5 November, 1993

LU TP 93-23

Dependency Analysis and Neural Network Modeling of Currency Exchange Rates

Hong Pi[1]

Department of Theoretical Physics, University of Lund
Sölvegatan 14A, S-223 62 Lund, Sweden

Abstract[2]: In this paper we explore function dependences in multivariate currency exchange rate data using the δ-test. Despite large noise levels and the high dimensional nature of the data set, the δ-test captures the dependency structures in the data quite efficiently. Dependences of a currency on its own history, interdependencies between currencies, and dependences of a currency on the Dow Jones Industrial Index are established from the data by the δ-test. When the variables identified in this way are used as inputs to a feedforward neural network to model the exchange rate series, a prediction performance better than status quo is obtained. The results confirm that exchange rates series are not random walk processes and that a certain degree of short term predictability is possible if the dependency structure is properly modeled.

[1]pihong@thep.lu.se

[2]Talk presented at the first International Workshop on Neural Networks in the Capital Markets, London Business School, November 18-19, 1993.

1 Introduction

In real world situations one often has access to a large collection of historical data (time series), with no knowledge about the physical laws or dynamics governing the system that generates the data. To gain partial understanding about the system and to be able to predict its future course, one would then rely on identifying and modeling empirical regularities in the data. Artificial neural networks (ANN), with its ability of adaptively exploring a class of linear and nonlinear models, are particularly suitable to the task [1]. Among various applications several authors [2, 3, 4] have applied the ANN techniques to time series data in economic and business problems, and found favorable results when comparing to more conventional approaches.

In this paper we report an analysis on multivariate exchange rate and stock index data series. While we use ANN to model and test predictability in the exchange rates, our emphasis will be on a exploratory data analysis using a statistical method, the δ-test [5], which examines the dependency structure that may exist in the data. The result of the analysis, coupled to the prediction performance of the ANN model, confirms that an exchange rate sequence is not simply a random walk process. Similar analysis can potentially be used to test the efficient market hypothesis [6].

Neural network models, though adaptive, are not entirely of "black box" nature. One has to select (sometimes by making an educated guess) relevant variables to feed the network as inputs. While it is viable to feed blindly as many input variables as the network can handle, the resulting large network may be difficult to train and if some of the inputs turn out to be noise, it can significantly reduce the ability of the network to generalize. Although one can use elaborated algorithms such as weight elimination [8, 9, 10] to cut down the effect of fitting to noise, or use recurrent networks to perform automatic temporal pattern recognition [11], one still can not avoid the problem of how to properly select inputs. It is desirable to identify in advance as precisely as possible the relevant variables.

The δ-test [5] is a statistical tool devised to detect function dependences between measurements of variables. Its use and effectiveness have been demonstrated on a number of synthetic time series [5], as well as on a real world utility consumption data set [7]. The test constructs correlations between the independent-dependent variables in the form of conditional probabilities, which are then checked against the expectations of null or full functional dependence. The identified variables and dimensionality serve as a useful guide for selecting inputs to construct models.

We conclude from the results of this paper that interdependencies exist between currencies and between currency and stock index, and these dependences can be modeled successfully by neural networks. This raises the question that to what extent modeling can be improved if one can identify and employ a more complete set of relevant variables such as the relevant interest rates.

We briefly describe the formulation of the δ-test in Section 2. In Section 3 we apply the δ-test to the multivariate exchange rate data series. In Section 4 we use a feedforward multi layer perceptron (MLP) to model the currency data.

2 The δ-test

Assume that we have sequences of measurements on a *dependent* variable z_0 and a set of *independent* variables z_1, z_2,...z_m. These measurements can manifest as multivariate time series, or as a univariate time series, in which case z_k's should be understood as the time lag variables of z_0: $z_k(t) = z_0(t-k)$. We wish to test that if there exist function dependences in the form

$$z_0 = f(z_1, z_2, ..., z_m) + r \tag{1}$$

where r represents an indeterminable part which originates either from insufficient dimension of the measurements or from real noise. r should be zero if the system is completely deterministic in terms of the set of independent variables. In case of a univariate time series, the dimension $d_{min} = m + 1$ which minimizes r is the minimum embedding dimension.

We approach the problem by constructing conditional probabilities in embedding spaces of various dimensions d. The time series can be represented as a series of N points $\mathbf{z}(i)$ in a $(d+1)$-dimensional space $(d = 0, 1, 2, ...)$

$$\mathbf{z}(i) = (z_0(i), z_1(i), .., z_k(i), ., z_d(i)). \tag{2}$$

The distances between the k:th components of two vectors $\mathbf{z}(i)$ and $\mathbf{z}(j)$ are defined as

$$l_k(i,j) = |z_k(i) - z_k(j)|, \quad k = 0, 1, ..., d \tag{3}$$

Given a set of positive numbers, ϵ and δ, one can construct from the data the conditional probabilities

$$P_d(\epsilon|\,\vec{\delta}) \equiv P(l_0 \leq \epsilon|\,\vec{l} \leq \vec{\delta}) = \frac{n(l_0 \leq \epsilon, \vec{l} \leq \vec{\delta})}{n(\vec{l} \leq \vec{\delta})}, \tag{4}$$

where the notation $\vec{l} \leq \vec{\delta}$ is short for $\{(l_1 \leq \delta), (l_2 \leq \delta), ..., (l_d \leq \delta)\}$, $n(l_0 \leq \epsilon, \vec{l} \leq \vec{\delta})$ and $n(\vec{l} \leq \vec{\delta})$ are the number of vector pairs satisfying the corresponding distance constraints. The quest of testing dependences then becomes to examine how the probability structure of the dependent variable is changed conditional to the independent variables. We will make use the following important observations:

1. For a completely random time series, no dependence exists and one has

$$P_0(\epsilon) = P_1(\epsilon|\vec{\delta}) = ... = P_d(\epsilon|\vec{\delta}) = ... \tag{5}$$

   This identity, which should be understood in a statistical sense, holds for any choice of positive ϵ and δ.

2. If a continuous map exists as in eq. (1) with no intrinsic noise, then for any $\epsilon > 0$ there exists a δ_ϵ such that

$$P_d(\epsilon|\vec{\delta}) = 1 \quad \text{for} \quad \delta \leq \delta_\epsilon \quad \text{and} \quad d \geq d_0 \tag{6}$$

   where d_0 represents some minimum dimension which covers all the relevant variables.

3. In the presence of noise r, $P_d(\epsilon|\vec{\delta})$ will no longer saturate to 1 as ϵ becomes smaller than the width Δr_{max} of the noise.

Eq. (6) is a direct consequence of the definition of function uniform continuity, which states that if $z_0 = f(z_1, ..., z_d)$, then for any $\epsilon > 0$ there exists a $\delta > 0$ such that the conditions $|z_1 - z_1'| < \delta$,...,$|z_d - z_d'| < \delta$ guarantee $|z_0 - z_0'| < \epsilon$. With the presence of noise, however, $|z_0 - z_0'| = |f(z_1, ...) - f(z_1', ...) + r - r'|$, which as $\delta \to 0$ can not be made smaller than $\Delta r = |r - r'|$. This justifies the statement number 3. This fact can be used to estimate the size of the noise. At the region where $P_d(\epsilon|\vec{\delta})$ just starts to drop off from one, we expect

$$P_d(\epsilon|\vec{\delta}) \approx 1 - Prob(\Delta r > \epsilon) = Prob(\Delta r \leq \epsilon). \tag{7}$$

If the noise is a flat distribution extending from $-R$ to R with standard deviation $\sigma_r = R/\sqrt{3}$, we obtain

$$Prob(\Delta r \leq \epsilon) = \frac{\epsilon}{R}\left(1 - \frac{1}{4}\frac{\epsilon}{R}\right) \tag{8}$$

The point at which $P_d(\epsilon|\vec{\delta}) = 0.99$ corresponds to $\epsilon_0 = 1.8R$. So for a flat noise distribution its standard deviation can be estimated as

$$\sigma_r = \frac{\epsilon_0}{1.8\sqrt{3}} = \frac{\epsilon_0}{3.1}. \tag{9}$$

If the noise obeys a Gaussian distribution then $Prob(\Delta r)$ is also a Gaussian and has a standard deviation $\sqrt{2}\sigma_r$. The point at which $P_d(\epsilon|\vec{\delta}) = 0.99$ would then correspond to $\epsilon_0 = 2.57\sqrt{2}\sigma_r$, and the estimate for σ_r becomes

$$\sigma_r = \frac{\epsilon_0}{2.57\sqrt{2}} = \frac{\epsilon_0}{3.6}. \tag{10}$$

The behaviour of $P_d(\epsilon|\vec{\delta})$ as a function of δ and ϵ for various d are shown schematically in fig. 1. The consequences of randomness (eq. (5)) and complete determination (eq. (6))

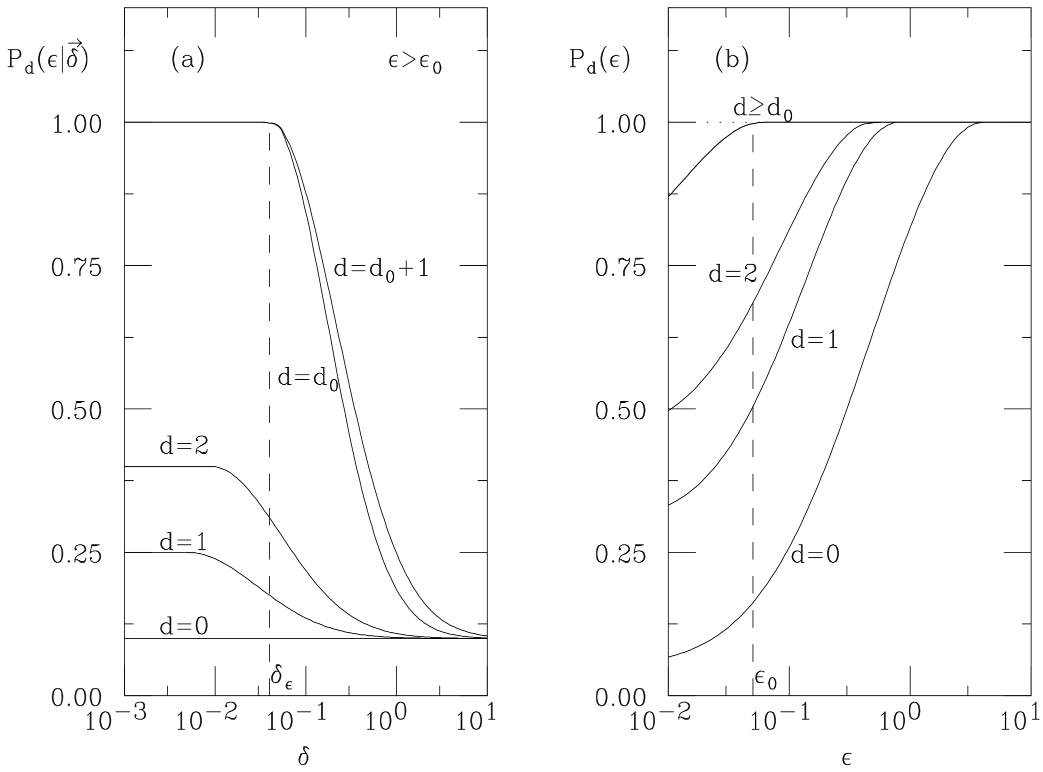

Figure 1: (a). $P_d(\epsilon|\,\vec{\delta})$ as a function of δ for fixed ϵ. Saturation to 1 would be observed if the width of noise is less than ϵ. (b). Behaviour of the maxima $P_d(\epsilon) \equiv \max_{\delta>0} P_d(\epsilon|\,\vec{\delta})$ as a function of ϵ. The region saturating to 1 would be pushed toward smaller ϵ if the d:th conditional variable is relevant. The point ϵ_0 at which the saturation deviates from 1 can be identified approximately as the width of the noise $\Delta r_{max} \sim \epsilon_0$.

provide us a yardstick to measure the degree of dependency between the variables. An interesting quantity to examine is the maxima

$$P_d(\epsilon) \equiv \max_{\delta>0} P_d(\epsilon|\,\vec{\delta}) = P_d(\epsilon|\,\vec{\delta})|_{\delta\leq\delta_\epsilon}. \tag{11}$$

How $P_d(\epsilon)$ changes with d and ϵ provides basically all the information we need (see fig. 1b). To quantify the dependence on each of the variables, it is convenient to define a *dependability index*

$$\lambda_d(\epsilon) = \frac{P_d(\epsilon) - P_{d-1}(\epsilon)}{1 - P_0(\epsilon)}, \quad d = 1, 2, ... \tag{12}$$

and its average over ϵ

$$\overline{\lambda}_d = \frac{\int_0^\infty d\epsilon \, \lambda_d(\epsilon)(1 - P_0(\epsilon))}{\int_0^\infty d\epsilon \, (1 - P_0(\epsilon))} = \frac{\int_0^\infty d\epsilon \, (P_d(\epsilon) - P_{d-1}(\epsilon))}{\int_0^\infty d\epsilon \, (1 - P_0(\epsilon))}. \tag{13}$$

In general $1 \geq \lambda_d \geq 0$, while $\lambda_d = 1$ (or $\sum \lambda_d = 1$) signals a completely deterministic relationship and $\lambda_d \approx 0$ singles out irrelevant variables[3].

One advantage of constructing statistical quantities out of pairs of points is a better utilization in available statistics [one gets $N(N-1)/2$ pairs out of N points]. Nevertheless limited statistics can become problematic, especially if the noise in data is large. We estimate statistical errors using the standard estimator

$$\Delta P_d(\epsilon | \vec{\delta}) = 2\sqrt{\frac{P_d(1 - P_d)}{n(\vec{l} \leq \vec{\delta})}}. \tag{14}$$

This expression is not entirely adequate for error estimation when correlations exist in the data points. It however serves the purpose to signal the onset of statistically unreliable regions. In case statistics are at a premium (a usual occurrence) we will adopt an option to set an irrelevant variable k *inactive*, which means that the condition $l_k \leq \delta$ is omitted when computing $P_d(\epsilon | \vec{\delta})$ for $d > k$. By removing the unnecessary restrictive conditions in this way, improved statistics are obtained without affecting the evaluation of P_d's in higher dimensions.

3 Results from the δ-test

In this section we analyze possible dependency structures in multivariate exchange rates data using the δ-test. The data at our disposal are the daily exchange rates with respect to the US dollar for six currencies: Canadian dollar (C), French Franc (F), German Mark (G), Japanese Yen (Y), Swiss Franc (S), and British Pound (P). An additional data set contains the daily closing values of the Dow Jones Industrial Index (W). The same exchange rates data (less the French Franc series) have been analyzed in the context of neural network forecasting by Weigend *et al.* [2], but to our knowledge an analysis on this data set combined with stock index data has not been done before.

Given these data we wish to uncover any nonstochastic structures. To be specific we pick the German Mark G as the target, and examine:

1. How G depends on its own history (autodependencies)?

[3] One should keep in mind however that these conditions are only necessary but not sufficient.

2. Does G depend on other currencies (interdependencies)?
3. Will the stock indices be useful to modeling exchange rates?
4. Is there any predictability in the currency?

Answers to these questions are of large interest by themselves, and they will offer us a starting base for selecting input variables to the ANN models.

A usual first approximation to the exchange rates is a random walk process, in which the optimum prediction to make would be a status quo $G_t = G_{t-1}$. In this simplest scenario the exchange rates are assumed to follow a lognormal distribution, *i.e.* the *rate of return* defined by

$$g_t = \ln \frac{G_t}{G_{t-1}}, \tag{15}$$

is a random Gaussian variable and there would be no serial correlation. The latter property is also a consequence of the weak form of the efficient market theory [6], which states that the current price/return impounds all the information contained in a past history of prices/returns. Since our aim is to probe any correlations and predictability beyond the efficient market theory, it is more appropriate for us to work on the scale of g_t instead of the original price series G_t. We thus convert all the exchange rates and stock index series into corresponding series of *returns* (denoted by corresponding little case letters). In the following all ϵ and δ are specified in unit of σ_g, the standard deviation of the g_t series.

Clearly major policy changes such as the formation of EMU would have profound effects on the inter dependences (if any) between currencies. To avoid this problem to some extent we restrict the data to be between January 1979 and December 1990 (with 3022 data in each time series), a period in which the EMU is reasonably stable.

We remark that the reason we choose these data sets is simply because they are the only ones available to us. We would wish to expand the analysis to include all possibly influencing factors, such as the relevant interest and inflation rates etc, should they become available to us.

3.1 Autodependencies

To investigate the autodependencies we construct vectors in embedding spaces of the type in eq. (2) with the components $z_d := g_{t-d}$. Fig. 2 shows the dependability indices for time lags d up to 16. A true random walk process requires all $\overline{\lambda}_d$ to conform with 0, and this is evidently not the case for the German Mark returns. In particular we can pick out g_{t-1}, g_{t-2}, g_{t-3}, g_{t-8}, g_{t-9}, g_{t-10} as the likely variables that g_t depends upon.

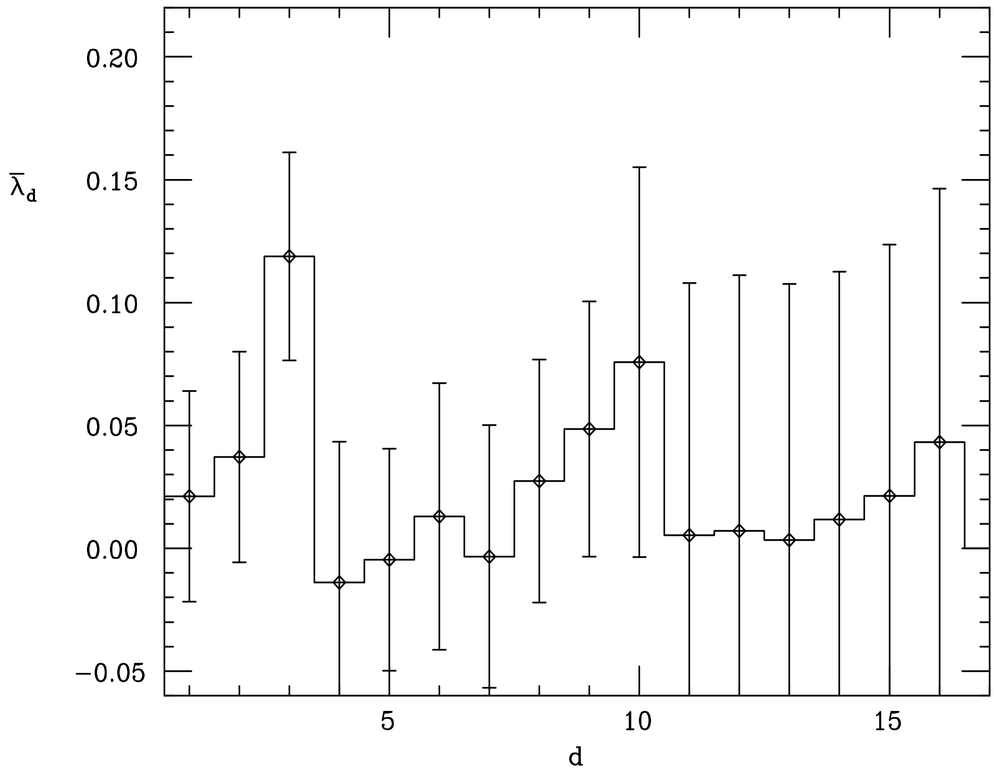

Figure 2: Dependency indices $\overline{\lambda}_d$ for g_t on its time lags g_{t-d}, shown for $d = 1, 2, ...16$.

3.2 Dependences on the Dow Jones Index

Both the currency exchange rate and the stock market are known to be affected by fluctuations in interest rates. So it is not unreasonable to ask if exchange rates in any way depend on the stock performances. Specifically we check if the Deutche Mark returns are related to the past stock returns. The vectors in eq. (2) are assigned the components $z_0 := g_t$ and $z_d := w_{t-d}$, $d = 1, 2, ...$, where w_t is the returns on the Dow at the time step (day) t.

Fig. 3 gives the computed dependability indices on w_{t-d} for d up to 8. This figure brings in a number of features resulting from low statistics. We must proceed with care when interpreting the results.

While the indices appear to be significant for $d = 2$ and $d = 3$, they drop to large negative

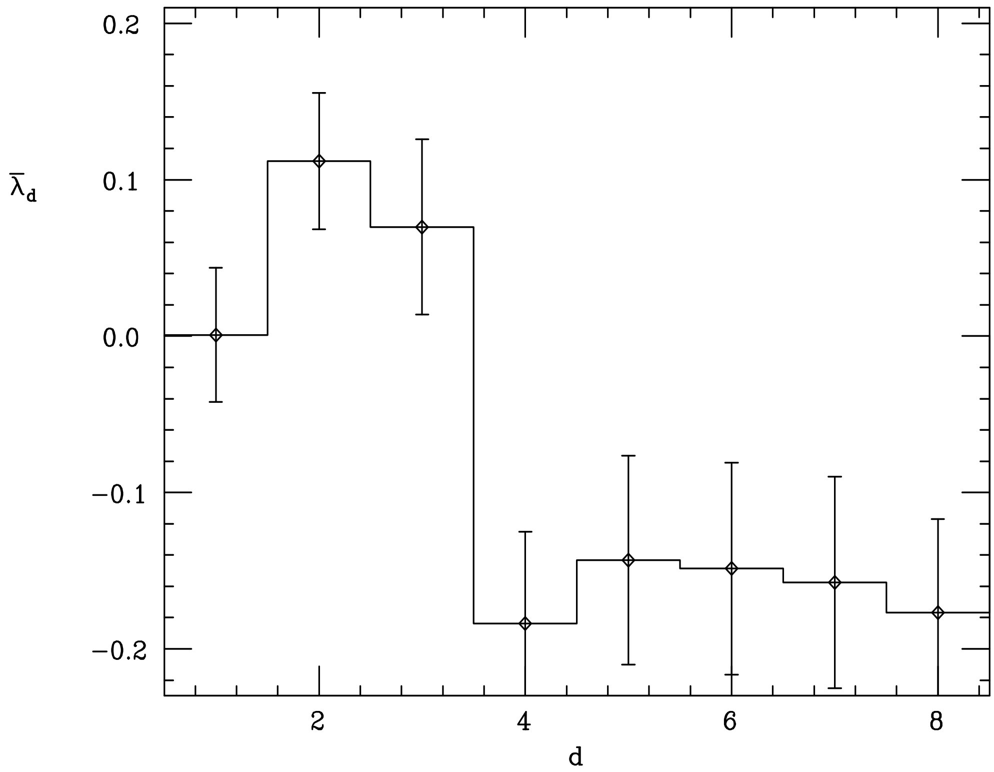

Figure 3: Dependency indices $\overline{\lambda}_d$ for g_t on the stock returns w_{t-d}, shown for $d = 1, 2, ...8$.

values for $d \geq 4$. We recall that in eq. (12) $\overline{\lambda}_d$ is defined to be proportional to the difference between the maxima P_d and P_{d-1}. Statistical fluctuations can result in false identification of these maxima, producing errors that are not accounted for in eq. (14). In this case, say for $d = 4$, either P_3 is identified to be too high, in which case the significance of $\overline{\lambda}_3$ should be discounted, or P_4 is underestimated due to insufficient statistics for probing the region where the maximum can be found. The situation can be resolved by inspecting directly on the conditional probabilities. Fig. 4 plots $P_d(\epsilon|\delta)$ as a function of δ at fixed $\epsilon = 1.7$. We see that the region where $P_2(\epsilon|\delta)$ and $P_3(\epsilon|\delta)$ are large can not be accessed for $d \geq 4$. So the negative indices at $d \geq 4$ should be ignored, and the effects at $d = 2$ and $d = 3$ can be regarded as real.

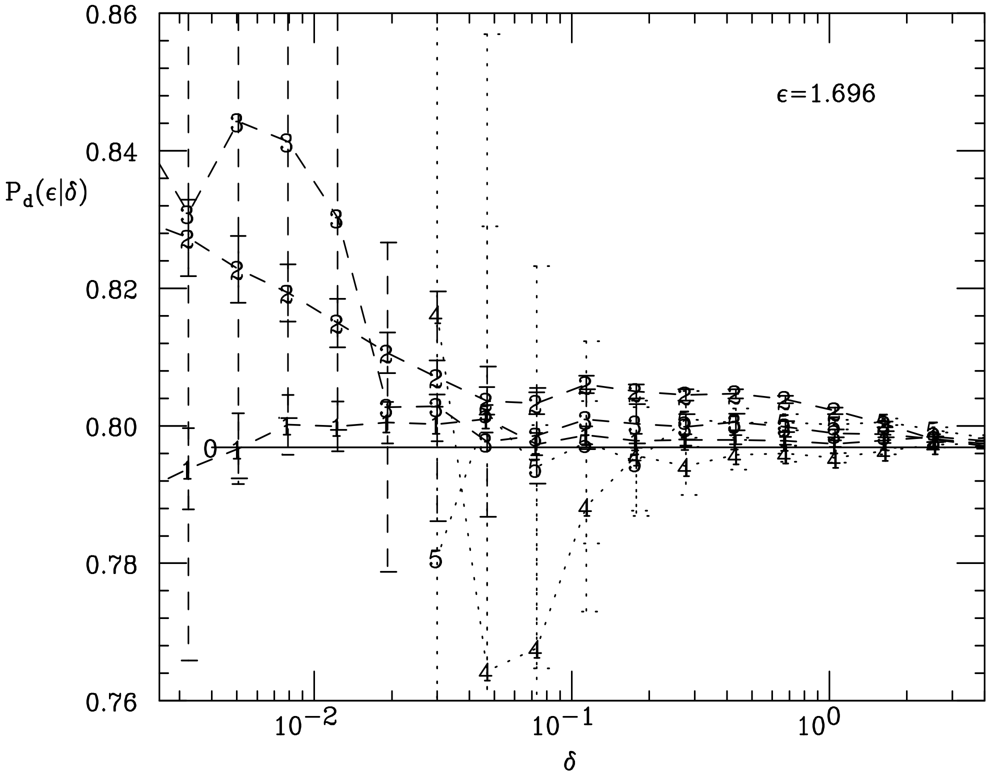

Figure 4: $P_d(\epsilon|\delta)$ as a function of δ for various d (marked on the points). Here $d = 1$, 2,..,5 correspond to w_{t-d}, the returns in the Dow Jones Index.

3.3 Dependences on the other currencies

We now examine whether g_t depends on returns in other currencies. The vectors in eq. (2) are constructed with the components $z_0 := g_t$ and $z_d := o_{t-d}$, $d = 1, 2, ...$, where o_t is the returns in one of the five *other* currencies f_t (French), s_t (Swiss), p_t (Pounds), y_t (Yen) and c_t (Canadian).

Figure 5 shows the dependability indices on f_{t-d}. The Mark appears to be dependent on the French Franc in some appreciable manner. The δ-test suggests dependences on f_{t-1},...,f_{t-8}. Similar analysis on the Swiss Franc (see figure 6) indicates possible dependences on s_{t-1},...,s_{t-6}.

For the British Pound, we find dependences on p_{t-1}, p_{t-2}, p_{t-3} and p_{t-7}, p_{t-10}. The useful variables in terms of the Canadian dollar appear to be c_{t-1}, c_{t-2} and c_{t-3}. Dependences on

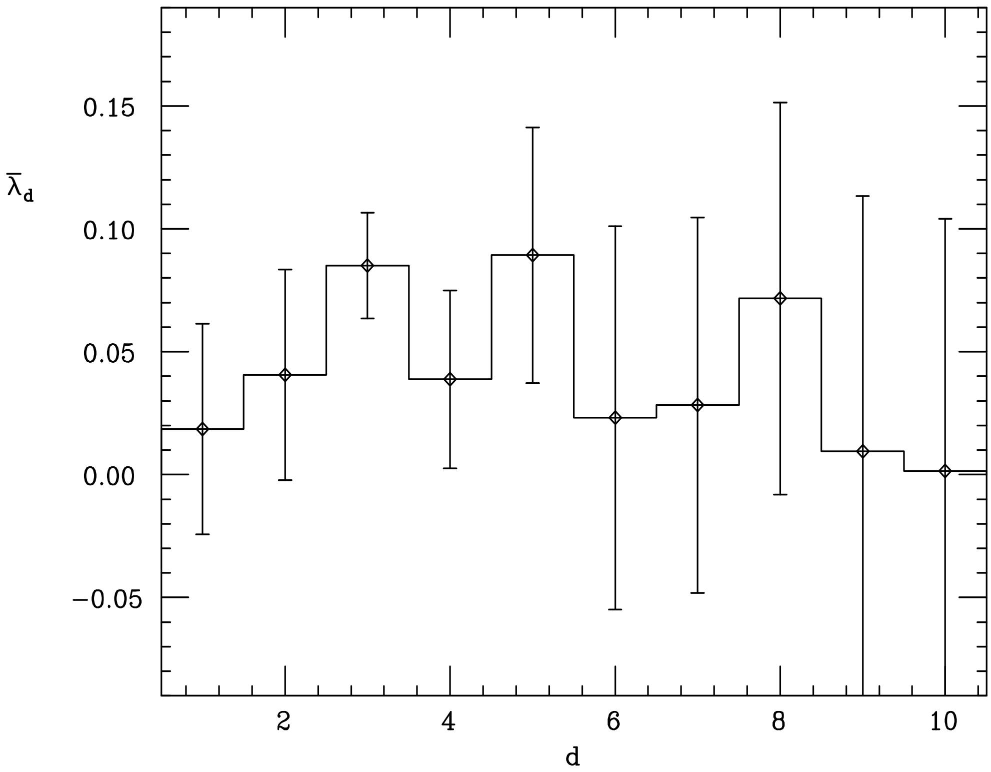

Figure 5: Dependency indices $\overline{\lambda}_d$ for g_t on returns in French Franc f_{t-d}, shown for $d = 1, 2, ...10$.

the Japanese Yen appear to be insignificant and we will neglect the Yen when modeling returns in German Mark.

3.4 Looking from a grand perspective

What we have done so far was examining the dependences of g_t on each set of time lag variables separately. Such an analysis, for example while indicating dependences of g_t on the French Franc and on the DJ index w_{t-d}, does not say whether or not w_{t-d} represents information extra to the information in the French Franc. After all the variables w_{t-d} may be redundant if their information content is contained in the French Franc variables. To recognize these redundancies we need to make a grand analysis taking into account all the variables. As usual we construct d-dimensional vectors with increasing d. We order the

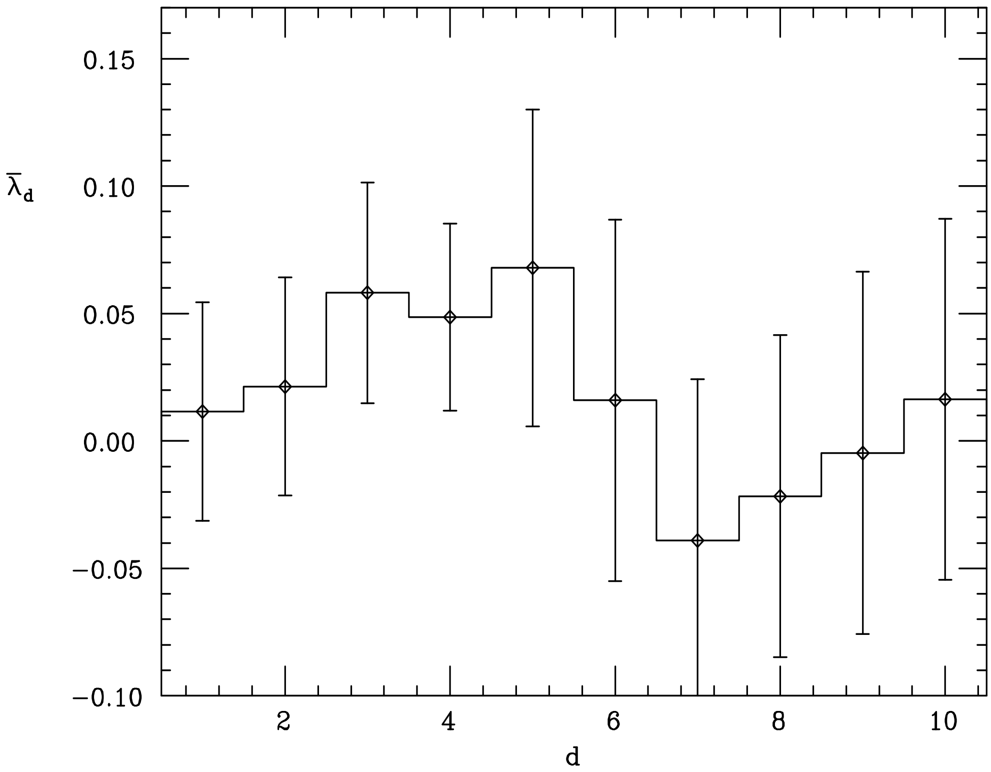

Figure 6: Dependency indices $\overline{\lambda}_d$ for g_t on returns in Swiss Franc s_{t-d}, shown for $d = 1, 2, ...10$.

variables (the components of the vector) in the following manner:

d:	1	2	3	4	5	6	7	8	9	10
	g_{t-1}	f_{t-1}	s_{t-1}	c_{t-1}	y_{t-1}	p_{t-1}	w_{t-1}	g_{t-2}	f_{t-2}	s_{t-2}
d:	11	12	13	14	15	16	17	18	19	20
	c_{t-2}	y_{t-2}	p_{t-2}	w_{t-2}	g_{t-3}	f_{t-3}	s_{t-3}	c_{t-3}	y_{t-3}	p_{t-3}
d:	21	22	23	24	25	26	27	28	29	30
	w_{t-3}	g_{t-4}	f_{t-4}	s_{t-4}	c_{t-4}	y_{t-4}	p_{t-4}	w_{t-4}	g_{t-5}	f_{t-5}
d:	31	32	33	34	35	36	37	38	39	40
	s_{t-5}	c_{t-5}	y_{t-5}	p_{t-5}	w_{t-5}	g_{t-6}	f_{t-6}	s_{t-6}	c_{t-6}	y_{t-6}
d:	41	42	43	44	45	46	47	48	49	50
	p_{t-6}	g_{t-7}	f_{t-7}	s_{t-7}	p_{t-7}	g_{t-8}	f_{t-8}	s_{t-8}	p_{t-8}	g_{t-9}

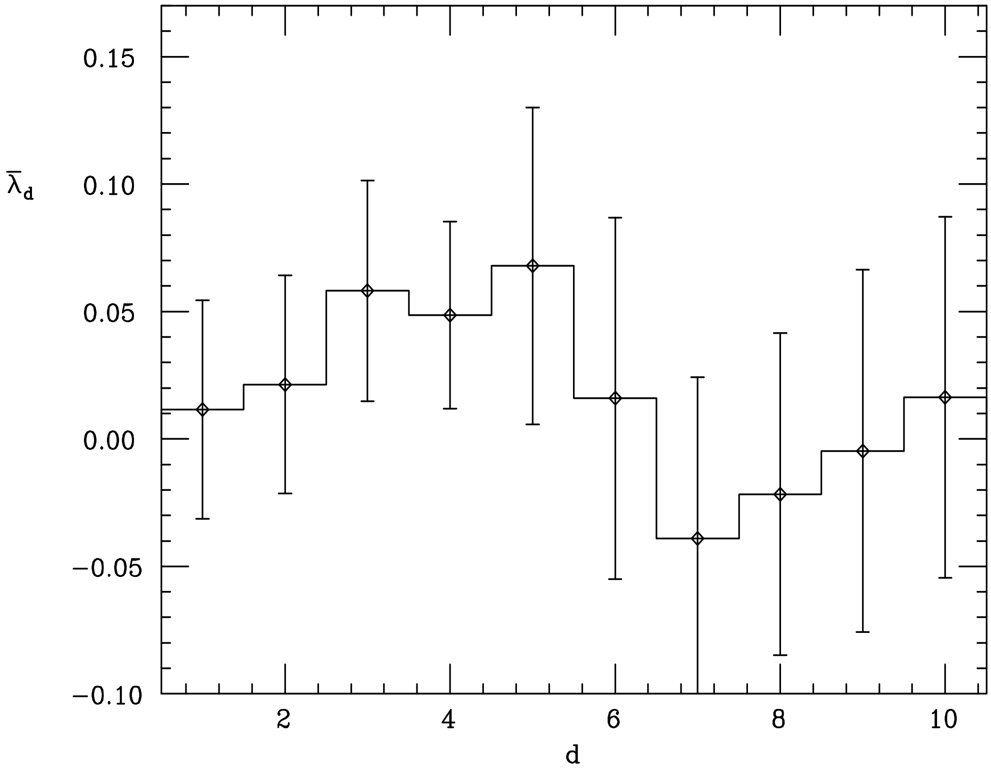

Figure 7: **SORRY, THIS IS A WRONG FIGURE!** Dependency indices $\overline{\lambda}_d$ for g_t on each variable d. See text for the actual variable corresponding to d.

Such a high dimensional analysis would demand huge statistics and it becomes essential to employ the option of removing a restrictive condition once a variable is judged of being irrelevant (see Section 2).

The resulting dependability indices are shown in fig. 7. Appreciable indices appear at $d = 1$, 7, 8, 14, 16, 23, 34, 37, 46 and 50, corresponding to variables g_{t-1}, w_{t-1}, w_{t-2}, g_{t-2}, f_{t-3}, f_{t-4}, p_{t-5}, f_{t-6}, g_{t-8} and g_{t-9} .

4 Forecasting with neural networks

In this section we construct ANN models suitable for making short term predictions for the exchange rate series. We use Multilayer Perceptron (MLP) networks with one hidden layer of 5 sigmoidal $(-1, +1)$ nodes and one linear output node. Following the procedures of Weigend *et al.* [9] and to avoid possible mix up of dynamics from different week days, we restrict the networks to specialize for only one particular week day, chosen as Mondays in the following. The networks utilize available information through a Monday and attempt to predict the return in German Mark for the next day.

The data are divided into three sets. The train set dates between January 1979 and December 1986 and contains 390 Mondays, among which 49 randomly chosen patterns are reserved as a validation set. The *early* test set dates between January 1987 and November 1988 and has 93 Mondays. The *late* test set has 96 Mondays from December 1988 to December 1990. The network is trained with the back propagation algorithm [12] as implemented in the JETNET program [13]. The weights are updated after every 50 patterns being presented to the network. The learning rate is fixed at 0.1. We measure the performance of the network by $NMSE$, which is the usual mean squared error normalized by an appropriate factor for each data set such that $NMSE = 1$ results from predicting status quo (random walk hypothesis).

We use the ten variables identified by the δ-test (see section 3.4) as inputs to the network. Figure 8 shows the learning curves. The validation set is constructed normally for the purpose of signaling an optimum solution, a task it fails to fulfill in this case. Nevertheless the fact that the performances on the test sets are better than status quo shows that the network is able to extract sensible information out of the inputs.

The predictability in the day-to-day exchange rate changes appears to be rather low. Questions on the possibility of further improvement remain open. The results we obtained suggest that there exist dependences of the exchange rates on external factors such as the stock performances. Naturally one would expect there can be more factors playing a role. A more complete analysis would incorporate the relevant interest rates, inflation rates, and the forward exchange rates etc into the information set. By exploring the full dimensionality, uncovering and including all relevant variables improved modeling on currency rates may very well be attainable.

Acknowledgements:

I wish to express my gratitude to Carsten Peterson for discussions and valuable comments on the manuscript. I would also like to thank Blake LeBaron for providing the currency data set, and thank Doug Hewko for making his stock data series available to public.

References

[1] For a recent review on the traditional and neural network models for time series analysis see Gershenfeld, N.A., and Weigend, A.S. 1993. The Future of Time Series. In *Time Series Prediction: Forecasting the Future and Understanding the Past*, edited by A.S. Weigend and N.A. Gershenfeld, 1-70 (Addison-Wesley, 1993).

[2] Weigend, A.S., Huberman, B.A., and Rumelhart, D.E. 1990. Predicting Sunspots and Currency Rates with Connectionist Networks. In *Proc. of the 1990 NATO Workshop on Nonlinear Modeling and Forecasting (Santa Fe)*, edited by M. Casdagli and S. Eubank (Addison-Wesley, 1991).

[3] Refenes, A.N., Azema-Barac, M., Chen, L., and Karoussos, S.A. 1993. Currency Exchange Rate Prediction and Neural Network Design Strategies, *Neural Computing & Applications* **1**, 46-58.

[4] Hoptroff, R.G. 1993. The Principles and Practice of Time Series Forecasting and Business Modelling Using Neural Nets. *Neural Computing & Applications* **1**, 59-66.

[5] Pi, H., and Peterson, C., 1993. Finding the Embedding Dimension and Variable Dependences in Time Series. University of Lund Preprint LU TP 93-4, *to appear in Neural Computation.*

[6] Macdonald, R., Floating Exchange Rates: Theories and Evidence (Unwin Hyman, London 1988).

[7] Ohlsson, M., Pi, H., Peterson, C., Rögnvaldsson, T., and Söderberg, B. 1993. Predicting Utility Loads with Artificial Neural Networks - Methods and Results from "the Great Energy Predictor Shootout". To appear in the *1994 Annual Proceedings of the American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE).*

[8] Weigend, A.S., Huberman, B.A., and Rumelhart, D.E. 1990. Predicting the Future: A Connectionist Approach, *International Journal of Neural Systems* **1**, 193-209.

[9] Weigend, A.S., Rumelhart, D.E., and Huberman, B.A. 1990. Generalization by Weight-Elimination with Application to Forecasting. In *Advances in Neural Information Processing III (NIPS 90)*, edited by R.P. Lippmann, J. Moody and D.S. Touretzky (San Mateo, CA: Morgan Kaufmann 1991).

[10] Nowlan, S.J., and Hinton, G. 1992. Simplifying Neural Networks by Soft Weight-Sharing, *Neural Computation* **4**, 473-493.

[11] Mozer, M.C., 1992. Neural Net Architectures for Temporal Sequence Processing. In *SFI Studies in the Sciences of Complexity, Proc. Vol. XV*, edited by A.S. Weigend and N.A. Gershenfeld (Addison-Wesley, 1993).

[12] Rumelhart, D.E., Hinton, G.E., and Williams, R.J. 1986. Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, edited by D.E. Rumelhart and J.L. McClelland (MIT Press, Cambridge MA, 1986).

[13] Lönnblad, L., Peterson, C., and Rögnvaldsson, T. 1992. Pattern Recognition in High Energy Physics with Artificial Neural Networks - JETNET 2.0, *Computer Physics Communications* **70**, 167-182.

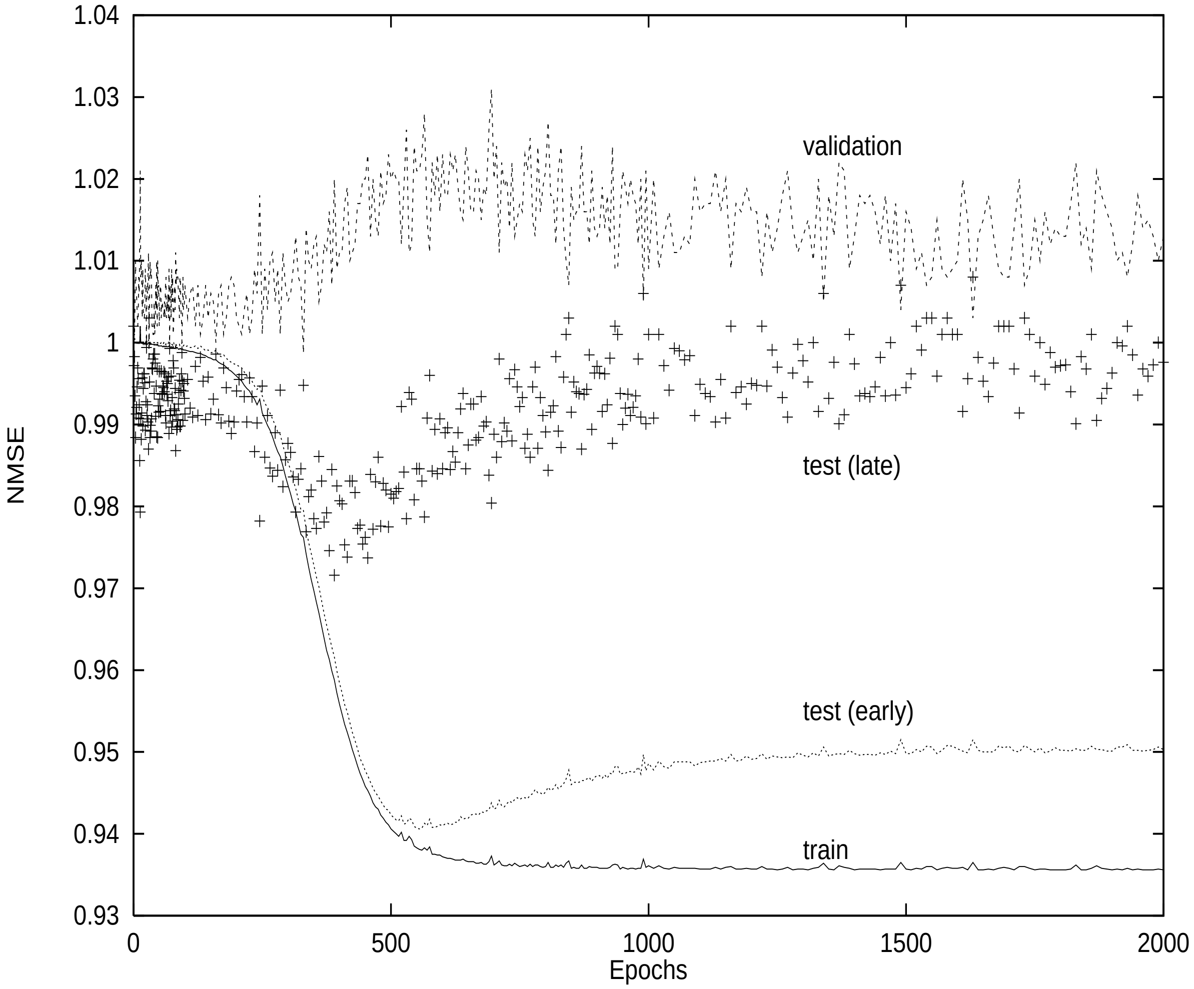

Figure 8: Learning curves plotted as the Normalized Mean Squared Error (NMSE) versus the number of training epochs. The curves are given for the training set, validation set, early and late test set, respectively.