictive regression model is that potential interdependencies between various forecast variables are not considered.[14]

One way to overcome this may be to use the multivariate "kitchen sink" regression model, which, as its name suggests, incorporates all available forecast variables simultaneously. While this model often performs poorly in terms of a mean squared forecast error due to an overparameterization (e.g., Welch and Goyal, 2008; Rapach et al., 2010), it can still potentially generate profits (Rapach and Zhou, 2013).

One promising approach to avoid overparameterization is the diffusion index approach, where all available forecast variables are aggregated into a relatively small number of diffusion indices using a principal component method. Several studies have reported positive results in terms of equity risk premium forecasts when the predictive regressions are implemented in this way (e.g., Ludvigson and Ng, 2007; Rapach and Zhou, 2013; Neely et al., 2014).

Due to the positive results reported, the simple average-based forecast combination (FC), the kitchen sink forecast model (KSF), and two predictive regressions based on diffusion indices are also considered here. To avoid overparameterization, we follow Rapach and Zhou (2013) and implement the diffusion index forecast approach with one principal component (PC1F). Additionally, a second predictive regression model is considered, which is based on the first two principal components (PC2F). Together with the fourteen bivariate regression models, we have a total of eighteen forecast models based on fundamental variables.

[14] Baur (2013) discusses this issue in more detail in his factor analysis of gold returns.

In order to evaluate the economic profit of the predictive regression models, the forecasted excess returns of the S&P 500 total return index are transformed into a trading strategy, as follows (e.g., Pesaran and Timmermann, 1995; Hong et al., 2007):

$$S_{i,t} = \begin{cases} 1 & if \ \hat{r}_{t+1}^{exc} \geq 0 \\ 0 & if \ \hat{r}_{t+1}^{exc} < 0 \end{cases} \tag{5}$$

This means that we invest in the S&P 500 if the predicted excess return is positive $\left(S_{i,t} = 1\right)$. Otherwise, we allocate to cash $\left(S_{i,t} = 0\right)$.

4. A data-snooping-resistant simulation of investment strategies

One drawback of many traditional backtest studies that perform their trading strategies on a single historical return path is that the result may be purely from chance, and not due to any genuine merit (Sullivan et al., 1999, 2001). In order to avoid problems from data-snooping in statistical inferences that consider only surviving rules, White (2000) proposes a "Reality Check" (RC) method, which allows for testing the possible superior performance of certain rules. The basic idea behind this test is to compare trading strategies not only against a benchmark, but also to draw statistical inferences from an empirical distribution of a performance measure by considering the full universe of strategies, from which the best is selected. However, a shortcoming of this method is that its power is rapidly reduced if many poor and irrelevant models are included.

Hansen (2005) develops a similar test with several refinements in order to improve the test power in most cases.[15] His "superior predictive ability" (SPA) test is based on real valued loss functions, and compares the performance of two or more forecasting models. When evaluating trading strategies, an adequate loss function $L_{k,t}$ can be defined for model k as the negative

[15] The subsequent description of Hansen's (2005) SPA test comes primarily from Dichtl and Drobetz (2014).

continuously compounded return at time t so that $L_{k,t} = -r_{k,t}$ (Hansen, 2005). The model forecasts are evaluated in terms of their expected losses, $E[L_k]$, measured as the mean values in the sample from $t = 1, \dots, n$ (here, the mean negative return). When testing for SPA, the relevant question is whether any (or, equivalently, the best) of the $k = 1, \dots, m$ forecasting models is better than the benchmark ($k = 0$). Therefore, the loss values are transformed into relative performance values, labeled $d_{k,t} = L_{0,t} - L_{k,t}$, with $k = 1, \dots, m$ and $t = 1, \dots, n$.

This issue can be addressed by testing the null hypothesis that the benchmark is not inferior to any alternative forecasting model: H_0: $\mu_k = \mathrm{E}\left[d_{k,t}\right] \leq 0$ for all $k = 1, \dots, m$. Defining a stacked m-dimensional vector of expected excess performances $\boldsymbol{\mu}$, the null hypothesis is thus: H_0: $\boldsymbol{\mu} \leq \mathbf{0}$. The SPA test uses the following studentized test statistic:

(6) $$T_n^{SPA} = \max\left[\max_{k=1,\dots,m} \frac{n^{1/2}\bar{d}_k}{\widehat{\omega}_k}, 0\right]$$

where $\bar{d}_k = n^{-1}\sum_{t=1}^{n} d_{k,t}$ (the average relative performance of model k), and $\widehat{\omega}_k^2$ is some consistent estimator for $\omega_k^2 \equiv var\left(n^{1/2}\bar{d}_k\right)$.[16]

Hansen (2005) also proposes a bootstrap simulation approach to generate a null distribution based on $N_m\left(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Omega}}\right)$, where $\widehat{\boldsymbol{\mu}}$ is an appropriate estimator for $\boldsymbol{\mu}$, and $\widehat{\boldsymbol{\Omega}}$ is a consistent estimate of the asymptotic $m \times m$ covariance matrix $\boldsymbol{\Omega}$ of average relative performances. In particular, for $k = 1, \dots, m$, the estimator for the expected excess performance $\widehat{\boldsymbol{\mu}}$ is specified as $\hat{\mu}_k = \bar{d}_k \mathbb{1}_{\{n^{1/2}\bar{d}_k/\widehat{\omega}_k \leq -\sqrt{2\log\log n}\}}$, where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. In order to obtain the distribution of the SPA test statistic under the null hypothesis, we follow Hansen (2005) and implement the stationary bootstrap simulation approach of Politis and Romano (1994). Their simulation technique involves combining blocks with random lengths. The block length is chosen

[16] This estimator can be computed as in Hansen (2005, p. 372).

to be geometrically distributed with the parameter $q \in (0,1]$, thus resulting in a mean block length of q^{-1}.

Note that $b = 1, \dots, B$ resamples are generated from the bootstrap population $\boldsymbol{d}_t = (d_{1,t}, \dots, d_{m,t})'$, with $t = 1, \dots, n$. These bootstrap variables are recentered around $\hat{\boldsymbol{\mu}}$ as follows: $Z^*_{k,b,t} = d^*_{k,b,t} - g(\bar{d}_k)$ for $b = 1, \dots, B$ and $t = 1, \dots, n$, where g(x) is:

(7) $$g(x) = x \cdot \mathbb{1}_{\left\{x \geq -\sqrt{(\widehat{\omega}_k^2/n) 2 \log\log n}\right\}}.$$

For each resample $b = 1, \dots, B$, $T^{SPA*}_{b,n} = \max\left\{\max_{k=1,\dots,m}\left[n^{1/2}\bar{Z}^*_{k,b}/\widehat{\omega}_k\right], 0\right\}$ is calculated, where $\bar{Z}^*_{k,b} = n^{-1}\sum_{t=1}^{n} Z^*_{k,b,t}$ for all $k = 1, \dots, m$. A consistent estimate of the *p*-value is:

(8) $$\hat{p}_{SPA} = \sum_{b=1}^{B} \frac{\mathbb{1}_{\{T^{SPA*}_{b,n} > T^{SPA}_{n}\}}}{B},$$

where the null hypothesis of the SPA test that the benchmark model is the best forecasting model is rejected for small *p*-values.

In addition to this consistent *p*-value, Hansen (2005) shows that both an upper and a lower bound can be obtained. The upper bound is the *p*-value of a conservative test, whose null hypothesis is that all competing models are exactly as good as the benchmark model. For this test, $\hat{\mu}^u_k = 0$, and the corresponding recentering function for the bootstrap variables is $g^u(x) = x$.

In contrast, the lower bound is the *p*-value of a liberal test that assumes the models with worse performance than the benchmark are poor models in the limit. For this test, $\hat{\mu}^l_k = \min(\bar{d}_k, 0)$, and the appropriate recentering function for the bootstrap variables is $g^l(x) = \max(0, x)$. A large difference between the upper and lower bound *p*-value is an indication of many poor models.

Given the robustness of the SPA test regarding the inclusion of poor models, it is prefered to White's (2000) RC in this analysis. We do not include only a single investment strategy here, but many alternative trading models (see section 3). Following Sullivan et al. (1999, 2001), the buy-and-hold model is chosen as benchmark.[17]

For an assessment of the large number of monthly trading strategies, we use the negative continuously compounded return as an adequate loss function, i.e., $L_{k,t} = -r_{k,t}$. In order to incorporate the risk dimension, another loss function $L_{k,t} = -r_{k,t}^{exc}/\sigma_k$ is implemented, where $r_{k,t}^{exc}$ denotes the excess return above the risk-free rate (defined as a simple return), and σ_k is the volatility of the excess returns of strategy k.[18] The mean values (for $k = 1, \dots, m$) based on these loss functions in the bootstrap population exactly represent the (negative) Sharpe (1994) ratio.

Using Politis and White's (2004) test procedure and the corrections made in Patton et al. (2009), we set $q = 0.5$ in the stationary bootstrap approach.[19] In order to derive stable and robust results, it is conducted using 10,000 resamples.

5. Simulation results

5.1 Baseline simulations

In order to make our results comparable to various other studies (e.g., Neely et al., 2014; Baetje and Menkhoff, 2016), we perform the simulations on an updated data set from the Welch and Goyal (2008) study.[20] This data set comprises monthly data from the S&P 500 (inclusive dividends), the Treasury-bill rate, as well as all fundamental predictor variables discussed in section

[17] In a robustness check, the results are also compared with a 50:50 constant mix benchmark.

[18] Because the excess returns of the cash buy-and-hold strategy and the corresponding volatility are always zero, this strategy is deleted from the set of models when we implement simulations based on this loss function.

[19] Following Hansen (2005), we repeat the analysis with $q = 0.25$. The results are robust to this change.

[20] These data are available from Amit Goyal's webpage at http://www.hec.unil.ch/agoyal/.

3.3.[21] For the calculation of our volume-based indicators we use the monthly volume data from Yahoo Finance which are available since January 1950.[22]

Given our available data set, we follow Neely et al. (2014) as well as Baetje and Menkhoff (2016) and use the time span from 1950:01 to 1965:12 as our initial estimation period for the fundamental predictive regression models. Our baseline out-of-sample period then starts in 1966:01 and ends in 2014:12.

Various studies provide evidence that technical indicator models (especially the trend following moving average strategies) substantially reduce volatility and drawdowns against a passive buy-and-hold strategy (e.g., Clare et al., 2016; Faber, 2007, 2009; Kilgallen, 2012). In order to verify these findings (particularly in comparison with seasonality- and fundamental-based strategies), we determine an out-of-sample period which is characterized by large drawdowns in the S&P 500 stock market. By starting our "crisis sample period" in 2000:01, we have a comparatively short investment period which is suffered from two serious crises: the dot.com bubble from 2000-2002 and the 2008/2009 global financial markets crises (Hurst et al., 2017).[23]

Due to the availability of historical returns for the S&P 500 index and corresponding cash rates, we also implement a long-term backtest for our strategies based on seasonalities and technical indicators ranging from 1931:01 to 2014:12 as one of our robustness checks.[24]

---

[21] Technical indicators are often calculated with weekly or daily data. Due to the monthly availability of our fundamental/macroeconomic factors, we follow Neely et al. (2014) and compute the technical indicators also with monthly data. Furthermore, there are several studies available comparing technical indicator models implemented with daily data within a data-snooping-resistant test-framework (e.g., Sullivan et al., 1999; Hsu and Kuan, 2005; Hsu et al., 2010; Bajgrowicz and Scaillet, 2012).

[22] See http://de.finance.yahoo.com.

[23] When we start our "crisis subsample" before 2000 (e.g., in 1996 as motivated by the study from Doidge et al., 2017) we have the unwanted side effect that we dilute the large drawdowns during of the 2000-2002 and the 2008/2009 crisis with a preceding high positive equity risk premium. For example, from 1995:12 to 1999:12 the S&P 500 index (without dividends) raised by 138.5 percent (from 615.9 to 1,469.2 index points).

[24] In the Welch and Goyal (2008) data set, the monthly returns of the S&P 500 index (with dividends) starts in January 1926. While we document the results for the moving average strategies up to 48 months, we also tested a 60 months

Exhibit 2 reports descriptive statistics for the continuously compounded monthly returns of the S&P 500 total return index as well as the results of some simple weak-form market efficiency tests (i.e., autocorrelation tests and runs tests).

Exhibit 2: Descriptive statistics and simple tests for weak-form market efficiency

Panel A: Descriptive Statistics							
Period	Mean	Std.	Min	Max	Skewness	Kurtosis	Jarque-Bera
1966:01 – 2014:12	0.79	4.39	-24.31	15.54	-0.66	2.55	198.23***
2000:01 – 2014:12	0.36	4.44	-18.27	10.35	-0.74	1.40	29.64***
1931:01 – 2014:12	0.80	5.42	-33.89	34.66	-0.41	8.40	2,960.31***

Panel B: Autocorrelations							
Period	AC(1)	Q(1)	Q(3)	Q(6)	Q(12)	Q(24)	Q(36)
1966:01 – 2014:12	0.05	1.43	2.75	8.87	11.11	19.82	28.91
2000:01 – 2014:12	0.13	3.11*	6.42*	10.61*	13.49	32.58	48.04*
1931:01 – 2014:12	0.08	5.80**	14.73***	27.32***	32.92***	63.96***	71.75***

Panel C: Runs Test			
Period	Cutoff: Mean	Cutoff: Median	Cutoff: Zero
1966:01 – 2014:12	-0.71	-0.33	-0.57
2000:01 – 2014:12	-0.28	-0.60	-1.25
1931:01 – 2014:12	-0.60	-0.69	-0.79

Notes: This table reports descriptive statistics (panel A) and the results of simple tests for weak-form market efficiency (panels B and C) for the returns of the S&P 500 total return index during the three backtest periods: January 1966-December 2014, January 2000-December 2014, and January 1931-December 2014. All calculations are based on monthly log returns. The mean returns and standard deviations in panel A are not annualized. Kurtosis is adjusted so that the normal distribution exhibits a kurtosis of zero. The Jarque-Bera test statistic tests the null hypothesis that the stock returns are normally distributed. AC(1) in panel B denotes the autocorrelation in returns at lag 1. The statistical significance of autocorrelation is tested using the Ljung-Box test. Q(·) denotes the Q-test statistic testing for an autocorrelation up to the lag shown in brackets. Panel C lists the test statistic of the runs test, where the definition of a run is based on three different cutoff values: mean, median, and zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

moving average strategy as an additional robustness test. Therefore, we start our long-term simulations in January 1931.

The values in panel A show that the mean monthly return of the S&P 500 index from 2000-2014 is markedly lower than for the other backtest periods (0.36 versus 0.79 and 0.80). This observation is attributable to the 2000-2002 dot.com crisis, as well as the 2008/2009 global financial crisis. The monthly returns in all three backtest periods exhibit negative skewness and positive kurtosis, with the consequence that the normal distribution assumption must be rejected at a 1% significance level.

Panel B lists the autocorrelations at a lag of one (AC(1)) and the results of the Ljung-Box test in order to test for higher-order autocorrelations (Q(·)). We observe statistically significant autocorrelations for both the backtest periods beginning in 2000 and 1931, but not for the 1966 period. The runs test applied to the same three backtest periods provides a different picture, however. This test detects no statistically significant run in any of the three periods. The result holds whether the mean, the median, or zero is used as cutoff value for defining runs. Apart from the fact that both weak-form market efficiency tests provide different results, neither the test for autocorrelation nor the runs test explains whether the inefficiencies detected can be successfully exploited within a realistic investment strategy.

In a next step, we attempt to simulate all the investment strategies discussed in section 3 in a realistic way. In terms of transaction costs, we follow Lynch and Balduzzi (2000) and Blitz and van Vliet (2008), and assume 25 basis points as turnover-dependent costs. The fundamental regression models are estimated with a rolling window approach that consists of 180 observations.[25] This means that the parameter estimation for the first forecast in January 1966 depends on the observations from January 1951 through December 1965.

[25] The specific advantages of a rolling window approach (versus possibly an expanding window approach) are discussed in Giacomini and White (2006). The application of Hansen's (2005) SPA test depends on the stationary assumption that would be violated within an expanding window approach but not within a rolling window approach (see the discussion in Hansen, 2005).

The results of the investment strategies are measured in terms of mean absolute returns and risk-adjusted mean excess returns (Sharpe, 1994). In order to test the results for statistical significance, we thus apply Hansen's (2005) SPA test (as described in section 4).[26] Exhibit 3 lists the ten best investment strategies for the three types (monthly seasonality, technical indicator, and fundamental).

Exhibit 3: Traditional backtest results (1966:01–2014:12)

Panel A: Monthly Seasonality Models			
Models	Mean Return	Models	Sharpe Ratio
SEA759	0.8434	SEA2482	0.1283
SEA1530	0.8321	SEA3154	0.1257
SEA279	0.8302	SEA3484	0.1244
SEA69	0.8267	SEA1565	0.1238
SEA9	0.8209	SEA2279	0.1204
SEA1299	0.8082	SEA1530	0.1194
SEA1565	0.8049	SEA759	0.1170
SEA269	0.7982	SEA2741	0.1166
SEA2244	0.7969	SEA2244	0.1153
SEA265	0.7958	SEA779	0.1142

Panel B: Technical Indicator Models			
Models	Mean Return	Models	Sharpe Ratio
VOL2-12	0.9002	VOL2-12	0.1503
VOL2-9	0.8847	VOL2-9	0.1477
VOL1-12	0.8846	VOL1-12	0.1451
MA2-12	0.8714	MA2-12	0.1401
VOL3-12	0.8696	VOL3-12	0.1386
MA1-12	0.8542	MA1-12	0.1346
MA3-9	0.8209	VOL1-9	0.1233

[26] Hansen's (2005) SPA test is conducted using the MULCOM 3.0 toolbox provided by P.R. Hansen and A. Lunde, combined with the Ox 7.0 programming language (see Doornik, 2009). All other simulations and calculations are coded with the free programming language Python.

VOL1-9	0.8187	MA3-9	0.1231
MOM12	0.8131	MA1-10	0.1200
MA2-9	0.8062	MA2-9	0.1200

Panel C: Fundamental Models

Models	Mean Return	Models	Sharpe Ratio
FC	0.8538	DE	0.1187
DE	0.8448	FC	0.1114
TMS	0.8342	TMS	0.1085
LTR	0.8073	TBL	0.1082
PC1F	0.8066	PC2F	0.1061
PC2F	0.8050	LTR	0.0997
TBL	0.7662	KSF	0.0917
NTIS	0.7567	PC1F	0.0903
DFY	0.7427	NTIS	0.0901
KSF	0.7088	DFY	0.0855

Panel D: Benchmark Models

Models	Mean Return	Models	Sharpe Ratio
Buy-and-Hold	0.7894	Buy-and-Hold	0.0837

Notes: This table reports the results for the traditional backtests during the January 1966-December 2014 period for the monthly seasonality models (panel A), the technical indicator models (panel B), the fundamental models (panel C), and the buy-and-hold benchmark strategy (panel D). All strategies are evaluated in terms of mean monthly absolute return (column 2) and Sharpe ratio (column 4), where the Sharpe ratio is defined as the mean of monthly excess returns divided by their standard deviations (Sharpe, 1994). The mean returns and Sharpe ratios are not annualized. Panels A, B, and C list the ten best models for each category of investment strategies in terms of mean absolute return and Sharpe ratio in descending order. The monthly seasonality models are described in section 3.1 and Appendix A1, the technical indicator models are described in section 3.2, and the fundamental models are in section 3.3.

Note that the ten monthly seasonality strategies with the highest mean monthly returns (panel A) all provide higher absolute returns than the buy-and-hold strategy (panel D). For example, the monthly seasonality strategy SEA759 exhibits a mean monthly return of 0.8434%, compared to 0.7894% with a buy-and-hold strategy. Interestingly, the three models with the highest mean monthly returns (SEA759, SEA1530, and SEA279) exhibit similar allocations, which are also comparable to that of the Halloween strategy (see Appendix A1). While the Halloween strategy invests in the cash market from May to October, the SEA759 model is allocated to cash from June to September, and the SEA1530 model from May to September. The SEA279 model (the model with the third highest mean return) is invested predominantly in the stock market, except for the months of July, August, and September. When examining the Sharpe ratio, we observe that the SEA2482 model (the Halloween strategy) dominates with a Sharpe ratio of 0.1283. Similar to the allocation of the Halloween strategy are the monthly allocations of the second and third best models in terms of the Sharpe ratio. The SEA3154 strategy is also invested in the cash market from May to October, but features an additional cash market allocation in December and February. In contrast to this model, the model with the third best Sharpe ratio (SEA3484) is invested in the cash market in January and February.

The performance values of the technical indicator models are listed in panel B of Exhibit 3. A comparison with the monthly seasonality models shows a clear superiority of the technical analysis models. This holds for the mean return and the Sharpe ratio criteria. Logically, all the technical models would also dominate the buy-and-hold strategy in terms of both criteria. But, within this group, those based on the volume indicator provide the highest absolute and risk-adjusted returns (Sharpe ratio).

The three best models in terms of the mean return and the Sharpe ratio are VOL2-12, VOL2-9, and VOL1-12. Within the group of eight best models (sorted by either absolute returns or risk-adjusted returns), we only observe those based on the volume indicator or the moving average trading rule. And it seems that both indicators provide better results than the simple momentum trading strategies. Only the momentum strategy based on the last twelve months (MOM12) is listed among the ten strategies with the highest mean monthly returns (in the ninth position). Moreover, it is surprising to observe that historical prices up to twelve months are obviously more important than prior prices. Strategies based on historical index levels prior to twelve months (e.g., MOM24, MOM36, MA1-36, MA1-48) do not appear on the top-ten list.

Panel C of Exhibit 3 shows the performance of the fundamental models. A comparison with the technical models shows they clearly dominate. For example, in terms of mean return, all six technical models with the highest yields dominate the best fundamental model (using a forecast combination approach (FC) with a monthly mean return of 0.8538). Furthermore, the best fundamental model that assesses the dividend payout ratio (DE) exhibits a lower Sharpe ratio (0.1187) than all ten listed technical models.

For the monthly seasonality and technical analysis models, we observe a clear dominance of the ten best models over the buy-and-hold strategy (for both the mean return and the Sharpe ratio). However, this result does not hold for the fundamental models. When examining the mean returns, only the FC, DE, TMS, LTR, PC1F, and PC2F models provide a higher mean return than the buy-and-hold.

While the results presented in Exhibit 3 offer a good overview of the pros and cons of the different models, these analyses do not account for the data-snooping problem (Lo and MacKinlay, 1990). In order to address this issue, we apply Hansen's (2005) SPA test (as presented in section

4) to all three model clusters. In each test, the corresponding models are compared with the performance of the buy-and-hold strategy. Thus, the first test considers all 4,095 different monthly seasonal allocation models, the second test all 24 technical indicator models, and the third test all 18 fundamental models.[27] The results are in Exhibit 4.

Exhibit 4: Tests for superior predictive ability (1966:01 – 2014:12)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Models	Consistent p-Value Lower p-Value Upper p-Value		Model	Loss Value	t-Statistic	p-Value
		Panel A: Loss Function $L_{k,t} = -r_{k,t}$				
Monthly	0.9737	Benchmark model	B&H	-0.0079		
Seasonality	0.7629	Most significant model	SEA0009	-0.0082	0.5709	0.2772
Models	0.9892	Best model	SEA0759	-0.0084	0.5041	0.2957
Technical	0.5089	Benchmark model	B&H	-0.0079		
Indicator	0.3802	Most significant model	VOL2-12	-0.0090	0.8438	0.1914
Models	0.5089	Best model	VOL2-12	-0.0090	0.8438	0.1914
Fundamental	0.8002	Benchmark model	B&H	-0.0079		
Models	0.6277	Most significant model	FC	-0.0085	0.7449	0.2188
	0.8114	Best model	FC	-0.0085	0.7449	0.2188
		Panel B: Loss Function $L_{k,t} = -r_{k,t}^{exc}/\sigma_k$				
Monthly	0.9292	Benchmark model	B&H	-0.0837		
Seasonality	0.6926	Most significant model	SEA2482	-0.1283	1.3692	0.0834
Models	0.9617	Best model	SEA2482	-0.1283	1.3692	0.0834
Technical	0.1157	Benchmark model	B&H	-0.0837		
Indicator	0.0900	Most significant model	VOL2-12	-0.1503	2.1095	0.0173
Models	0.1157	Best model	VOL2-12	-0.1503	2.1095	0.0173
Fundamental	0.5938	Benchmark model	B&H	-0.0837		
Models	0.4557	Most significant model	FC	-0.1114	1.3780	0.0843
	0.5938	Best model	DE	-0.1187	1.3403	0.0916

Notes: This table reports SPA p-values for the investment strategies based on monthly seasonalities, technical indicators, and fundamental factors compared to the buy-and-hold benchmark (B&H) in the 1966:01-2014:12 backtest period. Panel A uses a loss function based on negative continuously compounded absolute returns; panel B models the loss function as the negative risk-adjusted excess return. Column (2) shows the consistent p-value of the SPA test, as well as the lower and upper bounds for p-values. The table also reports the sample loss for the buy-and-hold benchmark and the two investment strategies of each category that have the smallest sample loss value and the largest t-statistic for average relative performance $(\bar{d}_k)$. These two models are referred to as the "best" and "most significant" model, respectively. The loss value is shown in column (5), and the corresponding t-statistic (of their sample loss relative to the benchmark) is given in column (6). Finally, column (7) reports the "p-values" from the pairwise comparisons of "best" and "largest t-statistic" models with the benchmark. These p-values (unlike the SPA p-value) ignore the search over all models that preceded the selection of the model being compared to the benchmark, i.e., they do not account for the entire universe of models.

[27] Note that, overall, we have 4,096 monthly seasonal models (section 3.1). However, the SEA0 model represents the buy-and-hold benchmark strategy.

Panel A of Exhibit 4 shows the results for a loss function based on continuously compounded absolute returns, and panel B shows those for a loss function based on risk-adjusted excess returns. Column (2) gives the consistent *p*-value and the lower and upper bound *p*-values of the SPA test. Column (5) gives the loss values of the benchmark model (buy-and-hold), the most significant model (the model with the highest *t*-statistic), and the best model (the model with the lowest loss value). The corresponding *t*-statistics for the most significant and the best models are in column (6).

Note that the *p*-values in column (7) result from a pairwise comparison of the best and the most significant models with the benchmark. In contrast to the *p*-values of the SPA test reported in column (2), these *p*-values do not account for the entire universe of all models. The consistent *p*-value of 0.9737 in panel A shows that the null hypothesis of the SPA test cannot be rejected, i.e., there is no statistically significant evidence that any monthly seasonality model is better than the buy-and-hold benchmark. This result holds even for the liberal test with a lower bound *p*-value of 0.7629.

Interestingly, the most significant and the best model are not identical. The most significant model, SEA9, represents a monthly allocation model that holds a cash allocation in September, but invests in the stock market in all other months (see Appendix A1). Obviously, an explicit consideration of the September swoon effect (Haug and Hirschey, 2011) would seem to be good advice for a monthly trading strategy. The SEA759 model exhibits the minimal loss value (in other words, it exhibits the highest mean monthly return (see Exhibit 3)). This model is allocated to cash from June to September, and to stocks from October to May.

The second row in panel A of Exhibit 4 reports the results for the technical indicator models. The consistent and lower bound *p*-values in column (2) demonstrate that no technical indicator model significantly outperforms the buy-and-hold benchmark (consistent *p*-value of 0.5089 and lower bound *p*-value of 0.3802). Within this model group, the best model VOL2-12 is also the most significant one. However, the *t*-statistic of 0.8438 is not significant even when we do not consider the entire universe of all technical models (*p*-value of 0.1914).

The third row in panel A of Exhibit 4 lists the findings for the fundamental investment models. Within this model cluster, we also observe that there is no single model that leads to a rejection of the null hypothesis of Hansen's (2005) SPA test. The forecast combination approach (FC) is not only the most significant model, it is also the best one. However, this model is also not significant when we don't take all the other fundamental models into account as well (a *p*-value of 0.2188 in column (7)).

Panel B in Exhibit 4 provides the results for the risk-adjusted excess returns. Note that the "Sell in May and Go Away" model (SEA2482) represents the model with the highest Sharpe ratio and the highest degree of significance. The *t*-statistic, with a value of 1.3692, is significant at a 10% level (*p*-value 0.0834) if we do not consider the full universe of monthly seasonal models. If all models are taken into consideration, the SPA test cannot reject the null hypothesis (consistent *p*-value of 0.9292 and a lower bound *p*-value of 0.6926). In other words, the Halloween strategy is not statistically superior to the buy-and-hold benchmark.

The VOL2-12 model (second row in panel B) is the best model, both in terms of absolute returns and risk-adjusted excess returns. With a *t*-statistic of 2.1095, we observe a high significance (*p*-value of 0.0173) when the full universe of technical analysis models is not considered. When we include the entire universe, we note that the *t*-statistic remains nearly significant at a 10% level within the SPA test (a consistent *p*-value of 0.1157). A lower bound *p*-value of 0.09 even indicates a statistically significant rejection of the null hypothesis at a 10% level when the SPA test is conducted in a more liberal way. Given that Hansen's SPA test is a very conservative test (e.g., Hsu et al., 2014), a look at the lower bound p-value seems quite sensible. In this respect, the VOL2-12 model statistically significantly dominates the buy-and-hold benchmark.

The last row in panel B lists the results for the fundamental models. The DE model (which predicts the excess return of the S&P 500 by means of the dividend payout ratio) represents the best model in terms of the Sharpe ratio. However, the FC approach is the most significant without considering the entire universe of fundamental models (p-value of 0.0843 in column (7)). But this result does not hold when we consider all other fundamental models. The SPA test provides a consistent p-value of 0.5938 and a lower bound p-value of 0.4557. Overall, we only observe a statistically significant dominance, in the sense of Hansen's (2005) SPA test, for the technical analysis model VOL2-12 compared to the buy-and-hold benchmark when considering risk-adjusted excess returns.

In a next step, the analysis is repeated for our "crisis period" ranging from 2000:01-2014:12. Because this period comprises the dot.com bubble and the 2008/2009 global financial crisis, it poses a real challenge to forecast-based investment strategies.

Exhibit 5 shows the five best models in terms of mean monthly returns and Sharpe ratios for each of the three categories of investment strategies.

Exhibit 5: Traditional backtest results (2000:01 – 2014:12)

Panel A: Monthly Seasonality Models			
Models	Mean Return	Models	Sharpe Ratio
SEA1761	0.6258	SEA3484	0.1573
SEA2706	0.6198	SEA2706	0.1567
SEA325	0.6086	SEA3909	0.1514

SEA884	0.6033	SEA1761	0.1486
SEA860	0.6026	SEA2741	0.1444

Panel B: Technical Indicator Models			
Models	Mean Return	Models	Sharpe Ratio
MA1-9	0.7430	MA1-9	0.2325
VOL3-12	0.7106	VOL3-12	0.2196
MA1-10	0.7010	MA1-10	0.2094
MOM6	0.6913	MA1-12	0.2054
MA1-12	0.6762	MOM6	0.2034

Panel C: Fundamental Models			
Models	Mean Return	Models	Sharpe Ratio
KSF	0.5939	KSF	0.1305
DFY	0.4883	DFY	0.0977
DE	0.4832	EP	0.0902
EP	0.4700	DE	0.0859
FC	0.4251	FC	0.0700

Panel D: Benchmark Models			
Models	Mean Return	Models	Sharpe Ratio
Buy-and-Hold	0.3582	Buy-and-Hold	0.04560

Notes: This table reports the results of the traditional backtests in the 2000:01-2014:12 period for the monthly seasonality models (panel A), the technical indicator models (panel B), the fundamental models (panel C), and the buy-and-hold benchmark strategy (panel D). All strategies are evaluated in terms of the mean monthly absolute return (column 2) and the Sharpe ratio (column 4), where the Sharpe ratio is defined as the mean of the monthly excess returns divided by their standard deviations (Sharpe, 1994). The mean returns and the Sharpe ratios are not annualized. Panels A, B, and C list the five best models for each category of investment strategies in terms of mean absolute returns and Sharpe ratios in descending order. The monthly seasonality models are described in section 3.1 and Appendix A1, the technical indicator models in section 3.2, and the fundamental models in section 3.3.

The five monthly seasonality strategies with the highest mean monthly returns all provide a higher absolute return than the buy-and-hold strategy. The fifth best model, SEA860, still provides a 0.6026% mean monthly return, but the buy-and-hold strategy only exhibits a 0.3582% return over the same time frame. However, a comparison with Exhibit 3 shows that the five best models listed in Exhibit 5 are not identical to the best models in the 1966:01-2014:12 time frame. The five best seasonality models in terms of mean monthly return all invest in the cash market in January, February, June, and September, and in the stock market in March, April, October, November, and December. The similar allocations of these five models differ more or less clearly from the allocation of the Halloween strategy.

The same holds when we measure success in terms of risk-adjusted mean excess returns. While all five seasonal allocation models dominate the buy-and-hold strategy in terms of the Sharpe ratio, they differ clearly from the Halloween strategy (the dominating strategy in the 1966:01-2014:12 period) in terms of allocations.

Panel B in Exhibit 5 shows the performance of the five best technical indicator models. Just as in the longer 1966-2014 backtest period, the technical indicator models clearly dominate the monthly seasonal models in terms of absolute and risk-adjusted returns. Logically, all technical models are also superior to the buy-and-hold strategy with respect to these two performance measures. Interestingly, the best technical indicator model type differs between the two backtest periods. In the longer one (1966 to 2014), volume-based technical indicators dominate; in the shorter one (2000-2014), the moving average strategies dominate. And, again, historical stock market prices up to twelve months seem to be the most important. Strategies based on historical index levels prior to twelve months (e.g., MOM24, MOM36, MA1-36, MA1-48) do not appear on the top-five list.[28]

Panel C contains the results for the fundamental models. We observe that the five best models in terms of mean monthly returns and Sharpe ratios are clearly better than the buy-and-hold benchmark. And, just as in the longer backtest period, the fundamental models are consistently dominated by the monthly seasonal, and particularly by the technical indicator models. In contrast to the preceding backtest, however, now the kitchen sink forecast model (KSF) provides the best results in terms of mean monthly return and Sharpe ratio. Regressions based on the dividend payout ratio (DE) also provide good results here (the third best model in terms of mean return, and the fourth best model in terms of Sharpe ratio).

In order to test the statistical significance of the results, again Hansen's (2005) SPA test is applied.

[28] Exhibit 5 lists only the five best models. However, this evidence also holds for the ten best models.

Exhibit 6: Tests for superior predictive ability (2000:01 – 2014:12)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Models	Consistent p-Value Lower p-Value Upper p-Value		Model	Loss Value	t-Statistic	p-Value
		Panel A: Loss Function $L_{k,t} = -r_{k,t}$				
Monthly	0.8612	Benchmark model	B&H	-0.0036		
Seasonality	0.7015	Most significant model	SEA0325	-0.0061	1.2076	0.1177
Models	0.8767	Best model	SEA1761	-0.0063	1.0809	0.1470
Technical	0.2508	Benchmark model	B&H	-0.0036		
Indicator	0.2073	Most significant model	MA1-9	-0.0074	1.3627	0.0921
Models	0.2508	Best model	MA1-9	-0.0074	1.3627	0.0921
Fundamental	0.7247	Benchmark model	B&H	-0.0036		
Models	0.5050	Most significant model	KSF	-0.0059	1.0526	0.1470
	0.7363	Best model	KSF	-0.0059	1.0526	0.1470
		Panel B: Loss Function $L_{k,t} = -r_{k,t}^{exc}/\sigma_k$				
Monthly	0.8216	Benchmark model	B&H	-0.0460		
Seasonality	0.6491	Most significant model	SEA2706	-0.1567	1.7839	0.0420
Models	0.8422	Best model	SEA3484	-0.1573	1.5583	0.0642
Technical	0.0244	Benchmark model	B&H	-0.0460		
Indicator	0.0194	Most significant model	MA1-9	-0.2325	2.8073	0.0050
Models	0.0244	Best model	MA1-9	-0.2325	2.8073	0.0050
Fundamental	0.3876	Benchmark model	B&H	-0.0460		
Models	0.2492	Most significant model	KSF	-0.1306	1.6498	0.0527
	0.4004	Best model	KSF	-0.1306	1.6498	0.0527

Notes: This table reports SPA p-values for the investment strategies based on monthly seasonalities, technical indicators, and fundamental factors compared to the buy-and-hold benchmark (B&H) in the 2000:01-2014:12 backtest period. Panel A uses a loss function based on negative continuously compounded absolute returns; panel B models the loss function as the negative risk-adjusted excess return. Column (2) shows the consistent p-value of the SPA test, as well as the lower and upper bounds for the p-values. The table also reports the sample loss for the buy-and-hold benchmark and the two investment strategies in each category that exhibit the smallest sample loss value and the largest t-statistic for average relative performance $(\bar{d}_k)$. These two models are referred to as the "best" and "most significant" models, respectively. The loss value is shown in column (5), and the corresponding t-statistic (of their sample loss relative to the benchmark) is in column (6). Finally, column (7) reports the "p-values" from the pairwise comparisons of "best" and "largest t-statistic" models with the benchmark. These p-values (unlike the SPA p-value) ignore the search over all models that preceded the selection of the model being compared to the benchmark, i.e., they do not account for the entire universe of models.

In panel A of Exhibit 6, we observe that the best monthly seasonal model, SEA1761, differs from the most significant model (SEA325). The consistent and lower bound p-values confirm that the null hypothesis cannot be rejected. Among the technical indicator models, MA1-9 is the best and the most significant (p-value of 0.0921 and significant at a 10% level), but without considering all the others. When we include the other technical indicator models, this model is no longer significantly superior (consistent p-value of 0.2508).

Within the fundamental models, the best model, KSF, is also the most significant one. As one would expect, the null hypothesis of the SPA test cannot be rejected. Panel B in Exhibit 6 provides the results of the SPA test for risk-adjusted excess returns. In contrast to the monthly seasonality and fundamental models, we note a statistically significant rejection of the null hypothesis within the technical indicator model group. The MA1-9 model dominates the buy-and-hold benchmark in terms of risk-adjusted excess returns at a 5% level (consistent *p*-value of 0.0244). This result also holds for the conservative variant of Hansen's SPA test (as documented by the upper bound *p*-value). Taken together, we only observe a statistically significant better model than the buy-and-hold within the cluster of technical indicator models. However, this result (observable for both backtest periods) only holds when performance is measured using risk-adjusted excess returns. Unfortunately, the two "best" models are not identical for both backtest periods (VOL2-12 for the longer backtest period, and MA1-9 for the shorter period).

5.2 Robustness tests

In order to validate the robustness of the results, a battery of tests is conducted. Due to the longer availability of historical price data for the S&P 500 index and cash market rates, we first conduct a long-term test for the monthly seasonality and technical indicator-based (momentum and moving average indicator) strategies.[29] Beginning from the initial estimation period for the technical indicator models (1926:01-1930:12), we now have a backtest period ranging from 1931:01 to 2014:12.[30]

[29] Historical volume data are not available for this longer backtest period.

[30] Due to data availability, we could begin the simulations in 1930:01. However, as an additional robustness check, the MA1-60 strategy is also tested, where an initial estimation period of five years is needed. This is why the simulations begin in 1931:01 and not 1930:01. However, because of poor performance, the MA1-60 strategy does not affect the results.

Exhibit 7: Traditional backtest results (1931:01 – 2014:12)

Panel A: Monthly Seasonality Models

Models	Mean Return	Models	Sharpe Ratio
SEA9	0.8550	SEA9	0.1104
SEA73	0.8202	SEA2063	0.1103
SEA54	0.8180	SEA73	0.1099
SEA30	0.8110	SEA54	0.1099
SEA69	0.8054	SEA258	0.1098

Panel B: Technical Indicator Models

Models	Mean Return	Models	Sharpe Ratio
MA2-12	0.8125	MA2-12	0.1475
MOM9	0.7948	MA1-10	0.1443
MA1-12	0.7846	MA2-9	0.1406
MA1-10	0.7827	MA1-9	0.1398
MA2-9	0.7754	MA1-12	0.1392

Panel C: Benchmark Models

Models	Mean Return	Models	Sharpe Ratio
Buy-and-Hold	0.8024	Buy-and-Hold	0.0941

Notes: This table reports the results of the traditional backtests in the 1931:01-2014:12 period for the monthly seasonality models (panel A), the technical indicator models (panel B), and the buy-and-hold benchmark strategy (panel C). All strategies are evaluated in terms of mean monthly absolute returns (column 2) and Sharpe ratios (column 4), where the Sharpe ratio is defined as the mean of the monthly excess returns divided by their standard deviations (Sharpe, 1994). The mean returns and the Sharpe ratios are not annualized. Panels A and B list the five best models for each category of investment strategies in terms of mean absolute return and Sharpe ratio in descending order. The monthly seasonality models are described in section 3.1 and Appendix A1, and the technical indicator models are in section 3.2.

Panel A in Exhibit 7 shows that the five monthly seasonality strategies with the highest mean returns all provide higher absolute returns than the buy-and-hold. While these five models differ from those listed in Exhibit 5 (backtest period from 2000:01-2014:14), the SEA9 and SEA69 models are also in the top-five for the 1966:01-2014:12 period. And, although it is not listed, the sixth best model in terms of mean return is the buy-and-hold benchmark model (SEA0).

Examining the Sharpe ratios, we observe a dominance of all five listed models compared to the buy-and-hold benchmark. The SEA9, SEA73, and SEA54 models are also superior in terms of mean returns. The strategy definition (Appendix A1) illustrates that these three seasonal strategies have similar allocations. While the SEA9 strategy only invests in the cash market in September, SEA73 additionally invests in October, and SEA54 in May. All of these strategies are predominantly invested in the stock market, and more or less exploit the September swoon effect.

The results for the technical indicator models are in panel B. For the mean return, only the MA2-12 model dominates the buy-and-hold benchmark. All other models provide lower mean absolute returns over the 1931:01-2014:12 period. Obviously, in the long run (approximated for with a long investment horizon), the buy-and-hold represents a strong absolute return benchmark. When we look at risk-adjusted returns, however, all listed moving average models clearly dominate the buy-and-hold benchmark. This dominance compared to the momentum models is observable not only for the Sharpe ratio, but also for the mean return criterion.

While Exhibit 7 only lists the five best models, the next five best models are also examined. And, again, neither a moving average model nor a momentum model included historical S&P 500 prices prior to twelve months. In order to validate the statistical significance of these results, again the SPA test is applied.

Exhibit 8: Tests for superior predictive ability (1931:01 – 2014:12)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Models	Consistent p-Value Lower p-Value Upper p-Value		Model	Loss Value	t-Statistic	p-Value
		Panel A: Loss Function $L_{k,t} = -r_{k,t}$				
Monthly	0.8882	Benchmark model	B&H	-0.0080		
Seasonality	0.4905	Most significant model	SEA9	-0.0086	0.9211	0.1774
Models	0.9578	Best model	SEA9	-0.0086	0.9211	0.1774
Technical	0.7556	Benchmark model	B&H	-0.0080		
Indicator	0.5727	Most significant model	MA2-12	-0.0081	0.0750	0.4712
Models	0.7894	Best model	MA2-12	-0.0081	0.0750	0.4712
		Panel B: Loss Function $L_{k,t} = -r_{k,t}^{exc}/\sigma_k$				
Monthly	0.8679	Benchmark model	B&H	-0.0941		
Seasonality	0.5674	Most significant model	SEA9	-0.1105	1.5394	0.0676
Models	0.9367	Best model	SEA9	-0.1105	1.5394	0.0676
Technical	0.1329	Benchmark model	B&H	-0.0941		
Indicator	0.0997	Most significant model	MA2-12	-0.1475	1.9697	0.0237
Models	0.1329	Best model	MA2-12	-0.1475	1.9697	0.0237

Notes: This table reports SPA p-values for the investment strategies based on monthly seasonalities and technical indicators compared to the buy-and-hold benchmark (B&H) in the 1931:01-2014:12 backtest period. Panel A uses a loss function that is based on negative continuously compounded absolute returns; panel B models the loss function as the negative risk-adjusted excess returns. Column (2) shows the consistent p-value of the SPA test, as well as the lower and upper bounds for p-values. The table also reports the sample loss for the buy-and-hold benchmark and the two investment strategies in each category that have the smallest sample loss value and the largest t-statistic for average relative performance $(\bar{d}_k)$. These two models are referred to as the "best" and "most significant" models, respectively. The loss value is shown in column (5), and the corresponding t-statistic (of their sample loss relative to the benchmark) is in column (6). Finally, column (7) reports the "p-values" from the pairwise comparisons of "best" and "largest t-statistic" models with the benchmark. These p-values (unlike the SPA p-value) ignore the search over all models that preceded the selection of the model being compared to the benchmark, i.e., they do not account for the entire universe of models.

Exhibit 8 shows that the best (and most significant) model, SEA9, cannot reject the null hypothesis of Hansen's (2005) SPA test. This result holds for the mean return (the first row in panel A), and for the risk-adjusted excess return (first row in panel B). However, for the best and most significant technical analysis model, MA2-12, we observe a statistically significant outperformance in terms of risk-adjusted excess returns when applying the SPA test in its liberal version (lower bound p-value of 0.0997, reported in the second row of panel B).[31]

[31] The consistent p-value is 0.1329, and so it is slightly too high compared to the 10% significance level.

As an interim result, it is noticeable that the technical indicator models dominate the monthly seasonality and fundamental models in all three backtest periods. However, while the volume-based technical models are superior in the 1966:01-2014:12 period, we observe a dominance of the technical indicator models based on historical prices in the other two periods. In order to verify the potential of the investment strategies that exploit historical prices in the 1966:01-2014:12 period in more detail, the simulations are repeated, but without the six volume indicators (the same set of technical indicator models used in the 1931:01-2014:12 backtest). The results are in panel A of Exhibit 9.

Exhibit 9: Further robustness tests

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Loss Function/ Period	Consistent p-Value Lower p-Value Upper p-Value		Model	Loss Value	t-Statistic	p-Value
Panel A: Technical Indicator Models Based on Historical Prices (1966:01 – 2014:12)						
$L_{k,t} = -r_{k,t}$	0.5876	Benchmark model	B&H	-0.0079		
	0.4157	Most significant model	MA2-12	-0.0087	0.6384	0.2552
	0.5876	Best model	MA2-12	-0.0087	0.6384	0.2552
$L_{k,t} = -r_{k,t}^{exc}/\sigma_k$	0.1714	Benchmark model	B&H	-0.0837		
	0.1302	Most significant model	MA2-12	-0.1401	1.8442	0.0332
	0.1714	Best model	MA2-12	-0.1401	1.8442	0.0332
Panel B: Technical Indicator Models Compared to a 50:50 Constant Mix Benchmark (in terms of risk-adjusted returns)						
1966:01 – 2014:12	0.2091	Benchmark model	50:50 CM	-0.0943		
	0.1452	Most significant model	VOL2-12	-0.1503	1.7726	0.0371
	0.2091	Best model	VOL2-12	-0.1503	1.7726	0.0371
2000:01 – 2014:12	0.0365	Benchmark model	50:50 CM	-0.0565		
	0.0276	Most significant model	MA1-9	-0.2325	2.6656	0.0068
	0.0365	Best model	MA1-9	-0.2325	2.6656	0.0068
1931:01 – 2014:12	0.3013	Benchmark model	50:50 CM	-0.1071		
	0.2254	Most significant model	MA2-12	-0.1475	1.4779	0.0702
	0.3013	Best model	MA2-12	-0.1475	1.4779	0.0702

Notes: This table reports SPA p-values for the investment strategies based on technical indicators compared to the buy-and-hold benchmark (B&H) and the 50:50 constant mix benchmark (CM). The 50:50 benchmark allocates 50% to the S&P 500 total return index, and 50% to cash, with monthly rebalancing under consideration of transaction costs. Panel A exhibits the results of the investment strategies based on momentum and moving average indicators (but with no volume-based indicators) in the 1966:01-2014:12 backtest period. The loss function in row 1 is based on negative continuously compounded absolute returns; the loss function in row 2 is modeled as the negative risk-adjusted excess return. Panel B lists the results of the investment strategies based on all technical indicators (momentum, moving average, and volume-based) compared to the 50:50 constant mix benchmark (CM). Rows 1, 2, and 3 contain the results for the three backtest periods beginning in 1966:01, 2000:01, and 1931:01. All end in 2014:12. Column (2) of the table shows the consistent p-value of the SPA test as well as the lower and upper bounds for p-values. The table also reports the sample loss for the buy-and-hold benchmark and the two investment strategies of each category that have the smallest sample loss value and the largest t-statistic for average relative performance $(\bar{d}_k)$. These two models are referred to as the "best" and "most significant" models, respectively. The loss value is shown in column (5), and the corresponding t-statistic (of their sample loss relative to the benchmark) is given in column (6). Finally, column (7) reports the "p-values" from the pairwise comparisons of "best" and "largest t-statistic" models with the benchmark. These p-values (unlike the SPA p-value) ignore the search over all models that preceded the selection of the model being compared to the benchmark, i.e., they do not account for the entire universe of models.

We observe that the MA2-12 model is not only the best model, but also the most significant. Unfortunately, this model is not statistically significantly superior in terms of Hansen's (2005) SPA test when a 10% significance level is assumed. This holds not only for the mean return loss function (first row in panel A), but also for the loss function based on the risk-adjusted excess return (second row in panel A) (consistent p-value of 0.1714, and lower bound p-value of 0.1302). While the best and most significant technical indicator model varies among the three backtest samples, we always observe a statistically significant outperformance against the buy-and-hold benchmark in terms of risk-adjusted excess returns. Furthermore, in contrast to the active investment strategies, the buy-and-hold benchmark is always fully invested in the S&P 500, which implies systematically higher volatility compared to our market timing models. In order to verify whether this explains the risk-adjusted outperformance of the best technical indicator models, in repeated simulations, the buy-and-hold benchmark is substituted with a 50:50 constant mix benchmark.[32]

The corresponding results are in panel B of Exhibit 9. We observe that the superiority of the technical models is now no longer statistically significant in two of the three backtest periods (1966:01-2014:12 and 1931:01-2014:12). The risk-adjusted outperformance against the benchmark remains statistically significant at a 5% level only for the 2000:01-2014:12 period (consistent p-value of 0.0365%). These results confirm the preceding results: Even if the market timing models are consistently better than the buy-and-hold benchmark – in terms of mean absolute

[32] This means that the benchmark strategy is 50% invested in the S&P 500 index and 50% in the cash market. Monthly rebalancing is conducted with transaction costs of 25 basis points (as with all other portfolio shifts).

returns or risk-adjusted excess returns – it continues to be difficult to obtain statistical significance using Hansen's (2005) SPA test.

In order to evaluate the influence of transaction costs on our result, we follow Blitz and van Vliet (2008) and repeat our baseline simulations (section 5.1) with transaction costs of 10 basis points. As Blitz and van Vliet (2008) emphasize, these are realistic transaction costs for highly liquid instruments, such as futures. As one would expect, with these lower transaction costs all active strategies become more attractive compared to the buy-and-hold benchmark strategy. With turnover-dependent transaction costs of 10 basis points, our best technical indicator model VOL2-12 in terms of risk-adjusted returns now becomes significant at a level of 10 percent. Specifically, within the backtest period ranging from 1966:01 to 2014:12, the consistent *p*-value of the VOL2-12 model reduces from 0.1157 (Panel B in Exhibit 4) to 0.0917 when performance is measured as risk-adjusted returns. All other results regarding the statistical significance in terms of Hansen's (2005) SPA test are not affected. This result shows up in both backtest periods, starting in 1966:01 and in 2000:01, respectively.

In terms of our monthly seasonality models, one can criticize that we consider all possible monthly allocation models and not only those for which empirical evidence exists.[33] While Hansen's (2005) SPA test is much more robust about irrelevant models compared to White's (2000) Reality Check, adding many irrelevant models also reduces the power of this test. In order to address this issue, we repeat our simulations with only three well documented calendar anomaly strategies, each trying to exploit one specific calendar effect: A strategy trying to exploit the January effect (investing in the stock market in January and otherwise in cash), another strategy based on the September swoon effect (investing in cash in September and in the stock market in

[33] We are grateful to an anonymous referee for this hint.

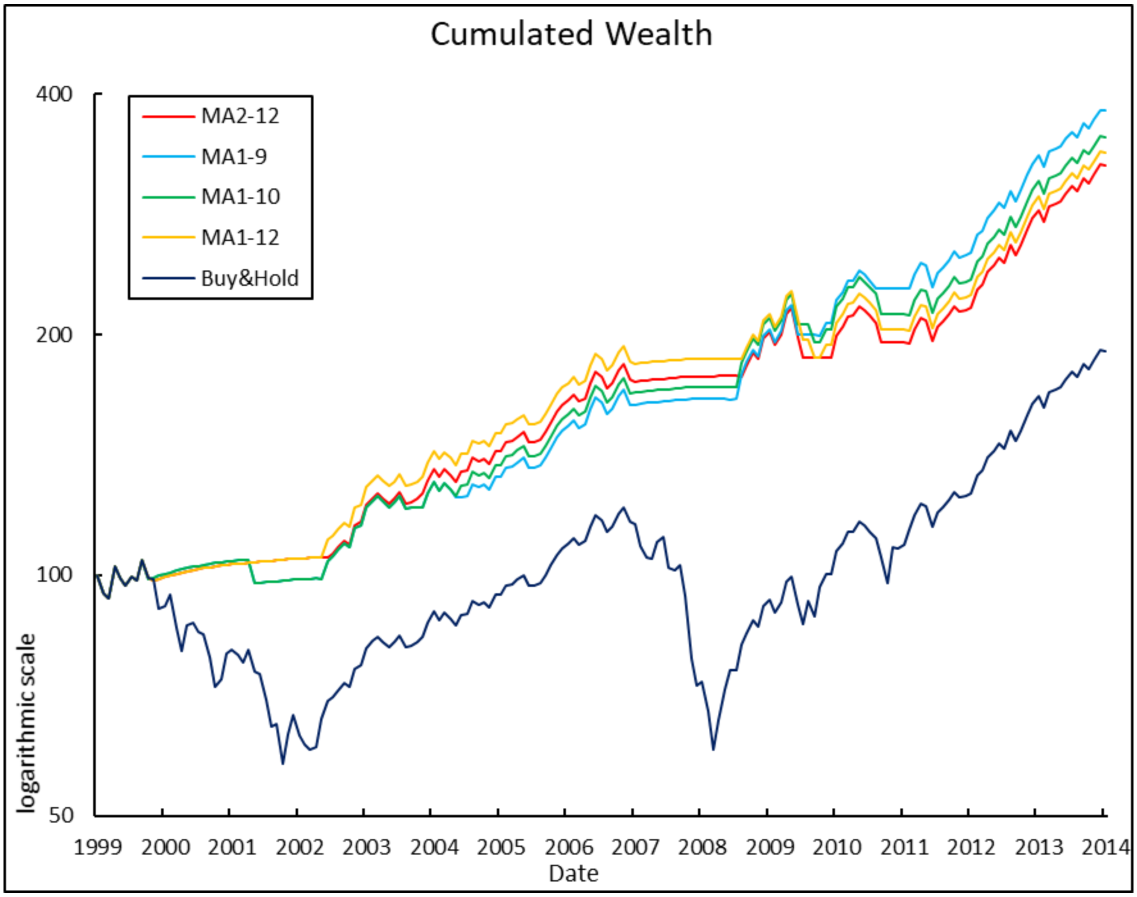
all other months), and the well known “Sell in May and Go Away” strategy, investing in the stock market from November to April and holding cash from May to October. Regarding the simulations in the time span from 1966 to 2014, we observe the following results: For our September swoon strategy (most significant strategy in terms of absolute returns) and the “Sell in May and Go Away” strategy (most significant strategy in terms of risk-adjusted returns) the consistent p-values decrease with our reduced strategy population. However, still with low transaction costs of 10 basis points, both consistent p-values are above the 10 percent significance level. In our “crisis period” ranging from 2000 to 2014, all three tested strategies are far beyond a statistical significant dominance of the buy-and-hold strategy in terms on Hansen’s SPA test.

6. Further insights and explanations

Despite the fact that the statistical significance of the dominance of the moving average strategies depends on the backtest period or the chosen benchmark strategy, we observe in each simulation period that moving average strategies dominate the buy-and-hold benchmark. This holds for mean absolute returns as well as for risk-adjusted excess returns. However, despite this issue, some studies find that moving average strategies can improve skewness and reduce the drawdowns compared to a buy-and-hold investment (e.g., Faber, 2009; Kilgallen, 2012; Clare et al., 2016; Ebert and Hilpert, 2016).

Next, we verify whether these properties are observable in the simulation setup. Because the market experienced two severe crises over the 2000:01-2014:12 period (the dot.com bubble from 2000 to 2002, and the global financial crisis in 2008/2009), we initially conduct a visual analysis of this time period.

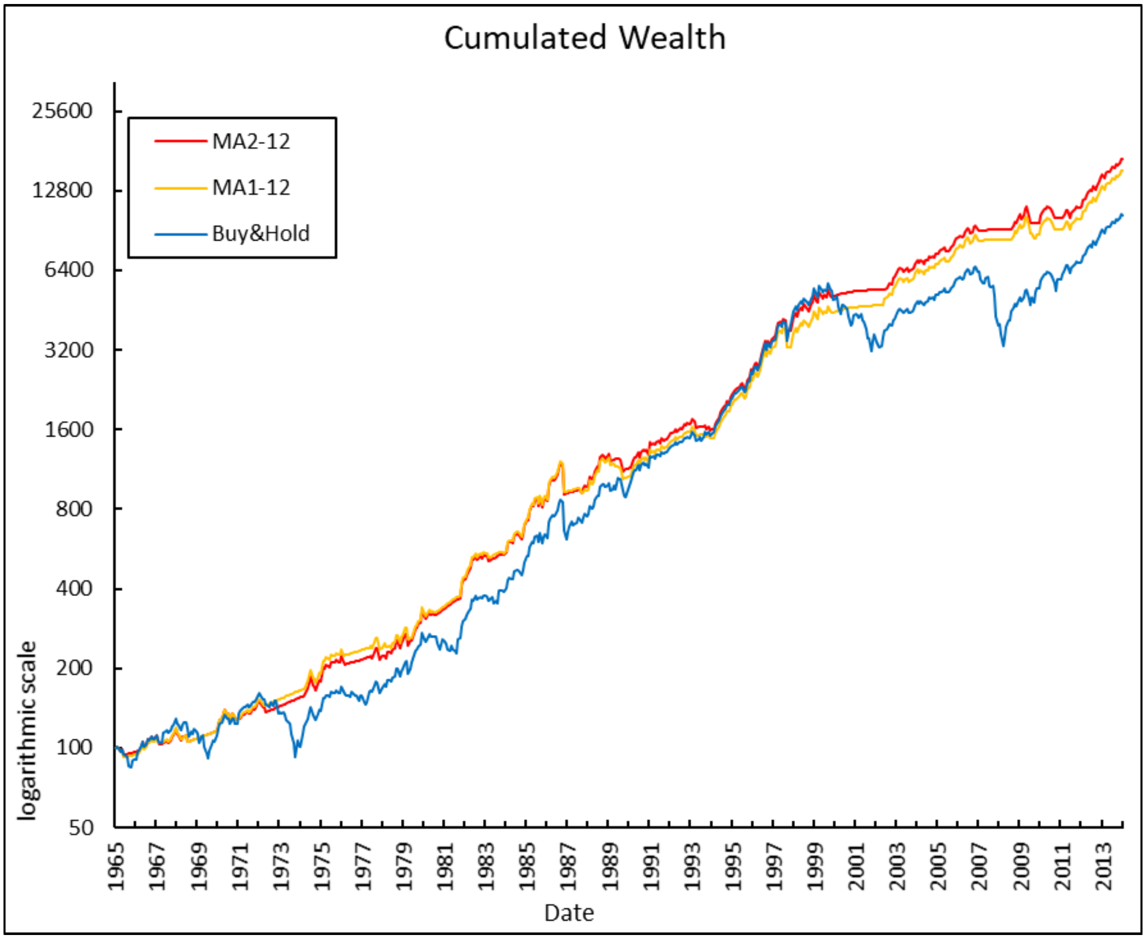
Exhibit 10: Cumulative wealth of moving average strategies compared to buy-and-hold benchmark (2000:01 to 2014:12)



Notes: This figure shows the development of an initial investment of $100 in January 2000 invested in the moving average strategies MA2-12, MA1-9, MA1-10, and MA1-12, as well as in the buy-and-hold benchmark (Buy&Hold) till December 2014.

Exhibit 10 shows the development of an initial investment of $100 in January 2000 invested in the moving average strategies MA2-12, MA1-9, MA1-10, and MA1-12, as well as in the buy-and-hold benchmark (Buy&Hold) till December 2014. The graphical representation clearly documents the higher cumulative wealth of all four moving average strategies (MA2-12, MA1-9, MA1-10, and MA1-12) compared to the buy-and-hold benchmark. While the technical indicator strategy VOL2-12 represents the best (and also most significant) strategy in our sample period from 1966:01 to 2014:12, the both moving average strategies MA2-12 and MA1-12 are also beneath the sixth best technical indicator models in this evaluation period (see Exhibit 3).[34] Exhibit 11 shows, that those both moving average strategies also dominate the buy-and-hold strategy in terms of cumulative wealth between 1966 and 2014 clearly.

Exhibit 11: Cumulative wealth of moving average strategies compared to buy-and-hold benchmark (1966:01 to 2014:12)



Notes: This figure shows the development of an initial investment of $100 in January 1966 invested in the moving average strategies MA2-12 and MA1-12, as well as in the buy-and-hold benchmark (Buy&Hold) till December 2014.

[34] The MA2-12 model is also the best strategy in our long-term simulations starting in 1931 and dominates the buy-and-hold strategy in terms of absolute and risk-adjusted return (see Exhibit 7).

Because this dominance in terms of absolute returns is not statistically significant in our analysis (see panel A in Exhibit 4 and 6), this confirms that Hansen's (2005) SPA test constitutes a high hurdle for an outperformance to be considered statistically significant.[35] Furthermore, Exhibit 10 and Exhibit 11 show that all moving average strategies provide obviously consistently lower drawdowns than a buy-and-hold investment during the dot.com crisis and the 2008/2009 financial crisis.

To verify this result more rigorously, the skewness, minimum monthly return, and maximum drawdown are analyzed for all four moving average strategies in all three backtest periods. The results are in Exhibit 12.

Exhibit 12: Skewness, minimum monthly return, and maximum drawdown

	MA2-12	MA1-9	MA1-10	MA1-12	B&H
Panel A: 1966:01 – 2014:12 Period					
Skewness	-1.00	-1.26	-1.21	-1.02	-0.66
Robust skewness ($\alpha = 0.01$)	-0.05	-0.16	-0.15	-0.05	-0.11
Robust skewness ($\alpha = 0.05$)	0.05	0.04	0.03	0.05	-0.12
Min. monthly return (%)	-24.56	-24.56	-24.56	-24.56	-24.31
Max. drawdown (%)	-23.49	-23.49	-23.57	-23.76	-50.21
Panel B: 2000:01 – 2014:12 Period					
Skewness	-0.16	-0.10	-0.19	-0.19	-0.74
Robust skewness ($\alpha = 0.01$)	-0.02	-0.01	-0.01	-0.02	-0.20
Robust skewness ($\alpha = 0.05$)	0.09	0.16	0.11	0.11	-0.20
Min. monthly return (%)	-8.35	-8.60	-8.60	-8.35	-18.27
Max. drawdown (%)	-13.33	-8.42	-13.03	-17.70	-50.21

[35] By now this phenomenon is known and motivated the development of less conservative data-snooping-resistant hypothesis tests (e.g., Hsu et al., 2014).

Panel C: 1931:01 – 2014:12 Period					
Skewness	-1.20	-0.93	-0.91	-1.20	-0.41
Robust skewness ($\alpha = 0.01$)	-0.05	-0.10	-0.09	-0.06	-0.18
Robust skewness ($\alpha = 0.05$)	0.07	0.05	0.05	0.05	-0.18
Min. monthly return (%)	-26.58	-24.56	-24.56	-26.58	-33.89
Max. drawdown (%)	-44.34	-37.71	-35.09	-47.21	-74.28

Notes: This table reports traditional skewness, robust skewness based on 1% and 5% quantiles, the minimum monthly return (in percent), and the maximum drawdown (in percent) for the moving average-based investment strategies MA2-12, MA1-9, MA1-10, and MA1-12, as well as for the buy-and-hold benchmark (B&H). All calculations are based on monthly log returns. Panel A shows the results for the 1966:01-2014:12 backtest period, panel B shows the results for the 2000:01-2014:12 period, and panel C the results for the 1931:01-2014:12 period.

Besides traditional skewness, a robust skewness measure proposed by various economic studies (Ebert and Hilpert, 2016; Green and Hwang, 2012; Kim and White, 2004) is also evaluated. It is defined as follows:

(9) $$S^R = \frac{(P_{1-\alpha} - P_{0.50}) - (P_{0.50} - P_\alpha)}{(P_{1-\alpha} - P_\alpha)},$$

where P_α is the α-percentile of the monthly return distribution. The robust skewness is evaluated with $\alpha = 0.01$ (Green and Hwang, 2012; Ebert and Hilpert, 2016) and $\alpha = 0.05$. The results confirm Clare et al.'s (2016) findings that the application of moving average strategies has a positive effect on the skewness of the return distribution compared to the buy-and-hold strategy. This holds especially for robust skewness with $\alpha = 0.05$, where we observe the positive effect for all four moving average strategies in all three periods.

For the 2000-2014 period (which includes the two severe stock market crises), this effect is extraordinarily pronounced, as indicated by all three skewness measures. In order to analyze the maximum drawdown, we follow Pedersen (2015), and define the percentage drawdown (DD_t) as $DD_t = (HWM_t - P_t)/HWM_t$, where HWM_t represents the high water mark (i.e., the highest cumulative return achieved in the past), and P_t is the cumulative return (or alternatively the price) at time t. The maximum drawdown (MDD_n) over some past time period ($t = 1, \ldots, n$) is then defined as:

(10) $$MDD_n = max_{t \leq n} DD_t.$$

As shown in Exhibit 12, maximum drawdown is significantly lower for all four moving average strategies than for the buy-and-hold strategy. This result is in line with the evidence in Kilgallen (2012) and Clare et al. (2016). But, in order to obtain further insights, a risk factor analysis based on the factors proposed by Fama and French (1992) and Carhart (1997) is conducted. The Fama and French (1992) factors include the excess return of the U.S. equity market (MKT-RF), the small minus big market capitalization factor (SMB), and the high minus low book-to-market ratio factor (HML). These are complemented by Carhart's (1997) momentum factor UMD (up minus down, or winner minus loser stocks). The corresponding regression coefficients and Newey-West-adjusted *t*-statistics are in Exhibit 13 (Newey and West, 1987).

Exhibit 13: Risk factor analysis

Model	Alpha	MKT-RF	SMB	HML	UMD	Prob. F
Period: 1966:01 – 2014:12						
MA2-12	0.11 [1.08]	0.57 [12.57]	-0.12 [-2.80]	0.07 [1.52]	0.20 [7.89]	0.00
MA1-9	0.07 [0.62]	0.53 [11.27]	-0.08 [-1.81]	0.07 [1.37]	0.14 [5.47]	0.00
MA1-10	0.07 [0.63]	0.53 [11.44]	-0.09 [-2.04]	0.09 [1.73]	0.16 [6.12]	0.00
MA1-12	0.09 [0.84]	0.57 [12.54]	-0.11 [-2.57]	0.09 [1.84]	0.19 [7.68]	0.00
Period: 2000:01 – 2014:12						
MA2-12	0.39 [2.38]	0.40 [5.93]	-0.13 [-1.95]	0.09 [1.42]	0.14 [4.63]	0.00
MA1-9	0.50 [3.25]	0.41 [6.04]	-0.15 [-2.36]	0.06 [0.92]	0.11 [3.93]	0.00
MA1-10	0.44 [2.69]	0.43 [6.16]	-0.13 [-2.02]	0.08 [1.16]	0.12 [4.06]	0.00
MA1-12	0.40 [2.35]	0.41 [6.05]	-0.11 [-1.59]	0.09 [1.49]	0.13 [4.49]	0.00

Period: 1931:01 – 2014:12						
MA2-12	0.05 [0.59]	0.50 [12.38]	-0.03 [-0.56]	0.08 [1.97]	0.30 [9.94]	0.00
MA1-9	0.06 [0.69]	0.46 [11.69]	-0.02 [-0.32]	0.07 [1.65]	0.25 [8.30]	0.00
MA1-10	0.06 [0.72]	0.46 [11.73]	-0.02 [-0.36]	0.08 [2.05]	0.26 [8.70]	0.00
MA1-12	0.03 [0.34]	0.50 [12.37]	-0.03 [-0.52]	0.08 [1.98]	0.29 [9.68]	0.00

Notes: This table reports the results of a risk factor analysis for the moving average-based investment strategies MA2-12, MA1-9, MA1-10, and MA1-12. The returns of these strategies are regressed on the three Fama and French (1992) risk factors MKT-RF (market risk premium), SMB (small minus big market capitalization), HML (high minus low book-to-market ratio), and on Carhart's (1997) momentum factor UMD (up minus down). "Alpha" denotes the alpha return (regression constant), and "Prob. F" is the *p*-value of the regression model *F*-statistic. The table lists the regression coefficients and the Newey-West-adjusted *t*-statistics (in squared brackets) (Newey and West, 1987). All calculations are based on monthly log returns. Panel A shows the results for the 1966:01-2014:12 backtest period, panel B shows the results for the 2000:01-2014:12 period, and panel C the results for the 1931:01-2014:12 period.

As one would expect, the main and statistically significant drivers of the moving average strategy returns are the U.S. equity market risk premium and Carhart's (1997) momentum premium.[36] This result is observable for all three analyzed time periods. Especially interesting is the 2000-2014 span, where we also observe a statistically significant and economically relevant alpha return. Obviously, moving average strategies can provide a significant alpha in times of turbulent stock market cycles with large drawdowns.

In light of these results, a natural question is why moving average strategies should work. Obviously, as trend-following strategies they depend strongly on the existence of market trends. Hurst et al. (2013) provide a thorough explanation of this occurrence, and of how market trends emerge. They argue that an initial underreaction to a shift in fundamental value offers a trend-following strategy the opportunity to invest before new information is fully reflected in prices.

[36] This finding is in line with evidence provided in Moskowitz et al. (2012). It is important to recognize that Carhart's (1997) momentum factor represents cross-sectional momentum. However, Moskowitz et al. (2012) demonstrate vividly that cross-sectional momentum and time series momentum are linked. They decompose cross-sectional momentum into three terms and time series momentum into two terms. Both decompositions contain an autocovariance term that in both cases is the most important component. This autocovariance term explains the (comparatively strong) link between the two concepts.

Besides certain market frictions, behavioral phenomena such as anchoring (Tversky and Kahneman, 1974) and the disposition effect (Shefrin and Statman, 1985; Frazzini, 2006) provide a convincing explanation for this initial underreaction effect. Once a trend begun, other phenomena, like for example the herding mentality (e.g., Bikhchandani et al., 1992), may extend it beyond the fundamental value resulting in an overreaction effect. However, at some point, prices extend too far beyond fundamental value, and the trend reverses. Beyond these explanations for either underreaction or overreaction effects, the behavioral finance literature also provides models with the power of explaining both effects simultaneously. While Barberis et al. (1998) explains under- and overreaction effects with representativeness and conservatism, the model provided by Daniel et al. (1998) is based on overconfidence and self-attribution. [37]

7. Concluding remarks

This study explores whether it is possible to beat the buy-and-hold of the S&P 500 index with an active investment strategy. In contrast to many existing studies, a large set of investment strategies that are based on calendar anomalies, technical indicators, and fundamental factors is considered, and the results are verified in terms of statistical significance with a data-snooping-resistant test.

Within the monthly seasonality strategies, we observe a dominance of the “Sell in May and Go Away” strategy and the “September swoon” strategy over the buy-and-hold benchmark in specific backtest periods. This result is in line with existing studies. The result holds for models with similar allocations (e.g., the SEA73 model, which invests in the cash market in September and in October). However, we observe no statistically significant dominance of any monthly

[37] See also the model from Hong and Stein (1999) that grounds both effects with boundedly rational agents.

seasonality model against the buy-and-hold benchmark for any of the three backtest periods. Furthermore, the superiority of the various monthly seasonality models depends on the time period under investigation. We observe a dominance of the "Sell in May" strategy (and some similar strategies) for the 1966-2014 period, whereas the September swoon strategy (and similar strategies) dominates from 1931 to 2014.

In terms of fundamental models, we observe a dominance of the forecast combination approach and the predictive regression model based on the dividend payout ratio against the buy-and-hold benchmark from 1966 to 2014. However, this dominance is not statistically significant based on Hansen's (2005) SPA test. During the 2000-2014 time period, the predictive regression model based on all fundamental variables (the kitchen sink model) provides the best and most significant results, albeit the dominance of this model is again statistically insignificant. This model also exhibits poorer performance than the best investment strategies based on monthly calendar effects. In light of these results, we believe future research should consider some of the various improvements that have been suggested in the recent past, e.g., the implementation of time-varying coefficients (Dangl and Halling, 2012) or regime shifts (Hammerschmid and Lohre, 2018).

The best results are observed for technical indicator-based models in all three backtest periods. Their dominance in terms of risk-adjusted excess returns against the buy-and-hold benchmark is statistically significant in the period from 2000 to 2014. When substituting for the S&P 500 buy-and-hold benchmark with a risk-reduced 50:50 constant mix strategy, the statistically risk-adjusted outperformance still persists over this "crisis period". In our longer backtest period ranging from 1966 to 2014, we also observe a risk-adjusted outperformance of our technical indicator strategies, but only at a significance level around 10 percent. However, given that Hansen's SPA test is a very conservative test, this superiority could also be judged as significant.

Further analysis shows that the popularity of the moving average strategies is justified by their potential to enhance absolute and risk-adjusted returns as well as by their positive effects on drawdown. This holds particularly true when we compare the maximum drawdown and skewness with the corresponding values of the buy-and-hold strategy. A further risk factor analysis shows that moving average strategies are suitable for generating positive alpha returns against commonly used risk factors, especially during turbulent stock market environments. Behavioral finance research provides various plausible explanations for why stock market trends exist, and explains the positive return and risk properties of the trend-following technical indicator strategies. Therefore, this study gives a further potential explanation for the popularity of these strategies in the investment industry (e.g., Menkhoff, 2010).

As already mentioned, because of time and capacity limitations, within the group of fundamental forecast models and technical indicator models we only covered the (in our view) most popular strategies. However, the integration of further investment strategies into our test-framework, as well as the additional application of a less conservative multiple test-framework (e.g., Hsu et al., 2014), is straightforward and is left for future research.

References

Andrade, S.C., Chhaochharia, V., and Fuerst, M.E. (2013): "Sell in May and Go Away" Just Won't Go Away, in: *Financial Analysts Journal*, Vol. 69, 94-105.

Baetje, F., and Menkhoff, L. (2016): Equity Premium Prediction: Are Economic and Technical Indicators Unstable? Discussion Paper 1552, DIW Berlin.

Bajgrowicz, P., and Scaillet, O. (2012): Technical trading revisited: False discoveries, persistence tests, and transaction costs, in: *Journal of Financial Economics*, Vol. 106, 473-491.

Barberis, N., and Thaler, R. (2003): A Survey of Behavioral Finance, in: *Handbook of the Economics of Finance*, eds. G.M. Constantinides, M. Harris, R.M. Stulz, Chapter 18, 1051-1121.

Barberis, N., Shleifer, A., and Vishny, R. (1998): A model of investor sentiment, in: *Journal of Financial Economics*, Vol. 49, 307-343.

Baur, D.G. (2013): Gold – Fundamental Drivers and Asset Allocation, Working paper, University of Technology, Sydney.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992): A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades, in: *Journal of Political Economy*, Vol. 100, 992-1026.

Blitz, D.C., and van Vliet, P. (2008): Global Tactical Cross-Asset Allocation: Applying Value and Momentum Across Asset Classes, in: *Journal of Portfolio Management*, Vol. 35, 23-38.

Blume, L., Easley, D., and O'Hara, M. (1994): Market Statistics and Technical Analysis: The Role of Volume, in: *Journal of Finance*, Vol. 49, 153-181.

Bouman, S., and Jacobsen, B. (2002): The Halloween Indicator, "Sell in May and Go Away": Another Puzzle, in: *American Economic Review*, Vol. 92, 1618-1635.

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

Campbell, J.Y. (1987): Stock returns and the term structure, in: *Journal of Financial Economics*, Vol. 18, 373-399.

Campbell, J.Y., and Thompson, S.B. (2008): Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?, in: *Review of Financial Studies*, Vol. 21, 1509-1531.

Cao, M., and Wei, J. (2005): Stock market returns: A note on temperature anomaly, in: *Journal of Banking & Finance*, Vol. 29, 1559-1573.

Carhart, M.M. (1997): On Persistence in Mutual Fund Performance, in: *Journal of Finance*, Vol. 52, 57-82.

Clare, A., Seaton, J., Smith, P.N., and Thomas, S. (2016): The trend is our friend: Risk parity, momentum and trend following in global asset allocation, in: *Journal of Behavioral and Experimental Finance*, Vol. 9, 63-80.

Dangl, T., and Halling, M. (2012): Predictive regressions with time-varying coefficients, in: *Journal of Financial Economics*, Vol. 106, 157-181.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998): Investor Psychology and Security Market Under- and Overreactions, in: *Journal of Finance*, Vol. 53, 1839-1885.

Dichtl, H., and Drobetz, W. (2014): Are stock markets really so inefficient? The case of the "Halloween Indicator," in: *Finance Research Letters*, Vol. 11, 112-121.

Dichtl, H., and Drobetz, W. (2015): Sell in May and Go Away: Still good advice for investors? in: *International Review of Financial Analysis*, Vol. 38, 29-43.

D'Mello, R., Ferris, S.P, and Hwang, C.Y. (2003): The tax-loss selling hypothesis, market liquidity, and price pressure around the turn-of-the-year, in: *Journal of Financial Markets*, Vol. 6, 73-98.

Doidge, C., Karolyi, G.A., and Stulz, R.M. (2017): The U.S. listing gap, in: *Journal of Financial Economics*, Vol. 123, 464-487.

Doornik, J.A. (2009): An Object-oriented Matrix Programming Language: OX™ 7. London, Timberlake Consultants Ltd.

Ebert, S., and Hilpert, C. (2016): Skewness Preference and the Popularity of Technical Analysis, Working paper, Tilburg University.

Faber, M.T. (2007): A Quantitative Approach to Tactical Asset Allocation, in: *Journal of Wealth Management*, Vol. 9, 69-79.

Faber, M.T. (2009): A Quantitative Approach to Tactical Asset Allocation – An Update, *Cambria Investment Management*.

Fama, E.F., and French, K.R. (1988): Dividend yields and expected stock returns, in: *Journal of Financial Economics*, Vol. 22, 3-25.

Fama, E.F., and French, K.R. (1992): The Cross-Section of Expected Stock Returns, in: *Journal of Finance*, Vol. 47, 427-465.

Fama, E.F., and Schwert, G.W. (1977): Asset returns and inflation, in: *Journal of Financial Economics*, Vol. 5, 115-146.

Frazzini, A. (2006): The Disposition Effect and Underreaction to News, in: *Journal of Finance*, Vol. 61, 2017-2046.

Giacomini, R., and White, H. (2006): Tests of Conditional Predictive Ability, in: *Econometrica*, Vol. 74, 1545-1578.

Glabadanidis, P. (2014): The Market Timing Power of Moving Averages: Evidence from US REITS and REIT Indexes, in: *International Review of Finance*, Vol. 14, 161-202.

Goetzmann, W.N., Kim, D., Kumar, A., and Wang, Q. (2015): Weather-Induced Mood, Institutional Investors, and Stock Returns, in: *Review of Financial Studies*, Vol. 28, 73-111.

Granville, J.E. (1963): *Granville's New Key to Stock Market Profits*, Prentice-Hall, New York.

Green, T.C., and Hwang, B.-H. (2012): Initial Public Offerings as Lotteries: Skewness Preference and First-Day Returns, in: *Management Science*, Vol. 58, 432-444.

Hammerschmid, R., and Lohre, H. (2018): Regime shifts and stock return predictability, in: *International Review of Economics & Finance*, forthcoming.

Hansen, P.R. (2005): A Test for Superior Predictive Ability, in: *Journal of Business & Economic Statistics*, Vol. 23, 364-380.

Harvey, C.R. (2017): Presidential Address: The Scientific Outlook in Financial Economics, in: *Journal of Finance*, Vol. 72, 1399-1440.

Harvey, C.R., Liu, Y., and Zhu, H. (2016): … and the Cross-Section of Expected Returns, in: *Review of Financial Studies*, Vol. 29, 5-68.

Haug, M., and Hirschey, M. (2006): The January Effect, in: *Financial Analysts Journal*, Vol. 62, 78-88.

Haug, M., and Hirschey, M. (2011): The September Swoon, in: *Journal of Applied Finance*, No. 1, 118-139.

Hong, H., and Stein, J.C. (1999): A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets, in: *Journal of Finance*, Vol. 54, 2143-2184.

Hong, H., Torous, W., and Valkanov, R. (2007): Do industries lead stock markets?, in: *Journal of Financial Economics*, Vol. 83, 367-396.

Hsu, P.-H., and Kuan, C.-M. (2005): Reexamining the Profitability of Technical Analysis with Data Snooping Checks, in: *Journal of Financial Econometrics*, Vol. 3, 606-628.

Hsu, P.-H., Hsu, Y.-C., and Kuan, C.-M. (2010): Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias, in: *Journal of Empirical Finance*, Vol. 12, 730-755.

Hsu, Y.-C., Kuan, C.-M., and Yen, M.-F. (2014): A Generalized Stepwise Procedure with Improved Power for Multiple Inequalities Testing, in: *Journal of Financial Econometrics*, Vol. 12, 730-755.

Hurst, B., Ooi, Y.H., and Pedersen, L.H. (2013): Demystifying Managed Futures, in: *Journal of Investment Management*, Vol. 11, 42-58.

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

Jacobsen, B., and Zhang, C.Y. (2014): The Halloween Indicator, “Sell in May and Go Away”: An Even Bigger Puzzle, Working paper, University of Edinburgh Business School.

Jones, C.P., and Lundstrum, L.L. (2009): Is “Sell in May and Go Away” a Valid Strategy for U.S. Equity Allocation? in: *Journal of Wealth Management*, Vol. 12, 104-112.

Kamstra, M.J., Kramer, L.A., and Levi, M.D. (2003): Winter Blues: A SAD Stock Market Cycle, in: *American Economic Review*, Vol. 93, 324-343.

Keim, D.B., and Stambaugh, R.F. (1986): Predicting returns in the stock and bond markets, in: *Journal of Financial Economics*, Vol. 17, 357-390.

Kilgallen, T. (2012): Testing the Simple Moving Average across Commodities, Global Stock Indices, and Currencies, in: *Journal of Wealth Management*, Vol. 15, 82-100.

Kim, T.-H., and White, H. (2004): On more robust estimation of skewness and kurtosis, in: *Finance Research Letters*, Vol. 1, 56-73.

Lo, A.W., and MacKinlay, A.C. (1990): Data-Snooping Biases of Financial Asset Pricing Models, in: *Review of Financial Studies*, Vol. 3, 431-467.

Lo, A.W., Mamayski, H., and Wang, J. (2000): Foundation of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation, in: *Journal of Finance*, Vol. 55, 1705-1765.

Lucey, B.M., and Zhao, S. (2008): Halloween or January? Yet another puzzle, in: *International Review of Financial Analysis*, Vol. 17, 1055-1069.

Ludvigson, S.C., and Ng, S. (2007): The empirical risk-return relation: A factor analysis approach, in: *Journal of Financial Economics*, Vol. 83, 171-222.

Lynch, A.W., and Balduzzi, P. (2000): Predictability and Transaction Costs: The Impact on Rebalancing Rules and Behavior, in: *Journal of Finance*, Vol. 55, 2285-2309.

Marshall, B.R., Nguyen, N.H., and Visaltanachoti, N. (2017): Time series momentum and moving average trading rules, in: *Quantitative Finance*, Vol. 17, 405-421.

McLean, R.D., and Pontiff, J. (2016): Does Academic Research Destroy Stock Return Predictability?, in: *Journal of Finance*, Vol. 71, 5-32.

Mele, A. (2007): Asymmetric stock market volatility and the cyclical behavior of expected returns, in: *Journal of Financial Economics*, Vol. 86, 446-478.

Menkhoff, L. (2010): The use of technical analysis by fund managers: International evidence, in: *Journal of Banking & Finance*, Vol. 34, 2573-2586.

Moskowitz, T.J., Ooi, Y.H., and Pedersen, L.H. (2012): Time series momentum, in: *Journal of Financial Economics*, Vol. 104, 228-250.

Neely, C.J., Rapach, D.E., Tu, J., and Zhou, G. (2014): Forecasting the Equity Risk Premium: The Role of Technical Indicators, in: *Management Science*, Vol. 60, 1772-1791.

Newey, W., and West, K. (1987): A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, in: *Econometrica*, Vol. 55, 703-708.

Patton, A., Politis, D.N., and White, H. (2009): Correction to “Automatic Block-Length Se-lection for the Dependent Bootstrap” by D. Politis and H. White, in: *Econometric Reviews*, Vol. 28, 372-375.

Pedersen, L.H. (2015): Efficiently Inefficient. How Smart Money Invests and Market Prices Are Determined. Princeton (New Jersey), Princeton University Press.

Pesaran, M.H., and Timmermann, A. (1995): Predictability of Stock Returns: Robustness and Economic Significance, in: *Journal of Finance*, Vol. 50, 1201-1228.

Politis, D.N., and Romano, J.P. (1994): The Stationary Bootstrap, in: *Journal of the American Statistical Association*, Vol. 89, 1303-1313.

Politis, D.N., and White, H. (2004): Automatic Block-Length Selection for the Dependent Bootstrap, in: *Econometric Reviews*, Vol. 23, 53-70.

Rapach, D.E., Strauss, J.K., and Zhou, G. (2010): Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy, in: *Review of Financial Studies*, Vol. 23, 821-862.

Rapach, D.E., and Zhou, G. (2013): Forecasting Stock Returns, in: G. Elliott and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 2A, Amsterdam, Elsevier, 328-383.

Sharpe, W.F. (1994): The Sharpe Ratio, in: *Journal of Portfolio Management*, Vol. 21, 49-58.

Shefrin, H., and Statman, M. (1985): The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence, in: *Journal of Finance*, Vol. 40, 777-790.

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

Sullivan, R., Timmermann, A., and White, H. (2001): Dangers of Data Mining: The Case of Calendar Effects in Stock Returns, in: *Journal of Econometrics*, Vol. 105, 249-286.

Swinkels, L., and van Vliet, P. (2012): An Anatomy of Calendar Effects, in: *Journal of Asset Management*, Vol. 13, 271-286.

Tversky, A., and Kahneman, D. (1974): Judgment under Uncertainty: Heuristics and Biases, in: *Science*, Vol. 185, 1124-1131.

Welch, I., and Goyal, A. (2008): A Comprehensive Look at The Empirical Performance of Equity Premium Prediction, in: *Review of Financial Studies*, Vol. 21, 1455-1508.

White, H. (2000): A Reality Check for Data Snooping, in: *Econometrica*, Vol. 68, 1097-1126.

Zakamulin, V. (2014): The real-life performance of market timing with moving average and time-series momentum rules, in: *Journal of Asset Management*, Vol. 15, 261-278.

Zhang, C.Y. (2014): Vacation behaviours and seasonal patterns of stock market returns, Working paper, Nottingham University Business School China.

Zhu, Y., and Zhou, G. (2009): Technical analysis: An asset allocation perspective on the use of moving averages, in: *Journal of Financial Economics*, Vol. 92, 519-544.

Appendix A1: Description of all documented monthly seasonal models

Model	January	February	March	April	May	June	July	August	September	October	November	December
SEA0000	1	1	1	1	1	1	1	1	1	1	1	1
SEA0009	1	1	1	1	1	1	1	1	0	1	1	1
SEA0030	1	0	1	1	1	1	1	1	0	1	1	1
SEA0054	1	1	1	1	0	1	1	1	0	1	1	1
SEA0069	1	1	1	1	1	1	1	0	0	1	1	1
SEA0073	1	1	1	1	1	1	1	1	0	0	1	1
SEA0258	1	1	1	1	0	1	1	1	0	0	1	1
SEA0265	1	1	1	1	1	0	0	1	0	1	1	1
SEA0269	1	1	1	1	1	0	1	0	0	1	1	1
SEA0279	1	1	1	1	1	1	0	0	0	1	1	1
SEA0325	0	0	1	1	1	0	1	1	0	1	1	1
SEA0759	1	1	1	1	1	0	0	0	0	1	1	1
SEA0779	1	1	1	1	1	1	0	0	0	0	1	1
SEA0860	0	0	1	1	0	0	1	1	0	1	1	1
SEA0884	0	0	1	1	1	0	1	0	0	1	1	1
SEA1299	1	0	1	1	1	0	0	0	0	1	1	1
SEA1530	1	1	1	1	0	0	0	0	0	1	1	1
SEA1565	1	1	1	1	1	0	0	0	0	0	1	1
SEA1761	0	0	1	1	1	0	0	0	0	1	1	1
SEA2063	1	0	0	0	0	1	1	1	0	0	1	1
SEA2244	1	0	1	1	0	0	0	0	0	1	1	1
SEA2279	1	0	1	1	1	0	0	0	0	0	1	1
SEA2482	1	1	1	1	0	0	0	0	0	0	1	1
SEA2706	0	0	1	1	0	0	0	0	0	1	1	1
SEA2741	0	0	1	1	1	0	0	0	0	0	1	1
SEA3154	1	0	1	1	0	0	0	0	0	0	1	0
SEA3484	0	0	1	1	0	0	0	0	0	0	1	1
SEA3909	0	0	1	1	0	0	0	0	0	0	0	1

Notes: This table presents a description of the investment strategies based on monthly seasonalities (SEA) that are discussed here. The entry “1” (“0”) indicates a stock (cash) allocation in a given month. Model SEA0000 denotes the buy-and-hold strategy.