NON-LINEAR EXCHANGE RATE FORECASTING: THE ROLE OF MARKET MICROSTRUCTURE VARIABLES

by

NIKOLA GRADOJEVIC

B.Sc., University of Novi Sad, 1996
M.A., Central European University, 1998

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in
THE FACULTY OF GRADUATE STUDIES

Department of Economics

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 2002
© Nikola Gradojevic, 2002
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Economics
The University of British Columbia
Vancouver, Canada

Date March 28, 2003
Abstract

In this dissertation, we conduct a study of exchange rate models for the Canada/U.S. exchange rate. More specifically, we focus on their intra-day (high-frequency) and, subsequently, weekly forecast performances. All attempts to explain equilibrium exchange rates suffer from various problems: structural (macroeconomic) models used for out-of-sample forecasting produce poor forecasts. Given that different market participants trade based on private as well as public information sets, it is natural to assume that equilibrium exchange rate expectations are formed from a combination of macroeconomic fundamentals and market microstructure variables.

Chapter 1 motivates research in the area of non-linear microstructure exchange rate modeling, reviews the recent literature and introduces the general ideas behind this thesis.

Chapter 2 outlines Artificial Neural Networks (ANNs) and other non-linear modeling approaches used in this research.

Chapter 3 introduces a non-linear Canada/U.S. exchange rate microstructure model and provides a strong evidence for the microstructure effects. Our horse race for forecast performance results in a non-linear ANN model as the winner. ANN models outperform random walk and linear models in a number of recursive out-of-sample forecasts. The daily forecasts produced by ANN models are statistically significant
according to Diebold and Mariano (1995) statistics. Apart from the nearest neighbours model, other linear and non-linear models are unable to generate significant predictions. The inclusion of a microstructure variable, order flow, substantially improves the predictive power of both the linear and non-linear models. Our findings also indicate the necessity of embodying (in a non-linear sense) information not only from interbank order flows, but also from commercial client and foreign institution transactions. No matter which non-linear model is used, there is always a slight forecast gain when dealer's private order flows are included into a set of explanatory variables.

Chapter 4 describes fuzzy logic technology in the form of approximate reasoning as a method that can be used in economics when dealing with continuous and imprecise economic variables, insufficient data for analysis and when a mathematical model of the process is unknown.

Chapter 5 develops an original and novel approach to generating trading strategies in the foreign exchange (FX) market based on forecasts from the ANN. Neuro-fuzzy (NF) decision-making technology is designed and implemented to obtain the optimal daily currency trading rule. We find that a non-linear ANN exchange rate microstructure model combined with a fuzzy logic controller (FLC) generates a set of trading strategies that, on average, earn a higher rate of return compared to the simple buy-and-hold strategy. We also find that after including transaction costs, the gains from the NF technology do not decline and increase on some periods. Finally, we successfully apply the NF model to the problem of determining the FX market's sentiment as reflected by the chartists' trading signals during periods of strong depreciation.
Contents

Abstract ................................................................. ii

Contents ........................................................................ iv

List of Tables ............................................................. viii

List of Figures ............................................................. ix

Acknowledgments ........................................................ x

1 Introduction ................................................................... 1

2 Artificial Neural Networks (ANNs) and Other Non-linear
Methodologies .............................................................. 9

2.1 ANNs ........................................................................ 9
  2.1.1 Definition and structure ........................................ 9
  2.1.2 Learning and adapting ........................................ 15
  2.1.3 Backpropagation ANN ....................................... 16
  2.1.4 Recurrent ANN ................................................ 21
2.2 NN method ................................................................. 23
2.3 TSK method .............................................................. 27
  2.3.1 Introduction .......................................................... 27
  2.3.2 TSK fuzzy model building ....................................... 29

3 Exchange Rate Forecasting: A Market Microstructure Approach 34
  3.1 Models of Exchange Rate Determination ................. 34
  3.2 Model Specification .................................................. 45
    3.2.1 Data description .............................................. 45
    3.2.2 Assessment of forecast performance .................. 51
  3.3 Empirical Results ................................................... 56
    3.3.1 Models based on individual order flows when controlled for the day-of-the-week-effects 56
    3.3.2 Parameter estimates and relative significance of inputs 63
    3.3.3 Models based on individual and aggregate order flows when not controlled for the day-of-the-week-effects 74
  3.4 Conclusions .......................................................... 87

4 An Introduction to Fuzzy Logic ....................................... 89
  4.1 Introduction .......................................................... 89
  4.2 Fuzzy System .......................................................... 92

5 Neuro-Fuzzy Decision-Making in Foreign Exchange Trading and
Other Applications ............................................................. 98
  5.1 Neuro-Fuzzy Design ............................................... 98
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Market Environment</td>
<td>106</td>
</tr>
<tr>
<td>5.3 Estimation Results</td>
<td>109</td>
</tr>
<tr>
<td>5.4 Excess Volatility and the NF Model</td>
<td>116</td>
</tr>
<tr>
<td>5.5 &quot;Continuous&quot; Trading Decisions and a Risk-Averse Investor</td>
<td>125</td>
</tr>
<tr>
<td>5.6 Conclusions and Further Research</td>
<td>128</td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
<td>131</td>
</tr>
</tbody>
</table>
List of Tables

3.1 Augmented Dickey-Fuller (ADF) unit root t-tests. 47
3.2 Out-of-sample forecast performance. 61
3.3 Two-step-ahead and three-step-ahead forecast performance of the ANNB2 model. 62
3.4 The ANN models with the greatest forecast improvements. 64
3.5 Estimated ANN's connection weights and node biases (Model 1). 65
3.6 Estimated ANN's connection weights and node biases (Model 2). 66
3.7 Estimated ANN's connection weights and node biases (Model 3). 66
3.8 Estimated ANN's connection weights and node biases (Model 4). 67
3.9 Relative contribution (pseudo-weights) of inputs for the ANN models. 68
3.10 Estimation Results For Linear Models. 76
3.11 RMSE (exp 10^{-3}) for ANN, linear, and random walk models (j=1). 81
3.12 RMSE (exp 10^{-2}) for ANN, linear, and random walk models (j=7). 82
3.13 The average percentages of correctly predicted signs for linear models 1 and 2 (LM 1 and LM 2), ANN models 1 and 2 (ANN 1 and ANN 2), and the random walk model. 83
3.14 PERC and RMSE (exp 10^{-3}) statistics for the recursive estimation over the whole testing set (k=225). 85
3.15 PERC(POS) and PERC(NEG) for ANN models 1 and 2 (ANN 1 and ANN 2). 86

5.1 The intervals for discrete decisions on traded volume (based on a defuzzified output z). 105
5.2 The number of moving windows when the NF technology outperforms a simple buy-and-hold strategy. 110
5.3 The intervals for discrete trading signals the traders receive from the NF system (based on a defuzzified output z). 123
5.4 Observed ANN (model 2) direction-of-change forecasting statistics (frequencies) for July/August 1998 inputs (34 observations). 123
5.5 Observed ANN (model 2) direction-of-change forecasting statistics (frequencies) for November/December 1994 inputs (27 observations). 124
5.6 The number of moving windows when the NF technology outperforms risk-averse investor's strategies. 127
## List of Figures

2.1 Computational element or node with N inputs and one output (weighted sum of inputs). 10  
2.2 A three-layered ANN architecture. 11  
2.3 A three-layered backpropagation ANN. 17  
2.4 A three-layered Elman recurrent ANN. 22  

3.1 Three-stage exchange rate determination from Lyons and Evans (2002). 41  
3.2 Above: Aggregate (cumulative) order flow and log Canada/U.S. real exchange rate. Below:  
    IB (cumulative) order flow and log Canada/U.S. real exchange rate. 43  
3.3 Non-linear pattern of response (ANNBP2 model). 71  
3.4 Non-linear pattern of response for the 1994 depreciation episode (ANNBP2 model). 72  
3.5 ANN errors: training, validation, and testing. 77  
3.6 An illustration of linear and ANN model exchange rate forecasts. Actual values are denoted  
    by circles. 78  
3.7 RMSE for ANN model 1, linear model 1, and random walk (j=1). 79  
3.8 RMSE for ANN model 2, linear model 2, and random walk (j=1). 79  
3.9 RMSE for ANN model 1, linear model 1, random walk (j=7). 80  
3.10 RMSE for ANN model 2, linear model 2, and random walk (j=7). 80  
3.11 Example: recursive estimation PERC for ANN model 2, linear model 2, and random walk  
    model (1-day forecast). 84  

4.1 Fuzzy control system. 93  
4.2 The example for the relationship between variables, fuzzy sets and triangular membership  
    functions for the variable "trading action". 95  

5.1.a Gaussian membership functions for the variable "exchange rate change forecast". 100  
5.1.b Triangular membership functions for the variable "FX trader's action". 101  
5.2 NF system for the FX trading. Two nonlinear models are used for forecasting. 102  
5.3 An example of Mamdani fuzzy inference. 104
5.4.a NF and the buy-and-hold rates of return (ROR) for the 10-day moving-window (model 2 without the transaction costs) and NF excess returns. 111
5.4.b NF and the buy-and-hold rates of return (ROR) for the 10-day moving-window (model 2 with the transaction costs) and NF excess returns. 112
5.4.c NF and the buy-and-hold rates of return (ROR) for the 10-day moving-window (model 1 without the transaction costs) and NF excess returns. 112
5.4.d NF and the buy-and-hold rates of return (ROR) for the 10-day moving-window (model 1 with the transaction costs) and NF excess returns. 113
5.5.a NF and the buy-and-hold rates of return (ROR) for the 20-day moving-window (model 2 without the transaction costs) and NF excess returns. 113
5.5.b NF and the buy-and-hold rates of return (ROR) for the 20-day moving-window (model 2 with the transaction costs) and NF excess returns. 114
5.5.c NF and the buy-and-hold rates of return (ROR) for the 20-day moving-window (model 1 without the transaction costs) and NF excess returns. 114
5.5.d NF and the buy-and-hold rates of return (ROR) for the 20-day moving-window (model 1 with the transaction costs) and NF excess returns. 115
5.6 Gaussian membership functions for the variable “exchange rate change forecast”. 121
5.7 Triangular membership functions for the variable “FX trader’s action.” 122
5.8 Tan-sigmoid function used to estimate the risk-averse investor’s strategies. 126
Acknowledgments

First and foremost, I would like to thank my thesis supervisor, Dr. John Cragg, who played a key role in the conception and development of the entire project. It has been an honour to have the opportunity to work with John, first as a research assistant, and later as a Ph.D. candidate. In working with John, I learned something new and exciting not only about economics, but also about the world each time we would meet. His expertise, support, patience, and friendship have set an example I hope to match some day.

Furthermore, I would like to express my gratitude to Glen Donaldson for his constructive and invaluable ideas that have made their mark throughout the work. Michael Devereux also provided some critical and helpful comments.

A very special thanks goes out to Angela Redish, without whose support and advice I would not have not been able to materialize this thesis.

Thanks to all of my friends at the Bank of Canada including Peter Thurlow, Toni Gravelle, Jing Yang, and many more, who initialized and greatly contributed to this research.

I also owe thanks to Dragan Kukolj (University of Novi Sad) and my former colleagues at the Faculty of Engineering who introduced me to the world of artificial
intelligence and its applications in economics, and who helped me with the empirical part of my work.

Marko Vujicic, Martin Berka, Christoph Schleicher and my other classmates have made the Ph.D. program fun and interesting. Marko also provided support and friendship during some difficult times.

Many thanks to the economics department secretaries for their positive energy and assistance.

Finally, this thesis would have never been possible without the love and support of my parents, Jovan and Ljubica Gradojevic, and my sister, Tina, who have never doubted my success on this journey.

I would like to dedicate this thesis to my “princesses”: my wife, Zana and my daughter, Marina. Thank you for all that I have become from the moment you have enlightened my life.
Chapter 1

Introduction

Understanding exchange rate movements has long been an extremely challenging and important task for academic and business researchers. Not only is the exchange rate a significant determinant of aggregate demand in a small open economy such as Canada’s, but it also responds immediately and materially to monetary policy (e.g., Clinton 2001). However, the response is not always predictable. This makes it sometimes difficult to achieve the desired results of monetary policy implementation.

Efforts to deepen our understanding of exchange rate movements have taken on a number of approaches. Initially, efforts centred on the development of low-frequency macroeconomic (fundamental) empirical models. More recently, efforts have been aimed at the development of high-frequency models of the foreign exchange (FX) market, based on microeconomic (microstructure) variables. Throughout, however, forecasting models have been developed for obvious utilitarian purposes, but they have also served as a gauge of our understanding of exchange rate movements. Moreover, they can sometimes help to pinpoint where the gaps in our knowledge may lie, and therefore suggest new avenues of research. The exchange rate forecasting model developed in this thesis serves all of these purposes.
Given the failure of traditional FX rate models to explain and predict exchange rate fluctuations correctly (Meese and Rogoff 1983), we turn to the market microstructure of FX markets in chapter 3. In recent years there has been a lot of evidence that the behavior of dealers and other market participants can influence equilibrium exchange rates (Lyons and Evans 2002, Yao 1997, Covrig and Melvin 1998). Inventory adjustments and bid-ask spread reactions to informative incoming order flows are two examples in which dealer behavior affects exchange rate determination. Indeed, given that different market participants trade based on private as well as public information sets, it is natural to assume that equilibrium exchange rate expectations are formed based on a combination of macroeconomic fundamentals and market microstructure variables (Goldberg and Tenorio 1997). The highest frequency used in this thesis is a daily frequency as we try to balance the tension between macroeconomic and microstructure effects.\footnote{The development of electronic trading and the Internet now allows us to follow price formation in real-time. Ideally, in order to truly understand foreign exchange markets one can attempt to follow our approach on tick-by-tick basis. In financial markets, the DGP is a complex network of layers where each layer corresponds to a particular frequency. As the highest available frequency for the microstructure data is daily, we leave this complete characterization of the true DGP to further research and aggregate to daily (and later weekly) information assuming the influence of random noise does not have a significant impact on our findings. Thus, the messages from this thesis to the mainstream paradigm of possible data generating mechanism of FX rates correspond to a particular frequency and require further investigation, once the data become available.} Another motivation for using high-frequency data is the capability to develop and estimate a model based on one decade of data. It is unrealistic to achieve a consistent and reliable FX model based on decades or even a century of data: markets are rapidly changing and one could be in fact analyzing a number of
fundamentally different markets. Also, by using a great amount of data we can explore more complicated (non-linear) models and reduce a risk of missing important features of the short-term data generating process (DGP).

This thesis is oriented toward generating new and useful theoretical and empirical insights about trader behavior, the informational role of trades and the way that information is processed in the FX market. First, in chapter 3, we extend the portfolio shifts FX model of sequential trading (Lyons and Evans 2002) to find that the fusion of fundamental exchange rate determinants used by the Bank of Canada and different order flow types matter in a highly non-linear fashion.

In chapter 2 we introduce the artificial neural network (ANN) model. The application of such a model brings examination of high-frequency exchange rate determination to a new level. It is quite plausible that the informational content of trades and public information are processed in a sophisticated, biological neural network-like model. In contrast to Lyons and Evans (2002), first, we apply ANNs (non-linear model), then, we use a larger amount of data (more than ten years of daily data) and, finally, we receive statistically significant forecasts. The inclusion of additional traditional and microstructure variables (to Lyons and Evans 2002) and ANNs results in out-of-sample forecasts superior to those from random walk and linear exchange rate models. This concept is pioneering in that we are able to establish a successful (and robust) non-linear ANN-based relationship between diverse traditional and microstructure variables and the Canada/U.S. exchange rate. Prior to this work, ANN exchange-rate models were built mostly on macroeconomic, technical trading or
autoregressive foundations (Gencay 1999, Kuan and Liu 1995, Plasmans et al. 1998) and the evidence from them was mixed. In chapter 3 we also compare ANN results to a number of other popular non-linear approaches and draw important conclusions from our empirical exercise. More specifically, we focus on nearest neighbours method (Diebold and Nason 1989), backpropagation ANN, recurrent ANN and the Takagi-Sugeno-Kang (TSK) fuzzy clustering model (Takagi and Sugeno 1985, Setnes et al. 1998) to estimate non-linear conditional-mean functions. Our exercise calls attention to non-linearities (ANNs) and order flows as important ingredients for exchange rate determination and very short-term forecast improvements. Thus, not only are strategic behavior and information asymmetries important, but, according to our findings, the very ability to understand and process that information as well.

In market microstructure theory, order flows are principal price determinants. Moreover, they are proxies for some of the short-run exchange rate volatility sources (technical trading, bandwagon effect, over-reaction to news, speculation, etc.). Having stated this, we do not preclude macroeconomic fundamentals from being the underlying determinant. Our findings strongly support the microstructure approach to exchange rates. Moreover, we point to important implications beyond those given by Lyons and Evans (2002) and Lyons (2001). What makes their dynamic, temporary equilibrium structure so easy to characterize is its linear structure. However, we find that market participants (i.e., dealers) are likely to pursue more complex or even mixed strategies. This brings us to the notion of a non-linear equilibrium where market makers create a non-linear relation between public information, microstructure effects and exchange
rates. One of the major strengths of this thesis is its strong empirical evidence that requires expansion of traditional microstructure linear characterization of equilibria as well as fundamental models of FX markets. We attempt to describe this new model only to a limited extent and more extensive research is required. In light of these findings, we challenge theoretical researchers to derive and explain this non-linear microstructure model more formally.

In chapter 5, we develop an original and novel approach to generating trading strategies in the foreign exchange (FX) market based on forecasts from the ANN (Gradojevic and Yang 2000). In recent literature that tackles the problem of generating profitable trading strategies, either the efficacy of technical analysis is tested (Lo and Wang 2000, LeBaron, Brock and Lakonishok 1992) or the search for improved technical trading models is conducted (Allen and Karjalainen 1999). We depart from recent research in this area in that our strategies contain macroeconomic and microstructure effects processed by the ANN. In addition to this, we employ a fuzzy logic approach to obtain a smoother decision surface. Previous studies were able to generate only buy, hold or sell trading signals while we model trading strategies as less discrete phenomena in such a form that recommends the fraction of our current endowment to be traded.

The fundamentals of fuzzy logic are presented in chapter 4. Fuzzy logic is based on natural language and represents reasoning related to realistic agent behaviour in the FX market. Our approach is motivated by existing evidence that implies human reasoning can be modeled as a fuzzy logic model (Smithson, 1987; Smithson and Oden, 1999). More precisely, we attempt to mimic the behavior of currency traders by
assigning them a set of continuous rules that explain the environment. This is more difficult to capture with conventional techniques. The idea is to extend the classical boolean logic by means of semantics that accounts for imprecision – the lack of sharply defined class membership criteria as opposed to the presence of random factors. This departs from the probability theory where the imprecision originates in the random nature of the process rather than in any vagueness of human decision-making. The advantage of this approach is that we eliminate the subjective views (i.e., precision errors) of the trading strategy given the exchange rate forecast.

Generally, we do not follow the neoclassical economic approach where economic agents are capable of logically inferring their price expectations and consequent trading strategy formation. In our view, real-world FX markets involve excessive market created uncertainty (Arthur 1994). This is a consequence of asymmetric expectations which force the agent to make subjective price predictions and corresponding strategies, given the subjective expectations of other agents. Fuzzy logic could provide an alternative form of reasoning to assist in agents' decision-making. We combine fuzzy reasoning with highly accurate ANN-generated forecasts into a dynamic neuro-fuzzy (NF) model. This enables us to investigate the evolution of NF-defined strategies over various time periods. The objectives of our this chapter are first to, model a fuzzy reasoning process as an extension to the ANN model, second, to investigate the NF model's implications for aggregate market behaviour (i.e., market efficiency) and, finally, to apply the NF model

---

2 See Tay and Linn 2001 who take similar assumptions.
to the problem of determining the FX market's sentiment, something any central bank needs to be able to assess.

The central part of our NF prediction and strategy-generating model is a rule-based deduction scheme in which agents (or more precisely traders) rely entirely on a fuzzy set of rules due to limitations in their ability to process public and private information. Our objectives inevitably force us to consider the notion of market efficiency. In its strong form the efficient market hypothesis (EMH) states that current prices reflect all known information (public and private) and no investor can consistently beat the market, i.e., earn extra profit (Fama 1965). A weak form of the EMH leads to the random walk hypothesis (RWH) which implies that on average today’s price is the best predictor of tomorrow’s price. In this thesis, we do not formally test the RWH or EMH by statistical means. Rather, we compare forecasts and corresponding strategies from our non-linear models to those generated by the random walk model (buy-and-hold strategies). In other words, we assume that the data have been generated by a non-linear DGP, possibly perturbed by additional noise.

Our contribution to the research in this area is threefold. First, we develop an innovative fuzzy decision-making system based on both public and private information. This gives us the ability to receive very precise strategy recommendations on not only how to trade but how much to trade as well. Therefore, our approach can also be seen as a natural extension to the classic portfolio selection model (Markowitz 1952) and applied to other decision-making problems. Having received the optimal portfolio weights from the Markowitz’s problem, the NF model can give more information about both asset-
specific and macroeconomic risks for particular assets. Second, our non-linear forecasts are significantly and consistently superior to the random walk forecasts. However, our daily trading strategies do not consistently make excess returns over simple buy-and-hold strategies. Further, when transaction costs are accounted for, the number of “inefficient” periods is roughly the same compared to when we neglect them. This is not an obvious indication supporting the EMH as there is a strong upward trend in the Canada/U.S. exchange rate testing series which clearly gives advantage to buy-and-hold trading strategies. In addition, the NF-based approach may incur transaction costs on a daily basis while in the case of buy-and-hold strategy it is applied only once (at the beginning of the buy-and-hold period). The findings support our approach to move from discrete (buy/hold/sell) market decisions to less discrete and more precise “fuzzy logic-based” recommendations. Finally, we successfully apply the NF system to assess the FX market’s sentiment, as reflected by the chartists’ trading signals, during periods of strong depreciations. We also find that the use of the continuous non-linear function that mimics the fuzzy logic decision-making can benefit a risk-averse trader.
Chapter 2

Artificial Neural Networks (ANNs) and Other Non-linear Methodologies

2.1 ANNs

2.1.1 Definition and structure

ANNs represent a general class of non-linear models that have been successfully applied to a variety of problems such as medical diagnostics, product selection, system control, pattern recognition, functional synthesis, and forecasting (e.g., econometrics). The most important financial applications of ANNs include option pricing, bankruptcy prediction, stock market prediction, and exchange rate forecasting.\(^3\)

\(^3\) More extensive review of empirical studies involving ANNs can be found in Qi (1996).
ANNs are composed of simple computational elements or nodes (Lippmann 1987). Figure 2.1 provides the simplest node, which sums N weighted inputs and conveys the outcome either to other nodes or out of the ANN. The node is characterized by an internal threshold or offset θ (or bias) and by its type of specified non-linearity. Figure 2.1 illustrates three common types of non-linearities used in ANNs: hard limiters, threshold logic elements, and sigmoid. More complex nodes might even include integration or other mathematical operations.

Neural network models differ in topology, node characteristics, and training or learning rules. These rules fix the initial set of weights and indicate how weights should be altered and adjusted during use to improve performance.
There are several ways to structure the neural networks. Typically, the elements are arranged in groups or layers. Fewer layers limit these networks when modeling a functional representation of data, a typical econometrics problem.

However, the development of learning algorithms has made it feasible for multi-layered networks. They are ideal for functional form determination and they are normally structured as three-layered networks. Figure 2.2 depicts such a network with four inputs, three nodes in the hidden layer, and one output.

The three layers are as follows:
• Input layer: The neural network receives its data in the input layer. The number of nodes (i.e., neurons) in this layer depends on the number of inputs to a model and each input requires one neuron. For example, in functional synthesis (this thesis’ scope of study), inputs are exogenous variables—that is, observations of interest.

• Hidden layer: The hidden layer lies between the input and output layers; there can be many hidden layers. They are analogous to the brain’s interneurons, a place where the hidden correlations of the input and output data are captured. This allows the network to learn, adjust, and generalize from the previously learned facts (e.g., data sets) to the new input. As each input-output set is presented to the network, the internal mapping is recorded in the hidden layer. Unlike any other classical statistical methodology, this gives the system intuitive predictability and intelligence. The number of hidden layers is determined by a trade-off between network intuitive ability and efficiency. A priori, the optimal number of hidden layers is not clear. With too many hidden layers, an overcorrecting problem arises: a network is “overtrained” or “overfitted,” which prevents it from learning a general solution. On the other hand, too few layers will inhibit the learning of the input-output pattern. Typically, the number of hidden layers and nodes inside the network is determined through experimentation, and this thesis follows that technique.
• Output layer: Having been trained, the network responds to new input by producing an output that represents a forecast. During training, the network collects the in-sample output values in the output layer.

Among non-linear methods, ANNs are one of the most recent techniques used in non-linear modeling. This is partly due to some modeling problems encountered in the early stage of development within the field of ANNs. In the earlier literature, the statistical properties of ANN estimators and their approximation capabilities were questionable. For example, there was no guidance in terms of how to choose the number of neurons and their configurations in a given layer and how to decide on the number of hidden layers in a given ANN. Recent developments in the ANN literature, however, have provided the theoretical foundations for the universality of feedforward ANNs as function approximators. The results in Cybenko (1989), Funahashi (1989), Hornik et al. (1989, 1990), and Hornik (1991) indicate that feedforward ANNs with sufficiently many hidden units and properly adjusted parameters can approximate an arbitrary function arbitrarily well. Hornik et al. (1990) and Hornik (1991) further show that the feedforward ANN can also approximate the derivatives of an arbitrary function.

forecasted the UK/DM exchange rate based on the exchange rates of several other currencies.

2.1.2 Learning and adapting

Neural networks develop an inner structure to solve problems. Through the training process, connection weights rearrange their values and reveal a data pattern. Thus, the fundamental feature of neural networks is that they are trained, not programmed.

A neural network learns with each new datum (an input-output combination) introduced into it during training. Every processing element responds to its input, adjusting its behavior. The network calculates the output in accordance with the elements’ transfer function and learns by adjusting the input weights and biases. The equation that explains this change is called the learning law.

There are two different learning modes: supervised and unsupervised.

The supervised learning mode presents input-output data combinations to the network. Consequently, the connection weights and node biases, initially randomly distributed, adjust their values to produce output that is as close as possible to the actual output, i.e., the learning method (recursive algorithm) tries to minimize the current errors of all processing elements. With each subsequent cycle the overall network error between the desired and the actual output will be lower. Eventually, the result is a minimized error between the network and actual output (or desired network accuracy), as well as the internal network structure, which represents the general input-output dependence. In a one-layered network, it is easy to control each individual neuron and observe the input-output pattern. In multi-layered neural networks, supervised learning becomes difficult. It is harder to monitor and correct neurons in hidden layers.
Supervised learning is frequently used for network decision, memorization, and generalization problems.

The unsupervised learning mode is independent of the external influences to adjust weights and biases. There are no concrete ANN output data to correct its pattern identification. In this mode, the ANN is able to recognize statistical regularities in its input space, automatically categorize them and develop behavior for different classes of inputs. Unsupervised learning can be used for classification, marketing and anomaly detection.

2.1.3 Backpropagation ANN

The backpropagation ANN, one of two network types applied in this thesis, is probably the most commonly used ANN. It is characterized by hidden layers and the generalized Delta rule for learning: If there is a difference between the actual and the desired output pattern during training, then the connection weights and biases, denoted by vector \( \lambda \), have to be recursively adjusted to minimize the mean-squared error (Qi 1996, Kuan and White 1994, Van Eyden 1996). Mathematically, \( \lambda \) is chosen to minimize the following loss function:
\[
\min_{\lambda} \left[ \frac{1}{M} \sum_{i=1}^{M} (y_t - \hat{y}_t)^2 \right]
\]

where \( M \) is the size of the training sample, \( y_t \) is sample output and \( \hat{y}_t \) is the estimated output value. We will show how to estimate \( \hat{y}_t \) in a three-layered backpropagation ANN (Figure 2.3).

Suppose that ANN is composed of \( p \) input and \( q \) hidden nodes whereas signals at each layer are denoted by \( x_i \) (\( i=1,...,p \)) and \( h_j \) (\( j=1,...,q \)), respectively. The layers are characterized by two arbitrary types of non-linearities: \( \psi \) and \( \sigma \). Backpropagation
learning algorithm requires continuous differentiable non-linearities and the most commonly used type is the sigmoid logistic (or logsig) function:

\[ f(w) = \frac{1}{1 + e^{-w}} \]

As specified before, nodes in the hidden layer produce \( h_j \) signals from the inputs while a single output node transforms them into \( y_t \). Specifically,

\[ h_{jt} = \psi \left( \alpha_{j0} + \sum_{i=1}^{p} \alpha_{ji} x_{it} \right), j=1,...,q \]

\[ y_t = \sigma \left( \beta_0 + \sum_{j=1}^{q} \beta_j h_{jt} \right) = \sigma \left( \beta_0 + \sum_{j=1}^{q} \beta_j \psi \left( \alpha_{j0} + \sum_{i=1}^{p} \alpha_{ji} x_{it} \right) \right) \]

where \( \alpha_{ij} \) and \( \beta_j \) denote appropriate connection weights between the adjacent layers. Subscripts 0 for \( \alpha \) and \( \beta \) stand for ANN biases.

Studies by Cybenko (1989), Hornik et al. (1989) and Funahashi (1989) show that the above non-linear representation with sigmoid non-linearities can approximate a large number of mappings between inputs and outputs reasonably well. This makes ANN a very useful non-parametric technique and there is no need for any unjustified restrictions often present in econometric modeling.

---

4 Other types of transfer functions used in this research are tan-sigmoid (\( f(w) = \frac{e^w - e^{-w}}{e^w + e^{-w}} \)) and purelin (\( f(w) = w \)).
Delta rule recursive algorithm readjusts connection weights and biases starting at the output nodes and working back to the first hidden layer (Lippmann 1987). Connection weights are adjusted by

$$\gamma_j(t+1) = \gamma_j(t) + \eta \delta_j x'_i$$

where:
- $\gamma_j(t)$ = the weight from hidden node $i$ or from an input to node $j$ at time $t$.
- $x'_i$ = either the output of node $i$ or an input.
- $\eta$ = the positive constant that measures the speed of the convergence of the weight vector (gain term or learning rate).
- $\delta_j$ = an error term for node $j$ ($k$ goes over all nodes in the layers before node $j$):

$$\delta_j = \begin{cases} y_j (1 - y_j)(d_j - y_j), & \text{when } j \text{ is an output node} \\ x'_j (1 - x'_j) \sum_k \delta_k \gamma_{jk}, & \text{when } j \text{ is a hidden node} \end{cases}$$

where:
- $d_j$ = the desired output of node $j$.
- $y_j$ = the actual output of node $j$.

The node biases are adapted through the similar iterative algorithm by assuming they are connection weights on links from constant-valued inputs.
At least three layers are required: input, hidden, and output. The hidden layer is very important, since it enables the ANN to extract patterns and to generalize. Even though a hidden layer should be large, one must be careful not to deprive the network of its generalizing ability when the network starts memorizing, rather than deducing. On the other hand, a hidden layer that is too small could reduce the accuracy of recall. The connections are only feedforwarded between the adjacent layers. For the transfer function, the backpropagation ANN employs the sigmoid function (mentioned above).

The learning algorithm that controls backpropagation follows a number of steps:

1) Initialization: Initialize connection weights and neurons to small random values.
2) Data introduction: Introduce the continuous data set of inputs and actual outputs to the backpropagation network.
3) Calculation of outputs: Calculate the outputs, and adjust the connection weights several times applying the current network error.
4) Adjustments: Adjust the activation thresholds and weights of the neurons in the hidden layer according to the Delta rule.
5) Repeat steps (2) to (4) to until a desired network accuracy (error) is reached.

Backpropagation estimation techniques are important for large samples and real-time applications since they allow for adaptive estimation. However, they may not fully utilize the information in the data. White (1989) showed that the recursive estimator is not as efficient as the non-linear least squares estimator. An important aspect of the
backpropagation methods is the choice of the learning rate \( \eta \). The inefficiency of the backpropagation originates in keeping the learning rate constant in an environment where the influence of random movements in inputs are not accounted by the target. This would lead the parameter vector to fluctuate indefinitely, i.e., there would be no convergence. A minimum requirement is to gradually drive the learning rate to zero to achieve convergence. In fact, White (1989) demonstrated that \( \eta \) has to be chosen not as vanishing scalar, but as a gradually vanishing matrix of a very specific form. These arguments on learning rates are only valid if the environment is stationary, which is the case in this work (as well as the constant learning rate). However, instead of tuning the network’s learning rate, we attempt to balance the bias and variance and avoid non-convergence with early stopping. This may slightly deteriorate the model’s performance and we acknowledge that probably by following White (1989) an optimal estimator would be achieved.

### 2.1.4 Recurrent ANN

In feedforward neural nets such as backpropagation ANNs there is no feedback between any two of its layers. In line with dynamic modeling, it would be useful to both detect and generate time-varying patterns. Backpropagation ANNs can account for lagged variables by simply including them into the set of inputs. However, this requires knowledge of a number of lags to be included. A sophisticated way to circumvent this
inconvenience is to allow for feedback connections within the ANN. Elman recurrent networks (Elman 1990) are backpropagation ANNs with the addition of a feedback connection from the output of the hidden layer to its input (Figure 2.4). Note that the Elman ANN has tan-sigmoid non-linearities (very similar to sigmoid logistic function, but can generate outputs between -1 and 1) in hidden neurons and linear output layer neurons.

The output of a recurrent ANN is calculated in a slightly different manner:

\[ h_{jl} = \psi \left( \alpha_{j0} + \sum_{i=1}^{p} \alpha_{ji} x_{il} + \sum_{k=1}^{q} \rho_{kj} h_{kl(t-1)} \right), \quad j=1,\ldots,q \]
\[
y_t = \sigma \left( \beta_0 + \sum_{j=1}^{q} \beta_j h_{j,t} \right) = \sigma \left( \beta_0 + \sum_{j=1}^{q} \beta_j \sigma \left( \alpha_{j0} + \sum_{i=1}^{p} \alpha_{ji} x_{it} + \sum_{k=1}^{q} \rho_{kj} h_{k(t-1)} \right) \right)
\]

where, in addition to the previous ANN type notation, \(\rho_{kj}\) denotes feedback connection weights (\(k=1,\ldots,q\)) that are inputted back into the \(j\)-th hidden neuron, and \(h_{k(t-1)}\) are lagged outputs from each of \(k\) hidden layer neurons.

By recursively substituting for \(h_{k(t-1)}\) into the last expression for \(h_j\), \(h_j\) can be expressed depending on the entire history of \(x_{it}\). Further, it follows that \(y_t\) is a function of \(x_{it}\) and its past lags. This “moving average” component is a crucial difference between feedforward and recurrent ANNs which can thus capture more of the dynamics of \(y_t\).

As in the case of the backpropagation ANNs, learning and adapting of the Elman ANN is performed with the generalized Delta rule. Naturally, due to the presence of the feedback connection, the iterative estimation becomes more complicated and the learning time for recurrent ANNs is typically much longer than that in backpropagation ANNs.

### 2.2 NN method

The NN method is based on an assumption that geometric patterns in the past of the time series, similar by some measure to the currently observed variables, can be used for forecasting the dependent variable (see, e.g., Yakowitz 1987, Cleveland 1979, Cleveland
and Devlin 1988). Specifically, we use locally-weighted regression (LWR) to estimate one-step-ahead exchange rate change from a weighting scheme in which weights are functions of the Euclidean distances. This approach is a philosophical departure from the Box-Jenkins methodology where the forecasts are extracted from lagged observations and error terms (or external variables as in ARMAX model), rather than from the set of ‘related’ observations of independent variables - nearest neighbours.

Suppose we are interested in forecasting $y_t$ one-period-ahead and $x_t$ is a $(m \times 1)$ vector of explanatory variables. Then, the non-linear model is

$$y_t = \phi(x_{t-1}) + \varepsilon_t, \quad E(\varepsilon_t | x_{t-1}) = 0; \quad t = 1, \ldots, T$$

where $\phi(.)$ is an arbitrary, but fixed non-linear function.

LWR estimates $\phi(x^*)$, i.e., the estimate of $\phi$ at the specific value $x = x^*$. This is performed in several steps as follows:

1. Let $y^* = \phi(x^*)$ denote our forecast of interest and $x^*$ the point where we estimate $\phi(.)$. We then organize the $x_{t-1}$ ($t = 1, \ldots, T$) series into $h$-histories defined by:

$$x_{t-1}^h = (x_{t-h+1}, \ldots, x_{t-1})$$
The parameter $h$ is also called the *embedding dimension*.

2. The next step is to choose $k$ $h$-histories "closest" to $x^*$ called the nearest neighbours ($x_k$). Let $\zeta$ be a smoothing constant such that $0<\zeta \leq 1$, and let $k=\text{int} \left( T\zeta \right)$, where $\text{int} \left( . \right)$ extracts the integer part of its argument. The LWR uses $k$ observations nearest to $x^*$, where the proximity is measured with the most commonly used distance measure, Euclidean distance:

$$d(x^*, x_k) = \left[ \sum_{i=1}^{m} (x^*_i - x_k) \right]^{\frac{1}{2}}$$

According to the Euclidean distance, $x_{t-1}^h$ are ranked and assigned specific weights. Observations not considered to be $x^*$'s nearest neighbours are assigned a weight of zero.

3. In this step, we follow Diebold and Nason (1990) to construct a tricube weighting function:

$$w_i = \left( 1 - \left( \frac{d(x_t, x^*)}{d(x^*, x_k)} \right)^3 \right)^3$$

4. Our forecast is simply computed as:

$$y^* = \phi(x^*) = x^* \hat{\beta},$$

where
In this thesis we decide to select the Euclidean distance measure as the one most commonly used in the literature. Note that other distance measures could be employed rather than the Euclidian norm, such as Mean Absolute Deviation or even Range, but we leave them to further research.

Further, the tricube weighting scheme is smooth in applying more confidence to the first nearest neighbour than the second nearest, and more to the second than to the third, and so forth. Thus, the weights are set inversely proportional to the Euclidean distances. Various weighting schemes other than Euclidean are presented in Robinson (1987). Stone (1977) proved the consistency (in the statistical sense) of NN estimators for different weighting schemes. Consistency of NN estimators (and therefore LWR) requires that the number of employed nearest neighbours goes to infinity with sample size, but at a slower rate, that is, as $T \to \infty$, $k \to \infty$, but $(k/T) \to 0$. Clearly, consistency depends on our choice of $\zeta$. As $\zeta$ increases, the bias in $\phi(x^*)$ increases and the sampling variability decreases. Thus, one needs to choose $\zeta$ to balance the trade-off between bias and variance.

We find the optimal $\zeta$ by in-sample cross-validation. Once determined, $\zeta$ is fixed for out-of-sample estimation.
2.3 TSK method

2.3.1 Introduction

The fuzzy model is based on a learning by examples approach; by establishing relations between the relevant variables in the form of if-then rules. One of the aspects that distinguish fuzzy modeling from black-box techniques like ANNs is that fuzzy models are to a certain degree transparent to interpretation and analysis. We present a data-driven Takagi-Sugeno-Kang (TSK) fuzzy rule-based model (Takagi and Sugeno 1985, Setnes et al. 1998). It requires only a small number of rules to describe highly complex non-linear models. The basic idea of the TSK model is the fact that an arbitrarily complex model is a combination of mutually interlinked sub-models. If $K$ regions, corresponding to individual sub-models, are determined in the state-space under consideration, then the behaviour of the system in these regions can be described with the simpler functional forms. If the sub-model is assumed to be linear and one rule is assigned to each sub-model, the TSK fuzzy model can be represented with $K$ rules of the following form:

\[ R_i: \text{If } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \text{ and } \ldots \text{ and } x_n \text{ is } A_{in} \text{ then } y_i = a_i x + b_i, \quad i = 1, 2, 3, \ldots, K \]  

(2.3.1)
where \( R_i \) is the \( i \)-th rule, \( x_1, x_2, \ldots, x_n \) are the input variables, \( A_1, A_2, \ldots, A_m \) are the fuzzy sets assigned to corresponding input variable, variable \( y_i \) represents the value of the \( i \)-th rule output, and \( a_i \) and \( b_i \) are parameters of the consequent function. The final output of the TSK fuzzy model for arbitrary \( x_k \) input can be calculated as a weighted average of the rule consequents:

\[
\hat{y}_k = \frac{\sum_{i=1}^{K} [\beta_i(x_k) \cdot y_i(x_k)]}{\sum_{i=1}^{K} \beta_i(x_k)} \quad k = 1, 2, 3, \ldots, N. \tag{2.3.2}
\]

where \( \beta_i \) is the weight assigned to the \( i \)-th rule.

Equation (2.3.2) can be written in a form that shows more explicitly the concept of the system description using local models. This expression then takes the form

\[
\hat{y}_k = \sum_{i=1}^{K} [w_{ik}(x_k)(x_ia_i + b_i)] \quad k = 1, 2, \ldots, N, \tag{2.3.3}
\]

where \( w_{ik} \) represents the normalized weighting factor (or firing strength) of the \( i \)-th rule for the \( k \)-th sample. Equation (2.3.4) shows that (2.3.2) and (2.3.3) are identical for the final output:

\[
w_{ik} = \frac{\beta_i(x_k)}{\sum_{j=1}^{K} \beta_j(x_k)} \quad k = 1, 2, 3, \ldots, N. \tag{2.3.4}
\]
The TSK fuzzy rule based model, as a set of local models, enables application of the linear least-squares (LS) method (Pedrycz et al. 1997). LS algorithm requires a model that is linear in parameters. However, that does not imply that the model itself must be linear.

We propose an extremely fast and accurate algorithm for generating the fuzzy model that is applicable for a variety of problems. The TSK algorithm combines clustering of the numerical input data and LS method. The first step is to identify the rule base structure. This step involves grouping of the input space using fuzzy clustering. Procedures for selection of the input variables and optimal number of rules are not discussed (these issues are covered in considerable depth in, for example, Sugeno and Yasukawa 1993). The next step consists of identification of parameters of the consequent part of the rule base. Keeping the conditional part of the rule fixed, parameters $a_i$ and $b_i$ are determined using the LS method on the global level. Globally formulated LS yields the minimal parameter error estimate of the global non-linear model.

2.3.2 TSK fuzzy model building

Step I - Structure identification

Partitioning of the input space is performed in the first step, using a selected clustering method which utilizes the training (learning) data: $N$ observations of the exogenous variables, i.e., inputs $x_k$, $k=1,\ldots,N$. Let the input vector be of dimension $n$. Each
obtained cluster represents a certain operating region of the system, where input data values are highly concentrated. The learning data, divided into these information clusters, are then interpreted as rules. Methods of fuzzy clustering, such as fuzzy c-means (FCM) and Gustafson-Kessel (GK), are convenient tools for partitioning of the input or input-output space (Forsythe et al. 1977, Jang et al. 1997).

A central part of the fuzzy clustering algorithm consists of calculation of the distance between the $k$-th sample, currently under consideration, and the centre of a given $i$-th cluster. In the case of the FCM algorithm, this distance is calculated using the Euclidean distance, which is in the $l$-th iteration of the following form:

$$d_{ik}^l = \sqrt{\sum_{j=1}^{n}(x_{jk} - v_{ij}^{(l)})^2} \quad i = 1, 2, ..., K; \quad k = 1, 2, ..., N \quad (2.3.5)$$

The GK algorithm evaluates this distance using the transformed Euclidian distance. Hence, in the $l$-th iteration of the GK algorithm, the distance is given with:

$$d_{ik}^l = [x_k - v_i^{(l)}]^T \left\{ \left[ \text{det}(F_i)^{1/2} \right] \left[ (F_i)^{1/2} \right] \right\} \left[ x_k - v_i^{(l)} \right], \quad i = 1, 2, ..., K; \quad k = 1, 2, ..., N \quad (2.3.6)$$

where $F_i$ is the covariance matrix for the $i$-th cluster and $v_i$ is a centre of the $i$-th cluster. The GK algorithm is of a more general nature, since it forms hyper-ellipsoid clusters in the $n$-dimensional input space (those created by the FCM method are hyper-
spherical). However, the FCM method is less numerically complex and therefore faster. For these reasons in this thesis we use the FCM method.

The result of the clustering process is, apart from the matrix of cluster centers, a fuzzy partition matrix $U_{K \times N}$. This matrix contains values $\mu_{ik} \in [0,1]$, for $1 \leq i \leq K$ and $1 \leq k \leq N$, that represent the degree of membership of the $k$-th sample to the $i$-th cluster. Elements of the matrix $U$ are calculated using the following formula:

$$
\mu_{ik}^{(t)} = \frac{1}{\sum_{j=1}^{K} \left( \frac{d_{ik}}{d_{jk}} \right)^{(m-1)}} \text{, for } d_{ik} > 0 \text{, or } \mu_{ik}^{(t)} = 0 \text{ for } d_{ik} = 0.
$$

(2.3.7)

The columns of the matrix $U$ must satisfy the following condition:

$$
\sum_{i=1}^{K} \mu_{ik} = 1, \quad k = 1,2,3,...,N.
$$

(2.3.8)

This condition corresponds to the condition of orthogonality of input vector fuzzy set.

**Step II - Identification of parameters of the consequent part of the rule**

The parameter identification step encompasses procedure for the calculation of parameters of the TSK fuzzy model: the parameters of the linear regressions in the consequent part of the rule. The crucial problem is to determine the normalized firing
strength of the rules, or, more precisely, to estimate the contribution of each rule to the final model output. First, it is necessary to determine one-dimensional fuzzy sets $A_{ij}$ from the conditional part of the rule. These fuzzy sets, denoted by $A_{ij}$, are obtained from the multidimensional fuzzy clusters (given by $U$) by projection onto the space of the input variable $x_i$.

$$\mu_{A_{ij}}(x_{kj}) = \text{proj}_j(\mu_{ik}) \quad (2.3.9)$$

where $\mu_{ik}$ is the level of belonging of the $k$-th observation (from the vector $x_k$) to the $i$-th cluster, while $\mu_{A_{ij}}(x_{kj})$ is the value of the level of belonging of the $j$-th input variable to the $k$-th observation ($j$-th coordinate of the vector $x_k$) of the fuzzy set $A_{ij}$. Since all the functions $\mu_{A_{ij}}(x_{kj})$ under consideration represent membership functions, the condition $\mu_{A_{ij}}(x_{kj}) : \mathbb{R} \rightarrow [0,1]$ must hold true. The value of the firing strength of the $i$-th rule for each $k$-th input sample is calculated as and-conjunction by means of the product operator:

$$\beta_{ik} = \beta_i(x_k) = \prod_{j=1}^{n} \mu_{A_{ij}}(x_{kj}), \quad k = 1,2,3,...,N \quad (2.3.10)$$

Finally, normalized firing strength $w_{ik}$ of a rule $R_i$ for the $k$-th sample is obtained using the equation (2.3.4).
The $k$-th element on the main diagonal of the $N \times N$ dimensional diagonal matrix $W_i (i=1,...,K)$ is formed using the values of the normalized firing strengths from (2.3.4). Thus, $N \times K(n+1)$ composite matrix $X'$ can be formed:

$$X' = [(W_1X_e), (W_2X_e), (W_3X_e), ..., (W_KX_e)]$$ (2.3.11)

where matrix $X_e = [X, 1]$ contains rows $[x^T_i, 1]$. Finally, the LS parameters $a_i^T$ and $b_i$ of (2.3.1) and (2.3.3), that are in the consequent part of all TSK fuzzy rules, can be grouped into a $K(n+1) \times 1$ vector $\theta'$. These parameters are organized in the vector $\theta'$ as follows:

$$\theta' = [\theta_1^T, \theta_2^T, ... \theta_K^T]^T$$ (2.3.12)

where $\theta_i^T = [a_i^T; b_i]$, for $i=1,...,K$. Problem described in (2.3.2) and (2.3.3) can now be rewritten as the LS regression of the form $Y = X\theta' + \epsilon$, where $\epsilon$ is an error term. $\theta'$ can be determined using the LS method:

$$\theta' = [(X')^T X']^{-1} (X')^T Y$$

As in the case of the NN method, we find the optimal number of clusters ($K$) by in-sample cross-validation. Then we estimate $\theta'$ and hold it fixed for out-of-sample estimation.
Chapter 3

Exchange Rate Forecasting: A Market Microstructure Approach

3.1 Models of Exchange Rate Determination

Various models aiming at explaining exchange rate fluctuations have been proposed. Meese and Rogoff (1983) found that a simple random walk model performed no worse than any of competing representative time series and structural exchange rate models. Out-of-sample forecasting power in those models was surprisingly low for various forecasting horizons (from 1 to 12 months).

Subsequent attempts to determine exchange rates shed very little light on the problem. Baillie and McMahon (1989) pointed out that exchange rates are in general not linearly predictable. They are described as highly volatile with an elusive data-generating process (DGP). Similarly, Hsieh (1988), Boothe and Glassman (1988) and Diebold and Nerlove (1989) observed that exchange rate changes are leptocurtic and may be non-linearly dependent. Further, the observed exchange rates seem to exhibit volatility clustering, i.e., high (low) volatility periods tend to be followed by high (low) volatile
periods. This conditional heteroskedasticity evidence was reported in Diebold (1988), Engle (1982), Hsieh (1989) and Engle et al. (1990). Nevertheless, excess kurtosis and conditional heteroskedasticity in the residuals may not improve point forecasts because these effects operate through even-ordered moments. To model the observed effects, parametric non-linear models such as ARCH (Hsieh 1989) and GARCH (Bollerslev 1990) were applied to exchange rates modeling, but with very little success. Gencay (1999) examined the predictability of spot foreign exchange rate returns using moving average technical trading rules and GARCH, nearest neighbours (NN) and artificial neural networks (ANN) models. GARCH models generated insignificant sign forecasts improvement (and less than 1 per cent mean-squared error improvement) over a simple random walk. As noted in Diebold and Nason (1989), the pre-specification of the GARCH model form may neglect other possible non-linearities resulting from a true DGP. Meese and Rose (1991) examined macroeconomic exchange rate models and found that poor explanatory power of the models cannot be attributed to non-linearities. They considered five non-linear structural exchange rate models in order to capture possible non-linearities. The application of several parametric and non-parametric techniques on these fundamentally-driven models did not show any improvement in our ability to understand the exchange rate fluctuations.

Meese and Rose (1990), Diebold and Nason (1989) and Mizrach (1992) used a NN non-parametric estimator called locally-weighted regression (LWR) in order to handle non-linearities, but they could not significantly improve upon a simple random walk in the out-of-sample exchange rate predictions. In contrast, using an NN method,
Gencay (1999) and Lisi and Medio (1997) were able to generate predictions superior to those generated by the random walk model. This mixed evidence could suggest an existence of non-linear patterns in the exchange rates which, if revealed, could be exploited to improve both point and sign predictions.

Kuan and Liu (1995) used backpropagation and recurrent artificial neural networks (ANN), a very powerful tool for detecting non-linear patterns, to investigate the ANN's out-of-sample forecasting ability on five exchange rates (British pound, Canadian dollar, Deutsche mark, Japanese yen, and Swiss franc) against the US dollar. The data were daily opening bid prices of the NY Foreign Exchange Market and the model of interest was non-linear AR where its performance against random walk and ARMA processes was measured. Their results showed the presence of non-linearities in exchange rates time series. For the Japanese yen and British pound, ANNs exhibited significant sign predictions and/or significantly lower out-of-sample MSPE (relative to the random walk model); for the remaining three currencies ANNs had inferior forecasting performance. Some other studies involving ANNs were less encouraging. Plasmans, Verkooijen and Daniels (1998) and Verkooijen (1996) used macroeconomic models, but they could not produce any satisfactory monthly forecasts. However, Zhang and Hu (1998) modeled exchange rate as depending non-linearly on its past values, and their model outperformed simple linear models, but they never compared it to a random walk. Hu et al. (1999) showed (using daily and weekly data) that ANNs are a more robust forecasting method than a random walk model. Hence, the application of ANNs to short-term currency behavior was successful in numerous cases and the results suggest
that ANN models may have some advantages when frequent short-term forecasts are required (Evans 1997, Jamal and Sundar 1997, Gencay 1999).\(^5\)

All the above-mentioned approaches try to find the exchange rate determinants among macroeconomic variables such as interest rates, money supplies, inflation rates, and trade balances. Flood and Rose (1995) concluded that exchange rate modeling based only on macroeconomic fundamentals might be insufficient to explain the exchange rate volatility. Recently, Cheung and Wong (2000) conducted a survey of practitioners in the interbank foreign exchange markets in Hong Kong, Tokyo, and Singapore. A majority of participants view short-term exchange rate variability closely related to non-economic forces including bandwagon effects, over-reaction to news, speculation, and technical trading. Only 1 per cent of the traders look at economic fundamentals to determine daily exchange rate movements.

Given the partial empirical success of the macroeconomic models, there is an increasing interest in the exchange rate microstructure. The microstructure approach investigates how specific trading mechanisms affect the exchange rate formation. Lyons and Evans (2002) incorporated a variable reflecting the microeconomics of asset pricing into a model of the exchange rate. They introduced the most important microstructure variable, "order flow" as the proximate determinant of the exchange rate (using daily data over a four-month period) and were able to significantly improve on existing

---

\(^5\) In this context, ANNs focus on daily (weekly) or less-than-a monthly forecasting frequency, while typical macroeconomic models are at a monthly or quarterly frequency.
macroeconomic models. More precisely, they managed to capture about 60% of the exchange rate daily changes using a linear model.

In general, there are two broad theories of exchange rate modeling: traditional macroeconomic models and the more recently developed market microstructure models. Macroeconomic models aim at modeling and estimating exchange rates at monthly or lower frequency. These models are in general of the following form:

$$\Delta \text{rpfx}_t = \phi(M_t) + \epsilon_t, \; t=1,..,N.$$  

where $\Delta \text{rpfx}_t$ is the change in the logarithm of the real exchange rate over the month or some lower frequency of observations, and $M_t$ is a vector of typical macroeconomic variables such as the difference between home and foreign nominal interest rates, the long-run expected inflation differential, and relative real growth rates.\(^6\) To control for the key Canadian macroeconomic variables, this thesis uses a variation of the model developed by Amano and van Norden (1995):

$$\Delta \text{rpfx}_t = \varphi(\text{rpfx}_t, \text{com}_t, \text{ene}_t, \text{intdiff}_t) + \delta_t, \; t=1,..,N.$$  

where $\text{rpfx}_t$ is real Canada/U.S. exchange rate deflated by GDP deflators, $\text{com}_t$ is the logarithm of non-energy commodity price index (deflated by the U.S. GDP deflator), $\text{ene}_t$ is the logarithm of energy commodity price index (deflated by the U.S. GDP

deflator) and intdiff represents the nominal 90-day commercial paper interest rate differential (Canada-U.S.).

Macroeconomic models provide no role for any "market microstructure" effects to directly enter into the estimated equation which are thus incorporated through the error term $\delta_t$. These models assume that markets are efficient in the sense that information is widely available to all market participants and all relevant and ascertainable information is already reflected in exchange rates. In other words, from this point of view, exchange rate changes are not informed by microstructure variables. However, typical macroeconomic models perform poorly. Moreover, empirical evidence from Lyons and Evans (2002), Yao (1997), Covrig and Melvin (1998), and this thesis suggests that a microstructure variable 'order flow' contains information relevant to exchange rate determination.

For the spot foreign exchange (FX) trader, what really matters is not the data on any of the macroeconomic fundamentals, but information about demand for currencies extracted from purchases and sales orders, or order flow (Cheung and Wong 2000). Any short-term exchange rate determinant such as portfolio shifts, over-reaction to news, or speculative trading would be recorded in order flow. It is presumed that certain FX

---

7. Microstructure literature examines the elements of the security trading process: the arrival and dissemination of information; the generation and arrival of orders; and the market architecture which determines how orders are transformed into trades. Prices are discovered in the marketplace by the interaction of market design and participant behaviour.
traders observe trades that are not observable to all the other traders and, in turn, the market efficiency assumption is violated at least in the very short term.\footnote{It may be that markets, absent these market microstructure frictions would be efficient, but trading frictions impede the instantaneous embodiment of all information into prices.}

Microstructure models directly rely on information regarding the order flow. Lyons and Evans (2002) approach an order flow/exchange rate relation through a very realistic framework - portfolio shifts model. Their model is a three-stage game between the dealers and the public (see Figure 3.1). In the first stage, non-dealer market participants (corporations, mutual and pension funds, etc.) analyze all the publicly available information, including macroeconomic fundamentals, and then decide on orders (dealer i's quote is $P_i^1$, $i=1,...,M$). Having observed their order flows (which thus reflect information about macroeconomic fundamentals), dealers re-set their price to $P_i^2$. The second stage is the interdealer trading (sharing the inventory risk) where each dealer simultaneously trades on other dealers' quotes. Aggregate interdealer trades ($\Sigma D_i$) are publicly observable while customer-dealer trades are not (the same is assumed for other order flows such as foreign institution transactions). The interdealer flows inform dealers about the total size of the currency stock (inventory imbalances) the public needs to absorb to achieve equilibrium. In the third stage, dealers simultaneously and independently quote a new price ($P_{i3}$) so that the public's inventory is in equilibrium and dealers end the day with no net position.
Lyons and Evans use the perfect Bayesian-Nash Equilibrium (BNE) concept where dealers choose quotes \( P_{t1}, P_{t2}, \) and \( P_{t3}, \) and the dealer's interdealer trade \( D_{t2}. \) They explicitly derive an equilibrium price change (between period \( t-1 \) and \( t \)) and equilibrium trading strategies. Intuitively, equilibrium price is determined from the stage 1 common information set (macroeconomic fundamentals, denoted \( r_t \)) and aggregate interdealer order flow (IB):\
\[
\Delta P_t = r_t + \lambda \text{IB}_t
\]
where \( \lambda \) is a positive constant.

This model was estimated over a four-month span of daily observations controlling for a key macroeconomic variable, interest rate differential. The results were in favour of microstructure approach with \( R^2 \) statistic over 50 per cent.

---

9 The sum of each dealer's interdealer order is an interdealer order flow, or interbank transactions (IB), as denoted later in the text.
However, there are several unanswered questions which require further research. First, can we restrict our model to only one macroeconomic determinant? In this research, we try to control (among other candidates available on daily basis) for another fundamental variable: crude oil price. Secondly, is there a non-linear conditional mean function that characterizes a true DGP? By employing several non-linear methods and relaxing the three-stage game restrictions, we attempt to find another equilibrium, non-linear by its structure.\textsuperscript{10} Indeed, in this new setting there is a possibility that other types of order flow might play a role in setting the price. Finally, can this model be successfully used for out-of-sample forecasting?\textsuperscript{11} In this thesis, we address all of these issues and provide adequate answers based solely on empirical grounds.

In general, the market microstructure approach assumes the following relationship between the exchange rate and the driving variables:

$$\Delta \text{rpf}_{t} = \psi (\Delta x_{t}, \Delta I_{t}, N_{t}) + \chi_{t}, \text{t}=1,...,N.$$  

where $\Delta x_{t}$ represents order flow, $\Delta I_{t}$ is a change in net dealer positions, while $N_{t}$ is any other microeconomic variable. Order flow can be positive (net dollar purchases), or negative (net dollar sales).\textsuperscript{12} Macroeconomic effects are incorporated into error term $\chi_{t}$.

\textsuperscript{10} For example, Bhattacharya and Spiegel (1991) and Rochet and Vila (1994) show that microstructure models can have multiple or non-linear equilibria.

\textsuperscript{11} Lyons and Evans (1999) are unable to generate statistically significant forecasts due to small sample size (and possibly because of misspecification and/or the linearity assumption).

\textsuperscript{12} Order flow is explained in the next section (3.2).
A positive relationship between the exchange rate and order flow is expected since informational asymmetries gradually affect the price until it reaches equilibrium. Figure 3.2 illustrates the explanatory power of an aggregate order flow and its IB component (the data cover the period from January 1990 to June 2000 at a daily frequency).

Figure 3.2. Above: Aggregate (cumulative) order flow and log Canada/U.S. real exchange rate. Below: IB (cumulative) order flow and log Canada/U.S. real exchange rate. Note: All values are normalized to [-1,1].

As the solid line indicates, the Canadian dollar has depreciated throughout most of the sample. The relationship between the exchange rate and order flow is quite clear as a positive correlation between cumulative purchases of U.S. dollars and the
depreciation. This relationship is more obvious from the three-month sample (December, 1999 – February, 2000) of IB order flow and log Canada/U.S. real exchange rate. However, it would be inappropriate to assume that order flow contains all the information that is relevant for exchange rates.\textsuperscript{13}

This thesis combines macroeconomic and microstructure approaches into a single high-frequency data model.\textsuperscript{14} More specifically, it embodies modified models from Amano and van Norden (1995, 1998) and Lyons and Evans (2002):

\[
\Delta \text{rpfx}_t = \Psi(\Delta \text{intdiff}_{t-1}, \Delta \text{oil}_{t-1}, \Delta x_{t-1}) + \eta_t, \quad t=1, \ldots, N.
\]

where $\Delta \text{intdiff}$ is the change in the differential between the Canada and U.S. nominal 90-day commercial thesis rates, $\Delta \text{oil}_t$ is the daily change in the logarithm of the crude oil price, and order flow is denoted by $\Delta x_t$. Later in this chapter, $\Delta x_t$ is substituted by either a vector of three different order flow types, aggregate order flow or interbank transactions.

ANNs, NN and TSK methods are employed to estimate a nonlinear relationship between exchange rate movements and these variables.

\textsuperscript{13} Models of this type can be found in Kyle (1985), Back (1992), etc.

\textsuperscript{14} Goldberg and Tenorio (1997) and Osler (1998) also follow this approach.
3.2 Model Specification

3.2.1 Data description

The order flow data were obtained from the Bank of Canada’s unique Daily Foreign Exchange Volume Report, which is coordinated by the Bank and organized through the Canadian Foreign Exchange Committee (CFEC). Details about the trading flows (in Canadian dollars) for six major Canadian commercial banks are categorized by the type of trade (spot, forward, and futures) and transaction type (i.e., with regard to trading partner). Because this thesis focuses on a short-term exchange rate forecast, spot transactions are of interest. In a spot transaction, a currency is traded for immediate delivery and payment is made within two business days of the contract entry date. Spot transactions vary, as follows:

- Commercial client transactions (CC) include all transactions with resident and non-resident non-financial customers.

- Canadian-domiciled investment transactions (CD) include all transactions with non-dealer financial institutions located in Canada.

- Foreign institution transactions (FD) include all transactions with foreign financial institutions, such as FX dealers.

15 The six banks used in this research account for approximately 83 per cent of all Canada/U.S. dollar transactions. The remaining transactions occur within Canada (4 per cent) and in the US and the rest of the world (13 per cent). Source: Bank of Canada.
• Interbank transactions (IB) include transactions with other chartered banks, credit unions, investment dealers, and trust companies in the interbank market.

Because it was unavailable prior to 1994, CD transactions are excluded as an explanatory variable in this work. However, according to a linear model by D'Souza (2002), this variable is not statistically significant in exchange rate forecasting and it is not considered to be a major part of aggregate order flow. Moreover, according to Reuters Dealing 2000-1 electronic dealing system, IB transactions account for about 75 per cent of total trading in major spot markets (Lyons and Evans 2002). Thus, the CD transactions contribution in daily exchange rate explanation is relatively small and our results support this conjecture.

Individual order flows (CC, FD, IB) are measured as the difference between the number of currency purchases (buyer-initiated trades) and sales (seller-initiated trades). Aggregate order flow (aggof) is the sum of individual order flows. As noted earlier, the other variables of interest are the crude oil closing price (in U.S. dollars) deflated by the U.S. consumer price index (CPI) (Δoil) and the change in the difference between nominal 90-day commercial paper rates in Canada and the United States (Δintdiff). The dependent variable data set comprises the logarithm of real Canada/U.S. exchange rate daily changes (Δrpfx) between January 1990 and June 2000. To control for the day-of-the-week effects, we focus on Wednesdays only: a total of 523 observations (after estimating the first differences). The first 470 observations represent the in-sample data and roughly the last 10 per cent (53 observations) are out-of-sample data. The real
exchange rate was calculated from the nominal exchange rate and CPI for the United States and Canada.

All variables are considered in first-difference terms, because the daily change (positive or negative) prediction is of interest. Also, by using the first differences we avoid theoretical problems of estimation of non-stationary non-parametric functions (see Diebold and Nerlove 1989b). To support this claim, standard Dickey-Fuller unit roots tests are performed on all variables and based on the entire data set (Table 3.1). The logarithm of the real exchange rate, interest rate differential and the logarithm of the crude oil price are found to be integrated of order one. In contrast, the null hypothesis of nonstationarity for daily order flows is rejected at the 99 per cent significance level. These findings are not surprising and in line with D'Souza (2002) who used a subset of our data set (from January 1996 through September 1999).

Table 3.1 Augmented Dickey-Fuller (ADF) unit root t-tests

<table>
<thead>
<tr>
<th>Variable</th>
<th>t-test</th>
<th>Lags (f)</th>
<th>Variable</th>
<th>t-test</th>
<th>Lags (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real exchange rate</td>
<td>-0.84</td>
<td>1</td>
<td>Change in real exchange rate</td>
<td>-34.31**</td>
<td>1</td>
</tr>
<tr>
<td>Interest rate differential</td>
<td>-1.48</td>
<td>8</td>
<td>Change in interest rate differential</td>
<td>-19.90**</td>
<td>5</td>
</tr>
<tr>
<td>Crude oil price</td>
<td>-2.42</td>
<td>10</td>
<td>Change in crude oil price</td>
<td>-31.35**</td>
<td>2</td>
</tr>
<tr>
<td>CC</td>
<td>-25.31**</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FD</td>
<td>-23.43**</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>-31.12**</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The sample is from January 1990 to July 2000. Critical values are from Hamilton (1994): *-3.43 (1 per cent), **-2.86 (5 per cent).
Given the systematic bias in currency forward rates (for more on this “conditional bias,” see Lyons 2001), and their unavailability on a daily basis (the minimum contract length is 30 days), we do not test if the inclusion of a forward rate into our set of explanatory variables improves the daily predictability of the Canada/U.S. exchange rate. However, we test for bias in forward rates based on the following equation as suggested by Lyons (2001):

\[ p_{t+30} - p_t = \alpha + \beta(f_{t,30} - p_t) + \epsilon_{t+30} \]

where \( p_{t+30} \) denotes the spot rate realized at time \( t+30 \) (Canada/U.S.), \( f_{t,30} \) denotes the 30-day forward rate of the Canadian dollar, settled at time \( t+30 \), and \( \epsilon_{t+30} \) is a random error term. The data for this exercise were obtained from the Statistics Canada CANSIM database (data range: January 1994 – May 2002).

The regression results can be summarized as follows:

\[ \hat{p}_{t+30} - \hat{p}_t = 0.000525 - 0.432875(f_{t,30} - p_t), \quad R^2 = 0.010 \]

\[ (2.937104) \quad (-3.449215) \]

We report the t-values in parentheses. The estimate of \( \beta \) is statistically significant and negative. In other words, when the forward rate predicts the spot rate will rise, it actually falls. Additionally, the model has a very poor fit with \( R^2 \) below one per cent. This confirms the bias in the 30-day forward rate of the Canadian dollar.

Initially, two non-linear models are considered:

**Model 1:** \( \Delta \text{rf}_t = f(\Delta \text{intdiff}_{t,1}, \Delta \text{oil}_{t,1}, \text{IB}_{t,1}) + \epsilon_t; \quad t=1,..,N. \)

**Model 2:** \( \Delta \text{rf}_t = g(\Delta \text{intdiff}_{t,1}, \Delta \text{oil}_{t,1}, \text{CC}_{t,1}, \text{IB}_{t,1}, \text{FD}_{t,1}) + \nu_t; \quad t=1,..,N. \)
Only for the purpose of ANN modeling, all data were normalized to the [-1,1] interval. Each of the above non-linear models was developed based on chapter 2 techniques and data sets of four variables (model 1) and six variables (model 2). The model was used to forecast the daily change of the Canada/U.S. real exchange rate one day into the future.

We *a priori* specify model parameters, i.e., ANN structure, the number of nearest neighbours, TSK clusters, etc. This pre-specification is based on the cross validation procedure which minimizes MSPE and maximizes PERC. We do not exactly follow Gencay (1999), but our approach imposes a fair restriction to all of the considered models. Essentially, once determined for the in-sample data (training set), the optimal model structure does not adapt with each out-of-sample observation (testing set). As pointed out by Gencay (1999) recalculation with each testing sample is computationally expensive and in this thesis we do not attempt to construct a flawless on-line exchange rate predictor. Rather, the intention is to explore the usefulness of an order flow variable and the possibility of presence of conditional-mean non-linearities.

Initially, our implementation of the cross-validation follows Gencay (1999). First, we select the 70 most recent training set observations (approximately 1/7 of the training set size). At this point, all the remaining training set observations are treated as “testing data.” Then, we repeatedly calculate MSPE and PERC statistics from these “testing data” until we determine the optimal model structure which yields the lowest MSPE and highest PERC. This structure is recorded and we proceed with the cross-validation by adding 10 more observations from the past to our initial training data. Again, we record
the optimal model parameters for a new smaller “testing set” and this procedure is
applied until the size of a “testing set” is zero. Finally, we choose our models’
parameters from the set of recorded model architectures. Specifically, we choose models
with lowest MSPEs and highest PERCs. Non-linear model parameters are held fixed for
the whole testing set. We consider the absence of model re-adjustment as a strong, but
fair restriction. If the applied non-linear model is truly capable of capturing non-
linearities in the exchange rate, the pattern should be quite clear from the basic training
set.

In order to investigate the robustness of our ANN-generated forecasts, in the next
step, we focus on daily and weekly forecasting with backpropagation ANNs where we
develop models of four variables (ANN model 1) and six variables (ANN model 2) using
the whole data set of 2,230 observations. Two models are considered:

ANN Model 1: \[ \Delta \text{rpfx}_t = f (\Delta \text{intdiff}_{t,j}, \Delta \text{oil}_{t,j}, \text{aggof}_{t,j}) + \varepsilon_t; \]
ANN Model 2: \[ \Delta \text{rpfx}_t = g (\Delta \text{intdiff}_{t,j}, \Delta \text{oil}_{t,j}, \text{CC}_{t,j}, \text{IB}_{t,j}, \text{FD}_{t,j}) + \nu_t; \]
\[ j=\{1, 7\}; t=1,\ldots,N. \]

The networks trained and tested were the three-layer and four-layer
backpropagation ANNs with the non-linear sigmoid and tan-sigmoid neuron activation
functions in hidden layers. The number of input neurons was three for model 1 and five
for model 2, while the number of hidden neurons varied between three and five for both
of the models. The last layer had one linear output neuron.
It is important to note that in this research we do not utilize a more complex scenario involving other ANN types and methods for determining optimal ANN structure. Our intention was to emphasize the role and usefulness of the microstructure variables (information) in a relatively simple setting, without refining the estimation procedures. It is possible that a more complicated approach such as Gencay and Qi (2001) would be more successful.

The ANN and NN analyses were performed using the software package Matlab, v. 5.2.0, The MathWorks, Inc. (1998). The TSK estimation was done with the C++ program created by Dr. Dragan Kukolj, University of Novi Sad, Yugoslavia.

### 3.2.2 Assessment of forecast performance

This thesis considers whether the ANN and other non-linear models can outperform linear models (with order flows included into the set of explanatory variables), Autoregressive Integrated Moving Average with eXternal process models (ARIMAX) and random walk (with a drift) models in terms of root mean squared error (MSPE) and the percentage of correctly predicted directions of exchange rate changes (PERC).

MSPE is defined as follows:

$$\text{MSPE} = \frac{1}{M} \sum_{t=1}^{M} (\hat{\Delta rrf_{x_t}} - \Delta rrf_{x_t})^2$$
where $M$ denotes the size of the out-of-sample testing set, $\Delta \hat{rpfx}_t$ is the model forecast at time $t$ ($t=1,\ldots,M$) and $\Delta rpfx_t$ is the actual (sample) exchange rate change.

In the second part of this chapter, in line with the Meese and Rogoff (1983) evaluation criterion, recursive estimation (or rolling regressions) is used to evaluate the models' predictive performance. The initial estimation starts with the first 90 per cent (chronologically) of the sample $N$, or, for instance, $m$ observations. That comprises training and validation sets for the ANN. The remaining 10 per cent is a testing (forecasting) set of initial size $k$; having estimated the model, $k$ forecasts are generated. Subsequent $(k-1)$ steps involve increasing the estimation sample (so that $m$ increases) and shrinking the testing set (so that $k$ decreases) by one period. In each subsequent step, $(k-s)$ forecasts are estimated ($s=1,\ldots,k-1$). Finally, $k$ sets of network responses (of size $1,\ldots,k$) can be compared to actual observations and other models.

This part of the chapter considers whether the ANN model can outperform linear and random walk models in terms of root-mean squared error (RMSE) and the percentage of correctly predicted directions of exchange rate changes. For most of this chapter, RMSE is defined as follows:

$$RMSE_{N_k} = \left[ \frac{1}{N_k} \sum_{t=0}^{N_k-1} \left( \Delta \hat{rpfx}_{k+m-t} - \Delta rpfx_{k+m-t} \right)^2 \right]^{1/2}$$

where $N_k$ denotes the size of the out-of-sample testing set ($N_k = 1,\ldots,k$), $\Delta \hat{rpfx}_{k+m-t}$ is the model forecast at time $(k+m-t)$, and $\Delta rpfx_{k+m-t}$ is the actual exchange rate change. The
ratio of data allocated to training, validation, and testing was maintained at 6:3:1 throughout the recursive experiment.

To reduce overfitting, i.e., inconsistency in calibration, which results from the network complexity (too many parameters to be estimated) and (possibly noisy) data length, one approach would be to use ANNs with hints (homogeneity, consistency, etc.). Recently, Garcia and Gencay (2000) showed that an ANN with homogeneity hint can produce a smaller out-of-sample error and a more robust estimator. Gencay and Qi (2001) found that early stopping can be as effective as the homogeneity hint in pricing options. By using early stopping in this thesis we assume that response functions are estimated consistently.

In addition to MSPE and RMSE, the percentage of correctly predicted signs (PERC) of the forecasted variable \( \Delta rpfx \), is considered; this is the total number of correctly forecasted positive and negative movements, defined as:

\[
PERC = \frac{1}{M} \sum_{t=1}^{M} p_t
\]

where \( p_t = \begin{cases} 
1 & \text{if} (\Delta rpfx_t \cdot \hat{\Delta rpfx}_t) > 0, \\
0 & \text{otherwise}. 
\end{cases} \)

Sometimes, the significance of the difference in the performance of alternative models has to be tested. We use Diebold-Mariano test (Diebold and Mariano 1995) to test the null hypothesis that there is no difference in the MSPE and PERC of two alternative models (in our case of the random walk and the other models).
Let $y_t (t=1,...,M)$ denote the time series to be forecasted and let $\hat{y}_1^t$ and $\hat{y}_2^t$ denote two different forecasts of $y_t$. For instance, $\hat{y}_1^t$ and $\hat{y}_2^t$ could originate from a non-linear ANN model and a random walk process, respectively. We define forecast errors as follows:

$$\epsilon_1^t = y_t - \hat{y}_1^t, \quad t=1,...,M$$

$$\epsilon_2^t = y_t - \hat{y}_2^t, \quad t=1,...,M$$

In general, a time-$t$ loss associated with a forecast is an arbitrary function of $y_t$ and $\hat{y}_1^t$ or $\hat{y}_2^t$. We denote those functions $l_1(y_t, \hat{y}_1^t)$ and $l_2(y_t, \hat{y}_2^t)$. In many cases such as MSPE or root mean squared error (RMSE), $l_1(\cdot)$ and $l_2(\cdot)$ will be functions of $\epsilon_1^t$ and $\epsilon_2^t$. Our loss functions are of the following form:

$$l_i = (\epsilon_i^t)^2, \quad i=1,2; \quad t=1,...,M.$$ 

$$l_i = \begin{cases} 0 & \text{if}(y_t - \hat{y}_i^t) > 0, \\ -1 & \text{otherwise} \end{cases}, \quad i=1,2; \quad t=1,...,M.$$ 

The first one is used to measure MSPE performance, and the second one for directional accuracy, or PERC.

The null hypothesis of interest is:

$H_0$: Equal forecast accuracy for $\hat{y}_1^t$ and $\hat{y}_2^t$ ($t = 1,...,M$),

or $E\left(l_1(y_t, \hat{y}_1^t)\right) = E\left(l_2(y_t, \hat{y}_2^t)\right)$, or $E(d_i) = 0$ ($d_i = l_1(y_t, \hat{y}_1^t) - l_2(y_t, \hat{y}_2^t)$),
and the alternative hypothesis is:

\[ H_i: \text{Different forecast accuracy or } \mathbb{E}(d_i) \neq 0. \]

The Diebold-Mariano test statistic for the equivalence of forecast errors will be

\[ S_1 = \frac{1}{M} \sum_{i=1}^{M} (d_i) \]

\[ \sqrt{\frac{2 \pi f(0)}{M}} \]

where M is the testing set size and f(0) is the spectral density of d_i at frequency zero. Diebold and Mariano show that S_i is asymptotically distributed as a N(0,1). If we use the above quadratic loss functions, we would have the following:

- When S_i < -1.96 model “1” has a MSPE significantly lower than model “2” at 95% confidence level (H_0 rejected at 5% significance level).
- When S_i > 1.96 model “2” has a MSPE significantly lower than model “1” at 95% confidence level (H_0 rejected at 5% significance level).
- When -1.96 < S_i < 1.96, H_0 is not rejected at 5% significance level.

For the directional accuracy loss functions the following would be true:

- When S_i < -1.96 model “1” has a PERC significantly lower than model “2” at 95% confidence level (H_0 rejected at 5% significance level).
- When S_i > 1.96 model “2” has a PERC significantly lower than model “1” at 95% confidence level (H_0 rejected at 5% significance level).
- When -1.96 < S_i < 1.96, H_0 is not rejected at 5% significance level.
As forecasts are done only one step ahead we do not have to include any of the sample autocovariances to calculate the long-run variance of $d_t$, $2\pi f(0)$. In this case, a consistent estimate of $2\pi f(0)$ will be the sample variance of $d_t$ (see Campbell et al. 1997 or Diebold and Mariano 1995). The estimated autocovariances of series $d_t$ (MSPE performance) for the ANN model 2 (backpropagation) for lags 0 to 7 are $10^{-12} \times (0.2847, 0.0577, 0.0233, 0.0134, -0.0396, -0.0192, -0.0003, -0.0300)$. Similarly, when directional accuracy is measured for the ANN model 2, the following autocovariances are estimated: $0.1381, -0.0133, -0.0334, -0.0144, 0.0463, -0.0178, -0.0185, -0.0143$. Clearly, except for lag 0, the estimated autocovariances are roughly zero and do not have to be included in the calculation of $2\pi f(0)$.

### 3.3 Empirical Results

#### 3.3.1 Models based on individual order flows when controlled for the day-of-the-week-effects

This section assesses the out-of-sample forecasting performance of a range of Canada/U.S. exchange rate models. The out-of-sample size is fixed at the last 53 observations. Generally, the random walk model performs better than any traditional
linear macroeconomic model that excludes microstructure variables; therefore, it can be viewed as the benchmark model. The following models are considered first:

**Random walk model (RW):**
\[ rpx_t = a_0 + rpx_{t-1} + \gamma_t, \quad t=1,\ldots,N. \]

**Linear model (LM1):**
\[ \Delta rpx_t = \gamma_0 + \gamma_1 \Delta \text{diff}_{t-1} + \gamma_2 \Delta \text{oil}_{t-1} + \gamma_3 \text{IB}_{t-1} + \epsilon_t, \quad t=1,\ldots,N. \]

**Linear model (LM2):**
\[ \Delta rpx_t = \beta_0 + \beta_1 \Delta \text{diff}_{t-1} + \beta_2 \Delta \text{oil}_{t-1} + \beta_3 \text{CC}_{t-1} + \beta_4 \text{IB}_{t-1} + \beta_5 \text{FD}_{t-1} + \nu_t, \quad t=1,\ldots,N. \]

To capture more of the model dynamics we use the Box-Jenkins approach. The ARMA models can typically provide an accurate description of stationary non-seasonal time series. This series may be forecasted by fitting an appropriate ARMA-type model. We employ the following two versions of this popular model:

**ARIMAX1(p,1,q):**
\[ A(L, \rho)(1 - L)rpx_t = \sum_{i=1}^{3} B_i(L, \alpha) X_{it} + C(L, \beta) \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \omega^2) \]

where
\[
A(L, \rho) = 1 + \rho_1 L + \ldots + \rho_p L^p.
\]
\[
B_i(L, \alpha) = 1 + \alpha_1 L + \ldots + \alpha_q L^q.
\]
\[
C(L, \beta) = 1 + \beta_1 L + \ldots + \beta_q L^q.
\]
\[
X_t = (\Delta \text{int diff}_{t-1}, \Delta \text{oil}_{t-1}, \text{IB}_{t-1}).
\]
\[t = 1, \ldots, N.\]

**ARIMAX2(p, l, q):**

\[
A(L, \rho)(1 - L)\rho f x_t = \sum_{i=1}^{5} B_i(L, \alpha)X_{it} + C(L, \beta)\varepsilon_t, \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)
\]

where
\[
A(L, \rho) = 1 + \rho_1 L + \ldots + \rho_p L^p.
\]
\[
B_i(L, \alpha) = 1 + \alpha_1 L + \ldots + \alpha_q L^q.
\]
\[
C(L, \beta) = 1 + \beta_1 L + \ldots + \beta_q L^q.
\]
\[
X_t = (\Delta \text{int diff}_{t-1}, \Delta \text{oil}_{t-1}, \text{IB}_{t-1}, \text{CC}_{t-1}, \text{FD}_{t-1}).
\]
\[t = 1, \ldots, N.\]

Naturally, we attempt to take advantage of the microstructure data set and that justifies the usage of microstructure external variables.

All of the non-linear models are of the form specified in Section 3.2:

\[
\Delta \rho f x_t = \phi (\text{int diff}_{t-1}, \Delta \text{oil}_{t-1}, \text{IB}_{t-1}) + \varepsilon_t; \quad t = 1, \ldots, N. \tag{3.3.1}
\]

\[
\Delta \rho f x_t = \gamma (\Delta \text{int diff}_{t-1}, \Delta \text{oil}_{t-1}, \text{CC}_{t-1}, \text{IB}_{t-1}, \text{FD}_{t-1}) + \nu_t; \quad t = 1, \ldots, N. \tag{3.3.2}
\]

We denote our set of non-linear models as follows:

- **ANNB1** (backpropagation ANN model for 3.3.1),
- **ANNB2** (backpropagation ANN model for 3.3.2),
• ANNR1 (recurrent ANN model for 3.3.1),
• ANNR2 (recurrent ANN model for 3.3.2),
• NN1 (NN model for 3.3.1),
• NN2 (NN model for 3.3.2),
• TSK1 (TSK model for 3.3.1),
• TSK2 (TSK model for 3.3.2).

The optimal networks trained and tested were the three-layer and four-layer backpropagation and recurrent ANNs with the non-linear sigmoid and tan-sigmoid neuron activation functions in hidden layers. The number of input neurons was three for model 1 and five for model 2, while the number of hidden neurons varied between three and five for both of the models. The last layer had one linear output neuron.

The results depend strongly on the ANN architecture. More specifically, the number of hidden layers, number of neurons in the hidden layer, type of activation function, and training algorithm are principal determinants of a good prediction model. To avoid overtraining, the ANN was trained with an *early stopping* technique, where the available data set was divided into three subsets (maintaining the ratio 6:3:1): a smaller training set (used for gradient calculation and weights and biases updating); a validation set (when the validation error starts increasing, the training is stopped); and a testing set (used to compare real and model output, or different models).
The optimal number of nearest neighbours was for $\zeta = 0.7$, i.e., approximately 270 nearest neighbours. The choice of seven clusters produced the most accurate TSK model.

The results of our out-of-sample forecasting exercise are summarized in Table 3.2. The first column represents the relative (to the benchmark) MSPE associated to each of the models; the benchmark model is an RW model and its MSPE is normalized to the value of one. Thus, the relative MSPE shows the exact relative gain (or loss) of our competing models compared to the RW model. The PERC statistics are reported in the second column. In the MSPE and PERC out-of-sample estimations, we use the optimal non-parametric model structure found from cross-validation procedure on the training set (470 observations). The Diebold-Mariano (DM) test statistics are reported in the parentheses below MSPEs and PERCs. If there are significant gains in MSPE (for instance, at 5% significance level) relative to the benchmark, the DM value should be less than -1.96. Similarly, if there are significant gains in PERC (at 5% significance level), the DM value should be greater than 1.96.\textsuperscript{16}

The PERC statistic of approximately 58 per cent for the RW model is very high. This suggests a significant amount of positive exchange rate movements, which is not surprising for the period 1990-2000. Relative to the benchmark model, linear models produce minimal gains. ARIMAX (2,1,1) models exert a 6 per cent improvement in PERC, but almost no improvement in MSPE. Adding more lags to the AR and MA

\textsuperscript{16} The critical values are +/- 1.64 and +/- 1.96 for a confidence level of 90% and 95%, respectively. (*) and (**) indicates the DM statistic is significant at 5% and 10% significance level, respectively.
components of the ARIMAX model only worsens these results. There is certainly a
degree of overfitting involved with the conditioning on past exchange rate changes and
eerrors.

Table 3.2 Out-of-sample forecast performance.

<table>
<thead>
<tr>
<th>Models</th>
<th>Relative MSPE</th>
<th>PERC [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(DM)</td>
<td>(DM)</td>
</tr>
<tr>
<td>RW</td>
<td>1.000</td>
<td>58.490</td>
</tr>
<tr>
<td>LM1</td>
<td>0.993</td>
<td>60.377</td>
</tr>
<tr>
<td>(-0.259)</td>
<td>(0.255)</td>
<td></td>
</tr>
<tr>
<td>LM2</td>
<td>0.994</td>
<td>62.264</td>
</tr>
<tr>
<td>(-0.206)</td>
<td>(0.467)</td>
<td></td>
</tr>
<tr>
<td>ARIMAX1(2,1,1)</td>
<td>0.991</td>
<td>64.150</td>
</tr>
<tr>
<td>(-0.269)</td>
<td>(0.553)</td>
<td></td>
</tr>
<tr>
<td>ARIMAX2(2,1,1)</td>
<td>0.993</td>
<td>62.264</td>
</tr>
<tr>
<td>(-0.197)</td>
<td>(0.362)</td>
<td></td>
</tr>
<tr>
<td>ANNB1</td>
<td>0.949</td>
<td>67.924</td>
</tr>
<tr>
<td>(-0.888)</td>
<td>(1.695)**</td>
<td></td>
</tr>
<tr>
<td>ANNB2</td>
<td>0.920</td>
<td>69.811</td>
</tr>
<tr>
<td>(-2.285)*</td>
<td>(2.196)*</td>
<td></td>
</tr>
<tr>
<td>ANNR1</td>
<td>0.971</td>
<td>67.924</td>
</tr>
<tr>
<td>(-0.827)</td>
<td>(1.938)*</td>
<td></td>
</tr>
<tr>
<td>ANNR2</td>
<td>0.952</td>
<td>69.811</td>
</tr>
<tr>
<td>(-2.357)*</td>
<td>(2.192)*</td>
<td></td>
</tr>
<tr>
<td>NN1</td>
<td>0.964</td>
<td>60.377</td>
</tr>
<tr>
<td>(-0.758)</td>
<td>(0.206)</td>
<td></td>
</tr>
<tr>
<td>NN2</td>
<td>0.917</td>
<td>66.037</td>
</tr>
<tr>
<td>(-1.590)</td>
<td>(0.813)</td>
<td></td>
</tr>
<tr>
<td>TSK1</td>
<td>1.024</td>
<td>58.491</td>
</tr>
<tr>
<td>(0.106)</td>
<td>(0.012)</td>
<td></td>
</tr>
<tr>
<td>TSK2</td>
<td>0.970</td>
<td>62.264</td>
</tr>
<tr>
<td>(-0.296)</td>
<td>(0.389)</td>
<td></td>
</tr>
</tbody>
</table>
One very clear pattern emerges: both types of ANNs are markedly better than any of the linear and non-linear competitors. They produce best forecasts when all three order flows are included (models ANNB2 and ANN2). The ANNB2 model provides statistically significant gains in MSPE (about 8 per cent) and PERC (about 12 per cent), while for the ANNB1 model there are only significant forecasting improvements in PERC. Further, backpropagation ANN appears to be more successful than the recurrent one. Thus, as in the linear case, lagged variables do not matter a lot for high-frequency forecasting. The NN2 model also provides almost border significant (at 10 per cent) reductions in MSPE relative to the RW model. TSK models do not offer any significant forecast gains, but, again, more accurate forecasts are related to a non-linear five-variable model.

In Table 3.3, we also present the out-of-sample performance (relative MSPE and PERC) for longer forecast horizons - two-step-ahead and three-step-ahead ANNB2 forecasts. These forecasts are notably inferior to the ones from the one-step-ahead experiment. The PERC statistic incurs losses of roughly 2 and 5 per cent, respectively, while the relative MSPE increases. Thus, the findings show that the optimal forecast horizon for daily frequency is 1 day.

<table>
<thead>
<tr>
<th>Horizon (k)</th>
<th>Relative MSPE</th>
<th>PERC [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>k=2</td>
<td>0.958</td>
<td>67.307</td>
</tr>
<tr>
<td>k=3</td>
<td>0.966</td>
<td>64.705</td>
</tr>
</tbody>
</table>

Table 3.3 Two-step-ahead and three-step-ahead forecast performance of the ANNB2 model.
3.3.2 Parameter estimates and relative significance of inputs

ANNs are often regarded as non-parametric “black-box” models because they involve highly complex non-linear functional forms and parameters which are difficult to interpret. However, once the ANN is trained (estimated), the relationship between inputs and outputs is uniquely defined by non-linear (transfer) functions, connection weights and node biases. This allows us to make ANNs somewhat transparent to interpretation and analysis. We extend the method of pseudo-weights proposed by Qi and Maddala (1995) to find the economic importance of the inputs. More precisely, we use the weighted average of the input weights to find the marginal contribution of each input variable to the output. Using the notation from chapter 2, we slightly modify their formula for pseudo-weights and define the four-layered ANN’s pseudo-weight for the \( i \)th input (\( PW_i \)) as

\[
PW_i = \sum_{h=1}^{q_1} \sum_{h'=1}^{q_2} \alpha'_{ij} \cdot \alpha''_{hj} \cdot \beta_{hj}
\]  

(3.3.3)

where \( q_1 \) is the number of nodes in the first hidden layer, \( q_2 \) is the number of nodes in the second hidden layer, \( \alpha'_{ij} \) denotes connection weights between the input layer and the first hidden layer (i.e., between the \( i \)th input node and the \( j_1 \)th first hidden layer node), \( \alpha''_{hj} \) denotes connection weights between hidden layers (i.e., between the \( j_1 \)th first hidden layer node and the \( j_2 \)th second hidden layer node), and \( \beta_{hj} \) represent the connection weights between the second hidden layer (\( j_2 \)th second hidden layer node) and the output.

\[17\] Qi and Maddala (1995) use ANNs to price call options and find that the economic implications of pseudo-weights are consistent with the call option properties.
layer (single node). Since we are dealing with a four-layered ANN here, we modify the above three-layered ANN notation by introducing \( q_1 \) and \( q_2 \) hidden nodes (as opposed to \( q \) nodes in a single hidden layer ANN), and two sets (matrices) of connection weights \( \alpha'_{ji} \) and \( \alpha''_{ji} \) associated with the connections inputted into hidden layers (as opposed to \( \alpha_{ji} \) for a single hidden layer). Also, for a four-layered ANN we denote non-linearities in hidden layers with \( \psi_1 \) and \( \psi_2 \), and their biases with \( \alpha'_{ji0} \) (for the \( j_1^{th} \) first hidden layer node) and \( \alpha''_{ji0} \) (for the \( j_2^{th} \) second hidden layer node). The recurrent ANN follows the same notation.

As mentioned before, the ANN models with two macroeconomic variables and three order flow types generated the greatest and statistically significant one-day-ahead forecast improvements. Table 3.4 uses the notation from 3.3.3 and presents the model's forecast performance for different ANN types.

### Table 3.4 The ANN models with the greatest forecast improvements.

<table>
<thead>
<tr>
<th>ANN type</th>
<th>Relative MSPE</th>
<th>PERC [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>[p ( \psi_1 ) ( \psi_2 ) 1 (( \sigma ))]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Backpropagation (ANNBP1)</td>
<td>0.92 (DM)</td>
<td>92.598 (DM)</td>
</tr>
<tr>
<td>2. Backpropagation (ANNBP2)</td>
<td>-2.285*</td>
<td>69.811</td>
</tr>
<tr>
<td>3. Recurrent (Elman1)</td>
<td>0.952 (DM)</td>
<td>69.811</td>
</tr>
<tr>
<td>4. Recurrent (Elman2)</td>
<td>-2.357*</td>
<td>69.811</td>
</tr>
</tbody>
</table>

Table 3.4 also gives more details on the ANN’s structure (the number of layers, neurons and the activation function’s type for each layer). For example, “[5 3 (logsig) 5 (tansig) 1 (purelin)]
(tansig) 1 (purelin)]" means there are four layers: five neurons in the input layer, three neurons in the first hidden layer (with log-sigmoid activation functions), five neurons in the second hidden layer (with tan-sigmoid activation functions) and one neuron in the output layer (with a linear activation function). Before we proceed with the calculation of the pseudo-weights for each of these four models, we will present the estimates for the connection weights and biases. It is worth mentioning that an estimated model with significant forecasting improvements in terms of MSPE does not necessarily involve significant improvements in PERC, and vice versa. Consequently, the sets of weights (and pseudo-weights) and biases for different estimated models vary. In Tables 3.5-3.8 we report the connection weights and biases for our models (using notation from chapter 2). These weights are used for the calculation of pseudo-weights for inputs.

Table 3.5 Estimated ANN's connection weights and node biases (Model 1).

<table>
<thead>
<tr>
<th>Layer</th>
<th>Connection weights matrix</th>
<th>Node biases vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hidden 1</td>
<td>$\alpha'_{bh} =$</td>
<td>$\alpha'_{b0} =$</td>
</tr>
<tr>
<td></td>
<td>9.8549 -2.0998 -2.9391 2.0196 0.0827</td>
<td>3.6983</td>
</tr>
<tr>
<td></td>
<td>-13.2714 -3.5306 0.7513 -1.8734 -1.4176</td>
<td>-8.6814</td>
</tr>
<tr>
<td></td>
<td>15.5297 2.8011 -2.2898 -1.8303 0.5049</td>
<td>12.5395</td>
</tr>
<tr>
<td></td>
<td>1.2344 -1.9592 -0.0203 -3.3835 -1.1381</td>
<td>3.7493</td>
</tr>
<tr>
<td></td>
<td>-9.8746 -2.5924 0.4338 0.3805 -0.8059</td>
<td>-11.5194</td>
</tr>
<tr>
<td>Hidden 2</td>
<td>$\alpha''_{bh} =$</td>
<td>$\alpha''_{b0} =$</td>
</tr>
<tr>
<td></td>
<td>-0.8239 -1.1924 -1.3401 -1.7941 -2.7949</td>
<td>6.4532</td>
</tr>
<tr>
<td></td>
<td>-1.6048 -2.6335 0.8536 1.2977 1.5261</td>
<td>0.9693</td>
</tr>
<tr>
<td></td>
<td>0.6845 0.2403 2.8913 2.3210 -0.6332</td>
<td>-2.6727</td>
</tr>
<tr>
<td></td>
<td>-0.1492 -1.6125 -0.4865 1.9818 2.8745</td>
<td>-1.2680</td>
</tr>
<tr>
<td></td>
<td>0.1582 0.2500 -0.8945 -2.1805 3.0275</td>
<td>1.9737</td>
</tr>
<tr>
<td>Output</td>
<td>$\beta_2 =$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3806 -0.0727 0.3939 0.4582 0.6165</td>
<td>-0.2192</td>
</tr>
</tbody>
</table>
Table 3.6 Estimated ANN's connection weights and node biases (Model 2).

<table>
<thead>
<tr>
<th>Layer</th>
<th>Connection weights matrix</th>
<th>Node biases vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hidden 1</td>
<td>(\alpha_1 = \begin{bmatrix} -5.9156 &amp; -0.6221 &amp; 0.1574 &amp; -0.1634 &amp; 1.3747 \ 8.2915 &amp; -0.0841 &amp; 0.1372 &amp; -0.8574 &amp; 0.2101 \ -2.0364 &amp; -0.2606 &amp; -1.8518 &amp; -0.6910 &amp; 1.1828 \end{bmatrix})</td>
<td>(\alpha_1^0 = -2.0194)</td>
</tr>
<tr>
<td></td>
<td>(\alpha_2 = \begin{bmatrix} 5.1096 \ -2.7705 \end{bmatrix})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(\alpha_4 = \begin{bmatrix} 7.8997 &amp; 3.5333 &amp; -1.7588 &amp; -0.4000 &amp; 1.3543 \ -5.3092 &amp; -0.6804 &amp; -0.6134 &amp; 2.3290 &amp; 0.3758 \ -4.6142 &amp; 0.1774 &amp; 1.6841 &amp; -0.3809 &amp; -0.0554 \ -10.4674 &amp; -0.4825 &amp; 2.8461 &amp; -0.6674 &amp; -0.1229 \end{bmatrix})</td>
<td>(\alpha_4^0 = -2.8190)</td>
</tr>
<tr>
<td></td>
<td>(\alpha_5 = \begin{bmatrix} 1.8518 &amp; 0.8631 &amp; -0.4264 &amp; -1.1182 &amp; 1.6506 \ 1.9171 &amp; 1.9706 &amp; 0.4564 &amp; 1.5809 &amp; -0.7884 \ 0.3869 &amp; 2.0378 &amp; 1.0049 &amp; -0.7275 &amp; 0.8327 \ 0.2741 &amp; 1.4992 &amp; 1.3473 &amp; -0.9813 &amp; 1.8014 \ 1.4491 &amp; 0.0813 &amp; -1.9059 &amp; -0.9378 &amp; -0.9777 \end{bmatrix})</td>
<td>(\alpha_5^0 = 1.3190)</td>
</tr>
<tr>
<td>Output</td>
<td>(\beta_2 = \begin{bmatrix} 0.3536 &amp; -1.0566 &amp; -0.7508 &amp; 0.2607 &amp; -0.9336 \ -0.1869 \end{bmatrix})</td>
<td>(\beta_2^0 = -0.1869)</td>
</tr>
</tbody>
</table>

Table 3.7 Estimated ANN's connection weights and node biases (Model 3).

<table>
<thead>
<tr>
<th>Layer</th>
<th>Connection weights matrix</th>
<th>Node biases vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hidden 1</td>
<td>(\alpha_1 = \begin{bmatrix} 11.8174 &amp; 0.7001 &amp; 2.2284 &amp; -0.4306 &amp; -3.9251 \ 7.8997 &amp; 3.5333 &amp; -1.7588 &amp; -0.4000 &amp; 1.3543 \ -5.3092 &amp; -0.6804 &amp; -0.6134 &amp; 2.3290 &amp; 0.3758 \ -4.6142 &amp; 0.1774 &amp; 1.6841 &amp; -0.3809 &amp; -0.0554 \ -10.4674 &amp; -0.4825 &amp; 2.8461 &amp; -0.6674 &amp; -0.1229 \end{bmatrix})</td>
<td>(\alpha_1^0 = 4.1052)</td>
</tr>
<tr>
<td></td>
<td>(\alpha_2 = \begin{bmatrix} 7.3423 \ -2.8190 \ -4.5682 \ -9.5688 \end{bmatrix})</td>
<td></td>
</tr>
<tr>
<td>Hidden 2</td>
<td>(\alpha_3 = \begin{bmatrix} 1.8518 &amp; 0.8631 &amp; -0.4264 &amp; -1.1182 &amp; 1.6506 \ 1.9171 &amp; 1.9706 &amp; 0.4564 &amp; 1.5809 &amp; -0.7884 \ 0.3869 &amp; 2.0378 &amp; 1.0049 &amp; -0.7275 &amp; 0.8327 \ 0.2741 &amp; 1.4992 &amp; 1.3473 &amp; -0.9813 &amp; 1.8014 \ 1.4491 &amp; 0.0813 &amp; -1.9059 &amp; -0.9378 &amp; -0.9777 \end{bmatrix})</td>
<td>(\alpha_3^0 = 1.6529)</td>
</tr>
<tr>
<td></td>
<td>(\alpha_4 = \begin{bmatrix} 1.3190 \end{bmatrix})</td>
<td>(\alpha_4^0 = -2.6040)</td>
</tr>
<tr>
<td></td>
<td>(\alpha_5 = \begin{bmatrix} 0.0153 \ -2.6040 \ 0.0153 \end{bmatrix})</td>
<td>(\alpha_5^0 = -1.6529)</td>
</tr>
<tr>
<td>Output</td>
<td>(\beta_2 = \begin{bmatrix} -0.3047 &amp; 0.1453 &amp; -0.7193 &amp; -0.1458 &amp; -0.1516 \ 0.5370 \end{bmatrix})</td>
<td>(\beta_2^0 = 0.5370)</td>
</tr>
</tbody>
</table>
Table 3.8 Estimated ANN’s connection weights and node biases (Model 4).

<table>
<thead>
<tr>
<th>Layer</th>
<th>Connection weights matrix</th>
<th>Node biases vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hidden 1</td>
<td>$\alpha_{1h}^\prime =$</td>
<td>$\alpha_{1h}^\prime 0 =$</td>
</tr>
<tr>
<td></td>
<td>-0.1283 -1.8204 0.5878 -0.0332 -1.0923</td>
<td>-5.3950</td>
</tr>
<tr>
<td></td>
<td>-7.3858 -0.5503 -1.0351 -0.5692 0.3137</td>
<td>-2.8995</td>
</tr>
<tr>
<td></td>
<td>-9.0746 1.8290 -0.1662 -0.8907 -0.4239</td>
<td>-6.0596</td>
</tr>
<tr>
<td></td>
<td>-2.0041 -1.9276 -0.8401 -0.3912 1.5673</td>
<td>1.1367</td>
</tr>
<tr>
<td></td>
<td>5.0291 1.7058 -2.4187 0.9187 -1.2042</td>
<td>7.3801</td>
</tr>
<tr>
<td>Hidden 2</td>
<td>$\alpha_{2h}'' =$</td>
<td>$\alpha_{2h}'' 0 =$</td>
</tr>
<tr>
<td></td>
<td>-1.8943 -0.8410 0.2635 0.5640 -1.6569</td>
<td>2.3129</td>
</tr>
<tr>
<td></td>
<td>0.9927 -0.6511 0.3522 0.5967 1.3380</td>
<td>-1.0230</td>
</tr>
<tr>
<td></td>
<td>1.4946 -1.6408 0.1485 0.5120 -0.1589</td>
<td>-1.2897</td>
</tr>
<tr>
<td></td>
<td>0.3754 1.0767 -0.0273 -1.1260 1.3149</td>
<td>-1.0367</td>
</tr>
<tr>
<td></td>
<td>-0.2060 -1.7259 -0.0312 0.2953 -0.4576</td>
<td>1.5937</td>
</tr>
<tr>
<td>Output</td>
<td>$\beta_h =$</td>
<td>-0.3593</td>
</tr>
<tr>
<td></td>
<td>-0.2092 0.2087 -0.3803 0.0371 -0.2552</td>
<td>-0.3593</td>
</tr>
</tbody>
</table>

Table 3.9 lists four sets of pseudo-weights, one for each considered model. Regardless of the statistic used for comparison, the sum of the absolute values of the pseudo-weights for the order flows is always greater than the sum of the absolute values of the interest rate differential and crude oil price. This is very strong evidence in favour of our conjecture that order flows are highly important for non-linear exchange rate modeling and forecasting. To support our conclusion about pseudo-weights we also report (from Tables 3.5-3.8) that there is consistently more node bias on the nodes topologically more proximate to the order flow inputs. Clearly, more node bias on these nodes indicates more impact of order flows on the exchange rate forecasts.
Table 3.9 Relative contribution (pseudo-weights) of inputs for the ANN models.

<table>
<thead>
<tr>
<th>Input</th>
<th>1. ANNBP1</th>
<th>2. ANNBP2</th>
<th>3. Elman1</th>
<th>4. Elman2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>-9.9739</td>
<td>-0.6871</td>
<td>-13.0805</td>
<td>-3.2867</td>
</tr>
<tr>
<td>FB</td>
<td>-2.1131</td>
<td>1.1735</td>
<td>-5.3067</td>
<td>-1.7364</td>
</tr>
<tr>
<td>IB</td>
<td>0.3797</td>
<td>10.6137</td>
<td>-0.0916</td>
<td>-2.9237</td>
</tr>
<tr>
<td>intdiff</td>
<td>3.6768</td>
<td>4.7441</td>
<td>0.3761</td>
<td>0.2750</td>
</tr>
<tr>
<td>oil</td>
<td>-0.1070</td>
<td>-6.1089</td>
<td>0.8910</td>
<td>-1.2431</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>order flows contribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12.4667</td>
<td>12.4743</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>macro contribution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.7838</td>
<td>10.8530</td>
</tr>
</tbody>
</table>

In models 1 and 3 we see that when we aim for minimum MSPE, CC and FB order flows are more important than IB order flow. To the contrary, when we use a PERC statistic as a measure of forecast performance, IB order flow becomes more significant in a "non-linear" fashion. It must be stressed that IB order flow accounts for most of the currency trading volume and that directional movements are more important for trading purposes. It appears that agents' expectations about future exchange rate sign movements are related to currency trading and originate in IB order flow, i.e., agents behave more like traders. Other types of order flows play a role in a more traditional forecasting framework based on the MSPE. Also, it is possible, since our models are estimated based on more than ten years of data, that the true DGP is changing over time and there is a degree of model misspecification involved (a few turning points are missed). Lastly, from models 3 and 4 we see that historical IB order
flow values (as well as past interest rates) are less important for both the MSPE and PERC statistic.

The sign effect of order flows on the exchange rate can be inferred from the following regression:

$$\Delta p_t = \beta_1 CC_t + \beta_2 FB_t + \beta_3 IB_t + \eta_t$$

where $\Delta p_t$ is the change in the logarithm of the Canada/U.S. spot rate, $CC_t$, $FB_t$ and $IB_t$ are order flows, as defined previously. We estimate this model, first, for the full data set, and then for the episodes when the Canadian dollar was highly volatile (subsample: November/December 1994 and July/August 1998). The least squares estimated equations (with t-values in parentheses) are

full sample: 2559 observations

$$\Delta \hat{p}_t = -5.95 \times 10^{-7} CC_t + 1.40 \times 10^{-4} FB_t - 1.26 \times 10^{-6} IB_t, R^2 = 0.268$$

$(-6.13)$ $(21.58)$ $(-9.89)$

subsample: 80 observations

$$\Delta \hat{p}_t = 1.17 \times 10^{-7} CC_t + 1.55 \times 10^{-6} FB_t - 6.12 \times 10^{-7} IB_t, R^2 = 0.179$$

$(0.13)$ $(3.00)$ $(-0.87)$

For the full sample, the coefficient on the interbank order flow is highly significant and the sign of the coefficient is as would be expected: net purchases of Canadian dollars by commercial banks should result in a decrease in the Canada/U.S. exchange rate, i.e., lead to an appreciation of the Canadian dollar. Commercial client transactions produce the same effect on the exchange rate as interbank flows, and the coefficient on the variable is statistically significant. By contrast, the estimate of the
coefficient on the foreign domiciled financial institution flow has a positive sign (and is statistically significant), suggesting that commercial bank purchases of Canadian dollars from this type of counterparty relate to a depreciation of the Canadian dollar.

The above conclusions are slightly different for the reduced sample. Both the coefficient on commercial client and the one on interbank transactions are not statistically significant. The only significant estimate is the coefficient on the foreign domiciled financial institution transactions. Moreover, commercial client transactions are found to have the opposite effect on the exchange rate: net purchases of Canadian dollars led to a currency depreciation. Our conjecture is that a two-stage version of "hot potato" trading took place in this instance. More precisely, in the second stage, dealers tried to induce FX customers – who appeared to have provided liquidity by buying Canadian dollars – to absorb their inventory imbalances. A simple look at the first-stage data reveals that the number of purchases of Canadian dollars from the foreign domiciled financial institutions was almost three times greater than the average for the full sample. This increased inventory had to be channeled through other dealers and, eventually, customers. The data support this hypothesis – averages of CC and IB order flows are negative and well below averages for the full sample. Thus, foreign domiciled financial institutions provided instability by selling Canadian dollars, while commercial banks provided both stability (buying from foreign domiciled financial institutions) and instability (excessive selling to other commercial banks). Commercial clients provided the currency stability and liquidity during the periods of strong depreciation. It would be
useful to gain more insight into the trade volume figures, but, unfortunately, data were not available.

To further investigate the pattern of response of the ANNBP2 network (the one with the best performance in terms of PERC) to the order flows for the whole sample, we plot response functions for IB, CC and FD in Figure 3.3. While the values of inputs are held constant at the following out-of-sample values - (CC: -0.7314; FB: -0.1020; IB: 0.5284; oil: 0.3898; interest rate: 0.3871) - the order flows are individually changed, and the response of the exchange rate to each order flow is recorded. The order flows take values from 0.1 to 3.60 (360 values; increment value is 0.01) and all of the other variables are normalized to zero mean and unity variance.

Figure 3.3 Non-linear pattern of response (ANNBP2 model).

* In some cases, setting the inputs to other out-of-sample values can change the pattern of responses, but typical responses are as shown for the above values.
The responses are non-linear and, as in the "linear case," net purchases of Canadian dollars by commercial banks from Canadian commercial banks and commercial clients result in a decrease in the Canada/U.S. exchange rate, i.e., lead to an appreciation of the Canadian dollar. Commercial bank purchases of Canadian dollars from foreign domiciled financial institutions relate to a depreciation of the Canadian dollar. Also, the effects of IB and FD order flows on the exchange rate appear to be stronger than is the case with transactions of commercial clients. This is consistent with the values of pseudo-weights for the ANNBP2 model.

Taking the data for the 1994 depreciation episode, the typical responses are different, but their directions (signs) are somewhat related to the intuition of the above linear exercise. Here, all of the inputs are set to -0.2 and again the responses can be substantially different for some other values of the input variables.

Figure 3.4 Non-linear pattern of response for the 1994 depreciation episode (ANNBP2 model).
Figure 3.4 shows that net purchases of Canadian dollars from foreign domiciled financial institutions and commercial clients are strongly correlated to both the appreciation (for smaller order flow values) and depreciation (for larger order flow values) of the Canadian dollar in a non-linear fashion. To the contrary, interbank flows are weakly (non-linearly) related to the appreciation of the Canadian currency during this depreciation episode.

We find it worthwhile to note that in the previous analyses we assumed that order flows are exogenous, i.e., that there is no feedback effect from the price to the order flows (sometimes referred to as “distressed selling”). Killeen, Lyons and Moore (2001) found that for the Germany/France exchange rate Granger causality ran from order flow to price, and not vice versa. However, in the case study on the collapse of the U.S./Japan exchange rate in 1998, Lyons (2001) found some evidence of falling prices inducing additional selling. If the effects we find in two depreciation episodes of the Canada/U.S. exchange rate are indeed due to “distressed selling” of the Canadian currency, it would be important to model and identify them. This is a difficult task, which we consider to be beyond the scope of this thesis, and thus leave it to further research.
We investigate the robustness of the forecasting performance of ANN models in this section. The following models are considered:

**Random walk model (RW):**

\[ \Delta \text{rpfx}_t = \alpha_0 + \Delta \text{rpfx}_{t-1} + \gamma_t, \ t=1,..,N. \]

**Linear model (LM1):**

\[ \Delta \text{rpfx}_t = \gamma_0 + \gamma_1 \Delta \text{intdiff}_{t,j} + \gamma_2 \Delta \text{oil}_{t,j} + \gamma_3 \text{iB}_{t,j} + \varepsilon_t, \ t=1,..,N. \]

**Linear model (LM2):**

\[ \Delta \text{rpfx}_t = \beta_0 + \beta_1 \Delta \text{intdiff}_{t,j} + \beta_2 \Delta \text{oil}_{t,j} + \beta_3 \text{CC}_{t,j} + \beta_4 \text{iB}_{t,j} + \beta_5 \text{FD}_{t,j} + \nu_t, \ t=1,..,N. \]

**ANN Model 1:**

\[ \Delta \text{rpfx}_t = f (\text{intdiff}_{t,j}, \Delta \text{oil}_{t,j}, \text{aggof}_{t,j}) + \varepsilon_t; \ t=1,..,N. \]

---

ANN Model 2:

\[ \Delta \text{rpfx}_t = g(\Delta \text{intdiff}_{t, j}, \Delta \text{oil}_{t, j}, \text{CC}_{t, j}, \text{IB}_{t, j}, \text{FD}_{t, j}) + \nu_t; \quad t=1,\ldots,N. \]

\[ j=\{1, 7\}; \quad t=1,\ldots,N. \]

Table 3.10 presents linear regression estimation results for these models based on the first 2,005 observations (initial estimation set). The impact of interest rate change is more significant for lower frequency models \((j=7)\), while the estimator of oil price change is more significant for one-day-ahead forecasting. Also, order flows are more important for higher frequency forecasting. Even though it is very small, as expected, the \(R^2\) increased when individual order flows were taken into account.\(^{20}\)

\(^{20}\) This thesis uses daily data over a ten-year period (as opposed to the four-month span used by Lyons and Evans 2002); therefore, the linear "microstructure" model's \(R^2\) is significantly lower.
Table 3.10 Estimation Results For Linear Models.

<table>
<thead>
<tr>
<th>Estimates (standard error)</th>
<th>Model</th>
<th>Linear Model 1</th>
<th>Linear Model 1</th>
<th>Linear Model 2</th>
<th>Linear Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(j=1)</td>
<td>(j=7)</td>
<td>(j=1)</td>
<td>(j=7)</td>
<td></td>
</tr>
<tr>
<td>$\gamma_0 \text{ (exp 10}^5)$</td>
<td>8.09</td>
<td>36.53</td>
<td>8.72</td>
<td>36.35</td>
<td></td>
</tr>
<tr>
<td>$\beta_0 \text{ (exp 10}^5)$</td>
<td>(2.93e-05)</td>
<td>(7.08e-05)</td>
<td>(2.93e-05)</td>
<td>(7.15e-05)</td>
<td></td>
</tr>
<tr>
<td>$\Delta \text{intdiff}_{ij} \text{ (exp 10}^-4)$</td>
<td>-1.13</td>
<td>-6.76</td>
<td>-1.89</td>
<td>-6.63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00025)</td>
<td>(0.00026)</td>
<td>(0.00025)</td>
<td>(0.00026)</td>
<td></td>
</tr>
<tr>
<td>$\Delta \text{oil}_{ij}$</td>
<td>-0.0090</td>
<td>-0.003</td>
<td>-0.0087</td>
<td>-0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0065)</td>
<td>(0.0073)</td>
<td>(0.0065)</td>
<td>(0.0073)</td>
<td></td>
</tr>
<tr>
<td>$\text{aggoft}_{ij} \text{ (exp 10}^-7)$</td>
<td>-1.35</td>
<td>-3.74</td>
<td>1.33</td>
<td>-4.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.96e-08)</td>
<td>(2.16e-07)</td>
<td>(1.28e-07)</td>
<td>(3.11e-07)</td>
<td></td>
</tr>
<tr>
<td>$CC_{ij} \text{ (exp 10}^-7)$</td>
<td>-1.015</td>
<td>-2.66</td>
<td>-1.96</td>
<td>-2.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.69e-07)</td>
<td>(4.07e-07)</td>
<td>(1.69e-07)</td>
<td>(4.07e-07)</td>
<td></td>
</tr>
<tr>
<td>$IB_{ij} \text{ (exp 10}^-7)$</td>
<td>-3.86</td>
<td>-3.78</td>
<td>-3.86</td>
<td>-3.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.98e-08)</td>
<td>(2.16e-07)</td>
<td>(8.98e-08)</td>
<td>(2.16e-07)</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.0021</td>
<td>0.005</td>
<td>0.0049</td>
<td>0.0055</td>
<td></td>
</tr>
</tbody>
</table>

Two above-specified non-linear models (ANN 1 and ANN 2) were estimated by feedforward backpropagation ANNs. Figure 3.5 shows the errors related to the training set, the validation set, and the testing set. As expected, all three errors decline during the learning process. Overtraining was prevented by stopping the training process when the validation set error started to increase.
Full sample estimation of 2,230 observations was used to compare the ANN and linear model's performance. Figure 3.6 shows that the linear model forecasts in a linear fashion (for an arbitrary 90-day period), whereas the ANN forecasts more in keeping with the pattern of actual exchange rate changes.
After the initial estimation of the models in the first 2,005 observations, a set of out-of-sample forecasts was used to generate RMSEs. Each recursive re-estimation added 10 observations, so that 18 RMSEs were calculated on out-of-sample data sets ranging in size from 225 to 55 observations. This led to the selection of an ANN model 1 and ANN model 2 for one-day-ahead (j=1) and for one-week-ahead (j=7) forecasts of exchange rate changes, which were compared to linear models 1 and 2 and the random walk model.

Figures 3.7, 3.8, 3.9, and 3.10 show that the ANN can produce promising short-run forecasts, since the RMSE for the ANN model for a given forecasting horizon is equal to or below both of the competing models.
Figure 3.7 RMSE for ANN model 1, linear model 1, and random walk (j=1).

Figure 3.8 RMSE for ANN model 2, linear model 2, and random walk (j=1).
Figure 3.9 RMSE for ANN model 1, linear model 1, random walk (j=7).

Figure 3.10 RMSE for ANN model 2, linear model 2, and random walk (j=7).
Tables 3.11 (for \( j=1 \)) and 3.12 (for \( j=7 \)) list the RMSE statistics illustrated in these figures.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Model</th>
<th>Random Walk (j=1)</th>
<th>Linear Model 1 (j=1)</th>
<th>ANN Model 1 (j=1)</th>
<th>Linear Model 2 (j=1)</th>
<th>ANN Model 2 (j=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>225</td>
<td>1.4321</td>
<td>1.4364</td>
<td>1.4270</td>
<td>1.4331</td>
<td>1.4216</td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>1.4168</td>
<td>1.4201</td>
<td>1.4121</td>
<td>1.4155</td>
<td>1.4060</td>
<td></td>
</tr>
<tr>
<td>205</td>
<td>1.4232</td>
<td>1.4272</td>
<td>1.4163</td>
<td>1.4218</td>
<td>1.4119</td>
<td></td>
</tr>
<tr>
<td>195</td>
<td>1.4216</td>
<td>1.4240</td>
<td>1.4143</td>
<td>1.4188</td>
<td>1.4114</td>
<td></td>
</tr>
<tr>
<td>185</td>
<td>1.3962</td>
<td>1.4020</td>
<td>1.3871</td>
<td>1.3969</td>
<td>1.3892</td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>1.3702</td>
<td>1.3782</td>
<td>1.3580</td>
<td>1.3717</td>
<td>1.3631</td>
<td></td>
</tr>
<tr>
<td>165</td>
<td>1.3482</td>
<td>1.3564</td>
<td>1.3353</td>
<td>1.3475</td>
<td>1.3428</td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>1.3368</td>
<td>1.3442</td>
<td>1.3237</td>
<td>1.3362</td>
<td>1.3322</td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>1.3381</td>
<td>1.3464</td>
<td>1.3238</td>
<td>1.3388</td>
<td>1.3337</td>
<td></td>
</tr>
<tr>
<td>135</td>
<td>1.3639</td>
<td>1.3708</td>
<td>1.3496</td>
<td>1.3622</td>
<td>1.3622</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>1.3696</td>
<td>1.3773</td>
<td>1.3540</td>
<td>1.3691</td>
<td>1.3671</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>1.3926</td>
<td>1.4026</td>
<td>1.3739</td>
<td>1.3933</td>
<td>1.3903</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>1.3057</td>
<td>1.3109</td>
<td>1.2963</td>
<td>1.3008</td>
<td>1.2907</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>1.3335</td>
<td>1.3380</td>
<td>1.3230</td>
<td>1.3256</td>
<td>1.3133</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>1.3652</td>
<td>1.3673</td>
<td>1.3515</td>
<td>1.3578</td>
<td>1.3508</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>1.3308</td>
<td>1.3353</td>
<td>1.3163</td>
<td>1.3243</td>
<td>1.3198</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>1.3851</td>
<td>1.3882</td>
<td>1.3738</td>
<td>1.3761</td>
<td>1.3743</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>1.4168</td>
<td>1.4202</td>
<td>1.4070</td>
<td>1.4156</td>
<td>1.4155</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.12 RMSE (exp 10² ) for ANN, linear, and random walk models (j=7).

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Model</th>
<th>Random Walk (j=7)</th>
<th>Linear Model 1 (j=7)</th>
<th>ANN Model 1 (j=7)</th>
<th>Linear Model 2 (j=7)</th>
<th>ANN Model 2 (j=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>225</td>
<td></td>
<td>0.3457</td>
<td>0.3461</td>
<td>0.3458</td>
<td>0.346</td>
<td>0.3448</td>
</tr>
<tr>
<td>215</td>
<td></td>
<td>0.3467</td>
<td>0.3474</td>
<td>0.3461</td>
<td>0.3473</td>
<td>0.3462</td>
</tr>
<tr>
<td>205</td>
<td></td>
<td>0.3504</td>
<td>0.3516</td>
<td>0.35</td>
<td>0.3515</td>
<td>0.3502</td>
</tr>
<tr>
<td>195</td>
<td></td>
<td>0.3431</td>
<td>0.3435</td>
<td>0.3427</td>
<td>0.3437</td>
<td>0.3428</td>
</tr>
<tr>
<td>185</td>
<td></td>
<td>0.3358</td>
<td>0.3367</td>
<td>0.335</td>
<td>0.3368</td>
<td>0.3358</td>
</tr>
<tr>
<td>175</td>
<td></td>
<td>0.3371</td>
<td>0.3381</td>
<td>0.3364</td>
<td>0.3382</td>
<td>0.3365</td>
</tr>
<tr>
<td>165</td>
<td></td>
<td>0.3333</td>
<td>0.3336</td>
<td>0.3328</td>
<td>0.3338</td>
<td>0.3331</td>
</tr>
<tr>
<td>155</td>
<td></td>
<td>0.3377</td>
<td>0.3381</td>
<td>0.337</td>
<td>0.3382</td>
<td>0.3373</td>
</tr>
<tr>
<td>145</td>
<td></td>
<td>0.3286</td>
<td>0.3308</td>
<td>0.3253</td>
<td>0.3307</td>
<td>0.3278</td>
</tr>
<tr>
<td>135</td>
<td></td>
<td>0.3374</td>
<td>0.3396</td>
<td>0.3338</td>
<td>0.3396</td>
<td>0.3366</td>
</tr>
<tr>
<td>125</td>
<td></td>
<td>0.3441</td>
<td>0.3461</td>
<td>0.3408</td>
<td>0.3462</td>
<td>0.3438</td>
</tr>
<tr>
<td>115</td>
<td></td>
<td>0.3554</td>
<td>0.3576</td>
<td>0.3518</td>
<td>0.3577</td>
<td>0.3553</td>
</tr>
<tr>
<td>105</td>
<td></td>
<td>0.3328</td>
<td>0.3351</td>
<td>0.33</td>
<td>0.3349</td>
<td>0.3326</td>
</tr>
<tr>
<td>95</td>
<td></td>
<td>0.3384</td>
<td>0.3404</td>
<td>0.3355</td>
<td>0.3404</td>
<td>0.3384</td>
</tr>
<tr>
<td>85</td>
<td></td>
<td>0.3523</td>
<td>0.3544</td>
<td>0.3491</td>
<td>0.3544</td>
<td>0.3521</td>
</tr>
<tr>
<td>75</td>
<td></td>
<td>0.3657</td>
<td>0.3682</td>
<td>0.3619</td>
<td>0.3683</td>
<td>0.3656</td>
</tr>
<tr>
<td>65</td>
<td></td>
<td>0.3768</td>
<td>0.3787</td>
<td>0.3743</td>
<td>0.3790</td>
<td>0.3764</td>
</tr>
<tr>
<td>55</td>
<td></td>
<td>0.3882</td>
<td>0.3894</td>
<td>0.3878</td>
<td>0.3895</td>
<td>0.3868</td>
</tr>
</tbody>
</table>

The experiments show that the ANN model forecasts one-day and seven-day-ahead exchange rate changes better than the linear and random walk models. Nevertheless, the primary indicator of good forecasting power is not necessarily RMSE, but the percentage of correctly forecasted directions of real exchange rate fluctuations. In this case, the estimation involves very small values (exp 10³ ) that might result in small
RMSEs. In turn, the presence of small RMSEs is not a guarantee that the prediction is accurate, and caution is required when interpreting the estimation results.

As noted above, the percentage of correctly forecasted exchange rate direction changes or good hits (PERC) is also considered. Recursive regression for horizons between 5 and 225 observations (step 5) reveals the superiority of the ANN model.\textsuperscript{21} ANN model 1 (2) correctly predicted, on average, 60.14 per cent (61.81 per cent) of the direction of daily exchange rate movements, while linear model 1 (2) correctly predicted 57.18 per cent (58.75 per cent) of such changes, and the random walk model predicted 54.88 per cent. One-week-ahead forecasts yield worse results for ANN model 1 and linear model 1 against random walk for \( j=7 \), but ANN model 2 has the best results. Also, the predictive power of both non-random walk models is lower. Table 3.13 compares all the models used in terms of the second comparison criterion.

Table 3.13 The average percentages of correctly predicted signs for linear models 1 and 2 (LM 1 and LM 2), ANN models 1 and 2 (ANN 1 and ANN 2), and the random walk model. One-day (\( j=1 \)) and one-week (\( j=7 \)) forecasts are considered.

<table>
<thead>
<tr>
<th>Model</th>
<th>Random Walk</th>
<th>LM 1</th>
<th>ANN 1</th>
<th>LM 2</th>
<th>ANN 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( j=1 )</td>
<td>54.88</td>
<td>57.18</td>
<td>60.14</td>
<td>58.75</td>
<td>61.81</td>
</tr>
<tr>
<td>( j=7 )</td>
<td>56.26</td>
<td>54.9</td>
<td>56.15</td>
<td>55.28</td>
<td>58.04</td>
</tr>
</tbody>
</table>

\textsuperscript{21} Step 5 is used instead of step 10 to impose a more demanding setting for ANN models.
Figure 3.11 illustrates the one-day results summarized in Table 3.13. The results show that the ANN models dominate in predicting the direction of exchange rate changes one day ahead.

Next, out-of-sample, one-step-ahead forecasts are considered. More precisely, ANN model 1 (2) is initially estimated for the first 2,006 observations. The forecast errors for the remaining 225 observations (a testing set) are calculated by extending the estimation set by one and recalculating the forecast errors until the whole testing set is exhausted. This differs from the preceding forecast experiment in that the earlier experiment did not re-estimate the model up to t-1 to forecast the exchange rate at t.
RMSEs and PERCs for the one-day-ahead forecasts are listed in Table 3.14. The striking result here is that ANN 2 correctly predicts almost 72 per cent of the directions of future exchange rate changes, while the random walk model stays at about a 55 per cent accuracy.

Table 3.14 PERC and RMSE (exp 10^3) statistics for the recursive estimation over the whole testing set (k=225). ANN models 1 and 2 (ANN 1 and ANN 2) and the random walk (RW) model for one-day-ahead (j=1) forecasts are considered.

<table>
<thead>
<tr>
<th>Model</th>
<th>RW</th>
<th>ANN 1</th>
<th>ANN 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERC (%)</td>
<td>54.88</td>
<td>67.56</td>
<td>71.56</td>
</tr>
<tr>
<td>RMSE</td>
<td>1.4321</td>
<td>1.4155</td>
<td>1.3988</td>
</tr>
</tbody>
</table>

To determine the percentage of correctly predicted changes (or good hits) that relates to positive changes, the following statistic was constructed for the initial testing sample size (k=225):

\[
\text{PERC}(\text{POS}) = \frac{\text{number of positive correct responses}}{\text{number of sample positive movements}}
\]

Similarly, for negative good hits another statistic was calculated:

\[
\text{PERC}(\text{NEG}) = \frac{\text{number of negative correct responses}}{\text{number of sample negative movements}}
\]
The term "positive changes" refers to values above the mean of estimation sample changes, while "negative changes" are values below the mean value. This corrects for the fact that there is a significantly greater number of positive changes in this sample. Taking zero as a mean value would affect the reliability of the criterion, since there were mostly positive changes in the sample.

According to Table 3.15, the ANN models forecast positive and negative changes roughly equally well. In comparison, failing to correct for the positive mean change would lead to the erroneous conclusion that the model predicts positive changes much better than negative changes.

Table 3.15 PERC(POS) and PERC(NEG) for ANN models 1 and 2 (ANN 1 and ANN 2). One-day (j=1) and one-week (j=7) forecasts are considered (k=225). Percentages without normalization are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Model</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANN 1</td>
<td>ANN 2</td>
<td></td>
</tr>
<tr>
<td>PERC(POS) (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j=1</td>
<td>42.98</td>
<td>38.02</td>
<td></td>
</tr>
<tr>
<td>(89.43)</td>
<td>(80.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>j=7</td>
<td>57.02</td>
<td>53.04</td>
<td></td>
</tr>
<tr>
<td>(98.39)</td>
<td>(98.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERC(NEG) (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>j=1</td>
<td>48.08</td>
<td>67.31</td>
<td></td>
</tr>
<tr>
<td>(2.05)</td>
<td>(36.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>j=7</td>
<td>41.44</td>
<td>56.36</td>
<td></td>
</tr>
<tr>
<td>(0.99)</td>
<td>(2.97)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.4 Conclusions

With this exercise, we conduct a study of exchange rate models for the Canada/U.S. exchange rate. More specifically, we focus on their intra-day (high-frequency) and, subsequently, weekly forecast performances using a set of non-linear microstructure models.

We find strong evidence for the microstructure effects. Our horse race for forecast performance results in a non-linear ANN model as the winner. ANN models are able to significantly improve upon a simple random walk model. The daily forecasts produced by ANN models are statistically significant according to Diebold and Mariano (1995) statistics. Apart from the NN model, other linear and non-linear models are unable to generate significant predictions. The results also indicate the necessity of embodying (in a non-linear sense) information not only from interbank order flows, but from CC and FD transactions. No matter which non-linear model is used, there is always a slight forecast gain when dealer’s private order flows are included into a set of explanatory variables.

In our second experiment which includes the whole data set, the ANN is consistently better in terms of RMSE than random walk and linear models for the various out-of-sample experiments. Moreover, ANN performs on average at least 3 per cent better than other models in its percentage of correctly predicted signs. This is true for both of the forecasting horizons. As expected, more accurate forecasts are generated for the shorter forecasting window, but they are still superior to the random walk model.
Recursive one-step-ahead forecasts lead to a considerable improvement in PERC compared to the random walk and linear models.

These findings point to important implications for the models for exchange rate determination. Particularly, Lyons and Evans (2002) partial equilibrium model could be extended to encompass these non-linear and microstructure effects. Balancing the tension between microeconomic and microstructure variables is crucial. The question of which macroeconomic variables to use remains open for further research as only prices make sense for high-frequency models. Similarly, the panel of microstructure variables can be extended to a set of non-fundamentals which would account for the bandwagon effect, over-reaction to news, speculation, etc. However, these factors are not easy to quantify and we can only rely on different proxies for them. To conclude, the “proper” selection of both macroeconomic and microstructure factors can improve these results. Lastly, the power of this approach can be tested on other currencies.
Chapter 4

An Introduction to Fuzzy Logic

4.1 Introduction

Advanced modeling techniques such as artificial neural networks (ANN), fuzzy logic controllers (FLC), and genetic algorithms (GA) can be applied to a vast variety of applications in finance and economics. Through hybridization of these techniques, more complex problems can be addressed. This thesis combines two of the most popular concepts: ANN and FLC.

Fuzzy logic represents a technology for designing sophisticated control systems (Cox 1992). It provides a convenient solution to the most complex modeling requirements. Fuzzy technology in the form of approximate reasoning is a method of easily representing analog processes on a digital computer. These processes involve continuous phenomena defined by imprecise terms such as “rapidly changing,” “significant risk,” “very cost effective,” and “large capital commitment.”

The difficulty of modeling these concepts, mathematically or by sets of rules, requires a tool that is nonlinear and tolerant of imprecise data. As an example, mutual
fund managers' investment decisions (Peray 1999) are subject to the above-mentioned problems. The rules to be used for equity fund investment decision-making could take into account the following variables: economic conditions, price-to-earnings ratio (P/E ratio), and the gap from the moving average (MA gap). Continuous variables, such as these, might be characterized by a number of different states. For instance, the variable “economic conditions” can be explained by a range of states: “very bad,” “good,” “extremely good,” “solid,” etc. The transition from one state to another is not precisely defined and it is very hard to say whether an increase in GNP of two percent would cause a “fairly good” economy to change to a “good” economy. Consequently, the idea of what is “fairly good” and what is “good” is subject to different interpretations by different economists observing the same value of a variable.

The subjectivity emerging here is resolved by introducing implications, i.e., the fuzzy inference method to continuous systems modeling. Fuzzy inference interprets the values (imprecise linguistic terms such as “good” or “bad”) as the states of the variable. Then, these linguistic terms could constitute a fuzzy trading rule for the mutual fund manager such as:

IF economic conditions are bad AND the P/E ratio is high, AND the MA gap is high,

THEN the trading action is strong sell.

22 The gap from the moving average is the percent difference between the current market index level and weighted moving average. Since the weighted moving average is the trend line for the market index, the gap is thus an indicator of the market moving ahead or behind the trend.
Fuzzy logic rules follow the natural language and are close to human reasoning. They "cover" a very broad fuzzy variable range, too. That is why one fuzzy rule can be a substitute for many conventional rules. In addition, since fuzzy logic creates a control surface by combining rules and fuzzy variable states, system control can be achieved even though the mathematical behavior of the system is incomplete. To sum up, fuzzy logic can be used when dealing with continuous and imprecise variables, when a mathematical model of the process is unknown, when the input-output relation is non-linear, and finally, when the set of rules which explains input-output dependence is available.

There has been a growing literature on FLCs and their use in financial economics. Many of these examples, such as ones for portfolio management and stock market trading, can be found in Deboeck (1994). Further, Bojadziev (1997) uses FLC for evaluating a client's risk tolerance level based on his/her annual income and total net worth. Peray (1999) determines an opportunity for equity fund investments using well-established technical indicators (e.g., the gap from the moving average) and market fundamentals (GDP, inflation rate, interest rate, etc.). Also, fuzzy logic can be used to explain non-linearities in interest rates (Ju et al. 1997). In Japan, fuzzy logic is used in a foreign exchange trading system to forecast the Japan/U.S. exchange rate. This system uses fuzzy logic rules to make inferences based on economic news that may affect the currency market (Yuize 1991). Recently, Tseng et al. (2001) integrate fuzzy and ARIMA models to forecast the Taiwan/U.S. exchange rate.
4.2 Fuzzy System

The conventional control system and fuzzy system are quite alike (Cox 1993). The only difference is that a fuzzy system contains a "fuzzifier," which converts inputs into "fuzzy variables," and a "defuzzifier," which converts the output of a fuzzy control process into numerical value output. Fuzzification and defuzzification processes are explained on the next page.

In a fuzzy system the process of generating the output (control) begins with taking the inputs, fuzzifying, and then executing all the rules from the rule base which are active. Active rules' outputs are aggregated into a single output and, after defuzzification, a new output is generated (Figure 4.1; Source: Cox 1993).
Fuzzy control system design is composed of four major steps:

1. Define the model’s functional and operational characteristics: This step comprises the definition of the system architecture, data transformation requests, inputs, outputs, and the position of the fuzzy system within the overall system.

2. Define the control surfaces (fuzzification): Each control and state variable is decomposed into a set of fuzzy regions (or states) called the “fuzzy sets”. These fuzzy
sets are assigned certain names from the set N (suppose N={"strong sell", "sell", "hold", "buy", "buy strong"}) that span the variables' domains. They do not have crisp, clearly defined boundaries. In the end each fuzzy set is represented by its "membership function" (example - Figure 4.2). A membership function is a curve which defines how the points in the input space (elements of N) are mapped to a membership value (or degree of membership), a real number between 0 and 1. Mathematically, a fuzzy set $\mathcal{N}$ is defined by a set of ordered pairs (Bojadziev 1997):

$$\mathcal{N} = \{ (x, \mu_\mathcal{N}(x)) | x \in N, \mu_\mathcal{N}(x) \in [0,1] \}$$

where $\mu_\mathcal{N}(x)$ denotes a membership function.

Membership functions can be of various types: triangular, trapezoidal, Gaussian, sigmoidal, polynomial, etc. This paper uses most common shapes, triangular and Gaussian, to build a fuzzy inference system. It should be noted that larger values of a membership function indicate higher degrees of membership. Basic operations on fuzzy sets (A and B) are the following:

- equality (A=B if and only if for every x, $\mu_A(x) = \mu_B(x)$ ),
- inclusion (A is included in B if and only if for every x, $\mu_A(x) \leq \mu_B(x)$ ),
- proper subset (A is a proper subset of B if and only if $\mu_A(x) < \mu_B(x)$ for every x, and $\mu_A(x) < \mu_B(x)$ for at least one x),
- complementation (A and B are complementary if $\mu_A(x) = 1 - \mu_B(x)$ ).
3. Define behavior of control surfaces (rules): To define a link between the input and output variables a rule base is created. Linguistic rules are of the form:

$$IF \ <x \ is \ A> \ THEN \ <y \ is \ B>.$$ 

where \(x\) and \(y\) are scalar variables and \(A\) and \(B\) are linguistic values defined by fuzzy sets. The phrase "\(x\) is \(A\)" is called the antecedent or premise, while "\(y\) is \(B\)" is called the

\[\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x))],\]

\[\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x))].\]
consequent or conclusion. Fuzzy rules form a fuzzy rule base. The number of rules varies with the number of control variables. The idea is to try to identify all the possible combinations of inputs. Thus, if there are three input variables, each described by five fuzzy sets, the required number of rules would be $5^3 = 125$. Fuzzified inputs cause some rules to be activated and to contribute to an overall output which is calculated using the so called “Mamdani inference”\footnote{See Mamdani (1975).}. Mamdani inference applies min and max operators for fuzzy AND (intersection) and OR (union) operators. It also requires that the output membership function is a fuzzy set (unlike “Sugeno inference” where the output is either constant or linear). To illustrate this, suppose there is one numeric fuzzy input $u_0$ ($u_0 \in U$) with the corresponding fuzzy membership function $\mu_{u}(u)$. The output variable has a membership function denoted by $\mu_{v}(v)$, $v \in V$. The $i$th active rule ($i=1,\ldots,I$) would produce the following output:

$$\alpha_i = \mu_{u_i}(u_0) \land \mu_{v_i}(v) \quad (4.1)$$

The overall output $Z$ is a union of all membership functions ($\alpha_i$s):

$$Z = \bigcup_{i=1}^{I} \alpha_i \quad (4.2)$$

\footnote{See Mamdani (1975).}
4. Select the defuzzification method: The last step to fuzzy system design is to choose between several defuzzification methods and extract a discrete value $z$ from $Z$. Perhaps the most popular and frequently used methods are the composite maximum and the composite moment or centroid of area (or centre of gravity). Centroid of area returns the point where the aggregate fuzzy set $Z$ under the curve is sliced into two equal masses (Jang and Sun 1995):

$$z_{COA} = \frac{\int z \mu_W(z) \, dz}{\int \mu_W(z) \, dz} \quad (4.3)$$

where $\mu_W(z)$ is the aggregated output membership function.

The whole process of a fuzzy decision-making system design for the FX trading will be described in greater detail in Chapter 5.
Chapter 5

Neuro-Fuzzy Decision-Making in Foreign Exchange Trading and Other Applications

5.1 Neuro-Fuzzy Design

The ANNs have been very successful in recognizing nonlinear patterns from noisy, high-frequency data and have been very useful forecasting tools. However, there has been criticism of their inability to transparently explain how a decision (forecast) is reached. Also, unlike fuzzy logic it is impossible to incorporate a priori information about the problem into the ANN system.

Fuzzy logic can quantify vague information and produce transparent decision-making logic. The drawbacks of fuzzy logic are the lack of a learning capability and the necessity for an expert knowledge about the system.

The aim of neuro-fuzzy (NF) systems is to combine the benefits of both approaches. Such a combination can be achieved in several ways, which include: neural fuzzy inference systems (Jang 1993), fuzzy neural networks (Gupta and Rao 1994), and neuro-fuzzy combination. Neural fuzzy inference systems introduce a parallel architecture and learning capability to a fuzzy inference system. Each fuzzy rule is created using the
ANN and it is a data driven process. Fuzzy neural networks embed fuzzy logic into the ANN by fuzzifying the learning algorithms. This paper uses a neuro-fuzzy combination where the ANN is used for identification (forecasting) and FLC extracts the decision from the ANN's output.

Finding a trading rule with FLC should not be confused with training of an ANN. An ANN uses a learning algorithm to map input variables (e.g., lagged interest rate, lagged order flow) into output variables (e.g., exchange rate change). In other words, an ANN is a nonlinear and dynamic system that learns from known input-output combinations. ANNs have been shown to have very good forecasting ability, but lack explanatory capability. By contrast, FLCs have no training capability and the mapping between inputs and output is generated from expert knowledge in the form of “if-then” rules. That is why it would be ideal to combine ANNs and FLCs to create a so-called neuro-fuzzy (NF) technology (Jang and Sun 1995, Nauck and Kruse 1997).

In this study the ANN estimates one-day-ahead forecasts (Gradojevic and Yang 2000) of the magnitude of the exchange rate change (positive or negative).24 Afterwards, that value is fuzzified on the interval $[-10^{-3}, 10^{-3}]$ that contains all of the FX rate change.

\[
\text{Model 1: } \Delta \text{rpfx}_t = \phi (\Delta \text{intdiff}_{t-1}, \Delta \text{oil}_{t-1}, \text{aggof}_{t-1}) + \epsilon_i; \\
\text{Model 2: } \Delta \text{rpfx}_t = \gamma (\Delta \text{intdiff}_{t-1}, \Delta \text{oil}_{t-1}, \text{CC}_{t-1}, \text{IB}_{t-1}, \text{FD}_{t-1}) + \nu_i;
\]

where, as before, $\Delta \text{rpfx}$ is a daily change of the logarithm of real Canada/US exchange rate, $\Delta \text{intdiff}$ is the change in the differential between the Canada-US nominal 90-day commercial paper interest rate, $\Delta \text{oil}$ is the daily change in the logarithm of the crude oil price, $\text{aggof}$ is aggregate order flow, and $\text{CC}$, $\text{IB}$, $\text{FD}$ are individual order flows. The data set covers the period between January 1990 and June 2000.

---

24
forecasts. We choose Gaussian-bell over triangular shaped membership functions since they produce a relatively smooth input/output mapping. They are defined by its mean and standard deviation. These parameters are arbitrary set to slice the variable domain into overlapping Gaussian functions that have the same shape and the highest degree of membership for the mean value. The following functions are assigned to each exchange rate change state (fuzzy set): “VERY NEGATIVE (V-NEG),” “NEGATIVE (NEG),” “MEDIUM-NEGATIVE (M-NEG),” “WEAKLY NEGATIVE (W-NEG),” “STABLE,” “WEAKLY POSITIVE (W-POS),” “MEDIUM-POSITIVE (M-POS),” “POSITIVE (POS),” and “VERY POSITIVE (V-POS)” (Figure 5.1.a).

Figure 5.1.a. Gaussian membership functions for the variable “exchange rate change forecast”.

![Gaussian membership functions](image-url)
The FX trader's action or trading strategy is also a fuzzy variable with the triangular fuzzy membership functions that represent five fuzzy sets between -1 and 1: "VERY STRONG SELL (VS-SELL)," "STRONG SELL (S-SELL)," "MEDIUM SELL (M-SELL)," "WEAK SELL (W-SELL)," "HOLD," "WEAK BUY (W-BUY)," "MEDIUM BUY (M-BUY)," "STRONG BUY (S-BUY)," "VERY STRONG BUY (VS-BUY)." These linear-shaped functions (Figure 5.1.b) can be interpreted as the "fuzzy utility functions" of a risk-neutral investor (the degree of satisfaction to the decision-maker with respect to the particular objective, e.g., expected target rate of return). These linear-shaped functions (Figure 5.1.b) can be interpreted as the "fuzzy utility functions" of a risk-neutral investor (the degree of satisfaction to the decision-maker with respect to the particular objective, e.g., expected target rate of return). These linear-shaped functions (Figure 5.1.b) can be interpreted as the "fuzzy utility functions" of a risk-neutral investor (the degree of satisfaction to the decision-maker with respect to the particular objective, e.g., expected target rate of return). Fuzzy rule base contains nine simple and intuitive IF-THEN type rules that link a fuzzy input (fuzzified exchange rate change forecast) and the FX trader's action:

25 See Ramaswamy (1998) for further discussion on "fuzzy utility functions."
• IF “exchange rate change forecast” is “V-NEG” THEN “FX trader’s action” is “VS-SELL”.
• IF “exchange rate change forecast” is “NEG” THEN “FX trader’s action” is “S-SELL”.
• IF “exchange rate change forecast” is “M-NEG” THEN “FX trader’s action” is “M-SELL”.
• IF “exchange rate change forecast” is “W-NEG” THEN “FX trader’s action” is “W-SELL”.
• IF “exchange rate change forecast” is “STABLE” THEN “FX trader’s action” is “HOLD”.
• IF “exchange rate change forecast” is “W-POS” THEN “FX trader’s action” is “W-BUY”.
• IF “exchange rate change forecast” is “M-POS” THEN “FX trader’s action” is “M-BUY”.
• IF “exchange rate change forecast” is “POS” THEN “FX trader’s action” is “S-BUY”.
• IF “exchange rate change forecast” is “V-POS” THEN “FX trader’s action” is “VS-BUY”.

Figure 5.2 shows how the FX trading strategy is determined.

<table>
<thead>
<tr>
<th>Figure 5.2. NF system for the FX trading. Two nonlinear models are used for forecasting.</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Diagram" /></td>
</tr>
</tbody>
</table>

102
NF estimation process begins after the introduction of the new input to the trained ANN which produces the logarithm of the exchange rate change forecast on its output. This is illustrated on Figure 5.3 where, using the notation from chapter 4, the ANN output is assumed to be \( u_0 = 3.9 \cdot 10^{-4} \). In the next stage, this output becomes a FLC input and when it is read, each rule that has any truth in its premise will be active. In this example, rules 7 and 9 are active. The rules use Mamdani inference as it is intuitive and well-suited to human input. Moreover, our computational requirements are very modest and there is no need for the Sugeno method which is more computationally effective. When the rule is active, its output will contribute towards the overall FLC output whereas the min operator is applied (equation 4.1). Two resulting membership functions (\( \alpha_1 = \mu_{\ell_1}(u_0) \land \mu_{\ell_1}(v) \) and \( \alpha_2 = \mu_{\ell_2}(u_0) \land \mu_{\ell_2}(v) \)) for each rule are combined together with the max (or logical “and”) operator (equation 4.2). Finally, the combined membership function \( Z \), i.e., the upper bound of the shaded area, is defuzzified with “centroid method” (equation 4.3) into a discrete value \( z = 0.417 \) from the \([-1,1]\) interval, the domain of the output membership function. This is a very reasonable estimate given the shape of \( Z \) and we assume that other defuzzification methods would not bring significant advantages.
By using the NF technology, we not only determine the FX trading strategy, but the traded volume too. We assess the traded volume using Table 5.1 and based on the shape of the membership functions for the output variable.
Table 5.1. The intervals for discrete decisions on traded volume (based on a defuzzified output z).

<table>
<thead>
<tr>
<th>Trading strategy</th>
<th>Interval for z</th>
</tr>
</thead>
<tbody>
<tr>
<td>VERY STRONG (VS) SELL-invest 80% of the U.S. currency</td>
<td>-1 ≤ z &lt; -0.875</td>
</tr>
<tr>
<td>STRONG (S) SELL-invest 60% of the U.S. currency</td>
<td>-0.875 ≤ z &lt; -0.625</td>
</tr>
<tr>
<td>MEDIUM (M) SELL-invest 40% of the U.S. currency</td>
<td>-0.625 ≤ z &lt; -0.375</td>
</tr>
<tr>
<td>WEAK (W) SELL-invest 20% of the U.S. currency</td>
<td>-0.375 ≤ z &lt; -0.125</td>
</tr>
<tr>
<td>HOLD-do not invest</td>
<td>-0.125 ≤ z ≤ 0.125</td>
</tr>
<tr>
<td>WEAK (W) BUY-invest 20% of the Canadian currency</td>
<td>0.125 &lt; z ≤ 0.375</td>
</tr>
<tr>
<td>MEDIUM (M) BUY-invest 40% of the Canadian currency</td>
<td>0.375 &lt; z ≤ 0.625</td>
</tr>
<tr>
<td>STRONG (S) BUY-invest 60% of the Canadian currency</td>
<td>0.625 &lt; z ≤ 0.875</td>
</tr>
<tr>
<td>VERY STRONG (VS) BUY-invest 80% of the Canadian currency</td>
<td>0.875 &lt; z ≤ 1</td>
</tr>
</tbody>
</table>

Since the output value z is always a number from [-1,1] interval, Table 5.1 uniquely defines the fraction of the current period endowment for either of the strategies that should be traded. In other words, a positive value indicates to buy the US currency using the fraction of domestic cash endowment from Table 5.1, while a negative value indicates a fraction of foreign currency to be sold. In this example, the trading rule ("MEDIUM BUY") implies buying the foreign currency using a fraction 0.40 of the current cash endowment. This paper aims to compare the NF daily trading strategy recommendation with the simple buy-and-hold strategy over certain periods. Lastly, it is important to note that a more complex FLC setting could probably achieve the optimal trading strategy estimator. However, we apply a very basic setting and leave this to further research.
5.2 Market Environment

Much of the previous trading rules literature has sought to test whether particular kinds of technical trading rules have useful forecasting ability. There are a few differences between technical trading rules studies and the NF trading rules. First, the searching space is different. For technical trading rules the searching space consists only of past prices, while in this paper the searching space of NF technology contains fundamental variables, such as interest rate differential, and market order flow variables. Second, the question of which of those trading rules are important is subtly different. Instead of checking whether specific rules work, this research tries to find out whether an optimized ANN combined with FLC can be used to generate a single trading strategy. Third, fuzzy logic is employed in this paper to link the forecasting value to a decision space. One advantage of fuzzy logic is that it can make a decision space more continuous. Instead of generating discrete “buy” or “sell” signals, fuzzy logic provides much finer trading decisions, for example, “strong buy,” (invest 60 per cent of the Canadian currency endowment) “strong sell,” (invest 60 per cent of the U.S. currency endowment) and “hold” (preserve the current position) etc.

In this paper two trading strategies are compared, namely the NF and buy-and-hold strategies. At the beginning of each period, currency dealers using both strategies are endowed with the same portfolio, which consists of M U.S. dollars, denoted $M_{USD}$ and

---

M Canadian dollars, denoted $M_{CAD}$. The NF method generates a daily trading signal, to buy a certain amount of foreign exchange using Canadian currency or to sell a certain amount of the foreign currency in the portfolio. Otherwise, the dealer's current position is preserved. The trading strategy from fuzzy logic specifies the position to be taken. The trading signal generated by FLC is located between -1 and 1, which suggests an asset position in the dealer's portfolio.  

We assume that the scenario where the NF strategy is compared to buy-and-hold benchmark strategies involves a lump sum initial endowment, and, if the currency is held, there is no reinvestment, i.e., it does not earn the overnight interest rate (domestic or foreign). The NF strategy competes with a conventional buy-and-hold strategy which suggests holding all of the lump sum initial endowments in U.S. dollars until the end of the period called the moving window. The rate of return at time $t$, $r_t$, for the buy-and-hold strategy is calculated from the amount held in Canadian dollars and according to the market price after each transaction as follows:

$$r_t = \frac{M_{USD} \cdot S_t + \left( \frac{S_{t}}{S_{t-W}} \cdot M_{CAD} \right)}{M_{USD} \cdot S_{t-W} + M_{CAD}} - 1$$

where $W$ is the size of the moving window and $S_t$ is the real Canada/U.S. exchange rate at time $t$.

---

$^{27}$ This trading strategy implies the assumption of risk neutrality.
Similarly, for the sequence of NF strategies over the window of size $W$, the rate of return ($r'_t$) is calculated by

$$r'_t = \frac{NF_{USD} \cdot S_t + NF_{CAD}}{M_{USD} \cdot S_{t-W} + M_{CAD}} - 1$$

where $NF_{CAD}$ and $NF_{USD}$ are the amounts of Canadian and U.S. dollars in the dealer’s portfolio position after $W$ days, respectively.

The rates of return defined above for the buy-and-hold strategies are compared to the day-to-day NF-guided trading strategy rate of return. Comparisons are conducted both with the transaction cost and without the transaction cost. The transaction cost imposed as a linear monotonic increasing function of the order size. For example, for a conventional size of a 5 million order, the transaction cost is 5 basis points. To calculate the returns for the buy-and-hold strategy, it is applied, first, when buying U.S. currency, and then, when converting back to Canadian currency. The NF strategies are subject to transaction costs every time the trade occurs, i.e., the “hold” signal is not received.

For the robustness of the results, the same comparison is conducted for a moving window with a different length of the periods. The results for 10-day, 20-day, 30-day, 40-day and 80-day moving window are reported in the next section.

---

28 Points are the smallest moves an exchange rate can make, i.e., digits added to or subtracted from the fourth decimal place.
5.3 Estimation Results

The major purpose of this study is to investigate how well the sequence of NF daily trading strategy recommendations performs against the buy-and-hold strategy over certain periods. The empirical results are discussed in this section. The focus is on the one-day-ahead out-of-sample analysis. Therefore, the NF-based strategy recommendation is re-estimated every time the observed window is shifted towards the end of the sample.

Observation windows are called moving windows because, in order to test the robustness of the NF technology, the trading period (window) was continuously moved (100 times) forward across the data set. For each moving window the winning strategy (in terms of the rate of return) was recorded. Thus, we measure the excess return \((r'_t - r_t)\) of the NF strategy on each of the 100 trading periods or windows. Table 5.2 contains the estimated number of windows (out of 100) when the NF strategy was superior to the simple buy-and-hold strategy.
Table 5.2. The number of moving windows when the NF technology outperforms a simple buy-and-hold strategy. Two models are considered: model 1 (with aggregate order flow) and model 2 (with individual order flows) with and without transaction costs.

<table>
<thead>
<tr>
<th>Window size (days):</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of winning windows:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>44</td>
<td>51</td>
<td>60</td>
<td>69</td>
<td>84</td>
</tr>
<tr>
<td>Model 2</td>
<td>46</td>
<td>53</td>
<td>66</td>
<td>72</td>
<td>75</td>
</tr>
<tr>
<td>Model 1 (with tr. costs)</td>
<td>46</td>
<td>48</td>
<td>55</td>
<td>67</td>
<td>80</td>
</tr>
<tr>
<td>Model 2 (with tr. costs)</td>
<td>50</td>
<td>57</td>
<td>66</td>
<td>72</td>
<td>74</td>
</tr>
</tbody>
</table>

In the absence of the transaction costs, it is quite evident that when the size of the moving window increases from 10 to 80, both of the NF models improve and perform better than the buy-and-hold strategy. They produce higher rates of return when the window size is 20 or greater on more than 50 per cent of the windows. This can be explained by the fact that when the window size is bigger the NF technology has more opportunities to compensate for its errors. Further, the inclusion of the transaction costs does not significantly reduce the number of winning windows. In some cases such as model 2 with the window size of 20, the percentage of winning windows increases (from 53 to 57). The explanation for this is that we define the NF technology (or more precisely, fuzzy membership functions) to produce a "hold" signal for an arbitrarily small exchange rate change forecast. For instance, there can be sequences that contain mostly "holds" and "weak-buys" or "weak-sells" with no or very little transaction costs.
incurred. On the other hand, buy-and-hold strategies involve trading the whole initial Canadian dollar endowment and applying the transaction cost would penalize the rate of return to the greater extent.

Figures 5.4 (a, b, c, d) present rates of return for a 10-day moving window. The first moving window is initially estimated for 2010 observations and one-day-ahead forecast and the resulting strategy are generated. This continues for the whole window (while increasing the estimation set by one) and the last action to be generated is based on 2019 observations. Subsequently, the moving window is shifted forward 100 times, these windows are indexed from 1 to 100, and the corresponding 100 rates of return are calculated and compared with the buy-and-hold rates of return. We also calculate excess returns for the NF strategies. Similarly, for 20-day moving windows, the rates of return are shown in Figures 5.5 (a, b, c, d).

Figure 5.4.a. NF and the buy-and-hold rates of return (ROR) for the 10-day moving-window (model 2 without the transaction costs) and NF excess returns.

The mean excess return is -0.00014.
Figure 5.4.b. NF and the buy-and-hold rates of return (ROR) for the 10-day moving-window (model 2 with the transaction costs) and NF excess returns.

The mean excess return is 0.00041.

Figure 5.4.c. NF and the buy-and-hold rates of return (ROR) for the 10-day moving-window (model 1 without the transaction costs) and NF excess returns.

The mean excess return is -0.0003.
Figure 5.4.d. NF and the buy-and-hold rates of return (ROR) for the 10-day moving-window (model 1 with the transaction costs) and NF excess returns.

The mean excess return is $-0.00008$.

Figure 5.5.a. NF and the buy-and-hold rates of return (ROR) for the 20-day moving-window (model 2 without the transaction costs) and NF excess returns.

The mean excess return is $0.00047$. 
Figure 5.5.b. NF and the buy-and-hold rates of return (ROR) for the 20-day moving-window (model 2 with the transaction costs) and NF excess returns.

The mean excess return is 0.00074.

Figure 5.5.c. NF and the buy-and-hold rates of return (ROR) for the 20-day moving-window (model 1 without the transaction costs) and NF excess returns.

The mean excess return is 0.00016.
The mean excess returns for the NF strategy are increasing with the window size in absence of transaction costs (model 1: $-0.0003$ for 10-day window, $0.00016$ for 20-day window, $0.0014$ for 30-day window, etc.; model 2: $-0.00014$ for 10-day window, $0.00047$ for 20-day window, $0.0018$ for 30-day window, etc.) and when transaction costs are included this still holds (model 1: $-0.00008$ for 10-day window, $-0.00027$ for 20-day window, $0.00024$ for 30-day window, etc.; model 2: $0.00041$ for 10-day window, $0.00074$ for 20-day window, $0.0018$ for 30-day window, etc.). As expected, model 1 is dominated by model 2 for the most of the windows and different window sizes. The longer, 20-day window implies higher mean excess returns without transaction costs for model 1, but surprisingly, as observed before, model 2 seems to be unaffected by the transaction costs.
This relationship persists for the various window sizes. To conclude, NF technology can be successful only on certain ranges and any FX trader attempting to use NF-advised dynamic strategies should be very cautious.

5.4 Excess Volatility and the NF Model

The issue of exchange rate overshooting is of deep concern to the Bank of Canada in that it may be the result of, or be manifested as, excess volatility in the FX market. Overshooting is viewed as being due to the prevalence of speculative, noise, or chartist traders. In periods of excess volatility, central banks, like the Bank of Canada, worry that FX volatility will spill over into domestic fixed income markets and in turn impact the real economy. Moreover, there is a concern that even a sudden depreciation in the currency, that was initially predicated on fundamentals, may become self-perpetuating and in turn lead to excess volatility (and excess depreciations). As noted in Murray et al. (2000), in periods where it is believed that overshooting is occurring, it may be necessary for the Bank of Canada to raise official rates in order to offset the self-perpetuating sentiment that exist in the market. In raising official rates (and in turn hopefully calming FX markets) the Bank hopes to avoid a more dramatic tightening of monetary conditions that may occur if market uncertainty (and volatility) continued to permeate.

The Bank of Canada has over the 1970s, 1980s, and 1990s on repeated occasion had to engage in such behaviour, the latest example being in August of 1998 when the
Bank of Canada raised rates by 1 per cent. During this most recent period, the Canadian dollar depreciation, that began rather slowly in mid-summer, started accelerating soon after the Russian ruble crisis.

It is rather difficult to know when a currency is overshooting its “fundamental” (equilibrium) value or when its fluctuation are excessive, since on a daily basis it is hard to ascertain what is the fair value of the currency.\(^{29}\) Moreover, in order for the Bank of Canada to appropriately apply contractionary (but FX stabilizing) medicine, it needs to know when self-perpetuating or destabilizing market sentiments are taking hold. Thus, models that are able to signal the advent of destabilizing trading activity are useful to the central bank. One such model developed at the Bank is the fundamentalist versus chartists model of Murray, van Norden, and Vigfusson (1996), in which Canadian dollar FX movements can be attributed to either fundamental based traders, who try to keep the exchange rate close to its true equilibrium value, or chartists traders, who often cause the exchange rate to deviate from its fair market value, using a Markov-switching econometric model. The fundamentalists are assumed to determine the Canada/U.S. real exchange rate \((rfx)\) from the following equation:

\[
\Delta \ln(rfx)_t = \alpha(\ln(rfx)_{t-1} - \beta_0 - \beta_c \text{comtot}_{t-1} - \beta_e \text{enetot}_{t-1}) + \gamma \text{intdif}_{t-1} + \varepsilon_t \tag{5.1}
\]

\(^{29}\) In line with Murray, van Norden, and Vigfusson (1996), we conjecture that the “fundamental” forces that drive the exchange rate are determined by the equation (5.1). It is important to stress that these forces are often very difficult to find.
where *comtot* denotes non-energy commodity terms of trade, *enetot* denotes energy terms of trade and *intdif* is Canada/U.S. interest rate differential. To the contrary, chartists have very little regard for fundamental variables and are assumed to base their exchange rate forecasts and trading strategies on technical indicators such as momentum.

Further, the expected exchange rate change can be modeled as a weighted average of both groups' expectations as follows:

$$E\Delta s_t = \omega_t E\Delta s^f_{t+1} + (1 - \omega_t) E\Delta s^c_{t+1} \quad (5.2)$$

The weights, $\omega_t$ and $1 - \omega_t$, assigned to each group are determined by a portfolio manager who favours the group that was most successful in the latest period.

$\Delta s^f_{t+1}$ and $\Delta s^c_{t+1}$ are determined by two forecasting equations:

$$\Delta s^f_t = \alpha^f + \phi(s_{t-1} - \bar{s}_{t-1}) + \gamma intdif_{t-1} + \epsilon^f_t, \quad \epsilon^f_t \sim \mathcal{N}(0, \sigma^f_t) \quad (5.3)$$

$$\Delta s^c_t = \alpha^c + \psi_{14} ma_{14} + \psi_{200} ma_{200} + \Gamma intdif_{t-1} + \epsilon^c_t, \quad \epsilon^c_t \sim \mathcal{N}(0, \sigma^c_t) \quad (5.4)$$

where $s$ is the logarithm of the nominal Canada/U.S. exchange rate, $\bar{s}$ is fundamentalists' forecast of $s$ based on (5.1), $ma_{14}$ and $ma_{200}$ are moving averages used as chartists' technical indicators, $\alpha^f$ and $\alpha^c$ are constants, and superscripts $f$ and $s$ stand for fundamentalists and chartists, respectively.
The probabilities that expected values of the exchange rate change will persist in fundamentalists' or chartists' regime (state) R at time t given that it was in that state at time t-1 are the following:

\[ \pi(R_t = f|R_{t-1} = f) = \Phi(\alpha_f) \]

\[ \pi(R_t = c|R_{t-1} = c) = \Phi(\alpha_c) \]

where \( \Phi \) is the normal cumulative density function.

The portfolio manager's objective is to maximize the following log-likelihood function:

\[ LLF = \sum_{i=1}^{T} \sum_{R_i} \pi(R_t)d(s_t|R_t) \]

where \( d(s_t|R_t) \) is the normal density function of the regime's residual.

In this model, it is shown that during tranquil periods, chartists are the dominant traders while during volatile periods it is the fundamentalist traders that drive FX price movements. This is somewhat counterintuitive, but is explained as the result of fundamentalist traders only participating in the market when they view the price of the currency as being sufficiently far from their view of what it should be, while at other times they choose to not participate (in a dominant manner). However, since chartists can in essence be viewed as being inactive during periods where fundamental traders are
dominant, discerning what signals the chartists are receiving during periods of sharp depreciations (attributed to fundamental traders) or what direction these traders would trade in, is not examined in this framework.

The NF model presented above can be viewed as also shedding light on activities of FX market participants. Specifically, given that the NF model is essentially a very sophisticated technical trading model, it allows one to understand what type of signals the chartist traders are getting during periods of Canadian dollar overshooting (excess depreciation). It would in turn shed light on how chartists were trading during periods that are thought to be periods where the Canadian dollar has overshot. In the case when one is uncertain whether the FX rate actually overshoots its fundamental value or not, it is important to note that the NF model could provide some indication that expectations are strong in one direction. To put it another way, this NF model allows one to get a partial assessment of the FX market's sentiment, as reflected by the chartists' trading signals, during periods of strong depreciations. As such, this model complements the previous work examining chartists and fundamentalist by providing information on the likely activity carried out by chartists during periods where fundamentalists are seen to dominate market activity.

In order to examine what is the sentiment of chartists during periods of strong depreciation, we review past episodes in which the Bank of Canada viewed the market as having destabilizing or self-perpetuating expectations. The most obvious one being the August 1998 periods in which the Bank of Canada intervened heavily in the FX market.
We choose to assign the following five fuzzy sets to the exchange rate change states: "VERY NEGATIVE," "NEGATIVE," "STABLE," "POSITIVE," and "VERY POSITIVE" (Figure 5.6). Similarly, the trading strategy variable has five states as follows: "SELL STRONG," "SELL," "HOLD," "BUY," and "BUY STRONG" (Figure 5.7). These states uniquely define the trading signal received by a currency trader.

Figure 5.6 Gaussian membership functions for the variable "exchange rate change forecast."
In accord with Section 5.1, there are five simple rules that link a fuzzy input and an output:

- IF "exchange rate change forecast" is "VERY NEGATIVE" THEN "FX trader's action" is "SELL STRONG".
- IF "exchange rate change forecast" is "NEGATIVE" THEN "FX trader's action" is "SELL".
- IF "exchange rate change forecast" is "STABLE" THEN "FX trader's action" is "HOLD".
- IF "exchange rate change forecast" is "POSITIVE" THEN "FX trader's action" is "BUY".
- IF "exchange rate change forecast" is "VERY POSITIVE" THEN "FX trader's action" is "BUY STRONG".

The other specifics of the NF design and its implementation in this section are the same as in Section 5.1. It is also worthwhile to note that we do not define any
portfolio decisions in this Section. Rather, we try to differentiate the types of signals the traders receive. For that purpose we use Table 5.3.

Table 5.3. The intervals for discrete trading signals the traders receive from the NF system (based on a defuzzified output z).

<table>
<thead>
<tr>
<th>Trading strategy</th>
<th>Interval for z</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELL STRONG</td>
<td>-1 ≤ z &lt; -0.75</td>
</tr>
<tr>
<td>SELL</td>
<td>-0.75 ≤ z &lt; -0.25</td>
</tr>
<tr>
<td>HOLD</td>
<td>-0.25 ≤ z &lt; 0.25</td>
</tr>
<tr>
<td>BUY</td>
<td>0.25 &lt; z ≤ 0.75</td>
</tr>
<tr>
<td>BUY STRONG</td>
<td>0.75 &lt; z ≤ 1</td>
</tr>
</tbody>
</table>

The ANN (model 2) is trained based on the first 1823 observations, i.e. until the end of June 1998. The remaining 34 observations, which cover July and August, are used for forecasting (testing) and generating NF system’s outputs. The ANN part of a NF system generates 34 forecasts (Table 5.4). The forecasting accuracy of an ANN is very high: roughly 82% of the overall changes are successfully forecasted, i.e., only six negative changes are missed.

Table 5.4. Observed ANN (model 2) direction-of-change forecasting statistics (frequencies) for July/August 1998 inputs (34 observations).

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive forecast</td>
<td>31</td>
</tr>
<tr>
<td>Actual positive</td>
<td>25</td>
</tr>
<tr>
<td>Negative forecast</td>
<td>3</td>
</tr>
<tr>
<td>Actual negative</td>
<td>9</td>
</tr>
</tbody>
</table>
Further, ANN forecasts are input into FLC to generate a sequence of 34 trading strategies. For this testing set the results are strikingly different from an average 30-day trading recommendations sequence (for a set of arbitrary 30-day windows without sharp increases or declines of the exchange rate) which consists of 11 buy (approximately 33%), 14 hold, and 5 sell signals. More precisely, NF generated 20 “buys” (approximately 60%), 14 “holds”, and not a single sell signal.

Another sharp exchange rate depreciation (5 cents) concerns November/December 1994 forecasting period which consists of 27 observations: 18 positive and 9 negative exchange rate changes. ANN statistics are reported in Table 5.5. High forecasting accuracy is maintained for this testing set and it is 81.5%. One positive and four negative changes are not successfully forecasted.

Table 5.5. Observed ANN (model 2) direction-of-change forecasting statistics (frequencies) for November/December 1994 inputs (27 observations).

<table>
<thead>
<tr>
<th>Frequency</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive forecast</td>
<td>21</td>
</tr>
<tr>
<td>Actual positive</td>
<td>18</td>
</tr>
<tr>
<td>Negative forecast</td>
<td>6</td>
</tr>
<tr>
<td>Actual negative</td>
<td>9</td>
</tr>
</tbody>
</table>

Again, the NF technology generated a similar “alert” set of signals: 14 “buys” (51%), 12 “holds” and one “sell”. The only sell signal is based on wrongly forecasted positive exchange rate movement. One could argue that in this case there is 10% less
buy signals generated by NF system and that there exists a wrong sell signal (whereas in July/August 1994 there are no sell signals). But, one must not forget that in the first case there was a sharper, 10-cent depreciation and that ANN estimation set in 1994 is twice shorter. The ANN's forecasting power strongly depends on the amount of data available for training and the lack of approximately 700 observations lowers that power.

This Section indicates that the NF system which generates discrete trading signals can play an important role in detecting strong and potentially dangerous (from the point of view of the central bank) sentiments in the FX market. These signals are assumed to characterize the possible activity carried out by chartists during periods of excessive depreciation where fundamentalists are seen to dominate market activity. Our analysis also suggests that the accuracy of the NF system is largely linked to the amount of data used for the ANN training.

5.5 “Continuous” Trading Decisions and a Risk-Averse Investor

In this section, we compare NF-generated strategies and decisions (strategies) made by a risk-averse investor, based exclusively on ANN forecasts. FLC provides a relatively smooth input/output surface. We assume that a risk-averse investor attaches a smooth non-linear tan-sigmoid function to ANN forecasts (Figure 5.8).
Figure 5.8. Tan-sigmoid function used to estimate the risk-averse investor's strategies.

\[ f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \]

Investment

This function establishes a familiar framework for determining a fraction of the investor's endowment to be traded. Using the setting from Section 5.4, an endowment fraction to be invested \((z \in [-1, 1])\) is received for every forecast, but unlike in Section 5.1, this fraction is directly applied to the current position. For example, \(z=0.417\) would be interpreted as a signal for buying the foreign currency with 41.7 per cent of the current domestic currency endowment. This can be thought of as an attempt to investigate if we can make finer and less discrete trading decisions than those defined in Section 5.1. Here, we proceed by recording and comparing the number of moving windows where one of the approaches generates higher returns. Of central interest is whether one of the strategies can consistently earn higher returns over different window sizes.
Table 5.6. The number of moving windows when the NF technology outperforms risk-averse investor’s strategies. ANN model 2 is considered (with individual order flows and without transaction costs) on 100 moving windows.

<table>
<thead>
<tr>
<th>Window size(days)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Number of winning windows:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td>54</td>
<td>46</td>
<td>32</td>
<td>25</td>
</tr>
</tbody>
</table>

The results (Table 5.6) show very clearly that it can, even though for small window sizes (5 and 10 days) both approaches produce similar returns. A risk-averse investor’s strategies have more substantial percentage advantage for 20-day, 30-day, and 40-day trading periods. The results suggest that fuzzy logic and its “relatively smooth” decision surface cannot outperform a smooth non-linearity such as tan-sigmoid in this simple one input/one output application. This indicates that a slightly different and simpler approach can produce better results and there are no grounds for “continuous” trading decisions. In a scenario involving the FLC with multiple inputs, given the difficulty to select the suitable multidimensional non-linearity, we anticipate the fuzzy logic approach to be more useful. ANN forecasts and some technical trading indicators as inputs can serve this purpose, but we leave this to further research.
5.6 Conclusions and Further Research

Exchange rate forecasting using artificial neural networks (ANN), fuzzy logic controllers (FLC) and genetic algorithms (Goldberg 1989) has recently received much attention. This chapter proposes a neuro-fuzzy (NF) technology to learn a single trading rule that specifies both the action (buy/hold/sell) and volume of currency to be traded. ANN produces one-day-ahead-forecasts of the Canada/US dollar exchange rate change based on lagged aggregate order flow (model 1)/individual order flows (model 2), interest rate, and crude oil price. Buy-and-hold periods chosen for this research are 10, 20, 30, 40, and 80 days in length. Each of these windows was shifted 100 times into the future and the recursive NF recommendations were estimated for all of the moving windows. The NF-based sequence of daily trading strategies earns superior returns over a simple buy-and-hold strategy on the most of the periods. After the transaction costs are included, this still holds. Also, it is shown that a risk-averse investor with a smooth non-linearity can earn similar or higher returns than a “fuzzy logic investor” making “continuous” trading decisions.

The purpose of this chapter is not to produce a flawless trading strategy that would always win over a simple buy-and-hold strategy. That is close to impossible when the exchange rate for the last moving window day is much higher than the exchange rate for the initial buy-and-hold day. Rather, the purpose is to examine the possibility of creating a successful decision-making model that accounts for both public and private

---

information. In other words, to apply and test a new approach to modeling the agent’s decision-making (or behavior) in FX market. By doing this we can identify certain trading patterns, other than those based on well-known technical indicators. Specifically, this approach is also beneficial for signaling purposes to detect destabilizing FX market sentiments. Potentially, NF technology can be used for the government portfolio management, conducting monetary policy, financial stability, etc.

There are a number of directions where one could pursue further research on this problem. Several remarks regarding ANN architecture were made in Gradojevic and Yang (2000). Further, the problem of selecting an adequate ANN input lag might be resolved not only by using ANNs, but some other method for determining the degree of “non-linear correlation” such as “average mutual information” and “false nearest neighbors”. The FLC consists of one input, one output, and a very simple rule base. A set of technical trading rules could be added to a rule base. Consequently, new inputs which may include moving averages, lagged exchange rates, would bring more decision-making power to an overall NF system. An adaptive FLC (Cox 1993) which allows for membership function changes (e.g., widening “HOLD” fuzzy region) and weighting or changing the rules in the rule base may improve the NF system’s performance. The percentage of winning strategies can be improved through FLC system modifications.

---

32 Murphy (1999).
with respect to membership functions types, fuzzy sets, and defuzzification method. One
final guideline involves a different aspect of NF hybridization or a NF-GA combination.\textsuperscript{33}

Bibliography


*Neural Networks* 4: 251-257.


*Neural Networks* 3: 551-560.


