An Empirical Analysis of Data Requirements for Financial Forecasting with Neural Networks

STEVEN WALCZAK

STEVEN WALCZAK is an Assistant Professor in the College of Business and Administration at the University of Colorado at Denver. Current research interests include novel applications of neural networks, expert systems, and other intelligent technologies, particularly in the domains of financial time series and medicine. His research has been published in the *Journal of Management Information Systems, Decision Support Systems, IEEE Transactions on Systems, Man, and Cybernetics, Expert Systems with Applications,* and many other journals.

ABSTRACT: Neural networks have been shown to be a promising tool for forecasting financial time series. Several design factors significantly impact the accuracy of neural network forecasts. These factors include selection of input variables, architecture of the network, and quantity of training data. The questions of input variable selection and system architecture design have been widely researched, but the corresponding question of how much information to use in producing high-quality neural network models has not been adequately addressed. In this paper, the effects of different sizes of training sample sets on forecasting currency exchange rates are examined. It is shown that those neural networks—given an appropriate amount of historical knowledge—can forecast future currency exchange rates with 60 percent accuracy, while those neural networks trained on a larger training set have a worse forecasting performance. In addition to higher-quality forecasts, the reduced training set sizes reduce development cost and time.

KEY WORDS AND PHRASES: forecasting, foreign exchange, neural networks, prediction accuracy, time series, training set size

FINANCIAL TIME SERIES—and foreign exchange rate forecasts in particular—are difficult to model [39]. Hsieh [15] and others [7] have demonstrated that foreign exchange and other financial time series follow a random walk and should therefore not be predictable much past 50 percent (the average performance of random walk models for foreign exchange markets). Neural networks provide a valuable tool for building nonlinear models of data, especially when the underlying laws governing the system are unknown [39]. Neural network forecasting models have outperformed

both statistical and other machine learning models of financial time series, achieving forecast accuracy of up to 58 percent [22, 26], and thus are being widely used to model the behavior of financial time series and to forecast future values for time series [13, 23, 37].

Development of high-quality neural network models is difficult [24, 39]. Most neural network designers develop multiple neural network solutions with regard to the network's architecture (quantity of nodes and arrangement in hidden layers). However, two critical design issues still face financial traders desiring to use neural networks: selection of appropriate variables and capturing a sufficient quantity of training examples to permit the neural network to adequately model the financial time series [38, 39].

Research by Walczak and Cerpa [32] and others [27, 39] has already focused on the question of selecting appropriate input values for modeling time series domains. However, the question of how much historical information is required to produce the best-performing model has not been addressed by neural network researchers. The research presented in this article investigates the requirements for training or modeling knowledge when building nonlinear financial time series forecasting models with neural networks. Homogeneous neural network forecasting models are developed for trading the U.S. dollar against various other foreign currencies. The differences between the neural network models for a specific currency lie solely in the quantity of training data used to develop each time series forecasting model.

A general heuristic for the design of neural networks in financial domains is that the more knowledge that is available to the neural network for forming its model, the better the ultimate performance of the neural network [11, 38], with a minimum of two years of training data a nominal starting point. Box et al. [5] indicate that time series models improve as more data is incorporated into the modeling process. Technical analysis estimates for various financial time series indicate that historical information is required for anywhere from one year [17] to six years [6]. Others [30] have indicated that currency exchange rates have a long-term memory, which implies that larger quantities (periods of time) of data will produce more comprehensive models and produce better generalization.

Neural network researchers have built forecasting and trading systems with training data from one year [22, 34] to sixteen years [39], including various training set sizes in between the two extremes [16, 26, 29]. However, once researchers have obtained their training data, they typically use all of the data in building the neural network forecasting model, with no attempt at comparing data quantity effects on the quality of the produced forecasting models. One of the few existing attempts to evaluate training set size effects has been performed by Zhang and Hu [39]. Zhang and Hu use a single comparison of a 16-year training set size to a 6-year training set size. Their results support Box et al. [5] and others [11], who claim that larger training sets produce better forecasting models, with the 16-year model outperforming the 6-year model.

This article reports research that critically examines the qualitative effect of training set size on neural network foreign exchange rate forecasting models. Training data sets of up to 21.75 years of data are used to predict 1-day future spot rates for several nominal exchange rates [1], representing data from the start of the floating

exchange rate regime to the near present. Multiple neural network forecasting models for each exchange rate forecasting model are trained on incrementally larger quantities of training data. The resulting outputs are used to empirically evaluate whether neural network exchange rate forecasting models achieve optimal performance in the presence of a critical amount of data used to train the network. Once this critical quantity of data is obtained, addition of more training data will not improve (and may in fact hinder) the forecasting performance of the neural network forecasting model.

Results of the research show that for most exchange rate predictions, a maximum of two years of training data produces the best neural network forecasting model performance. The smaller-than-expected training set sizes that produce the optimal neural network forecasting performance led to the induction of the Time Series Recency Effect. The Time Series Recency Effect states that constructing models with data that is closer in time to the data that is to be forecast by the model produces a higher-quality model.

The Time Series Recency Effect has several direct benefits for both neural network researchers and developers. These benefits are:

- Evidence that disproves the time-honored heuristic method of using the greatest quantity of data available for producing time series model.
- Production of higher-quality (with better forecasting performance) models through the use of smaller quantities of data. This will enable improvements in current neural network time series models of 5 percent (or more).
- A net reduction in the development costs of neural network time series models, since less training data is required.
- A net reduction in development time, since smaller training set sizes typically require fewer training iterations to accurately model the training data.

The presence of the Time Series Recency Effect will also serve as a call to research for reexamining previous claims of longevity effects in financial time series. The empirical evidence provided refutes existing heuristics.

Additionally, the 21.75 years of training data is the largest set of data to be used to date to develop neural network foreign exchange forecasting models and thus continues to expand the robustness of neural network time series modeling results. Finally, the empirical method used to evaluate and determine the critical quantity of training data for exchange rate forecasting is generalized for application to other financial time series, indicating the generalizability of the Time Series Recency Effect to other financial time series.

Neural Network Models of Foreign Exchange Rates

Common Features

BACKPROPAGATION NEURAL NETWORKS are the most popular neural network paradigm [10, 37]. Hornik et al. [14] and White [35] have demonstrated that backpropagation neural networks are universal approximators. Various researchers indicate

that backpropagation neural networks have performance superior to other neural network paradigms [3, 4, 31]. Hence, for ease of comparison to previous research, the training algorithm used to enable the neural networks to develop their forecasting models is backpropagation. Readers interested in details of backpropagation neural networks are referred to the references [10, 11, 14, 31, 35].

While some researchers [12, 34] claim that heterogeneous models are necessary, most neural network models of financial time series are homogeneous [21, 22, 23, 26, 39]. Homogeneous models utilize data from the specific time series being forecast or data that are directly obtainable from that time series (e.g., a k-day trend or moving average), while heterogeneous models utilize information from outside the time series (e.g., using the DJIA as part of a TOPIX forecasting model) in addition to the time series itself. Homogeneous models rely on the predictive capabilities of the time series itself, corresponding to a technical analysis as opposed to a fundamental analysis. To reduce the degrees of freedom in the developed neural network models and to maintain consistency with previous research efforts, all of the neural network input variables utilized in the presented research form homogenous sets.

Current neural network research with homogeneous models has achieved neural networks that predict foreign exchange rates with 58 percent accuracy for trading the British pound and 57 percent accuracy for trading the German mark [22, 26]. These values may be used as a benchmark for evaluating the quality of the various neural network models trained with the differing time periods of data.

Evaluating Forecasting Neural Networks

How should the quality of time series forecasting neural networks be measured? Most neural networks forecasting in the capital markets produce an output value that is the future price or exchange rate. Measuring the mean standard error of these neural networks may produce misleading evaluations of the neural networks' capabilities, since even very small errors that are incorrect in the direction of change will result in a capital loss. Instead of measuring the mean standard error of a forecast, Green and Pearson [12] and others [19, 28] argue that a better method for measuring the performance of neural networks is to analyze the direction of change. Therefore, the reported accuracy of the neural network forecasting models developed for the research presented in this paper is the percentage of correct direction of change forecasts made by the neural network.

The direction of change is calculated by subtracting today's price from the forecast price and determining the sign (positive or negative) of the result. The percentage of correct direction of change forecasts is equivalent to the percentage of profitable trades enabled by the ANN system.

Design of the Neural Network Financial Time Series Forecasting Models

As stated in the previous two sections, the neural network models developed for the research reported in this article are backpropagation neural networks that utilize ho-

mogeneous input variables. A preliminary effort to maximize the output performance of the developed backpropagation neural networks is conducted by ensuring adequate domain knowledge representation from the input variables [25, 27, 32] (since analyzing the effect of training set size on neural network models that did not provide any advancement over a standard random walk model would not be very interesting). Hence, following Walczak et al.'s [34] suggestion that multiple time lags provide a significant trading advantage for foreign exchange time series models, for each currency forecasting model to be developed, different lag variable combinations are tested on the full 21.75-year training data.

The input to each neural network model is one or more spot rate lags. Spot rates are determined from the closing price on the New York market. A k-day lag is calculated as:

$$\text{Lag}_k = x_t - x_{t-k}, \tag{1}$$

where x_t is the currency exchange spot rate value at time t, such that a 1-day lag is the difference between the closing spot rates from yesterday and the day before. Lag values for each of the currencies to be forecast are calculated up to a 10-day lag (two business weeks). Lags are then selected for possible inclusion in the final neural network model by selecting significant or the three largest autocorrelation coefficients as reported by Cornell and Dietrich [7] (and later by Levich and Thomas [20]) for each currency, along with the 1-day lag for that currency (since the 1-day lag is the future value to be forecast by the neural networks). All combinations of the identified lags are then trained by neural networks and the neural network that produces the highest accuracy for direction of change predictions on that currency is selected as the final neural network architecture (input variable set). The forecast value for each neural network is the 1-day future lag at time $t+1$ ($x_{t+1} - x_t$).

For example, between the two research studies mentioned [7, 20], the spot rate for trading the U.S. dollar to the British pound had significant or high autocorrelation coefficients for 2-day, 5-day, and 10-day lags, thus including the 1-day lag produces 15 different combinations of lags (calculated as $\Sigma C(4, m)$, where C is the standard combinatorial analysis equation for number of combinations and m varies from 1 to 4). Fifteen neural networks are designed, trained against the full training data set of 21.75 years of data, and evaluated on the final six months of test data. The neural network using a combination of 1-day, 2-day, and 5-day lags produced the best forecasting performance over the test set, so this neural network model is used for the remainder of the testing that uses different sizes of training data sets.

The finding, that a 1-day, 2-day, and 5-day lag combination produces the best possible forecasting model is consistent with the findings of Walczak et al. [34]. Walczak et al. used a heterogeneous model (including 18 additional cross rates) and a fixed training set size of 1 year to determine that a 1-day and 2-day lag combination and a 1-day and 5-day lag combination produced the two best results for trading the U.S. dollar and British pound (they did not evaluate the 1-day, 2-day, and 5-day lag combination found by this research). It should be noted at this point that the selection of the input variables (independent variables), which is one of the two critical problems facing neural network researchers and designers (but is not the focus of this research report),

Table 1. Neural Network Training Sets

Training Group	Data Quantity (in years)	Start Date of Training Set
1	21.75	1 March 1973
2	15	2 January 1980
3	13	4 January 1982
4	11	3 January 1984
5	9	2 January 1986
6	7	4 January 1988
7	5	2 January 1990
8	4	2 January 1991
9	3	2 January 1992
10	2	4 January 1993
11	1	3 January 1994

is based upon performance over the full 21.75-year training data set. The input variable selection method may bias the results toward the use of the full training data set.

Method

THE INITIAL RESEARCH DESIGN produces evidence for the claim that "once an initial minimal amount of training data (from one to five years of data) produces the best forecasting performance, addition of more training data will not improve, and may decrease, the performance of the neural network." First, neural network models are developed to forecast the spot rates of the U.S. dollar against the British pound, German mark, and Japanese yen. These four currencies are the four most heavily traded currencies [9] in the trillion-dollar foreign exchange market [16]. Historic spot rate data values for each of these currency exchange rates are obtained from 1 March 1973 through 30 June 1995. This data sample represents data from the time that a floating exchange rate policy was adopted until the near present.

Various-sized training sets are formulated, as shown in Table 1, ranging from 21.75 years to 1 year of historic training data, with each training group ending at 30 December 1994. The test or evaluation data set is the last two quarters of data, from 2 January 1995 through 30 June 1995, yielding 125 test cases. Dividing the data in this fashion simulates the proactive use of the neural network models for performing foreign exchange trading.

Input variables consisting of different homogeneous combinations of spot rate lag values are selected as described in the "Method" section. Both one- and two-hidden-layer backpropagation neural networks are developed. The research results achieved by the reported research indicate that the single-hidden-layer neural networks consistently outperform the two-hidden-layer neural networks, hence only the single-hidden-layer neural network results are reported.

Table 2. Input Variables Selected As Producing the Best Forecasting Performance

Currency (Traded Against the U.S. Dollar)	Optimal Lag Combination
British pound	1, 2, 5
German mark	1, 2, 3
Japanese yen	1, 3

For the three initial currencies evaluated during the reported research, the lag combinations identified as producing the best forecasting accuracy on the training group 1 (21.75 years) data set are shown in Table 2. For the dollar/pound and the dollar/mark neural networks, a single architecture with the three input values, five hidden-layer perceptrons, and a single output value, as shown in Figure 1, is used for all neural network applications of the different-sized training sets, so that any variation on the output performance of the neural networks is directly attributable to the size of the historic training data set. The dollar/yen neural network also used a single architecture of two input values, three hidden-layer perceptrons, and a single output value. Likewise, all neural network models are trained for an identical quantity of iterations over the training examples (50,000 iterations) and then tested a single time against the 1995 spot rate test set.

Results and Discussion

Evaluating Different Training Set Sizes for Three Common Foreign Exchange Rates

FOR EACH OF THE THREE FOREIGN EXCHANGE RATES TO BE FORECAST, eleven different neural network models are trained (one for each of the eleven different training data sets as shown in Table 1) resulting in 33 different neural network forecasting models. Recall that output accuracy is measured as the percentage of the 1-day future lag neural network forecasts that correctly match the direction of change of the actual 1-day future. Forecasting results for all 33 neural network models are shown in Table 3, with forecasting accuracy greater than 60 percent highlighted in bold. Forecast accuracy values at or above 60 percent are statistically significant, with p-values below 0.001 (the Z-test value for the group 11 German mark values was 8.48 and is similar for the other reported forecasting accuracy values around 60 percent).

It can immediately be seen from the results presented in Table 3 that the best forecasting accuracy for each of the three different currency models is produced with two years of data. While not always the second best with regard to accuracy, forecasting performance with only a single year of data approaches the performance achieved by the 2-year training data set. The effect on the quality of the neural network model forecasting outputs achieved from the quantities of training data is termed the Time Series (TS) Recency Effect. The TS Recency Effect states that for time series data,
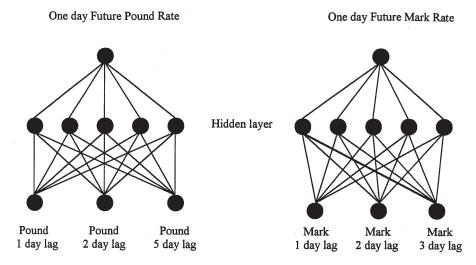
Figure 1. Neural Network Models Used to Test the Effect of Information Quantity

Table 3. Neural Network Foreign Exchange Model Forecasting Accuracy

Training Group (Years of Data)	Dollar/Pound Forecast Accuracy (%)	Dollar/Mark Forecast Accuracy (%)	Dollar/Yen Forecast Accuracy (%)
1 − (21.75)	57.6	59.2	59.2
2 − (15)	56.0	60.0	59.2
3 − (13)	57.6	56.8	59.2
4 − (11)	62.4	58.4	40.8
5 − (9)	52.8	59.2	59.2
6 − (7)	51.2	58.4	59.2
7 − (5)	54.4	60.8	59.2
8 − (4)	52.8	55.2	59.2
9 − (3)	53.6	56.8	56.0
10 − (2)	62.4	61.6	59.2
11 − (1)	61.6	60.0	59.2

model construction data that are closer in time to the values to be forecast produce better forecasting models. This effect is similar to the concept of a random walk model that assumes future values are only affected by the previous time period's value, but the TS Recency Effect is capable of using a wider range of proximal data for formulating the forecasts. The dollar/yen neural network models are the only ones that have a similar (identical) performance result between the full 22-year training set and either of the 1- or 2-year training set models. Modifications to the training data set alone are responsible for an 11.2 percent change in forecasting performance for the dollar/pound neural networks (comparing the 2-year training performance against the worst-case performance of larger training sets), a 6.4 percent change for the dol-

Table 4. Net Gain Enabled by the Different Neural Network Forecasting Models

Neural Network Training Set	Net Gain for Pound Options Trading	Net Gain for Mark Options Trading
1973–1994 (21.75 years)	$762.50	$23,850.00
1993–1994 (2 years only)	$7,550.00	$34,425.00

lar/mark networks, and an 18.4 percent change in the dollar/yen neural network forecasting accuracy.

Another method for evaluating the improvement in forecasting accuracy enabled through the use of smaller training set sizes is to use the neural network forecasts on the test (out-of-sample) set to simulate foreign exchange option trading. The simulation assumes that $250,000 is used in an account to enable two option contracts between the U.S. dollar and either the British pound or German mark to be purchased or sold for each of the 125 days in the test period at the direction of the neural network (pound options are for 31,250 pounds [8] at an average cost of $49,674.14 per contract and mark options are for 62,500 marks [8] at an average cost $43,520.65 per contract during the test period). Japanese yen options contracts are not simulated since there is no difference in the neural network forecast accuracy. Results of the simulation are shown in Table 4. The smaller-training-set-size neural networks provide a net gain on investment of almost ten times the full-training-set neural network forecasts for the pound and almost one and a half times the full-training-set neural network forecasts for the mark.

The research results offer an explanation as to why previous research efforts using neural network models have not surpassed the 60 percent prediction accuracy threshold claimed by Lequarré [18] and demonstrated as a realistic threshold by other research [22, 26]. The difficulty in most prior neural network research is that too much data is typically used. In attempting to build (what at the time was perceived as) the best possible forecasting model, too much training data is used (typically 4 to 6 years of data), thus violating the TS Recency Effect by introducing data into the model that is not representative of the current time series behavior. The 58 percent neural network forecasting performance achieved previously for the dollar/pound is similar to what is achieved by using the full 22-year training set (and the 13-year training set), but still significantly less, with a p-value of 0.0373, than the performance of the 2-year training set. Likewise, the 57 percent accuracy of previous neural network research for the dollar/mark exchange rate is approximated by the 13-and 3-year data training sets.

The pound and the yen neural network models both have at least one other training data set that equals the performance of the 2-year training set (the 11-year set for the pound and most of the other sets for the yen). Bansal et al. [2] have noted that data (training, test, and standard use) represents an important and recurring cost for information systems in general and neural networks in particular. Thus, if the 2-year training set (or the 1-year for the dollar/yen network) produces the best performance and

represents the minimal quantity of data required to achieve this level of performance, then this minimal amount of data is all that should be used to minimize the costs of neural network development and maintenance. The Chicago Mercantile Exchange (CME) reports that it sells historical data on commodities (including currency exchange rates) at the cost of $100 per year per commodity. At this rate, using one to two years of data instead of the full 22 years of data provides an immediate data cost savings of $2000 to $2100 for producing the neural network models. Finally, the dollar/yen neural network results, although apparently resilient to the effects of data quantity, do indicate that selection of the incorrect quantity of data may negatively impact the performance of a neural network solution (as in the 11- and 3-year training sets).

Examination of Data Quality Versus Data Quantity Effects

Recall that the only variation in the neural network models reported in Table 3 is the quantity of data used to build the neural network models. It may be argued that certain years of training data contain noise and would thus adversely affect the forecasting performance of the neural network model. If this is the case, the addition of more (older) training data that is error-free should compensate for the noise effects in middle data, creating a U-shaped performance curve (with the most recent data providing high performance and the largest quantity of data available also providing high performance due to drowning out the noise in middle time frame samples), as appears to be the case with the Japanese yen. The 11-year training set for the dollar/yen exchange rate produces a short-lived U-shape. However, the quality of the neural network predictions for the pound and the mark networks appears to be more cyclical in nature (rising and falling in cycles as the data become more distant from the present). Figure 2 displays the forecasting accuracy for all of the dollar/pound and dollar/mark neural networks.

A test for the possibility of noisy data samples—which would cause a reduction in forecasting performance caused by the errors in the data sample as opposed to the addition of error-free, but nonrelevant data—is performed. As an example, examine the 7-year training set for the dollar/pound network, which produces the poorest forecasting performance for this particular time series. The addition of the 1988 and 1989 data to the more recent training data has a definite negative impact on the forecasting performance of the neural network. If the data are inherently flawed with noise, then similar poor performance should be the result of using this data for modeling and forecasting against any other out-of-sample test set.

Using the discovered principle of the TS Recency Effect, neural network forecasting models are trained with data from 1988 and previously (back to 1985 in 1-year increments) to forecast the 1-day future values for the year 1989. The 1988 data produced a 61.9 percent forecasting accuracy and the use of a 2-year training set, 1987 to 1988, produces the best performance for the 1989 test data at a **69.33** percent forecasting accuracy (a previously unheard-of forecasting accuracy!). Addition of more training data from both 1986 and 1985 to 1986 causes the performance of the new neural networks to decline toward previously reported levels. The 1988 data

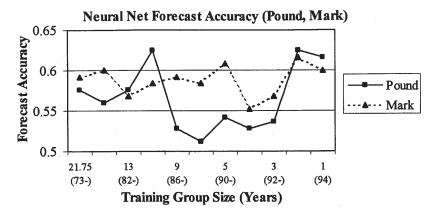**Neural Net Forecast Accuracy (Pound, Mark)**

Figure 2. Graphical View of Time Effect for Training Data on Performance

were also used alone as a training group to forecast the 1995 test data and produced a forecast accuracy of 38.4 percent, indicating that the phenomenal forecasting results of this data with respect to the 1989 test data is due in part, if not wholly, to the TS Recency Effect.

Further evidence for the TS Recency Effect is provided by building a series of neural networks that use rolling training data sets for forecasting the dollar/pound exchange rates. Twenty new neural networks are constructed, with the identical input variables from the previous neural network models, using 2-year training periods to forecast the following year (e.g., a neural network trained on 1977 to 1978 data forecasts the 1979 test data set). Additionally, once a large enough data set is available (a minimum of 6 years), neural network models for alternating forecasting years are trained using all possible previous data (e.g., a neural network to forecast the 1979 data set would be trained on the full 1973 to 1978 training data set). Results are displayed in Table 5. For ease of comparison the 1995 test data set results shown in Table 3 are repeated in the last column of Table 5.

The neural networks that utilize 2-year rolling training groups for forecasting the dollar/pound exchange rate do not always achieve greater than 60 percent forecasting accuracy. Eight out of the 21 (nearly 40 percent) neural networks trained with only two years of data do achieve forecast accuracies of greater than 60 percent and two other neural network forecasting models are at a nearly 60 percent level. However, the purpose of examining the rolling data sets is to compare the 2-year training set models with the larger training set models. Each of the neural networks trained with only a 2-year sample size consistently outperforms the corresponding neural network trained with the largest possible training sample size for any given test set. The average difference for all nine comparisons is 8.6 percentage points, with a maximum difference of 16.37 percentage points in favor of the 2-year training set models.

An identical rolling comparison is also made for the dollar/mark forecasting neural networks and produces similar results. The dollar/mark forecasting neural networks trained on two years of data outperform identical neural network architectures trained

Table 5. Rolling Training Data Sets Forecasting the Dollar/Pound

Group Training (Years)	Dollar/Pound Forecast Accuracy (%)	Forecast Accuracy Using Data from 1973 (%)
1973–1974	57.67	N/A
1974–1975	57.92	N/A
1975–1976	59.46	N/A
1976–1977	56.67	N/A
1977–1978	62.28	52.28 (1973–1978)
1978–1979	59.98	N/A
1979–1980	56.05	43.05 (1973–1980)
1980–1981	57.17	N/A
1981–1982	57.65	56.32 (1973–1982)
1982–1983	63.30	N/A
1983–1984	59.84	55.74 (1973–1984)
1984–1985	58.15	N/A
1985–1986	63.69	47.32 (1973–1986)
1986–1987	56.67	N/A
1987–1988	69.33	55.11 (1973–1988)
1988–1989	58.56	N/A
1989–1990	63.75	51.13 (1973–1990)
1990–1991	64.98	N/A
1991–1992	55.66	54.75 (1973–1992)
1992–1993	60.90	N/A
1993–1994	62.4	57.6 (1973–1994)

on the maximum data sets, with an average difference of 6.03 percentage points and a maximum difference of 11.88 percentage points in favor of the 2-year training set models. Six of the 21 dollar/mark forecasting neural networks trained on only two years of sample data achieved forecast accuracies above 60 percent, with four additional neural networks performing at nearly 60 percent accuracy.

Extending the Foreign Exchange Results to Other Currencies

The TS Recency Effect has been demonstrated for the three most widely traded currencies against the U.S. dollar [9]. These results contradict current methodology, which states that as the quantity of training data used in constructing neural network models increases, the forecasting performance of the neural networks correspondingly improves. The results are tested for robustness by extending the research method to other foreign currencies. Three additional currencies are selected: the French franc, the Swiss franc, and the Italian lira. These three currencies are chosen to approximate the set of nominal currencies used in the previous research of Baillie and Bollerslev [1].

Optimal input lags for each currency's model are determined the same way as for the pound, mark, and yen neural networks, by testing all possible combinations of lags that had previously identified [7, 20] significant or large coefficients. The Swiss

Table 6. Performance of French Franc, Swiss Franc, and Italian Lira Neural Networks

Training Data Set (Years)	Dollar/French Franc Forecast Accuracy (%)	Dollar/Swiss Franc Forecast Accuracy (%)	Dollar/Italian Lira Forecast Accuracy (%)
1 (21.75 years, 1973–94)	59.2	54.4	56.0
2 (5 years, 1990–1994)	57.6	50.4	53.6
3 (4 years, 1991–1994)	44.8	57.6	49.6
4 (3 years, 1992–1994)	59.2	56.8	48.8
5 (2 years, 1993–1994)	57.6	56.8	52.8
6 (1 year, 1994 only)	59.2	55.2	57.6

franc and Italian lira neural network input variables used the 1-day, 2-day, and 3-day lags similar to the German mark neural network shown in Figure 1 and the French franc uses a 1-day and 2-day combination of lags. After the input variable set is determined, neural network models are trained on the full 22-year training data set and the five training data sets containing the recent 1990 through 1994 data values. All of these neural networks follow the same training and testing protocol as used in the original research, with the test data being the corresponding lag values for the first two quarters of 1995.

Results for the six different neural network models for each of the three new currencies are displayed in Table 6. As shown in Table 6, the full 22-year training data set continues to be outperformed by either the 1-year or 2-year training sets (except for the French franc, which has equivalent performance for the most recent and the largest training data sets). The result that the 22-year data set cannot outperform the smaller one or two year training data sets provides further empirical evidence that a critical amount of training data (less than the full 22 years for the foreign exchange time series) produces optimal performance for neural network financial time series models.

The French franc (similar to the Japanese yen) neural network models have identical performance between the largest (22-year) data set and the smallest (1-year) data set. Since no increase in performance is provided through the use of additional data, economics dictates that the smaller 1-year set be used as the training paradigm for the French franc (producing a possible $2100 savings in data costs).

Additionally, the TS Recency Effect is supported by all three currencies. However, the Swiss franc achieves its maximum performance with four years of training data. Notice that the quality of the neural network outputs for the Swiss franc model continually increases as new training data years are added, through the fourth year, then precipitously drops in performance as additional data are added to the training set. Again, the Swiss franc results still support the research goal of determining a critical training set size and the discovered TS Recency Effect. However, the Swiss franc

results indicate that validation tests should be performed individually for all financial time series to determine the minimum quantity of data required to produce the best forecasting performance.

Generalizing the Foreign Exchange Results to Other Financial Time Series

WHILE A SIGNIFICANT AMOUNT OF EVIDENCE has been acquired to support the TS Recency Effect for neural network models of foreign exchange rates, can the TS Recency Effect be generalized to apply to other financial time series? The knowledge that only a few years of data are necessary to construct neural network models with maximum forecasting performance would serve to save neural network developers significant development time, effort, and costs. Additionally, the dollar/Swiss franc neural networks described in the previous section indicate that a cutoff of two years of training data may not always be appropriate.

A methodology for determining the optimal training set size for financial time series neural network models is displayed in Figure 3. Figure 3 uses a flowchart to depict the process of the training set size determination methodology. The methodology does not attempt to address other important issues in developing neural network models, such as selecting appropriate input variables [25, 27, 32, 39], but labels them as preparation (preprocess) steps.

The methodology may be described as an iterative technique that starts with a single year of training data and continues to add additional years of training data until the trained neural network's performance begins to decrease (i.e., the process continues to search for better training set sizes as long as the performance increases or remains the same). The optimal training set size is then set to be the smallest quantity of training data to achieve the best forecasting performance.

Since the described methodology is a result of the empirical evidence acquired using foreign exchange rates, it stands to reason that testing the methodology on additional neural network foreign exchange rate forecasting models will continue to validate the methodology. Therefore, three new financial time series are used to demonstrate the robustness of the specified methodology. The DJIA stock index closing values, the closing price for the individual DIS (Walt Disney Co.) stock, and the CAC-40 French stock index closing values serve as the three new financial time series. Data samples from January 1977 to August 1999 (to simulate the 22 years of data used in the foreign exchange neural network training) are used for the DJIA and DIS time series and data values from August 1988 to May 1999 are available for the CAC-40 index.

Input variables are selected by performing an autocorrelation on lag values of the daily data from a 1-day lag to a 10-day lag and selecting the lags with the most significant coefficients. The DJIA index neural network model used the 1-day and 5-day lags as input variables, while the DIS stock neural network used the 1-day, 3-day, 4-day, 5-day, 6-day, and 7-day lags as input variables, and the CAC-40 index neural

May need to test multiple architectures

```
┌─────────┐     ┌──────────────┐         ┌──────────────────┐
│  Start  │ ──▶ │  Determine   │  ────▶  │     Select       │
└─────────┘     │ Input Variables│        │ Neural Network   │
                │ (see [25,27,32,39]) │   │  Hidden Node     │
                └──────────────┘         │  Architecture    │
                                         │  (see [10,25,32]) │
                                         └──────────────────┘
```

Data Samples For Training & Testing

Create 1-year training set (using most recent data) and appropriate test set

Train with newest training set and test on original test set

Add 1 additional year of training data (closest to current training set)

Train with 1-year set and test (baseline)

Is the performance of the new larger training set ≥ previous performance?

Yes

No

Finish

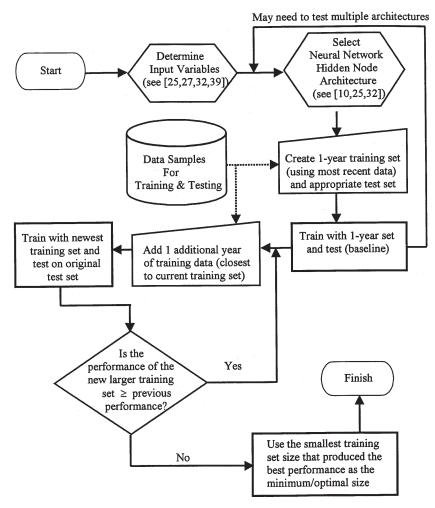Use the smallest training set size that produced the best performance as the minimum/optimal size

Figure 3. Methodology for Determining Minimum/Optimal Training Set Size

network model used the 1-day and 3-day lags as input variables. Multiple single- and two-hidden-layer architectures are evaluated, duplicating the experimental method for the foreign exchange backpropagation neural networks, and the neural network (for each of the three new time series) that produced the highest forecasting performance is reported below. The output for all of the neural networks is the 1-day future change in value, evaluated for direction of change (as with the previous foreign exchange neural networks).

Following the methodology displayed in Figure 3, three backpropagation neural networks—one for each of the two time series–are trained on the 1998 data set and tested a single time on the 1999 data values (164 cases, 123 cases for the CAC-40). Then a single year is added to the training set, a new neural network model is trained and tested a single time, with the process repeated until a decrease in forecasting

Table 7. Evaluation of the Method to Determine Minimum Training Sets for DJIA and DIS

Training Data Set (Years)		DJIA Forecasting Accuracy (%)	DIS Forecasting Accuracy (%)	CAC-40 Forecasting Accuracy (%)
1998	(1)	56.71	54.88	52.85
1997−1998	(2)	54.27	45.73	56.10
1996−1998	(3)	55.49	46.34	54.47
1995−1998	(4)	54.88	46.34	52.85
1994−1998	(5)	55.49	46.34	52.85
1993−1998	(6)	N/A	N/A	53.66
1988/9−1998	(10)	56.10	48.78	53.66
1977−1998	(22)	54.27	48.78	N/A

performance occurs. An additional three years of training data, in 1-year increments, is added to the training sets and evaluated to strengthen the conclusion that the optimal training set size has been acquired. A final test of the usefulness of the generalized methodology for determining minimum optimal training set sizes is performed by training similar neural network models on the full 22-year training set for the DJIA index and DIS stock neural networks and on the 10-year training set for all networks, which is the maximum data quantity available for the CAC-40. Then each of the neural networks trained on the "largest" training sets is tested on the 1999 test data set to evaluate the forecasting performance. Results for the three new backpropagation neural network financial time series models are displayed in Table 7.

For both the DJIA and the DIS stock, the 1-year training data set is immediately (as soon as the neural network trained on the 2-year data set is tested) identified as the best size for a training data set. The CAC-40 neural network forecasting model however, achieves its best performance with a 2-year training data set size. While the forecasting accuracy for these three new financial time series is not remarkable (not achieving the 60 percent forecasting accuracy as do many of the foreign exchange forecasting neural networks), it does support the generalized method for determining minimum necessary training data sets and consequently lends support to the Time-Series Recency Effect. Once the correct or best performing minimum training set is identified by the generalized method, no other neural network model trained on a larger training set is able to outperform the "minimum" training set. The two neural network models that forecast stock index future value changes do achieve an accuracy over 56 percent, which is a significant result at the 0.10 level using a single tailed Z-test.

The results for the DIS stock value are slightly better, but still similar to results reported by White [36] for a backpropagation neural network that forecast the IBM stock during the late 1970s. White's conclusion was that his neural network model, which used approximately four years of training data, emulated a simple efficient market. A random walk model of the DIS stock produced a 50 percent prediction accuracy and so the DIS neural network forecasting model does outperform the ran-

dom walk model, but not by a statistically significant amount. An improvement to White's neural network model to predict IBM stock price changes may be achieved by following the generalized methodology for determining the best size of the training set and reducing the overall quantity of training data, thus limiting the effect of nonrelevant data.

Again, as an alternative evaluation mechanism, a simulation is run with the CAC-40 stock index data. A starting value of $10,000 with sufficient funds and/or credit is assumed to enable a position on 100 index options contracts. Options are purchased or sold consistent with the neural network forecasts for the direction of change in the CAC-40 index. All options contracts are sold at the end of the year-long simulation. The 2-year training data set model produces a net gain of $16,790, while using the full 10-year training data set produces a net loss of $15,010. The simulation results yield a net average difference between the TS Recency Effect model (of 2 years) and the heuristic greatest quantity model (10 years) of $31,800, or three times the size of the initial investment.

Conclusions

EVIDENCE HAS BEEN PRESENTED that contradicts the current financial neural network development heuristic, which implies that greater quantities of training data necessarily produce better-quality forecasting models. A new time series model effect, termed the Time-Series Recency Effect, has been demonstrated to work consistently across neural network models for six different currency exchange time series. The TS Recency Effect claims that model building data that is nearer in time to the out-of-sample values to be forecast produces more accurate forecasting models.

The empirical results discussed in this article show that frequently a smaller quantity of training data will produce a better-performing backpropagation neural network model of a financial time series. Other problematic issues related to the development of the best possible neural network model—such as selection of input variables or selection of the neural network training algorithm and design of the neural network architecture—are not discussed. The prudent financial time series neural network developer will realize that these other factors will affect the neural network model's performance and should utilize existing guidelines [10, 25, 27, 32, 39] to solve these other issues.

The presented research results indicate that for financial time series two years of training data is frequently all that is required to produce optimal forecasting accuracy. Results from the Swiss franc models alert the neural network researcher that the TS Recency Effect may extend beyond two years. A generalized methodology is developed for determining the minimum training set size that produces the best forecasting performance. The generalized method was tested on two stock indexes (DJIA and CAC-40) and an individual stock value (DIS) to demonstrate the functionality and practicality of the methodology. Neural network researchers and developers using the generalized method for determining the minimum necessary training set size will be

able to implement neural networks with the highest forecasting performance at the least cost.

Great quantities of financial time series exist that were not examined in this article. Future research can continue to provide evidence for the TS Recency Effect by examining the effect of training set size for additional financial time series (e.g., any other stock or commodity and any other index value). The TS Recency Effect may not be limited only to financial time series, and evidence from nonfinancial time series domain neural network implementations already indicates that smaller quantities of more recent modeling data are capable of produce high-performance forecasting models [33].

Additionally, the TS Recency Effect has been demonstrated with neural network models trained using backpropagation. Future research will attempt to demonstrate that the TS Recency Effect holds for all supervised learning neural network training algorithms (e.g., radial basis function, fuzzy ARTMAP, probabilistic, etc.) and is therefore a general principle for time series modeling and not restricted to backpropagation neural network models.

In conclusion, Bansal et al. [2] have noted that neural network systems incur costs from training data. This cost is not only financial, but also impacts the development time and effort. Empirical evidence demonstrates that frequently only one or two years of training data will produce the "best" performing backpropagation trained neural network forecasting models. The proposed methodology for identifying the minimum necessary training set size for optimal performance enables neural network researchers and implementers to develop the highest-quality financial time series forecasting models in the shortest amount of time and at the lowest cost.

REFERENCES

1. Baillie, R.T., and Bollerslev, T. Cointegration, fractional cointegration, and exchange rate dynamics. *Journal of Finance, 49* (1994), 737–745.

2. Bansal, A.; Kauffman, R.J.; and Weitz, R.R. Comparing the modeling performance of regression and neural networks as data quality varies: a business value approach. *Journal of Management Information Systems, 10,* 1 (1993), 11–32.

3. Barnard, E., and Wessels, L. Extrapolation and interpolation in neural network classifiers. *IEEE Control Systems, 12,* 5 (1992), 50–53.

4. Benjamin, C.O.; Chi, S.; Gaber, T.; and Riordan, C.A. Comparing BP and ART II neural network classifiers for facility location. *Computers and Industrial Engineering, 28,* 1 (1995), 43–50.

5. Box, G.E.P.; Jenkins, G.M.; and Reinsel, G.C. *Time Series Analysis 3rd Edition.* Englewood Cliffs, NJ: Prentice Hall, 1994.

6. Cassetti, M.D. A neural network system for reliable trading signals. *Technical Analysis of Stocks and Commodities, 11,* 6 (1993), 78–84.

7. Cornell, W.B., and Dietrich, J.K. The efficiency of the market for foreign exchange under floating exchange rates. *The Review of Economics and Statistics, 60* (1978), 111–120.

8. Daigler, R.T. *Financial Futures and Options Markets.* New York: Harper Collins, 1994.

9. Federal Reserve Bank of New York. *April 1995 Survey of Foreign Exchange Market Activity in the United States.* New York: Federal Reserve Bank, 1995.

10. Fu, L. *Neural Networks in Computer Intelligence.* New York: McGraw-Hill, 1994.

11. Gately, E. *Neural Networks for Financial Forecasting.* New York: Wiley, 1996.

12. Green, H., and Pearson, M. Neural nets for foreign exchange trading. In G.J. Deboeck (ed.), *Trading on the Edge.* New York: Wiley, 1994, pp. 123–129.

13. Hammerstrom, D. Neural networks at work. *IEEE Spectrum, 30,* 6 (1993), 26–32.

14. Hornik, K.; Stinchcombe, M.; and White, H. Multilayer feedforward networks are universal approximators. *Neural Networks, 2,* 5 (1989), 359–366.

15. Hsieh, D.A. Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business, 62,* 3 (1989), 339–368.

16. Jamal, A.M., and Sundar, C. Modeling exchange rates with neural networks. *Journal of Applied Business Research, 14,* 1 (1998), 1–5.

17. Kean, J. Treasury bond yields: a neural net analysis approach. *Technical Analysis of Stocks and Commodities, 11,* 4 (1993), 78–84.

18. Lequarré, J.Y. Foreign currency dealing: a brief introduction. In N.A. Gershenfeld and A.S. Weigend (eds.), *Time Series Prediction: Forecasting the Future and Understanding the Past.* Reading, MA: Addison-Wesley, 1993, pp. 131–137.

19. Levich, R.M. How to compare chance with forecasting expertise. *Euromoney* (August, 1981), 61–78.

20. Levich, R.M., and Thomas, L.R. The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach. *Journal of International Money and Finance, 12* (1993), 451–474.

21. Mehta, M. Foreign exchange markets. In A. Refenes (ed.), *Neural Networks in the Capital Markets.* New York: Wiley, 1995, pp. 177–198.

22. Refenes, A.N. Constructive learning and its application to currency exchange rate forecasting. In R. Trippi and E. Turban (eds.), *Neural Networks in Finance and Investing.* New York: Irwin, 1993, pp. 465–493.

23. Ruggiero, M.A. Getting the lag out. *Futures, 23,* 4 (1994), 46–48.

24. Ruggiero, M.A. Neural networks: Tahiti or bust. *Artificial Intelligence in Finance, 2,* 1 (1995), 15–20.

25. Smith, M. *Neural Networks for Statistical Modeling.* New York: Van Nostrand Reinhold, 1993.

26. Steurer, E. Nonlinear modelling of the DEM/USD exchange rate. In A. Refenes (ed.), *Neural Networks in the Capital Markets.* New York: Wiley, 1993, pp. 199–211.

27. Tahai, A.; Walczak, S.; and Rigsby, J.T. Improving artificial neural network performance through input variable selection. In P. Siegel, K. Omer, A. deKorvin, and A. Zebda (eds.), *Applications of Fuzzy Sets and the Theory of Evidence to Accounting II.* Stamford, CN: JAI Press, 1998, pp. 277–292.

28. Taylor, S.J. *Modelling Financial Time Series.* New York: Wiley, 1988.

29. Tenti, P. Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence, 10* (1996), 567–581.

30. Van de Gucht, L.M.; Dekimpe, M.G.; and Kwok, C.C. Persistence in foreign exchange rates. *Journal of International Money and Finance, 15,* 2 (1996), 191–220.

31. Walczak, S. Neural network models for a resource allocation problem. *Transactions on Systems, Man and Cybernetics, 28B,* 2 (1998), 276–284.

32. Walczak, S., and Cerpa, N. Heuristic principles for the design of artificial neural networks. *Information and Software Technology, 41,* 2 (1999), 107–117.

33. Walczak, S., and Krause, J. Chaos, neural networks, and gaming. In E.A. Yfantis (ed.) *Intelligent Systems Volume 1.* Dordrecht, Netherlands: Kluwer Academic, 1995, pp. 457–466.

34. Walczak, S.; Tahai, A.; and Karim, K. Improved cash flows using neural network models for forecasting foreign exchange rates. In P. Siegel, K. Omer, A. deKorvin, and A. Zebda (eds.), *Applications of Fuzzy Sets and the Theory of Evidence to Accounting II.* Stamford, CN: JAI Press, 1998, pp. 293–310.

35. White, H. Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings. *Neural Networks, 3,* 5 (1990), 535–549.

36. White, H. Economic prediction using neural networks: the case of IBM daily stock returns. In R.R. Trippi and E. Turban (eds.), *Neural Networks in Finance and Investing.* New York: Irwin, 1993, pp. 315–328.

37. Widrow, B.; Rumelhart, D.E.; and Lehr, M.A. Neural networks: applications in industry, business and science. *Communications of the ACM, 37,* 3 (1994), 93–105.

38. Zahedi, F. A meta-analysis of financial applications of neural networks. *International Journal of Computational Intelligence and Organizations, 1,* 3 (1996), 164–178.

39. Zhang, G., and Hu, M.Y. Neural network forecasting of the British pound/US dollar exchange rate. *Omega, International Journal of Management Science, 26,* 4 (1998), 495–506.