The intraday performance of market timing strategies and trading systems based on Japanese candlesticks

Matthieu Duvinage[a] Paolo Mazza[b] Mikael Petitjean[c]

January 14, 2013

Abstract

We develop market timing strategies and trading systems to test the intraday predictive power of Japanese candlesticks at the 5-minute interval on the 30 constituents of the DJIA index. Around a third of the candlestick rules outperforms the buy-and-hold strategy at the conservative Bonferroni level. After adjusting for trading costs, just a few rules remain significant however. When we correct for data snooping by applying the SSPA test on double-or-out market timing strategies, no single candlestick rule beats the buy-and-hold strategy after transaction costs. We also design fully automated trading systems by combining the best performing candlestick rules. No evidence of outperformance is found after transaction costs. Although Japanese candlesticks can somewhat predict intraday returns on large US caps, we show that such predictive power is too limited for active portfolio management to outperform the buy-and-hold strategy when luck, risk, *and* trading costs are correctly measured.

JEL Classification: G14, G17, C58

Keywords: Technical analysis, bootstrap, SSPA, Japanese candlesticks.

---

[a] TCTS Lab, Faculté Polytechnique de Mons (FPMs), UMons, 31 Boulevard Dolez - 7000 Mons (Belgium). Email: matthieu.duvinage@umons.ac.be.

[b] Louvain School of Management, Université Catholique de Louvain, 151 Chaussée de Binche - 7000 Mons (Belgium). E-mail: paolo.mazza@uclouvain.be.

[c] Louvain School of Management and CORE, Université Catholique de Louvain, 151 Chaussée de Binche - 7000 Mons (Belgium). E-mail: mikael.petitjean@uclouvain.be.

1 Introduction

Are investors smart when they try to beat the market? According to the efficient market theory, they are not. Once costs, risk, and luck are correctly measured, outperforming the benchmark in the long run is not possible. Short-term local outperformance is just due to statistical fluctuations and the best portfolio management technique consists in tracking the benchmark. Nevertheless, fund managers and private bankers in the real world are still advocating the use of dynamic strategies. The performance of these strategies is typically measured against a benchmark in terms of raw and (sometimes) risk-adjusted returns. However, surviving practitioners seldom provide an in-depth statistical performance analysis of their success. They rarely address the following question: Is luck, hidden risk or underestimated trading costs the main explanation behind these 'abnormal' returns?

In deciding upon the timing of short-term transactions, many active managers focus on technical analysis. Technical analysis is based on the study of historical asset prices. It includes numerical methods, chartist graphical methods or Japanese candlestick pattern recognition. This paper focuses on the latter. Japanese candlesticks are extensively used by practitioners nowadays. For example, Marshall et al. (2006) write that 'since its introduction to the Western World, candlestick technical analysis have become ubiquitous, available in almost every software and online charting package' (p. 2304) . According to Nison (1994, p. 7), 'candle charting techniques are among the most discussed form of technical analysis in the world'. Japanese candlesticks characterize price dynamics with a candle and two shadows. They display the close, open, high and low prices over a given timescale (minutes, days, weeks, or even months). Specific candlestick patterns are bullish or bearish and lead to buy or sell transactions.

No previous study has looked at the information content of Japanese candlesticks on an intraday basis by using up-to-date statistical tools, on so many rules. In this paper, 83 Japanese candlestick rules are tested at the 5-minute interval on the 30 stocks of the Dow Jones Industrial Average (DJIA) index. We first test the market timing power of Japanese candlesticks by computing the return realized over a given number of periods after a buy or sell signal is generated. Even if Japanese candlestick rules appear to be performing well, we must ensure that luck is not just the explanation. Otherwise, investors may be recommended to blindly apply strategies that are not 'true outperformers'. To determine whether the returns generated by the Japanese candlestick rules are spurious or not, we rely on the bootstrap methodology and assume different return generating models, such as the random walk, AR(1), and GARCH-in-mean processes.

Data snooping is also a serious issue when a large number of trading rules are tested on the same sample. In such a case, some rules will inevitably produce 'false positives'. For example, 5% of randomly chosen trading rules will turn out to be significant at the 5% level by chance alone. In this paper, we correct for this bias by using the Superior Predictive Ability (SPA) test and its stepwise version (SSPA). Interestingly, this bias correction algorithm is able to identify every significant rule that beats the benchmark. The SSPA is applied on a double-or-out market timing strategy which consists into holding a long position modulated by one buy or sell transaction depending upon the next Japanese candlestick signal, as defined in Bessembinder and Chan (1998). In other words, the buy-and-hold strategy is followed except when the trading rule provides a signal. A double-or-out strategy is typically recommended when there is a high variability in the number of generated signals between rules. Such a market timing strategy is very convenient to correct for data snooping because it delivers the same number of observations for each simulation, which is required by the SSPA test.

Finally, we combine the best performing trading rules into fully automated trading systems. Such trading systems have never been tested before using up-to-date statistical tools. Practitioners often justify the use of trading systems because the combination of all rules may give better results than the simple application of individual rules. As a robustness check, a sensitivity analysis is also performed by taking trading costs into account and by varying some intrinsic characteristics of the rules. A contrarian strategy is also developed by generating the opposite signal that the rule recommends to follow.

In the market timing application, the ratio of profitable trades based on the original time series is 56 % for bullish patterns and 22 % for bearish patterns. 64 % of bullish patterns and 39 % of bearish patterns deliver positive mean returns. Statistical testing shows that some Japanese candlesticks have significant explanatory power at the conservative Bonferroni level which counteracts the problem of multiple hypothesis testing. Out of 83 rules, 26 are significant based on raw returns and 27 are significant based on risk-adjusted returns. Whatever the parameter configuration and the underlying return generating model, no real difference is detected, pointing to robust results. When trading costs are included, trading profits are eroded in the vast majority of Japanese candlestick rules. Only five rules out of 83 rules exhibit a higher average profit than the average trading cost per trade. From a risk-adjusted point of view, there are three significant Japanese candlestick patterns only. When contrarian rules are allowed, five Japanese candlestick patterns deliver significant results.

Fully automated trading systems are then developed in three steps. We first identify the top ten candlestick rules for each of the 30 stocks. To be selected, these top ten rules must be significant at the Bonferroni level in the double-or-out market timing strategy. We then retain only the rules that are listed at least twice on average across the 30 stocks. 11 candlestick patterns pass the filter. These 11 candlestick patterns are finally combined in a double-or-out trading system in order to potentially detect profitable complex trading strategies. As a robustness check, we include contrarian rules and use different parameter configurations. Over the 24,232 tested trading systems *on average per stock*, no evidence of statistical outperformance is found after trading costs.

Once luck, risk, and trading costs are taken into account in an intraday environment, we conclude that markets on large caps are not sufficiently inefficient for the buy-and-hold strategy to be beaten by active trading rules based on Japanese candlesticks.

The remainder of the paper is organized as follows. Section 2 provides a brief literature review on technical analysis and candlesticks. Section 3 describes the dataset and the methodology that we apply. Section 4 includes the main empirical findings of the paper. The final section concludes.

2 Literature review

In practical terms, the Efficient Market Hypothesis (EMH) implies that one cannot consistently outperform the market once risk, luck, and trading costs are correctly measured. If the EMH is violated however, active trading strategies such as technical analysis may beat a purely passive benchmarking strategy. Technical analysis consists in predicting future price trends based on historical data on prices (and volumes). Compared to fundamental analysis, no balance-sheet information is strictly required. Technical analysis also provides specific signals to enter and leave the market, i.e. to determine the best moment to open or close a position. Not surprisingly, the shorter the forecasting horizon, the more emphasis is given by practitioners on technical analysis compared to fundamental analysis (Marshall et al. 2008).

Technical analysis includes the study of High-Low-Open-Close price dynamics (henceforth HLOC). This analysis can be done through the study of Japanese Candlesticks. A daily candlestick, also referred to as ‘single line’, is a graphical representation of the day’s opening, high, low and closing prices. As depicted in Figure 1, a typical candlestick exhibits a body (black/red or white/green) as well as upper and a lower shadows. The area between the open and the close is called the real body. Price movements above and below the real body are called shadows.

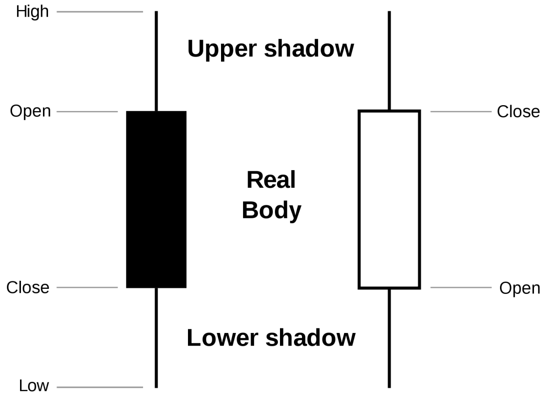


Figure 1: Candlesticks represent low, high, opening and closing prices of a day in one simple figure. The color indicates whether the close is above the open.

Single lines are said to have forecasting power. For example, a bullish (bearish) pattern is believed to lead to a future price increase (decrease). However, some single line candlesticks are regarded as more powerful than others. Some single lines are even said to have forecasting power regardless of the underlying trend in the market. In contrast, some other single lines require the existing trend to be identified. Consecutive single lines form continuation and reversal patterns.[1]

The performance of technical analysis has been first measured by Brock et al. (1992) (henceforth BLL) who use a bootstrap methodology on the DJIA index and find that moving averages and trading-range breaks generate statistically significant abnormal returns compared to four benchmark models. Bessembinder and Chan (1995) also test the profitability of technical analysis in Asian markets. The authors find evidence that their rules lead to abnormal returns and make the distinction between emerging and developed Asian markets, pointing out that returns are higher in emerging markets than in developed countries. Nevertheless, higher returns do not compensate for higher trading costs. Bessembinder and Chan (1998) confirm the results of BLL but argue that market efficiency cannot be rejected due to return measurement errors arising from nonsynchronous trading, among other things. Sullivan et al. (1999) go beyond the analysis of BLL by testing 7,846 trading rules with White's Reality Check bootstrap methodology which offers a better control for data-snooping biases (White 2000). They find that some of their trading rules perform even better than those of BLL. Hsu and Kuan (2005) study a more

[1] See Table 1 in Section 4.1 for the list of patterns covered in this paper.

complete universe of trading techniques on four main indices at the daily level. They apply the more robust Superior Predictive Ability (SPA) test (Hansen 2005) and find some evidence of profitability as well. More recently, Neuhierl and Schlusche (2012) and Shynkevich (2012) apply both the SPA test and its stepwise version developed by Hsu et al. (2010). While Hsu et al. (2010) and Neuhierl and Schlusche (2012) find some significant outperformance based on a set of technical and fundamental indicators respectively, Shynkevich (2012) find no evidence that technical analysis leads to profitable results in different segments of the US equity market.

The performance of Japanese candlesticks is studied in a few papers only. Marshall et al. (2006) apply the BLL bootstrap methodology on Japanese candlesticks and find no evidence of predictability on the DJIA stocks. Using White's Reality Check bootstrap methodology, Marshall et al. (2008) again find no evidence of predictability on the US equity market. Lu et al. (2012) take a long-term perspective by studying six bearing and bullish patterns at the daily level on a single tracker fund; they find mixed results. Although Fock et al. (2005) use intraday data on two futures contracts, they do not apply any bootstrapping methodology. As such, they do not correct for the luck factor.

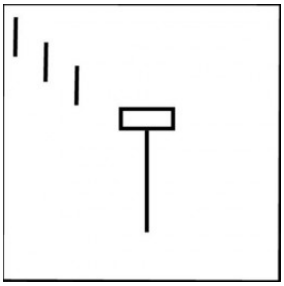
In this paper, we fill this gap and use the most up-to-date statistical tests on a data sample of intraday prices for the 30 DJIA stocks. We also apply the SSPA test for data snooping to measure the performance of each candlestick rule by applying double-or-out intraday strategies. Finally, we combine the best performing trading rules into fully automated trading systems. No previous study has tested the performance of such trading systems based on Japanese candlesticks using up-to-date statistical tools.

3 Data and methodology

The data sample is extracted from Bloomberg and includes 5-minute intraday HLOC prices from April 1, 2010 to April, 13 2011 for the 30 components of the DJIA index. Around 20,550 HLOC prices are available for each stock.

The 83 Japanese candlestick rules that we test and combine are defined in the TA-Lib MATLAB Toolbox. TA-Lib is an open-source technical analysis library which is widely used by trading software developers who perform technical analysis of financial market data. For each type of configuration and for each record, the TA-lib library returns "1" if the bullish part of the structure is identified, "-1" for the bearish part and "0" otherwise. As the structures are bullish, bearish or both, for each event type, the values that may appear are [0 ; 1], [-1 ; 0] or [-1 ; 0 ; 1]. The TA-lib allows some flexibility in the recognition of the configurations. As it is an open source C-code library, we have been able to check the parametrization of the structures. Events are recognized according to the standard flexibility rules. For instance, the Hammer configuration is identified when the real body is small, the lower shadow is long and, the body is near the lows of the previous candles (Figure 2). All these criteria are programmed in the C-code of the Ta-lib. The meaning of the words "small", "long" and "near the lows" is pre-programmed using the recommendations of Nison (1991), Nison (1994) and Morris (1995). We therefore do not modify the standard pattern recognition parameters used in the library. The list of candlestick rules tested in the paper is given in Table 1.

Figure 2: Hammer



The Hammer implies that control has shifted from sellers to buyers. By construction, it presents a long lower shadow and almost no upper shadow. The Hammer occurs at the end of a downtrend. This structure is said to be highly reversal.

We follow Brock et al. (1992) and study three return generating models: the random walk (RW), autoregressive (AR), and generalized autoregressive conditional heteroskedasticity in mean (GARCH-M) models. To obtain the parameters estimates and the associated p-values from the bootstrapped models, we proceed as follows.

1. If a generating model of returns is assumed (e.g. random walk, AR, GARCH-M), standardized residuals from the estimated model should be realizations of i.i.d. innovations under the H_0 hypothesis that this model is the correct one. Otherwise, the return itself is used. We follow Marshall et al. (2008) and include overnight returns.

2. Resampling with replacement of the estimated standardized residuals or direct returns is applied. The new price history is derived by considering resampled standardized residuals

as innovation or by considering directly resampled returns. The resulting time series owns the same length and underlying distribution of the original data.

3. The trading rule or strategy is applied on the generated time series to obtain statistics of interest.

4. Step 2 and 3 are repeated B times to obtain the empirical distribution of the statistics of interest (such as mean profit) of the trading strategy under the chosen return generating model. $B = 500$ is often chosen (Brock et al. 1992; Marshall et al. 2006, 2007).

5. The p-value is obtained by computing the fraction of generated statistics greater than the one obtained on the original data.

Under the null hypothesis, the statistic of interest obtained by the trading strategy is just explained by the return generating model. If the null is rejected, the return model fails to explain the performance of the trading strategy.

Standard bootstrapping has to be adapted to take into account the fact that candlestick signals are known to be valid over a maximum of 10 periods only (Morris 1995 and Marshall et al. 2006). In our market timing bootstrap methodology, positions are held for a given number of periods and profitability is then computed accordingly. The procedure is repeated for every candlestick rule. In addition, bootstrapping returns is not sufficient to measure the performance of candlestick-based trading rules. The four HLOC prices have to be simulated. As described in Marshall et al. (2006, 2007), (high-closing)/closing and (closing-low)/closing percentages have to be computed and bootstrapped. These simulated percentages are then added or subtracted to the simulated closing price to form simulated high and low prices. A similar process is used to generate simulated opening prices but it is resampled if it happens to be higher or lower than high and low values, respectively.

3.1 Superior Predictive Ability test

Data-snooping bias is a serious issue when a high number of hypotheses is tested on the same time series. To the best of our knowledge, the best algorithm to adjust for data-snooping bias is the Superior Predictive Ability (SPA) test. The SPA test compares the performance of one benchmark model to m alternative forecasting models, while adjusting explicitly for data-snooping. It is an improvement of White's Reality Check for data-snooping which is less powerful and more sensitive to the inclusion of irrelevant alternatives. These enhancements are done by studentizing the test statistic and by invoking a re-centered sample dependent null distribution based on the bootstrap methodology. Under the null hypothesis H_0, the benchmark is not inferior to any alternative forecast.

The SPA test requires the use of a loss function for a model k at a time $t = \{1, ..., T\}$. Considering a situation where a decision must be made h periods in advance and letting $\delta_{k,t-h},\ k = \{0, 1, ..., m\}$ be a finite set of possible decision rules, namely the model k and the return r_t, the loss function is formally defined as $L_{k,t} = L(r_t, \delta_{k,t-h})$. Forecasts are compared based on their expected loss $E[L_{k,t}(\xi_t, \delta_{k,t-h})]$. The kth trading rule $\delta_{k,t-1}$, which instructs a trader to take either a short position ($\delta = -1$), a long position ($\delta = 1$) or no position ($\delta = 0$) in an asset at time $t - 1$, leads to a profit $\pi_{k,t} = \delta_{k,t-1} r_t$. This formalism gives:

$$L(r_t, \delta_{k,t-h}) = -\delta_{k,t-1} r_t \tag{3.1}$$

As the null hypothesis is that the benchmark is not inferior to any alternative, the main variables of interest are the relative performance variables given by the model k compared to the benchmark: $d_{k,t} = L_{0,t} - L_{k,t},\ k = \{1, ..., m\}$. Provided that $\lambda_k = E(d_{k,t})$, the formal proposition of H_0 is:

$$H_0 : \max_{k=1,...,m} \lambda_k \leq 0 \tag{3.2}$$

The associated test statistic is defined as:

$$\tau_{H_0} = \max_{k=1,...,m} \frac{\sqrt{T}\bar{d}_k}{\hat{\omega}_{kk}} \tag{3.3}$$

with $\hat{\omega}^2_{kk}$ is a consistent estimate of ω^2_{kk} and where

$$\bar{d}_k = \frac{1}{T}\sum_{t=1}^{T} d_{k,t},\ \omega^2_{kk} = \lim_{T\to\infty} var(\sqrt{T}\bar{d}_k) \tag{3.4}$$

To obtain a sample-based distribution under the null hypothesis, the most adequate resampling method according to Hansen (2005) is the stationary bootstrap, developed by Politis and Romano (1994). This method relies on resampling the pseudo time-series of the relative performance vector $d_{k,t}$ by building new sample subseries of different lengths. The subseries length M is obtained by a geometric distribution of parameter Q and lengths are independent between them. For a specific subseries, the first element is randomly chosen, then, the $M - 1$ next elements in the original series are concatenated to obtain the subseries. Finally, this operation is repeated and subseries are concatenated to achieve the original time series size of T elements. Obviously, lengths are ideally small but sufficiently large to reflect the serial dependence in the $d_{k,t}$ time series. The resulting bootstrap samples for $d_{k,t}^{b}$ considering B bootstraps, $b = \{1, ..., B\}$ lead to the bootstrapped empirical distribution. The sample mean of each bootstrap and the variance estimation are computed as:

$$\begin{cases} \bar{d}_k^i = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} d_{k,t}^i, \ i = \{1, ..., B\} & (3.5) \\ \hat{\omega}_{kk}^2 = \dfrac{1}{B}\displaystyle\sum_{i=1}^{B}(\sqrt{T}\bar{d}_k^i - \sqrt{T}\bar{\bar{d}}_k)^2, \ \bar{\bar{d}}_k = \dfrac{1}{B}\displaystyle\sum_{i=1}^{B}\bar{d}_k^i & (3.6) \end{cases}$$

Under the null hypothesis, the distribution of T can be empirically determined considering:

$$\bar{Z}_k^i = \bar{d}_k^i - g_j(\bar{d}_k), \ j = l, c, u \tag{3.7}$$

where $g_l(x) = \max(0, x)$, $g_c(x) = x \times 1_{x_k \leq A_{k,c}}$ with $A_{k,c} = -T^{-\frac{1}{2}}\sqrt{2loglogT}\hat{\omega}_{kk}$ and $1_{\{\}}$ is an indicator function and $g_u(x) = x$. As there is no perfect estimation of the mean distribution, Hansen proposed Lower, Central and Upper p-value estimations. This re-centering ensures that irrelevant models do not asymptotically influence the distribution of the test statistic. The empirical distribution of T_{H_0} is obtained by:

$$T_{H_0}^i = \max_{k=1,...,m} \frac{\sqrt{T}\bar{Z}_k^i}{\tilde{\omega}_{kk}} \tag{3.8}$$

converges to the distribution of T_{H_0} under the null hypothesis. P-value is determined as:

$$\frac{1}{B}\sum_{i=1}^{B} 1_{\{T^i > T\}} \tag{3.9}$$

Intuitively, the p-value of a SPA test indicates the relative performance of a reference model in comparison with alternative models $k = \{1, .., m\}$. A high p-value means that the null hypothesis (according to which the base model is not outperformed) is not rejected.

However, the test is not applicable as such in a market timing framework. While the SPA test requires the same number of observations for each competing model or rule, market timing does not. By construction of the buy and sell signals, rules will typically exhibit a different number of observations in a market timing setting. To circumvent this problem, Bessembinder and Chan (1998) propose a double-or-out strategy which is a benchmark following strategy modulated by one additional market timing position. The buy-and-hold strategy is followed except when the short-time trading rule provides a signal. When no further signal is generated by the short-term rule for a maximum of 10 periods, the additional position is closed. Thereby, the number of open positions is always between 0 and 2. As outlined in Hsu and Kuan (2005) and Bessembinder and Chan (1998), another advantage of this strategy is to compare trading rules which may generate a significantly different number of signals.

3.2 Stepwise SPA extension

Practitioners are not only interested in knowing whether there is any significant trading rule but they also want to identify all such trading rules. This is the purpose of the Stepwise SPA test (SSPA). As defined in Hsu et al. (2010), the SSPA is a three-step procedure:

1. At step j, re-label all models in descending order of corresponding $T^k_{H_0}$, $T^k_{H_0}$ being the kth component of the T_{H_0} vector before maximization.
2. Reject individual model k if $T^k_{H_0} > q_j(\alpha_0)$ at the α_0 significance level and where $q_j(\alpha) = \max(\inf\{q_j | p - \text{value}(T_{H_0} = q_j) = \alpha\}, 0)$ at step j.
3. If none of the null hypotheses are rejected, the process stops. If the first $k_1(> 1)$ models are rejected in the second step, those models are removed from the data and the remaining models are the new original data leading to a modification of the T_{H_0} distribution for further steps. The process restarts at step 2 with $j = j + 1$.

3.3 Trading systems

Another original way to measure the performance of Japanese candlestick rules is to combine them in a trading system based on the double-or-out strategy. The trading system considers all possible combinations of buy and sell signals sent by every candlestick rule. Each rule may lead to a buy signal (+1), a sell signal (-1), or no signal (0). When combining the rules, we use a majority vote to normalize the buy and sell vector δ at each time t as follows: $\delta_t = \{-1, 0, 1\}$. The position is reversed ($\delta = -1$), maintained ($\delta = 0$), or reinforced ($\delta = +1$), respectively. A maximum number of c holding periods for the additional position is considered, with $1 < c \leq 10$. Finally, the vector of returns and Sharpe ratios are fed into the SPA test to determine whether one of the combinations is superior to the benchmark. If there is significant evidence of outperformance, the Stepwise SPA test can then identify each of the outperforming trading system.

Testing trading systems based on candlesticks is very resource intensive. If 100 candlestick rules are included in the system, we obtain $1.27e^{30}$ possible combinations per observation and per stock. Computation time would be very substantial indeed. To reduce the curse of dimensionality, we only include the best performing rules. In particular, we first identify the top ten double-or-out candlestick rules for each of the 30 stocks. These top ten rules must be significant at the Bonferroni level in the market timing application. Second, we retain only the rules that are listed at least twice on average across the 30 stocks. 11 candlestick patterns pass the filter. Third, these 11 candlestick patterns are combined in a double-or-out trading system in order to potentially detect profitable complex trading strategies. As a robustness check, the analysis is done with different parameter configurations.

4 Empirical results

We report the empirical results for the market timing application before looking at the performance of the trading systems.

4.1 Market timing

To test the intraday market timing performance of Japanese candlesticks, we proceed as follows. First, we enter the market at the closing price when one of the 83 tested candlestick signals comes out of the data. Second, we compute the trend over the past ten periods to determine if the transaction is a buy or a sell. This is required for candlestick signals that may be bullish or bearish depending on the past trend. Finally, we hold the position over the next ten periods, except if the same signal (bullish or bearish) is generated. In such a case, the holding period is extended by ten periods. In Table 1, we report some basic statistics for each of the 83 candlestick rules.

The number of trades varies significantly between candlestick rules. While three patterns (namely 'three stars in South', 'concealing baby swallow', and 'mat hold') do not lead to any transaction, the 'bullish long line' rule involves 22,459 trades over 630,000 observations approximately.

Rule No.	Candlestick rule (Bearish or bullish/No.)	NT		RPT		MR	
		OS	BS	OS	BS	OS	BS
1	2CROWS (Bearish/1)	12	57	0.417	0.494	-8.47E-05	-2.69E-06
2	3BLACKCROWS (Bearish/2)	70	2	0.514	0.502	1.29E-05	1.04E-05

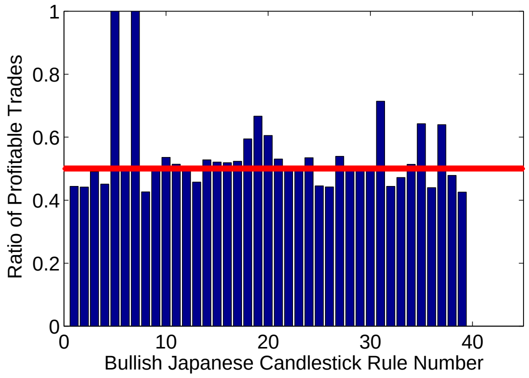
Rule No.	Candlestick rule (Bearish or bullish/No.)	NT		RPT		MR	
		OS	BS	OS	BS	OS	BS
3	3INSIDE (Bullish/1)	874	1571	0.444	0.503	-1.69E-05	4.06E-06
4	3INSIDE (Bearish/3)	906	794	0.500	0.496	2.17E-05	-4.91E-06
5	3LINESTRIKE (Bullish/2)	129	22	0.442	0.501	-9.43E-05	-7.60E-06
6	3LINESTRIKE (Bearish/4)	71	2	0.507	0.477	1.81E-04	-2.34E-05
7	3OUTSIDE (Bullish/3)	2016	1015	0.491	0.504	-1.03E-05	6.71E-06
8	3OUTSIDE (Bearish/5)	2989	3794	0.465	0.498	-1.72E-06	-2.80E-06
9	3STARSINSOUTH (9/)	0	0	NaN	NaN	NaN	NaN
10	3WHITESOLDIERS (Bullish/4)	641	24	0.451	0.504	-3.15E-06	3.41E-06
11	ABANDONEDBABY (Bullish/5)	3	8	1.000	0.507	9.47E-05	1.45E-05
12	ABANDONEDBABY (Bearish/6)	3	11	0.333	0.489	5.17E-05	-2.05E-05
13	ADVANCEBLOCK (Bearish/7)	2305	281	0.480	0.495	6.66E-06	-5.53E-06
14	BELTHOLD (Bullish/6)	20257	10702	0.498	0.516	-5.87E-07	4.74E-06
15	BELTHOLD (Bearish/8)	20624	10409	0.478	0.507	-1.33E-06	-4.83E-06
16	BREAKAWAY (Bullish/7)	1	0	1.000	0.403	1.53E-04	-8.96E-05
17	BREAKAWAY (Bearish/9)	2	1	1.000	0.503	2.19E-04	-6.10E-05
18	CLOSINGMARUBOZU (Bullish/8)	15683	13545	0.426	0.487	-4.91E-06	5.24E-06
19	CLOSINGMARUBOZU (Bearish/10)	15061	12651	0.396	0.481	-1.86E-05	-4.72E-06
20	CONCEALBABYSWALL (20/)	0	0	NaN	NaN	NaN	NaN
21	COUNTERATTACK (Bullish/9)	85	940	0.494	0.503	6.04E-05	3.09E-06
22	COUNTERATTACK (Bearish/11)	73	883	0.493	0.499	-3.36E-05	-2.06E-06
23	DARKCLOUDCOVER (Bearish/12)	462	389	0.461	0.496	-5.13E-05	-4.04E-06
24	DOJISTAR (Bullish/10)	1291	8850	0.536	0.518	4.57E-05	4.03E-06
25	DOJISTAR (Bearish/13)	1530	9585	0.488	0.511	5.73E-06	-4.98E-06
26	DRAGONFLYDOJI (Bullish/11)	6464	15105	0.514	0.528	1.39E-05	4.48E-06
27	ENGULFING (Bullish/12)	4196	1956	0.493	0.505	-8.14E-06	5.70E-06
28	ENGULFING (Bearish/14)	6388	7048	0.477	0.503	1.09E-06	-3.45E-06
29	EVENINGDOJISTAR (Bearish/15)	290	516	0.462	0.496	-4.81E-06	-4.54E-06
30	EVENINGSTAR (Bearish/16)	735	685	0.454	0.495	-1.98E-05	-4.79E-06
31	GAPSIDESIDEWHITE (Bullish/13)	774	2956	0.457	0.497	-7.32E-07	3.39E-06
32	GAPSIDESIDEWHITE (Bearish/17)	351	3099	0.402	0.494	-3.76E-05	-1.94E-06
33	GRAVESTONEDOJI (Bullish/14)	6835	14826	0.528	0.526	1.57E-05	4.60E-06
34	HAMMER (Bullish/15)	9184	10798	0.521	0.522	2.76E-06	4.90E-06
35	HANGINGMAN (Bearish/18)	7604	10177	0.510	0.512	1.24E-05	-5.77E-06
36	HARAMI (Bullish/16)	4215	7024	0.519	0.508	9.59E-06	3.96E-06
37	HARAMI (Bearish/19)	4469	7297	0.505	0.502	1.92E-05	-4.12E-06
38	HARAMICROSS (Bullish/17)	1605	5255	0.523	0.507	1.25E-05	3.84E-06
39	HARAMICROSS (Bearish/20)	1767	5450	0.498	0.502	2.25E-05	-3.70E-06
40	HIKKAKEMOD (Bullish/18)	37	165	0.595	0.506	5.32E-05	5.08E-06
41	HIKKAKEMOD (Bullish/19)	9	26	0.667	0.511	4.76E-05	1.37E-05
42	HIKKAKEMOD (Bearish/21)	29	165	0.517	0.497	3.62E-05	-7.28E-06
43	HIKKAKEMOD (Bearish/22)	12	35	0.250	0.494	-2.51E-04	-1.20E-05
44	HOMINGPIGEON (Bullish/20)	185	1480	0.605	0.506	7.99E-05	6.06E-06
45	IDENTICAL3CROWS (Bearish/23)	771	2	0.429	0.514	-3.96E-05	2.07E-05

Rule No.	Candlestick rule (Bearish or bullish/No.)	NT		RPT		MR	
		OS	BS	OS	BS	OS	BS
46	INNECK (Bearish/24)	430	628	0.447	0.496	-2.36E-05	-4.87E-06
47	INVERTEDHAMMER (Bullish/21)	2453	10592	0.531	0.523	2.28E-05	4.75E-06
48	KICKING (Bullish/22)	2	1	0.500	0.509	1.37E-05	-1.52E-05
49	KICKING (Bearish/25)	3	2	0.000	0.519	-5.37E-04	1.01E-05
50	KICKINGBYLENGTH (Bullish/23)	2	1	0.500	0.489	1.37E-05	-1.94E-05
51	KICKINGBYLENGTH (Bearish/26)	3	2	0.000	0.501	-5.37E-04	1.77E-06
52	LADDERBOTTOM (Bullish/24)	43	35	0.535	0.506	1.67E-05	9.41E-06
53	LONGLINE (Bullish/25)	22459	22672	0.445	0.505	-6.87E-06	5.09E-06
54	LONGLINE (Bearish/27)	22373	22292	0.430	0.496	-7.73E-06	-4.73E-06
55	MARUBOZU (Bullish/26)	11167	4468	0.442	0.500	-3.64E-06	4.93E-06
56	MARUBOZU (Bearish/28)	10317	4093	0.413	0.492	-1.41E-05	-5.45E-06
57	MATCHINGLOW (Bullish/27)	3691	3607	0.539	0.506	6.38E-05	3.60E-06
58	MATHOLD (58/)	0	0	NaN	NaN	NaN	NaN
59	MORNINGDOJISTAR (Bullish/28)	296	495	0.493	0.502	1.82E-05	3.42E-06
60	MORNINGSTAR (Bearish/29)	727	658	0.509	0.502	3.82E-05	4.46E-06
61	ONNECK (Bullish/29)	441	1944	0.438	0.495	-4.75E-05	-3.08E-06
62	PIERCING (Bearish/30)	380	375	0.508	0.502	2.83E-05	3.45E-06
63	RISEFALL3METHODS (Bearish/31)	14	0	0.714	0.544	1.20E-04	2.94E-05
64	RISEFALL3METHODS (Bullish/30)	14	4	0.643	0.500	1.61E-04	-3.91E-06
65	SEPARATINGLINES (Bearish/32)	865	136	0.444	0.505	-2.06E-05	6.79E-06
66	SEPARATINGLINES (Bullish/31)	2991	381	0.432	0.497	-2.16E-05	-4.20E-06
67	SHOOTINGSTAR (Bullish/32)	2449	11091	0.477	0.515	-9.50E-06	-6.07E-06
68	STALLEDPATTERN (Bearish/33)	1038	256	0.491	0.497	2.88E-05	-5.62E-06
69	STICKSANDWICH (Bullish/33)	341	196	0.472	0.501	1.57E-05	3.26E-06
70	TAKURI (Bearish/34)	6419	15421	0.514	0.528	1.25E-05	4.55E-06
71	TASUKIGAP (Bearish/35)	42	230	0.643	0.506	1.07E-04	7.30E-06
72	TASUKIGAP (Bullish/34)	40	207	0.450	0.493	-5.42E-05	-6.74E-06
73	THRUSTING (Bullish/35)	878	651	0.459	0.496	-2.79E-05	-4.31E-06
74	TRISTAR (Bullish/36)	75	4074	0.440	0.508	-3.98E-05	4.62E-06
75	TRISTAR (Bearish/36)	71	4000	0.423	0.502	-1.83E-05	-4.40E-06
76	UNIQUE3RIVER (Bullish/37)	25	183	0.640	0.507	5.42E-05	6.57E-06
77	UPSIDEGAP2CROWS (Bearish/37)	2	13	0.000	0.495	-4.42E-05	-9.14E-06
78	XSIDEGAP3METHODS (Bullish/38)	282	54	0.479	0.503	-3.00E-05	6.13E-06
79	XSIDEGAP3METHODS (Bearish/38)	283	50	0.431	0.495	-6.04E-05	-7.16E-06
80	DRAGONFLYDOJIneg (Bearish/39)	7545	15257	0.502	0.519	5.64E-06	-4.95E-06
81	GRAVESTONEDOJIneg (Bearish/40)	6990	14986	0.492	0.518	9.91E-07	-5.21E-06
82	OPENINGMARUBOZU (Bullish/39)	16010	10855	0.425	0.492	-1.60E-06	5.17E-06
83	OPENINGMARUBOZU (Bearish/41)	15647	10268	0.395	0.485	-1.65E-05	-4.53E-06

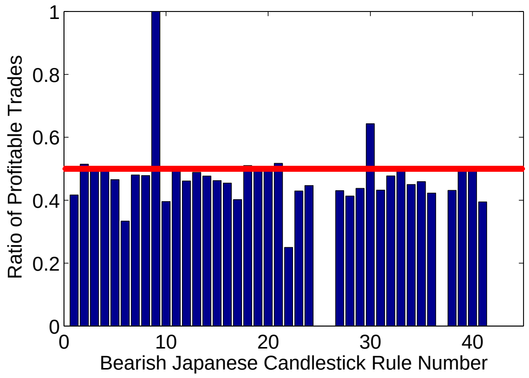
Table 1: Candlestick signals: Number of trades (NT) summed over all equities, ratio or profitable trades (RPT), and mean return (MR) for both the original and GARCH-M bootstrapped series (OS and BS, respectively).

Table 1 and Figure 3 indicate that the number of profitable trades does not cross the 50%

'luck' threshold on average very often. In fact, 56% of bullish patterns (i.e. 22 over 39 rules) and only 22% of bearish patterns (i.e. 9 over 41 rules) pass the threshold. If we exclude the candlestick rules which have a very low number of trades (such as the 'breakaway' rule), fewer candlestick rules would even pass the test.



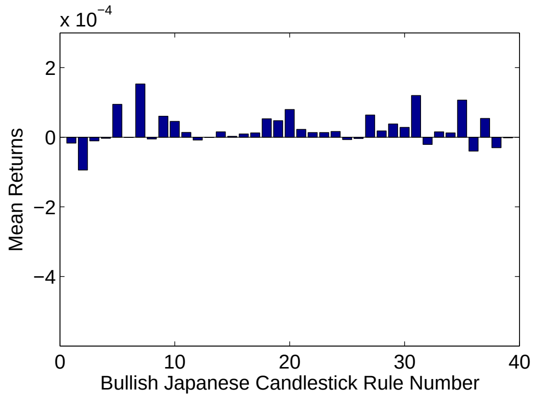
(a) Profitability ratios in market timing of bullish candlesticks



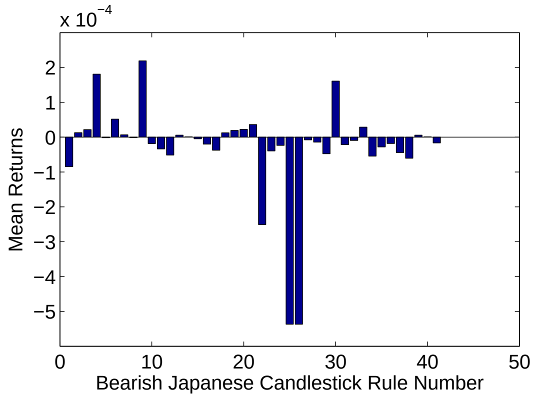
(b) Profitability ratios in market timing of bearish candlesticks

Figure 3: Figure (a) and Figure (b) respectively show the ratio of profitable trades for bullish and bearish candlestick rules.

The mean returns of each candlestick rule are reported in Table 1 and summarized in Figure 4. 64% of bullish patterns (i.e. 25 over 39 rules) and 39% of bearish patterns (i.e. 16 over 41 rules) deliver a positive mean return. All in all, bullish signals seem to perform better than their bearish counterparts. The general market trend may explain such this asymmetry. From early April 2010 to late June 2010 (i.e. the first 3 months), the DJIA index was down by around 11 %. However, the index then rose by around 28% for the next 9 months.



(a) Mean returns in market timing of bullish candlesticks



(b) Mean returns in market timing of bearish candlesticks

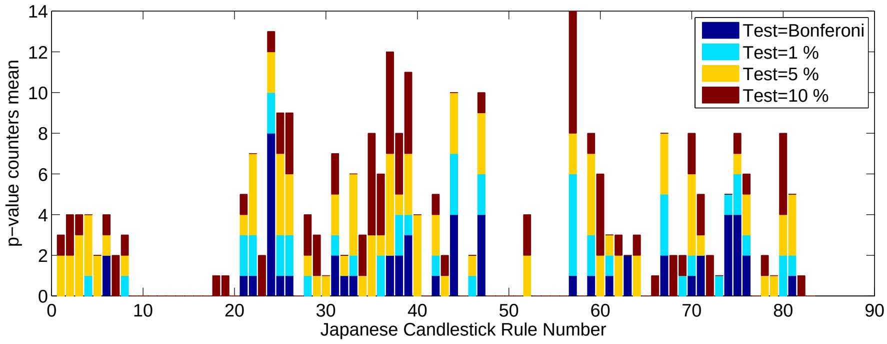
Figure 4: Figure (a) and Figure (b) show the mean returns for bullish and bearish candlestick rules respectively.

4.1.1 Bootstrapped results

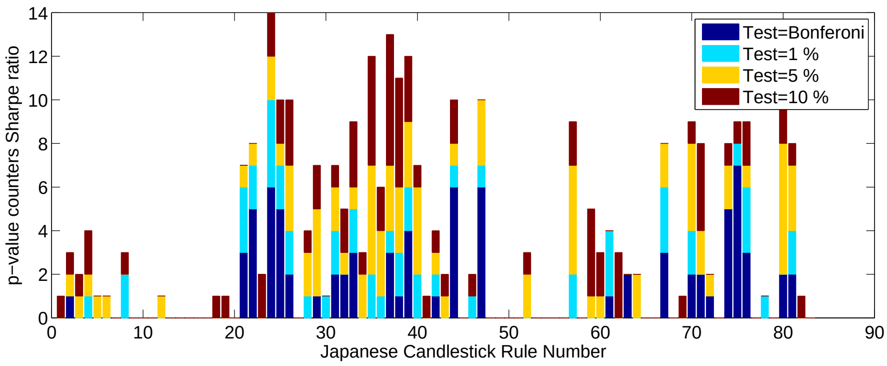
The number of trades on the GARCH-generated bootstrapped series is obtained by summing across all DJIA stocks the mean number of signals per bootstrapped series for each individual stock. Table 1 indicates that the number of trades on each original Dow stock series (OS) is not always consistent with the average number of signals per bootstrap series for each stock (BS).

In Figure 5, we report the p-value counters based on the significance of the mean return and the Sharpe ratio for each of the 83 Japanese candlestick rules. The counters indicate the number of stocks for which the mean return (Figure 5a) and the Sharpe ratio (Figure 5b) are statistically greater on the original series than on the bootstrapped series. The three classical levels of statistical significance are considered, i.e. the 1%, 5% and 10% levels. We also include a test at the 5 % level, which integrates the Bonferroni correction useful to counteract the problem of multiple comparisons. For example, the mean return obtained by rule n°57 is statistically significant for 1, 6, 8, and 14 stocks at, respectively, the Bonferroni, 1%, 5% and 10% levels. The Sharpe ratio of rule n°57 is statistically significant for 2, 7, and 9 stocks at, respectively, the 1%, 5% and 10% levels.

It appears that a relatively high number of candlestick rules are significant. Even at the conservative Bonferroni level, 26 and 27 rules are significant based on mean returns and Sharpe ratios respectively, pointing to robust results with respect to data-snooping.



(a) Mean return - p-value counters



(b) Sharpe Ratio - p-value counters

Figure 5: Figure (a) and Figure (b) indicate the p-value counters of the mean return and the Sharpe ratio respectively for each of the 83 Japanese candlestick rules.

4.1.2 Robustness checks

We first test the sensitivity of our results to each key parameter used in the trading strategies (Figure 6). The basic scenario implies that trades are entered at the closing price, positions are held for 10 periods (i.e. 50 minutes), and the previous trend is computed over the last 10 periods (Scenario 5). In Scenario 1, positions are instead entered at the following opening price.

In Scenarios 2 and 6, the previous trend is computed over the last 5 and 15 periods, respectively. In Scenarios 3 and 4, the holding period is changed to 2 and 5 periods, respectively. At this stage, the return generating model is still the GARCH-M model. As shown in Figure 6, there is no significant impact on the average p-value counters for the mean return or the Sharpe ratio, whatever scenario is considered. Interestingly, risk adjustment through the Sharpe ratio does not worsen the overall picture since its p-value counter is on average never below the p-value counter of the mean return.

We also study the robustness of our results to the chosen return generating model. The p-value counter is computed for the Random Walk (Scenario 1), Auto-Regressive (Scenario 2) and GARCH-M models (Scenario 3). As shown in Figure 7, there is again no significant difference between models when the standard trading parameters are used.

4.1.3 Cost analysis

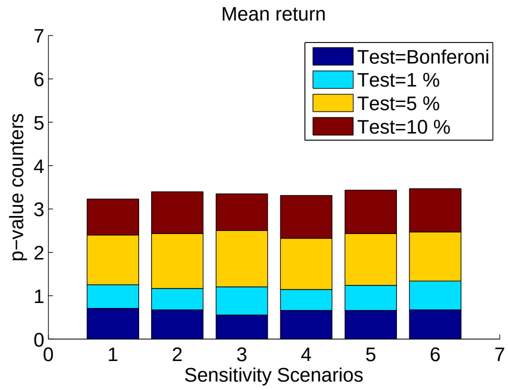
Trading costs are typically hard to evaluate. McSheery (2011) find that overall US Institutional equity trading costs are close to 50 basis points, including commissions, fees, and implementation shortfall costs. One-way brokerage commission averages in the US alone would be 7.5 basis points approximately. McSheery (2010) also estimate that US equity composite (all trading) commissions were 2.27 cents in 1Q2010. Given the median and average NYSE stock prices of 15.87$ and 53.19$ respectively in August 2011, trading commissions would be between 14.3 and 4 basis points. In the SSPA tests, one-way trading commissions are estimated at the 5 basis point conservative level.

Following Bessembinder and Chan (1995), we also determine the level of trading costs that eliminate the ex post difference between cumulative returns to traders using the candlestick rules and cumulative returns to traders using the buy-and-hold strategy. The so-called breakeven one-way trading costs (BEC) is computed as follows:

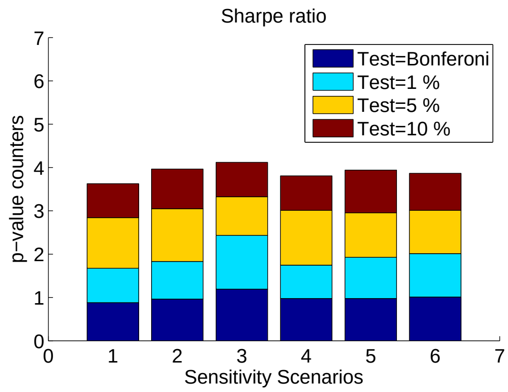
$$\mathrm{BEC} = \frac{\pi}{2(N)} \tag{4.1}$$

where π is the total trade profit derived from the active trading strategy and N is the total number of signals.

When compared to the average of BECs across candlestick rules, trading profits are eroded in almost every case. Only 5 out of 83 rules exhibit a mean profit higher than the average



(a) Sensitivity analysis of the mean return w.r.t. trading parameters



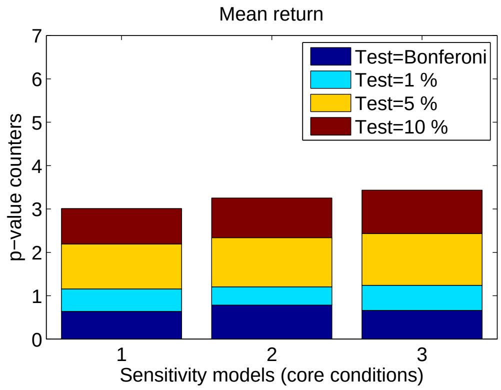
(b) Sensitivity analysis of the Sharpe ratio w.r.t. trading parameters

Figure 6: Figure (a) and Figure (b) display the average p-value counters for the mean return and Sharpe ratio respectively based on five trading scenarios. Scenario 5 is the base scenario: trades are entered at the closing price, positions are held for 10 periods, and the previous trend is computed over the last 10 periods. In Scenario 1, positions are instead entered at the following opening price. In Scenarios 2 and 6, the previous trend is computed over the last 5 and 15 periods, respectively. In Scenarios 3 and 4, the holding period is changed to 2 and 5 periods, respectively. The return generating model is the GARCH-M model in all scenarios.

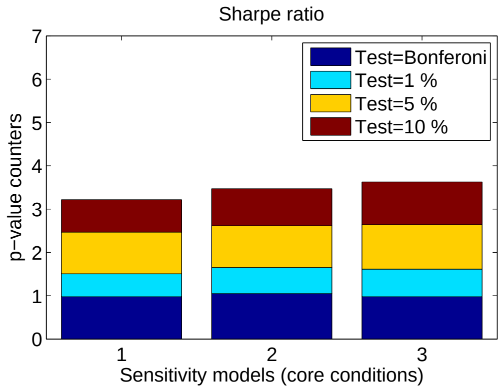
trading cost per trade, namely 3LINESTRIKE (Bearish/4), ABANDONEDBABY (Bullish/5), BREAKAWAY (Bullish/7), BREAKAWAY (Bearish/9), and RISEFALL3METHODS (Bullish/30). From a risk-adjusted point of view, there are three significant Japanese candlestick patterns only.

4.2 SSPA test

We also correct for the data snooping bias by using the Superior Predictive Ability (SPA) test and its stepwise version (SSPA) which enables the identification of every significant rule that



(a) Sensitivity analysis of the mean return w.r.t the return generating model (standard trading parameters)



(b) Sensitivity analysis of the Sharpe ratio w.r.t the return generating model (standard trading parameters)

Figure 7: Figure (a) and Figure (b) show the average p-value counters for the mean return and Sharpe ratio respectively based on three return generating models, i.e Random Walk (Scenario 1), Autoregressive of order 1 (Scenario 2) and GARCH-M (Scenario 3). The standard trading parameters are used: trades are entered at the closing price, positions are held for 10 periods, and the previous trend is computed over the last 10 periods.

beats the buy-and-hold benchmark strategy. The SSPA test is nevertheless not applicable when market timing rules lead to different number of observations. To circumvent this problem, we develop a double-or-out strategy as described in Bessembinder and Chan (1998) and detailed in Section 3.1.

To detect whether candlestick rules beat the buy-and-hold strategy, five-minute log returns $(r_{k,t})$ and the corresponding Sharpe ratios $(s_{k,t})$ are fed in the SSPA algorithm as explained in Section 3.1. In a double-or-out strategy,

$$r_{k,t} = D_{k,t} ln(\frac{D_{k,t}\pi_{k,t} + P_{t-1}}{P_{t-1}}) \tag{4.2}$$

where $D_{k,t}$ is a dummy variable of candlestick rule k (+1 for long positions and -1 for short positions) and $\pi_{k,t}$ is the profit delivered by rule k at time t. It is computed as follows:

$$\pi_{k,t} = D_{k,t}(P_t - BEP_{k,t-1}) + (P_t - P_{t-1} - Cost_t) \tag{4.3}$$

where $BEP_{k,t}$ is the break-even price at time t for rule k (i.e. the stock price at time t leading to no profit for rule k) and $Cost_t$ is the trading cost at time t.

Finally, the periodic Sharpe ratio $s_{k,t}$ is defined as:

$$s_{k,t} = \frac{r_{k,t}}{\sigma_k} \tag{4.4}$$

where σ_k is the return volatility of rule k.

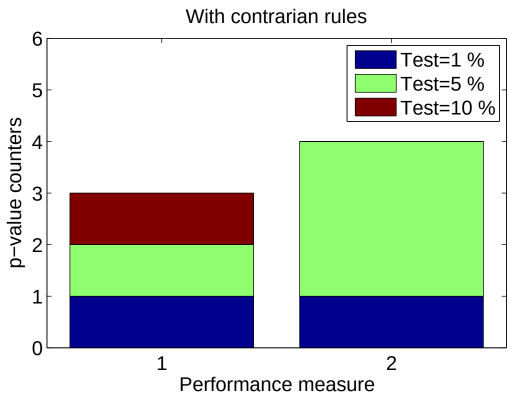
We define several sensitivity parameters. First, two trading profiles are assessed, i.e. aggressive and conservative. Aggressive traders submit a market order at the closing price as they anticipate the signal and the related period close. Alternatively, conservative traders wait for the signal to be completed before submitting a market order at the next opening price. Second, three different holding periods are defined. The base parameter is 10. The two alternatives are 2 and 5. When the same pattern is detected during the holding period, the period is extended accordingly. Third, if the detection of a specific Japanese candlestick pattern requires the trend to be identified, the Exponential Moving Average (EMA) is used. A bearish trend is identified when EMA(t) > close(t), and vice versa. The EMA is computed in different ways: 5 to 15 observations are included. Those sensitivity parameters are assessed in one single SSPA test to avoid data-snooping bias that would result from parameter optimization.

Contrarian rules are also investigated. As poor or irrelevant rules can decrease the power of the SPA test (albeit to a lesser extent than in White's test), we follow the suggestion in Hansen (2005) and estimate the sensitivity of the SPA test by running it twice, with and without contrarian rules.

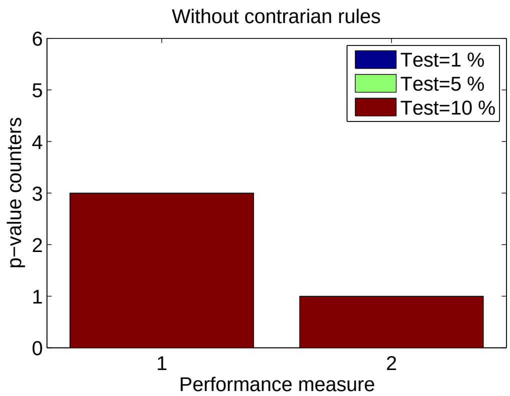
When the gross mean return is used as the performance measure, Figure 8 shows that there is at least one candlestick rule beating the buy-and-hold strategy in 10% of the cases (i.e. 3 stocks out of 30). When contrarian rules are included, the p-value counter is still equal to three but the significance level is improved, being equal to 5% for two stocks out of three. Based on the Sharpe ratio, Figure 8 also shows that the inclusion of contrarian rules leads to a higher

p-value counter, going up from 1 to 4 with an improved level of significance as well. Whatever the performance measure considered, there is at least one stock for which candlestick rules beat the buy-and-hold strategy.

Figure 9 gives the mean number of the outperforming candlestick rules across the 30 DJIA stocks. When contrarian rules are included, around one candlestick rule per stock outperforms the buy-and-hold strategy. If contrarian rules are excluded, the number is closer to zero.



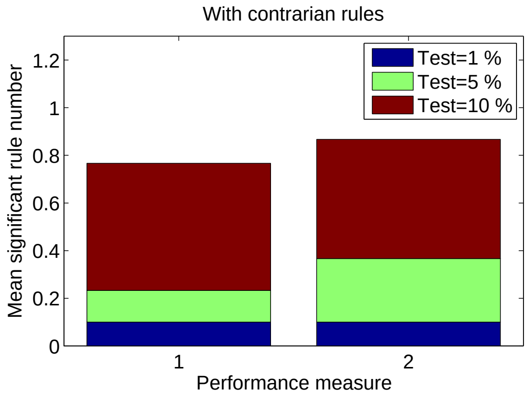
(a) p-value counters in a double-or-out strategy including contrarian rules



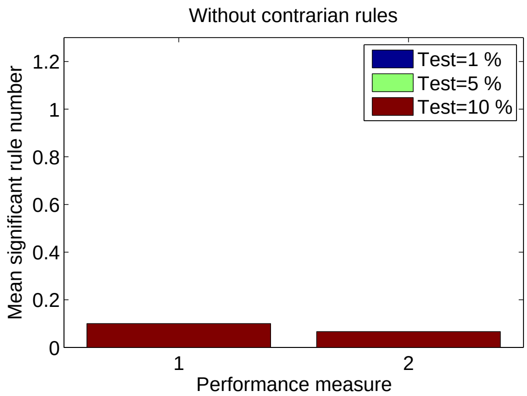
(b) p-value counters in a double-or-out strategy without contrarian rules

Figure 8: Figure (a) and Figure (b) respectively show the number of stocks for which there is at least one candlestick rule beating the buy-and-hold strategy based on the gross mean return (1) or the Sharpe ratio (2) when contrarian rules are included or not.

In Table 2, we identify the candlestick rules which outperform the buy-and-hold strategy at least once across the 30 stocks, based either on the mean return or Sharpe ratio, and at the 10% significance level at worst.



(a) Number of significant rules in a double-or-out strategy including contrarian rules



(b) Number of significant rules in a double-or-out strategy without contrarian rules through the 30 equities

Figure 9: Figure (a) and Figure (b) respectively show the average number of significant rules across the 30 equities when the gross mean return (1) or the Sharpe ratio (2) is considered and when contrarian rules are included or not.

With contrarian rules	Without contrarian rules
CLOSINGMARUBOZU	GRAVESTONE DOJI
GRAVESTONENEGATIVE	HANGINGMAN
LONGLINE	HARAMICROSS
MARUBOZU	
OPENINGMARUBOZU	

Table 2: Names of the candlestick rules which outperform the buy-and-hold strategy when contrarian rules are included or not.

Most interestingly, no single candlestick rule beats the buy-and-hold strategy when conservative trading costs (at 0.05 %) are taken into account.

4.3 Trading Systems

To design the trading system and reduce the curse of dimensionality as explained in Section 3.3, we first keep a maximum of 10 double-or-out rules per stock, selecting only the rules which are significant at the Bonferroni level *after trading costs*. Table 3 identifies the rules that pass the filter. In a second step, we only retain the rules that are listed at least twice on average across the 30 stocks. In fact, 7 rules pass the filter. They are indicated in italics in Table 3. Among these rules, we can identify some two-candle patterns and some rules based on gaps. Gaps are potentially informative since they correspond to price jumps. Since 4 out of these 7 rules can be bearish or bullish (as indicated in Table 1), we can include up to 11 patterns in the trading system. If we combine them all in a double-or-out trading system in order to potentially detect profitable complex trading strategies (as explained in Section 3.3), the total number of trading rule combinations is equal to 2,047 (= 2^{11} -1 rules), plus all the sensitivity scenarios. In the SSPA test, we also include all the previous market timing strategies to mitigate the risk of data snooping resulting from the fact that both the trading systems and the individual candlestick rules are tested on the same dataset.

Top-ten ranked rules	
3BLACKCROWS	MATCHINGLOW
3LINESTRIKE	MORNINGDOJISTAR
ADVANCEBLOCK	ONNECK
COUNTERATTACK	PIERCING
DARKCLOUDCOVER	RISEFALL3METHODS
DOJISTAR	SEPARATINGLINES
EVENINGDOJISTAR	*SHOOTINGSTAR*
EVENINGSTAR	STALLEDPATTERN
GAPSIDESIDEWHITE	STICKSANDWICH
HARAMICROSS	TASUKIGAP
HIKKAKEMOD	THRUSTING
HOMINGPIGEON	*TRISTAR*
INNECK	UNIQUE3RIVER
KICKING	XSIDEGAP3METHODS
LADDERBOTTOM	

Table 3: This figure presents the ten-best ranked rules per equity when trading costs are considered.

As in Section 4.2, two SSPA tests are performed. The first test excludes the contrarian rules. Over the 12,116 tested rules *on average per stock* (including the sensitivity scenarios), none is found to outperform the buy-and-hold strategy after trading costs. The second test includes the contrarian rules and deal with 24,232 rules. The key conclusion holds: when trading cots are taken into account, no complex strategy delivers statistically higher economic performance than the buy-and-hold strategy, even at the 10 % significance level.

5 Conclusion

Although Japanese candlesticks are extensively used by practitioners nowadays, the intraday predictive power of Japanese candlestick rules has not yet been seriously tested in the existing literature. When luck, risk or trading costs are not measured correctly, the illusion of intraday outperformance leads investors to blindly apply strategies that are doomed to failure. This paper fills this gap by testing 83 Japanese candlestick rules at the 5-minute interval on the 30 components of the DJIA index. To determine whether the statistical and economic performance of Japanese candlestick rules is spurious or not, we design both market timing strategies and trading systems.

Market timing strategies are tested against the buy-and-hold strategy by relying on the bootstrap methodology and assuming different return generating models, such as the random walk, AR(1), and GARCH-in-mean processes. Statistical testing shows that some Japanese candlesticks have significant explanatory power. Even at the conservative Bonferroni level, 26 and 27 rules (out of 83) are significant based on mean returns and Sharpe ratios respectively. Whatever the parameter configuration and the underlying return generating model, no real difference is detected, pointing to robust results. When trading costs are included, trading performance is very much eroded in the vast majority of Japanese candlestick rules. Only five out of 83 rules exhibit a higher average profit than the average trading cost per trade. From a risk-adjusted point of view, there are three significant Japanese candlestick patterns only.

We also correct for data snooping by using the SSPA test, i.e. the stepwise extension of the Superior Predictive Ability test which enables the identification of every significant rule that beats the buy-and-hold benchmark strategy. As the SSPA test is not applicable when market timing rules lead to different number of observations, we apply a double-or-out strategy which consists in buying and holding the underlying asset modulated by one additional market timing position. If we exclude contrarian rules, three candlestick rules outperform the buy-and-hold strategy at least once across the 30 stocks. If contrarian rules are included instead, five candlestick rules are identified. Most interestingly, no single candlestick rule beats the buy-and-hold strategy when trading costs are taken into account.

Finally, we design automated trading systems in three steps. First, we select a maximum of ten double-or-out rules for each of the 30 DJIA stocks. To be selected, these rules must be significant at the Bonferroni level. Second, we retain only the rules that are listed at least twice on average across the 30 stocks. This gives 11 candlestick patterns which are finally combined in 2,047 different trading systems. As a robustness check, we also include contrarian rules and use different parameter configurations. Over the 24,232 rules *on average per stock* (including the 83 original market timing strategies), no evidence of statistical outperformance is found when trading costs are taken into account. An interesting avenue for future research would nevertheless consist in using a more flexible approach based on stop losses, instead of using a maximum number of 10 holding periods as suggested by Morris (1995).

While we hold the view that Japanese candlesticks can somewhat predict intraday returns, we show that such predictive power is too limited for active portfolio management to outperform the buy-and-hold strategy. How can we then reconcile our findings with the widespread continued use of active trading rules? In a seminal paper, Grossman and Stiglitz (1980) show that it is impossible for markets to be informationally efficient in a world with costly information. If we believe prices are right, practitioners would not spend time and energy researching securities so that prices would eventually fail to reflect the intrinsic value of the securities. As a consequence, there will always exist pockets of inefficiency. Our findings imply that there is just enough mispricing on large US caps to tempt practitioners into actively trading securities. However, we show that such mispricing is insufficient for the active investors to beat their passive counterparts when luck, risk, *and* trading costs are taken into account.

6 Acknowledgements

We gratefully acknowledge the detailed and insightful suggestions from the editor and two anonymous referees. The usual disclaimer applies.

References

Bessembinder, H. and K. Chan (1995). The profitability of technical trading rules in the asian stock markets. *Pacific-Basin Finance Journal 3*(2-3), 257–284.

Bessembinder, H. and K. Chan (1998). Market efficiency and the returns to technical analysis. *Financial Management 27*(2), 5–17.

Brock, W., J. Lakonishok, and B. LeBaron (1992, December). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance 47*(5), 1731–64.

Fock, J., C. Klein, and B. Zwergel (2005). Performance of candlestick analysis on intraday futures data. *Journal of Futures Markets 13*(1), 28–40.

Grossman, S. J. and J. Stiglitz (1980). On the impossibility of informationally efficient markets. *American Economic Review 70*(3), 393–408.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics 23*(4), 365–380.

Hsu, P.-H., Y.-C. Hsu, and C.-M. Kuan (2010). Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. *Journal of Empirical Finance 17*(3), 471–484.

Hsu, P.-H. and C.-M. Kuan (2005). Reexamining the profitability of technical analysis with data snooping checks. *Journal of Financial Econometrics 3*(4), 606–628.

Lu, T.-H., Y.-M. Shiu, and T.-C. Liu (2012). Profitable candlestick trading strategies - the evidence from a new perspective. *Review of Fianncial Economics 21*(2), 63–68.

Marshall, B., M. Young, and L. Rose (2006). Candlestick technical trading strategies: Can they create value for investors? *Journal of Banking and Finance 30*(8), 2303 – 2323.

Marshall, B., M. Young, and L. Rose (2007). Market timing with candlestick technical analysis. *Journal of Financial Transformation 20*(1), 18–25.

Marshall, B. R., R. H. Cahan, and J. M. Cahan (2008). Does intraday technical analysis in the u.s. equity market have value? *Journal of Empirical Finance 15*(2), 199 – 210.

McSheery, E. (2010). Equity brokerage commission averages increase slightly in the major markets. *Newsletter 7*(2), 4.

McSheery, E. (2011). High frequency trading volumes continue to increase throughout the world. *Newsletter 8*(3), 5.

Morris, G. L. (1995). *Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures*. McGraw-Hill.

Neuhierl, A. and B. Schlusche (2012). Data snooping and market-timing rule performance. *Journal of Financial Econometrics 9*(3), 550–587.

Nison, S. (1991). *Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Technique of the Far East*. New-York: New York Institute of Finance.

Nison, S. (1994). *Beyond Candlesticks: New Japanese Charting Techniques Revealed*. New-York: John Wiley and Sons.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Shynkevich, A. (2012). Performance of technical analysis in growth and small cap segments of the us equity market. *Journal of Banking and Finance 36*(1), 193–208.

Sullivan, R., A. Timmermann, and H. White (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance 54*(5), 1647–1691.

White, H. (2000). A reality check for data snooping. *Econometrica 68*(5), 1097–1126.