STUDIECENTRUM VOOR ECONOMISCH EN SOCIAAL ONDERZOEK

VAKGROEP MACRO-ECONOMIE

Long-run exchange rate determination : a neural network study

W.J.H. Verkooijen*
J.E.J. Plasmans**
H.A.M. Daniëls*

report 95/330

November 1995

* *Tilburg University*
** *UFSIA (University of Antwerp) - e-mail : fte.plasmans.j@alpha.ufsia.ac.be*

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

D/1995/1169/17

Abstract

In the economics literature on exchange rate determination no theory has yet been found that performs well in prediction experiments. Until today the simple random walk model has never been significantly outperformed. We have identified a set of fundamental long-run exchange rate models from the literature. The aim of this paper is to investigate whether neural network (nonlinear) model specification improves prediction performance of the structural exchange rate models, which are traditionally estimated by linear regression methods or by (transfer function) time series methods. The empirical results for the dollar-deutsche mark, dollar-guilder, dollar-pound, and dollar-yen exchange rates, indicate that neglected nonlinearities are not a likely cause for the generally bad prediction performance of the structural exchange rate models.

Key words : foreign exchange rates, long-run prediction, structural models, neural networks

1. Introduction

"It is now recognized that empirical exchange rate models of the post-Bretton Woods era are characterized by parameter instability and dismal forecast performance..." [MR91]. The pessimism about the prediction quality of exchange rate models has become generally accepted after the publication of the influential paper by Meese and Rogoff [MR83]. These authors performed a large number of statistical tests, indicating that not a single economic model of exchange rates was better in predicting bilateral exchange rates during the floating-rate period than the simple random walk model, which posits that all future values of the exchange rate are equal to today's rate.

However, in [MT92]–a good survey paper on exchange rate determination – it is stated that foreign exchange rate participants focus more on fundamentals at longer prediction horizons, and that more attention might be paid to modelling these fundamental determinants of long-term prediction.

Several approaches have been tried to improve the quality of existing structural exchange rate models. Some of these approaches have considered the incorporation of nonlinearities in the models. Diebold and Nason in [DN90], for example, state that "...In summary, there appears to be strong evidence, consistent with rigorous economic theory, that important nonlinearities may be operative in exchange rate determination...". Diebold and Nason further observe that, despite the routinely occurring statistical rejections of linearity in exchange rate models, no nonlinear model has been found in the literature (yet) that can significantly outperform even the simplest linear model in out-of-sample forecasting. Although Diebold and Nason used a powerful nonparametric prediction technique (locally-weighted regression), they were generally unable to improve upon a simple random walk in out-of-sample prediction of ten major dollar spot rates in the post-1973 period in which the dollar exchange rates are floating. Also Meese and Rose [MR91] end up with a negative conclusion: "...we do conclude that incorporating non-linearities into existing structural models of exchange rate determination does not at present appear to be a research strategy which is likely to improve dramatically our ability to understand how exchange rates are determined".

The exchange rate literature usually restricts the application of nonparametric approaches to locally-weighted regression techniques [MR91, MR90, DN90], which are in principle generalisations of the standard nearest neighbour technique. It is generally recognised that nonparametric modelling based on local approximations becomes difficult in high-dimensional spaces due to the increasing sparseness of the data (see [Hub85]). In macroeconomic models most data are typically sparsely distributed; at best data on economic fundamentals are available on a monthly basis, which limits the amount of data available to (say) a few hundred observations. Consequently, the principle of local averaging is likely to fail in macroeconomic modelling problems.

The foregoing does not necessarily imply that model-free regression modelling is impossible in economics. When a low-dimensional representation is embedded in the data, dimensionality reduction methods may successfully be applied. One such method is neural network regression, which we will use throughout this study. During the past few years there is a noticeable increase of neural network applications in economics and finance (see [TT93, BvdBW94]). However, to the best of our knowledge, no studies have been performed yet that apply neural networks to structural exchange rate modelling.

This paper examines whether introducing nonlinearities into theoretical models of exchange rate determination improves the prediction power of these models. In the empirical study neural networks are employed to investigate the nonlinearity hypothesis for the exchange rates of the Japanese yen-US dollar, the British pound-US dollar, the Deutsche mark-US dollar, and the Dutch guilder-US dollar.

More specifically, we will test whether the hypothesised fundamental determinants of the structural models that we consider, do in fact affect the exchange rate, without making auxiliary assumptions on the functional form of the relationship.

The outline of this paper is as follows. Section 2 introduces the theoretical structural exchange rate models which form the basis for the analyses in subsequent sections. In section 3 empirical (testable) models of exchange rate determination are formulated, based on the theoretical models. Section 4 introduces the neural network methodology. In section 5 the characteristics of the collected data are examined. Section 6 outlines the methodology for assessing predictive performance, and examines the long-run and short-run prediction power of the selected exchange rate models, specified in linear and in neural network form respectively. Section 7 concludes the paper.

2. Theoretical Models of Exchange Rate Determination

There are several different theories on exchange rate determination [BM89, MT92]. In many theories two general hypotheses play a prominent role, the Purchasing

Power Parity (PPP-) hypothesis and the Uncovered Interest rate Parity (UIP-) hypothesis. The main idea of the PPP-hypothesis is that exchange rates and national consumption price indices will adjust proportionally so as to maintain a given currency's purchasing power across boundaries, which means that under the assumption of strict PPP the real value of a given currency will be the same in all countries at any moment in time. The UIP-hypothesis states that, in equilibrium, the interest rate differential among countries must be equal to the expected rate of change of the exchange rate. In the next subsections we will make these assumptions more explicit and explain their impacts on exchange rate models.

2.1. The PPP-hypothesis

Assume two countries i and j, each with a bundle of n tradeable goods with average (consumer) prices P_i and P_j:

$$P_i := \sum_{k=1}^{n} \alpha_k p_{i,k} \quad \text{and} \quad P_j := \sum_{k=1}^{n} \beta_k p_{j,k},$$

and define the percentage (consumer) price differential between countries i and j as:

$$dp_{ij} := \log P_i - \log P_j - \log S_{ij},$$

with S_{ij} the nominal exchange rate between i and j's currencies (expressed as units of i's currency per unit of j's currency). Then, under PPP, dp_{ij} is zero if, for example, the bundle weights between the two countries are identical for corresponding goods.

In reality, countries utilise different bundles of goods and price indices $P_i/P_{i,0}$, where 0 indicates the base year. Hence, the percentage (consumer) price *index* differential between countries i and j can be written as:

$$q_{ij} = \log \frac{P_i}{P_{i,0}} - \log \frac{P_j}{P_{j,0}} - \log S_{ij} \qquad (1)$$

To simplify our notation, we will denote $\log(P_i/P_{i,0})$ by p_i, and $\log S_{ij}$ by s_{ij}. Obviously, for any sample observation t, the time differentials satisfy:

$$q_{ij,t} - q_{ij,t-1} = dp_{ij,t} - dp_{ij,t-1},$$

which implies that when modelling in time differences, the distinction between prices and price indices becomes irrelevant. Under the PPP-hypothesis, q_{ij} is assumed to be zero, and the nominal exchange rate satisfying this hypothesis will be denoted, henceforth, by s^*_{ij}. Notice that, due to transportation costs, trade restrictions, speculation, governmental stabilisation policies, etc., the observed (spot) exchange rate s_{ij} will generally be different from the s^*_{ij} satisfying the PPP-hypothesis (see [WP95]).

2.2. The CIP- and UIP-hypotheses

Consider an economic agent who requires a certain amount of foreign currency, say, dollars, for use after a specific period of time, say, one month. If this economic agent is risk averse, this agent is expected to to buy foreign currency now if he expects that buying at the current spot exchange rate is more favourable than buying at the one month's forward rate. This forward rate $f_{ij,t}$ is the rate agreed upon now for an exchange of currencies at an agreed specific future point in time. The consequence of buying at the current spot rate is that the foreign interest rate instead of the domestic interest rate is received (assuming the money is held in a foreign deposit). Since both options are riskless, it is expected that they yield the same rate of return; otherwise, arbitrage would generate riskless profits, thereby assuming that there are no barriers to arbitrage across international financial markets. The forward premium (or the opposite forward discount) at a certain maturity is the percentage difference between the current forward rate of that maturity and the current spot rate. Hence, under the Covered Interest rate Parity hypothesis (CIP-hypothesis) this interest rate differential is assumed to be equal to the forward premium (at any time period):

$$\log f_{ij} - s_{ij} = r_i - r_j, \tag{2}$$

where r_i denotes the nominal (short term) interest rate of country i.

When a trader expects the future spot exchange rate to be lower than the current forward rate, it may be attractive for this trader to wait until the next month; thereby taking the risk of the spot rate being higher than the current forward rate. In this case actors on the forward market are prepared to pay for a risk premium, which equals the difference between the forward rate and the expected future exchange rate. If no risk premium exists in the currency market, which means that the expected future exchange rate and forward rate coincide, CIP implies the Uncovered Interest rate Parity (UIP) condition.

Under the UIP-hypothesis capital markets are assumed to be fully integrated, so that the domestic and the foreign assets are perfect substitutes and international capital is perfectly mobile. Furthermore, financial markets are assumed to be fully efficient. This assumption implies that there are no transaction costs, no differences in national tax systems on capital income, and no risk premia in forward markets. Then, under the UIP-hypothesis, the rates of return on domestic and foreign assets expressed in the same currency are equal:

$$r_{i,t} = r_{j,t} + s^{e}_{ij,t+k} - s_{ij,t}, \tag{3}$$

where the superscript "e" denotes the market's expectation based on information at time t ($s^{e}_{ij,t+k} := E[s_{ij,t+k}|I_t]$, where I_t denotes the information available at time t), and k denotes the period of maturity. The UIP-hypothesis is the cornerstone parity condition for testing foreign exchange rate market efficiency; it assumes rational expectations and risk neutrality. In an efficient market prices should fully reflect the information available to the market participant and it should be impossible for a trader to earn excess returns due to speculation. It is important noticing that only if the nominal interest rate differential is identically equal to a constant and if expectations are rational, the UIP implies a random walk in the exchange rate (with drift if the constant is non-zero). In general, the random walk model is inconsistent with the UIP-hypothesis.

2.3. Monetary and Portfolio Models

Monetary models of exchange rate determination were developed after the collapse of the (Bretton Woods) fixed exchange rate regime in March 1973. These models are descendants of the Mundell-Fleming type of models (see [Mun63, Fle62]).

Several versions of these monetary exchange rate models have been put forward, giving rise to two main types of models. These are the Flexible-Price Monetary Model (FPMM) due to Frenkel [Fre76] and Bilson [Bil78], and the Sticky-Price Monetary Model (SPMM) due to Dornbusch [Dor76] and Frankel [Fra79]. The modelling strategy is similar for both cases. Aggregated macroeconomic relationships are used to obtain a semi-reduced form equation which specifies the level of the (logarithmic) nominal exchange rate as a log-linear function of fundamentals.

The starting point for both basic monetary models of exchange rates is Cagan's money demand function for hyperinflation (see [Cag56]) for any country, where the (logarithmic) demands for real monetary balances are assumed to be linear functions of the (logarithmic) real national income and the nominal interest rate:

$$m^d = p + \alpha y - \beta r + \alpha_0 \quad (\alpha, \beta > 0), \tag{4}$$

with m^d the logarithm of a country's nominal money demand, p the logarithm of the price index, y the logarithm of the real national income, r the nominal short term interest rate level, α the domestic income elasticity, and β the domestic interest rate semi-elasticity of the demand for money.

In the following subsections the two monetary exchange rate models will be explicitly derived and compared. Additionally, the portfolio balance model (PBM), which is non-monetary, will be discussed.

2.3.1. The Flexible-Price Monetary Model (FPMM)

Consider the following assumptions:

1. prices fully adjust such that foreign and domestic commodity markets clear instantaneously,
2. there exists complete equilibrium in the domestic and foreign money markets, or for any country: $m^d = m^s = m$,
3. national incomes are at their full-employment levels,
4. the PPP-hypothesis is (continuously) valid with a corresponding exchange rate s_{ij}^*.

Then the (spot) nominal exchange rate can be expressed under the FPMM by substituting Cagan's money demand function (4) into the PPP-hypothesis of section 2.1, yielding:

$$s_{ij}^* = (\alpha_{0,i} - \alpha_{0,j}) + (m_i - m_j) - \alpha_i y_i + \alpha_j y_j + \beta_i r_i - \beta_j r_j, \qquad (5)$$

which is the fundamental flexible price monetary equation. In this equation an increase in the domestic money supply, relative to the foreign money stock, will lead to a depreciation of the domestic currency in terms of the foreign currency. Also a rise in the domestic real income, other things equal, will lead to an appreciation of the domestic currency. Similarly, a depreciation of the domestic currency follows from an increase in the domestic interest rate.

If the income elasticities on the one side and the interest rate semi-elasticities on the other side are assumed to be equal for both countries ($\alpha_i = \alpha_j$; $\beta_i = \beta_j$), (5) reduces to

$$s_{ij}^* = (\alpha_{0,i} - \alpha_{0,j}) + (m_i - m_j) - \alpha (y_i - y_j) + \beta (r_i - r_j), \qquad (6)$$

where the (logarithmic) nominal exchange rates are determined as a linear combination of differences between domestic and foreign fundamentals.

A basic problem with the FPMM is that it assumes continuous PPP, so that the (logarithm of the) real exchange rate cannot vary over time, even not in the short run.

This is in contrast with reality: although PPP existed during the 1920s, it largely collapsed during the recent floating rate period since March 1973 (see [Fre81] and for further evidence until the end of the 1980s [MP91]).

Therefore, a monetary model for nominal exchange rates with incomplete competition in the market of tradeable goods with sticky prices was needed, at least for the short run. The Sticky-Price Monetary model, treated in the next subsection, remains fundamentally monetary since attention remains focused on equilibrium conditions in the money market.

2.3.2. The Sticky-Price Monetary Model (SPMM)

The SPMM is built on the assumptions of

1. a finite adjustment speed in the commodity market with sluggish prices (sometimes leading to short-term 'overshooting', because of a slow adjustment of these commodity prices; see [Dor76]),
2. clearance of the commodity market in the long run,
3. instantaneous money and asset market equilibrium with perfect substitutability of domestic and foreign non-money assets and perfect capital mobility (reflected in the UIP - hypothesis).

Assume incomplete competition in the commodity market. Then, following Mundell [Mun63] and Fleming [Fle62], country i's commodity demand is assumed to be dependent on variables such as real exchange rates, the real national income of country j, and (short term) real interest rates:

$$y^d_{i,t} = \beta_{0,i} + \beta_{1,i}(s_{ij,t} - p_{i,t} + p_{j,t}) + \beta_{2,i}y_{j,t} - \beta_{3,i}(r_{i,t} - \pi_{i,t}), \tag{7}$$

where $\pi_t := p_{i,t} - p_{i,t-1}$. When country j acts as domestic country, i and j need to be interchanged in equations (8) and (7).

The general principle of SPMM is that prices do not adjust instantaneously. The price-adjustment equation is assumed to be dependent on the commodity market disequilibrium, that is,

$$\pi_{i,t} := \gamma_i(y^d_{i,t} - y_{i,t}), \tag{8}$$

with γ_i the positive price adjustment speed for country i, $y^d_{i,t}$ country i's commodity demand, and $y_{i,t}$ country i's national income. Hence, a shortage of demand will evoke decreasing prices, which, according to (7), will result in a rise of aggregate demand. This process will repeat itself, until the domestic commodity market is cleared; the higher the adjustment speed, the quicker the commodity market equilibrium will be reached.

The exchange rate regime is determined by the UIP-assumption. After an initial disturbance, a new equilibrium exchange rate will emerge in the long run (the 'target-exchange rate' $\bar{s}_{ij}$); in the short run the exchange rate adjustment for country i will take place at adjustment speed θ_i:

$$s^e_{ij,t+1} - s_{ij,t} = \theta_i \left(\bar{s}_{ij} - s_{ij,t}\right), \ \ 0 < \theta_i < 1. \tag{9}$$

Considering again Cagan's money demand function (4), the UIP-hypothesis (3) of expected exchange rate movement being equal to the interest rate spread, and the above relationships (7-9), we find after substitution and definition of the equilibrium commodity price and the long run PPP-hypothesis (see equation (6)):

$$\begin{aligned} p_{i,t} &= m_{i,t} - \alpha_{1,i} y_{i,t} + \alpha_{2,i} r_{j,t} + \alpha_{2,i}\theta_i \left(\bar{s}_{ij} - s_{ij,t}\right) - \alpha_{0,i} && (10) \\ \bar{s}_{ij} &= (m_i - m_j) - \delta_1 (y_i - y_j) + \delta_2 (r_i - r_j) + \delta_0 && (11) \\ s_{ij,t} &= \frac{1}{\beta_{1,i}}[y_i - \beta_{2,i} y_j + \beta_{3,i} r_i - \beta_{3,i} p_{j,t} && (12) \\ &\quad +(\frac{1}{\gamma_i} - \beta_{3,i} + \beta_{1,i}) p_{i,t} - (\frac{1}{\gamma_i} - \beta_{3,i}) p_{i,t-1} - \beta_{0,i}]. \end{aligned}$$

Note that the last equation (12) is also included with country j acting as domestic country, that is, with i interchanged with j.

When the above equations (corresponding to the sticky price monetary model) are written into a single reduced form equation, country i's nominal exchange rate for one unit of country j's currency satisfies

$$s_{ij,t} = f\left(m_{i,t}, m_{j,t}, y_{i,t}, y_{j,t}, r_{i,t}, r_{j,t}, p_{i,t}, p_{i,t-1}, p_{j,t}, p_{j,t-1}\right). \tag{13}$$

As we have already indicated, the above models are called monetary because the equilibrium conditions in the money market are the focus. They also assume perfect substitutability of domestic and foreign non-money assets so that the corresponding markets can be aggregated into a single extra market (of 'bonds'). This perfect substitutability assumption will be relaxed now in the Portfolio Balance Model of exchange rate determination (see [BH85]). This model will be stock-flow consistent, in that it allows for current account imbalances to have a feedback effect on wealth and, hence, on long run equilibrium.

2.3.3. The Portfolio Balance Model (PBM)

The key feature of the PBM is the assumed imperfect substitutability between domestic and foreign assets.

The net financial wealth of the private sector can be subdivided into three components: nominal domestic money M_i, domestically issued bonds B_i, which can be government debt held by the domestic private sector, and foreign bonds B_j denominated in foreign currency and held by domestic residents, which can be interpreted as net claims on foreigners held by the private sector. In a regime of floating exchange rates, a current account surplus on the balance of payments must be exactly matched by a capital account deficit, i.e., by capital outflow and, hence, by an increase in the net foreign indebtedness B_j to the domestic economy. Therefore, current account imbalances will determine exchange rate changes.

Furthermore, the imperfect substitutability of domestic and foreign assets is equivalent to the assumption of a risk premium, separating expected depreciation and the domestic-foreign interest rate differential (implying a collapse of the UIP-hypothesis). In the PBM this risk premium will be a function of relative domestic and foreign debt outstanding.

Summarising, the reduced form equation for the nominal exchange rates may be written under the PBM as:

$$S_{ij,t} = f(M_{i,t}, M_{j,t}, B_{i,t}, B_{j,t}, FB_{i,t}, FB_{j,t}), \tag{14}$$

where $FB_{i,t}$ and $FB_{j,t}$ denote foreign holdings of domestic and foreign bonds, respectively. Taking account of the above mentioned arguments, the four last terms may be replaced by the domestic and foreign accumulated current account surplus.

The derived (logarithmic) nominal exchange rate models (6), (13), and the logarithmic version of (14) can be compared through appropriate statistical testing (Lagrange Multiplier). Sufficient room should be left for a synthesis of the monetary and portfolio balance models, where aspects of various models should be considered simultaneously.

3. Empirical Models

In the previous section we introduced the three main types of structural exchange rate models and discussed the underlying hypotheses. The models are known as the flexible price monetary model, the sticky-price monetary model, and the portfolio balance model. These models are among the models that are often selected in the recent literature [MR83, MR91, MT92, CT95], perhaps due to relative tractability of their data requirements.

Since we examine exchange rates against the US dollar, the notation used in the previous section is slightly simplified; the subscripts i and j are omitted, instead of which all fundamentals corresponding to the U.S. carry a "*" mark.

The three models of exchange rate determination, which we will test empirically in the remainder, are subsumed in

$$s_t = f(r_t - r_t^*, m_t - m_t^*, ip_t - ip_t^*, \pi_t - \pi_t^*, TB_t, TB_t^*) + \epsilon_t, \qquad (15)$$

where s is the logarithm of the bilateral spot exchange rate (e.g. DM/$); $m - m^*$ the logarithm of the relative (ratio of foreign to domestic) nominal money supply; $ip - ip^*$ the logarithm of the relative industrial production; $r - r^*$ the nominal short-term interest rate differential; $\pi - \pi^*$ the inflation rate differential; TB and TB^* the cumulated trade balances, and ϵ is a disturbance term. (It should be noted that, theoretically, GNP is to be preferred as proxy of real income; GNP data, however, are available on a quarterly basis, whereas industrial production data are available on a monthly basis. Therefore, following Meese and Rogoff, we use industrial production data in our experiments.)

The flexible price monetary model (FPMM) includes only the first three terms, that is, $r_t - r_t^*$,$m_t - m_t^*$, and $ip_t - ip_t^*$. The sticky price monetary model (SPMM) adds the inflation rate differential $\pi_t - \pi_t^*$ to this set of variables. The portfolio balance model (PBM) adds the cumulated domestic and foreign trade balances to this set of variables.

Imposing the constraint that domestic and foreign variables (except for trade balances) enter the structural models in differential form, implicitly assumes that the parameters of the corresponding domestic and foreign variables are equal in absolute size, in the case of linear regression. While this parsimoniousness assumption is conventional in empirical applications, it is a potential source of misspecification. In the subsequent sections we will investigate whether this misspecification occurs.

4. Neural Networks

In section 1 we have pointed out that one purpose of the study is to assess the employability of neural networks and their success in predicting exchange rates. In order to provide the necessary background knowledge, we discuss the methodological aspects of neural networks, and indicate some difficulties that may arise when applying neural networks to data modelling problems.

Cognitive scientists have recently developed a class of nonlinear models inspired by the neural architecture of the human brain ('neural network models'). These models are capable of learning through interaction with their environment by a process that can be viewed as a recursive statistical estimation procedure. The neural networks field is developing rapidly since the influential work of Rumelhart and McLelland [RM86]. Neural network models (also known as multi-layer perceptrons) are now widely used for regression and classification purposes.

4.1. Representation

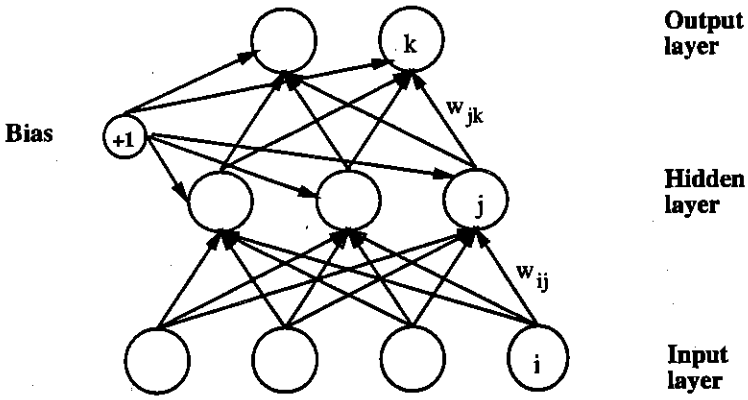

Figure 1: A generic feed-forward neural network with a single hidden layer; to prevent the graph from getting overcrowded, the bias neuron has been removed from the input layer.

A neural network model is a particular type of input-output model. Given an input vector $\mathbf{x}_t = (x_1, \ldots, x_l)'_t$ the network produces an output vector $\hat{\mathbf{y}}_t = (\hat{y}_1, \ldots, \hat{y}_q)'_t$ where q indicates the number of output units and l the number of input units. In statistics it is common practice to denote estimated variables, which neural network outputs in fact are, by a hat. We conform to this notation.

A widely studied network is the feed-forward neural network, which is depicted in figure 1. In graphical form, feed-forward neural networks consist of directed graphs without cycles. Each node represents a "unit", also called artificial neuron, which is the building brick of the artificial neural network. The functionality of each unit is as follows. Each non-input unit j sums its incoming signals and adds a constant term to form the total incoming signal and applies a function ϕ to this total incoming signal to construct the output of the unit. The links have weights

w_{ij} which multiply the signal travelling through them by that factor. Figure 2 shows the functionality of an artificial neuron. The function ϕ, which is called transfer function or squashing function, is usually taken to be logistic (with $\phi(z) = \exp(z)/1 + \exp(z)$). The input units only distribute the input vector, so their ϕ is the identity function.

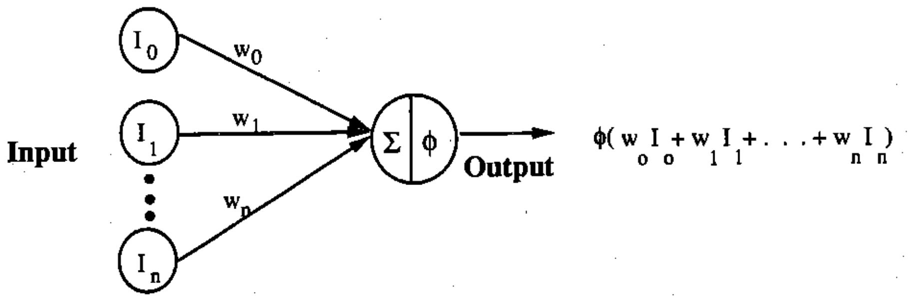

Figure 2: Graphical representation of a neuron

In mathematical notation a feed-forward neural network, as depicted in Figure 1, is expressed by

$$\hat{y}_{kt} = \phi_O(\sum_{j \to k} w_{jk} \phi_H(\sum_{i \to j} w_{ij} x_{it})) \tag{16}$$

where we used $\hat{y}_{kt}$ to denote the value of k'th output unit when input $\mathbf{x}_t$ (the t'th input pattern) is fed into the network, and $\sum_{i \to j}$ stands for the sum over neurons i connected to j. Usually identical squashing functions are used within the same layer; ϕ_O denotes the squashing function of the output units, and ϕ_H denotes the squashing function of the hidden units. The indexing presumes that neurons are numbered sequentially in the order: input units, hidden units, and output unit(s).

It may be preferable to include direct 'skip-layer' connections from the input layer to the output layer to explicitly incorporate the basic linear model, which leads to

$$\hat{y}_{kt} = \phi_O(\sum_{i \to k} w_{ik} x_{it} + \sum_{j \to k} w_{jk} \phi_H(\sum_{i \to j} w_{ij} x_{it})) \tag{17}$$

The expression $f(\mathbf{x}, \mathbf{w})$ is convenient short-hand for network output since this depends only on inputs and weights, given a fixed network architecture. The symbol $\mathbf{x}$ represents the input vector, and the symbol $\mathbf{w}$ represents a vector of all the weights.

In our applications the dimension of the output vector $\hat{y}$ is one, the squashing function of the output unit ϕ_O is assumed to be linear and the squashing function of the hidden units ϕ_H logistic. Let ϕ (without subscript) denote the squashing function of the hidden units. So that (17) can be rewritten in notation that is more familiar to econometricians as follows:

$$\hat{y}_t = \alpha_0 + \alpha^T \mathbf{x}_t + \sum_{i=1}^{N_h} \beta_i \, \phi(\mathbf{w}_i^T \mathbf{x}_t) \tag{18}$$

where N_h denotes the number of hidden units. Note that we have used different parameter names for different parts of the neural network. Without comment $\mathbf{w}$ refers to the whole set of parameters (weights) of the neural network.

Equations (16) and (17) represent quite general classes of functions. A number of authors, e.g. [Cyb89], have shown that single hidden layer feed-forward neural networks with a particular type of nonlinear squashing function (e.g. sigmoid) in the hidden units and linear output units can approximate any continuous function f uniformly on compact sets, by increasing the size of the hidden layer. This justifies the use of neural networks for function approximation and pattern recognition.

4.2. Learning

Given a particular neural network architecture, the role of learning is to find suitable values for the network weights $\mathbf{w}$ to approximate an underlying regression function g of $\mathbf{x}$ by $f(\mathbf{x}, \mathbf{w})$. Traditionally, the weight vector $\mathbf{w}$ is estimated by minimizing the sum of observed squared errors [RHW86]. Suppose we have n observations $\{(\mathbf{x}_t, y_t)\}_{t=1}^{n}$ where y_t represents the "target" value[1] that the neural network should generate when the t'th input vector $\mathbf{x}_t$ is fed in. Then the parameter vector $\mathbf{w}$ is chosen to minimize

$$E(\mathbf{w}) = \sum_{t=1}^{n} [y_t - f(\mathbf{x}_t; \mathbf{w})]^2 \tag{19}$$

as in statistical non-linear regression. This is a general minimization problem, so in principle we can use general purpose algorithms from unconstrained optimisation [Sca85, Nas90]).

The neural network community, however, has developed its own learning algorithm known as error back-propagation. It permits weights to be learned from experience in a process resembling trial and error [RHW86]. Experience is based on the empirical observations on the phenomenon of interest. From now on we assume that the error function $E(\mathbf{w})$ is a differentiable function, which implies that the squashing functions have to be differentiable, thereby excluding threshold units.

[1] In statistical terms this would be called response value.

According to back-propagation, we start with a set of random weights $\mathbf{w}_0$ and then update them by the formula

$$\mathbf{w}_l = \mathbf{w}_{l-1} + \eta \sum_{t=1}^{n} \nabla f(\mathbf{x}_t, \mathbf{w}_{l-1})(y_t - f(\mathbf{x}_t, \mathbf{w}_{l-1})),\ l = 1,\ 2, \ldots \tag{20}$$

where η is a learning rate and ∇f is the gradient (the vector containing the partial derivatives) of f with respect to the weights $\mathbf{w}$ [RHW86]. This simply is the application of the gradient descent technique to the minimization of (19). The distinguishing feature of backprop is that the calculation of $\nabla f(\mathbf{x}, \mathbf{w})(y - f(\mathbf{x}, \mathbf{w}))$ is carried out by a sequence of local computations on the network itself. The specific structure of the neural network is utilized in this calculation. Input data is fed into the network and is forwarded through the connections to the output unit(s), the error between the network output and the desired output is calculated and propagated backwards through the connections. During this backpropagation phase the weights are updated. For a good description of feed-forward neural networks and their specific learning algorithms we refer to [HKP91].

Backpropagation, like any other nonlinear optimisation algorithm, suffers from getting stuck into local minima. Therefore, in practice multiple restarts are performed to ensure a 'good' local minimum has been found. The actual number of restarts employed in practice is generally limited by the computing time required to train a neural network. In our applications we use 5 restarts.

4.3. Generalisation

In practice, the approximation feature of neural networks very often results in estimators with low bias and high variance. Hence, each relationship, characterized by a finite set of observations, can be approximated arbitrarily close by a neural network if enough hidden units are used. Although the trade-off between bias and variance is well-known in statistics, its relevance for neural network regression has been explicitly pointed out only recently [GBD92]. Two concepts from the neural network community which are directly related to the bias/variance dilemma are: generalisation and "overfitting". A network has good generalisation qualities if it performs well on future ('unseen') examples. A network is said to overfit the data if too many characteristics are drawn from the data set at hand. A network that overfits the data sample at hand generally has low bias but high variance, and consequently displays bad generalisation behaviour. The interest obviously is in neural network models that generalize well.

Several approaches have been taken to improve generalisation. A well known approach, known as weight decay, is to add a penalty term to the original error function (19)

$$E'(\mathbf{w}) = E(\mathbf{w}) + \lambda \sum_{ij} w_{ij}^2 \tag{21}$$

and estimate the parameters $\mathbf{w}$ by minimizing this compound error function. Now, each connection w_{ij} has a tendency to decay to zero[2], so that connections disappear unless reinforced. The rationale behind weight decay is the same as that of all shrinkage methods in statistics such as, for example, ridge regression [Rip93, FF93]. So the weight decay parameter λ can be used for the regularisation of bias and variance. A higher λ results in smaller weights, which in turn leads to smoother approximations, and smoother approximators are assumed to display better generalisation.

Another factor that influences both the bias and the variance of the neural network estimator is the size of the hidden layer [GBD92]. This follows directly from the approximation theorem for neural networks. Due to the flexibility of neural networks, it makes no sense to make in-sample comparisons between the modelling errors of linear models and neural network models. It would, namely, be very easy to let neural networks outperform linear alternatives, simply by inserting many units into the hidden layer. Therefore, in the experiments we will focus on out-of-sample instead of in-sample prediction performance.

4.4. Cross-validation

The decay parameter λ in (21) and the number of hidden units, N_h, are typical "smoothing" parameters that have to be "optimized" in some way. Cross-validation [Sto74, HT90, FF93] is a practical approach that is often used for this purpose. It works by leaving points $(\mathbf{x}_i, y_i)$ out one at a time, estimating the weights $\mathbf{w}$ on the remaining $n-1$ points, and calculating the prediction error on the left out point. Finally, the average out-of-sample prediction error is calculated. This is an attempt to mimic the use of training and test samples for prediction. Since neural network training requires a relatively large amount of computation time, in practice not one but a subset of points are left out, where the number of subsets, k, to be left out, is typically chosen to be 5 or 10.

For neural networks calculation of the cross-validation error proceeds as follows. The set of observations, denoted by D, is randomly permuted and decomposed into

[2] This follows directly from the updating of the weights in the direction of the negative gradient of E'.

k mutually exclusive subsets D_i of roughly equal size where $i = 1, \ldots, k$. When working with time series data we do not permute the data, but keep the original time ordering. Initially one repeatedly trains a neural network with different random starting weight vectors on the complete data set until a 'good' (local) minimum $\mathbf{w}^*$ has been obtained. Then the observations from subset D_i are left out from the complete data set[3], the weights are re-trained (starting from $\mathbf{w}^*$), the observations from D_i are predicted, and the average prediction squared error on D_i is calculated. Foregoing is repeated until each subset D_i has been left out from the complete data set. Finally, the average prediction error is calculated from the k repetitions. The cross-validation error, CV, is calculated as follows:

$$CV(\lambda, N_h) = 1/k \sum_{i=1}^{k} \left\{ \frac{1}{|D_i|} \sum_{t \in D_i} \left(f_{\lambda, N_h}(\mathbf{x}_t, \mathbf{w}(D^{-i}, \mathbf{w}^*)) - y_t \right)^2 \right\}. \qquad (22)$$

We have explicitly added to the network output function notation $f(\mathbf{x}_t, \mathbf{w})$ the smoothing parameters λ and N_h, and stressed the dependence of the (re-trained) weights $\mathbf{w}$ on the training set D^{-i} and the starting weights $\mathbf{w}_0$ by including them in parentheses.

Notice that k networks have to be constructed to calculate the cross-validation error for one particular combination of the "smoothing" parameters, which can still be computationally demanding.

In the remainder this cross-validation procedure is used to compare models out-of-sample, and to select the "optimal" set of smoothing parameters.

5. Data and Preliminary Diagnostics

In econometric modelling it is important to distinguish between stationary and non-stationary time series data. The worst consequence of modelling with nonstationary time series data is that standard statistical tests provide evidence for a supposed relationship between economic fundamentals, whereas in fact the relationship is purely spurious. Tests for cointegration have been developed to guard against making these erroneous conclusions (see, e.g., [BDGH93]).

The first step in a modelling excercise therefore incorporates the characterisation of the data. Unit root tests are normally used for this purpose. When the various time series contain a unit root, the next step is to investigate whether these nonstationary time series drift together (are cointegrated) or drift apart (are not cointegrated).

[3] We will use D^{-i} to denote data set D without the observations from D_i

5.1. Data Sources

We take most of the monthly data from the OECD series (using Datastream), including bilateral exchange rates, industrial production index, consumer price index (total), foreign trade balance, money supply (M1), short-term interest rate, and long-term interest rate. The items not available in the OECD series, were taken from the National Accounts. The data source of each variable is reported in Table 8 in the appendix. The Figures 3 through 6 in the appendix depict the variables corresponding to each country; the monthly series range from January 1974 until July 1994.

To facilitate neural network training with weight decay (explained in section 4.3), we have rescaled the data corresponding to each explanatory variable in such a way that at least 95 percent of the data lies within the [0, 1] range and the average equals 0.5. This rescaling makes the signal transferred by each input unit comparable with the outputs of internal units, which is required for weight decay to have effect.

5.2. Unit-roots

In [BDGH93] the characterisation of time series by the order of integration is discussed. We tested all series for possible nonstationarity by simple Augmented Dickey-Fuller (ADF) tests. Table 1 reports the characterisations suggested by these tests for the variables in differential form of the structural exchange rate models. The numerical test results are presented in Table 9 in the appendix.

Table 1: Conclusions based on unit-root tests

	Japan	*U.K.*	*Germany*	*Netherlands*
Exchange Rate	I(1)	I(1)	I(1)	I(1)
Nominal Interest Rate	I(1)	I(1)	I(1)	I(1)
Money Supply	I(1)	I(1)	I(1)	I(0)+c+t
Industrial Production	I(1)	I(0)+c+t	I(0)	I(0)
Inflation	I(1)	I(0)	I(1)	I(1)
Cumulated Trade Balances	I(1)	I(1)	I(1)	I(1)

Most variables are intergrated of order one (I(1)) involving a trend, although the industrial production differential appears to be (trend) stationary, in three out of four cases. Trend stationarity is denoted by "I(0)+c+t" in Table 1. It should be noted that discriminating between a trend stationary series and a random walk with drift, is difficult on the basis of a limited sample.

5.3. Cointegration

The tests for unit roots suggest that most of the variables included in the models, can be assumed to be I(1), although some variables seem to be (trend) stationary. Modelling with *levels* of variables that are I(1) can give misleading results [BDGH93]. The next step tests whether some linear combination of the variables is stationary. The variables are said to be cointegrated if this is the case. Table 2 reports the ADF test for cointegration between the variables in the various models. The cointegrating relationship is estimated in *PcGive*[4] as the long run static solution of a dynamic autoregressive distributed lag (ADL) model including 6 lags[5] for each variable, a constant term, and a trend. The residuals of this static long run solution are then tested for stationarity by ADF tests, using the critical values calculated by MacKinnon [Mac91].

Table 2: Cointegration tests

model	*Japan*	*U.K.*	*Germany*	*Netherlands*	Critical Value
flexible-price	2.09	2.64	2.52	3.05	4.20
sticky-price	2.32	3.57	2.38	4.29	4.50
portfolio	1.79	3.27	1.52	2.61	4.77

note: The critical values are for $\alpha = 0.1$

The number of lags included in the auxiliary regression of the residuals, is determined by the highest significant ($\alpha = 0.05$) lag. No constant term was added, since it was already included in the long-run relationship.

Table 2 shows that in all cases the null hypothesis of no (linear) cointegration cannot be rejected. The collected data do not seem to confirm the three theoretical models of exchange rate determination. This conclusion is not altered when the models are estimated in unrestricted form, which incorporates foreign and domestic variables separately. In particular, the evidence for cointegration is weakened since

[4]Econometric software package developed by Hendry and his co-workers [Hen93].

[5]The number of lags was determined by capacity constraints of PcGive.

the number of variables is doubled; additionally the corresponding critical values are not available in the literature.

Another step in the cointegration analysis is the test for the presence of nonlinear cointegration by neural networks, in the way it is described in the appendix. One main drawback of this approach is the huge computational effort required especially when simulating the critical values, which depend on several neural network parameters. We have to make some concessions regarding optimality and efficiency of the test. To reduce the computational burden, we adopt the same neural network parameters for each exchange rate model and each country. In this way we have to simulate three critical values; for four, five, and seven series. The neural network consists of three hidden units. The weight decay parameter is taken to be 0.001, and the number of observations equals 246. Further, no multiple restarts are employed in the neural network training process. The residuals of the neural network versions of the flexible-price, sticky-price, and portfolio models are tested for a unit root using the neural network ADF test. How the required critical values are actually generated is found in the appendix. Corresponding to the linear ADF tests for cointegration, the number of lags in the neural network ADF test is determined as the highest lag (maximum 13) that is significant at a 5% level. The results are shown in Table 3.

Table 3: Neural network ADF tests

model	*Japan*	*U.K.*	*Germany*	*Netherlands*
flexible-price	4.51	5.16	3.05	2.70
sticky-price	3.54	5.23	4.88	4.73
portfolio	4.04	4.90	3.78	5.03

note: The critical values (α=0.01, 0.05, 0.10) are:
flexible-price: 6.02, 5.46, 5.19
sticky-price: 6.27, 5.73, 5.39
portfolio: 6.67, 6.07, 5.72

The tests for nonlinear cointegration do not reject the null hypothesis of no cointegration at reasonable significance levels. Functional form misspecification does not seem to be an important explanation for the weak evidence for a long-run relationship between the economic fundamentals and the exchange rate.

When determining the number of lags to include in the ADF test of the residuals, we observed that the absolute value of the "t-ADF"-statistic increases, when the number of lags included decreases. The highest values are observed for the "t-DF" statistic (meaning that no lags are incorporated). The DF-statistics would reject the null hypothesis of no cointegration for each model. However, leaving out lagged terms that are significant makes the ADF regression misspecified, which invalidates the DF test.

Nevertheless, Sephton in [Sep94] performs the cointegration tests on the MARS algorithm by a DF test on the residuals, even though the sample sizes of the data employed in their applications seem too small to neglect the effects of the lagged terms in the ADF tests. Sephton's evidence for the existence of nonlinear cointegration can thus be questioned.

6. Predictive Performance Assessment

In this section we investigate the predictive power of the various exchange rate models –both in levels (long-run) and changes (short-run). Our main objective is to examine whether nonlinear specification of the supposed relationship between the economic fundamentals and the exchange rate improves upon the predictive performance of the benchmark random walk model, and upon linear specifications.

As we have indicated already neural networks have the potential to 'overfit' the data. To prevent neural networks from overfitting we apply neural network training with a weight decay term added to the least squares error function (see equation (21). The effect of this is that large weights are penalised. Varying the weight decay parameter from low to high transforms the approximating function from highly flexible to rigid. There exists a value for the weight decay parameter that restricts the network weights so much that the approximating function closely resembles the linear model estimated by OLS; a further increase of the value makes the approximating function resemble 'penalised OLS' (also known as ridge regression). So, by the weight decay parameter we determine the level of flexibility.

Cross-validation has been introduced in section 4 as a procedure for selecting the value of the weight decay parameter. The weight decay value indicated by cross-validation and the corresponding cross-validation MSE, immediately indicate whether (strong) linearities are present in the data, or that even the OLS estimates of the parameters in the linear model have to be shrinked. In the following two subsections we have sometimes used this information to skip the neural network results, when cross-validation indicated that no flexibility was allowed for. In some cases we deliberately choose a weight decay parameter smaller than the one suggested by cross-validation to enforce differences in performance between the linear models and the neural network models; of course risking bad predictions due to overfitting.

6.1. Methodology for Out-of-sample Model Comparison

In line with the Meese and Rogoff study we will compare the models using rolling regressions (also called recursive estimation or running regressions). This means that we start with an initial estimation of the model, using (say) the first n_0 observations. Then we make predictions for the remaining part of our sample (n-n_0). Hereafter, we include the next observation to the parameter estimation set, which now consists of $n_0 + 1$ observations, and again predict the values of the response variables in the remaining set of observations. This procedure is repeated, until the training set equals the total sample. In this way we have constructed a set of $(n - n_0)$ 1-step ahead predictions, $(n - n_0 - 1)$ 2-steps ahead predictions; or generally $n_k = (n - n_0 - k + 1)$ k-steps ahead predictions ($k < n - n_0$).

It must be noted that the structural models require forecasts of their predictor variables in order to generate predictions of the exchange rate. In line with what is usually done in this case, we use the actually realised values of predictor variables. Consequently, the results are optimistically biased.

We take RMSE as our principal criterion for comparison

$$RMSE(k) = \left\{ \sum_{p=n_0}^{N-k} [\hat{y}_{p+k} - y_{p+k}]^2 / (N - k) \right\}^{1/2} \tag{23}$$

where k denotes the prediction horizon (in months), y_{p+k} the observed value of the response variable at time $p + k$, and $\hat{y}_{p+k}$ the estimated response value from a model with parameters estimated from the data set $\{(\mathbf{x}_i, y_i)\}_1^p$.

6.2. Long-Run Prediction

The cointegration analyses indicate that if there is a relationship between the exchange rate and the selected economic fundamentals, it is tenuous at best. In this section we examine whether –despite the weak evidence for cointegration– the exchange rate models are able to tell us more about the future than the random walk model ($\hat{s}_{t+k} = s_t$, $k = 1, 2, \ldots$) does.

When the models in the levels of the variables in fact are spurious, the out-of-sample prediction excercise will show no improvement over the prediction accuracy of the random walk model. Examining the prediction accuracy in levels of variables is regarded as a complementary test to the existence of the supposed (equilibrium) relationships between the economic fundamentals and the exchange rate. Theoretically it is possible that a cointegration relationship escapes the Engle-Granger cointegration test, applied in section 3. Hence, the assumptions of the cointegration test were not all satisfied; some of the variables were I(0) rather than I(1). Furthermore, Mark [Mar95] finds evidence for long-horizon predictability (in levels) of the exchange rate by some economic fundamentals.

To examine the possible existence of a long-run relationship both linear and neural network specification of the exchange rate models in levels are employed. The models are compared on basis of the RMSE criterion described in the previous section. The prediction performance of the random walk model is included as a benchmark in the comparison; random walk models are often used for this purpose in the literature.

The following procedure has been followed in constructing Table 4. The initial models –both linear and neural network– have been estimated on the first 140 observations; including the determination of the weight decay value for the neural networks. The number of hidden units was fixed at four. Five restarts have been used to find a good neural network representation of a particular exchange rate model. The cross-validation procedure to select the weight decay value suggested a value between 0.01 and 0.001; the corresponding cross-validation error was smaller than the cross-validation error of a linear model estimated by OLS. The initial model has then be used to predict the remaining part of the data. Hereafter, the next observation has been added to the estimation set, and the parameters (weights) have been updated using the latest values to depart from. The remaining part of the data has been predicted again. This procedure has been repeated until all observations were in the estimation set. Then, all one-month-ahead predictions have been collected, and the corresponding RMSE has been calculated; the first column in Table 4. The same has been done for 6, 12, 18, and 24 months-ahead predictions, which are shown in the next columns of the table.

From Table 4 two conclusions can be derived. First, no structural exchange rate model –linear or neural network– gave better predictions than the random walk model for prediction horizons up to two years ahead. It should be recalled that actual values were inserted for the independent variables, which makes the results even less promising. The results, however, are in line with the findings of other studies [MR83, MR91, DN90], and support the very weak evidence we found for linear and nonlinear cointegration. Second, the neural networks outperformed the linear models in most cases. However, with a large prediction horizon (18 and 24 months) the neural network's predictions generally are worse than the linear model's predictions. This may be due to extrapolation difficulties, which seems to hurt neural networks more than the linear models.

We also investigated the out-of-sample prediction capacity of the models in unrestricted form, i.e., foreign and domestic variables are included separately. The results are shown in Table 5. Since the neural network predictions closely approximate the predictions from the linear model, we left them out. The most striking observation in Table 5 is that the predictions for the Japanese Yen against the US dollar exchange rate have been improved considerably. Again the random walk models could not be beaten by the structural models, either specified by a linear model or by a neural network. Despite these disappointing results, we have made some observations on modelling for prediction that are worth mentioning.

In unrestricted form the number of variables is doubled, which increases the variance of the OLS-parameter estimates. This may lead to bad long-run predictions. The neural networks weights are determined by minimizing the compound loss function consisting of the sum of squared errors and the sum of squared weights. In section 4 we have called this learning with weight decay. The effect of the penalty term is to reduce the variance of the weights, at the expense of a (somewhat) higher bias. Weight decay is particularly effective in the case of many connections and relatively few observations. Applying biased estimation to the linear model by adding the same penalty term -known as ridge regression- can possibly improve the long-run predictions as well. For instance, when the linear unrestricted flexible-price model for the UK is estimated by penalized OLS, i.e. with a weight decay term of size 10, the corresponding row in Table 5 becomes

flexible-price 0.18 0.19 0.22 0.24 0.26.

The prediction performance has increased significantly. Despite the positive impact regularization has on the predictions, the performance of the random walk model is still out of range.

Another observation concerns the chance of drawing faulty conclusions from the one-period-ahead prediction criterion when used for discerning between the prediction power of neural networks, or flexible regression methods in general, and the prediction power of linear models, in the case of slowly moving I(1) variables. When modelling with I(1) variables, it pays off to overfit the data exemplars in the training set, presumed that the performance assessment is done on one-period-ahead prediction errors. To illustrate this statement, we fitted a redundant neural network (eight hidden units and weight decay value λ=0.0001) to the model for the Yen-Dollar exchange rate including ip, m, ip^*, r^*, and m^*. These particular variables were selected on basis of their sluggishness in changing. The resulting one-period-ahead RMSE was 0.06, which compared to the values of the first part of Table 4 and Table 5, is clearly the best among the structural models for the Yen-Dollar case. The 12, 24, and 36-periods ahead prediction RMSE, however, have dramatically increased to 0.49, 0.94, and 1.56, respectively. Compared to the values of the corresponding rows in Table 4 and Table 5 these values are excessively high.

An intuitive explanation of the foregoing proceeds as follows. Assume the series of interest y_t is generated by

$$y_t = y_{t-1} + \nu_t \quad \nu_t \sim \text{i.i.d.}(0, \sigma^2).$$

Let $\mathbf{x}_t$ denote series presumed to be useful for predicting y_t, but which in reality are *not*. Assume $\mathbf{x}_t$ to be generated by

$$\mathbf{x}_t = \mathbf{x}_{t-1} + E_t \quad E_t \sim \text{i.i.d.}(0, \sigma^2 I).$$

The hypothesized relationship f between y and $\mathbf{x}$ is determined on the data set $\{(y_i, \mathbf{x}_i)\}_{i=1}^{t}$ by a very flexible method on the one hand, and a linear model on the other hand. The flexible method will generally be able to approximate the last observation in the training set closely; assuming that this is the case implies $f(\mathbf{x}_t) \approx y_t$. The linear model will in general be less close to individual observations. Using a flexible f in predicting y_{t+1} given $\mathbf{x}_{t+1}$ then results in

$$\hat{y}_{t+1} = f(\mathbf{x}_{t+1}) = f(\mathbf{x}_t + E_{t+1}) \approx f(\mathbf{x}_t) \approx y_t,$$

assuming E_t is sufficiently small. Since y_t, by construction, is a random walk, i.e., the best prediction of the next value is the present one, the one-step-ahead prediction performance of f will be close to the prediction performance of the random walk model. The linear model, which has a much larger bias, will fit the data less precisely. So in that case $f(\mathbf{x}_t) \approx y_t$ will not hold, making the linear model's prediction performance worse than that of the flexible model. Hence, y_t is the best predictor of y_{t+1} by construction. The investigator should be alert not to conclude from this that y_t is nonlinearly related to $\mathbf{x}_t$ by f, with the argument that combining $\mathbf{x}_t$ nonlinearly yielded better predictions than combining them linearly.

The spurious relationship comes to light when the prediction horizon is enlarged. In case a real fundamental relationship was found the performance would not decrease much. However, if the relationship was spurious the performance would decrease rapidly when increasing the prediction horizon.

6.3. Short-Run Prediction

Previous sections showed no evidence for the presence of a long-run relationship between the exchange rate and the economic fundamentals proposed by theories on exchange rate determination. In this section we examine whether short-run predictions can be made from the various exchange rate models with the variables included in first-differenced form.

Applying standard econometric inference to a dynamic (ADL) form of the exchange rate models with first-differenced variables revealed a strong significance of the exchange rate change in the previous period for all four countries. This is evidence against the most simple theory of "no change" in the level of the exchange rate. We will therefore also consider a univariate model for exchange rate changes. Hereafter we will examine whether adding changes in the economic fundamentals alters the prediction power. Econometric analysis also provides some evidence of economic fundamentals having a significant influence on the exchange rate change. The details on the econometric analysis are presented in the Appendix.

Table 6 presents the RMSEs of one-period-ahead predictions made by the parsimonious models displayed in Table 10, the complete (portfolio) models with two lags for each variable, a univariate time series model; all estimated by OLS and a neural network. Additionally, Table 6 gives results for the random walk model $\Delta s_t = \epsilon_t$ with ϵ_t i.i.d(0,σ^2). The column headings of Table 6 refer to these models respectively. In the recursive estimation procedure the initial models were estimated on the first 180 observations. The neural network versions of the parsimonious models have been estimated with two hidden units and a weight decay parameter 0.1; the neural network versions of the complete models with two hidden units and a weight decay value of 5. These network parameters have been determined by cross-validation. These network parameter values indicate that if nonlinearities are present the effects are tenuous. Hence, the small number of hidden units and the relatively large value of the weight decay parameter suggested by cross-validation are attempts to reduce overfitting, more than exploring real nonlinearities. From Table 6 it is seen that some short-run prediction is possible; the RMSEs of the one-period-ahead predictions were smaller than the RMSEs of the 'no-change' random walk models.

To investigate whether the economic fundamentals have effect on the short-run prediction of the models, a univariate time series AR(6) model is fitted to the data as well; its one-month-ahead prediction results are displayed in the penultimate column of Table 6. Testing the univariate models for neglected nonlinearities by the neural network test, revealed no signs of nonlinearities. The one-month-ahead predictions from a neural network with inputs consisting of six time lags of the exchange rate, were not better than those from the linear univariate model. The prediction results from the univariate model compare favourably with those from the parsimonious structural model. This indicates that the economic fundamentals are of no help in predicting the exchange rates movements.

Are all terms necessary in the AR(6) model? To examine this, we repeated the same prediction procedure with an AR(1) model instead of the AR(6) model. The RMSEs obtained from the AR(1) models for the exchange rate movements are similar to the results obtained from the AR(6) models; in three out of four cases even slightly better. Concluding, the only regularity that has been found in the data is that the next exchange rate movement has the same sign as the previous movement, but is damped by a factor of approximately 0.4.

The rolling prediction experiment has revealed that over the last 62 months of observations some structure is present in the exchange rate movements. The only factor that seemed important is the change in the exchange rate from the previous month. To assess the possible impact of the economic fundamentals that were selected in the parsimonious models in the total period 1974-1994, we performed an additional 10-fold cross-validation test on the linear models. This test has been performed as follows. Two years of monthly observations are repeatedly left out from model estimation, and are then predicted from the model constructed on the remaining observations. The so obtained out-of-sample predictions are then compared with the actual values. The results are shown in Table 7. The table shows which part of the variance in Δs_t is explained by the parsimonious linear models from Table 10 and the univariate model $\Delta s_{t+1} = \alpha_0 + \alpha_1 \Delta s_t$, respectively. We conclude that over the whole period 1974-1994 some economic fundamentals helped a bit in explaining part of the variance in the exchange rate changes. Over the last 5 years, however, their effect on prediction performance was small at best.

As said before, the parsimoniousness assumption, may introduce a misspecification into the exchange rate models. We have examined whether this is indeed the case for the models in first differenced form. Incorporating each domestic and foreign variable separately, as well as two lags of each, has a negligible impact on the prediction quality of the models with first-differenced variables.

7. Conclusions

We have applied neural network specification and linear specification to three structural exchange rate models (flexible price and sticky price monetary models and the portfolio balance model), and have compared the out-of-sample prediction qualities. From this study we conclude the following.

First, no evidence has been found that confirms the existence of a long-run relationship (linear or nonlinear) between the economic fundamentals included in the flexible price, sticky-price, and portfolio models, and the exchange rate. Including foreign and domestic variables as separate explanatory variables did not alter this. Consequently, long-run predictions were worse than predictions made from the 'no-change' model.

Second, when estimated in first differenced form we have found some evidence of a weak structure underlying monthly exchange rate changes. The two main determinants are the previous month's exchange rate change and the change in the interest differential between two countries. The 'no-change' model, which implies that changes in exchange rates are random-and can therefore not be predicted-, is outperformed by linear models for all four countries. Exploring the possible existence of nonlinearities in the short-run models by neural networks, did not show any evidence thereof. Including foreign and domestic variables as separate explanatory variables did not alter this finding.

Third, in general biased estimation improves the prediction quality of the various models, especially for the long-run. The experiments with neural networks revealed the large effect of the weight decay parameter on the prediction quality of the neural network models. In those cases where the neural network showed better prediction performance than the corresponding linear model estimated by OLS, it was the regularization by weight decay rather than the introduction of nonlinearities that was responsible for this. Hence, biased estimation also improved the prediction quality of the linear models considerably, especially when modelling with (nearly) collinear independent variables or with a high number of independent variables and relatively small set of observations.

Exchange rate determination has always been a difficult problem [MR83, MR91, MT92] that is characterised by very weak underlying relationships, which are consequently problematic to quantify empirically by any regression method –neural networks included. Introducing nonlinearities in exchange rate models does not seem to be a research direction where high payoffs can be expected.

Appendix

Data

Table 8 presents the data source of the variables which are used in the structural exchange rate models. The first column gives the variable symbol, the second column the variable description, the third column the unit in which the variable has been measured, the fourth column indicates the published data series it originates from, and the last column refers to the DATASTREAM code. To obtain data series of considerable length we had to switch from money supply definition M1 to M0 for the United Kingdom case.

In figures 3 through 6 the time paths of the variables in the structural exchange rate models are displayed. These monthly series start at January 1974 and end at June 1994.

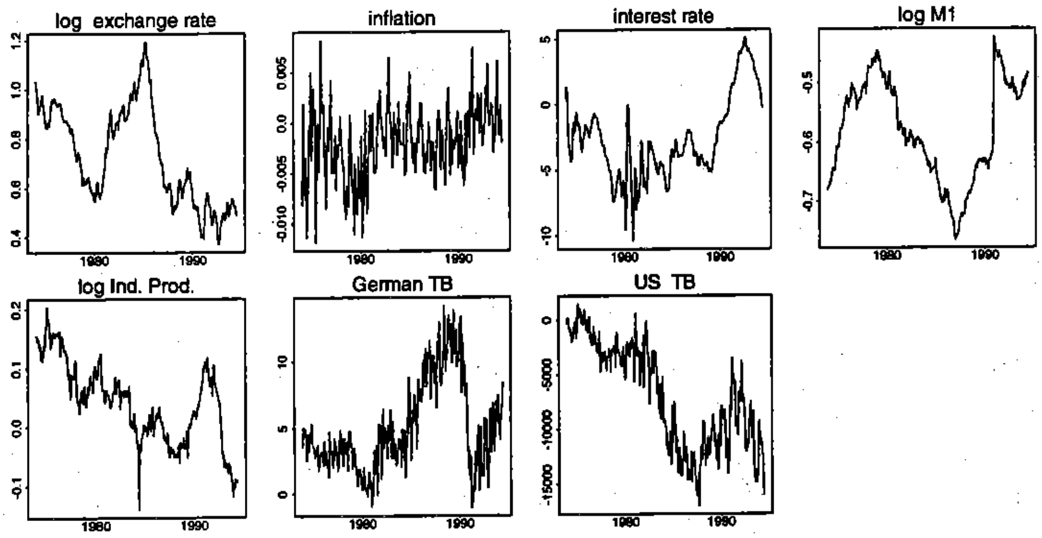

Figure 3: The data for the Germany-US situation; All data (except for the exchange rate and the trade balances) are in differential form (Germany-US).

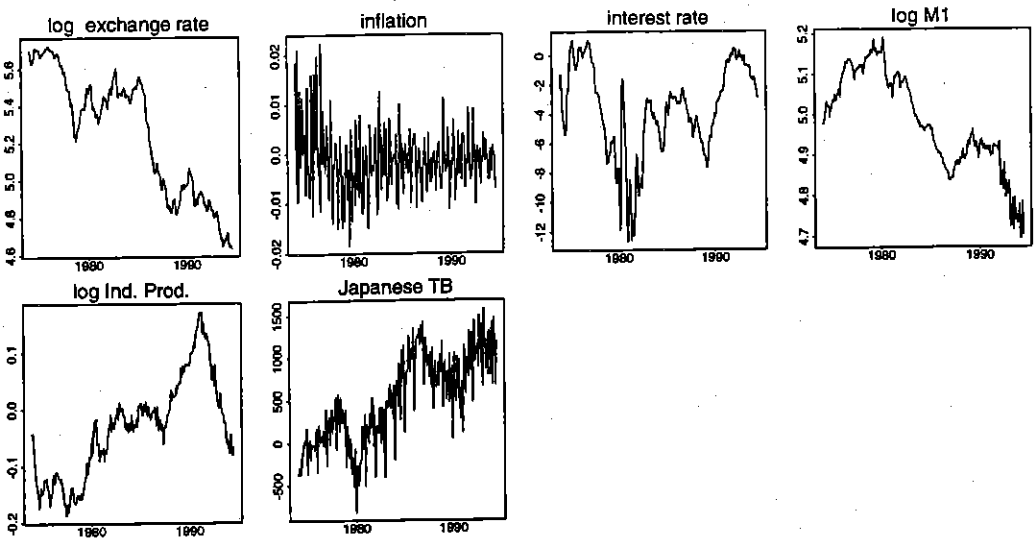

Figure 4: The data for the Japan-US situation; All data (except for the exchange rate and the trade balances) are in differential form (Japan-US).

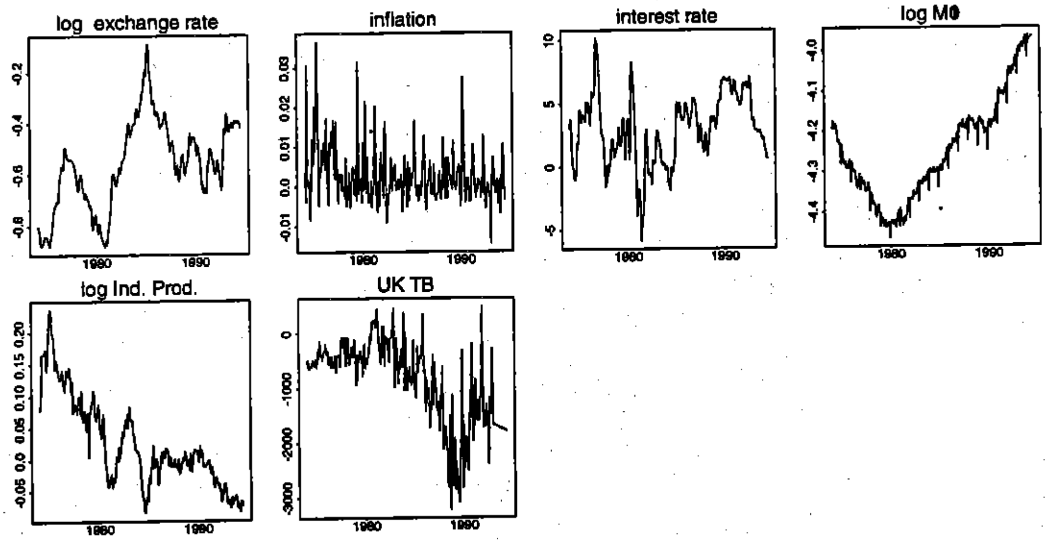

Figure 5: The data for the United Kingdom-US situation; All data (except for the exchange rate and the trade balances) are in differential form (UK-US).

Table 4: Long-run predictions

model	1 month	6 months	12 months	18 months	24 months
			Japan		
flexible-price	0.23/0.09	0.26/0.15	0.30/0.22	0.33/0.28	0.35/0.34
sticky-price	0.23/0.10	0.27/0.19	0.31/0.25	0.34/0.32	0.38/0.38
portfolio	0.23/0.09	0.27/0.17	0.31/0.24	0.35/0.33	0.38/0.38
random walk	0.03/0.03	0.08/0.08	0.11/0.11	0.14/0.14	0.17/0.17
			United Kingdom		
flexible-price	0.21/0.11	0.24/0.15	0.28/0.17	0.32/0.19	0.35/0.24
sticky-price	0.20/0.15	0.23/0.13	0.26/0.17	0.29/0.20	0.31/0.21
portfolio	0.19/0.12	0.23/0.15	0.26/0.16	0.31/0.17	0.33/0.21
random walk	0.04/0.04	0.13/0.13	0.17/0.17	0.19/0.19	0.20/0.20
			Germany		
flexible-price	0.30/0.25	0.35/0.38	0.39/0.49	0.41/0.56	0.41/0.57
sticky-price	0.29/0.18	0.35/0.27	0.41/0.36	0.44/0.41	0.44/0.44
portfolio	0.27/0.18	0.33/0.27	0.38/0.36	0.40/0.43	0.41/0.45
random walk	0.04/0.04	0.12/0.12	0.16/0.16	0.20/0.20	0.20/0.20
			The Netherlands		
flexible-price	0.21/0.21	0.25/0.28	0.28/0.34	0.30/0.38	0.32/0.41
sticky-price	0.17/0.16	0.20/0.24	0.23/0.34	0.24/0.40	0.24/0.43
portfolio	0.20/0.13	0.23/0.19	0.27/0.23	0.29/0.27	0.29/0.29
random walk	0.04/0.04	0.11/0.11	0.15/0.15	0.19/0.19	0.19/0.19

note: displayed values are: RMSE(OLS)/RMSE(neural network)

Table 5: Long-run predictions with unrestricted models

model	1 month	6 months	12 months	18 months	24 months
			Japan		
flexible-price	0.13	0.16	0.20	0.23	0.26
sticky-price	0.10	0.13	0.14	0.17	0.18
portfolio	0.10	0.13	0.14	0.16	0.18
			United Kingdom		
flexible-price	0.25	0.30	0.35	0.40	0.44
sticky-price	0.23	0.28	0.33	0.38	0.42
portfolio	0.20	0.24	0.27	0.32	0.35
			Germany		
flexible-price	0.23	0.30	0.35	0.35	0.34
sticky-price	0.22	0.28	0.33	0.34	0.34
portfolio	0.24	0.27	0.32	0.33	0.32
			The Netherlands		
flexible-price	0.21	0.26	0.30	0.32	0.33
sticky-price	0.20	0.24	0.28	0.29	0.31
portfolio	0.19	0.22	0.24	0.25	0.24

note: displayed values are: RMSE(OLS)

Table 6: Short-Run Predictions

country	parsimonious (OLS/NN)	complete (OLS/NN)	univariate (OLS/NN)	random walk
Japan	0.221/0.221	0.230/0.214	0.217/0.216	0.223
UK	0.289/0.291	0.305/0.303	0.290/0.290	0.323
Germany	0.247/0.224	0.269/0.254	0.243/0.273	0.264
Netherlands	0.250/0.250	0.266/0.259	0.248/0.248	0.269

Table 7: Out-of-sample explained variance -R^2

country	R^2-parsimonious	R^2-univariate
Japan	0.18	0.11
UK	0.26	0.15
Germany	0.22	0.14
Netherlands	0.20	0.11

Table 8: The sources of the data

variable	description	unit	series	code
		United States		
cpi	Consumer Prices	Index	OECD	USOCPCONF
r_s	short-term interest rate	Percentage	OECD	USOCTBL%
r_l	long-term interest rate	Percentage	OECD	USOCLNG%
m	money supply M1	US $ Bln	OECD	USOCM1MNA
	money supply M1	US $ Bln	OECD	USOCM1MNB
	monetary base M0	US $ Bln	GOV	USMONBASA
ip	industrial production -total	Index	OECD	USOCIPRDG
TB	Foreign Trade Balance	US $ Mln	OECD	USOCVBALA
		Germany		
cpi	Consumer Prices	Index	OECD	BDOCPCONF
r_s	short-term interest rate	Percentage	OECD	BDOCTBL%
r_l	long-term interest rate	Percentage	OECD	BDOCLNG%
m	money supply M1	DM Bln	OECD	BDOCM1MNB
ip	industrial production -total	Index	OECD	BDOCIPRDG
TB	Foreign Trade Balance	DM Bln	OECD	BDOCVBALA
		United Kingdom		
s	exchange rate -Pound to 1 US $		GOV	USX$UK..
cpi	Consumer Prices	Index	OECD	UKOCPCONF
r_s	short-term interest rate	Percentage	OECD	UKOCTBL%
r_l	long-term interest rate	Percentage	OECD	UKOCLNG%
m	money supply M0	Pound Bln	GOV	UKM0....A
ip	industrial production -total	Index	OECD	UKOCIPRDG
TB	Foreign Trade Balance	Pound Mln	OECD	UKOCVBALA
		Netherlands		
s	exchange rate -DFL to 1 US $		GOV	USX$DFL
cpi	Consumer Prices	Index	OECD	NLOCPCONF
r_s	short-term interest rate	Percentage	GOV	NLEURO3
r_l	long-term interest rate	Percentage	IMF	NLI61...
m	money supply M1	DFL Bln	OECD	NLOCM1MNA
ip	industrial production -total	Index	OECD	NLOCIPRDG
TB	Foreign Trade Balance	DFL Mln	OECD	NLOCVBALA
		Japan		
s	exchange rate -Yen to 1 US $		GOV	USX$YEN
cpi	Consumer Prices	Index	OECD	JPOCPCONF
r_s	short-term interest rate	Percentage	OECD	JPOCTBL%
r_l	long-term interest rate	Percentage	OECD	JPOCLNG%
m	money supply M1	Yen Bln	OECD	JPOCM1MNB
ip	industrial production -total	Index	OECD	JPOCIPRDG
TB	Foreign Trade Balance	Yen Mln	OECD	JPOCVBALA

note M0 is the money base; M1 adds money of account to M0

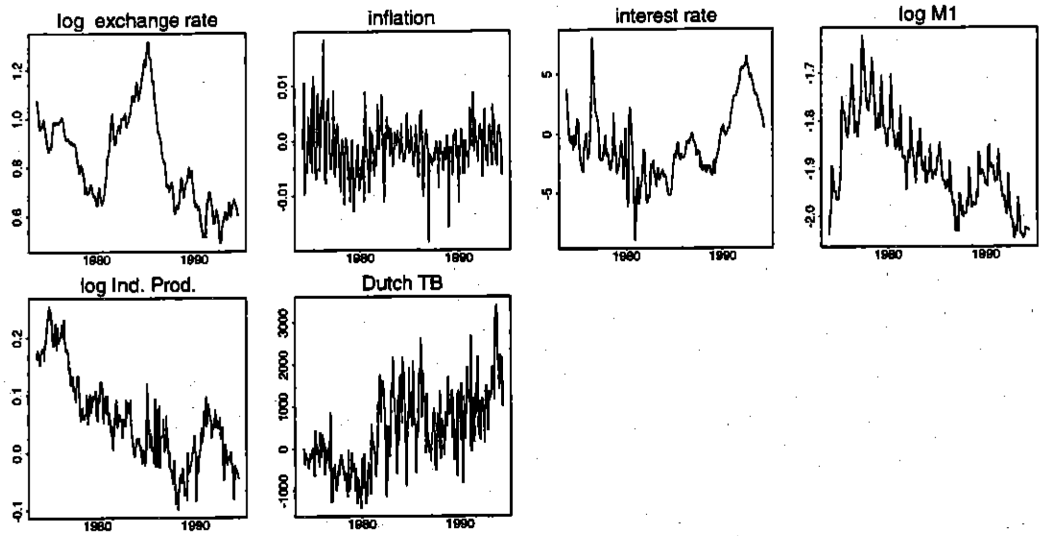

Figure 6: The data for the Netherlands-US situation; All data (except for the exchange rate and the trade balances) are in differential form (Netherlands-US).

Unit Root Test Results

Table 9 presents, for each country, the results of the ADF unit root tests[6] for the variables in the exchange rate models. Recall that under the null hypothesis of a unit root $\gamma_0 = 0$. The first column of Table 9 refers to the variable in differential form (except for the exchange rates and the trade balances); for example r denotes $r - r^*$. The second through fourth column respectively gives the "t-value" (negative sign omitted) of γ_0 in the following transformed regressions

$$
\begin{aligned}
(1)\ \Delta y_t &= \alpha_0 + \alpha_1 t + \gamma_0 y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \nu_t \\
(2)\ \Delta y_t &= \alpha_0 + \gamma_0 y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \nu_t \\
(3)\ \Delta y_t &= \gamma_0 y_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \nu_t
\end{aligned}
$$

The number of lags p is determined by the highest possible lag (with a maximum of 13) which is significant at a 10 % error level. The corresponding critical values at the 1%, 5%, and 10% error levels are calculated according to MacKinnon [Mac91], and are displayed at the bottom of each column. The last two columns give the t-values of α_0 and α_1 in either

$$
\Delta y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{p} \gamma_i \Delta y_{t-i} + \nu_t
$$

if the null hypothesis of $\gamma_0 = 0$ was not rejected, or in (1), (2), or (3), if the null is rejected. Substituting $\gamma_0 = 0$ in (1) removes the possible multicollinearity between the trend and y_{t-1}, which makes the estimation of α_0 and α_1 more accurate.

[6] The tests have been performed in PcGive 8.0

Table 9: Unit root tests

variable	$t_{\gamma_0}^{(1)}$	$t_{\gamma_0}^{(2)}$	$t_{\gamma_0}^{(3)}$	t_{α_0}	t_{α_1}
United Kingdom-US					
s	2.44	2.45	1.37	0.41	0.58
r	2.71	2.64	1.67	0.04	0.16
Δcpi	3.39	2.74	2.62**		
m	2.26	0.24	1.42	0.69	1.33
ip	3.95*	2.91*	3.27**	2.44*	2.61**
TB_{uk}	2.25	0.83	0.06	0.13	0.50
TB_{us}	1.95	0.85	0.66	0.52	0.26
Netherlands-US					
s	1.78	1.25	0.79	0.13	0.26
r	2.07	1.54	1.25	0.56	0.50
Δcpi	2.26	2.27	1.70	0.74	0.75
m	4.40**	1.60	0.34	4.17**	3.96**
ip	2.50	1.90	2.39**		
TB	1.83	0.58	0.03	0.80	0.33
Germany-US					
s	1.80	1.25	0.93	1.07	0.69
r	2.06	1.23	0.86	0.93	1.20
Δcpi	1.96	1.54	1.46	0.13	0.14
m	1.37	1.44	0.65	0.22	0.47
ip	2.88	2.04	2.28*		
TB	2.09	1.77	0.75	0.07	0.17
Japan-US					
s	2.41	0.50	1.62	0.40	0.45
r	2.16	2.08	1.05	0.73	0.73
Δcpi	2.34	2.44	1.58	1.26	1.10
m	2.73	0.35	1.11	0.50	1.13
ip	0.90	1.60	1.60	1.27	1.29
TB	2.95	1.46	0.43	0.31	0.03
critical values:					
1%	4.00	3.46	2.57	2.60	2.60
5%	3.43	2.87	1.94	1.97	1.97

Testing for nonlinear cointegration

To test for the existence of a nonlinear cointegrating relationship, we generalise the Dickey-Fuller tests, which have been presented in the previous section. The main issue is to construct critical values for cointegration tests on neural networks. In doing this, we adopt a similar procedure as employed in [EY87, Sep94]. The cointegrating relationship is assumed to be represented by

$$x_{1t} = f(x_{2t}, x_{3t}, \ldots, x_{kt}) + \epsilon_t \tag{24}$$

In (24) f is estimated by a neural network, but any other flexible regression method can be used instead.

Critical values of the ADF test on neural networks depend on several factors. When employing neural networks to construct a test for nonlinear cointegration, it is necessary to condition critical values on the following neural network factors: number of hidden units, value of weight decay parameter, number of inputs, number of observations, and total number of restarts employed in finding the final network. All factors are somehow related to the temptation of neural networks to overfit the training data, which would result in unjustly small sized residuals. When the influence of the neural network factors is neglected, the ADF test would too often reject the null hypothesis of no cointegration.

The present soft- and hardware makes it computationally infeasible to construct tables for critical values for all possible combinations of neural network factors. In the applications, we will select the "best" neural network factors for a particular case, and will calculate the critical values that correspond to that particular combination of neural network factors.

The critical values are constructed under the null hypothesis of no cointegration through thousand replications of the following procedure. Construct k independent random walks of length n, using the data generating mechanism

$$\mathbf{x}_t = \mathbf{x}_{t-1} + U_t, u_{it} = 0.8\, u_{it-1} + \nu_{it}, \quad i = 1, \ldots, k, \tag{25}$$

with $\nu_{it} \sim IN(0, 1)$. Train a neural network with a particular set of factors to approximate f in (24). Then estimate the parameters in

$$\Delta\hat{\epsilon}_t = \alpha_0 + \gamma_0\hat{\epsilon}_{t-1} + \sum_{i=1}^{p} \gamma_i \Delta\hat{\epsilon}_{t-i} + \zeta_t \quad \zeta_t \sim \text{i.i.d}(0, \sigma^2), \tag{26}$$

using the residuals from (24) and calculate the t-statistic of $\hat{\gamma}_0$ (usually minus signs are omitted). The thousand t-statistics, so obtained, give an empirical distribution, which is used in calculating the critical values required for testing the null hypothesis of no cointegration.

Econometric Analysis of Short-Run Models

The following procedure has been followed to arrive at a parsimonious short-run model for Δs_t. According to the general to specific approach we depart from the most general model (sticky price II) extended with price changes, and with three lags for each variable included. We then use F-tests to test for zero restrictions on a subset of variables. The cumulated trade balance terms are tested first on their relevance, then the price change differentials, followed by the elements of the flexible price model (r, m, and ip). Only the end results of this 'testing down' process have been reported in Table 10. Table 10 reports the variable name, its coefficient, the corresponding t-value, the R^2 of the estimated model, and the Durbin-Watson statistic DW. The Durbin-Watson statistic is defined as:

$$DW := \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

where e_t denote the observed residuals from the estimated model. A value of the DW-statistic close to 2 indicates that no autocorrelation is present in the residuals. Some diagnostic tests have been performed as well.

It should be noted that the dominant factor in all models is Δs_{t-1}. A recursive estimation exercise learned that most parameter estimates were stable, with some exception for the money supply variables. Stable parameter estimates are a prerequisite for reliable predictions.

We performed the neural network test to test for possible neglected nonlinearities in the models presented in Table 10. The adjusted p-value of the test was smaller than 0.000 for the Japan-, German-, and UK-cases; for the Dutch-case the adjusted p-value was 0.92. Additionally, RESET tests were performed which add $\widehat{\Delta s_t}^2$ to the models. The probabilities on the observed F-statistics were for Japan, Germany, U.K, and the Netherlands, 0.036, 0.136, 0.121, and 0.842, respectively. Both tests suggest that possible nonlinearities are present in the exchange rate models for Japan, Germany, and the U.K., and none in the model for the Netherlands. The evidence displayed by the neural network test is stronger than that of the RESET test.

Table 10: Model Estimates

variable	coefficient	t-value
Japan (R^2=0.19; DW=1.97)		
Δs_{t-1}	0.36	6.11
Δs_{t-3}	0.10	1.76
Δm_{t-2}	0.17	1.94
Δr_t	-0.0075	-3.96
United Kingdom (R^2=0.25; DW=2.00)		
Δs_{t-1}	0.47	7.70
Δs_{t-2}	-0.16	-2.53
Δm_{t-2}	0.20	2.43
ΔTB_t^*	3.07e-6	3.54
ΔTB_t	-7.91e-6	-2.58
ΔTB_{t-2}	-1.18e-5	-3.31
Germany (R^2=0.21; DW=1.96)		
Δs_{t-1}	0.34	5.99
Δm_{t-2}	0.29	2.76
Δr_t	-0.0086	-4.55
Δip_{t-1}	-0.19	-2.39
The Netherlands (R^2=0.20; DW=1.99)		
Δs_{t-1}	0.34	5.90
Δm_{t-1}	0.13	2.51
Δm_{t-3}	0.11	2.20
Δr_t	-0.0064	-4.27

References

[BDGH93] A. Banerjee, J. Dolado, J. Galbraith, and D. Hendry. *Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data.* Oxford University Press, Oxford, 1993.

[BH85] W. Branson and D. Henderson. The specification and influence of asset markets. In R.W. Jones and P.B. Kenen, editors, *Handbook of International Economics, vol. 2*, pages 749–805. North Holland, Amsterdam, 1985.

[Bil78] J.F. Bilson. Rational expectations and the exchange rate. In J. Frenkel and H. Johnson, editors, *The economics of exchange rates.* Addison-Wesley Press, Reading, 1978.

[BM89] R. Baillie and P. McMahon. *The foreign exchange market: theory and econometric evidence.* Cambridge University Press, Cambridge, 1989.

[BvdBW94] D. Baestaens, W. van den Bergh, and D. Wood. *Neural Network Solutions for Trading in Financial Markets.* Pitman, London, 1994.

[Cag56] P. Cagan. In M. Friedman, editor, *Studies in the Quantitative Theory of Money.* University of Chicago Press, Chicago, 1956.

[CT95] M. Camarero and C. Tamarit. A rationale for macroeconomic policy coordination: evidence based on the spanish peseta. *European Journal of Political Economy*, 11:65–82, 1995.

[Cyb89] Cybenko. Approximation by superposition of a sigmoidal function. *Math. Control systems Signals*, 2:303–314, 1989.

[DN90] F. Diebold and J. Nason. Nonparametric exchange rate prediction? *Journal of International Economics*, 28:315–332, 1990.

[Dor76] R. Dornbusch. Expectations and exchange rate dynamics. *Journal of Political Economy*, 84:1161–1176, 1976.

[EY87] R. Engle and S. Yoo. Forecasting and testing in co-integrated systems. *Journal of Econometrics*, 35:143–159, 1987.

[FF93] I.E. Frank and J.H. Friedman. A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135, 1993.

[Fle62] J.M. Fleming. Domestic financial policies under fixed and floating exchange rates. *International Monetary Fund Staff Papers*, 9(4):369–380, 1962.

[Fra79] J.A. Frankel. On the mark: A theory of floating exchange rates based on real interest differentials. *The American Economic Review*, 69:611–622, 1979.

[Fre76] J.A. Frenkel. A monetary approach to the exchange rate: doctrinal aspects and empirical evidence. *Scandinavian journal of economics*, 76:200–224, 1976.

[Fre81] J.A. Frenkel. The collapse of purchasing power parities during the 1970's. *European Economic Review*, 16(4):145–165, 1981.

[GBD92] S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58, 1992.

[Hen93] D. Hendry. *Econometrics: Alchemy or Science?* Blackwell, Oxford, 1993.

[HKP91] J. Hertz, A. Krogh, and R. Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, Redwood City, CA, 1991.

[HT90] T. Hastie and R. Tibshirani. *Generalized additive models*. Chapman and Hall, London, 1990.

[Hub85] P.J. Huber. Projection pursuit. *The Annals of Statistics*, 13:435–475, 1985.

[Mac91] MacKinnon. Critical values for cointegration tests. In R. Engle and C. Granger, editors, *Long-Run Economic Relationships*, pages 267–276. Oxford University Press, 1991.

[Mar95] N.C. Mark. Exchange rates and fundamentals: Evidence on long-horizon predictability. *The American Economic Review*, 85(1):201–218, 1995.

[MP91] M. Meulepas and J. Plasmans. Estimating and forecasting exchange rates by means of ppp and uip. Technical Report SESO 91/262, UFSIA University of Antwerp, 1991.

[MR83] R. Meese and K. Rogoff. Empirical exchange rate models of the seventies: Do they fit out-of-sample? *Journal of International Economics*, 14:3–24, 1983.

[MR90] R. Meese and A. Rose. Non-linear, non-parametric, non-essential exchange rate determination. *American Economic Review*, 192-196:601–619, 1990.

[MR91] R. Meese and A. Rose. An empirical assessment of non-linearities in models of exchange rate determination. *Review of Economic Studies*, 58:601–619, 1991.

[MT92] R. MacDonald and M. Taylor. Exchange rate economics: a survey. *IMF Staff Papers*, 39:1–57, 1992.

[Mun63] R.A. Mundell. Capital mobility and stabilization policy under fixed and flexible exchange rates. *Canadian Journal of Economic and Political Science*, 29(4):475–485, 1963.

[Nas90] J. Nash. *Compact Numerical Methods for Computers*. Adam Hilger, New York, 1990.

[RHW86] D. Rumelhart, G. Hinton, and R. Williams. Learning internal representations by error propagation. In D. Rumelhart and J. McClelland, editors, *Parallel distributed processing: explorations in the microstructures of cognition*, pages 318–362. MIT Press, Cambridge, MA, 1986.

[Rip93] B.D. Ripley. Statistical aspects of neural networks. In O.E. Barndorff-Nielsen, J.L. Jensen, and W.S. Kendall, editors, *Chaos and Networks-Statistical and Probabilistic Aspects*, pages 40–123. Chapman & Hall, London, 1993.

[RM86] D. Rumelhart and J. McClelland. *Parallel distributed processing: explorations in the microstructures of cognition*. MIT Press, Cambridge, MA, 1986.

[Sca85] L.E. Scales. *Introduction to non-linear optimization.* MacMillan Publishers, London, 1985.

[Sep94] P. Sephton. Cointegration tests on MARS. *Computational Economics*, 7(1):23–35, 1994.

[Sto74] M. Stone. Cross-validatory choice and asssessment of statistical predictions. *Journal of the Royal Statistical Society*, B36:111–147, 1974.

[TT93] R. Trippi and E. Turban. *Neural Networks in Finance and Investing: using artificial intelligence to improve real world performance*. Probus Publishing, Chicago, Illinois, 1993.

[WP95] S. Wei and D. Parsley. Purchasing power disparity during the floating rate period: exchange rate volatility, trade barriers and other culprits. Technical Report 5032, National Bureau of Economic Research, 1995.

LIJST VAN RECENTE SESO-RAPPORTEN

DE BORGER B., FERRIER G. en K. KERSTENS, *The choice of a technical efficiency measure on the free disposal hull reference technology : a comparison using US banking data,* January 1995, 34 p. **(95/315)**

KONINGS J. en H. VANDENBUSSCHE, *How does increased foreign competition affect UK employment and wages ? An empirical exploration,* February 1995, 19 p. **(95/316)**

HEYLEN F. en A. VAN POECK, *Central bank independence : only part of the inflation story,* February 1995, 20 p. **(95/317)**

VANHOUDT P., *Het belang van investeringen in mensen, materieel en know-how voor economische groei : empirische evidentie,* april 1995, 29 p. **(95/318)**

NONNEMAN W. en P. VANHOUDT, *A further augmentation of the Solow model and the empirics of economic growth for OECD countries,* April 1995, 12 p. **(95/319)**

MEERSMAN H. en D. CASSIMON, *Investment in the EC-countries : does reducing uncertainty matter ?,* April 1995, 35 p. **(95/320)**

VANDENBUSSCHE H., *Is European antidumping protection against Central Europe too high ?,* June 1995, 33 p. **(95/321)**

CORTENS I. en W. NONNEMAN, *Parental education and the demand for higher education : some evidence for Belgium,* June 1995, 22 p. **(95/322)**

SCHROYEN F., *Preference characterisation and indirect Allais coefficients,* June 1995, 40 blz. **(95/323)**

DOUVEN R. en J. PLASMANS, *S.L.I.M. - A small linear inter-dependent model of eight EU-member states, the USA and Japan,* June 1995, 35 blz. **(95/324)**

DOUVEN R. en J. PLASMANS, *Convergence and international policy coordination in the EU : a dynamic games approach,* June 1995, 39 blz. **(95/325)**

VAN REETH D., *How the Belgian treasure should auction its debt,* July 1995, 32 blz. **(95/326)**

DUCHESNE I. en W. NONNEMAN, A neo-classical demand model for higher education in Belgium, October 1995, 52 blz. **(95/327)**

DE BORGER B. en S. WOUTERS, Transport externalities and optimal pricing and supply decisions in urban transportation : a simulation analysis for Belgium, October 1995, 52 blz. **(95/328)**

VAN HOOYDONK C., A TSP-procedure to test for the order of integration of a time series by means of (A)DF, PP and KPSS tests, October 1995, 44 blz. **(95/329)**