

# Revisiting Non-Parametric Exchange Rate Prediction

Nikola Gradojevic, Lakehead University, Canada  
Marko Caric, Faculty of Foreign Trade, Business Academy, Geri Karolja, Serbia

## ABSTRACT

*Given the large body of research addressing exchange rate predictability, the inability of the non-linear model by Diebold and Nason (Journal of International Economics 1990; 28: 315-332) to forecast better than a random walk is puzzling. This paper examines the forecasting performance of Diebold and Nason's non-parametric model for six major spot Canadian dollar exchange rates for the period 1987-2004. The findings suggest that a more flexible non-parametric estimation technique (artificial neural networks) is required and draw into question the choice of lagged dependent variables as explanatory factors. This paper also proposes a pure microstructure exchange rate model as an alternative to non-linear autoregressive models. Such a model sheds new light on the current evidence on linear/non-linear exchange rate predictability based on market microstructure variables.*

**Keywords:** Exchange Rates; Market Microstructure; Artificial Neural Networks; Nearest-neighbors Regression; Forecasting.

## 1. INTRODUCTION

A major challenge to international economic theory is to produce a short-run exchange rate model that would be more accurate than a simple random walk in out-of-sample forecasting. As the literature shows, initially, the efforts were focused on linking the macroeconomic (fundamental) variables to exchange rates at medium to long forecast horizons. For example, Meese and Rogoff (1983) showed that a simple random walk model performed no worse than any of the standard macroeconomic exchange rate models. Even after including ex-post data on the fundamentals, out-of-sample forecasting performance at 1-, 6- and 12-month horizons was surprisingly low. These findings were reinforced by a number of authors such as Baxter and Stockman (1989), Meese and Rose (1991), Flood and Rose (1995), Frankel and Rose (1995) and, very recently, Qi and Wu (2003). Contrary to this literature, a more favorable evidence was found in Mark (1995), Chinn and Meese (1995), Kilian and Taylor (2003) and Mark and Sul (2001). While there has been a significant criticism related to the statistical robustness of these results (e.g., Kilian, 1999), it has become apparent that some of the gains in the forecasting performance were due to accounting for non-linearities in the data (e.g., Kilian and Taylor, 2003). Noteworthy, Meese and Rose (1991) and Qi and Wu (2003) did not find non-linearities and market fundamentals useful for lower frequency forecasting. Obstfeld and Rogoff (2000) referred to this kind of weak relationship between the exchange rate and market fundamentals as the “exchange rate disconnect puzzle.”

Recently, a new approach attempted to shed more light on this puzzle. Instead of focusing solely on the macroeconomic fundamentals, “New Micro Exchange Rate Economics” (Lyons, 2001) calls attention to imperfections of financial markets: incomplete markets, sticky prices and various deviations from rational expectations due to over-reaction to news, noise or technical trading. This idea was pioneered by Kyle (1985) and extended by many other authors.<sup>1</sup> In the foreign exchange market context, Flood and Rose (1995) suggested that more microeconomic detail should be taken into account. Similarly, Cheung and Wong (2000) conducted a survey of practitioners in the interbank foreign exchange markets and reported significant deviations from rational expectations: only 1% of the traders look at macroeconomic fundamentals to determine short-run exchange rate movements. Further, when Jeanne and Rose (2002) incorporated noise traders into the general equilibrium framework, they showed that, for a fixed level of volatility of fundamentals, based on different levels of noise

trading, different levels of exchange rate volatility can occur. Within the partial equilibrium framework, Lyons and Evans (2002) included a variable reflecting the microeconomics of asset pricing to an exchange rate model. They introduced “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 macroeconomic models. More precisely, they managed to capture about 60% of the daily exchange rate changes using a linear model. However, the model was unable to generate statistically significant forecasts due to small sample size (and possibly because of misspecification and/or the linearity assumption).<sup>2</sup> More recently, Payne (2003) showed that 40% of the variation in the USD/DEM exchange rate was order flow driven.

A number of other scholars have pursued exchange rate modelling and forecasting using various methodologies, but with mixed success. To model the observed conditional heteroskedasticity of exchange rates, ARCH (Hsieh, 1989) and GARCH (Bollerslev, 1990) models were employed, but the results were very discouraging. For instance, in Gençay (1999), GARCH model generated insignificant directional and mean-squared prediction error forecast improvements over a simple random walk. Baillie and McMahon (1989), Hsieh (1988), Boothe and Glassman (1987) and Diebold and Nerlove (1989) reported that exchange rate changes are leptocurtic and might be non-linearly dependent. Kuan and Liu (1995) used backpropagation and recurrent artificial neural networks (ANNs) and detected non-linearities in the daily Japanese yen and British pound time series for the period 1980-1985. Some other studies involving ANNs such as Zhang and Hu (1998) and Hu *et al.* (1999) showed similar results: non-linear exchange rate forecasting based on its lagged values can be fruitful. Alternatively, technical trading signals can be constructed from the time series of spot exchange rates and used as forecasting variables in both linear and non-linear models. This intriguing possibility was researched and documented in Gençay (1999), Lisi and Medio (1997), Lo, Mamaysky and Wang (2000) and Levich and Thomas (1993). In all of these studies, the results contradicted the weak form of market efficiency.

Given the above findings, it appeared as though the optimal strategy for short-run forecasting was to employ non-linear autoregressive or technical trading exchange rate models. Surprisingly, Meese and Rose (1990), Diebold and Nason (1990), and, more recently, Cao and Soofi (1999) strongly rejected the existence of non-linearities and could not forecast better than a simple random walk in out-of-sample exchange rate predictions.

This paper focuses on Diebold and Nason (1990) (D&N from hereafter) and looks for explanations for their model’s poor forecasting performance. Specifically, for the post-D&N period (1987-2004), a nearest-neighbors (NN) technique called locally-weighted regression (LWR) is applied to six daily spot exchange rates (*vis-à-vis* the Canadian dollar): U.S. dollar, Japanese yen, Swiss franc, British pound, Australian dollar and Euro.<sup>3</sup> Thus, in addition to a different sample period, this paper also differs from the study by D&N in the number of exchange rates that are considered and the numeraire currency. D&N focused on the U.S. dollar and ten major exchange rates (1973-1987).

The most important findings of this paper are:

1. For the post-D&N period, the original NN approach could not improve upon the random walk model, i.e., the performance pattern is very similar as for the 1973-1987 period.
2. Allowing for more flexibility in the modeling technique (while fixing the input variables) produces forecasts that are superior to those from the random walk model. More precisely, the inputs from D&N were fed into the backpropagation ANN model and this substantially improved its out-of-sample performance for the Japanese yen, Swiss franc and Australian dollar. This result contradicts Gençay (1999) who found no significant difference in forecasting performance between ANN and NN models. It is worthwhile to mention, however, that the variables used in Gençay (1999) were fundamentally different.
3. The results of fusing a set of market microstructure variables for the Canada/U.S. dollar exchange rate with the NN technique demonstrate forecasting superiority of the non-linear microstructure model at a daily frequency. Thus, the evidence suggests that the NN model captures non-linearities created by the microstructure of the foreign exchange market on a daily basis. On average, the NN microstructure model accounts for 5.4% of the daily returns variance.
4. When the data are aggregated to a lower frequency (weekly), the advantages from using a non-linear model diminish. That microstructure variables drive currency returns in a linear fashion at the growth frequencies

was suggested in Evans and Lyons (2005), but, using a rich data set from the Bank of Canada, this paper provides a broader insight into their findings. Sub-section 3.4 compares the data set from this paper to the one used by Evans and Lyons (2005). Some new evidence also arises from the weekly data: across the three out-of-sample data sets the microstructure model explains on average almost 10% of the exchange rate movements. This is substantially higher than the result from Evans and Lyons (2005) who received the best estimate of 5.7% for the same forecasting horizon. In addition, the findings indicate that the forecasting power of the linear microstructure model is linked to controlling for the day-of-the-week effect.<sup>4</sup>

5. The nature of the informational content of the microstructure effects observed in this paper is such that they are relevant only for one-step-ahead forecasting, which is also found in Gradojevic (2006) for a linear microstructure model.

The remainder of the paper has three sections. Section 2 explains estimation techniques. Section 3 briefly describes the data and provides the results of the in-sample and out-of-sample empirical exercises. Conclusions and some recommendations for future research follow in Section 4.

## **2. ESTIMATION TECHNIQUES**

### **2.1 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, as in D&N, LWR is used to estimate one-step-ahead exchange rate returns 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 neighbors.

Suppose one is 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

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

where  $\phi(\cdot)$  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 the forecast of interest and  $x^*$  the point where we estimate  $\phi(\cdot)$ . Then, the  $x_{t-1}$  ( $t = 1, \dots, T$ ) series is organized into  $h$ -histories defined by:

$$x_{t-1}^h = (x_{t-h+1}, \dots, x_{t-1}) \tag{2}$$

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 neighbors ( $x_k$ ). Let  $\zeta$  be a smoothing constant such that  $0 < \zeta \leq 1$ , and let  $k = \text{int}(T\zeta)$ , where  $\text{int}(\cdot)$  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:

$$d(x^*, x_k) = \left[ \sum_{i=1}^m (x_{ki}^* - x_i^*)^2 \right]^{\frac{1}{2}} \tag{3}$$

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

3. In this step, a tricube weighting function is constructed:

$$w_t = \left( 1 - \left( \frac{d(x_t, x^*)}{d(x^*, x_k)} \right)^3 \right)^3 \tag{4}$$

4. The forecast is computed as:

$$y^* = \phi(x^*) = x^{*'} \hat{\beta} \tag{5}$$

where

$$\hat{\beta} = \arg \min \left[ \sum_{t=1}^T w_t (y_t - x_t' \beta)^2 \right] \tag{6}$$

The tricube weighting scheme is smooth in applying more confidence to the first nearest neighbor 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 neighbors goes to infinity with sample size, but at a slower rate, that is, as  $T \rightarrow \infty$ ,  $k \rightarrow \infty$ , but  $(k/T) \rightarrow 0$ . Clearly, consistency depends on the 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.

Initially, to keep the methodology comparable to D&N, when  $x_t$  is a vector of lagged dependent variables,  $\zeta$  is varied, but, when  $x_t$  is a vector of lagged microstructure variables, the optimal  $\zeta$  is found through in-sample cross-validation that minimizes the mean-squared prediction error (MSPE) on the in-sample data. This paper follows a cross-validation procedure from Gençay (1999). Essentially, based on the in-sample data (training set), all out-of-sample (testing)  $y^*$ 's are recursively estimated.

In the implementation of the cross-validation, first, the 50 most recent training set observations (1/8 of the training set size) are selected. At this point, all the remaining training set observations are treated as “testing data.” Then, the MSPE statistic is calculated from these “testing data” until the optimal  $\zeta$  is determined, i.e., the one that yields the lowest MSPE. This  $\zeta$  is recorded and the cross-validation proceeds by adding one more observation from the past to the initial training data. Again, the optimal  $\zeta$  is recorded for a new smaller “testing set” and this procedure is applied until the size of a “testing set” is zero. Finally, the  $\zeta$  related to the lowest overall MSPE is chosen and held fixed for estimating the first forecast observation in the actual testing set. Note that this method also determines the in-sample data length that is used for estimation. This procedure is then performed for each out-of-sample observation until the whole testing set is exhausted.

**2.2 Backpropagation ANNs**

ANNs represent a general class of non-parametric, non-linear models that had been originally conceptualized for pattern recognition and system control purposes, but subsequently found applications in finance and economics.<sup>5</sup>

Suppose that a single hidden layer ANN is composed of  $s$  input and  $q$  hidden nodes whereas the  $i^{th}$  independent variable is denoted by  $x_{it}$  ( $i=1, \dots, s$ ). The hidden and the output layers are characterized by two arbitrary types of non-linearities:  $\psi$  and  $\kappa$ , respectively. Backpropagation learning algorithm requires continuous differentiable non-linearities and the most commonly used type is the sigmoid logistic (or logsig) function:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{7}$$

The dependent variable ( $y_t$ ) is written as

$$y_t = \kappa \left( \beta_0 + \sum_{j=1}^q \beta_j \psi \left( \alpha_{j0} + \sum_{i=1}^s \alpha_{ij} x_{it} \right) \right) + \varepsilon_t \tag{8}$$

where  $\alpha_{ij}$  and  $\beta_j$  denote appropriate connection weights between the adjacent layers. Subscripts 0 for  $\alpha$  and  $\beta$  stand for ANN biases. Other types of transfer functions used in this paper are hyperbolic sigmoid tangent and linear.

Studies by Cybenko (1989) and Funahashi (1989) show that the non-linear representation given by equation (8), with  $\psi$  given by equation (7) can approximate a large number of mappings between  $x_t$ 's and  $y_t$  reasonably well.

In the first step of the implementation of the ANN model, the data is divided into training, validation, and testing parts, roughly in the ratio 6:3:1. The selection of the number of hidden layers and nodes in them is guided by the ANN's performance on the validation data with respect to the MSPE. The optimal number of hidden layers is found to be {1,2} and the number of hidden nodes {3,5}, depending on the forecasted exchange rate. The parameters are estimated using the standard Levenberg-Marquardt algorithm. Overfitting is prevented by early stopping, i.e., stopping the training process when the validation set error starts to increase. To control for data snooping biases, as in Garcia and Gençay (2000), the robustness of the ANN model is explored from the aspect of repeating the parameter estimation from five different sets of starting values.

**2.3 Empirical Measures Of Forecasting Performance**

The out-of-sample performance of the non-linear (and linear) models is measured based on the significance of the difference of the forecasting performance relative to the random walk model. The random walk model is summarized as:

$$y_t = y_{t-1} + e_t \tag{9}$$

In the first part of the empirical analysis the Diebold and Mariano (1995) statistic is utilized. However, due to the possible unreliability of the Diebold-Mariano statistic (West, 1996), a new test statistic, called the "projection statistic" is constructed. Evans and Lyons (2005) show that this statistic can be obtained from the following regression:

$$\hat{r}_{t+h|t} = a + br_{t+h} + e_{t+h} \tag{10}$$

where  $r_{t+h} = \log(S_{t+h}) - \log(S_t)$  are the actual returns,  $\hat{r}_{t+h|t}$  denotes the forecast of  $r_{t+h}$  based on the information available at time  $t$  and  $h$  is the forecasting horizon. If the considered models are unable to significantly improve upon the random walk, the estimate of  $\beta$  will be insignificant. As in Evans in Lyons (2005), the variance of  $\beta$  is estimated with the Newey-West (1987) estimator and the number of lags is set to  $h-1$ .

By definition, the estimate of  $\beta$  can be written as:

$$\hat{\beta} = \frac{\text{Cov}(r_{t+h}, \hat{r}_{t+h|t})}{\text{Var}(r_{t+h})} \tag{11}$$

Therefore, the estimate of  $\beta$  can also be viewed as the percentage of the currency returns variance that is explained by the model over the forecasting period.

### 3. EMPIRICAL RESULTS

#### 3.1 The Data

All exchange rate data are at a daily frequency and are obtained from the Bank of Canada. The following six nominal spot rates vis-à-vis the Canadian dollar are studied: U.S. dollar (USD), Japanese yen (JY), Swiss franc (SF), British pound (BP), Australian dollar (AD) and Euro (EU). Following D&N, to control for the day-of-the-week effects, the sample is composed of Wednesdays, 12:00 p.m. from 29 September 1987 to 26 October 2004. This represents a total of 886 observations of daily returns ( $r_t$ ). If  $S_t$  denotes an exchange rate at time  $t$ , then  $r_t = \log(S_t) - \log(S_{t-1})$ . The last 86 observations are held out-of-sample and estimated using the NN and ANN methodologies.

In the final empirical exercise, the variables from the field of microstructure are utilized to find the non-parametric conditional mean estimates for the Canada/U.S. exchange rate. Due to data unavailability at present time, the microstructure model is not applied to other exchange rates. For the same reason, the sample is shorter: from January 1990 to July 2000, and covers 494 observations (initially, 370 in-sample and 124 out-of-sample, for a daily frequency). The order flow data were obtained from the Bank of Canada and they are daily trading flows (in Canadian dollars) for six major Canadian commercial banks:

- Commercial client transactions (CC) include all transactions with resident and non-resident non-financial customers.
- Foreign institution transactions (FD) include all transactions with foreign financial institutions, such as FX dealers.
- Interbank transactions (IB) include transactions with other chartered banks, credit unions, investment dealers, and trust companies in the interbank market.

According to the Bank of Canada, these trading flows represent approximately 60% of all Canada/U.S. dollar transactions. Using the definition from Lyons (2001), 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).

#### 3.2 D&N: Revisited

The explanatory variables in D&N are lagged dependent variables and, thus, the forecasting model becomes a non-linear autoregressive one:

$$r_t = f(r_{t-1}, \dots, r_{t-p}) + e_t, p \in \{1, 3, 5\} \tag{12}$$

Table 1 reports the ratio of the smallest in-sample and out-of-sample MSPEs to the random walk model's MSPE when  $p=1$  for all of the studied exchange rates.

**Table 1. In-sample and out-of-sample performance of the NN model,  $p=1$ .**

	USD	JY	SF	BP	AD	EU
<b>In-sample MSPE ratio</b>	0.9743	0.9678	0.9794	0.9537	0.9728	0.9314
<b>Optimal in-sample <math>\zeta</math></b>	0.5	0.9	0.1	1.0	1.0	0.6
<b>Out-of-sample MSPE ratio</b>	1.0071	0.9878	0.994	0.9752	1.0037	0.9526
<b>Optimal out-of-sample <math>\zeta</math></b>	0.3	0.7	0.5	0.3	0.2	1.0
<b>DM</b>	0.41	-0.60	-0.24	-1.13	0.17	-0.57

Notes: Rows (1) and (3) are the ratios of the NN model’s MSPE to that of the random walk model. Rows (2) and (4) are optimal values for  $\zeta$ , corresponding to the smallest MSPE. DM denotes the Diebold and Mariano (1995) test statistic. Its one-sided critical values are -2.33, -1.645 and -1.282 for confidence levels of 99%, 95% and 90%, respectively.

Similarly, Tables 2 and 3 show the above ratio when  $p=3$  and 5. The optimal  $\zeta$  is not determined through a cross-validation procedure. Rather, as in D&N, a range of values for  $\zeta$  is tested until the one that minimizes the MSPE is found. These optimal MSPEs are also reported in Tables 1, 2 and 3. In addition, for the out-of-sample MSPEs, the Diebold and Mariano (1995) statistics are listed. In Table 1, when  $p=1$ , the MSPE for each exchange rate is minimized for  $\zeta$  taking different values, without any clear pattern of variation. This holds for both in-sample and out-of-sample estimations, and, also, for  $p=3$  and 5. This contradicts D&N in that they find the in-sample performance linked to high values of  $\zeta$  (0.9 and 1.0). Nonetheless, the in-sample performance for all exchange rates is similar, i.e., the MSPE is consistently higher than the random walk’s MSPE. The in-sample MSPE improvements are more substantial when  $p$  increases. For example, for  $p=5$ , the in-sample MSPE for the EU is almost 50% lower.

**Table 2. In-sample and out-of-sample performance of the NN model,  $p=3$ .**

	USD	JY	SF	BP	AD	EU
<b>In-sample MSPE ratio</b>	0.9165	0.8944	0.9065	0.886	0.9082	0.7844
<b>Optimal in-sample <math>\zeta</math></b>	0.1	0.3	0.5	0.5	0.5	0.4
<b>Out-of-sample MSPE ratio</b>	1.0341	1.0025	1.0146	1.0087	1.027	1.0433
<b>Optimal out-of-sample <math>\zeta</math></b>	0.8	0.3	0.3	0.6	0.8	0.2
<b>DM</b>	0.51	0.07	0.29	0.13	1.72	0.57

Notes: Rows (1) and (3) are the ratios of the NN model’s MSPE to that of the random walk model. Rows (2) and (4) are optimal values for  $\zeta$ , corresponding to the smallest MSPE. DM denotes the Diebold and Mariano (1995) test statistic. Its one-sided critical values are -2.33, -1.645 and -1.282 for confidence levels of 99%, 95% and 90%, respectively.

As many authors, including D&N, have found, a successful in-sample performance is never a guarantee that the out-of-sample performance will be comparable. From Tables 1-3, it is quite obvious that, if present, the forecast improvements are indeed modest and statistically insignificant according to the Diebold and Mariano (1995) statistic in almost all of the cases.

**Table 3. In-sample and out-of-sample performance of the NN model,  $p=5$ .**

	USD	JY	SF	BP	AD	EU
<b>In-sample MSPE ratio</b>	0.7993	0.7628	0.8054	0.7846	0.8093	0.5382
<b>Optimal in-sample <math>\zeta</math></b>	0.1	0.4	0.4	0.2	1.0	0.6
<b>Out-of-sample MSPE ratio</b>	1.0556	0.9896	0.9999	1.0312	1.0231	1.0027
<b>Optimal out-of-sample <math>\zeta</math></b>	0.2	0.8	0.8	0.7	0.2	0.6
<b>DM</b>	1.18	-0.27	-0.00	0.46	0.61	0.00

Notes: Rows (1) and (3) are the ratios of the NN model’s MSPE to that of the random walk model. Rows (2) and (4) are optimal values for  $\zeta$ , corresponding to the smallest MSPE. DM denotes the Diebold and Mariano (1995) test statistic. Its one-sided critical values are -2.33, -1.645 and -1.282 for confidence levels of 99%, 95% and 90%, respectively.

It is important to stress that the fact the optimal  $\zeta$  is not chosen ex-ante signifies the inability of the autoregressive NN models to forecast better than a random walk. To conclude, the results for the Canadian dollar for 1987-2004 data strongly support the findings of D&N.

### 3.3 The ANN Model

The ANN model employs the same data and explanatory variables as the NN model from the previous subsection. The objective is to improve upon the NN model in the out-of-sample part of the data. The forecasts are estimated using equation (8) with  $y_t=r_t$  and vector  $x_t=(r_{t-1}, \dots, r_{t-p})$ . As in the case of the NN model, it is found that ANNs produce considerable in-sample gains for  $p=1, 3, 5$  and insignificant out-of-sample MSPE improvements for  $p=1$ .<sup>6</sup>

Table 4. Out-of-sample performance of the ANN model,  $p=3$  and 5.

(a) U.S. dollar			
USD	MSPE ratio	$\beta$	p-value
<i>p=3</i>			
Experiment (1)	0.926	0.091	0.066
Experiment (2)	0.972	0.010	0.118
Experiment (3)	0.978	0.063	0.311
Experiment (4)	0.988	0.007	0.493
Experiment (5)	0.977	0.028	0.154
<i>p=5</i>			
Experiment (1)	0.955	0.038	0.052
Experiment (2)	0.988	0.011	0.603
Experiment (3)	0.991	0.027	0.218
Experiment (4)	0.983	0.011	0.305
Experiment (5)	0.969	0.022	0.291
(b) Japanese yen			
JY	MSPE ratio	$\beta$	p-value
<i>p=3</i>			
Experiment (1)	0.976	0.015	0.174
Experiment (2)	0.975	0.023	0.151
Experiment (3)	0.959	0.027	0.044
Experiment (4)	0.991	0.021	0.249
Experiment (5)	0.979	0.029	0.187
<i>p=5</i>			
Experiment (1)	0.938	0.045	0.041
Experiment (2)	0.946	0.054	0.060
Experiment (3)	0.947	0.048	0.043
Experiment (4)	0.943	0.038	0.011
Experiment (5)	0.978	0.022	0.066
(c) Swiss franc			
SF	MSPE ratio	$\beta$	p-value
<i>p=3</i>			
Experiment (1)	0.977	0.019	0.273
Experiment (2)	0.988	0.012	0.407
Experiment (3)	0.983	0.012	0.259
Experiment (4)	0.984	0.013	0.350
Experiment (5)	0.997	0.003	0.496
<i>p=5</i>			
Experiment (1)	0.951	0.065	0.051
Experiment (2)	0.947	0.051	0.021
Experiment (3)	0.969	0.032	0.136
Experiment (4)	0.964	0.028	0.037
Experiment (5)	0.957	0.036	0.026



<b>(d) British pound</b>			
<b>BP</b>	<b>MSPE ratio</b>	<b><math>\beta</math></b>	<b>p-value</b>
<i>p=3</i>			
Experiment (1)	1.004	0.003	0.748
Experiment (2)	0.999	0.000	0.899
Experiment (3)	1.009	-0.004	0.224
Experiment (4)	0.991	0.013	0.323
Experiment (5)	1.016	-0.003	0.760
<i>p=5</i>			
Experiment (1)	0.999	0.008	0.474
Experiment (2)	0.984	0.014	0.178
Experiment (3)	0.993	0.006	0.425
Experiment (4)	1.017	0.002	0.889
Experiment (5)	0.985	0.017	0.305
<b>(e) Australian dollar</b>			
<b>AD</b>	<b>MSPE ratio</b>	<b><math>\beta</math></b>	<b>p-value</b>
<i>p=3</i>			
Experiment (1)	0.987	0.012	0.099
Experiment (2)	0.975	0.017	0.022
Experiment (3)	0.947	0.036	0.040
Experiment (4)	0.981	0.032	0.040
Experiment (5)	0.955	0.032	0.101
<i>p=5</i>			
Experiment (1)	0.985	0.011	0.372
Experiment (2)	0.983	0.013	0.229
Experiment (3)	0.987	0.009	0.290
Experiment (4)	0.963	0.030	0.053
Experiment (5)	0.958	0.034	0.087
<b>(f) Euro</b>			
<b>EU</b>	<b>MSPE ratio</b>	<b><math>\beta</math></b>	<b>p-value</b>
<i>p=3</i>			
Experiment (1)	0.997	0.000	0.958
Experiment (2)	0.983	0.018	0.254
Experiment (3)	0.997	0.002	0.871
Experiment (4)	0.982	0.009	0.505
Experiment (5)	0.999	0.000	0.940
<i>p=5</i>			
Experiment (1)	0.931	0.060	0.210
Experiment (2)	0.968	0.029	0.198
Experiment (3)	0.969	0.052	0.175
Experiment (4)	0.968	0.089	0.120
Experiment (5)	0.963	0.017	0.161

Notes: Experiment (1)-(5) denotes that the out-of-sample MSPEs are obtained from ANNs for which the parameters were estimated from five random seeds. The MSPE ratios column are the ratios of the ANN model's MSPE to that of the random walk model. The estimate of the slope coefficient from the projection regression (Evans and Lyons, 2005) is denoted by  $\beta$ . The p-value is for the significance of  $\beta$  where its variance is estimated using the Newey-West (1987) estimator with  $h$  (forecasting horizon) – 1 lags.

For some exchange rates, the findings are different for  $p=3$  and 5. The performance of the ANN model for all six exchange rates is presented in Table 4 (a, b, c, d, e, f). The significant forecast improvements in MSPEs can be observed for the JY ( $p=5$ ), SF ( $p=5$ ) and AD ( $p=3$ ) exchange rates. These improvements range from about 2-6% and are robust to different starting values. The estimates of statistically significant  $\beta$ 's are correctly signed and show that the ANN model can explain from 1.2% to 6.5% of the daily returns variance. A noteworthy result is that the findings indicate predictability power that contradicts the weak form of market efficiency. Not surprisingly, in general, the forecast improvements are more substantial when  $p=5$ . More lags enable the ANN model to extract

more information from the past data. Thus, the out-of-sample results strongly favor the ANN autoregressive model over the NN autoregressive model.

As noted earlier, to control for possible data snooping, the robustness of the performance of the ANN model is also evaluated and presented in Table 4. Similar to the approach by Garcia and Gençay (2000), the ANNs are estimated over five experiments that involve five different (random) sets of initial parameter values. These experiments show that, when the estimates of  $\beta$  are insignificant, the forecasting performance of the non-linear ANN autoregressive models is quite sensitive to different starting values. Taken together, the results of the robustness exercise indicate that either the model is misspecified or that more lags of the dependent variable have to be included in vector  $x_t$ . Increasing  $p$  to values greater than five is beyond the scope of this paper, but several alternative model specifications will be explored in the next sub-section.

**3.4 The Microstructure Model**

This part of the paper is motivated by Evans and Lyons (2005) and Gradojevic and Yang (2005) who employed microstructure variables – order flows – and find that they have a significant short-run predictive power. In line with these papers, out-of-sample forecasts are estimated recursively, relying only on information actually available. The forecasts are generated for daily and weekly frequencies.

The non-linear (NN) microstructure-only model for the Canada/U.S. exchange rate returns is specified as follows:<sup>7</sup>

$$r_t = f(CC_{t-1}, \dots, CC_{t-p}, IB_{t-1}, \dots, IB_{t-p}, FD_{t-1}, \dots, FD_{t-p}) + e_t, p \in \{1, 2, 3\} \tag{13}$$

In the traditional microstructure literature the price is linked to order flow in a linear manner. The rationale behind using a non-linear model is to allow for the possibility that market participants could act strategically and pursue more complex trading strategies (O’Hara, 1995).

To compare the forecasting performance of equation (13) to a linear forecasting model by Evans and Lyons (2005), the following linear regression is estimated for  $p=1$ :

$$r_t = a + b_1CC_{t-1} + b_2IB_{t-1} + b_3FD_{t-1} + e_t \tag{14}$$

Note that there are fundamental differences between equation (14) and the model by Evans and Lyons (2005). In this study, the forecasted exchange rate is the Canada/U.S. dollar, the Bank of Canada data account for 85% of the transactions for 1990-2000 data and the order flow types are as defined in sub-section 3.1. In contrast, Evans and Lyons (2005) forecast the USD/Euro exchange rate using 1993-1999 data, while accounting for only 10-15% of the transactions (six market segments covered by Citibank). In addition, the size of their out-of-sample part is roughly the second part of the data set.

**Table 5. One-day-ahead forecasting performance of the microstructure-only models,  $p=1$ .**

	Out-of-sample size		
	124	94	74
<b>MSPE ratio -NN model</b>	0.9597	0.9498	0.9608
$\beta$	0.072	0.054	0.035
<b>p-value</b>	(0.136)	(0.151)	(0.061)
<b>MSPE ratio -Linear model</b>	1.0010	1.0111	1.0178
$\beta$	0.004	-0.000 (0.988)	0.001
<b>p-value</b>	(0.744)		(0.869)

Notes: MSPE ratio is the ratio of the MSPE of the NN (and the linear) model to that of the random walk model. The estimate of the slope coefficient from the projection regression (Evans and Lyons, 2005) is denoted by  $\beta$ . The p-value is for the significance of  $\beta$  where its variance is estimated using the Newey-West (1987) estimator with  $h$  (forecasting horizon) – 1 lags.

The results for one-day-ahead forecasting and  $p=1$  are presented in Table 5. Relative to the random walk model, the NN model provides robust improvements that range from 4% to 5% across three different out-of-sample sizes (74-124 observations). On average, the NN model is able to explain about 5.4% of the daily returns variance. For the linear model there are no forecast improvements relative to the random walk. Although the performance of the NN model is not impressive, the comparison of the NN and the linear models shows that, for a daily frequency, a non-linear model is superior.

**Table 6. One-week-ahead forecasting performance of the microstructure-only models,  $p=1$ .**

	Out-of-sample size		
	150	100	80
<b>MSPE ratio -NN model</b>	0.9459	0.9802	0.9840
<b><math>\beta</math></b>	0.087	0.073	0.129 (0.003)
<b>p-value</b>	(0.000)	(0.003)	
<b>MSPE ratio -Linear model</b>	0.9235	0.9053	0.8564
<b><math>\beta</math></b>	0.093 (0.000)	0.083	0.116
<b>p-value</b>		(0.000)	(0.000)

Notes: MSPE ratio is the ratio of the MSPE of the NN (and the linear) model to that of the random walk model. The estimate of the slope coefficient from the projection regression (Evans and Lyons, 2005) is denoted by  $\beta$ . The p-value is for the significance of  $\beta$  where its variance is estimated using the Newey-West (1987) estimator with  $h$  (forecasting horizon) – 1 lags.

Moving to a longer, weekly forecasting horizon (5 trading days) gives quite a different picture (Table 6;  $p=1$ ). For this exercise, utilizing 480 observations, the in-sample size varies from 330 to 400 observations (consequently, the out-of-sample size is from 150 to 80). The dependent variable  $r_t$  is now a weekly returns series (from Wednesday to Wednesday) and the order flows are cumulative over five trading days. To clarify, the weekly returns for the week  $t$  are forecasted based on the aggregated transactions from the week  $t-2$  to  $t-1$ . Table 6 shows that the forecast improvements of both the linear and the NN models are significantly superior to the random walk model. For all of the considered out-of-sample sizes, the MSPE ratio is smaller than one. The average forecast improvement across four out-of-sample sets is about 10% for the linear model and for the NN model it is about 3%. In addition, the forecasts from the linear model account for almost 10% of the returns variance. Therefore, there is no forecast gain from applying a non-linear model at a lower frequency.

Increasing  $p$  to values greater than 1 deteriorates the forecasting performance of both the linear and the NN models. This suggests that the predictive ability of the microstructure model is limited to one-step-ahead forecasting and the information extracted from the time series of transactions may not be useful. In other words, the impact of the information, regardless if fully private or not, contained in order flow is quickly reflected in spot prices. This is an important and novel finding that, to the authors’ best knowledge, has not been researched by both linear and non-linear methods in any previous studies.

#### 4. CONCLUSIONS

This paper has analyzed why the exchange rate model by D&N shows a poor forecasting performance relative to the random walk model. In light of the findings from this analysis, the results suggest that improvements in predicting exchange rate movements might be realized through further market microstructure research.

The evidence from the post-D&N sample for the Canadian dollar price of the six major currencies is consistent with the findings by D&N that a non-linear autoregressive (NN) model is unable to forecast exchange rate returns at a daily frequency better than a random walk. Two possible causes for this performance were further investigated: the first related to the non-parametric modelling technique and the second related to the selection of inputs. It was shown that an ANN autoregressive model outperformed the random walk model and that the daily exchange rate returns were to some extent predictable. Thus, the results do not support the claims by D&N that non-linearities cannot be used for improved short-run point prediction.

Building on the recent empirical success of the microstructure-based models for exchange rate determination and forecasting, a model in the spirit of Evans and Lyons (2005) was constructed. A rich data set from the Bank of Canada provided an alternative panel of explanatory variables for the model. Although the variables of the model were not identical to those from Evans and Lyons (2005), based on the definitions of individual order flows, a significant degree of overlap is present. The results, in general, support Evans and Lyons (2005) and offer more insight into building microstructure models for forecasting. The results strongly favor microstructure-only forecasting models, especially at lower frequencies. The significant improvements over the simple random walk model are present at both daily and weekly frequencies. At a daily frequency, a non-parametric model performed superior to a linear model while this was not the case at a weekly frequency, where a linear model yielded a lower average MSPE. Finally, the findings signify that microstructure models are useful only for one-step-ahead forecasting.

There have been numerous concerns raised about microstructure models. The main one concerns their relationship to fundamental as well as general equilibrium models and endogeneity of order flow variables. It should be noted that the goal of this paper is not to test the usefulness of additional fundamental variables, but to test if the very short-run Canada/U.S. dollar exchange rate can be forecasted by a pure microstructure model.<sup>8</sup> Consequently, this paper does not investigate whether fundamental variables, or, more generally, public information is aggregated and channelled through order flows. This paper also does not study the implications of private information (noise traders), reflected in order flows, for exchange rates and macroeconomic fundamentals in a general equilibrium framework. More detailed examination along all of these research avenues is left to a future paper. As for the concern related to endogeneity, Granger causality tests were performed and it was found that the causality ran strictly from order flows to the price.<sup>9</sup> One possible extension to the current study could be a market microstructure and time series combined approach which might improve the forecasting performance further.

Finally, as order flow represents a proxy for private (and to a certain extent public) information flow, looking beyond what drives order flow is a research direction that can undoubtedly help resolving the “exchange rate disconnect puzzle.” Within the macroeconomic announcements context, order flows may be viewed as a gauge for how market participants perceive the arrival of new information to the market. Order flows reflect the actual market sentiment, i.e., aggregate expectations about the future direction of the economy. Investigating the surprise effect of macroeconomic announcements on order flows could be a promising research direction. Further, a number of papers show that technical trading rules can be useful for generating excess profits in the FX market, even after adjusting for transaction costs (Neely et al., 1997). However, the question to what extent FX traders employed technical indicators in their trading still remains. Examining order flows and determining the technical trading content in actual trades can undoubtedly answer this question.

#### **AUTHOR INFORMATION**

**Dr. Nikola Gradojevic** has been an associate professor of finance at Lakehead University in Thunder Bay, Canada since July 2007 (assistant professor: July 2003-June 2007) . He completed his Ph.D. in economics (area of specialization: finance) at the University of British Columbia (UBC) in Vancouver, Canada in January 2003. During his career he took various positions at UBC, Bank of Canada, Federal Reserve Bank of St. Louis and in the private sector (as a consultant). Dr. Gradojevic’s research interests include international finance (market microstructure), empirical asset pricing (option pricing), technical trading, asset price volatility and bubbles, and artificial intelligence (neural networks, fuzzy logic, etc.).

**Dr. Marko Caric** is an associate professor at the Faculty of Foreign Trade of the Business Academy in Novi Sad, Serbia. He obtained his undergraduate training from the University of Novi Sad, Faculty of Technical Sciences, Electrical Engineering in 1996, and the Faculty of Economics in 1999. In 2002, he obtained a PhD degree in the field of management in telecommunications, and in 2005 a PhD degree in Management at the Business Academy in Novi Sad. Dr. Caric is an author of numerous books and journal publications. Outside academia, he worked as a business development manager for Lucent Technologies in Hilversum, Netherlands.

**Footnotes**

- <sup>1</sup> See O’ Hara (1995) for more information on this approach, generally known as market microstructure theory.
- <sup>2</sup> In a very recent paper (Evans and Lyons, 2005), using a linear model on different data, they managed to generate statistically significant forecast improvements.
- <sup>3</sup> The sample for the Euro is 1999-2004.
- <sup>4</sup> By utilizing the whole sample, Gradojevic (2007) reports inferior predictive ability for the weekly forecasting horizon.
- <sup>5</sup> More extensive review of various applications of ANNs in finance and economics can be found in Qi (1996).
- <sup>6</sup> These results are not presented and are available upon request from the author.
- <sup>7</sup> For the pure microstructure non-linear model no significant difference in the forecasting performance was found between the NN and the ANN models, i.e., unlike in the D&N’s model, the nature of the non-linearity is found to be invariant to the choice of estimation method. Noteworthy, the term “pure” refers to a model that contains only microstructure explanatory variables.
- <sup>8</sup> Interest rate differential and crude oil price were found to be informative in Gradojevic and Yang (2006).
- <sup>9</sup> In a recent paper that covers the period 1990-2005, we have investigated the causality issue further. This paper estimates VECM/VAR systems to capture the dynamics of order flow and exchange rates. We find that CD and FD contemporaneously drive the exchange rate while we identify periods when CC and IB order flows seem to respond to exchange rate movements. This can be viewed as an evidence that commercial customers and home dealers occasionally act as passive liquidity providers.

**REFERENCES**

1. Baillie, R., McMahon, P., 1989. *The foreign exchange market: theory and econometric evidence*. Cambridge University Press, New York.
2. Baxter, M., Stockman, A., 1989. Business cycles and exchange-rate systems. *Journal of Monetary Economics* 23, 377-400.
3. Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model.” *Review of Economics and Statistics* 72, 498-505.
4. Boothe, P., Glassman, D., 1987. The statistical distribution of exchange rates. *Journal of International Economics* 22, 297-319.
5. Cheung, Y-W., Wong., C.Y-P., 2000. A survey of market practitioners’ views on exchange rate dynamics. *Journal of International Economics* 51, 401-423.
6. Chinn, M.D., Meese, R.A., 1995. Banking on currency forecasts: how predictable is change in money? *Journal of International Economics* 38 (1-2), 161-178.
7. Cleveland, W.S., 1979. Robust-locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74, 829-836.
8. Cleveland, W. S., Devlin, S. J., 1988. Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83, 596-610.
9. Cybenko, G., 1989. Approximation by superposition of a sigmoidal function. *Mathematics of Control, Signals and Systems* 2, 303-314.
10. Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253-263.
11. Diebold, F.X., Nason, J., 1990. Nonparametric exchange rate prediction. *Journal of International Economics* 28 (3-4), 315-332.
12. Diebold, F.X., Nerlove, M., 1989. The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics* 4, 1–21.
13. Evans, M.D.D., Lyons, R.K., 2005. Meese-Rogoff redux: micro-based exchange rate forecasting. *American Economic Review* 95 (2), 405-414.
14. Flood, R., Rose, A.K., 1995. Fixing exchange rates: A virtual quest for fundamentals. *Journal of Monetary Economics* 36, 3-37.
15. Frankel, J.A., Rose, A.K., 1995. Empirical research on nominal exchange rates. In Grossman, G., Rogoff, K. (Eds.). *Handbook of International Economics*, Vol. III. Elsevier Science, Amsterdam, 1689-1729.
16. Funahashi, K.-I., 1989. On the approximate realization of continuous mappings by neural networks. *Neural Networks* 2, 183-192.

17. Garcia, R., Gençay, R. 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics* 94 (1-2), 93-115.
18. Gençay, R., 1999. Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules. *Journal of International Economics* 47, 91-107.
19. Gradojevic, N. 2006. The microstructure of the canada/u.s. dollar exchange rate: A robustness test, *Economics Letters* 94 (3), 426-432.
20. Gradojevic, N., Yang, J., 2006. Non-linear, non-parametric, non-fundamental exchange rate forecasting. *Journal of Forecasting* 25 (4), 227-245.
21. Hsieh, D.A., 1988. The statistical properties of daily foreign exchange rates. *Journal of International Economics* 24, 129-145.
22. Hsieh, D.A., 1989. Testing for nonlinear dependence in daily foreign exchange rate changes. *Journal of Business* 62, 329-368.
23. Hu, M.Y., Zhang, G., Jiang, C.X., Patuwo, B.E., 1999. A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting. *Decision Sciences* 30 (1), 197-216.
24. Jeanne, O., Rose, A.K., 2002. Noise trading and exchange rate regimes. *The Quarterly Journal of Economics* 117 (2), 537-569.
25. Kilian, L., 1999. Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions? *Journal of Applied Econometrics* 14 (5), 491-510.
26. Kilian, L., Taylor, M.P. Taylor., 2003. Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics* 60 (1), 85-107.
27. Kuan, C.-M., Liu, T., 1995. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics* 10 (4), 347-364.
28. Kyle, A., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315-1335.
29. Levich, R., Thomas, L., 1993. The significance of technical trading rule profitability: A bootstrap approach. *Journal of International Money and Finance* 12, 451-474
30. Lisi, F. Medio, A., 1997. Is a random walk the best exchange rate predictor? *International Journal of Forecasting* 13, 255-267.
31. Lo, A.W., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. *Journal of Finance* 55 (4), 1705-1765.
32. Lyons, R.K., 2001. *The microstructure approach to exchange rates*. MIT Press, Cambridge.
33. Lyons, R.K. Evans, M.D.D., 2002. Order flow and exchange rate dynamics. *Journal of Political Economy* 110 (1), 170-180.
34. Mark, N.C., 1995. Exchange rates and fundamentals: evidence on long-horizon prediction. *American Economic Review* 85, 201-218.
35. Mark, N.C., Sul, D., 2001. Nominal exchange rates and monetary fundamentals: evidence from a small post-Bretton Woods panel. *Journal of International Economics* 53, 29-52.
36. Meese, R.A., Rogoff, K., 1983. Empirical exchange rate model of the seventies: do they fit out of sample? *Journal of International Economics* 14, 3-24.
37. Meese, R.A., Rose, A.K., 1990. Non-linear, nonparametric, nonessential exchange rate estimation. *American Economic Review* 80, 192-196.
38. Meese, R.A., Rose, A.K., 1991. An empirical assessment of non-linearities in models of exchange rate determination. *Review of Economic Studies* 58 (3), 603-619.
39. Neely, C., Weller, P., Dittmar, R. 1997. Is technical analysis in the foreign exchange market profitable?: A genetic programming approach. *Journal of Financial and Quantitative Analysis* 32, 405-426.
40. Newey, W.K., West, K.D., 1987. A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.
41. Obstfeld, M., Rogoff, K., 2000. The six major puzzles in international macroeconomics: is there a common cause? NBER Working Paper 7777, National Bureau of Economic Research, Inc.
42. O'Hara, M., 1995. *Market microstructure theory*. Blackwell Publishers Ltd., Oxford.
43. Payne, R., 2003. Informed trade in spot foreign exchange markets: an empirical investigation. *Journal of International Economics* 61, 307-329.
44. Qi, M., 1996. Financial application of artificial neural networks. *Handbook of Statistics* 14, 529-552.

45. Qi, M., Wu, Y., 2003. Nonlinear prediction of exchange rates with monetary fundamentals. *Journal of Empirical Finance* 10, 623-640.
46. Robinson, P., 1987. Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875-892.
47. Soofi, A., Cao, L., 1999. Nonlinear deterministic forecasting of daily dollar exchange rates. *International Journal of Forecasting* 15, 421-430.
48. Stone, C.J., 1977. Consistent non-parametric regression. *Annals of Statistics* 5, 595-645.
49. West, K.D., 1996. Asymptotic inference about predictive ability. *Econometrica* 64, 1067- 1084.
50. Yakowitz, S.J., 1987. Nearest neighbour methods for time series analysis. *Journal of Time Series Analysis* 8 (2), 235-247.
51. Zhang, G., Hu, M.Y., 1998. Neural network forecasting of the British pound/US dollar exchange rate. *International Journal of Management Science* 26 (4), 495–506.

**NOTES**

NOTES