Trading the Foreign Exchange Market with Technical Analysis and Bayesian Statistics

October 2018

Arman Hassanniakalager*, Georgios Sermpinis, Charalampos Stasinakis*****

In this study, the profitability of technical analysis and Bayesian Statistics in trading the EUR/USD, GBP/USD, and USD/JPY exchange rates are examined. For this purpose, seven thousand eight hundred forty-six technical rules are generated, and their profitability is assessed through a novel data snooping procedure. Then, the most promising rules are combined with a Naïve Bayes, a Relevance Vector Machine, a Dynamic Model Averaging, a Dynamic Model Selection and a Bayesian regularised Neural Network model. The findings show that technical analysis has value in foreign exchange trading, but the profit margins are small. On the other hand, Bayesian Statistics seems to increase the profitability of technical rules up to five times.

Keywords: Trading; Technical Analysis; Foreign Exchange; Bayesian Averaging; Relevance Vector Machines;

*** School of Management, University of Bath, Bath BA2 7AY, United Kingdom (ah2493@bath.ac.uk)**

**** Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ, United Kingdom (Georgios.Sermpinis@glasgow.ac.uk)**

***** Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ, United Kingdom (Charalampos.Stasinakis@glasgow.ac.uk)**

1. Introduction

Technical analysis is the study of past market data in order to forecast the direction of financial asset prices. Its origins can be traced back to the Dow theory in 1900 when Charles H. Dow argued that the financial markets follow repetitive trends. Practitioners apply this principle in practice and many technical trading rules were developed over the next decades aiming to identify the future direction of financial assets. An industry was created based on the application of mathematics in trading. Today thousands of professionals trade financial series with mathematical models.

The most heavily traded assets are Foreign eXchange (FX) with a turnover of up to $5.3 trillion daily in 2013 (Jorion, 1996; BIS, 2013). The enormous size of the FX market, the competition among market participants and the advent of technology have led to a continuous search for more advanced and complex trading rules. Researchers and practitioners borrow algorithms from mathematics, physics, genetics and computer science in an attempt to model series that have a non-linear and non-stationary structure. Some apply simple technical rules (Gençay *et al*. 2003; Qi and Wu, 2006; Neely *et al.*, 2009; Cialenco and Protopapadakis, 2011) while others explore complex non-linear models (Neely *et al.*, 1997; Gehrig and Menkhoff, 2006; Gardojevic 2007; Sermpinis *et al.*, 2015). There are also academics that believe FX series follow a random walk and any profitable trading rules are due to luck (Meese and Rogoff, 1983; MacDonald and Taylor, 1994; Kilian and Taylor, 2003).

This study utilizes the latest developments in time-series modelling and statistics in order to discover whether simple technical rules are profitable in FX trading series. It also explores whether it is possible to combine simple technical rules with a set of some of the most up-to-date Bayesian models (Relevance Vector Machine (RVM), Dynamic Model Averaging (DMA), Dynamic Model Selection (DMS) and a Bayesian regularised Neural Network (BNN)) and derive superior trades.

For this purpose, seven thousand eight hundred forty-six technical rules are generated and applied to three exchange rates (EUR/USD, GBP/USD, and USD/JPY). Next, the genuinely profitable trading rules are identified based on the Romano *et al.* (2008) test combined with the balancing procedure of Romano and Wolf (2010). These profitable rules are then combined with Naïve Bayes (NB), RVM, DMA, DMS and a BNN. It is worth noting that the RVM, DMA, DMS and BNN have not been used in a trading application[1]. Our results show that superior trading performance is achievable by combining a data snooping procedure and Bayesian learning models. We find that BNN, DMA, and DMS have the highest performance across the study periods. The profitability is robust over the time and the performance does not deteriorate over the more recent years.

[1] To our best knowledge, RVM has only one related application (Fletcher *et al.*, 2009) on FX carry trade. In this paper the RVM is used as a part of a set of AI models and its individual performance is not assessed. The BNN also has only one application in financial forecasting in Ticknor (2013). In his study, BNN is not evaluated in trading terms. We did not identify any related trading application of DMA and DMS although there are several studies with them in financial and economic modelling (such as Koop and Korobilis, 2012; and Byrne *et al.*, 2016).

The motivation for this study derives from four sources: the Adaptive Market Hypothesis (AMH), the contradicting reports on the value of technical analysis in trading, the popularity of Bayesian techniques in financial forecasting, and the increased use of computational techniques in trading. The AMH has three main principles: traders need to be adaptive, the performance of trading models varies through time and in competitive environments the opportunities for profits are scarce. In other words, in highly efficient markets simple trading strategies have small power and traders need to seek complex statistical methods that are adaptive to the changing environment. The FX market – the biggest capital market – is most competitive and it is heavily affected by the intervention of central banks. It is interesting to check the effectiveness of simple trading rules in this environment and if possible, to generate Bayesian combinations of simple rules that can beat a market. We also examine if the performance of the trading models varies through time and whether their profitability is less in "popular" exchange rates, as the AMH proposes.

Technical analysis is considered a universal trading practice across different markets (Blume *et al.*, 1994). Although theories around technical analysis vary, all of them are based on the idea of the recurrent nature of patterns in the securities' price charts. Chartists believe that understanding these patterns can facilitate the prediction of future prices (Fama, 1965). This approach to prediction of financial markets can be traced back to Dow theory. The theory argues that the average values represent net interactions of all market participants over day-to-day activities and discount all kind of news and events, even the unpredictable ones. It proposes three bands of trends known as primary, secondary and minor trends. The primary trends are major market movements known as bull and bear market. The secondary trend represents the corrections and recoveries over bull and bear markets respectively. Finally, the minor trends are daily meaningless fluctuations (Edwards *et al.*, 2007). Several studies, such as Sweeney (1988), Brock *et al.* (1992) and Blume *et al.* (1994) demonstrate the utility and the profitability of technical analysis in financial markets. In these studies, a large universe of simple trading rules is generated, and their average performance is evaluated on stocks or stocks indices over a large period of time. Gençay (1998) uses technical rules as inputs to Artificial Neural Networks (ANNs) and generates profitable models, while Allen and Karjalainen (1999) use a genetic algorithm to identify profitable technical trading rules for the Standard and Poor's (S&P) 500 index. Although these preliminary studies seem promising, they ignore the data snooping bias.

Data snooping occurs when a given dataset is used more than once for purposes of inference and model selection (White, 2000). This bias is prominent in trading applications where researchers rely on the same data set to test the significance of different trading rules individually. These individual statistics are generated from the same dataset and relate to each other. White (2000) formalises this bias and introduces the Bootstrap Reality Check (BRC), which considers the dependence of individual statistics. The introduction of BRC test allowed researchers to revisit technical analysis from a new angel. Sullivan *et al.* (1999) claim that, based on the BRC test, technical analysis has no value on Dow Jones Industrial Average (DJIA) index. Hansen (2005) argues that the BRC is too conservative and checks only whether there is any significant model. The BRC does not identify all such models. As a solution, Hansen (2005) introduces the Superior Predictive Ability (SPA) test, which is less conservative and seems more powerful (Hansen and Lunde, 2005). Hsu and Kuan (2005) study technical rules after taking into account data snooping with the SPA test and claim that it is possible to beat the market with complex rules. Romano and Wolf (2005) and Hsu *et al.* (2010) improve the BRC and the SPA test respectively and introduce stepwise procedures Step-BRC and Step-SPA. These tests can identify all possible significant models. Further improvements in MHT procedures are made by Romano and Wolf (2007), Romano *et al.* (2008), Bajgrowicz and Scaillet (2012) and Hsu *et al.* (2014). The trend in recent data snooping literature is to relax the statistics by controlling the probability of making multiple false rejections (falsely "found" profitable strategies) and at the same time improve the efficiency of the tests (Kearney *et al*., 2014). This is beneficial in trading applications, where large groups of technical rules are under study and the ability to make true rejections is the main concern. Based on the latest tests, Romano and Wolf (2007), Romano *et al.* (2008), Bajgrowicz and Scaillet (2012) and Hsu *et al.* (2014) conclude that it is possible to identify genuinely profitable trading rules by using an efficient MHT procedure. However, the same studies argue that the profit margins are small, and the trading performance varies through time.

In FX market specifically, Gehrig and Menkhoff (2006) argue that technical analysis has by far the greatest importance for FX trading. Gençay *et. al.* (2003) generate positive annualized returns on four currency pairs with a real-time trading based on simple exponential Moving Average (MA) models. Baillie and Chang (2011) and recently Elaut *et al.* (2018) also suggest that momentum trading strategies can be applied to capture the volatility of the FX market. However, Cialenco and Protopapadakis (2011) argue that simple trading rules do not report statistically significant profitability in fourteen currencies. Meese and Rogoff (1983), Baillie and Bollerslev (1989), and Chinn and Meese (1995) claim that major exchange rates follow a random walk (at least in the short-run). Taylor (1992) reports that 90% of the chief FX dealers based in London place some weight to technical analysis in their decision processes. Yilmaz (2003) suggests that FX prices do not always follow a martingale[2] process, especially during the periods of central banks interventions. Yang *et al.* (2008) argue that martingale behaviour cannot be rejected for major exchange rates. Contrary to these, MacDonald and Taylor (1994) develop a monetary model which outperforms the random walk for the GBP/USD exchange rate over short-run periods. Kilian and Taylor (2003) find strong evidence of predictability over horizons of 2 to 3 years with a similar model, but not over shorter horizons. Hsu *et al.* (2010) and Hsu *et al.* (2014) argue that technical analysis can beat the FX market. The same statement is made by Neely and Weller (2013) who add that traders need to be adaptive in their portfolios.

Developments in statistics and computer science offer new potentials for wealth management. The developments include the advent of new tools in the fields of MHT and AI. Chui *et al.* (2016) and the Boston Consulting Group (2015) estimate that by 2025 the field of wealth management will be dominated by ML. In academia, there is a plethora of studies in this

[2] Martingale corresponds a sequence of random variables where the expected value for the next observation is equal to the present one or $E(\zeta_{t+1}|\zeta_1, \dots, \zeta_t) = \zeta_t$.

field. Gençay (1998), Fernández-Rodríguez *et al.* (2000), Jasic and Wood (2004), Gradojevic (2007), Sermpinis *et al.* (2013), and Sermpinis *et al.* (2015) apply ANNs – a form of non-linear regression algorithms – to the task of forecasting and trading financial series with some success. Alvarez-Diaz and Alvarez (2003), Pai *et al.* (2006), and Huang *et al.* (2010) develop models inspired by the evolution of species to financial forecasting with good results. Allen and Karjalainen (1999) use a genetic algorithm to identify profitable technical trading rules for the S&P 500 index. Lin and Pai (2010), Bekiros (2010) and Gradojevic and Gençay, (2013) apply FL in order to generate trading signals. Other studies, such as Ticknor (2013), Gramacy *et al.* (2014) and Baltas and Karyampas (2018) use Bayesian Statistics in financial forecasting problems. The literature in the area is extensive and promising. In the papers that have a trading application (see among others, Jasic and Wood, 2004; Gradojevic and Gençay, 2013) the proposed complex models significantly outperform simple trading rules. An explanation can be offered by the AMH which argues that complex models can survive better in informative markets.

In a nutshell, the literature in technical analysis, data snooping and computational applications in trading, is wealthy and contradicting. Studies that do not consider the data snooping bias and involve models that require parametrization should be treated with scepticism. The data snooping bias should be examined with recent related tests that are not strict. Computational techniques seem able to generate profitable trades. However, it is not clear from the previous studies if computational models can outperform technical analysis, as the AMH claims.

The rest of this manuscript is as follows. In Section 2, the MHT procedure and the Bayesian models used for trading are studied. Section 3 presents the description of the application our trading system is applied. Section 4 exhibits the empirical results for the alternative models. Section 5 offers the concluding marks and finally the Appendix includes the technical trading pool used in this study and robustness checks.

2. Methodology

In this study, a large set of technical trading rules on FX data is generated. The genuine profitable rules are identified with the Romano *et al.* (2008) test as modified based on the balancing procedure of Romano and Wolf (2010). Then, the profitable rules are combined with an NB, an RVM, a DMA, a DMS, and a BNN. The next sections contain a short description of the data snooping procedure and the Bayesian techniques[3].

2.1 Data Snooping Test

At the first stage for the modelling, the genuinely profitable trading rules are identified from a pool of 7846 technical rules. For this purpose, the Romano *et al.* (2008) test is combined with

[3] These algorithms are characterized by their complexity (except for NB). For the sake of space and as their mathematical derivation already exist in the relevant literature, we present the general framework. For the data snooping procedure, the reader is referred to Romano *et al.* (2008) for the FWER control with one sided setup and for the balancing procedure to Romano and Wolf (2010). A detailed description of RVM is provided by Tipping (2001) while a complete mathematical derivation of DMA and DMS is provided by Raftery *et al.* (2010). The Bayesian training procedure of BNN is described in detail in Ticknor (2013).

the balancing procedure of Romano and Wolf (2010). The benefits of the proposed approach are threefold. Firstly, it considers different measures of errors. Secondly, it is balanced since each individual hypothesis is treated fairly. Finally, it involves a resampling and subsampling approach that considers the dependence structure of the individual test statistics. These facts make it highly applicable in trading applications and more efficient compared to the Step-BRC and Step-SPA tests (Romano and Wolf, 2005; Hsu *et al.*, 2010).

The data snooping test is an MHT procedure in which a set of models are tested to identify the statistically different ones. As in any statistical test, there is the chance that a hypothesis is falsely rejected (Type I error). Familywise Error Rate (FWER) is the probability of having at least one false rejection. Traditional data snooping tests are too strict as they are attempting to control (asymptotically) the FWER. If the number of hypotheses is very large (as in our case), it is very difficult to make true rejections. In the asset management industry, professionals diversify their risk by investing in a large portfolio of models. The performance of any bad model is diluted by the much larger set of profitable rules. k-FWER determines the probability of having at least k false rejections. The data snooping approach of Romano *et al.* (2008) tries to control the k-FWER.

Let us consider a set of S trading strategies over T sample periods. For each trading strategy s (where $s = 1 \; to \; 7846$), the aim is to test the hypothesis that a model s beats a benchmark (ς) in terms of profitability. Then we can define μ_s as the unconditional average profit of the strategy s and $\theta_s = \mu_s - \mu_\varsigma$ as its difference from the benchmark. The null hypothesis is $H_{0,s}\colon \theta_s \leq \theta_\varsigma$, while the alternative is $H_{1,s}\colon \theta_s > \theta_\varsigma$. This setting tests the hypothesis that the technical rules have an equal or worse profitability compared to the benchmark ς . The test statistic is set as:

$$\mathrm{Z}_{T,s} = \frac{\bar{r}_{T,s} - \bar{r}_{T,\varsigma}}{\hat{\sigma}_{T,s}} \tag{1}$$

where the historical mean $\bar{r}_{T,s}$ and standard deviation $\hat{\sigma}_{T,s}$ in Eq. (1) are given by:

$$\bar{r}_{T,s} = \frac{1}{T} \sum_{t=1}^{T} \hat{r}_{t,s}, \tag{2}$$

$$\hat{\sigma}_{T,s} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} \left(r_{t,s} - r_{t,\varsigma} \right)^2}. \tag{3}$$

The k-FWER is controlled through the one-sided setup of the k-StepM method of Romano and Wolf (2005). Firstly, the strategies are sorted in a descending order based on the test statistics. After this is done, if b_k is the k-largest test statistic, then $\mathrm{Z}_{T,b_1} \geq \cdots \geq \mathrm{Z}_{T,b_S}$. Next, the k-th largest test statistic and the $1 - a$ (where a is the significance level) percentile of its sampling distribution are estimated. The individual hypotheses outside the confidence region are rejected. For the hypotheses not rejected, the process is repeated until the number of rejections is smaller than the desired k. For more details on the Step-M methods and the relevant bootstrap approach, see Romano *et al.* (2008) or Mazzocco and Saini (2012). In order to control the k-FWER, the innovations of Romano and Wolf (2010) are followed. They introduce an asymptotically balanced method that controls the average number of false rejections. Implicitly this approach considers the dependence structure of the individual test statistics, which leads to a more efficient control of false null hypotheses (Type II error). In this application, most technical trading rules have some form of weak dependency. (For instance, two MA cross-over strategies with different fast-MA of 2 and 5 periods but a similar slow-MA of 75 periods).

The selection of k depends on the problem under study and the practitioner's approach. If k is 1, the method can be overly conservative and inefficient. For this study, the k is set to 39 (roughly 0.5% of the 7846 technical rules under study[4]). As a benchmark to the data snooping test (ς), a basic random walk model is applied since major exchange rates are widely known to follow a random walk (see among others, Meese and Rogoff, 1983; Baillie and Bollerslev, 1989; and Chinn and Meese, 1995).

2.2 RVM

The RVM approach proposed by Tipping (2001), seeks to find the most effective inputs based on probabilistic approaches to classification and regression problems. Throughout this process, the determined effective points are defined as relevance vectors. This Section summarizes the RVM structure.

Assuming a supervised learning framework, we define a dataset D with ν predictors and T training points, an input series set $\boldsymbol{x} = \{x_i: i = 1, \dots, T\}$ and a target series set $\boldsymbol{y} = \{y_i: i = 1, \dots, T\}$. The general predictive formulation can be specified as:

$$y_i = f(x_i) + \varepsilon_i \tag{4}$$

where ε_i is the zero-mean Gaussian error term with distribution $\varepsilon_i = y_i - \hat{y}_i \sim \mathcal{N}(0, \sigma^2)$, $\hat{y}_i$ is the target point forecast, and f is the transfer function.

Given the basis function set $\boldsymbol{\varphi}(\boldsymbol{x})$ and the weight vector $\boldsymbol{w}$, the RVM's prediction under the linear model assumption can be expressed as:

$$\hat{\boldsymbol{y}} = f(\boldsymbol{x}, \boldsymbol{w}) = \sum_{j=1}^{\nu} w_j \varphi_j(x) + w_0 \tag{5}$$

where $\boldsymbol{\varphi}(\boldsymbol{x}) = [1, K(\boldsymbol{x}, x_1), \dots, K(\boldsymbol{x}, \boldsymbol{x}_T)]'$, w_0 is the bias, and $\boldsymbol{w} = [w_1, \dots, w_\nu]$.

In the context of RVM, Radial Basis Function (RBF) is mostly considered as the basis function K. This is due to its simplicity and superior optimization performance (Park and Sandberg, 1991). Subsequently, the multivariate Gaussian likelihood of the dataset can be written as:

$$Pr(\boldsymbol{y}|\boldsymbol{w}, \sigma^2) = (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{\|y - \Phi w\|^2}{2\sigma^2}\right) \tag{6}$$

[4] The choice of 0.5% is based on *approximating* the set of initial rejections (including both true and false discoveries) with the top 5 percent of trading rules and allowing 10% of rejections to be the Type I error. This *approximation* can be improved by alternative statistical approaches to the rejections based on the test statistics and the bootstrap *p*-value. However, this *approximation* is chosen to find the most profitable trading rules.

where $\mathbf{\Phi}$ is the $T \times (T+1)$ 'design' matrix with $\Phi_{nm} = K(x_n, x_{m-1})$ and $\Phi_{n1} = 1$.

Over-fitting can be expected in the maximum-likelihood estimation of $\boldsymbol{w}$ and σ^2 in Eq. (6). To overcome this, Tipping (2001) recommends setting prior constraints on parameters $\boldsymbol{w}$ by adding a complexity term inspired by the traditional margin concept of Support Vector Machine (SVM) modelling. Gaussian priori in RVM context for an individual w_j can be expressed as:

$$Pr(w_j|\alpha_j) = (\frac{\alpha_j}{2\pi})^{1/2}\exp(-\frac{\alpha_j w_j^2}{2}) \tag{7}$$

Similarly, for the whole set of $\boldsymbol{w}$: $Pr(\boldsymbol{w}|\boldsymbol{\alpha}) = \prod_{i=1}^{T} \mathcal{N}(w_i|0, \alpha_i^{-1})$, where $\boldsymbol{\alpha} = [\alpha_0, \dots, \alpha_T]'$ is a hyperparameter vector governing the prior defined over the weight $\boldsymbol{w}$ to control deviation of each w_j from the zero mean.

Given priori information controlling the generalisation ability and the likelihood distributions, applying Bayes' rule generates the posterior over $\boldsymbol{w}$ as:

$$Pr(\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma^2) = \frac{Pr(\boldsymbol{y}|\boldsymbol{w}, \sigma^2)Pr(\boldsymbol{w}|\boldsymbol{\alpha})}{Pr(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2)} \tag{8}$$

In the case of a multivariate Gaussian distribution, the posterior takes the following form:

$$Pr(\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{\alpha}, \sigma^2) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{9}$$

The covariance and the mean of the distribution are estimated respectively by the following analytical solution of Eq.s (10 and 11):

$$\boldsymbol{\Sigma} = (\boldsymbol{\Phi}'\mathbf{B}\boldsymbol{\Phi} + \mathbf{A})^{-1} \tag{10}$$

$$\boldsymbol{\mu} = \boldsymbol{\Sigma}\boldsymbol{\Phi}'\boldsymbol{B}\boldsymbol{t} \tag{11}$$

where $\mathbf{A} = (\alpha_0, \dots, \alpha_T)$ and $\mathbf{B} = \sigma^{-2}\boldsymbol{I}_T$.

To estimate the weights, the missing set $\boldsymbol{\alpha}$ in the above equations is treated as a hyperparameter. Therefore, the relevance vector learning model approximates the mode for the hyperparameter posterior i.e. maximization of $Pr(\boldsymbol{\alpha}, \sigma^2) \propto Pr(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2)Pr(\boldsymbol{\alpha})Pr(\sigma^2)$ given $\boldsymbol{\alpha}$ and σ^2. Assuming uniform hyperperiors, the model optimization can be thought equivalent to the maximization of $Pr(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2)$. By Integrating out the weights, the following is derived:

$$Pr(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2) = \int Pr(\boldsymbol{y}|\boldsymbol{w}, \sigma^2)\, Pr(\boldsymbol{w}|\boldsymbol{\alpha}) d\boldsymbol{w} \tag{12}$$

where $Pr(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2)$ can be computed by the following equation:

$$Pr(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2) = (2\pi)^{-T/2}\left|\boldsymbol{B}^{-1} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}'\right|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{y}'(\boldsymbol{B}^{-1} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}')^{-1}\boldsymbol{y}\right\} \tag{13}$$

The marginal likelihood for hyperparameters in the Gaussian distribution form is given by:

$$Pr(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2) = N(0, \boldsymbol{B}^{-1} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}') \tag{14}$$

The estimation of the above hyperparameters is conducted through an iterative procedure similar to the gradient ascent on the objective function for Maximum A Posteriori (MAP) estimate of the weights (for more details refer to Ghosh and Mujumdar, 2008; and Candela and Hansen, 2004). The numerical approximation is adopted because there is no closed form solution. The MAP estimation is dependent on the hyperparameters $\boldsymbol{\alpha}$ and σ^2 in other words **A** and **B** in Eq.s (10 and 11).

Following Tipping (2001) the solution to Eq.s (10 and 11) is estimated through differentiating and setting Eq. (14) to zero. After rearranging we yield:

$$\alpha_m^{new} = \frac{\gamma_m}{\mu_m^2} \tag{15}$$

where μ_m is the m-th posterior mean-weight from the equation set and $\gamma_m \equiv 1 - \alpha_m \Sigma_{mm}$.

The Σ_{mm} is the m-th diagonal element of the covariance $\boldsymbol{\Sigma}$ matrix calculated by the updated $\boldsymbol{\alpha}$ and σ^2. Parameter γ_m is interpreted as the degree to which associated w_m is well-determined by the training data (MacKay, 1992). When the fit is not appropriate, the w_m is constrained by priori with small σ_m^2. For example, for a high value of α_m, Σ_{mm} will tend to α_m^{-1} and consequently γ_m approaches zero. On the other hand, when the fit is good, $\alpha_m \approx 0$, this leads to $\Sigma_{mm} \approx 0$, and finally $\gamma_m \approx 1$. Consequently, the range for γ_m is $[0,1]$. For the other hyperparameter σ^2 differentiation results in the update of the noise variance estimation as:

$$(\sigma^2)^{new} = \frac{\|\boldsymbol{t} - \boldsymbol{\Phi}\boldsymbol{\mu}\|^2}{T - \Sigma_m \gamma_m} \tag{16}$$

The learning process advances by re-estimating the hyperparameters and updating the mean and covariance of the posterior in each iteration. This continues until the convergence is met at an iteration step or until the incorporated stop criteria are activated to avoid reaching redundant loops. In practice during the iterative update of the hyperparameters, many α_js approach infinity. In that way w_js tend to form a delta function around zero. Consequently, many elements in $\boldsymbol{w}$ and associated elements in $\boldsymbol{\varphi}(\boldsymbol{x})$ would be discarded from the operational model. The remaining basis functions that are associated with training points within the sample dataset produce a sparse solution for the RVM model. These remaining examples are the so-called RVs. Tipping (2001) claims that the above predictive estimations are found to be robust by most empirical evidence. The predictive distribution for a given new point x_* complemented by a y_* class label is given by:

$$Pr(y_*|x_*, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) = \int Pr(y_*|x_*, \boldsymbol{w}, \sigma_{MP}^2)\, Pr(\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) d\boldsymbol{w} \tag{17}$$

The Gaussian form is expressed as:

$$Pr(y_*|x_*, \boldsymbol{\alpha}_{MP}, \sigma_{MP}^2) = \mathcal{N}(\hat{y}_*, \sigma_*^2) \tag{18}$$

where $\hat{y}_* = \boldsymbol{\mu}'\boldsymbol{\Phi}(x_*)$ is the mean estimate of the target and $\sigma_*^2 = \sigma_{MP}^2 + \boldsymbol{\Phi}(x_*)'\boldsymbol{\Sigma}\boldsymbol{\Phi}(x_*)$ is the corresponding uncertainty. The $\boldsymbol{\alpha}_{MP}$ and σ_{MP}^2 are the most probable hyperparameter values obtained from Eq. (13).

The predictive mean is generated through the reduced basis function and the input explanatory variables. The predictive variance confirms that the Out-Of-Sample (OOS) prediction is consistently higher than the In-Sample (IS) one due to extra uncertainty caused in the process of the weights' prediction.

2.3 DMA and DMS

Financial trading series are dominated by structural breaks. Models with fixed coefficients work only for short periods. Time-Varying Parameter (TVP) models consider the parameters as a function of time and are estimated using state-space methods such as Kalman filter. Despite the benefits of the TVP models over the static methods, the assumption is that the initial set of explanatory variables remains relevant over time. This can be undesirable in real environment applications.

The DMA proposed by Raftery *et al.* (2010) allows selecting different subsets of explanatory variables over time along with variable coefficients. Consider a candidate input set $u = 1, \ldots, U$, then the state-space model at time $t = 1, \ldots, T$ for the dependent variable y_t can be presented under observational and state equations as:

$$y_t = F_t^{(u)'}\zeta_t^{(u)} + \varepsilon_t^{(u)}, \tag{19}$$

$$\zeta_t^{(u)} = \zeta_{t-1}^{(u)} + \eta_t^{(u)}, \tag{20}$$

$$\begin{pmatrix} \varepsilon_t^{(u)} \\ \eta_t^{(u)} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} R_t^{(u)} & 0 \\ 0 & V_t^{(u)} \end{pmatrix}, \tag{21}$$

where $F_t^{(u)}$ in Eq. (19) is a subset from the ν potential predictors at each time. The $\zeta_t^{(u)}$ is a $p \times 1$, $p \leq \nu$ vector of time-varying regression coefficients evolving over time by Eq. (20). From the specification provided, it is immediately visible that the total number of candidate models is $U = 2^\nu$. Unless ν is very small, updating the parameters becomes demanding and computationally very slow using a full Bayesian approach. Raftery *et al.* (2010) approximates the solutions of Eq.s (19 to 21) and thus makes the algorithm more efficient. However, the computational burden still increases exponentially when ν is large. This makes DMA impractical with standard computer processing when ν is larger than 20.

The DMA averages the forecasts across candidate combination of models based on predictive likelihood through a recursive updating scheme. The predictive likelihood estimates the ability of model u to predict y_t. The models containing better predictors receive higher predictive likelihood and are associated with higher weights in the averaging process. Respectively, at each time t two vectors of weights for the model u are calculated as $\omega_{t|t-1,u}$ and $\omega_{t|t,u}$. The first quantity denotes the weight of a specific model given information available at time $t - 1$, while the latter one represents the dedicated weight to the specific model after the model update at time t. The DMS makes the prediction based on the highest value of weight which is calculated through the updating process. This can be mathematical expressed as:

$$\omega_{t|t,u} = \frac{\omega_{t|t-1,u}\, L_u(y_t|y_{1:t-1})}{\sum_{l=1}^{U} \omega_{t|t-1,l}\, L_l(y_t|y_{1:t-1})}, \tag{22}$$

where $L_u(y_t|y_{1:t-1})$ is the predictive likelihood measured by the realized value of y_t. By using a forgetting factor δ, as suggested by Raftery *et al.* (2010), the weights for the following period are formulated as:

$$\omega_{t+1|t,u} = \frac{\omega_{t|t,u}^{\delta}}{\sum_{l=1}^{U} \omega_{t|t,l}^{\delta}}. \tag{23}$$

The δ controls the 'forgetting' of the entire model set and it can take values in the range of $0 < \delta \leq 1$. Raftery *et al.* (2010) suggest $\delta = 0.99$ as a benchmark, while Koop and Korobilis (2012) recommend $\delta \in [0.95,0.99]$. The recursive calculation starts with a non-informative choice for the initial weight $\omega_{0|0,u} = \frac{1}{U}$ for $u = 1, \dots, U$. The other approximation is used in the estimation of the $V_t^{(u)}$. The second forgetting factor, λ, explains the information loss over time. Representing the variance estimator $\zeta_t^{(u)}$ by $C_t^{(u)}$, the conditional variance, $V_t^{(u)}$(there is no need to be estimated for each individual model), is calculated as:

$$V_t^{(u)} = (1 - \lambda^{-1}) C_{t-1}^{(u)}. \tag{24}$$

In other words, the λ controls the amount of shock affecting the coefficients $\zeta_t^{(u)}$. Identical to δ, λ may also take values near to one. This determines the rate of which information loses effect on the model coefficients. Here, it should be noted that by setting $\delta = 1$, the DMA is transformed to a TVP model with no change in the subset selection over time. Additionally, by setting $\delta = \lambda = 1$, the DMA is simplified to conventional Bayesian Model Averaging with no time-varying characteristic.

The term "forgetting factors" stems from the fact that observations at *j* periods ago have a contribution with factor λ^j to the model. As a simple analogy, in the case of having $\lambda = 0.99$, it takes 69 periods for the shock from each observation to lose half of its effect on the coefficients. The half-life of the shock to the model is reduced to 14 periods for $\lambda = 0.95$ and further to 6 in the case of $\lambda = 0.90$. The values of the forgetting factors can considerably affect the way models react to the changes of the environment. Various surveys recommend the direct use of $\delta = \lambda = 0.99$ as a benchmark (Raftery *et al.*, 2010; Aye *et al.*, 2015). Koop and Korobilis (2012) argue that the performance of competing models with different forgetting factors are robust and perform efficiently. They also conduct a sensitivity analysis for the parameters that shows that the best OOS forecasting results are obtained by setting $\delta = 0.95$ and $\lambda = 0.99$. The study of Koop and Korobilis (2012) is conducted on macroeconomic data and indicates that appropriate selection of parameters under volatile conditions can enhance the predictive ability of the DMA and DMS models. In this study, a wider variety of values e.g. $\{0.90, 0.95, 0.99, 1\}$ for the parameters are experimented to accommodate a more rapid update of the model specification. This choice accommodates the dynamics and nonlinearities of the market. The wide range for parameters also replicates the behaviour of expert traders on the market floor that constantly revise their trading strategy and if necessary rapidly switch from one approach to another[5].

2.4 BNN

BNN is specific extension of ANNs that are a class of non-linear models inspired by the work and functioning of biological neurones. In the most common set-up, an ANN has at least three layers. The first layer is called the input layer (where the technical rules are fed). The last layer is called the output layer (where the forecasted value is extracted). An intermediary layer of nodes, the hidden layer, separates the input from the output layer. The number of nodes in the hidden layer controls the complexity the model is able to fit. In addition, the input and hidden layer contain an extra node called the bias node. This node has a fixed value of one and has the same function as the intercept in traditional regression models. Normally, each node of one layer has connections to all the other nodes of the next layer.

The training of the network is to adjust its weights so that the network maps the input value of the training data to the corresponding target value. It begins with randomly chosen weights and proceeds by applying a learning algorithm. The most common procedure is the backpropagation of errors (Shapiro, 2000) which looks for the minimum of the error function (commonly the Mean Squared Error (MSE) between the actual and forecasted values) in weight space using the method of gradient descent.

Ticknor (2013) modifies the training procedure by applying Bayesian regularisation which trains the ANN based on:

$$\Omega = \gamma_1 E_{SE} + \gamma_2 E_{we} \tag{25}$$

where E_{SE} is the sum of the squared errors, E_{we} is the sum of the squared network weights and γ_1 and γ_2 are objective function parameters. In this framework, the ANN's weights are considered random variables and their density function based on the rule of Bayes is:

$$Pr(w|D, \gamma_1, \gamma_2, M) = \frac{Pr(D|w, \gamma_1, M) Pr(w|\gamma_2, M)}{Pr(D|\gamma_1, \gamma_2, M)} \tag{26}$$

where w is a vector of the network weights, D is a vector with the dataset (technical rules in our case) and M is the underlying model (the ANN in this case). Based on Forsee and Hagan (1997), the optimization of parameters γ_1 and γ_2 requires solving a Hessian matrix based on the Levenberg–Marquardt training algorithm. In order to protect the ANN from over-fitting, the early stopping procedure in the IS is applied.

In BNN, overly complex models are penalized as unnecessary linkage weights and are effectively driven to zero. Burden and Winkler (2009) argue that the network calculates and trains on the nontrivial weights which converges to a constant as the network grows.

[5] The reported results in Section 4 are based on the best performance measured in the IS. The IS in the DMA and DMS cases is the first 80% of observations.

Parsimonious ANNs limit the training time and the danger of over-fitting. Additionally, they do not require the validation step which is otherwise necessary on the traditional back-propagated ANNs.

2.5 NB

The RVM models the posterior $Pr(y|x)$ from the attribute variable set $\boldsymbol{x}$ to the class label set $\boldsymbol{y}$. In the context of probabilistic classification, this approach is termed as discriminative learning. In discriminative classifiers, all training observations from any class y_i are considered in establishing the model. Despite evidence in favour of discriminative classification (Vapnik, 1998), there is a reverse approach to the probabilistic classification regarded as generative learning. The generative classifiers learn the joint probability $Pr(x, y)$, by using Bayes rules to calculate $Pr(y|x)$. Then, a classification model is obtained, which classifies each data point to the label y with the highest posterior probability. Generative learning is particularly useful, when there is missing information in the dataset.

The NB is a simple classifier that allocates each point of the dataset to the most likely class according to the generative approach. The model is named naïve because of its simplifying assumption that all variables x_i are conditionally independent for a certain class y_0. For a test sample with attribute variables $x = x_0$ of v dimension and class label $y = y_0$, the probability of each class can be calculated by the observed values of the predictive attributes $x_{j,t}, j = 1, \ldots, v\ ;\ t = 1, \ldots, T$. By using the Bayes rule, the posterior can be calculated as:

$$Pr(y = y_0|x = x_0) = \frac{Pr(y=y_0)Pr(x = x_0|y = y_0)}{Pr(x=x_0)} \tag{27}$$

The predicted label is the most probable class given by (27). Under the class-conditional independence assumption, we have:

$$Pr(x = x_0|y = y_0) = \prod_{i=1}^{v} Pr(x_i = x_{i,0}|y = y_0) \tag{28}$$

The conditional distribution $Pr(x = x_0|y = y_0)$ may take a multinomial (Gaussian) form for discrete (continuous) variables. Based on the training dataset and plugging the empirical probabilities in Eq.s (27 and 28), it is easy to make a natural classification as the naïve benchmark.

In this paper, NB is used as a benchmark for other types of Bayesian probabilistic models. The attribute variables are the signals generated by the trading rules from a set of $x \in \{-1, 0, 1\}$. This set represents short, hold or long positions respectively. Similarly, the class label is the one-step-ahead direction of the market change. For example, a class label $y \in \{-1, 0, 1\}$ represents the fall, no change or rise respectively of the market in the next period.

3. Empirical Section

3.1 Dataset

The proposed methodology is applied to the daily price (open, high, low, and close) and volume series for EUR/USD, GBP/USD, and USD/JPY exchange rates. The period under study is the start of 2010 until the end of 2016 and it is divided into four trading exercises. In each trading exercise, the first three years act as IS and the following year as OOS (i.e. in the first exercise the IS covers the years 2010 until 2012 and the year 2013 acts as OOS).

At the first stage, 7846 simple trading rules are generated for each of the three exchange rates at the IS periods of the four exercises. The trading rules consist of FIlter Rules (FIRs), MAs, Support and Resistance levels (S&Rs), Channel Breakouts (CBs), and On-Balance Volume indicators (OBVs). It is the same set of rules applied in the studies of Sullivan *et al.* (1999) and Bajgrowicz and Scallet (2012). For a description of these rules see Appendix A.1. All trading rules are generated through the logarithmic returns of the exchange rates. The summary statistics of the logarithmic returns on daily close for the exchange rates under study are presented in the Table 1.

Table 1: Descriptive statistics

Period	Statistic	EUR/USD	GBP/USD	USD/JPY
2010.01.04 - 2013.12.31	Mean (bp)	-4.59	-0.28	1.25
	Standard Deviation (bp)	62.2	50.9	63.1
	Kurtosis	3.78	3.25	8.91
	Skewness	-0.12	-0.07	-0.55
	JB p-value	0.00	0.00	0.00
	ADF p-value	0.00	0.00	0.00
2011.01.03 - 2014.12.31	Mean (bp)	-0.97	-0.02	0.38
	Standard Deviation (bp)	53.8	43.9	59.3
	Kurtosis	4.28	3.51	9.27
	Skewness	-0.14	-0.11	0.06
	JB p-value	0.00	0.00	0.00
	ADF p-value	0.00	0.00	0.00
2012.01.02 - 2015.12.31	Mean (bp)	-1.68	-0.13	4.30
	Standard Deviation (bp)	54.5	45.5	56.7
	Kurtosis	4.95	3.96	6.51
	Skewness	0.17	0.17	0.05
	JB p-value	0.00	0.00	0.00
	ADF p-value	0.00	0.00	0.00
2013.01.02 - 2016.12.30	Mean (bp)	-2.19	-0.04	0.29
	Standard Deviation (bp)	54.7	62.1	64.8
	Kurtosis	5.48	4.47	6.81
	Skewness	0.11	-3.42	-0.27
	JB p-value	0.00	0.00	0.00
	ADF p-value	0.00	0.00	0.00

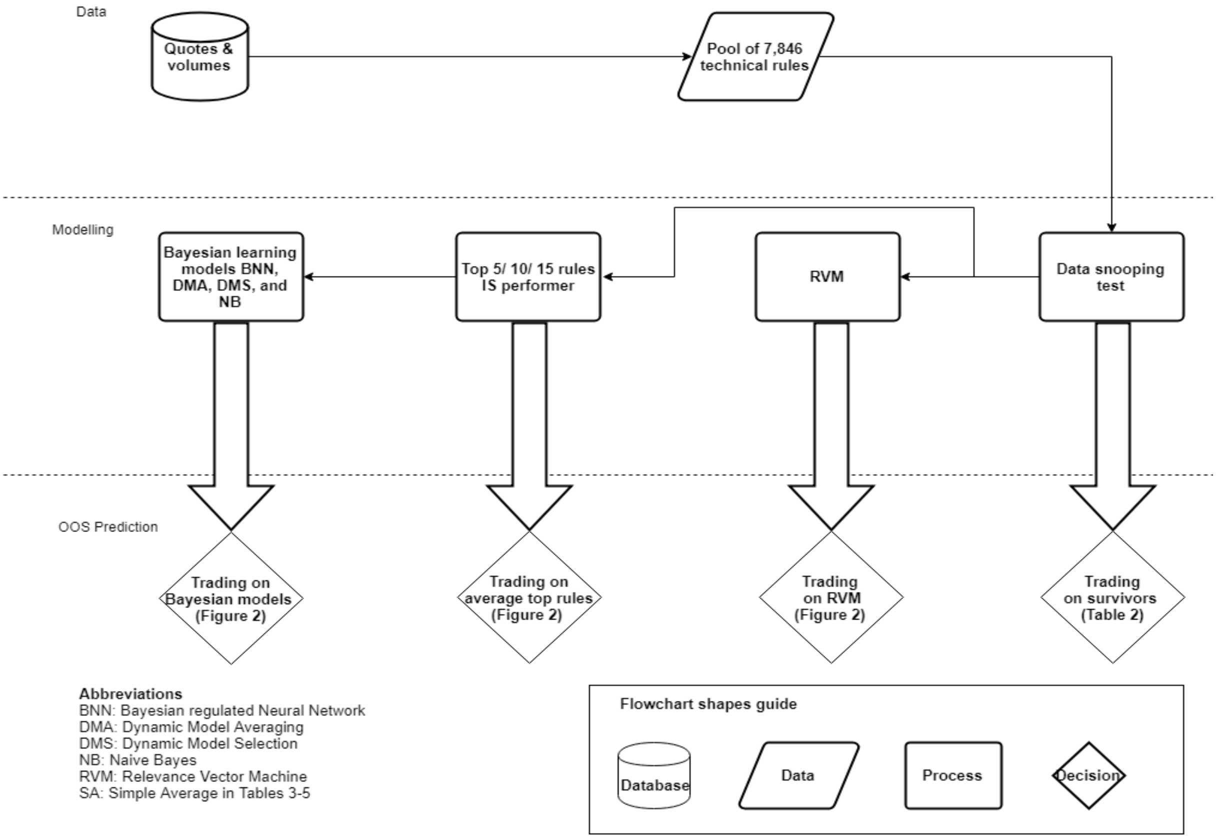
Note: The mean and standard deviations are reported in basis points. Reported values of zero for the Jarque–Bera and augmented Dickey–Fuller (without any lagged difference) tests correspond to p-values less than 1 over 100 ($p<0.01$)

All series exhibit positive kurtosis while the skewness is mixed but generally close to zero. The Jarque and Bera (JB) (1980) test reveals that the return series do not follow a normal distribution, while the Augmented Dicky-Fuller (ADF) (1979) shows that they are stationary. Each trading rule generates a daily trading signal for the relevant exchange rate and IS period. The trading signal can be long (buy), short (sell) or hold (no action). Based on these signals, the trading performance of each of the 7846 rules is generated. As transaction costs, we consider three basis points per trade[6].

3.2 Empirical design

Our empirical study is designed to show the merits of combining Bayesian Statistics with data snooping procedures. This Section describes how the models are intertwined to and how their corresponding performance is measured. Figure 1 presents the model synthesis used to quantify the differences between the alternative predictive models.

Figure 1: Modelling flowchart



Note: The study utilizes the generated technical trading universe to produce four sets of predictions. In the modelling process of this study, the RVM model is fed directly the set of discoveries from the data snooping procedure to generate forecasts. The other Bayesian models are fed with top rules based IS accuracy in predicting one step ahead. The Bayesian models are benchmarked against the average of survivors from the test and the average of top rules' performances. The performance is measure in terms of excess return and Sharpe ratio.

In the first step, the technical trading pool is generated. Then, the data snooping test is conducted with a zero-rate benchmark[7]. The survivals of the test are considered as the potential pool of superior models. We measure the number of survivors and their profitability within the IS and OOS in Section 4.1. In the next step we want to check if the Bayesian models are able to learn from the market and replicate an experience trader's behaviour. The three main methods — BNN, DMA, and DMS — are computationally demanding and

[6] See among others, www.interactivebrokers.com and www.fxall.com

[7] We chose this benchmark since the mean return on the studied assets are approximately zero for all cases.

combining them with all the identified trading rules is not feasible[8]. Thus, for BNN, DMA, DMS and NB the best 5, 10 and 15 technical rules are used as inputs based on accuracy in the IS[9]. For RVM, the algorithm requires a large set of potential predictors in order to identify the optimum relevant subset of inputs. Therefore, it is fed with all the identified genuine profitable technical rules. The performance of the best 5, 10, and 15 rules are also measured without using any learning algorithm.

4. Trading Application

We assess the behaviour of the technical trading pool at different stages. First we see how many rules are able to statistically generate a positive test statistics as in Eq.s (1 to 3). Then trading portfolios are constructed based on the significant ones. The portfolios are studied for their number of constituents, and the corresponding IS and OOS profitability. Three main exchange rates and their average as a portfolio of major pairs are studied over exercises with length of one calendar year.

4.1 Genuine Technical Rules

The equal weight[10] trading performance of the identified genuine technical rules is presented below.

Table 2: Properties of data-snooping test survivors

Period	Asset	IS Performance	OOS Performance	Surviving Rules Count
2013	EUR/USD	3.07% (0.29)	-1.23% (-0.72)	839
	GBP/USD	1.1% (0.29)	0.21% (0.05)	788
	USD/JPY	3.24% (0.07)	-3.37% (-0.65)	870
2014	EUR/USD	5.71% (1.01)	3.69% (0.96)	819
	GBP/USD	5.67% (2.62)	4.41% (1.56)	368
	USD/JPY	6.7% (0.58)	6.19% (1.07)	439
2015	EUR/USD	3.18% (1.45)	2.21% (0.41)	1144
	GBP/USD	6.42% (1.85)	1.47% (0.27)	1047
	USD/JPY	12.5% (2.13)	-6.58% (-0.94)	523
2016	EUR/USD	10.31% (0.25)	2.04% (0.41)	1147
	GBP/USD	4.85% (0.14)	1.91% (0.22)	981
	USD/JPY	8.77% (1.43)	-1.31% (-0.33)	337

[8] For example, for DMA and the EUR/USD in the first forecasting exercise, the algorithm would have to estimate 2^{839} combinations. This task is feasible with the help of supercomputers (which were not available in this project) but it is unrealistic from a trading perspective where speed is essential. For an up-to-date personal computer (Intel core-i5 3470 64-bit processor with 8 GB memory), DMA needs around 30 mins to produce the results for one experiment with fifteen inputs (out of the twelve similar experiments, for each of the three exchange rates). This is increased to twenty-one hours for twenty inputs. Similarly, in BNNs, when the number of inputs is very large their algorithm becomes insufficient, they become prone to overfitting and their forecasting performance is crippled (Zhang *et al.*, 1998; and Yegnanarayana, 2009).

[9] We repeat the same practice with the top performers based on excess return and Sharpe ratio within the IS. The results with these metrics are presented in Appendix B.

[10] The equal weight corresponds to investing the $1/C$ of the total wealth to each of the C trading rules identified from the data snooping procedure. The portfolio construction approach as presented in Bajgrowicz and Scallet (2012) has also been explored. In Bajgrowicz and Scallet (2012), the buy and sell signals counter each other while the neutral signs are considered risk free investments. The portfolios derived from this approach do not change the view of Table 1. However, as the scope of this study is to check the efficiency of technical analysis in FX and whether Bayesian techniques can improve their trading performance, the annualized averages are presented. Following this approach, the results of Table 2 can also be compared with the results of Section 4.2 where the best trading rules are combined with the Bayesian techniques.

Total	Average	5.96% (1.01)	0.8% (0.19)	775.17

Note: The table presents the excess annualized returns (above the risk-free rate) of the technical rules after transaction costs. The values in parentheses correspond to the Sharpe ratios. Trading rules correspond to the number of genuine trading rules identified in the in-sample by the Romano *et al.* (2008) test combined with the balancing procedure of Romano and Wolf (2010).

From Table 2, we note that in all cases the data snooping procedure was able to identify genuinely profitable trading rules based on the IS observations. The number of significant rules corresponds roughly to 5%-15% of the total number of trading rules under study. The average trading performance of the genuine trading rules is positive in all IS cases and in most OOS cases.

These results allow us to argue that technical analysis seems to have value on the exchange rates and periods under study. There are genuine profitable simple technical rules in the IS. It is possible for investors and researchers to identify these rules with the help of recent developments in SI. This performance supports AMH which argues that investors need to be adaptive in highly competitive trading environments. These results agree with Hsu *et al.* (2010), Neely and Weller (2013) and Hsu *et al.* (2014) that argue that technical analysis has some value. However, in line with the previous studies, we note that the performance of these rules is volatile probably due to the time-varying market conditions. we also note that the profit margins are low, and the OOS trading performance is not always above the risk-free rate[11].

4.2 Bayesian Trading

Although technical rules seem unreliable for trading, Bayesian techniques can offer an advantage to investors. Arguably, they could combine different trading signals and derive strongly positive trading performances. Complex models should be capable of encompassing the simple trading rules. Additionally, these models should be able to offer an advantage to highly competitive markets. More specifically, the dynamic nature of DMS and DMA and the non-linear adaptive nature of BNN should be able to handle the changing trends of the FX series under study. By using the flowchart in Figure 1, in this Section we present our empirical results for the studied Bayesian models. In Tables 3 to 5, the trading performance of all the Bayesian methods in the OOS are presented for three major exchange rates. Table 6 replicates the same practice for an equal-weight portfolio of the studied pairs. A Simple Average (SA) of the best rules is also estimated as a naïve benchmark.

Table 3: Trading performance for EUR/USD

Model	2013	2014	2015	2016	Average
BNN (15)	6.12% (0.59)	5.37% (0.4)	5.37% (0.64)	7.28% (0.76)	6.04% (0.6)
DMA (15)	5.3% (0.57)	6.15% (0.48)	6.01% (0.62)	6.74% (0.65)	6.05% (0.58)
DMS (15)	4.8% (0.47)	5.4% (0.44)	6.12% (0.68)	**6.8% (0.74)**	5.78% (0.58)
BNN (10)	6.48% (0.6)	**6.42% (0.51)**	6% (0.77)	6.01% (0.65)	**6.23% (0.63)**
DMA (10)	**6.89% (0.62)**	6.08% (0.44)	6.85% (0.71)	5.07% (0.56)	6.22% (0.58)
DMS (10)	5.02% (0.48)	4.24% (0.37)	6.97% (0.69)	5.29% (0.73)	5.38% (0.57)
BNN (5)	6.01% (0.59)	3.77% (0.36)	5.08% (0.7)	6.21% (0.69)	5.27% (0.59)

[11] As risk free rate, the effective federal funds rate is considered. The interest rate at which US depository institutions trade federal funds with each other overnight.

DMA (5)	4.34% (0.44)	3.86% (0.32)	**7.79% (0.8)**	6.36% (0.64)	5.59% (0.55)
DMS (5)	3.74% (0.5)	3.99% (0.35)	6.23% (0.57)	4.5% (0.4)	4.62% (0.46)
RVM	4.35% (0.5)	3.28% (0.31)	4.41% (0.63)	4.24% (0.47)	4.07% (0.48)
NB (15)	1.8% (0.3)	1.72% (0.17)	1.18% (0.27)	3.33% (0.41)	2.01% (0.29)
NB (10)	2.09% (0.33)	2.74% (0.26)	0.86% (0.25)	3.18% (0.38)	2.22% (0.31)
NB (5)	1.06% (0.18)	1.82% (0.18)	4.23% (0.42)	2.28% (0.36)	2.35% (0.29)
SA (15)	-1.44% (-0.14)	0.16% (0.08)	-0.99% (-0.12)	-1.32% (-0.12)	-0.9% (-0.08)
SA (10)	1.17% (0.2)	1.89% (0.19)	-2.18% (-0.22)	-2.46% (-0.2)	-0.4% (-0.01)
SA (5)	-2.88% (-0.22)	0.27% (0.09)	0.26% (0.19)	0.21% (0.17)	-0.54% (0.06)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample accuracy. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table 4: Trading performance for GBP/USD

Model	2013	2014	2015	2016	Average
BNN (15)	**7.29% (0.75)**	6.29% (0.76)	6.17% (0.71)	6.01% (0.64)	6.44% (0.72)
DMA (15)	6.66% (0.65)	**6.53% (0.72)**	**7.19% (0.79)**	**7.78% (0.83)**	**7.04% (0.75)**
DMS (15)	5.2% (0.56)	5.57% (0.69)	6.22% (0.68)	6.4% (0.67)	5.85% (0.65)
BNN (10)	6.38% (0.61)	5.21% (0.65)	4.55% (0.48)	5.83% (0.57)	5.49% (0.58)
DMA (10)	3.89% (0.38)	4.03% (0.5)	5.68% (0.62)	5.36% (0.51)	4.74% (0.5)
DMS (10)	3.92% (0.37)	4.48% (0.57)	6.01% (0.62)	4.39% (0.42)	4.7% (0.5)
BNN (5)	4.89% (0.42)	5.16% (0.61)	5.14% (0.53)	5.03% (0.52)	5.06% (0.52)
DMA (5)	4.57% (0.43)	3.64% (0.37)	5.96% (0.64)	4.7% (0.48)	4.72% (0.48)
DMS (5)	4.6% (0.44)	3% (0.45)	6.07% (0.64)	3.21% (0.36)	4.22% (0.47)
RVM	4.61% (0.41)	4.19% (0.43)	3.72% (0.39)	4.27% (0.48)	4.2% (0.43)
NB (15)	3.82% (0.38)	1.88% (0.19)	3.44% (0.37)	2.85% (0.31)	3% (0.31)
NB (10)	2.43% (0.33)	2.62% (0.28)	2.55% (0.31)	3.05% (0.29)	2.66% (0.3)
NB (5)	2.7% (0.32)	2.42% (0.26)	1.85% (0.27)	2.78% (0.26)	2.44% (0.28)
SA (15)	-0.53% (-0.15)	2.4% (0.24)	1.58% (0.19)	1.02% (0.09)	1.12% (0.09)
SA (10)	0.8% (0.12)	2.22% (0.2)	1.66% (0.23)	1.98% (0.18)	1.67% (0.18)
SA (5)	1.92% (0.3)	0.71% (0.09)	-0.48% (-0.07)	-0.13% (-0.05)	0.51% (0.07)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample accuracy. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table 5: Trading performance for USD/JPY

Model	2013	2014	2015	2016	Average
BNN (15)	**6.15% (0.68)**	6% (0.62)	4.23% (0.43)	6.84% (0.69)	**5.81% (0.61)**
DMA (15)	5.18% (0.53)	**6.03% (0.67)**	5.4% (0.58)	6.08% (0.63)	5.67% (0.6)
DMS (15)	4.14% (0.43)	5.42% (0.56)	4.91% (0.53)	5.67% (0.6)	5.04% (0.53)
BNN (10)	5.03% (0.54)	5.74% (0.63)	4.85% (0.51)	6% (0.62)	5.41% (0.58)
DMA (10)	4.43% (0.47)	5.97% (0.64)	4.67% (0.46)	5.36% (0.55)	5.11% (0.53)
DMS (10)	5.09% (0.55)	4.06% (0.41)	4.79% (0.42)	4.92% (0.49)	4.72% (0.47)
BNN (5)	4.1% (0.46)	4.91% (0.5)	**5.99% (0.62)**	6.78% (0.71)	5.45% (0.57)
DMA (5)	3.84% (0.38)	5% (0.53)	4.42% (0.45)	**7.33% (0.76)**	5.15% (0.53)
DMS (5)	5.62% (0.6)	3.04% (0.32)	4.83% (0.46)	5.7% (0.59)	4.8% (0.49)
RVM	5.22% (0.54)	4.75% (0.49)	4.01% (0.41)	5.47% (0.56)	4.86% (0.5)
NB (15)	3.07% (0.34)	3.06% (0.36)	2.77% (0.27)	2.12% (0.23)	2.76% (0.3)
NB (10)	2.18% (0.22)	2.92% (0.28)	3.58% (0.39)	1.87% (0.2)	2.64% (0.27)

NB (5)	2.44% (0.26)	2.12% (0.24)	2.09% (0.17)	1.61% (0.18)	2.07% (0.21)
SA (15)	-1.16% (-0.21)	0.1% (0.06)	0.21% (0.03)	0.86% (0.09)	0% (-0.01)
SA (10)	-0.2% (-0.09)	-3.63% (-0.42)	-1.72% (-0.15)	-2.09% (-0.23)	-1.91% (-0.22)
SA (5)	-2.24% (-0.3)	-2.28% (-0.25)	-2.75% (-0.21)	-3.33% (-0.4)	-2.65% (-0.29)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample accuracy. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table 6: Trading performance for the equal-weight portfolio

Model	2013	2014	2015	2016	Average
BNN (15)	**6.52% (0.67)**	5.89% (0.59)	5.26% (0.59)	6.71% (0.7)	6.09% (0.64)
DMA (15)	5.71% (0.58)	**6.24% (0.62)**	**6.2% (0.66)**	**6.87% (0.7)**	**6.25% (0.64)**
DMS (15)	4.71% (0.49)	5.46% (0.56)	5.75% (0.63)	6.29% (0.67)	5.55% (0.59)
BNN (10)	5.96% (0.58)	5.79% (0.6)	5.13% (0.59)	5.95% (0.61)	5.71% (0.6)
DMA (10)	5.07% (0.49)	5.36% (0.53)	5.73% (0.6)	5.26% (0.54)	5.36% (0.54)
DMS (10)	4.68% (0.47)	4.26% (0.45)	5.92% (0.58)	4.87% (0.55)	4.93% (0.51)
BNN (5)	5% (0.49)	4.61% (0.49)	5.4% (0.62)	6.01% (0.64)	5.26% (0.56)
DMA (5)	4.25% (0.42)	4.17% (0.41)	6.06% (0.63)	6.13% (0.63)	5.15% (0.52)
DMS (5)	4.65% (0.51)	3.34% (0.37)	5.71% (0.56)	4.47% (0.45)	4.54% (0.47)
RVM	4.73% (0.48)	4.07% (0.41)	4.05% (0.48)	4.66% (0.5)	4.38% (0.47)
NB (15)	2.9% (0.34)	2.22% (0.24)	2.46% (0.3)	2.77% (0.32)	2.59% (0.3)
NB (10)	2.23% (0.29)	2.76% (0.27)	2.33% (0.32)	2.7% (0.29)	2.51% (0.29)
NB (5)	2.07% (0.25)	2.12% (0.23)	2.72% (0.29)	2.22% (0.27)	2.28% (0.26)
SA (15)	-1.04% (-0.17)	0.89% (0.13)	0.27% (0.03)	0.19% (0.02)	0.07% (0)
SA (10)	0.59% (0.08)	0.16% (-0.01)	-0.75% (-0.05)	-0.86% (-0.08)	-0.21% (-0.02)
SA (5)	-1.07% (-0.07)	-0.43% (-0.02)	-0.99% (-0.03)	-1.08% (-0.09)	-0.89% (-0.06)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample accuracy. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Tables 3 to 6 show that all Bayesian combinations are capable of producing positive returns and Sharpe ratios, after transaction costs for the exchange rates and the periods under study. The DMA and the BNN seem to outperform their Bayesian counterparts and the SAs. This can be explained by the dynamic nature and time-varying coefficients of DMA and the highly non-linear features of BNN. On the other hand, the RVM that explores the whole set of genuine rules presents a marginally better performance than NB. DMS presents a consistent lower trading performance than the relevant DMA models. In trading, model averaging almost always works better than model selection and thus these results are not surprising. In general, we find that the trading performance increases three to four times with DMA and BNN, compared to the pool of surviving technical rules (see Table 2). It is also worth noting that all Bayesian combinations are statistically different from a random walk forecasts based on the Giacomini and White (2006) test.

Based on these results, Bayesian Statistics has value in trading and can considerably increase the profitability of the underlying trading systems. The models under study (DMA, DMS, BNN and RVM) are characterized by their complexity but can offer investors substantially increased returns. Similar to the concept of AMH, in highly competitive markets (such as FX) simple rules have a small value. Traders should seek complex non-linear models that can offer them an advantage over their competitors. DMA and DMS search all possible input combinations and select the optimal subset at each step, while BNN imitates the work of biological neurons and maps the non-linear dataset through Bayesian statistics. Unlike the simple technical rules that can be estimated by hand, none of the three Bayesian models can be used without the help of computer processors. However, complexity is always translated to an increased computational burden. This study was limited to subsets of the genuine profitable technical rules for DMA, DMS and BNN. While this protects the models from over-fitting, applying the whole set of genuine rules might have led to better results.

5. Conclusions

In this study, we explore the utility of technical analysis and Bayesian statistics in trading. For this purpose, 7846 technical rules are generated for the EUR/USD, GBP/USD and the USD/JPY exchange rates. Then, the genuinely profitable trading rules are identified with the help of the Romano *et al.* (2008) data snooping test combined with the balancing procedure of Romano and Wolf (2010). Finally, the profitable rules are combined with NB, RVM, DMA, DMS and BNN models. The motivation for this research is the AMH which states that complex models should have an advantage in highly competitive markets. The promising forecasting performance of Bayesian Statistics in our study confirms the proposals of the AMH.

In the results, we find that our data-snooping procedure identifies 5% to 15% of the technical rules as genuinely profitable. However, the generated portfolios based on them, present small annualized returns and Sharpe ratios over the OOS. When subsets of these rules are combined with the Bayesian models, we find that all Bayesian techniques increase the trading performance of the simple technical rules to 6% per annum after transactions costs and risk-free rate. Among the competing models, the DMA and the BNN clearly outperform their benchmarks. These results allow us to argue that market efficiency is variable, and it is possible to benefit from market inefficiencies with Bayesian Statistics.

Our results should go forward to convince traders and academics, to explore the recent development in statistics for procedures capable of providing an advantage in financial markets. These procedures might be characterized by complexity and are therefore inappropriate for high-frequency trading or large experiments. Nevertheless, the complex procedures in this paper can provide an edge in comparison to the traditional trading models.

References

Allen, F. and Karjalainen, R. 1999. Using genetic algorithms to find technical trading rules. Journal of Financial Economics, 51(2), pp. 245-271.

Alvarez-Diaz, M. and Alvarez, A. 2003. Forecasting exchange rates using genetic algorithms. Applied Economics Letters, 10(6), pp.319-322.

Aye, G., Gupta, R., Hammoudeh, S. and Kim, W.J., 2015. Forecasting the price of gold using dynamic model averaging. International Review of Financial Analysis, 41, pp.257-266.

Baillie, R.T. and Bollerslev, T., 1989. Common stochastic trends in a system of exchange rates. The Journal of Finance, 44(1), pp.167-181.

Baillie, R.T. and Chang, S.S., 2011. Carry trades, momentum trading and the forward premium anomaly. Journal of Financial Markets, 14(3), pp.441-464.

Bajgrowicz, P. and Scaillet, O., 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. Journal of Financial Economics, 106(3), pp.473-491.

Baltas, N. and Karyampas, D., 2018. Forecasting the equity risk premium: The importance of regime-dependent evaluation. Journal of Financial Markets, 38, pp.83-102.

Bekiros, S. D., 2010. Heterogeneous trading strategies with adaptive fuzzy actor–critic reinforcement learning: A behavioral approach. Journal of Economic Dynamics and Control, 34(6), pp.1153-1170.

BIS, 2013. Triennial Central Bank Survey of foreign exchange and derivatives market activity in 2013, Bank for International Settlement, online report.

Boston Consulting Group, 2015. Where machines could replace humans—and where they can't (yet), Boston Consulting Group, online report

Blume, L., Easley, D. and O'hara, M., 1994. Market statistics and technical analysis: The role of volume. The Journal of Finance, 49(1), pp.153-181.

Brock, W., Lakonishok, J. and LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. The Journal of finance, 47(5), pp.1731-1764.

Burden, F., & Winkler, D., (2009). Bayesian regularization of neural networks. Artificial Neural Networks: Methods and Applications, 23-42.

Byrne, J. P., Korobilis, D. and Ribeiro, P. J., 2016. Exchange rate predictability in a changing world. Journal of International Money and Finance, *62*, 1-24.

Candela, J. Q. and Hansen, L. K., 2004. Learning with uncertainty-Gaussian processes and relevance vector machines, Doctoral Dissertation,

Chinn, M. D. and Meese, R. A., 1995. Banking on currency forecasts: How predictable is change in money?. Journal of International Economics, 38(1), pp.161-178.

Chui, M., Manyika, J. and Miremadi, M., 2016. Where machines could replace humans—and where they can't (yet), Mckinsey Quarterly

Cialenco, I. and Protopapadakis, A., 2011. Do technical trading profits remain in the foreign exchange market? Evidence from 14 currencies. Journal of International Financial Markets, Institutions and Money, 21(2), pp.176-206.

Edwards, R. D., Magee, J. and Bassetti, W. C. 2007. Technical analysis of stock trends. CRC Press.

Elaut, G., Frömmel, M. and Lampaert, K., 2018. Intraday momentum in FX markets: Disentangling informed trading from liquidity provision. Journal of Financial Markets, 37, pp. 35-51.

Fama, E. F. (1965). The behaviour of stock-market prices. The Journal of Business, 38(1), pp. 34-105.

Fernández-Rodríguez, F., González-Martel, C and Sosvilla-Rivero, S., S. 2000. On the profitability of technical trading rules based on artificial neural networks:: Evidence from the Madrid stock market. Economics letters, 69(1), pp.89-94.

Fletcher, T., Redpath, F. and D'Alessandro, J., 2009. Machine learning in FX carry basket prediction. In Proceedings of the World Congress on Engineering (Vol. 2).

Gençay, R. 1998. The predictability of security returns with simple technical trading rules. Journal of Empirical Finance, 5(4), pp. 347-359.

Gencay, R., Dacorogna, M., Olsen, R. and Pictet, O., 2003. Foreign exchange trading models and market behaviour. Journal of Economic Dynamics and Control, 27(6), pp.909-935.

Gehrig, T. and Menkhoff, L. 2006. Extended evidence on the use of technical analysis in foreign exchange. International Journal of Finance & Economics, 11(4), pp. 327-338.

Ghosh, S. and Mujumdar, P.P., 2008. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Advances in water resources, 31(1), pp.132-146.

Gradojevic, N., 2007. Non-linear, hybrid exchange rate modelling and trading profitability in the foreign exchange market. Journal of Economic Dynamics and Control, 31(2), pp.557-574.

Gradojevic, N. and Gençay, R. 2013. Fuzzy logic, trading uncertainty and technical trading. Journal of Banking & Finance, 37(2), pp.578-586.

Gramacy, R., Malone, S. W. and Horst, E. T. 2014. Exchange rate fundamentals, forecasting, and speculation: Bayesian models in black markets. Journal of Applied Econometrics, 29(1), pp.22-41.

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

Hansen, P. R. and Lunde, A. 2005. A forecast comparison of volatility models: does anything beat a GARCH (1, 1)?. Journal of applied econometrics, 20(7), pp. 873-889.

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. Journal of Empirical Finance, 17(3), pp.471-484.

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

Hsu, Y., Kuan, C. and Yen, M., 2014. A generalized stepwise procedure with improved power for multiple inequalities testing. Journal of Financial Econometrics, 12(4), pp. 730–755.

Huang, S.C., Chuang, P.J., Wu, C.F and Lai, H.J. 2010 Chaos-based support vector regressions for exchange rate forecasting, Expert Systems with Applications, 37(12), pp. 8590-8598.

Jasic, T. and Wood, D. 2004. The profitability of daily stock market indices trades based on neural network predictions: Case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965–1999. Applied Financial Economics, 14(4), pp.285-297.

Jorion, P., 1996. Risk and turnover in the foreign exchange market. In Frankel, J., Galli, G. and Giovannini, A. (eds) The microstructure of foreign exchange markets, Chicago, Ill.: University of Chicago Press, pp.19-40.

Kearney, F., Cummins, M. and Murphy, F., 2014. Outperformance in exchange-traded fund pricing deviations: Generalized control of data snooping bias. Journal of Financial Markets, 19, pp.86-109.

Kilian, L. and Taylor, M.P., 2003. Why is it so difficult to beat the random walk forecast of exchange rates?. Journal of International Economics, 60(1), pp.85-107.

Koop, G. and Korobilis, D., 2012. Forecasting inflation using dynamic model averaging. International Economic Review, 53(3), pp.867-886.

Lin, K.P. and Pai, P.F. 2010. A fuzzy support vector regression model for business cycle predictions, Expert Systems with Applications, 37 (7), pp. 5430-5435.

Lo, A.W., 2004. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. Journal of Portfolio Management, 30, pp. 15-29.

MacDonald, R. and Taylor, M.P., 1994. The monetary model of the exchange rate: long-run relationships, short-run dynamics and how to beat a random walk. Journal of International Money and finance, 13(3), pp.276-290.

MacKay, D.J., 1992. Bayesian interpolation. Neural Computation, 4(3), pp.415-447.

Mazzocco, M. and Saini, S., 2012. Testing efficient risk sharing with heterogeneous risk preferences. The American Economic Review, 102(1), pp.428-468.

Meese, R.A. and Rogoff, K., 1983. Empirical exchange rate models of the seventies: Do they fit out of sample?. Journal of international economics,14(1), pp.3-24.

Neely, C., Weller, P. and Dittmar, R. 1997. Is technical analysis in the foreign exchange market profitable? A genetic programming approach. Journal of Financial and Quantitative Analysis, 32(4), pp.405-426.

Neely, C.J., Weller, P.A. and Ulrich, J.M., 2009. The adaptive markets hypothesis: evidence from the foreign exchange market. Journal of Financial and Quantitative Analysis, 44(2), pp.467-488.

Neely, C.J. and Weller, P.A., 2013. Lessons from the evolution of foreign exchange trading strategies. Journal of Banking & Finance, 37(10), pp.3783-3798.

Pai, P.F., Lin, C.S., Hong, W.C. and Chen, C.T. 2006 A hybrid support vector machine regression for exchange rate prediction, International Journal of Information and Management Sciences, 17 (2), pp. 19-32.

Park, J. and Sandberg, I.W., 1991. Universal approximation using radial-basis-function networks. Neural Computation, 3(2), pp.246-257.

Qi, M. and Wu, Y., 2006. Technical trading-rule profitability, data snooping, and reality check: evidence from the foreign exchange market. Journal of Money, Credit, and Banking, 38(8), pp.2135-2158.

Raftery, A.E., Kárný, M. and Ettler, P., 2010. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics, 52(1), pp.52-66.

Romano, J.P. and Wolf, M., 2005. Stepwise multiple testing as formalized data snooping. Econometrica, 73(4), pp.1237-1282.

Romano, J.P. and Wolf, M., 2007. Control of generalized error rates in multiple testing. The Annals of Statistics, pp.1378-1408.

Romano, J.P., Shaikh, A.M. and Wolf, M., 2008. Formalized data snooping based on generalized error rates. Econometric Theory, 24(02), pp.404-447.

Romano, J.P. and Wolf, M., 2010. Balanced control of generalized error rates. The Annals of Statistics, pp.598-633.

Sermpinis, G., Theofilatos, K.A, Karathanasopoulos, A.S., Georgopoulos, E.F. and Dunis, C.L. 2013. Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and Particle Swarm Optimization, European Journal of Operational Research, 225 (3), pp. 528–540.

Sermpinis, G., Laws, J. and Dunis, C.L., 2015. Modelling commodity value at risk with Psi Sigma neural networks using open–high–low–close data. The European Journal of Finance, 21(4), pp.316-336.

Shapiro, A. F. (2000), A Hitchhiker's Guide to the Techniques of Adaptive Nonlinear Models, Insurance, Mathematics and Economics, 26(2), pp. 119-132.

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

Sweeney, R. J. 1988. Some new filter rule tests: Methods and results. Journal of Financial and Quantitative Analysis, 23(03), pp. 285-300.

Taylor, M.P. 1992. The use of technical analysis in the foreign exchange market. Journal of International Money and Finance, 11(3), pp.304-314.

Ticknor, J. L. 2013. A Bayesian regularized artificial neural network for stock market forecasting. Expert Systems with Applications, 40(14), pp. 5501-5506.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. Journal of machine learning research, 1(Jun), pp.211-244.

Vapnik, V.N., 1998. *Statistical learning theory*. New York: Wiley.

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

Yang, J., Su, X. and Kolari, J.W., 2008. Do Euro exchange rates follow a martingale? Some out-of-sample evidence. Journal of Banking & Finance,32(5), pp.729-740.

Yegnanarayana, B. 2009. *Artificial neural networks*. PHI Learning Pvt. Ltd.

Yilmaz, K., 2003. Martingale Property of Exchange Rates and Central Bank Interventions. Journal of Business & Economic Statistics, pp.383-395.

Zhang, G., Patuwo, B. E. and Hu, M. Y. 1998. Forecasting with artificial neural networks: The state of the art. International journal of forecasting, 14(1), 35-62.

Appendices

A. Technical trading rules

Technical trading strategies involve using quotes for open, high, low and close prices along with the trading volumes of every ticker under study. Their purpose is to recognize patterns or trends in the price charts. In this paper, 7846 rules are studied, including filter rules, moving averages, support and resistance levels, channel breakouts, and on-balance volumes indicators. Short descriptions of these rules are presented in the following subsections.

A.1 Filter Rules (FRs)

The filter strategy is based on making a financial decision to undertake a long or short position when the security price moves a certain amount e.g. x percent upward or downward. A buy order is placed when the x percentage upward movement is seen in the market and this position is held until the price falls x percent where the position is first neutralized. Then a short position is opened and kept until a subsequent upward movement is seen. Movements that are smaller than the filter level in either direction are discarded as noise.

Tuning the FIRs is at the analyst's discretion. The definition of upward/downward movement, the holding process and liquidating/closing the position can be subjective. Upward (downward) movements are recognized as an uptrend when the price exceeds the last high (low) by x percent. The last high (low) can be defined either as the highest (lowest) close price observed in a long (short) position; or the maximum (minimum) close price over the last d days. The position holding can also be modified. Another case is considered in which the position is opened by the FIR and held for h days where signals over this period are ignored. The strategy may also include a neutral position where positions are closed in case of y percent backward movement compared to extrema level. The filter level for liquidating the position must be less than the filter level for opening a position.

Considering the following sets of possible x, y:

$$x \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30, 40, 50\}\ in\ \%$$

$$y \in \{0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 7.5, 10, 15, 20\}\ in\ \%,$$

$$(\#x = 24, \#y = 12)$$

Then, the number of $x - y$ combinations, given that $y < x$, are $\#(x - y) = 185$.

For these combinations, we experiment with the following d, h:

$d \in \{1, 2, 3, 4, 5, 10, 15, 20\}, (\#d = 8)$

$$h \in \{5, 10, 25, 50\}, (\#h = 4)$$

Based on the above, in this application we examine a total of 497 filter rules as calculated below:

$$\#F = \#x + \#x \times \#d + \#x \times \#h + \#(x - y) = 24 + 192 + 96 + 185 = 49 \qquad (A.1)$$

A.2 Moving averages (MAs)

In technical trading MAs play an integral role. Trends are not considered robust until they are reflected in MAs. When a change in price is also visible in the MAs, it means that the news or change source is important enough to last over a period of time and can be taken into account. Uptrends start to form when a fast MA exceeds the slow MA. The fast MA can simply be the price quote or a short-term average. The long positions are kept so long as the price remains above the MA benchmark. When the price falls below the MA, the downtrend is initiated. At this point the previous position is liquidated and a sell position is opened. The new position remains open until another upward penetration is observed. MA crossover-based strategies may appear from a wide variation. In this paper, simple forms of MA crossover along with some filter and delays are taken into account. A common application of MA is having a fast and slow MA and looking for their crossovers signalling/which signal up and downtrends.

A buy signal can be generated when the fast MA goes beyond the slow MA. The fast and slow MAs come with parameters n and m respectively showing the number of days taken into consideration ($n < m$). Similarly, a sell signal is created when the fast MA drops below the slow MA, which suggests the formation of a downtrend. The fast and slow MA strategy can be accompanied by a band (b) filter to avoid the noise in trend detection. Trends are deemed solid only if the fast MA can exceed the slow MA by b percent. Alternatively, the time lag l is considered between opening a position and taking any action. During the lag period, all signals are ignored. Another innovation is holding each position for a fixed period of h days no matter what the signals are after the opening of the positions. In our application, we consider innovations separately and impose only one filter at a time.

We consider the following sets of possible n, m:

$$n \in \{2, 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 200, 250\}, (\#n = 15)$$

$$m \in \{2, 5, 10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, 200\}, (\#m = 14)$$

Then, the number of $n - m$ combinations, given are $m < n$ are $\#(n - m) = 105$.

For these combinations, we experiment with the following b, l, h:

$$b \in \{0.1, 0.5, 1, 1.5, 2, 3, 4, 5\}\ in\ \%, (\#b = 8)$$

$$l \in \{2, 3, 4, 5\}, (\#l = 4)$$

$$h \in \{5, 10, 25, 50\}, (\#h = 4)$$

The band filter is set at 1% . A 10-day holding period is applied to all combinations of MA crossovers. For the fast and slow MA we set respectively $n = 1, 2, 5$ and m =50, 150, 200. We also include 9 cases of double-filters. Based on the above, in this application we examine in total 2049 MA rules as calculated below:

$$\#MA = \#n + \#(n - m) + \#b \times \big(\#n + \#(n - m)\big) + \#l \times \big(\#n + \#(n - m)\big) + \#h \times \big(\#n + \#(n - m)\big) + 9 = 15 + 105 + 960 + 480 + 480 + 9 = 2049 \quad \text{(A.2)}$$

A.3 Support and Resistance (S&R) levels

The S&R trading rules are based on the premise that the price should remain in a trading range capped by a resistance and floored by a support level. Breaching these levels suggests that a stock or an exchange rate would move in the same direction. The S&R rules are constructed similarly to the FIRs. The only difference is that trading signals are generated when the rate under study breaks the support or resistance barriers by a certain percentage.

The S&R levels can be defined as the intra-day low and intra-day high quotes over the past n days. Another variation in the definition of the S&R is to calculate the support and resistance level based on the minimum and maximum closing prices over the past e days. Alternative S&Rs are set by using a fixed band filter for noise removal: the holding period h, the l-day lag before making any decisions, a combination of a fixed holding period on a position, and a delay in decision making before undergoing any new positions.

Based on the above, we consider the following possible sets:

$$n \in \{5, 10, 15, 20, 25, 50, 100, 150, 200, 250\}, (\#n = 10)$$

$$e \in \{2, 3, 4, 5, 10, 20, 25, 50, 100, 200\}, (\#e = 10)$$

$$b \in \{0.1, 0.5, 1, 1.5, 2, 3, 4, 5\}\ in\ \%, (\ \#b = 8)$$

$$l \in \{2, 3, 4, 5\}, (\#l = 4)$$

$$h \in \{5, 10, 25, 50\}, (\#h = 4)$$

In accordance with these sets, we examine a total of 1220 S&R rules as calculated below:

$$\#S\&R = [(\#n + \#e) \times (1 + \#h)] + [(\#n + \#e) \times (1 + \#h) \times \#b] + [(\#n + \#e) \times \#h \times \#l] = 100 + 800 + 320 = 1220 \quad \text{(A.3)}$$

A.4 Channel Breakout (CBs)

Based on the principles of S&R, practitioners can detect time-varying support and resistance levels that drift together within a certain range. This creates the so-called trading channel. Once a trading channel is formed, then a CB rule can be applied. The premise behind the CB rule is that once the trading channel is breached, there will be a substantial trend towards the same direction. A channel is formed when the highest observed price remains within a c% range above the lowest price over the past n days. The trend is considered significant, when the price breaks one of the channel borders, which generates a buy (sell) order after an upward (downward) breakout. As in the previous categories discussed in Sections A.1.1 to A.1.3, we also consider CB alternatives with a fixed filter band b and holding period h.

Here we look at the following possible sets:

$$n \in \{5, 10, 15, 20, 25, 50, 100, 150, 200, 250\}, (\#n = 10)$$

$$c \in \{0.5, 1, 2, 3, 5, 7.5, 10, 15\}\ in\ \%, (\#c = 8)$$

$$b \in \{0.1, 0.5, 1, 1.5, 2, 3, 4, 5\}\ in\ \%, (\#b = 8)$$

$$h \in \{5, 10, 25, 50\}, (\#h = 4)$$

Given $b < c$, the number of $c - b$ combinations are $\#(c - b) = 43$.

In this application we examine in total 2040 CB rules:

$$\#CB = \#n \times \#c \times \#h + \#n \times \#(c - b) \times \#h = 320 + 1720 = 2040 \qquad \text{(A.4)}$$

A.5 On-Balance Volumes (OBVs)

In the technical trading context, prices and trading volumes are expected to move together. Trading volumes confirm the potential significance of price moves. In case of major economic events or important news, increased trading volumes reflect decisions in favour of or against the price change. Therefore, monitoring the volumes and their changes can be a useful source of information for the practitioner. The OBV line is simply a running total of positive and negative volumes. In other words, if the closing price is above (below) the prior close price, then the current OBV is the sum (difference) of the previous OBV and the current volume. When the volume is not increasing during bullish days, it is a sign that buying pressure is weakening and the upward trend is probably not sustainable. OBVs are usually used with MAs to generate trading signals. In this scenario, the average OBV is calculated and then combined with slow and fast MAs. In our application, we use the MAs as in Section A.1.2, excluding the 9 double-filter cases. Based on these, we examine a total of 2040 OBV rules as calculated below:

$$\#OBV = \#n + \#(n - m) + \#b \times \big(\#n + \#(n - m)\big) + \#l \times \big(\#n + \#(n - m)\big) + \#h \times \big(\#n + \#(n - m)\big) = 15 + 105 + 960 + 480 + 480 = 2040 \qquad \text{(A.5)}$$

A.6 Trading universe

The Trading Universe (TU) consists of the total number of trading rules reported in the previous subsections:

$$\#TU = \#F + \#MA + \#S\&R + \#CB + \#OBV = 497 + 2049 + 1220 + 2040 + 2040 = 7846$$

B. Robustness checks

In this appendix, we present the results from Bayesian trading with the top rule selection metric of in-sample return (Tables A.1 to A.4) and Sharpe ratio (Tables A.5 to A.8). The presented tables are comparable to Tables 3 to 6 in the main text.

Table A.1: Trading performance for EUR/USD – profitability metric

Model	2013	2014	2015	2016	Average
BNN (15)	6% (0.58)	5.04% (0.33)	4.9% (0.68)	6.05% (0.51)	5.5% (0.53)
DMA (15)	5.49% (0.59)	3.43% (0.32)	**6.49% (0.69)**	5.28% (0.49)	5.17% (0.52)
DMS (15)	4.99% (0.49)	3.25% (0.24)	4.36% (0.56)	5.94% (0.44)	4.64% (0.43)
BNN (10)	6.51% (0.63)	5.38% (0.34)	5.29% (0.66)	6.01% (0.5)	**5.8% (0.53)**
DMA (10)	4.26% (0.44)	3.82% (0.3)	4.54% (0.58)	**6.28% (0.52)**	4.73% (0.46)
DMS (10)	5.06% (0.55)	3.14% (0.26)	4.22% (0.54)	5.8% (0.41)	4.56% (0.44)
BNN (5)	5.91% (0.57)	3.01% (0.22)	4.77% (0.58)	5.12% (0.41)	4.7% (0.45)
DMA (5)	4.55% (0.46)	**6.97% (0.41)**	4.95% (0.61)	5.46% (0.47)	5.48% (0.49)
DMS (5)	**6.86% (0.67)**	3.26% (0.29)	3.92% (0.49)	3.35% (0.38)	4.35% (0.46)
RVM	4.35% (0.5)	3.28% (0.31)	4.41% (0.63)	4.24% (0.47)	4.07% (0.48)
NB (15)	1.31% (0.26)	1.96% (0.26)	1.74% (0.37)	3.02% (0.33)	2.01% (0.31)
NB (10)	1.58% (0.29)	1.79% (0.26)	1.52% (0.36)	1.8% (0.26)	1.67% (0.29)
NB (5)	0.63% (0.14)	2.04% (0.28)	2.14% (0.4)	-2.66% (-0.24)	0.54% (0.15)

SA (15)	-0.68% (-0.11)	1.78% (0.22)	-1.08% (-0.28)	1.08% (0.18)	0.28% (0)
SA (10)	-2.11% (-0.18)	-0.82% (-0.11)	-1.46% (-0.07)	1.24% (0.2)	-0.79% (-0.04)
SA (5)	-0.6% (-0.08)	-0.7% (-0.07)	1.24% (0.34)	-0.32% (-0.08)	-0.1% (0.03)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample excess-return. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table A.2: Trading performance for GBP/USD – profitability metric

Model	2013	2014	2015	2016	Average
BNN (15)	**6.48% (0.66)**	5.14% (0.61)	6.9% (0.7)	**7.54% (0.8)**	**6.52% (0.69)**
DMA (15)	5.95% (0.63)	5.29% (0.63)	**7.24% (0.81)**	7.4% (0.78)	6.47% (0.71)
DMS (15)	5.04% (0.48)	5.05% (0.58)	6.92% (0.72)	6.17% (0.66)	5.8% (0.61)
BNN (10)	5% (0.5)	**5.87% (0.67)**	5.84% (0.63)	5.95% (0.61)	5.67% (0.6)
DMA (10)	5.63% (0.57)	3.85% (0.42)	4.79% (0.49)	5.61% (0.62)	4.97% (0.53)
DMS (10)	5.67% (0.59)	3.04% (0.33)	4.42% (0.46)	5.2% (0.5)	4.58% (0.47)
BNN (5)	5.28% (0.51)	4.16% (0.45)	5.35% (0.6)	6.02% (0.64)	5.2% (0.55)
DMA (5)	4.87% (0.51)	3.64% (0.39)	4.17% (0.45)	4.87% (0.51)	4.39% (0.47)
DMS (5)	4.18% (0.4)	3% (0.31)	3.05% (0.32)	3.91% (0.41)	3.54% (0.36)
RVM	4.61% (0.41)	4.19% (0.43)	3.72% (0.39)	4.27% (0.48)	4.2% (0.43)
NB (15)	3.78% (0.39)	4.04% (0.43)	4.09% (0.43)	4.02% (0.48)	3.98% (0.43)
NB (10)	2.7% (0.28)	3% (0.41)	3.8% (0.39)	3.43% (0.37)	3.23% (0.36)
NB (5)	1.56% (0.21)	2.42% (0.33)	2% (0.22)	2.3% (0.26)	2.07% (0.26)
SA (15)	-0.05% (-0.02)	0.06% (0.04)	-0.99% (-0.13)	0.34% (0.07)	-0.16% (-0.01)
SA (10)	-0.77% (-0.08)	0.9% (0.14)	-1.92% (-0.24)	1.25% (0.14)	-0.14% (-0.01)
SA (5)	-3.1% (-0.4)	0.71% (0.11)	1.25% (0.14)	1.99% (0.18)	0.21% (0.01)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample excess-return. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table A.3: Trading performance for USD/JPY – profitability metric

Model	2013	2014	2015	2016	Average
BNN (15)	6.01% (0.63)	6.12% (0.64)	**6.58% (0.69)**	5.09% (0.52)	5.95% (0.62)
DMA (15)	**6.96% (0.74)**	5.54% (0.58)	6.36% (0.67)	7.32% (0.75)	**6.55% (0.69)**
DMS (15)	6.19% (0.63)	5.91% (0.62)	6.44% (0.68)	6.04% (0.62)	6.15% (0.64)
BNN (10)	4.88% (0.51)	5.06% (0.53)	4.9% (0.53)	5.82% (0.59)	5.17% (0.54)
DMA (10)	4.45% (0.48)	**6.76% (0.73)**	4.21% (0.45)	**8.3% (0.86)**	5.93% (0.63)
DMS (10)	3.86% (0.4)	2.75% (0.33)	3.09% (0.33)	7.11% (0.7)	4.2% (0.44)
BNN (5)	5.09% (0.55)	6.36% (0.7)	5.43% (0.56)	5.23% (0.54)	5.53% (0.59)
DMA (5)	5.88% (0.64)	5.4% (0.56)	4.5% (0.48)	7.07% (0.74)	5.71% (0.61)
DMS (5)	4.74% (0.49)	4.45% (0.47)	3.91% (0.4)	6.07% (0.62)	4.79% (0.5)
RVM	5.22% (0.54)	4.75% (0.49)	4.01% (0.41)	5.47% (0.56)	4.86% (0.5)
NB (15)	2.04% (0.25)	2.24% (0.2)	2.7% (0.26)	2.04% (0.21)	2.26% (0.23)
NB (10)	2.86% (0.29)	1.37% (0.14)	0.32% (0.07)	2.15% (0.2)	1.68% (0.18)
NB (5)	3.85% (0.42)	2.48% (0.26)	2.26% (0.25)	2.6% (0.24)	2.8% (0.29)

SA (15)	0.98% (0.15)	0.44% (0.1)	-0.93% (-0.12)	-0.62% (-0.11)	-0.03% (0.01)
SA (10)	3.54% (0.4)	2.46% (0.25)	-1.45% (-0.16)	-2.96% (-0.28)	0.4% (0.05)
SA (5)	2.72% (0.31)	1.88% (0.17)	-1.77% (-0.19)	-2.63% (-0.24)	0.05% (0.01)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample excess-return. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table A.4: Trading performance for the equal-weight portfolio – profitability metric

Model	2013	2014	2015	2016	Average
BNN (15)	**6.16% (0.62)**	**5.43% (0.53)**	6.13% (0.69)	6.23% (0.61)	5.99% (0.61)
DMA (15)	6.13% (0.65)	4.75% (0.51)	**6.7% (0.72)**	6.67% (0.67)	**6.06% (0.64)**
DMS (15)	5.41% (0.53)	4.74% (0.48)	5.91% (0.65)	6.05% (0.57)	5.53% (0.56)
BNN (10)	5.46% (0.55)	5.44% (0.51)	5.34% (0.61)	5.93% (0.57)	5.54% (0.56)
DMA (10)	4.78% (0.5)	4.81% (0.48)	4.51% (0.51)	**6.73% (0.67)**	5.21% (0.54)
DMS (10)	4.86% (0.51)	2.98% (0.31)	3.91% (0.44)	6.04% (0.54)	4.45% (0.45)
BNN (5)	5.43% (0.54)	4.51% (0.46)	5.18% (0.58)	5.46% (0.53)	5.14% (0.53)
DMA (5)	5.1% (0.54)	5.34% (0.45)	4.54% (0.51)	5.8% (0.57)	5.19% (0.52)
DMS (5)	5.26% (0.52)	3.57% (0.36)	3.63% (0.4)	4.44% (0.47)	4.23% (0.44)
RVM	4.73% (0.48)	4.07% (0.41)	4.05% (0.48)	4.66% (0.5)	4.38% (0.47)
NB (15)	2.38% (0.3)	2.75% (0.3)	2.84% (0.35)	3.03% (0.34)	2.75% (0.32)
NB (10)	2.38% (0.29)	2.05% (0.27)	1.88% (0.27)	2.46% (0.28)	2.19% (0.28)
NB (5)	2.01% (0.26)	2.31% (0.29)	2.13% (0.29)	0.75% (0.09)	1.8% (0.23)
SA (15)	0.08% (0.01)	0.76% (0.12)	-1% (-0.18)	0.27% (0.05)	0.03% (0)
SA (10)	0.22% (0.05)	0.85% (0.09)	-1.61% (-0.16)	-0.16% (0.02)	-0.18% (0)
SA (5)	-0.33% (-0.06)	0.63% (0.07)	0.24% (0.1)	-0.32% (-0.05)	0.06% (0.02)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample excess-return. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table A.5: Trading performance for EUR/USD – Sharpe ratio metric

Model	2013	2014	2015	2016	Average
BNN (15)	5.96% (0.55)	**7.49% (0.62)**	6.08% (0.72)	6.19% (0.59)	**6.43% (0.62)**
DMA (15)	5.4% (0.55)	5.72% (0.59)	5.23% (0.65)	4.14% (0.37)	5.12% (0.54)
DMS (15)	2.63% (0.26)	6.12% (0.49)	5.86% (0.69)	4.01% (0.42)	4.66% (0.47)
BNN (10)	5.01% (0.52)	5.38% (0.48)	**7.27% (0.74)**	**7.44% (0.7)**	6.28% (0.61)
DMA (10)	**6.02% (0.58)**	6.18% (0.55)	6.98% (0.72)	5.8% (0.42)	6.25% (0.57)
DMS (10)	5.43% (0.58)	6.03% (0.55)	4.44% (0.63)	5.93% (0.4)	5.46% (0.54)
BNN (5)	4.82% (0.48)	6.32% (0.52)	4.82% (0.71)	5.57% (0.54)	5.38% (0.56)
DMA (5)	4.99% (0.48)	7.09% (0.53)	4.59% (0.54)	5.14% (0.39)	5.45% (0.49)
DMS (5)	4.09% (0.52)	5.17% (0.53)	4.3% (0.61)	5.02% (0.38)	4.65% (0.51)
RVM	4.35% (0.5)	3.28% (0.31)	4.41% (0.63)	4.24% (0.47)	4.07% (0.48)
NB (15)	1.82% (0.31)	2% (0.15)	3.07% (0.49)	2.77% (0.28)	2.42% (0.31)
NB (10)	2.18% (0.35)	3.04% (0.18)	2.72% (0.47)	1.49% (0.11)	2.36% (0.28)
NB (5)	1.1% (0.18)	1.11% (0.13)	2.27% (0.44)	1.82% (0.25)	1.58% (0.25)
SA (15)	-2.22% (-0.21)	2.26% (0.17)	-0.64% (-0.2)	-1.44% (-0.1)	-0.51% (-0.09)
SA (10)	-0.19% (0.06)	1.6% (0.14)	-3.12% (-0.37)	-3.01% (-0.29)	-1.18% (-0.12)
SA (5)	-0.09% (-0.02)	1.95% (0.12)	-2.37% (-0.22)	-2.69% (-0.22)	-0.8% (-0.09)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample Sharpe ratio. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table A.6: Trading performance for GBP/USD – Sharpe ratio metric

Model	2013	2014	2015	2016	Average
BNN (15)	3.7% (0.38)	4.21% (0.49)	**6.29% (0.68)**	3.15% (0.36)	4.34% (0.48)
DMA (15)	3.76% (0.36)	4.09% (0.49)	4.44% (0.48)	5.09% (0.52)	4.35% (0.46)
DMS (15)	3.02% (0.32)	3.59% (0.34)	5.37% (0.57)	4.82% (0.52)	4.2% (0.44)
BNN (10)	4.74% (0.45)	5.14% (0.57)	3.11% (0.32)	5.31% (0.56)	4.58% (0.48)
DMA (10)	**6.22% (0.65)**	**5.67% (0.61)**	5.04% (0.51)	**5.98% (0.64)**	**5.73% (0.6)**
DMS (10)	3.69% (0.38)	4.5% (0.48)	5.3% (0.57)	4.44% (0.49)	4.48% (0.48)
BNN (5)	6.12% (0.63)	4.94% (0.53)	5.02% (0.54)	5.28% (0.57)	5.34% (0.57)
DMA (5)	3.98% (0.41)	5.57% (0.56)	4.32% (0.46)	5.34% (0.59)	4.8% (0.51)
DMS (5)	4.16% (0.42)	5.42% (0.57)	3.41% (0.39)	4.09% (0.43)	4.27% (0.45)
RVM	4.61% (0.41)	4.19% (0.43)	3.72% (0.39)	4.27% (0.48)	4.2% (0.43)
NB (15)	1.67% (0.2)	0.77% (0.15)	1.22% (0.17)	3.6% (0.37)	1.82% (0.22)
NB (10)	3.09% (0.34)	1.15% (0.2)	3.39% (0.42)	3.18% (0.34)	2.7% (0.33)
NB (5)	1.1% (0.13)	1.94% (0.23)	2.52% (0.26)	2.79% (0.25)	2.09% (0.22)
SA (15)	-2.48% (-0.26)	-0.86% (-0.11)	-2.58% (0.31)	-0.17% (-0.04)	-1.52% (-0.03)
SA (10)	-3.32% (-0.39)	1% (0.14)	2.61% (0.27)	2.58% (0.27)	0.72% (0.07)
SA (5)	0.08% (-0.03)	-2.49% (-0.29)	-1.32% (-0.21)	0.36% (0.05)	-0.84% (-0.12)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample Sharpe ratio. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table A.7: Trading performance for USD/JPY – Sharpe ratio metric

Model	2013	2014	2015	2016	Average
BNN (15)	5.7% (0.6)	5.88% (0.61)	5.51% (0.5)	4.9% (0.48)	5.5% (0.55)
DMA (15)	6.29% (0.65)	3.22% (0.35)	4.02% (0.41)	6.54% (0.67)	5.02% (0.52)
DMS (15)	4.3% (0.46)	5.06% (0.53)	3.26% (0.3)	5.28% (0.54)	4.48% (0.46)
BNN (10)	5.47% (0.57)	5.35% (0.57)	5.7% (0.62)	**6.71% (0.7)**	5.81% (0.62)
DMA (10)	5.51% (0.61)	5.88% (0.6)	5.7% (0.59)	5.91% (0.58)	5.75% (0.6)
DMS (10)	4.93% (0.52)	4.31% (0.46)	4.85% (0.47)	4.82% (0.49)	4.73% (0.49)
BNN (5)	4.84% (0.5)	3.42% (0.4)	4.36% (0.45)	4.55% (0.43)	4.29% (0.45)
DMA (5)	**7.22% (0.76)**	**6.03% (0.64)**	**6.19% (0.63)**	4.69% (0.43)	**6.03% (0.62)**
DMS (5)	6.65% (0.69)	2.11% (0.25)	2.72% (0.28)	4.63% (0.45)	4.03% (0.42)
RVM	5.22% (0.54)	4.75% (0.49)	4.01% (0.41)	5.47% (0.56)	4.86% (0.5)
NB (15)	3.62% (0.34)	3.24% (0.36)	0.45% (0.06)	1.28% (0.15)	2.15% (0.23)
NB (10)	4.24% (0.46)	2.18% (0.24)	0.3% (0.04)	2.44% (0.21)	2.29% (0.24)
NB (5)	3.45% (0.42)	1.15% (0.11)	-1.59% (-0.19)	-1.37% (-0.16)	0.41% (0.05)
SA (15)	3.74% (0.4)	0.8% (0.05)	-1.52% (-0.17)	-3.1% (-0.35)	-0.02% (-0.02)
SA (10)	4.07% (0.45)	1.65% (0.18)	1.04% (0.1)	-1.82% (-0.21)	1.24% (0.13)
SA (5)	3.09% (0.32)	2.31% (0.25)	-2.96% (-0.31)	-2.79% (-0.3)	-0.09% (-0.01)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample Sharpe ratio. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.

Table A.8: Trading performance for the equal-weight portfolio – Sharpe ratio metric

Model	2013	2014	2015	2016	Average
BNN (15)	5.12% (0.51)	5.86% (0.57)	**5.96% (0.63)**	4.75% (0.48)	5.42% (0.55)
DMA (15)	5.15% (0.52)	4.34% (0.48)	4.56% (0.51)	5.26% (0.52)	4.83% (0.51)
DMS (15)	3.32% (0.35)	4.92% (0.45)	4.83% (0.52)	4.7% (0.49)	4.44% (0.45)
BNN (10)	5.07% (0.51)	5.29% (0.54)	5.36% (0.56)	**6.49% (0.65)**	5.55% (0.57)
DMA (10)	**5.92% (0.61)**	5.91% (0.59)	5.91% (0.61)	5.9% (0.55)	**5.91% (0.59)**
DMS (10)	4.68% (0.49)	4.95% (0.5)	4.86% (0.56)	5.06% (0.46)	4.89% (0.5)
BNN (5)	5.26% (0.54)	4.89% (0.48)	4.73% (0.57)	5.13% (0.51)	5.01% (0.53)
DMA (5)	5.4% (0.55)	**6.23% (0.58)**	5.03% (0.54)	5.06% (0.47)	5.43% (0.54)
DMS (5)	4.97% (0.54)	4.23% (0.45)	3.48% (0.43)	4.58% (0.42)	4.31% (0.46)
RVM	4.73% (0.48)	4.07% (0.41)	4.05% (0.48)	4.66% (0.5)	4.38% (0.47)
NB (15)	2.37% (0.28)	2% (0.22)	1.58% (0.24)	2.55% (0.27)	2.13% (0.25)
NB (10)	3.17% (0.38)	2.12% (0.21)	2.14% (0.31)	2.37% (0.22)	2.45% (0.28)
NB (5)	1.88% (0.24)	1.4% (0.16)	1.07% (0.17)	1.08% (0.11)	1.36% (0.17)
SA (15)	-0.32% (-0.02)	0.73% (0.04)	-1.58% (-0.02)	-1.57% (-0.16)	-0.68% (-0.04)
SA (10)	0.19% (0.04)	1.42% (0.15)	0.18% (0)	-0.75% (-0.08)	0.26% (0.03)
SA (5)	1.03% (0.09)	0.59% (0.03)	-2.22% (-0.25)	-1.71% (-0.16)	-0.58% (-0.07)

Note: The table presents the annualized return and Sharpe Ratio for the top rules out of the data-snooping procedure survivors. Three fixed levels (5, 10, and 15) are studied SA, NB, DMA, DMS, DMA and BNN while the RVM is selecting the most relevant rules endogenously. The best rules are selected based on three measures of in-sample Sharpe ratio. All returns are after transaction costs. The values in bold correspond to the best performing combination for each exercise. The Giacomini and White (2006) test is applied to all combinations. The benchmark of the test is a simple random walk with no trend.