Computational Intelligence Applied to Financial Price Prediction: A State of the Art Review

Javier Sandoval Archila*

javier.sandoval@uexternado.edu.co

1. Financial Price Prediction: Theory Review

The currently accepted paradigm in Finance has dominated the study of human interactions in financial markets since Samuelson’s seminal work, “Proof that Properly Anticipated Prices Fluctuate Randomly” in 1965. The emerging theoretical framework was built over three concepts: agents’ rationality, market efficiency and neglected or zero predictability. This paradigm has been called Neoclassical Finance and it has focused on studying market interactions in equilibrium, a steady state, a state that does not offer any new stimulus to change. Therefore, relevant questions have been what agents’ actions, interactions and strategies should be in order to be with line to the steady state.

The established paradigm has had to overcome an indeterminacy presented with its lines of reasoning. In financial markets, (as well as any other economic system) individuals behave collectively to create an outcome. However, the same outcome also influences the behavior previously mentioned. Therefore, the financial interaction problem is ill-posed. As pointed out by Arthur (2006), Neoclassicals proposed the following questions in order to overcome this problem. What prices and quantities of financial assets offered and bought are consistent with equilibrium? Moreover, what strategies and actions will also help to perpetuate this moveless state?

These two questions have defined the first pillar of modern Finance studies; agent’s rationality and homogeneity. On one side, rationality and homogeneity helped financial economics to developed tractable and analytical models. However, Finance has paid a high cost. Models created under the Neoclassical domain are not necessarily supported by realistic assumptions. Therefore, conclusions are elegant but sometimes wrong. Rationality drives financial economics to the so-called expression, rational decisions or rational expectations. When participants of a financial market make decisions, they continuously do so in accordance with available information. Furthermore, strong rationality assumes that agents always use the information at hand in the most optimal way.

Due to rationality, agents’ decision making process is reduced to an optimization problem, hoping that individuals have the capacity to make necessary calculations in an acceptable time. Ill-posed problems have been tamed. From a domain of potential individual expectation models, the one that would validate the whole model if every individual uses it, has been chosen. The answer is not difficult to guess, with expectations following previous rationality definition. So, if every agent is implementing the same decision making process, Neoclassicals rapidly came to the aggregation assumption which states that every individual in the market can
be summarized by one agent. The representative agent was born bringing with it the homogeneous rational expectation assumption.

Several dissident works from the rational expectation assumption have offered extended evidence that individuals’ decisions do not necessarily have to follow the Neoclassical’s view. For example, Tesfatsion (2006) argued that agents’ expectations can differ even if they are conditioned to the same set of information simply by behavioral noise, uncertainty regarding what others would do in response to my actions. Moreover, this auto referential framework can be destabilizing and decision making could never converge to the Neoclassical’s assumption of perfect rationality.

The debate around rational expectations is not a new issue. Keynes (1936), years before Neoclassical economics became a standard paradigm, already argued that ”Investment based on genuine long-term expectations is so difficult as to be scarcely practicable. He, who attempts it must surely lead much more laborious days and run greater risks than he who tries to guess how the crowd will behave; and, given equal intelligence, he may make more disastrous mistakes”. In response to Keynes critics, Simon (1957) proposed the concept of bounded rationality. Individuals are limited in their knowledge and capacity to process information. Therefore, they may often use simple rules of thumb that do not necessarily inhabit a well-posed optimization domain.

A relevant extension of the previous bounded rationality concept was released by Conlisk (1980). Individual decisions not only can vary in a non-simple manner from perfect rationality but financial markets give evidence of long-run coexisting between costly rational strategies and cheap rules of thumb. It seems that investors adapt and must learn how to do this switching between decision models. This idea contradicts a famous statement in Finance, The Friedman Hypothesis, ‘non-rational agents in the neoclassical sense, will no survive evolutionary competition’.

If markets are not formed by strict rational agents, but they are full of adaptive, perpetually learning individuals that interact among others, what consequences can be deduced on the widely-spread concepts of market efficiency and neglected or zero predictability?

Following Lebaron (2006), market efficiency is the condition stating that asset prices reflect market information as a direct consequence of individual perfect rational expectations. In its strongest version, the market efficiency hypothesis states that no available information would be useful to create abnormal earnings, extra earnings created without increasing risk exposure. Thus, Strong Market Efficien-
cy states that price prediction techniques studying information at hand are full of flaws that invalidate themselves.

However, a new approach to investors’ decision making framework could suggest that predictability may be greater than what Neoclassicals had previously thought. If investors do not behave as perfect optimization machines but use techniques in accordance with a soft decision making process, there could be an emergence of extra predictability even when using historical information.

One of the most prominent alternatives to the classical framework represented in the Neoclassical’s market efficiency definition was presented by Andrew Lo, an economist from the Sloan School, MIT in his paper, The Adaptive Market Hypothesis: Market Efficiency from a Evolutionary Perspective. As stated by Lo, prices do not reflect all information available but rather all information available conditional to what has been called the environmental conditions, or in biological terms, the market ecology. The level of efficiency of a specific market can vary by time and context. It can also fluctuate from market to market and in different time frames of the same market. As an analogy, predictability level and therefore, profit opportunities, can be compared to resource availability in ecosystems. The scarcer the resources, the more fierce the competition between participants. Because of market complexity, financial market agents perpetually adapt to changing market conditions implementing new and better heuristics which approximate optimal solutions while regimes become stable. However, if regimes switch, individuals accept that old heuristics will not work in the new environment and new techniques have to be derived again.

Under this new market efficiency definition, predictability is a manner of timing. It is necessary to identify changing (possibly hidden) states that govern what kind of strategies may or may not work. Depending on the investment horizon, timing can be measured in months or even years or can be as tiny as the next transaction tick now ranging in milliseconds. For instance, a classical example of a regime switch, on a large scale time dimension, was triggered by the Russian default in 1998. Investors ran to safety and liquidity overwhelmed basic arbitrage opportunities between assets of similar credit risk profile. Regime changes can also be observed on small time scales. A typical case is the arrival of aggressive market orders to limit order book driven markets. In a fraction of a second, investors react to the new micro regime, implementing strategies from a different domain.

Following the previous reasoning, several empirical studies also confirmed considerable predictability levels. Different from the Neoclassical’s view, market prices are more predictable than once thought. For example, Lo and MacKinlay
(1996) showed that there exist evidence of price prediction on short horizons. This prediction can be hypothesized to be connected to the type of strategy used by agents extrapolating in future prices.

On this context, Frankel and Froot (1987b), Frankel and Froot (1987a), Frankel and Froot (1990), Frankel (1990), Allen and Taylor (1990) and Taylor and Allen (1992) showed that investors tended to use chartist strategies\(^1\) for prediction at short periods of time and fundamentalist for long ones. As found by Frankel and Froot (1987a) p. 264, “It may be that each respondent is thinking to himself or herself, I know that in the long run the exchange rate must return to the equilibrium level dictated by fundamentals. But in the short run I will ride the current trend a little longer. I only have to be careful to watch for the turning point and to get out of the market before everyone else does”. These findings are still valid today.

Extra predictability are also supported by institutional investors’ performance which specialize on exploiting market opportunities. The Hedge Fund Weighted Composite Index (\textit{hfrifwi}) summarizes the performance of the hedge fund industry, private, actively managed institutional investors that exploit sophisticated strategies. As shown in Figure 1, from 1992 to 2010, the index has increased 530%. In contrast, the \textit{s&;p500} has gained, including dividends, 310%. This extra return can not be explained by high risk exposure. The standard deviation of \textit{hfrifwi}’s annual returns has been 12.08% compared to 20.04% on the \textit{s&;p500} index. This conclusion is even supported if analyzing the maximum dropback, the maximum decrease over a year observed on both indexes, 12% and 20% respectively.

\textbf{2 State of the Art on Computacional Intelligence Applied to Financial Price Prediction}

Computational Intelligence studies machines’ ability to learn, adapt, make decisions, and display behaviors not explicitly programmed into their original capabilities. As mentioned by Roy and Rada (2008), much of the classic, knowledge-based work in artificial intelligence does not appear in the recent work of financial price prediction applications. Pure knowledge-based models applied to financial price prediction were rapidly abandoned and replaced by machine learning, data driven techniques. This trend has been observed during the 2000’s, with machine learning

\(^1\) Chartism states that historical prices and their transformation can be an informative source for predicting future price performance. On the other hand, Fundamental analysis looks at company’s or real assets’s performance to predict future financial asset prices.
techniques, relegating expert knowledge to a secondary role in price prediction modeling or simply, ignoring it.

After 2009, a revival of the importance of expert knowledge has increased the use of hybrid models. This trend has been highly motivated by a dimensionality reduction produced by expert knowledge inclusion on model’s design. It is worth noting that the term hybrid refers to a combination of a data driven and knowledge-based techniques.

The next subsections will present Machine Learning techniques applied to the financial asset price prediction domain. Then, recent proposed hybrid models will be presented including studies on Dynamic Bayesian Networks, a newly explored field in financial price prediction.

2.1 Artificial Neural Networks (ANNs)

One of the first models applied to price prediction in computational intelligence were Artificial Neural Networks (ANNs). ANNs are comprised of family of mathematical modeling methodologies inspired by the function of human neural system, specifically, neurons’ interrelation. First publications on application of ANNs to price prediction can be traced back to 1988 with the seminal work of White (1988) which focused on ANNs for predicting the daily value of IBM stock based on historical prices. White’s results were not as astonishing as he expected, yet he inaugurated a list of countless articles devoted to studying market predictability using non-traditional techniques.

Focusing on a Market index, Trippi and DeSieno (1992) explored ANNs ability to predict the daily values of the S&P500. This article gave evidence of how a combination of rule-based expert system techniques and a ANN outperformed a simple passive strategy. In a short period of time, ANN models covered other financial assets. Grudnitski and Osburn (1993) studied again ANN’s ability to forecast the S&P500 index and they also tested ANN’s predictions on Gold futures. The latter article, not only used historical prices as input data but also included, open interest patterns which were thought to incorporate the beliefs of a majority of the traders in the corresponding market.

ANNs were also used to predict values of fundamental variables such as interest rates. Nikolopoulos and Fellrath (1994) and Swanson and White (1995) studied the ANN’s forecast predictability on the term structure of interest rates. The first paper focused on interest rate trends detection while the second one studied the interest
rate forward spread over spot rates as a potential predictor on future interest rate behavior.

Before 1997, articles on ANN applications to financial forecasting showed primitive frameworks such as multilayer perceptron (MLP) and mostly used historical prices on US stock markets. After 1997, researchers began to explore new sets of input data and sophisticated structures.

ANNs were adapted to forecasting models on foreign exchange markets. Frances and Homelen (1998) developed a model based on ANNs to be able to predict four major exchange rates which included the Dutch Guilder. Over the sample data, the paper reported unfavorable in-sample fits or forecast performance. On the contrary, Hu et al. (1999) found evidence of superior forecasting performance of ANNs over Random Walk predictions on major foreign exchange markets. This kind of contradictory conclusions were reached every time simple ANN models, fed with historical prices, were constructed.

Researchers moved forward and place together ANNs and technical analysis indicators as in Tan and Yao (2000). This paper left simple historical price datasets and implemented a MLP-ANN framework directed to predict major foreign exchange rates including, the CHF/USD and JPY/USD, using typical inputs from chartist. Prediction power was compared to ARIMA’s performance. The authors found that on almost every market the ANN outperforms the ARIMA benchmark. Other articles which combined technical indicators and ANNs are Baba and Kozaki (1992) and Armano et al. (2005).

Input data was also extended to cover international stock markets others than the US market. For example, Fernandez-Rodriguez et al. (2000) implemented an ANN-type model in order to forecast the general index of the Madrid Stock Market. Authors found that the trading rule developed had a superior performance than a buy-and-hold strategy specially on “bear” and “stable” market episodes. Another study on international markets was that of Olson and Mossman (2003) which used accounting ratios to predict stock prices on the Canadian market.

Other proposed assemblies on ANN literature applied to price prediction were Radial Basis function Neural Networks (RBF) combined with dimensionality reduction on input data as in A. Lendasse et al. (2000) and, local linear wavelet neural network (LLWNN) together with an estimation of distribution algorithm (EDA) as in Chen (2005).
2.2 Kernel Machines

Following the definition of (Chalup and Mitschele (2008), page 655), kernels methods can be regarded as machine learning techniques which are ‘kernelised” versions of other fundamental machine learning methods. The latter include traditional methods for linear dimensionality reduction such as principal component analysis (PCA), methods for linear regression and linear classification.

In finance and specially in price prediction, research has concentrated on adapting support vector machines (SVMS) from image, text and sound analysis to the new problem domain. One of the main drawbacks observed on implementing SVM models in price prediction is that financial price series are highly noisy and non-stationary. This leads to changes in the relationship of input and output variables over time which make the use of static models unfeasible.

Before 2000, there is no evidence to our knowledge of applications of SVMS to the price prediction problem. One of the first contributors was Cao and Tay (2001). This paper emphasized the over-fitting problem of ANNS and proposed the implementation of a new machine learning technique, SVMS. Cao and Tay (2001) found that SVMS were better suited for the financial forecast problem due to SVM’s adoption of the empirical risk minimization principle which seeks to minimize an upper bound of the generalization error rather than minimize the training error. Therefore, SVMS could perform better on price prediction than ANNS. The paper reported a better prediction performance of the SVM model than a Multilayer perceptron trained using back-propagation and historical and chartist indicators of the S&P 500 as input data.

Extending their previous results, Tay and Cao (2001) presented the same year a model’s architecture which combined a Support Vector Machine (SVM) and a self-organized feature map (SOM). First, the SOM was used to partition the input data space into disjoint regions. Then, multiple SVMs were trained using partitioned regions. The models were tested on prices from five future contracts. The authors reported a high prediction power and a better model performance when compared to results obtained from a single SVM. These two works determined the beginning of multiple publications devoted to SVMS applied to price prediction.

To overcome the static nature of SVM, Tay and Cao (2002) and Cao and Tay (2003) suggested the implementation of an Adaptative SVM applied on regression. The former extension of SVM was inspired by non-stationary financial prices in which most recent data provides the most relevant information. Researchers gave more weight to recent observations and less weight to distant ones letting the va-
values of the regularization and tube’s width change in the support vector regression framework. Their results were also validated using a standard ANN’s architecture, a back-propagation multilayer perceptron and a regularized RBF neural network using historical prices and chartist variables as the data input for values of the S&P500.

Although SVM algorithms implemented before 1999 were generally slower than artificial neural networks with similar generalization performance (Haykin (1999), p. 345), significant improvements were achieved in the following years. Therefore, ANNs became the typical benchmark to beat.

Huang et al. (2005) also implemented the SVM-ANN comparative study over the Nikkei 225 series examining the models’ ability to forecast the series direction over a week time interval. Their results presented a SVM model that beat, in terms of the hit ratio, a back-propagation neural network as well as a random walk and a linear discriminant analysis model. Other papers that showed superior performance of SVM models over ANNs were Ince (2000) and Sansom et al. (2002), the latter over an electricity market.

Although SVM models were highly outperforming ANNs, researchers moved to combined SVM with other knowledge extraction techniques that improve preprocessing of input data. A good example of these implementations was Ince and Trafalis (2006) which proposed joining a non-parametric technique, i.e. SVMs, together with a parametric model such as integrated moving average (ARIMA), vector autoregressive (VAR) and co-integration in order to implement a better input variable selection. Its finding showed that filtering input variables using a parametric method generated a better forecasting performance of a SVM model.

Several studies showed SVMs as a superior forecasting model compared to ANNs in the price prediction problem domain. For example, Kim (2003) examined SVM’s performance compared to predictions generated by a back-propagation neural network and case-based reasoning (CBR). Results showed that the SVM model outperformed ANN benchmark when tested on the daily Korea composite stock price index (KOSPI).

Another example of SVM enhancement is developed in Lee (2009). This paper combines a SVM framework with a filtering and wrapping method called F-score and Supported Sequential Forward Search (SSFS) to execute a better information extracting procedure over the set of input data. The authors reported a better performance of this hybrid structure than the one obtained by a standard SVM model when predicting the daily value of the NASDAQ index using a set of technical analysis variables as raw input variables.
Filtering and wrapping was only one of a possible data pre-processing option. A second alternative was dimensionality reduction using well known algorithms such as independent component analysis (ICA) as in Lu et al. (2009) which focused on reducing noise in input data before calibrating the SVM model. Specifically, the authors proposed a two stage modeling approach that firstly implemented an ICA algorithm in order to generate independent components from input variables. ICA-SVM was tested on Nikkei 225 and TAIEX indexes obtaining a better performance than a regular SVM and a Random Walk model.

So far, all presented SVM models have focused on historical prices, technical analysis and other macroeconomic variables. To the best of our knowledge, only one paper has experimented with the limit order book as input set for a SVM models directed to predict the market’s directional changes. Specifically, Fletcher et al. (2011) proposed the implementation of multiple kernel learning methods (Simple-MKL and LPBOOSTMKL) to train a multiclass SVM model in order to predict EUR/USD changes based on cumulative volumes shown in the limit order book. The authors found promising results specially when compared to a single SVM’s performance.

### 2.3 Evolutionary Methods

As defined by (Brabazon and O’Neill (2006), page 37), evolutionary methods comprise a series of algorithm inspired by the evolutionary metaphor. Researchers have specifically focused on Genetic Algorithm (GA), Genetic Programming (GP) and Grammatical Evolution (GE) regarding two kinds of strategies. Evolutionary models to directly evolve trading systems aimed to exploit price prediction capabilities and, implementing evolutionary methods to optimize non traditional models such as ANNS. Therefore, analyzing research on genetic methods applied to price prediction will result in an understanding of both approaches.

One of the first works to propose an evolutionary technique to tackle the price prediction problem was Deboeck (1994). In this work, the author proposed an ANN model calibrated using Genetic algorithms (GA) and the maximum profitability/maximum drawdown ratio as a performance variable to be optimized in a training data set. The paper by Versace Versace et al. (2004) used GA to train ANNS.

Another revolutionary technique applied to the price forecast problem domain has been Genetic Programming (GP). Genetic programming as popularized by Koza (1992), is an evolutionary method which directly evolves potential solutions commonly represented by structures such Lisp S-expressions or computer programs.
The literature shows that evidence of superior performance of GP models on the price prediction domain is not clear. One of the first works focused on Genetic programming was Chen and Yeh (1996) which studied the Efficient Market Hypothesis (EMH, as defined in section 1). The authors evolved some math expressions representing trading strategies in order to predict returns on the S&P500 and the TAIEX, a Taiwan stock exchange index. The paper concluded that there was no evidence of superior performance of the GP-based model vs a buy-and-hold strategy. Other works that also found no evidence of a better performance of GP based models are Allen and Karjalainen (1999) Evolving Technical Rules on S&P500 Values From 1928 to 1995 and Potvin et al. (2004), which studied Canadian individual stocks using historical prices and transaction volumes.

In contrast, Li and Tsang (1999) showed a different perspective. The latter work, studied the performance of GP-based models on foreign exchange markets. The authors reported superior performance of GP models compared to a buy-and-hold strategy. They also gave evidence that this superior return was not due to a greater exposure to risk. This conclusion was also supported by Becker and Seshadri (2003) implementing a fitness function that considers consistency of performance and takes into account the transaction cost effect. Similar conclusions have been reached recently by Lohpetch and Corne (2010).

A different form of evolutionary technique is grammatical evolution (GE). Grammatical evolution as defined by Brabazon and O’Neill (2006), page 73, is a grammar-based form of Genetic Programming which extends the biological analogy by employing a distinction between the genotype and phenotype similar to that which exists in nature. Studies focused on GE applied to the price prediction domain were presented for the first time in O’Neill et al. (2001a). This paper proposed the uncovery of useful technical trading techniques using data from the FTSE 100 index for the period 26/4/1984 to 4/12/1997. This first paper revealed much potential in GE based models. GE models were also tested in foreign exchange series. Brabazon and O’Neill (2002) presented a GE model tested over the GBP/USD daily series from 1993 to 1997. Technical indicator were evolved using a grammar defined model. Experimental results showed that the GE model was able to beat buy-and-hold strategies even on out of sample data using a fitness function which emphasized excess returns and the maximum drawdown of each trading strategy. Another study specialized on GE focused on exchange rate prediction was Brabazon and O'Neill (2004).
2.4 Dynamic Bayesian Networks (DBNs)

Dynamic Bayesian networks represent graphically a stochastic process using Bayesian networks which include directed edges pointing in the direction of time (Murphy (2002)). The graph contains a qualitative part which represent the a priori expert knowledge expressed in a set of junctions which facilitate conditional probability calculation. Following Bengtsson (1999), the entire model can be thought of as “a compact and convenient way of representing a joint probability distribution over a finite set of variables”.

One of the first special cases of DBNs implemented to tackle the price prediction problem were Hidden Markov Models (HMMs). HMMs assumed that the underlying modeled system exists in one of a finite number of states. The latter states are hidden and are responsible for producing a sequence of observable variables. Hassan (2005) is one of the first works which extracted HMMs from speech and image recognition problems and place it in the stock price prediction domain. In this paper, a 4-hidden state system is assumed to represent the performance of an airline stock price. Special emphasis is made on today’s open, high, low and closing prices and the work aims to predict the next day’s closing price. Performance is evaluated using MAPE and $R^2$ between predicted and realized closing values. Results are informally compared to those obtained using a ANNS which followed the same architecture. Hassan et al. (2007) extended the previous work and combined GA and ANN with a HMM. Authors reported superior performance of the combined model compared to the one obtained on previous HMM model calibrated with standard techniques.

As a natural extension to HMM, Hierarchical Hidden Markov (HHMM) models were also implemented to represent financial markets and solve price prediction problems. For example, Jangmin et al. (2004) presented a 5 state model that describes market trends; strong bear, weak bear, random walk, weak bull and strong bull market phases. Each state has different second level abstract states which at the same time, call its own third level variables which are responsible for output emission. Calibration is based on k-means clustering and a semi-supervised training technique. Dataset included daily stock prices of the 20 most active companies of the Korean stock market from 1998 to 2003. Authors reported that the HHMM outperformed on average a simple buy and hold strategy and a trading strategy following TRIX, a commonly used technical analysis indicator.

Following recent works by Tayal (2009) and Wisebourt (2011), the potential of HHMM on forecasting stock price changes were investigated using a preprocessing
technique called zigzag aggregation. A two state model which captures runs and reversal is coupled with a second hidden variable layer which defines the emitted zig-zag interval. The latter is a production state and therefore, produces observable stock returns. The first remarkable detail on these two works is that both implemented asynonymously time models. Therefore, homogeneous data in time is not a requirement. Second, both models are able to recognize regime change from uptrend to downtrend periods of time. The author concluded that the model’s performance is considerably affected by market liquidity. Thus, the less liquid an asset, the less likely a knowledge extraction technique outperform basic strategies. The input variable set went from historical prices and volume to elements of the order book.

Applications of DBNs to price prediction is a recently explored field of study. Therefore, general implementations are still to come. For example, learning structure from data could be considered or at least, leaving the basic structure to evolve as the financial environment changes. Chain graphs, which incorporate the notion of correlation, can strengthen the directed arc representation of DBNs. Furthermore, multi-dynamic Bayesian Networks can include features from different time scales enriching the information extraction process.

2.5 Hybrid Models

Although Machine Learning, data driven inference systems, offer better solutions for unexpected inputs, hybrid systems are preferred for more complex tasks. See Browne and Sun (2001). Standalone, data driven techniques are exposed to several difficulties that may be overcome when expert knowledge is taken into account. As presented by Webb et al. (1996), a machine learning approach is possible when most of the knowledge acquisition task has already been achieved. Machine learning models need a problem domain description and collection of example cases in order to be able to inductively generalize hypothesis. In this context, expert prior knowledge is crucial for every stage of the models’ development. During, pre-learning, expert knowledge can provide valuable inputs to the process. An expert can also audit rules generalized by the machine learning model. She can also strengthen possibly shallow knowledge acquired by machine learning models. Moreover, expert knowledge and judgement can help to discern between equally weighted choices from the point of view of the machine learning model.

In finance and especially in financial price prediction, hybrid models that include expert knowledge are almost in every case fuzzified versions of data driven models or data driven models with humans interpretable rules.
One of the first authors to implement a fuzzy-ANN joint was Nishina and Hagiwara (1997). The latter proposed a Fuzzy Inference Neural Network (FINN) whose goal was to extract fuzzy if-then rules from numerical data. The FINN was a two-layer network which was trained using Kohonen’s and least mean square learning algorithms. The authors tested FINN’s ability to predict daily prices of two stocks versus the forecast obtained from a regular Network trained by a back-propagation (BP) technique. They found that FINN performed much better on testing sample though the BP network had less hidden neurons and therefore, a better ability to generalize.

Another study which set out to compare relative performance between plain ANNS and Neuro Fuzzy NNS was Bekiros and Georgoutsos (2007). This paper presented a fuzzified Neural Network orientated to predict NASDAQ and NIKKEI 225 returns. It found, as in Fernandez-Rodriguez et al. (2000) that ANN models are more suitable for forecasting tasks in bearish and non-trend markets. Additionally, it also reported a consistent superior performance of Neuro Fuzzy Network frameworks over a traditional recurrent neural network. The study covered data from 1998 to 2002. Other research using Neuro Fuzzy Networks are Kuo et al. (2001) and Hadavandi et al. (2010). The latter also integrate, Fuzzy Neural Networks, Genetic Algorithms and Self Organized Maps (SOM).

Researchers also used GA to optimize fuzzified ANN as in the case of Doeksen et al. (2005). The model was used to predict daily movements of certain stocks such as Microsoft and Intel. Among other conclusions the author found that, based on input depending on close, open, high and low stock daily values, it was possible to design a profitable trading strategy which means net excess return.

There are plenty more cases of machine learning models trained using GA. Another relevant work is the one presented by Lam (2001) working on data extracted from the Hong Kong Stock Exchange. This work used GA to optimize a set of fuzzy trading rules. The author supported his work on the assumption that some market indicators embedded valuable information for generating buy and sell signals. However, those indicators can only predict price movements up to a certain extent. Therefore, they should be treated in a fuzzy sense.

Recent efforts creating hybrid systems applied to financial price prediction, involve combinations of Fuzzy logic and a data driven technique or the combination of two or more data driven techniques.
3 Conclusions

Computational Intelligence applied to the price prediction problem, covers a wide variety of models and input data domains.

As shown, the problem has migrated from a daily to a high frequency basis. It means that information has significantly increased. However, market data sampling in intraday periods has shown to be time heterogenous because observations are not evenly spaced in time. High frequency data also has exposed researchers to, in the classical sense, market microstructure noise. One of the best known sources of microstructure noise is the bid-ask bounce.

Transactions can be originated from both side of the order book. Therefore, high frequency series shows asset price variations that can not be executed in real life trading.

While time scale has decreased, studied asset price features have also changed. On first works, open, high, low and closing prices, including lag values were commonly included. Moreover, technical analysis variables were almost always present. If technical analysis variables are considered price transformation, it can be stated that historical transaction prices have been the main source of information on price prediction models. As information has become available, researchers have included trading volume and order book variables. Because of information limitations, limit order features mainly include best bid and offer and fixed interval liquidity levels. Though the order book is a rich information space, no other element has been extracted from it.

Research on price prediction has concentrated on developed markets. Available studies used financial market data from the USA, Europe, and Southeast Asia including Japan. This geographical restriction is due to information availability and stages of development of local financial markets. It is relevant to mention that financial markets in developed markets are characterized by the presence of automata trading systems, computerized systems making decisions in terms of buying or selling financial assets.

Techniques are highly concentrated on data driven models relegating prior expert’s knowledge to a second role. This second role especially focused on what kind of input data is used and how it is pre-processed. One of the main reasons for initial expert systems failure on predicting financial prices is the highly uncertain and changing financial environment.

There exists a novel approach to model nondeterministic dynamic systems when expert knowledge is present. This approach has been recently tested in other
domains and their results are quite enlightening. It consists of Dynamic Bayesian Models designed using expert’s knowledge.

Finally, almost every listed study gave evidence of profitable trading strategies developed based on predictions from previous models.

References


cce*, (5):357-364.


Figure 1. S&P500 vs HFRIFWI performance

Table 1: Summary of articles implementing ANNS applied to price prediction.

<table>
<thead>
<tr>
<th>Article</th>
<th>Input Data</th>
<th>Calibration</th>
<th>Network type and transfer functions</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grudnitski and Osburn (1993)</td>
<td>Technical analysis factor calculated over the S&amp;P500</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Article</td>
<td>Input Data</td>
<td>Calibration</td>
<td>Network type and transfer functions</td>
<td>Conclusions</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Nikolopoulos and Fellrath (1994)</td>
<td>Average rate of taxable money market funds of the 25 weeks preceding the current week</td>
<td>Genetic Algorithms</td>
<td>Feed-forward network with 33-3-21 layers.</td>
<td>Hitting ratio of 0.71 when predicting interest rate trend.</td>
</tr>
<tr>
<td>Tan and Yao (2000)</td>
<td>Major exchange rates such as USD/JPY, GBP/USD, USD/DSM and technical indica-tors.</td>
<td>N/A</td>
<td>Three layer feed forward neural network</td>
<td>Superior performance when compared to an ARIMA model in terms of NMSE.</td>
</tr>
<tr>
<td>Kuo et al. (2001)</td>
<td>Historical prices from the Taiwan Stock Market.</td>
<td>Genetic algorithms</td>
<td>Neuro-Fuzzy</td>
<td>Neuro-Fuzzy networks outperformed BP networks in terms of MSE.</td>
</tr>
<tr>
<td>Olson and Mossman (2003)</td>
<td>Financial ratios from stocks in the Toronto stock exchange</td>
<td>Delta Rule learning algorithm</td>
<td>Backpropagation network</td>
<td>The ANN forecast outperform results obtained from a OLS model using the same input variables.</td>
</tr>
<tr>
<td>Armano et al. (2005)</td>
<td>Error backpropagation</td>
<td>Hybrid system that integrates extended classifier system with Genetic Algorithms and ANNs</td>
<td>N/A</td>
<td>The neural network is able to beat a buy-and-hold strategy.</td>
</tr>
<tr>
<td>Hadavandi et al. (2010)</td>
<td>Historical prices of IT and Airline stock prices.</td>
<td>Genetic algorithms</td>
<td>Neuro-Fuzzy and self-organizing maps</td>
<td>When using the MAPE, the Neuro-Fuzzy model outperformed regular ANN models.</td>
</tr>
</tbody>
</table>
Table 2: Summary of articles implementing Kernel Machines applied to price prediction. 30

<table>
<thead>
<tr>
<th>Input Data</th>
<th>Calibration</th>
<th>SVM type</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ince (2000) IBM, Yahoo and American Online historical prices.</td>
<td>Direct search of parameters</td>
<td>SVM with a gaussian kernel.</td>
<td>N/A</td>
</tr>
<tr>
<td>Tay and Cao (2001) Technical indicators calculated over 5 future contracts in the US market.</td>
<td>N/A</td>
<td>Multiple SVMs with self-organizing maps</td>
<td>The improved SVM model outperformed a single SVM model in terms of mse, mae and hitting ratio.</td>
</tr>
<tr>
<td>Tay and Cao (2002) Technical indicators calculated over 3 future contracts in the US market.</td>
<td>Sequential minimal optimization (SMO) algorithm</td>
<td>SVM with an adaptive regularization parameter applied on regression</td>
<td>The SVM with adaptive regularization parameter outperformed a regular SVM in terms of NMSE.</td>
</tr>
<tr>
<td>Sansom et al. (2002) Historical prices and demand data of the Australian National Electricity Market (NEM), New South Wales regional data over the year 2002</td>
<td>Stefan Ruping routine and a fixed parameter C.</td>
<td>single SVM.</td>
<td>N/A.</td>
</tr>
<tr>
<td>Cao and Tay (2003) Technical indicators calculated over 3 future contracts in the US market.</td>
<td>Sequential minimal optimization (SMO) algorithm</td>
<td>SVM with an adaptive regularization and tube size parameters applied on regression.</td>
<td>The SVM with adaptive regularization and tube size parameters outperformed a regular SVM in terms of NMSE.</td>
</tr>
<tr>
<td>Kim (2003) Technical analysis factor calculated over the KOSPI, the Korea composite stock price index.</td>
<td>Direct search of parameters C and E.</td>
<td>SVM with a gaussian kernel</td>
<td>SVM predictions outperformed in terms of the hitting ratio a back-propagation network and a case-based reasoning model on predicting the index directional change.</td>
</tr>
<tr>
<td>Input Data</td>
<td>Calibration</td>
<td>SVM type</td>
<td>Conclusions</td>
</tr>
<tr>
<td>------------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Historical information of the s&amp;p500 index and the USD/JPY exchange rate.</td>
<td>N/A</td>
<td>SVM with a Gaussian Kernel</td>
<td>SVM outperformed A linear discriminant analysis, a quadratic discriminant analysis and an Elman backpropagation neural network when predicting Nikkei 225’s next week directional change.</td>
</tr>
<tr>
<td>Daily values of exchange rates for EUR/USD, GBP/USD, JPY/USD and AUD/USD were used from January 1, 2000 to May 26, 2004</td>
<td>N/A</td>
<td>SVM with filtered inputs using ARIMA and VAR models.</td>
<td>The SVM model with filtered inputs using ARIMA and VAR outperformed a backpropagation neural network trained over the same dataset. Moreover, the hybrid SVM model outperformed also ARIMA and VAR models when using in isolation.</td>
</tr>
<tr>
<td>Closing historical prices of 20 futures contracts and 9 spot indexes from 2001 to 2007.</td>
<td>Grid search approach</td>
<td>F-score and supported sequential forward search (FSSF) SVM.</td>
<td>The proposed model is compared to a Backpropagation network. It is found that FSSF SVM outperformed BPN in terms of hitting ratio.</td>
</tr>
<tr>
<td>Historical prices of the Nikkei 225 and the TAIX.</td>
<td>Grid search approach</td>
<td>ICA-SVR mode</td>
<td>The ICA-SVM model’s output outperformed a regular SVM and a random walk prediction when compared in terms of RMSE and directional change.</td>
</tr>
</tbody>
</table>
Table 3: Summary of articles implementing Evolutionary techniques applied to price prediction

<table>
<thead>
<tr>
<th></th>
<th>Input Data</th>
<th>Methodology</th>
<th>Model</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li and Tsang (1999)</td>
<td>Dow Jones Industrial Average (DIIA) index data from 7 April 1969 to 11 October 1980</td>
<td>Decision trees generation representing technical analysis rule combinations.</td>
<td>A genetic programming discovering a technical trading rule.</td>
<td>N/A</td>
</tr>
<tr>
<td>O’Neill et al. (2001a)</td>
<td>UK FTSE 100 stock index historical values from 1984 to 1997</td>
<td>The evolution of a technical analysis strategy.</td>
<td>Grammatical Evolution</td>
<td>Trading strategy evolved from technical analysis model evolution economically outperformed buy-and-hold models.</td>
</tr>
<tr>
<td>Versace et al. (2004)</td>
<td>Open-High-Low-Close-Volume of major US stock index, 2 major exchange rates, 5 bond and one commodity price series.</td>
<td>Combination of Networks predicting directional change based on a voting schema</td>
<td>GA and mixtures of networks</td>
<td>The Mixture model directional forecast can be significantly differentiated from chance using a Ki2 test.</td>
</tr>
<tr>
<td></td>
<td>Input Data</td>
<td>Methodology</td>
<td>Model</td>
<td>Conclusions</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>--------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Doeksen et al. (2005)</td>
<td>Intel and Microsoft historical open-high-close stock prices from 1997 through 2003</td>
<td>Neural networks which used filtered inputs in terms of importance.</td>
<td>GA and neural networks.</td>
<td>Proposed model was able to outperform a buy-and-hold strategy in terms of economic performance excluding transaction costs.</td>
</tr>
</tbody>
</table>