SESO — UNIVERSITAIRE FACULTEITEN ST.-IGNATIUS TE ANTWERPEN

STUDIECENTRUM VOOR ECONOMISCH EN SOCIAAL ONDERZOEK

VAKGROEP MACRO-ECONOMIE

Exchange Rate Modeling by Multivariate Nonlinear Cointegration Analysis using Artificial Neural Networks

A.J.T.M. WEEREN
UFSIA (University of Antwerp)

F. DUMORTIER
UFSIA (University of Antwerp)

J.E.J. PLASMANS
UFSIA (University of Antwerp)
Tilburg University

report 97/343

March 1997

Universitaire Faculteiten St.-Ignatius
Prinsstraat 13 - B 2000 Antwerpen

D/1997/1169/003

Abstract

In this paper we investigate the merits of artificial neural networks in forecasting foreign exchange rates. From previous research it is known that it is hard to beat the random walk model using structural exchange rate models. In this paper we show that by using a suitable multivariate specification a structural model can be derived that beats the random walk. By introducing a new method for multivariate nonlinear cointegration analysis, based on the linear method of Johansen (1988), we construct a neural network error correction model for the yen/dollar, pound/dollar and DM/dollar exchange rates that significantly outperforms both the random walk model and a linear vector error correction model.

1 Introduction

Forecasting exchange rates still is a hot topic in modern financial econometrics. Numerous authors have tried a broad range of advanced econometric techniques but mostly without spectacular success. Since the publication of the paper by Meese and Rogoff (1983), it has become widely accepted that it is virtually impossible to achieve a better prediction performance than can be obtained by using the simple random walk model. Despite this pessimism a lot of effort is put into exchange rate modeling. Roughly one can identify two streams of research. One way of modeling exchange rates is by using advanced methods of time series analysis which produce predictions based primarily on the past movements of the exchange rate process. The other approach involves the use of structural models, which try to forecast exchange rates by making use of economic fundamentals like e.g. interest rates. The main advantage of the second method is the availability of several possible model specifications based on economic theory. The biggest disadvantage however, is the (un)availibility of data on the economic fundamentals. Although high frequency observations on exchange rates are available, most economic fundamentals are only observed on a monthly basis at best. This makes the second approach infeasible if forecasts are needed as part of a trading system. Nevertheless, in this paper we will pursue the second approach. We will use a very general model formulation, introduced by Verkooijen (1996a,b), which encompasses popular exchange rate models like the Flexible Price Monetary Model (FPMM), the Sticky Price Monetary Model (SPMM) as well as a Portfolio Balance Model (PBM). For details on the specification and interpretation of the model see Verkooijen (1996a,b).

One of the main problems one faces when modeling exchange rates is that the observed time series in general appear to be nonstationary. Also for the fundamentals often the null hypothesis of a unit root is not rejected. These findings invalidate the use of standard regression techniques. Instead techniques based on cointegration (see Engle and Granger (1987)) and error correction (see Hendry et al. (1984)) are often used. Many authors have argued that the main limitation of the before mentioned techniques is that these are linear techniques. Granger and Teräsvirta (1993) show that the standard cointegration tests can only find linear relationships between variables. Moreover (see Diebold and Nason (1990)), most authors find that when they test linearity of exchange rate models, this linearity is almost always rejected. Hence generalisation of the concepts of integration and cointegration towards a nonlinear setting is called for (see Granger and Teräsvirta (1993)). Verkooijen (1996b) proposes a neural network test for nonlinear cointegration based on the original Engle-Granger two step methodology. This test however is not successful in the sense that Verkooijen (1996b) does not find a nonlinear cointegrating relationship between the variables in his model. Another limitation of the original Engle-Granger framework is that it is of a univariate nature. Johansen (1988) introduces a more sophisticated, multivariate (linear) approach. Already by using a (linear) multivariate Johansen approach several authors indicate that some promising results for structural exchange rate modeling may be found (see Aftalion and Indjehagopian (1996); Girardin and Marimoutou (1996)). In this paper we will try to follow both ways, i.e. we develop a method which allows for a general nonlinear formulation which is based on a multivariate cointegration. In developing this method we will use the philosophy on nonlinear cointegration as introduced in Granger and Teräsvirta (1993).

The outline of this paper is as follows. In section 2 we will introduce the methodology. In section 3 we will use the approach introduced by Johansen (1988). In section 4 and 5 we will also incorporate nonlinear elements in the analysis. In section 4 nonlinearities only occur in the short term behavior, while in section 5 also nonlinear cointegration is used. Finally in section 6 we will state some conclusions.

2 Nonlinear cointegration

The starting point of our research is Granger and Teräsvirta (1993). In this section we shortly review the method described in Granger and Teräsvirta (1993). The main point raised in (Granger and Teräsvirta, 1993, chapter 5) is the generalisation of the linear notions integration and cointegration in a nonlinear setting.

Definition 2.1 Let X be a stochastic process adapted to the filtration $\{\mathcal{F}_t\}$. X is called *short memory in mean (SMM)* if there exists a random variable D, such that

$$\lim_{h\to\infty} \mathbb{E}(X_{t+h} \mid \mathcal{F}_t) = D,$$

where the distribution of D does not depend on $\mathcal{F}_t$.
A process that is not SMM, is called *long memory in mean (LMM)*.

Remark 2.2 Note that according to this definition a stationary process is always SMM, and that an $I(1)$ process (an integrated process) is LMM. The notions SMM and LMM can so be interpreted as generalisations of the linear notions stationary resp. integrated processes (see Granger and Teräsvirta (1993)). However, there exist processes that are both nonstationary and SMM, for example the process Y defined by $Y_t = 0.8\ \sin(t)\ Y_{t-1}$.

The approach, as suggested in Granger and Teräsvirta (1993), can be translated to the following scheme:

Let Y and X both be LMM processes.

Step 1 Estimate a long run relationship (i.e. nonlinear cointegration)

$$Y_t = f(X_t) + \varepsilon_t, \tag{1}$$

with $\varepsilon \in$ SMM.

Step 2 If there do not exist $d_{1,2} \in \mathbb{N}$, such that $\Delta^{d_1} Y$ and $\Delta^{d_2} X$ are SMM[1], try to find functions ψ_1 and ψ_2 such that $(Y_t - \psi_1(Y_{t-1}, \dots, Y_{t-d_1}))$ and $(X_t - \psi_2(X_{t-1}, \dots, X_{t-d_2}))$ are both SMM; ψ_1 and ψ_2 then determine the operators $\Box_1$ and $\Box_2$.

Step 3 Estimate the nonlinear error correction model (NECM)

$$\Box_1 Y_t = g(Y_{t-1} - f(X_{t-1}), \Box_1 Y_{t-1}, \dots, \Box_1 Y_{t-k}, \Box_2 X_t, \dots, \Box_2 X_{t-\ell}). \tag{2}$$

[1] In (Granger and Teräsvirta, 1993, section 5.5) it is noted that if Z is a LMM process, it does not follow that there exists a $d \in \mathbb{N}$, such that $\Delta^d Z$ is SMM. In theory this complicates the extension of cointegration theory and the specification of error correction models in the nonlinear case. That is why Granger and Teräsvirta (1993) introduce the $\Box$ operators in order to remain close to the linear theory.

Ideally one would like to use in step 1 a direct method to estimate a nonlinear relation of the form (1) such that $\varepsilon \in$ SMM. However, the fact that both processes Y and X are LMM processes complicates this step. For example, if one uses a feedforward artificial neural net to estimate (1), the parameter estimates will not be consistent and in general also the residuals will not be SMM (see White (1992)). So, unless a relationship of the form (1) is already *exactly* known in advance, this step can only be implemented approximately. Consistency of the estimated parameters is too much to ask for in this step, instead one should concentrate on the essential property that $\varepsilon \in$ SMM.

To obtain a long run relation (1) with SMM residuals we will use nonlinear cointegration analysis. Contrary to Verkooijen (1996b), where the univariate Engle-Granger approach towards cointegration is chosen as a starting point, we intend to generalize the well known multivariate Johansen-approach (see Johansen (1988)). The Johansen-method is based on the VAR model:

$$X_t = \Pi_1 X_{t-1} + \ldots + \Pi_k X_{t-k} + \varepsilon_t, \tag{3}$$

where $\Pi_1 \cdots \Pi_k$ are matrices of appropriate dimensions and $\varepsilon \overset{\text{i.i.d.}}{\sim} N(0, \Sigma)$. This VAR can then be rewritten as a VECM

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t, \tag{4}$$

where $\Pi = -(I - \Pi_1 - \ldots - \Pi_k)$. Π can now be factorised as $\Pi = \alpha\beta'$, such that $\beta' X$ is stationary. Now suppose we want to generalize this approach directly to the nonlinear case (1). Then we have a problem, for in general it is not possible to construct a nonlinear VECM out of a nonlinear VAR in order to obtain a (nonlinear) operator Π. Therefore we suggest the following two alternatives:

Alternative 1: Use a linear cointegration analysis as a starting point, which boils down to the assumption of a linear long run relationship of the form $\beta' \begin{pmatrix} Y \\ X \end{pmatrix} \in$ SMM;

Alternative 2: similar to the way Sephton (1994); Verkooijen (1996b) proceed, take the residuals ε_t of (1) as a starting point. As remarked before, we should concentrate on establishing the property that $\varepsilon \in$ SMM. Therefore we estimate using the original Johansen methodology a linear VECM for the residuals[2]

$$\Delta \varepsilon_t = \alpha\beta' \varepsilon_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta \varepsilon_{t-i} + \nu_t, \tag{5}$$

which results in the nonlinear long run relationship $\beta' (Y - f(X)) \in$ SMM. The main problem in using this second alternative is that, just as in Verkooijen (1996b) for a nonlinear Engle-Granger test, we have to construct critical values for a nonlinear LR trace test or a LR $\Lambda_{\max}$ test. Constructing these critical values is computationally very demanding (see Verkooijen (1996b)). Moreover, one should be careful in the interpretation of the matrix β. The only thing we know of this matrix β is that it ensures that $\beta'(Y - f(X)) \in$ SMM.

[2] In fact we should also use a $\Box$-operator in this specification. In this paper however we choose to restrict ourselves to $\Box$-operators of the form Δ^d.

Another problem in using the before mentioned approach in practice is caused by the $\Box$-operators. Although it is possible to construct (nonlinear) processes with a corresponding appropriate $\Box$-operator (see Granger (1995)), it is impossible to find an appropriate $\Box$-operator without assuming a priori that the process is contained in some fixed class of processes. The choice of $\Box$ to be a difference operator Δ^d for instance corresponds to the assumption that the process is a d times integrated SMM process. In this paper we will assume that this is the case, i.e. we will only consider $\Box$-operators of the form Δ^d, although one should be aware that this is a (possibly severe) restriction. As an alternative, one could actually try to estimate the functions ψ_i, for instance using artificial neural nets. In this case one should however realize that the same remark regarding the consistency of the estimates applies as the remarks previously made regarding the estimation of the long run relation (1) in step 1.

Assume that in the first step we have obtained a long run relation either of the form $\beta' \begin{pmatrix} Y \\ X \end{pmatrix} \in \text{SMM}$ or of the form $\beta'(Y - f(X)) \in \text{SMM}$. In the second step we have decided to restrict the $\Box$-operators to the class of difference-operators Δ^d, in which d has to be determined[3]. Now for the estimation of g in (2) a feedforward or recurrent artificial neural net can be used. Note that the parameter estimates in this last artificial neural net will be consistent (see Kuan and White (1992); White (1992)), due to the fact that all involved processes are now within the class of SMM processes.

Remark 2.3 After this procedure a consistent estimate of the long run relationship is obtained in the form, either

$$g(\beta' \begin{pmatrix} Y \\ X \end{pmatrix}, 0, \ldots, 0) \in \text{SMM}, \tag{6}$$

or

$$g(\beta'(Y - f(X)), 0, \ldots, 0) \in \text{SMM}. \tag{7}$$

In this paper we will use both of the before mentioned alternatives in step 1 to construct a structural model for forecasting foreign exchange rates. In section 3 we will estimate a linear vector error correction model using the Johansen approach. In section 4 we use the (linear) cointegration relationship found in section 3 to construct a NNECM (Neural Network Error Correction Model). Then in section 5 we will use the second alternative. Based on the residuals from (1), estimated with a feedforward neural network, we calculate using (5) a nonlinear long run relationship of the form $\beta'(Y - f(X)) \in \text{SMM}$. This long run relationship is then used to estimate another NNECM.

3 Linear multivariate cointegration analysis

In the following sections we will estimate a dynamic structural foreign exchange rate model. We use the methodology as developed by Johansen (1988); Johansen and Juselius (1988),

[3] In this paper we will perform ADF-tests to determine d, but of course also other unit-root tests could have been performed.

implemented in the software package PCFIML 8.10 (Doornik and Hendry (1993a)). As a starting point in deriving the empirical model, we use the bilateral structural exchange rate model specification as essentially[4] used in Verkooijen (1996a):

$$s_t = f(r_t, r_t^*, m_t, m_t^*, ip_t, ip_t^*, \pi_t, \pi_t^*, TB_t, TB_t^*) + \varepsilon_t. \tag{8}$$

This model encompasses popular monetary exchange rate models like the Flexible Price Monetary Model (FPMM), the Sticky Price Monetary Model (SPMM), as well as a Portfolio Balance Model (PBM). In (8) the variables marked with an asterix are the variables of the foreign country (in our case the USA), whereas the variables without asterix concern the home country. In (8) the following variables play a role:

s:	logarithm of nominal exchange rate vis à vis the foreign currency,
r:	nominal short term interest rate,
m:	logarithm of money supply M1,
ip:	logarithm of the industrial production index,
π:	inflation in consumer prices,
TB:	foreign trade balance.

Note that in (8) the function f is left unspecified. In our experiments we extend this model to a multivariate setting in the following way:

$$\begin{aligned}
s_t^{\mathrm{UK}} &= f^{\mathrm{UK}}(r_t^{\mathrm{UK}}, r_t^{\mathrm{US}}, m_t^{\mathrm{UK}}, m_t^{\mathrm{US}}, ip_t^{\mathrm{UK}}, ip_t^{\mathrm{US}}, \pi_t^{\mathrm{UK}}, \pi_t^{\mathrm{US}}, TB_t^{\mathrm{UK}}, TB_t^{\mathrm{US}}, s_t^{\mathrm{JP}}, s_t^{\mathrm{Ger}}) + \varepsilon_t, && (9)\\
s_t^{\mathrm{JP}} &= f^{\mathrm{JP}}(r_t^{\mathrm{JP}}, r_t^{\mathrm{US}}, m_t^{\mathrm{JP}}, m_t^{\mathrm{US}}, ip_t^{\mathrm{JP}}, ip_t^{\mathrm{US}}, \pi_t^{\mathrm{JP}}, \pi_t^{\mathrm{US}}, TB_t^{\mathrm{JP}}, TB_t^{\mathrm{US}}, s_t^{\mathrm{UK}}, s_t^{\mathrm{Ger}}) + \varepsilon_t, && (10)\\
s_t^{\mathrm{Ger}} &= f^{\mathrm{Ger}}(r_t^{\mathrm{Ger}}, r_t^{\mathrm{US}}, m_t^{\mathrm{Ger}}, m_t^{\mathrm{US}}, ip_t^{\mathrm{Ger}}, ip_t^{\mathrm{US}}, \pi_t^{\mathrm{Ger}}, \pi_t^{\mathrm{US}}, TB_t^{\mathrm{Ger}}, TB_t^{\mathrm{US}}, s_t^{\mathrm{UK}}, s_t^{\mathrm{JP}}) + \varepsilon_t. && (11)
\end{aligned}$$

Here the superscripts US, UK, JP and Ger denote the USA, the United Kingdom, Japan and Germany respectively[5]. The nominal exchange rates are all taken as the prices in the home currency for 1 US dollar. In specifying a multivariate model of the form (9–11) we explicitely take into account the direct dependence of the nominal exchange rate of one country on the other exchange rates. Especially in the EMS context it seems reasonable to take (nonlinear) dependence on the other exchange rates into account, which occur e.g. through informal target zones.

3.1 Preliminary data analysis

One possible approach in estimating (9–11) could be the more or less classical approach of imposing linear functions f^{UK}, f^{JP} and f^{Ger}, maybe add some lags, and basically estimate a VAR or VARMA model. This approach would be a valid approach, if all involved time-series were stationary (or in fact SMM). So as a first step in our analysis we have to check whether the variables entering the model are integrated or not. For this purpose we performed an ADF-test on all our variables, the results of which are summarized in table 1 in appendix B. The ADF-tests are all performed using PCGive 8.10 (Doornik and Hendry (1993b)). In these ADF-tests we used a maximal lag length of 13 and we included a trend whenever this was significant at a 5% level. The results are stated using the highest significant lag (at 5%).

[4] Verkooijen (1996a) specifies the model in differentials, e.g. s_t is assumed to depend on $r_t - r_t^*$, etc.

[5] Monthly data, January 1974–June 1995. For a description of the data and their sources see appendix A

From table 1 it follows that all variables, except r^{JP}, appear to be integrated of order 1. Regarding r^{JP}, which is found to be trend stationary, we should however note that discriminating between trend stationary processes and processes that are integrated of order 1 becomes difficult on smaller samples. Based on the results we can conclude that the stationarity assumption is rejected and hence an approach such as the one described in section 2 is called for.

3.2 The Johansen approach

Using PCFIML 8.10 (Doornik and Hendry (1993a)) we will now conduct a cointegration analysis in the spirit of Johansen (1988); Johansen and Juselius (1988). As a starting point we use linear versions of the equations (9–11), including at most 6 lags of each variable. Moreover only the first 246 observations were used, so that the last 12 observations (i.e. July 1994 – June 1995) can be used for out of sample comparisons with the nonlinear models. Based on (9–11) we looked for a system of equations, in which all the variables (and their lags) contribute significantly (using F-tests), and for which there exists at least one cointegrating relationship. After a systematic selection procedure we find a system including the following variables: $s^{\text{JP}}, s^{\text{UK}}, s^{\text{Ger}}, r^{\text{JP}}, m^{\text{US}}, ip^{\text{JP}}$. For this system the results of the cointegration analysis, as calculated by PCFIML 8.10, are listed in table 2. As can be concluded from table 2 the results of the (linear) cointegration analysis are not very clear. Depending on which column one chooses to use, the conclusions of no cointegration ($p = 0$), full rank Π ($p = 6$) or a relationship of rank 1, 2, 3, 4 or 5 all have some evidence. We are particularly interested in a model explaining the three exchange rates s^{JP}, s^{UK} and s^{Ger} in which the other variables are (weakly) exogenous. In column six we see that, using the LR trace test corrected for the degrees of freedom, the hypothesis H_0 : rank ≤ 3 cannot be rejected at a 10% level, whereas H_0 : rank ≤ 2 is rejected at a 10% level. Although any choice of rank can be defended using table 2 we choose the rank 3 model. Following Johansen and Juselius (1988) we tested some restrictions on α and β, in order to establish weak exogeneity of the explanatory variables $r^{\text{JP}}, m^{\text{US}}, ip^{\text{JP}}$. Using the facilities of PCFIML 8.10 for general restriction testing (see table 3), we find the following cointegrating vectors:

$$\hat{\beta}_1' = \begin{pmatrix} 1 & 0.5 & -1 & 0 & 1.239 & -0.47016 \end{pmatrix}, \tag{12}$$

$$\hat{\beta}_2' = \begin{pmatrix} -1 & 1 & 0 & 0.035234 & -0.67513 & 0.20621 \end{pmatrix}, \tag{13}$$

$$\hat{\beta}_3' = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & -5.247 \end{pmatrix}. \tag{14}$$

Using the cointegrating vectors $\hat{\beta}$ given by (12–14), a parsimonious linear VECM of the form (4) is estimated in a second step, using full information maximum likelihood. The results of this estimation are summarized in table 4. Note in particular that there is some evidence for neglected nonlinearities in the linear VECM, as can be concluded from the X_i^2 and $X_i * X_j$ tests (see Doornik and Hendry (1993a)).

Remark 3.1 The process of obtaining a parsimonious linear VECM of the form (4) involves a lot of difficult decisions. In particular the choice of which variables to eliminate and in which order variables are eliminated is always susceptable to discussion. However, it is not possible to give a detailed exposition and motivation of all the involved decisions for the obvious reasons of lack of space.

4 A NNECM based on the linear cointegration analysis

As suggested in section 2, we will estimate a nonlinear VECM based on the linear cointegrating relationship we found in the previous section. In fact this boils down to the assumption of a linear long run relationship, i.e. to the assumption that nonlinearities are only significant for the short term behavior. We will use Artificial Neural Networks (ANNs) (see e.g. White (1992); Kuan and White (1992)) to estimate the NNECM.

4.1 Feedforward Neural Nets

In this subsection we estimate a NNECM of the form (2):

$$\Delta Y_t = f(\hat{\beta}' \begin{pmatrix} Y_{t-1} \\ X_{t-1} \end{pmatrix}, \Delta Y_{t-1}, \Delta X_t, \Delta X_{t-1}, \theta) + \varepsilon_t, \tag{15}$$

in which

$$\begin{aligned} Y_t &:= \begin{pmatrix} s_t^{\mathrm{JP}} & s_t^{\mathrm{UK}} & s_t^{\mathrm{Ger}} \end{pmatrix}', \\ X_t &:= \begin{pmatrix} r_t^{\mathrm{JP}} & m_t^{\mathrm{US}} & ip_t^{\mathrm{JP}} \end{pmatrix}', \end{aligned}$$

and $\hat{\beta}$ is given by (12–14). For the estimation of (15) we choose $f(\cdot,\theta)$ in the class of feedforward ANNs. Specifically, $f(\cdot,\theta)$ is chosen in the class of one hidden layer multi layer perceptron networks. Estimation of the feedforward ANN is carried out with the Neural Net System Identification Toolbox developed by Nørgaard (1995) for Matlab 4.2. For the training of the ANN we use the Levenberg-Marquardt training method (see e.g. Nørgaard (1995); Sjöberg et al. (1995)). In order to prevent overfit we use explicit regularization in the form of weight decay as well as a pruning algorithm (see Nørgaard (1995); Sjöberg et al. (1995) for details). We try to avoid getting stuck in local minima by using multiple restarts. After some trial and error we selected the following parameters: 5 hidden units and a weight decay parameter $\lambda = 0.01$. To our experience, with these parameters about 5 restarts are needed to avoid getting stuck in local minima. For the pruning we used the obs-prune algorithm, introduced by Hassibi and Stork (1993), which is implemented in the toolbox by Nørgaard (1995). This method succesively eliminates the weights which have the smallest contribution in minimizing the validation error. Ultimately the network with the smallest validation error is selected. In order to do this we split up our sample of 258 observations (monthly data, January 1974 – June 1995) in the following way: the first 200 observations are used for the training of the ANN, observation 201–246 are used as a held out validation set for pruning purposes and observation 247–258 are used to evaluate the performance of the final network.

In order to evaluate the performance of the final network, we compare the prediction performance of the NNECM with the predictions made using the linear VECM we estimated in the previous section. As a benchmark we also compared the performance of both models with predictions made by a simple random walk model ($\hat{Y}_t^{RW} = Y_{246},\ t > 246$). As measures of the prediction quality we use:

$$\text{RMSE (Root Mean Square Error)} : \sqrt{\frac{1}{n}\sum\left(\hat{Y}_t - Y_t\right)^2},$$
$$\text{MAE (Mean Absolute Error)} : \frac{1}{n}\sum|\hat{Y}_t - Y_t|.$$

The results are summarized in table 7.

In table 7 we see that the NNECM with linear long run outperforms the random walk model and the linear VECM for all three exchange rates. The linear VECM in its turn already succeeds beating the random walk model. This shows that, at least in the short term, including nonlinearities contributes in enhancing the prediction performance of the error correction model.

5 A nonlinear Johansen approach

In this section we will use the second alternative as discussed in section 2. As we have seen in section 3.2 the results of the linear cointegration analysis (see table 2) were not very clear. Some evidence for the existence of a (linear) cointegration was found, but as remarked before, especially regarding the rank of the relation several choices could have been defended. Therefore it might be a more fruitful approach to look for a nonlinear cointegrating relationship of the form $\beta'(Y - f(X)) \in$ SMM. In this section we proceed as follows. First we will estimate an ANN for equation (1) and calculate its residuals $\hat{\varepsilon}$. These residuals are then used to calculate a VECM of the form (5). From this model we calculate the decomposition $\alpha\beta'$, and the LR test statistics. The resulting long run relationship (if present) is then used to construct a NNECM.

5.1 The long run relationship

Using the toolbox of Nørgaard (1995) we start by estimating a feedforward ANN

$$Y_t = f(X_t, \theta) + \varepsilon_t, \tag{16}$$

in which

$$\begin{aligned} Y_t &:= \begin{pmatrix} s_t^{\text{JP}} & s_t^{\text{UK}} & s_t^{\text{Ger}} \end{pmatrix}', \\ X_t &:= \begin{pmatrix} r_t^{\text{JP}} & m_t^{\text{US}} & ip_t^{\text{JP}} \end{pmatrix}'. \end{aligned}$$

We choose to include the same variables as before in our model in order to be able to compare the results of the NNECM based on a nonlinear cointegrating relationship to the NNECM and the linear VECM we derived in the previous sections. Obviously, it would be better to start from scratch when searching for a long run specification, by using the specification (9–11) and for instance a pruning algorithm. However, this would make a direct comparison with the other models very difficult.

To train the network (16) we again used the Levenberg-Marquardt algorithm, a weight decay of $\lambda = 0.01$ and 5 hidden units. For computational reasons we decided not to use the obs-prune algorithm and to make only three restarts. On the residuals from (16) we fitted the VECM (5) with 4 lags. The LR test statistics are given in table 5. From table 5 we see that there is clear evidence for a rank 2 nonlinear cointegrating relationship

$$\beta' (Y - f(X)) \in \text{SMM},$$

for which we find

$$\hat{\beta} = \begin{pmatrix} -0.0116 & -0.5905 \\ -0.8365 & 0.4045 \\ 0.2918 & 0.5285 \end{pmatrix}. \tag{17}$$

In order to conduct the analysis as summarized in table 5 we had to determine critical values for this nonlinear Johansen procedure. In general it can be expected that the use of ANNs has an influence on the distribution of ε and hence ultimately on the distribution of the LR test statistics. Hence, we have to generate critical values of the test statistics by simulation taking the testing procedure into account (see also Engle and Yoo (1987); Johansen (1988); Sephton (1994); Verkooijen (1996b)). The critical values are produced through 1000 replications of the following procedure:

- Generate p random walks u from a (random coefficient) VAR(k) process e_1 based on a Gaussian white noise innovation process. Note that e_1 is stationary and hence $e_1 \in$ SMM.
- Generate also n random walks y from a (independent) VAR(k) process e_2.
- Construct a nonlinear function f. For this we use an ANN with $10 * N_h$ hidden units and random weights. Due to the universal approximation property (see for instance Sontag (1993); White (1992)) we know that the class of one hidden layer ANNs is a very general class of nonlinear continuous functions.
- Replace $y_1(\cdot), \ldots, y_r(\cdot)$ with $f_1(u), \ldots, f_r(u)$. Then premultiply y with a random full rank $n \times n$ matrix S of norm 1. In this way y and u have a nonlinear cointegrating relation of rank r. The premultiplication with S guarantees that all components of y are related to u, not only the first r components.
- Determine the values of the LR test statistics.

As an illustration of the huge computational demands, for the parameters we used (i.e. 3 outputs, 3 inputs, 4 lags, 5 hidden units, weight decay parameter $\lambda = 0.01$, 3 restarts and no pruning) the time needed to obtain the critical values (see table 6) was about 125 hours[6].

[6]On a Pentium 166MHz machine (16MB RAM) running Matlab 4.2c under Windows 95.

5.2 A NNECM based on nonlinear cointegration

Based on the cointegrating relationship (16) we estimate a NNECM of the form

$$\Delta Y_t = g(\hat{\beta}'\left(Y_{t-1} - f(X_{t-1}, \hat{\theta}^{\text{longrun}})\right), \Delta Y_{t-1}, \Delta X_t, \Delta X_{t-1}, \theta^{\text{ECM}}) + \varepsilon_t, \tag{18}$$

where

$$\begin{aligned} Y_t &:= \begin{pmatrix} s_t^{\text{JP}} & s_t^{\text{UK}} & s_t^{\text{Ger}} \end{pmatrix}', \\ X_t &:= \begin{pmatrix} r_t^{\text{JP}} & m_t^{\text{US}} & ip_t^{\text{JP}} \end{pmatrix}', \end{aligned}$$

and $\hat{\beta}$ is given by (17). Again, for $g(\cdot, \theta^{\text{ECM}})$ we use a feedforward ANN, with five hidden units. This ANN is trained using the Levenberg-Marquardt algorithm, using 5 restarts, weight decay $\lambda = 0.01$, and is pruned using the obs-prune algorithm. The results are also summarized in table 7. From table 7 we see that both NNECMs are capable of beating the random walk model and that they also beat the linear model. The NNECM based on the linear cointegration outperforms the NNECM based on the NNECM based on the nonlinear cointegration. This may indicate that the extension to possible nonlinear long run relationships might (except for Japan) not be so much an improvement as the inclusion of nonlinearities on a short term error correction basis. However, we have to be careful with this conclusion because, as remarked before, we had to restrict ourselves in performing the nonlinear cointegration tests to computational limits. It is very well possible that we would be able to find a more suitable long run relation by comparing some different neural net architectures or by adding more inputs (i.e. by starting from scratch in our search for a long run relation).
As a graphical illustration of the results we have plotted the forecasts of the nonlinear NNECM, the linear VECM and the random walk model in figures 1,2,3. In these figures the superior performance of the NNECM becomes obvious. Based on visual inspection, the dynamic forecasting performance of the NNECM is clearly better than the performance of the linear VECM.

6 Conclusions

In this paper we have extended the test on nonlinear cointegration using artificial neural nets as introduced by Verkooijen (1996b). Instead of using the classical Engle-Granger method as a starting point, we departed from the multivariate approach as introduced by Johansen (1988). In our experiments we saw that already the generalisation towards linear multivariate cointegration leads to promising results for exchange rate forecasting. As we noted in section 2, it is not possible to find a straightforward generalisation of the Johansen approach towards a nonlinear formulation. We suggested two ways out. The first alternative starts from the linear long run relation, which can be found using standard Johansen techniques. Based on this linear long run relation a NNECM is estimated. In the second alternative we specified a new test to find a nonlinear long run relation. From the results of our experiments we can conclude that both nonlinear methods are indeed capable to improve upon the existing linear methodology. We also found that there is indeed some evidence for a nonlinear long run relationship, although the performance of both NNECMs is comparable. The main disadvantage of the second nonlinear method is that it is computationally very demanding to obtain critical values for the nonlinear cointegration test. However, in the exchange rate modeling application we noticed that the results of the nonlinear cointegration tests were less ambiguous than in the linear case in the sense that in this case a clear distinction between the ranks could be made. This suggests that the nonlinear test as proposed in this paper is a useful extension to the econometrician's toolbox.

Of course, one should keep in mind that the models as they are estimated in this paper are intented to be illustrations of the proposed methodology rather than final models to be used for trading or forecasting purposes. As we have noted before, the choice of variables is kept the same over all models to allow for a fair comparison, although it is better to depart with all the variables in the estimation of the NNECM with nonlinear long run relation. Also the effects of the sample time, the choice of different input variables, addition of other exchange rates, the choice between feedforward or recurrent neural networks, etc. need to be studied before a practically relevant model for reliable exchange rate prediction can be obtained.

References

Aftalion, F. and Indjehagopian, J. P. (1996). Purchasing power parity dynamics. In *Forecasting Financial Markets, New Advances for Exchange Rates, Interest Rates and Asset Management*, London. Chemical Bank and Imperial College.

Diebold, F. and Nason, J. (1990). Nonparametric exchange rate prediction? *Journal of International Economics*, 28:315–332.

Doornik, J. A. and Hendry, D. F. (1993a). *PCFIML 8.10.* International Thompson Publishing, London.

Doornik, J. A. and Hendry, D. F. (1993b). *PCGive 8.10.* International Thompson Publishing, London.

Engle, R. and Yoo, S. (1987). Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35:143–159.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55:251–276.

Girardin, E. and Marimoutou, V. (1996). Forecasting daily spot dollar exchange rates: Can long run relationships help? In *Forecasting Financial Markets, New Advances for Exchange Rates, Interest Rates and Asset Management*, London. Chemical Bank and Imperial College.

Granger, C. W. J. (1995). Modelling nonlinear relationships between extended memory variables. *Econometrica*, 63, no. 2:265–279.

Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships.* Oxford University Press.

Hassibi, B. and Stork, D. G. (1993). Second order derivatives for network pruning: Optimal brain surgeon. In S.J. Hanson et al., editor, *NIPS 5.* Morgan Kaufmann, San Mateo.

Hendry, D. F., Pagen, A. R. and Sargan, J. (1984). Dynamic specifications. In Grilliches, Z. and Intrilligator, M., editors, *Handbook of Econometrics*, pages 149–214.

Johansen, S. (1988). Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control*, 12:231–254.

Johansen, S. and Juselius, K. (1988). Hypothesis testing for cointegrating vectors – with an application to the demand for money in Denmark and Finland. Discussion Paper 88–05, Institute of Economics, University of Copenhagen.

Kuan, C. H. and White, H. (1992). Artificial neural networks: An econometric perspective. *Econometric Reviews*, 13, no. 1:1–91.

Meese, R. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out-of-sample? *Journal of International Economics*, 14:3–24.

Nørgaard, M. (1995). Neural network based system identification toolbox. Technical Report 95-E-773, Institute of Automation, Technical University of Denmark. Toolbox available via http://kalman.iau.dtu.dk/Projects/proj/nnsysid.html.

Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics*, 54, no. 3:461–472. Special Issue on Testing Integration and Cointegration.

Sephton, P. (1994). Cointegration tests on MARS. *Computational Economics*, 7, no. 1:23–35.

Sjöberg, J., Zhang, Q., Ljung, L., Beneviste, A., Deylon, B., Glorennec, P. Y., Hjalmarsson, H. and Juditsky, A. (1995). Nonlinear black-box modeling in system identification: A unified overview. *Automatica*, 31, no. 12:1691–1724.

Sontag, E. D. (1993). Neural networks for control. In Trentelman, H. L. and Willems, J. C., editors, *Essays on Control: Perspectives in the Theory and its Applications*, volume 14 of *Progress in Systems and Control Theory*, chapter 10, pages 339–380. Birkhäuser, Boston.

Verkooijen, W. (1996a). A neural network approach to long-run exchange rate prediction. *Computational Economics*, 9, no. 1:51–65.

Verkooijen, W. (1996b). *Neural Networks in Economic Modelling: An Empirical Study.* Ph.D. thesis, Tilburg University. CentER dissertation series, no. 10.

White, H. (1992). *Artificial Neural Networks, Approximation and Learning Theory.* Blackwell, Oxford.

A Data sources

In this appendix we have listed the source of the data used in this paper. All series have been extracted from the data sources of the IMF unless indicated otherwise. All series consist of monthly data ranging from Januari 1974 until June 1995. The dataset is available upon request from the authors (in EXCEL format).

Country	Variable	IMF Code	IMF Name	Note
US	M1			Source: US Department of Commerce
US	Rs	11160C..ZF...	Treasury Bill Rate	
US	cpi	11164...ZF...	Consumer Prices	
US	π			Calculated from cpi
US	ip	11163...ZF...	Wholesale Prices	
US	TB			Source: Datastream OECD
UK	Exr	112..AE.ZF...	Market Rate	
UK	M0			Source: ESRC
UK	Rs	11260C..ZF...	Treasury Bill Rate	
UK	cpi	11264...ZF...	Consumer Prices	
UK	π			Calculated from cpi
UK	ip	11263...ZF...	Prices: Industrial Output	
UK	TB			Source: Datastream OECD
Germany	Exr	134..AE.ZF...	Market Rate	
Germany	M1	13434...ZF...	Money	
Germany	Rs	13460B..ZF...	Call Money Rate	
Germany	cpi	13464...ZF...	Consumer Prices	
Germany	π			Calculated from cpi
Germany	ip	13463...ZF...	Wholesale Prices: Industrial	
Germany	TB			Source: Datastream OECD
Japan	Exr	158..AE.ZF...	Market Rate	
Japan	M1	15834...ZF...	Money	
Japan	Rs	15860B..ZF...	Call Money Rate	
Japan	cpi	15864...ZF...	Consumer Prices	
Japan	π			Calculated from cpi
Japan	ip	15863...ZF...	Wholesale Prices	
Japan	TB			Source: Datastream OECD

B Tables and figures

	US	UK	Japan	Germany
s	I(1)	I(1)	I(1)	I(1)
r	I(1)	I(1)	I(0)+c+t	I(1)
π	I(1)	I(1)	I(1)	I(1)
m	I(1)	I(1)	I(1)	I(1)
ip	I(1)	I(1)	I(1)	I(1)
TB	I(1)	I(1)	I(1)	I(1)

Table 1 : Results ADF-tests. I(1) indicates integrated of order 1, I(0)+c+t means trend stationary

eigenvalue $\hat{\mu}_i$	loglik for	rank
	5302.39	0
0.1687877	5324.57	1
0.1421822	5342.98	2
0.105741	5356.39	3
0.06644593	5364.64	4
0.05204428	5371.05	5
0.02513974	5374.11	6

H_0:rank=p	$-T\log(1-\mu)$	using $T-nm$	90%	95%	$-T\sum\log(1-\mu)$	using $T-nm$	90%	95%
$p=0$	44.37**	37.71	39.5	42.5	143.4***	121.9***	100.1	104.9
$p\leq 1$	36.81**	31.29	33.7	36.4	99.07***	84.21**	73.4	77.7
$p\leq 2$	26.82	22.8	27.8	30.3	62.26***	52.92*	50.7	54.6
$p\leq 3$	16.5	14.03	21.5	23.8	35.44**	30.12	31.4	34.6
$p\leq 4$	12.83	10.9	14.8	16.9	18.94**	16.1	16.1	18.2
$p\leq 5$	6.111**	5.194**	2.6	3.7	6.111**	5.194**	2.6	3.7

*: significant at 10% level
**: significant at 5% level
***: significant at 1% level
Critical values according to PC-Fiml 8.10, Doornik and Hendry (1993a).
These correspond to the critical values as given in Osterwald-Lenum (1992).

Table 2: Results cointegration analysis

$\hat{\beta}'$ eigenvectors:

s^{JP}	s^{UK}	s^{Ger}	r^{JP}	m^{US}	ip^{JP}
1.000	0.5000	−1.000	0.0000	1.239	−0.47016
−1.000	1.000	0.0000	0.035234	−0.67513	0.20621
1.000	1.000	1.000	0.0000	0.0000	−5.247

$\hat{\alpha}$ coefficients:

s^{JP}	−0.14093	0.076288	0.0000
s^{UK}	−0.10353	−0.1465	0.0000
s^{Ger}	0.016186	0.0000	−0.022181
r^{JP}	0.0000	0.0000	0.0000
m^{US}	0.0000	0.0000	0.0000
ip^{JP}	0.0000	0.0000	0.0000

loglik = 5347.0068 unrestricted loglik = 5356.3878
LR-test, rank=3: $\chi^2_{15} = 18.762[0.2247]$

Table 3: Test of cointegrating restrictions.

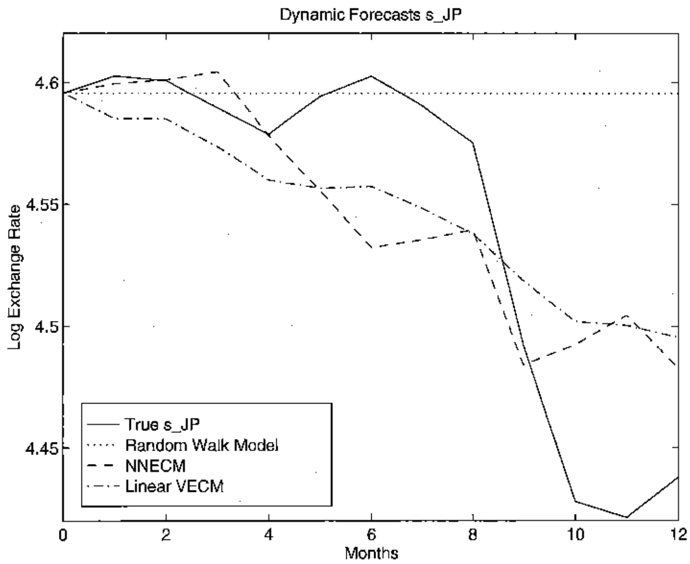

Figure 1: Dynamic forecasts of Yen/Dollar exchange rates

Equation 1 for Δs^{JP}:

Variable	Coefficient	Std.Error	t-value	t-prob	HCSE
Δs^{Ger}	0.774643	0.274037	2.827	0.0051	0.318702
$\hat{\beta}_{1,-1}$	−0.00889887	0.00499203	−1.783	0.0759	0.00483871
$\hat{\beta}_{2,-1}$	0.0928216	0.024299	3.82	0.0002	0.0257778
Δip^{JP}	2.3665	0.322008	7.349	0.0000	0.371961
$\Delta ip^{\mathrm{JP}}_{-1}$	−0.803395	0.320607	−2.506	0.0129	0.324915
Constant	0.908977	0.237696	3.824	0.0002	0.255142
	$\hat{\sigma} = 0.02491848$				

Equation 2 for Δs^{UK}

Variable	Coefficient	Std.Error	t-value	t-prob	HCSE
Δs^{JP}	0.492638	0.281431	1.75	0.0813	0.26731
$\hat{\beta}_{1,-1}$	0.00253636	0.00626514	0.405	0.686	0.00710597
$\hat{\beta}_{2,-1}$	−0.111205	0.0361561	−3.076	0.0023	0.0467275
$\Delta s^{\mathrm{UK}}_{-1}$	0.103889	0.0466032	2.229	0.0267	0.0462267
Δip^{JP}	−1.3593	0.581407	−2.338	0.0202	0.578115
Constant	−1.0086	0.366424	−2.753	0.0064	0.474207
	$\hat{\sigma} = 0.02873829$				

Equation 3 for Δs^{Ger}

Variable	Coefficient	Std.Error	t-value	t-prob	HCSE
$\hat{\beta}_{1,-1}$	−0.0197064	0.0066482	−2.964	0.0033	0.00644646
$\hat{\beta}_{3,-1}$	−0.0102997	0.00351364	−2.931	0.0037	0.00341793
$\Delta r^{\mathrm{JP}}_{-1}$	−0.00824626	0.00486312	−1.696	0.0913	0.00566105
	$\hat{\sigma} = 0.03384052$				

loglik $= 2615.6426 \log|\Omega| = -21.797\ |\Omega| = 3.41723 \cdot 10^{-10}\ T = 240$
LR test of over-identifying restrictions: $\chi^2_{24} = 28.2258[0.2506]$
Test-summary:

Vector portmanteau 12 lags	$= 94.368$
Vector AR 1–7 F^{63}_{633}	$= 0.885311[0.7222]$
Vector Normality χ^2_6	$= 33.89[0.0000]^{**}$
Vector X_i^2, F^{144}_{1205}	$= 1.489[0.0003]^{**}$
Vector $X_i \cdot X_j$, F^{510}_{871}	$= 1.2034[0.0089]^{**}$

Table 4: Linear VECM

H_0: rank=p	$-(T-nm)\log(1-\mu)$	95%	$-(T-nm)\sum\log(1-\mu)$	95%
$p=0$	40.74***	32.91	68.27***	51.93
$p\leq 1$	26.47***	21.87	27.54*	28.03
$p\leq 2$	1.07	13.24	1.07	13.24
	*: significant at 10% level **: significant at 5% level ***: significant at 1% level			

Table 5: Results of nonlinear Johansen test

H_0: rank=p	90%	95%	99%	H_0: rank=p	90%	95%	99%
$p=0$	29.63	32.91	38.69	$p=0$	47.74	51.93	59.67
$p\leq 1$	19.68	21.87	25.68	$p\leq 1$	24.94	28.03	34.31
$p\leq 2$	10.67	13.24	17.37	$p\leq 2$	10.67	13.24	17.37
	$\Lambda_{\max}$ statistic				Trace statistic		

Table 6: Critical values for nonlinear Johansen test

	s^{JP}		s^{UK}		s^{Ger}		Overall
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
NNECM, nonlinear coint.	0.0449	0.0350	0.0311	0.0233	0.0334	0.0281	0.0640
NNECM, linear coint.	0.0463	0.0391	0.0208	0.0171	0.0261	0.0216	0.0571
Linear VECM	0.0443	0.0389	0.0381	0.0346	0.0350	0.0284	0.0681
Random Walk	0.0889	0.0560	0.0325	0.0287	0.0896	0.0707	0.1304

Table 7: RMSE, MAE of predictions

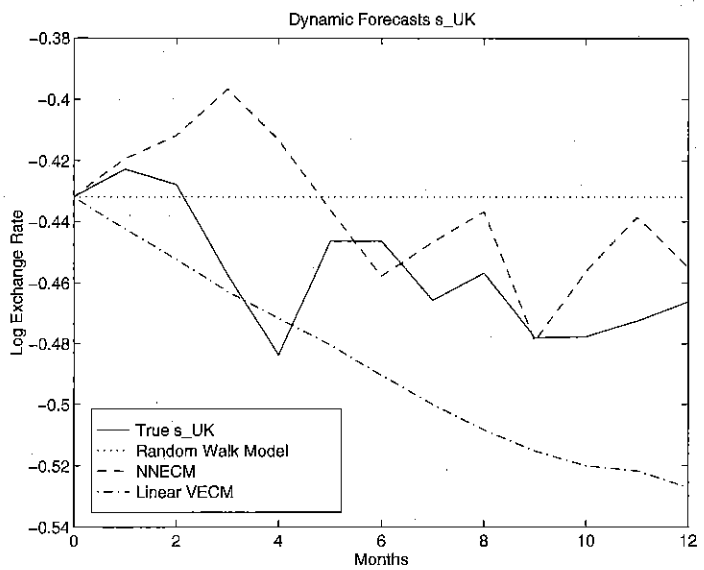

Figure 2: Dynamic forecasts of Pound/Dollar exchange rates

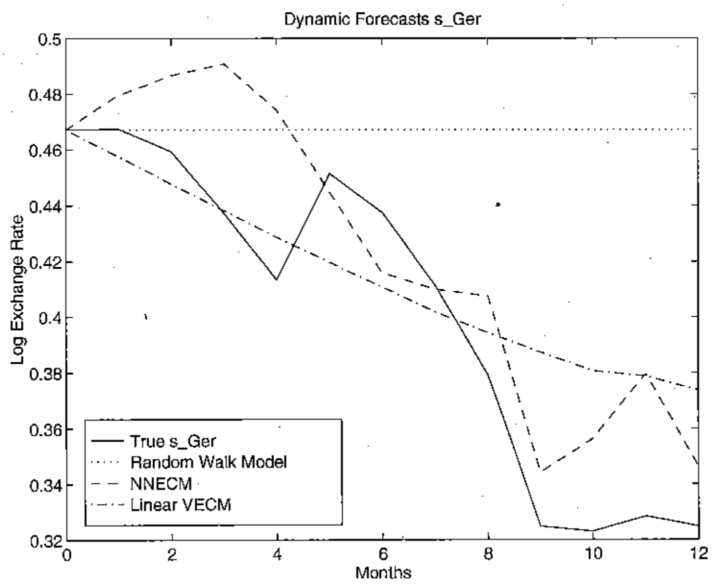

Figure 3: Dynamic forecasts of DM/Dollar exchange rates