

GLOBAL CONFERENCE ON  
BUSINESS AND FINANCE  
PROCEEDINGS

VOLUME 6, NUMBER 1  
2011

ISSN 1931-0285 CD

ISSN 1941-9589 ONLINE

*The Institute for Business and Finance Research*

*Las Vegas, Nevada  
January 2-5, 2011*

Global  
IBFR  
bal

*TheIBFR.ORG*

<b>GLOBAL PRODUCTION AND INTERNATIONAL COORDINATION COSTS .....</b>	<b>498</b>
WANIDA NGIENTHI, ASSUMPTION UNIVERSITY	
<b>CHSF AND ENGLISH LANGUAGE: A COMPETITIVE ADVANTAGE STRATEGY IN THE HOSPITALITY/TOURISM INDUSTRY .....</b>	<b>507</b>
CRISTINA G. GALLATO, UNIVERSITI UTARA MALAYSIA KRISTY AIREEN G. GALLATO-REAMILLO, SAINT MARY'S UNIVERSITY	
<b>BEHAVIORAL MODELING OF TIME, TREND AND TRADING OF FIIS IN INDIAN EQUITY MARKET.....</b>	<b>508</b>
SANTOSH KUMAR, AMITY UNIVERSITY RAJU G, GALGOTIAS COLLEGE OF ENGINEERING AND TECHNOLOGY JAPNIT WALIA, AMITY UNIVERSITY TAVISHI, AMITY UNIVERSITY	
<b>PROFILING AND THE STUDENT PROJECT PEER APPRAISAL PROCESS .....</b>	<b>517</b>
RODLEY C. PINEDA, TENNESSEE TECHNOLOGICAL UNIVERSITY BONITA B. BARGER, TENNESSEE TECHNOLOGICAL UNIVERSITY LINDA D. LERNER, TENNESSEE TECHNOLOGICAL UNIVERSITY	
<b>ENHANCING STUDETNT SATISFACTION WITH HIGHER EDUCATION SERVICES:.....</b>	<b>523</b>
<b>GOING BEYOND SERVICE QUALITY PERSPECTIVES .....</b>	<b>523</b>
JUNGKI LEE, KOREA UNIVERSITY AT SEJONG	
<b>FINANCIAL MARKET SUPERVISION AND CAPITAL REGULATION IN VIETNAM .....</b>	<b>528</b>
LE THI PHOUNG HOA, BRANDEIS UNIVERSITY	
<b>RE-THINKING FINANCIAL NEURAL NETWORK STUDIES: SEVEN CARDINAL CONFOUNDS .....</b>	<b>537</b>
CHARLES WONG, BOSTON UNIVERSITY MASSIMILIANO VERSACE, BOSTON UNIVERSITY	
<b>THE INFLUENCE OF ULTIMATE OWNERSHIP TO EARNINGS MANAGEMENT: EVIDENCE FROM INDONESIA .....</b>	<b>549</b>
I PUTU SUGIARTHA SANJAYA, ATMA JAYA YOGYAKARTA UNIVERSITY	
<b>ECHO ARTMAP: CONTEXT SENSITIVITY WITH NEURAL NETWORKS IN FINANCIAL DECISION-MAKING .....</b>	<b>559</b>
CHARLES WONG, BOSTON UNIVERSITY MASSIMILIANO VERSACE, BOSTON UNIVERSITY	
<b>FASHION FOR SUSTAINABILITY: A MULTI-PRONG APPROACH TO MARKETING RECYCLED TEXTILES .....</b>	<b>569</b>
KENNY K. CHAN, CALIFORNIA STATE UNIVERSITY, CHICO ZACK TREANOR, CALIFORNIA STATE UNIVERSITY, CHICO	
<b>COMPARISON OF THE EVA EFFECT TO OCF AND EPS ON MV OF LISTED COMPANIES ON THE TSE .....</b>	<b>570</b>
ZAHRA LASHGARI, ISLAMIC AZAD UNIVERSITY CENTRAL TEHRAN BRANCH AKBAR JAVADIAN KOOTANAE, ISLAMIC AZAD UNIVERSITY FIRUZKOOH BRANCH	
<b>TRANSFORMATIONAL LEADERSHIP IN LAS VEGAS CASINO-GAMING MANAGERS.....</b>	<b>577</b>
GREGORY W. GOUSSAK, UNIVERSITY OF SOUTHERN NEVADA JON K. WEBBER, UNIVERSITY OF PHOENIX	
<b>APPLYING JOHN RAWLS' DIFFERENCE PRINCIPLE: A CASE FOR A WEAK INTELLECTUAL PROPERTY REGIME IN GUATEMALA.....</b>	<b>589</b>
BRIAN J. QUARLES, UNIVERSITY OF SOUTHERN NEVADA	

# RE-THINKING FINANCIAL NEURAL NETWORK STUDIES: SEVEN CARDINAL CONFOUNDS

Charles Wong, Boston University  
Massimiliano Versace, Boston University

## ABSTRACT

*Neural networks are among the best adaptive analysis tools for decision-making. As the field approaches maturity, there should be a convergence of applied theoretical and empirical findings. In financial analysis studies, neural networks superficially appear to deliver. Closer examination however finds a theoretical intra-disciplinary mismatch between neural network models and an empirical inter-disciplinary mismatch between neural network and financial approaches and analyses. Based on a sample of 25 financial neural network studies from the past decade, this paper explores the different characteristics of the major network categories used and the cardinal confounds in applying them to finance. In so doing, this paper provides an updated snapshot on the state of the art in applied neural networks for financial decision-making and specific suggestions on maximizing the utility of neural networks.*

## INTRODUCTION: NEURAL NETWORKS AND DECISION-MAKING

Financial time series are collections of financial indicators over time. A prime example is the daily closing price of a given stock over time. At any time during the trading day, a buyer and a seller may agree upon a transaction price. The buyer ideally agrees only if the price reflects her best knowledge regarding the fair market value – for example, the bid may reflect an estimate of the net present value of all expected future cash flows (Chartered Financial Analyst Institute, 2009). Vice versa, the seller should only agree when the price accurately reflects his best estimate of fair market value. Barring cases where the traders are forced to sell with regret to meet extrinsic obligations, public trading prices usually reflect the mutually agreed assessment of the stock by human decision-makers (Kaizoji & Sornette, 2008). By extension, a series of publicly available prices reflects a series of human decisions over time. Unfolding this decision making process by analyzing the time series can lead to more efficient financial markets, a deeper understanding of biological intelligence, and the development of technological tools that seek to emulate human decision-making. Neural networks are among the best adaptive analysis tools due to their biologically inspired learning rules; they better accommodate non-linear, non-random, and non-stationary financial time series than alternative techniques (Lo, 2001; Lo & Repin, 2002; Lo, 2007). Much research in the past decade applies a variety of neural network tools to financial time series with typically strong and compelling empirical results (e.g. Duhigg, 2006; Hardy, 2008). See table 1.

Table 1: Survey Information

Category	Instances	Typical Result
Slow learning	10	Improved vs. market
Fast learning	6	Much improved vs. market
Maximum margin	3	Improved vs. market
Temporal sensitive	6	Improved vs. market
Mixture of networks	4	Much improved vs. market

*Table 1: A survey of 25 studies that applied neural networks to financial stock time series analysis over the past decade. Categories divide into five distinct models included with frequency and typical results. The total exceeds 25 because individual studies may include multiple networks.*

Two trends appear to be dominant from Table 1. First, neural networks appear to outperform the market, which is typically defined as a random walk or buying-and-holding of an index. Second, the results appear robust regardless of network category. The bias towards slow learning models in Table 1 probably reflects their earlier availability (Rumelhart, Hinton, & Williams, 1986).

This paper has a dual purpose: (1) to classify the main neural network models used in financial studies based on their learning mechanism; and (2) to identify and address the cardinal confounds that may reduce confidence in the typical study's findings that neural networks consistently and robustly outperform the market regardless of network model. Section II explores the five distinct network models and their applicability to financial time series. Section III discusses the seven confounds that can serve as guideposts for applying neural networks to financial decision-making studies. Section IV concludes with the lessons learned.

## WHAT ARE NEURAL NETWORKS?

A basic model of decision-making gathers evidence, weighs it, and aggregates it in favor of one decision or another (Siegler & Alibali, 2005). Figure 1 (left) shows a schematic of this decision making process. It conforms in all manners of operation to typical models of regression analysis (Bishop, 2006).

Figure 1: Schematic

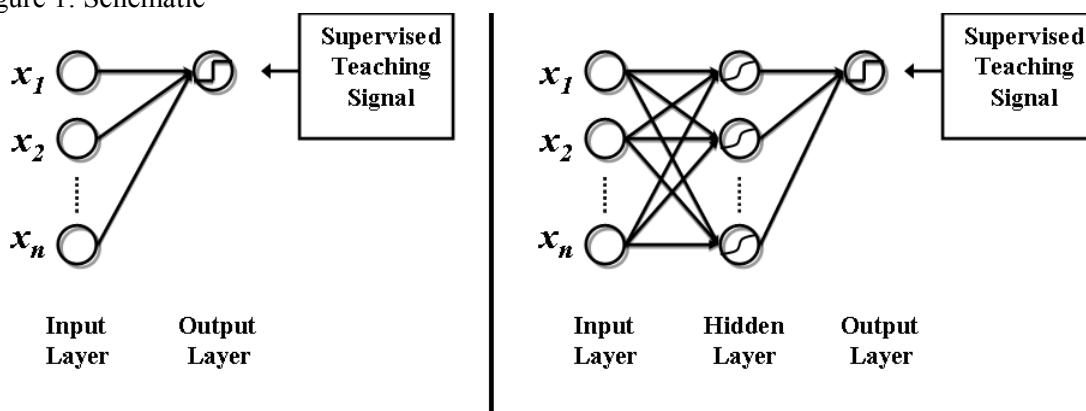


Figure 1: (left) A classic model of decision-making, which also conforms to a general model of regression. The evidence, or inputs ( $x_1 \dots x_n$ ) in the input layer, are weighted and gathered to form an output, which applies a threshold to decide among two choices. (right) A slow learning model of neural network decision-making. Inputs in the input layer are weighted and gathered for each hidden node. Hidden nodes process these sums via non-linear transfer functions, represented by the sigmoid symbol. The output layer weighs and gathers these hidden values to form the final decision, thresholded among two choices. The supervised training signal provides the expected output decision when training the models.

Over the past decade, an increasing number of studies have used various types of artificial neural networks for analyzing financial time series (e.g. Zhu et al., 2008; Chen et al., 2006; Leung and Tsoi, 2005; Thawornwong and Enke, 2004; Versace et al., 2004). The term artificial neural network (hereafter, termed neural network) refers to a class of flexible, non-linear, and adaptive models inspired by biological neural systems (Bishop, 2006; Hornik et al., 1989; Rumelhart et al., 1986; Rosenblatt, 1958). The key addition in neural network models is the brain-like hierarchical structure. Inputs typically aggregate in hidden layer nodes that perform further levels of processing before finally favoring a particular decision in the output layer. See Figure 1 (right). This additional processing adaptively explores hidden relationships among the evidence, but adds immense complexity. Exploring this complexity led to a divergence of multiple forms of neural networks divisible by learning rules and structures. This paper selects 25 neural network financial studies spanning 1997-2009 and categorizes the models used based on network learning rule. The five models are: slow learning models, fast learning models, maximum margin models, temporal sensitive models, and mixture of network models.

### A Slow Learning Models

Slow learning models are among the earliest forms of modern neural networks, relying on the work of McCulloch and Pitts (1943), Rosenblatt (1958), Rumelhart et al. (1986) and others. Processing proceeds in three steps. See Figure 1 (right). First, the input layer nodes weigh the input, with each node representing one input feature. The connections from each input node to each hidden node have individual weights. Second, the hidden layer nodes typically apply a non-linear, sigmoid shaped transfer function to the aggregate of these weighted inputs. The selection of non-linear, sigmoid transfer functions are inspired by physiological studies of nerve functions (e.g. McCulloch and Pitts, 1943) that allow enhanced exploration of complex relationships in the data. Finally, the output layer weighs and aggregates these hidden layer node values and applies a threshold to decide on one of the possible output options.

Slow learning models are highly distributed. All input and hidden nodes are involved for each output decision. If the output of the network does not match the correct output as dictated by the supervised teaching signal (e.g. decisions made by a human expert on the same data), the slow learning model typically adjusts all weights incrementally until the output values sufficiently match. Since all connection weights are inter-related, the adjustments are necessarily small and incremental, leading to slow and measured convergence on the correct output. Hence, these models are termed slow learning. Known complications of slow learning models include their tendency to overfit the data, failure to converge to acceptable rates of correct outputs, and their poor adaptability to new data. Overfitting refers to the network developing overly specific associations to in-sample data that generalize poorly on novel out-of-sample data sets. The failure to converge is caused by the inter-relatedness of all node connections. Adjusting one connection favorably can cause another connection to become unfavorable and vice versa until the network oscillates between optimal decisions. This also renders the trained network unable to incorporate new data without a complete reset and refresh of all prior learning. The majority of recent financial studies with neural networks use networks derived from this fundamental architecture (e.g. Kumar and Bhattacharya, 2006; Kaastra and Boyd, 1996; Zhu et al., 2008; Kirkos et al., 2007; Brockett et al., 2006). Recent research exploring different slow learning non-linear heuristics with kernel smoothing (Zhang, Jiang, and Li, 2005), genetic algorithms (Kim, 2006), and Lagrange multipliers (Medeiros, Terasvirta, and Rech, 2006) mitigate – but do not eliminate – the complications.

### B Fast Learning Models

Fast learning models address the complications of non-convergence and poor adaptability to new data of slow learning models. Primary examples include probabilistic and radial basis function networks (Moody & Darken, 1989). The key difference from slow learning models is the use of a storage layer instead of a hidden layer. Figure 2 (left) shows the typical form of a fast learning model.

Processing proceeds in three steps. First, the input layer nodes form the input. Each node represents one input feature; all nodes in aggregate form a single, complete input pattern. Each storage layer node represents one complete stored pattern. The second step identifies the single storage pattern most similar to the input pattern via a kernel, which is a mathematical function that determines the similarity between patterns (Bishop, 2006). Finally, the output node provides the decision associated with that storage pattern's node.

Fast learning models are highly independent. Typically only one storage node is involved in each output decision. If the output of the network does not match the correct output as dictated by the supervised teaching signal, the fast learning model can generate a new, independent storage node with the correct output. Training is immediate and there are no convergence issues. Fully trained models can incorporate

new data without affecting previously stored patterns. Hence, these models are termed fast learning. Overfitting remains a complication.

Figure 2: Typical Form of a Fast Learning Model

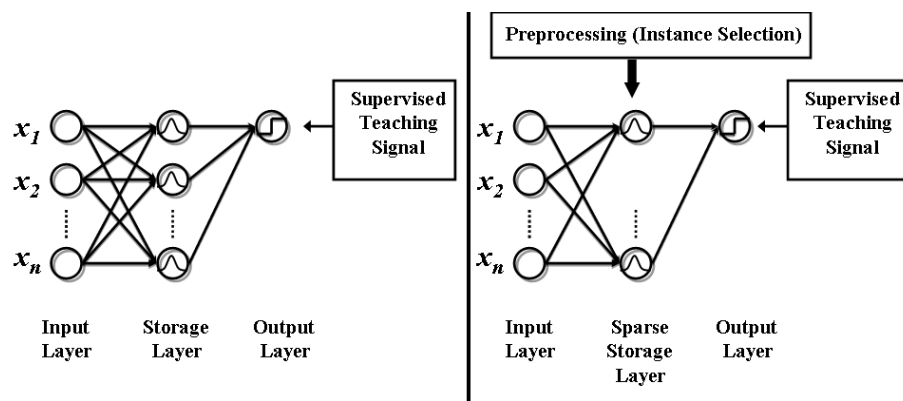


Figure 2: (left) A typical fast learning model. The storage layer replaces the hidden layer from slow learning. Each storage node is independent of each other. The Gaussian symbol represents a common non-linear distance kernel. See text for details (right) A typical maximum margin model, which extends the fast learning model. The preprocessing step limits the number of nodes in the sparse storage layer.

### C Maximum Margin Models

Maximum margin models, such as support vector machines (Cortes & Vapnik, 1995), address the overfitting complication of fast learning models. The key difference from fast learning models is the addition of preprocessing. Figure 2 (right) shows the typical form of a maximum margin model.

Processing proceeds in four steps. The first step, preprocessing, eliminates all unnecessary in-sample input patterns. The preprocessing identifies the key patterns that differ just nearly enough to delineate the margins between the correct, supervised output decisions. These are termed the support vectors. Preprocessing typically relies on slow learning and related model mechanisms such as Lagrange multipliers, quadratic programming, Newton's method, or gradient descent (Bishop, 2006). The final three steps are the same as in fast learning models.

Maximum margin models have very sparse storage layers that reduce the chances of overfitting the dramatically reduced in-sample training data support vectors. However, since the pre-processing relies on slow learning rules, maximum margin models are susceptible to poor adaptability and non-convergence during their pre-processing stages.

### D Temporal Sensitive Models

Temporal sensitive models allow past data and changes in the data to influence current decisions. This can be highly effective in a financial time series where a model needs to track changes in the data over time to identify temporal patterns. Primary examples include Jordan networks (Jordan, 1986), Elman networks (Elman, 1990), and time delay networks. Figure 3 shows the typical form of a temporal sensitive model.

Temporal sensitive models can track the input data as it changes over time. Variants allow the model to consider input data from time  $\{t, t-1, t-2, \dots, t-\infty\}$  to form an unprocessed temporal pattern or to consider the hidden or output layers values from  $\{t-1, t-2, \dots, t-\infty\}$  to form a highly processed temporal pattern. Processing is otherwise identical to slow learning models. While temporal sensitive models can

theoretically incorporate fast learning rules, this configuration tends to be under-explored (e.g. Leung and Tsoi, 2005) due to the extensive modifications required.

Figure 3: Typical Form of a Temporal Sensitive Model

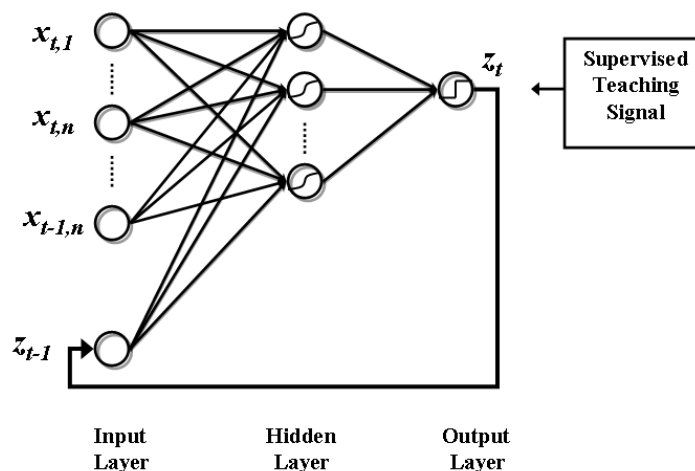


Figure 3: A typical temporal sensitive model, in this case, an out-to-input feedback model with slow learning. Inputs may include patterns over time and feedback from the prior output.  $x_{t,n}$  are the inputs from time  $t$ , and  $z_t$  is the output from time  $t$ .

Temporal sensitive models with slow learning inherit all vulnerabilities of slow learning models. In addition, they are also exposed to potential saturation issues. Saturation occurs, for example, when placing a speaker near a microphone in an infinite loop (Rabiner and Juang, 1993; Armstrong et al, 2009) that causes output intensities to continually increase towards infinity regardless of other microphone inputs. The practical result is that the speaker sustains its maximum output – a maximum volume screech. Similarly, in a temporal sensitive slow learning model with output-to-input (i.e. Jordan, 1986) or hidden-to-input feedback (i.e. Elman, 1990), each feedback connection transmits the product of its maximum thresholded output (e.g. typically one) and their node connection (e.g. up to infinity) to the input node. If the produce of the connection weights in the loop exceeds one, the network can automatically self-sustain a constant output decision regardless of changes in the input. This renders the network non-adaptive and non-reactive to the environment. Additional safeguards and mechanisms are required that either increase the complexity or reduce the sensitivity to time.

### E Mixture of Network Models

Mixtures of networks are meta-models that combine multiple neural networks to simulate a “divide and conquer” approach for the data set. In contrast to individual models that may overfit the data in their attempt to capture the relationships that underlie all the data, mixture of network models allow each network to specialize in specific patterns that fit a portion of the data. Primary examples include AdaBoost (Freund and Schapire, 1996) and its variants (e.g. West, 2000; West et al., 2005).

Processing proceeds in two steps. First, one network attempts to learn the correct decision rules using all in-sample data. The cases where it failed to produce correct output decisions form the in-sample data for the next network. This repeats iteratively for either a specified number of networks or until there are no more incorrect output decisions. Finally, a meta-network weighs the values from each network to arrive at a final decision.

In addition to inheriting all vulnerabilities of the component networks, mixture of network models are highly complicated and susceptible to hand-tuning where small parameter changes can have dramatic impacts.

Despite the theoretical differences, the studies appear to show robust results across all neural network types on a wide range of financial data. However, many of these studies include one or more significant confounds that limit the utility of these tools for out-of-sample use. The vulnerability to these confounds may explain why neural networks have, to the present day, had a more limited impact on financial decision making than previously estimated by the community (Kelly, 2007; Holden, 2009). The next section details seven cardinal confounds that may illuminate how neural networks may be better developed and evaluated for financial decision-making, thereby enhancing differences between network types.

## SEVEN CARDINAL CONFOUNDS IN NEURAL NETWORK

Addressing seven cardinal confounds may help researchers and traders make more informed choices in applying neural networks to financial decision-making. These are: risk neglect, survivorship bias, timing bias, scalability, data dredging, context blindness, and inconstancy.

### A. Risk Neglect

There exists much research in financial decision-making on evaluating the prudence of an investment. One aspect revolves around the capital asset pricing model (Ross et al., 2005; Chartered Financial Analysts Institute, 2009). This model adjusts the return by its accompanying risks, typically as measured by its volatility. For example, an investment that averages 20% gains including a -100% maximum loss is far less prudent than an investment that steadily produces 5% gains. A Sharpe Ratio (Sharpe, 1994), for example divides the excess gain by the standard deviation to provide a simple, objective ratio.

Few of the studies surveyed included any specific discussion on this issue (e.g. Lin et al., 2006; Yu et al., 2008). The remaining studies typically discussed either higher returns in absence of risk or lesser risk in absence of returns. To address this issue, a study needs to include risk-adjusted return metrics or include information that can lead to calculation of the same. Benchmark returns comparisons are insufficient by themselves as they may exhibit different levels of risk and returns.

### B. Survivorship Bias

One of the most famous of stock indices, the Dow Jones Industrial Average, by definition has 30 members. If any member fails or weakens, it is replaced by another stock. While selecting this index or members of this index intuitively should appear to provide a fair and transparent historical data set on which to study, doing so may inadvertently bias the sample. If a study selects an index member as of December 31 for back-testing or training during the preceding period January 1 – December 31, that study may become vulnerable to survivorship bias (Chartered Financial Analyst Institute, 2009). It is exploring data during a year in which it was impossible for the stock to become bankrupt or acquired. If it had, it would not have been a member as of December 31. As the recent bankruptcy de-listing of General Motors in 2009 shows, Dow Jones Industrial Average members are not immune to this possibility. Research shows that survivorship can significantly and artificially inflate nominal returns (Aggarwal and Jorion, 2010).

The majority of the studies used composite indices and may be susceptible to this survivorship bias. For the remaining studies that used non-member individuals, the selection process was unspecified. While

other studies relating neural network models to predicting fraudulent annual reports or bankruptcies are promising (e.g. Boyacioglu et al., 2009; Ng, Quek, and Jiang, 2008; Gaganis et al., 2007), it cannot be taken for granted that a neural network would automatically screen out stocks heading for failure.

Possible improvements may include randomly selecting some or all stocks that were members of the Dow Jones Industrial Average index as of a historical date and using data succeeding that date. While this still may inadvertently leverage off of the stock selection process for the index, it does help the simulation better reflect reality.

### C. Timing Bias

Financial time series data is notoriously highly heteroskedastic and non-stationary. A model that appears to deliver high risk-adjusted returns in one time period may perform inconsistently on another. The study test period may not accurately represent the overall market even shortly before or afterwards. For example, a model may be strong during a period of market expansion, but may not be robust over a period of market contraction. The majority of the studies showed results based on relatively short testing time periods. Sample size estimations aside (Berenson and Levine, 1998), the market environment changes too rapidly for short testing period results of less than one year – 250 trading days – to represent a replicable return.

Typically, investment manager and fund tracking includes a measure of the past performance over at least three years with statistics like proportion of successful years, the worst performance during a year, and how many years the fund has been in operation (Schwager, 1994; Morningstar, 2010). Including similar metrics may help bolster the impact of study results.

### D. Scalability and Complexity Concerns

Financial time series contain massive amounts of data, with new data arriving every trading day. Neural networks are adaptive and reliant on data. Training them on new data as they become available makes use of their flexibility. However, the slow learning neural networks and their derivatives require resetting the network and retraining on all accumulated data to incorporate new data. Only a fast learning neural network model may be suitable since it allows training on new data without overwriting prior learned patterns.

Complexity refers to a network's robustness. Neural networks are notoriously complex models that often incorporate many different settings and values. If modifying one assumed value can alter the results and there are dozens or hundreds of settings, this exposes the network to hand-tuning that relies more upon the user's specific knowledge and experience than on the powerful feature of neural networks to extract general trading rules from data streams. A streamlined slow or fast learning model is more favorable than a complex mixture of networks.

### E. Data Dredging

All neural networks, regardless of the learning rule and structure, are non-parametric and thus highly reliant on the quality and clarity of input data and any preprocessing (Bishop, 2006). The relevance of the input features is fundamental to neural network prediction success (Witten & Frank, 2002). As with nearly all decision tasks, there appears to be a limitless set of potential features in financial decision-making with only a subset being predictive and the remainder generating excessive noise and bias if included (Chartered Financial Analyst Institute, 2009; McQueen and Thorley, 1999).

Two of the studies (Kim, 2006; Versace et al., 2004) used a form of feature selection to remove potentially confounding variables. The remainder studies generally contained overwhelming numbers of

input features, often all highly related. While validation results are very important, additionally testing whether the variables conform to common sense may be vital to improving the explanatory power of a model. One common sense rule that can be automated to promising results is to cluster correlated input features together and select at most one feature per cluster (Schwager, 1994; Wong and Versace, 2010a). More research to improve feature selection techniques is needed.

#### F. Context Blindness

Given a set of clear, high quality features, a decision-making system still needs to interpret them within contexts that can alter their import. Human subjects often treat similar tasks differently under different contexts (e.g. Carraher, Carraher, & Schliemann, 1985; Bjorklund & Rosenblum, 2002). Working memory allows features to be tracked over time to extract a context (Kane & Engle, 2002; Baddeley & Logie, 1999). Context sensitivity enables the decision-maker to disambiguate different feature inputs that may be identical at single points in time (Kane & Engle, 2002).

To better model human decision-making with context sensitivity, a neural network model must be context sensitive. For example, tracking the price over time to determine whether the market is uptrending (bullish) or downtrending (bearish) provides contextual cues (Schwager, 1994). Outside of the temporal sensitive models, only two (Chen et al., 2006; Kim, 2006) used input features that embed temporal triggers or crossing points with limited context sensitivity. However, all temporal context sensitive studies relied on slow learning network models with their accompanying non-convergence and poor adaptability to new data. Temporal sensitive models with slow learning are also highly vulnerable to complexity and scalability concerns when the models include any form of feedback (Wong & Versace, 2010c). Future studies should explore combining temporal models with fast learning, which may be less vulnerable.

#### G. Inconstancy

The financial time series represent buyer and seller actions in a dynamic marketplace that is bound by the laws of supply and demand (Mas-Colell et al., 1995). A high volume buy order in a volatile or illiquid market can trigger a shift from a neutral market to a bullish, uptrending one. The prices may continue to rise such that the ordered trade may not be executed at the desired prices. These are slippage, bid/ask spread, and execution risks (Chartered Financial Analyst Institute, 2009). A trader in real life may not be able to meaningfully replicate the study results due to these effects. Calculating results net-of-trading-costs may better demonstrate more realistic risk-adjusted returns. Future work should also consider applying populations of neural networks to explore dynamic marketplaces, supply-and-demand relationships, and social learning.

### **LESSONS LEARNED**

To better unlock the potential of neural networks for financial decision-making, more research is needed to make informed choices on network type, network enhancement, and network usage. By identifying and addressing seven cardinal confounds and by analyzing which network types are most appropriate for financial data derived from chaotic, non-stationary human decisions, this paper can propose a series of recommendations for future work:

- To improve the probability of the results being replicated on out-of-sample data, include a risk/reward metric, a random sample of stocks instead of an index, and results over a sufficient time frame to include multiple market conditions.

- To improve the likelihood that the results are due to the network predictive power, include a discussion on the effects of parameterization and the network's ability to incorporate new or different data.
- To maximize neural network utility, adaptability, and transparency, select a fast learning model. The independent storage nodes allow for modular design, direct pattern analysis, and adaptability to new data. For these reasons, enhancing a fast learning model with context sensitivity or mixture of network rules may provide more intuitive and therefore effective behaviors and results than currently exist.
- To emulate a human trader's judgment, include network enhancements for modeling automatic feature selection (Wong & Versace, 2010b), context sensitivity (Wong & Versace, 2010c), and if possible, feedback effects from altering the supply and demand relationships.

In addition, more multi-disciplinary collaborations may prove fruitful for applied neural networks. For biologically inspired neural networks to effectively emulate biological financial decision-making, they should extend and conform to existing financial findings. The seven cardinal confounds are well known in the financial literature and serve as guiding posts for evaluating and informing financial studies. If neural networks can also adopt these guideposts, then the choices of model, data, and design may similarly be well-informed. This drives the future of neural network financial studies towards more intelligent decision-making.

## REFERENCES

- Aggarwal, R., and Jorion, P. (2010). Hidden survivorship in hedge fund returns, *Financial Analysts Journal*, 66(2), 69-74.
- Armstrong, S., Sudduth, J., & McGinty, D. (2009). Adaptive feedback cancellation: Digital dynamo, *Advance for Audiologists*, 11(4), 24.
- Baddeley, A. D., & Logie, R. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.
- Berenson and Levine (1998). Basic business statistics, *Prentice Hall*.
- Bishop, C. (2006). Pattern recognition and machine learning, *Springer*.
- Bjorklund, D. & Rosenblum, K. (2002). Context effects in children's selection and use of simple arithmetic strategies, *Journal of Cognition & Development*, 3, 225-242.
- Boyacioglu, M., Kara, Y., and Baykan, O. (2009). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey, *Expert Systems with Applications*, 36, 3355-3366.
- Brockett, P., Golden, L., Jang, J., and Yang, C. (2006). A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction, *The Journal of Risk and Insurance*, 73(3), 397-419.
- Carraher, T., Carraher, D., & Schliemann, A. (1985). Mathematics in the streets and in schools, *British Journal of Developmental Psychology*, 3, 21-29.

- Chartered Financial Analyst Institute (2009). CFA program curriculum, *CFA Institute, Pearson*.
- Chen, W., Shih, J., and Wu, S. (2006). Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets, *International Journal of Electronic Finance*, 1(1), 49-67.
- Cortes, C. and Vapnik, V. (1995). Support vector networks, *Machine Learning*, 20, 273- 297.
- Duhigg, C. (2006). Artificial intelligence applied heavily to picking stocks, *New York Times*, retrieved from: <http://www.nytimes.com/2006/11/23/business/worldbusiness/23iht-trading.3647885.html>
- Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211.
- Freund, Y., and Schapire, R. (1996). Experiments with a new boosting algorithm, *Machine Learning: Proceedings of the Thirteenth International Conference*, 148- 156.
- Gaganis, C., Pasiouras, F., and Doumpos, M. (2007). Probabilistic neural networks for the identification of qualified audit opinions, *Expert Systems with Applications*, 32, 114–124.
- Hardy, Q. (2008). How to make software smarter, *Forbes*, retrieved from: [http://www.forbes.com/2008/03/19/hawkins-brain-ai-tech-innovation08-cz\\_qh\\_0319innovations.html](http://www.forbes.com/2008/03/19/hawkins-brain-ai-tech-innovation08-cz_qh_0319innovations.html)
- Holden, J. (2009). Why computers can't--yet--beat the market, *Forbes*, retrieved from: <http://www.forbes.com/2009/06/18/fina-financial-markets-opinions-contributors-artificial-intelligence-09-joshua-holden.html>
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators, *Neural Networks*, 2, 359-366.
- Jordan, M.I. (1986). Serial order: A parallel distributed processing approach, *Institute for Cognitive Science Report 8604*, University of California, San Diego.
- Kaastra, I., and Boyd, M. (1996). Designing a neural network for forecasting financial time series, *Neurocomputing*, 10, 215-236.
- Kane, M. & Engle, R. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective, *Psychonomic Bulletin & Review*, 9(4), 637-671.
- Kelly, J. (2007). HAL 9000-style machines, Kubrick's fantasy, outwit trader, *Bloomberg*.
- Kim, K. (2006). Artificial neural networks with evolutionary instance selection for financial forecasting, *Expert Systems with Applications*, 30, 519-526.
- Kirkos, E., Spathis, C., and Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements, *Expert Systems with Applications*, 32, 995-1003.
- Kumar, K. and Bhattacharya, S. (2006). Artificial neural network vs linear discriminant analysis in credit ratings forecast, *Review of Accounting and Finance*, 5(3), 216- 227.

- Lin, C., Huang, J., Gen, M., and Tzeng, G. (2006). Recurrent neural network for dynamic portfolio selection, *Applied Mathematics and Computation*, 175, 139-1146.
- Lo, A. (2001). Bubble, rubble, finance in trouble?, *Journal of Psychology and Financial Markets*, 3, 76-86.
- Lo, A. (2007). The efficient markets hypothesis, in *The Palgrave Dictionary of Economics*, Palgrave Macmillan
- Lo, A. & Repin, D. (2002). The psychophysiology of real-time financial risk processing, *Journal of Cognitive Neuroscience* 14(3), 323- 339.
- Mas-Colell, A., Whinston, M., & Green, J. (1995). Microeconomic Theory, *Oxford University Press*.
- McCulloch and Pitts (1943). A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biology*, 5(4), 115-133.
- McQueen, G. and Thorley, S. (1999). Mining fool's gold, *Financial Analysts Journal*, 55(2), 61-72.
- Medeiros, M., Terasvirta, T., and Rech, G. (2006). Building neural networks for time series: a statistical approach, *Journal of Forecasting*, 25, 49-75.
- Moody, J. and Darken, J. (1989). Fast learning in networks of locally tuned processing units, *Neural Computation*, 1, 281-294.
- Morning Star (2010). Retrieved from <http://www.morningstar.com>
- Ng, G., Quek, C., and Jiang, H. (2008). FCMAC-EWS: A bank failure early warning system based on a novel localized pattern learning and semantically associative fuzzy neural network, *Expert Systems with Applications*, 34, 989-1003.
- Rabiner, L. and Juang, B. (1993). Fundamentals of speech recognition, *Upper Saddle River: Prentice Hall*.
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 65(6), 386-408.
- Ross, Westerfield, and Jordan (2005). Fundamentals of corporate finance, *McGraw Hill*.
- Rumelhart, D., Hinton, G., and Williams, R. (1986). Learning internal representations by error propagation, *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*, MIT Press.
- Schwager (1994). The new market wizards: Conversations with America's top traders, *Harper*.
- Sharpe, W. (1994). The sharpe ratio, *Journal of Portfolio Management*, 21(1), 49-58.
- Siegler, R. and Alibali, M. (2005). Children's Thinking, *Prentice Hall*.

Thawornwong, S. and Enke, D. (2004). The adaptive selection of financial and economic variables for use with artificial neural networks, *Neurocomputing*, 56, 205 – 232.

Versace, M., Bhatt, R., Hinds, O., and Schiffer, M. (2004). Predicting the exchange traded fund DIA with a combination of Genetic Algorithms and Neural Networks, *Expert Systems with Applications*, 27(3), 417-425.

West, D. (2000). Neural network credit scoring models, *Computer and Operations Research*, 27, 1131-1152.

West, D., Dellana, S., and Qian, J. (2005). Neural network ensembles for financial decision applications, *Computers and Operations Research*, 32, 2543-2559.

Witten, I. and Frank, E. (2005). Data Mining, *Morgan Kaufman*, San Francisco.

Wong, C., and Versace, M. (2010b). CARTMAP: A neural network method for automated feature selection in financial time series forecasting, submitted to *Neural Computing and Applications*.

Wong, C., and Versace, M. (2010c). Echo ARTMAP: Context sensitivity with neural networks in financial decision-making, *preparation for submission to Winter Global Conference On Business And Finance*, and from *Fourteenth International Conference on Cognitive and Neural Systems, Poster Session, Boston*.

Yu, L., Wang, S., and Lai, K. (2008). Neural network-based mean–variance–skewness model for portfolio selection, *Computers and Operations Research*, 35, 34– 46.

Zhang, D., Jiang, Q., and Li, X. (2005). A heuristic forecasting model for stock, *Decision Making Mathware and Soft Computing*, 12, 33-39.

Zhu, X., Wang, H., Xu, L., and Li, H. (2008). Predicting stock index increments by neural networks: The role of trading volume under different horizons, *Expert Systems with Applications*, 34, 3043–3054.

## BIOGRAPHY

Charles Wong is a PhD candidate at the Cognitive and Neural Systems program, Boston University. He has previously worked for Deutsche Bank AG and KPMG LLP in New York. He can be contacted at [cwong@bu.edu](mailto:cwong@bu.edu).

Max Versace (PhD, Cognitive and Neural Systems, Boston University, 2007) is a Senior Research Scientist at the Department of Cognitive and Neural Systems at Boston University, Director of Neuromorphics Lab, and co-Director of Technology Outreach at the NSF Science of Learning Center CELEST: Center of Excellence for Learning in Education, Science, and Technology. He is a co-PI of the Boston University subcontract with Hewlett Packard in the DARPA Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE) project.