Modeling Exchange Rates: Smooth Transitions, Neural Networks, and Linear Models

Marcelo C. Medeiros, Álvaro Veiga, and Carlos Eduardo Pedreira

***Abstract*—The goal of this paper is to test for and model nonlinearities in several monthly exchange rates time series. We apply two different nonlinear alternatives, namely: the artificial neural-network time series model estimated with Bayesian regularization and a flexible smooth transition specification, called the neuro-coefficient smooth transition autoregression. The linearity test rejects the null hypothesis of linearity in 10 out of 14 series. We compare, using different measures, the forecasting performance of the nonlinear specifications with the linear autoregression and the random walk models.**

***Index Terms*—Bayesian regularization, exchange rates, neural networks, smooth transition models, time series.**

I. Introduction

DURING the last two decades many different nonlinear models have been proposed in the literature to model and forecast exchange rates. Several authors claimed that exchange rates are rather unpredictable, and that a random walk model is often a better predictor than concurrent nonlinear models. See, for example, [3], [9], [21], and [22]. With this concern in mind, some questions should be raised: How relevant is the nonlinearity in the series? Is the nonlinearity uniformly spread? Are nonlinear models better predictors? If there are periods of the series with no nonlinearity, what is the lost (if any) of applying a nonlinear model? Without having the intention of solving these fundamental and complex questions, this paper addresses the problem by benchmarking two nonlinear alternatives against the linear autoregressive (AR) and the random walk (RW) models. Several monthly exchange rates time series are used. For similar papers, see [26] and [24].

The nonlinear alternatives considered in this paper are the artificial neural network (ANN) model [14] and a novel flexible model called the neuro-coefficient smooth transition autoregression (NCSTAR). The NCSTAR specification can be interpreted as a linear model where its coefficients are given by a single hidden layer feedforward neural network and has the main advantage of nesting several well-known nonlinear formulations, such as the self-exciting threshold autoregression (SETAR) [30], [31], the smooth transition autoregression (STAR) [2], [16], [28], and the ANN model. Furthermore, if the neural network is interpreted as a nonparametric universal approximation to any Borel-measurable function, the NCSTAR model is directly comparable to the functional coefficient autoregression (FAR) [4] and the single-index coefficient regression model [34]. A modeling strategy for the NCSTAR model was developed in [19] and [20].

The plan of this paper is as follows. Section II discusses the NCSTAR model and briefly describes the modeling strategy. Section III outlines the neural network models with Bayesian regularization. The benchmark models are described in Section IV. Section V gives a description of an experiment comparing the forecasting performance of the NCSTAR and the neural network models with the benchmark alternatives. The results are discussed in Section VI. Finally, Section VII concludes.

Manuscript received August 25, 2000; revised March 15, 2001.

M. C. Medeiros is with the Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil.

Á. Veiga and C. E. Pedreira are with the Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Brazil.

Publisher Item Identifier S 1045-9227(01)05015-9.

II. The NCSTAR Model

A. Mathematical Formulation

Consider a linear model with time-varying coefficients expressed as

$$y_t = \boldsymbol{\phi}_t' \mathbf{z}_t + \varepsilon_t \tag{1}$$

where $\boldsymbol{\phi}_t = [\phi_t^{(0)}, \phi_t^{(1)}, \ldots, \phi_t^{(p)}]'$ is a vector of real coefficients and $\mathbf{z}_t = [1, \tilde{\mathbf{z}}_t']'$. $\tilde{\mathbf{z}}_t \in \mathbb{R}^p$ is a vector of lagged values of y_t and/or some exogenous variables. The random term ε_t is a normally distributed white noise with variance σ^2. The time evolution of the coefficients $\phi_t^{(j)}$ of (1) is given by

$$\phi_t^{(j)} = \sum_{i=1}^{h} \lambda_{ji} F(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i) - \lambda_{j0}, \qquad j = 0, \ldots, p \tag{2}$$

where λ_{ji} and λ_{j0} are real coefficients.

The function $F(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i)$ is the logistic function, where $\mathbf{x}_t \in \mathbb{R}^q$ is a vector of input variables, $\boldsymbol{\omega}_i = [\omega_{1i}, \ldots, \omega_{qi}]'$ and β_i are real parameters. The norm of $\boldsymbol{\omega}_i$ is called the *slope parameter*. In the limit, when the slope parameter approaches infinity, the logistic function becomes a step function.

Note that (1) can be interpreted as a linear model where its coefficients are given by a single hidden layer feedforward neural network. The neural-network architecture representing (2) is illustrated in Fig. 1. The elements of $\mathbf{x}_t$, called the transition variables, can be formed by lagged values of y_t and/or any exogenous variables. In this paper, we assume that $\mathbf{x}_t$ is formed by a subset of the elements of $\mathbf{z}_t$ and that there are no exogenous variables in the model specification.

Equations (1) and (2) represent a time-varying model with a multivariate smooth transition structure defined by h hidden neurons.

Equation (1) can be rewritten as

$$y_t = G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}) + \varepsilon_t$$

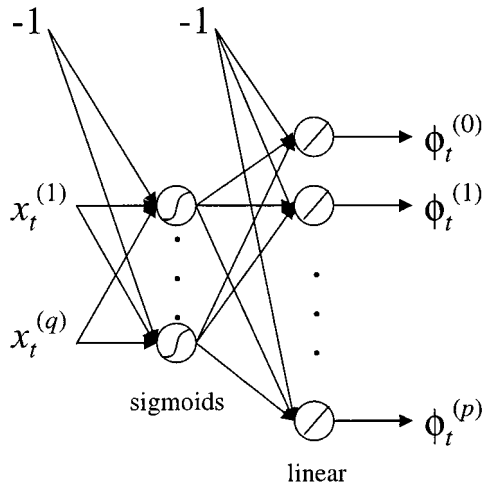

Fig. 1. Architecture of the neural network.

$$= \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{t-j} + \sum_{i=1}^{h} \lambda_{0i} F(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i) + \sum_{j=1}^{p} \left\{ \sum_{i=1}^{h} \lambda_{ji} F(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i) \right\} y_{t-j} + \varepsilon_t \quad (3)$$

or in vector notation

$$y_t = G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}) + \varepsilon_t = \boldsymbol{\alpha}' \mathbf{z}_t + \sum_{i=1}^{h} \boldsymbol{\lambda}_i' \mathbf{z}_t F(\boldsymbol{\omega}_i' \mathbf{x}_t - \beta_i) + \varepsilon_t \quad (4)$$

where

$$\boldsymbol{\Psi} = [\boldsymbol{\alpha}', \boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_h', \boldsymbol{\omega}_1', \ldots, \boldsymbol{\omega}_h', \beta_1, \ldots, \beta_h]'$$

is a parameter vector with $(q+1) \times h + (p+1) \times (h+1)$ elements

$$\boldsymbol{\alpha} = [\alpha_0, \ldots, \alpha_p]' = [-\lambda_{00}, \ldots, -\lambda_{p0}]', \quad \text{and}$$
$$\boldsymbol{\lambda}_i = [\lambda_{0i}, \ldots, \lambda_{pi}]'.$$

Note that model (4) is, in principle, neither globally nor locally identified. To ensure the identifiability of (4) we have to impose the following restrictions: $\beta_1 \leq \ldots \leq \beta_h$ and $\omega_{1i} > 0$, $i = 1, \ldots, h$. For details on identifiability concepts see, e.g., [27], [15], [12], [1], [32], [18].

The NCSTAR model has the main advantage of nesting several well-known nonlinear formulations, such as, for example, the SETAR, STAR, and ANN models.

B. Modeling Cycle

In this section, we briefly outline a modeling technique based on statistical inference to build the NCSTAR model. For more details, see [19] and [20]. This amounts to proceeding from a linear model to the smallest NCSTAR model and gradually toward larger ones through a sequence of Lagrange multiplier (LM) tests. Finally, after the model has been estimated, it is evaluated by some misspecification tests. For similar ideas, see [33], [28], and [7]. The modeling cycle can be summarized as follows:

1. Specification
   a) Select the variables of the model.
   b) Test linearity.
   c) If linearity is rejected, determine the number of hidden units.
2) Parameter estimation of the specified model.
3) Model evaluation based on misspecification testing.
   a) Test for parameter constancy.
   b) Test for serial independence of the error term.
   c) Test for homoscedasticity of the error term .

These three stages are briefly described below.

1) Specification:

a) Variable Selection: The first step of the specification stage is to select the variables of the model. In this step we will not distinguish between the variables in $\mathbf{z}_t$ and $\mathbf{x}_t$ in (4). Following [18], we adopt the simple procedure proposed by [25]. Their proposal uses global parametric least squares estimation and is based on a polynomial expansion of the model. We provide a brief overview of the method. For more details see [25].

Consider model (4). The first step is to expand function $G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi})$ into a k-order polynomial expansion around an arbitrary fixed point in the sample space. After merging terms, one obtains

$$G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}) = \boldsymbol{\pi}' \mathbf{z}_t + \sum_{j_1=1}^{p} \sum_{j_2=j_1}^{p} \rho_{j_1 j_2} z_{j_1,t} z_{j_2,t} + \ldots + \sum_{j_1=1}^{p} \cdots \sum_{j_k=j_{k-1}}^{p} \theta_{j_1 \ldots j_k} z_{j_1,t} \ldots z_{j_k,t} + R(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}) \quad (5)$$

where $R(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi})$ is the remainder and the θ's, ρ's, and $\boldsymbol{\pi} \in \mathbb{R}^{p+1}$ are parameters. Note that the terms involving $\mathbf{x}_t$ merged with the terms involving $\mathbf{z}_t$.

The second step is to regress y_t on all variables in the polynomial expansion and compute the value of a model selection criterion, such as, for example, the Akaike's information criterion (AIC). After that, remove one variable from the original model and regress y_t on all the remaining terms in the expansion and compute the value of the AIC. Repeat this procedure by omitting each variable in turn. Continue by simultaneously omitting two regressors of the original model and proceed in that way until the polynomial expansion becomes a function of a single regressor. Choose the combination of variables that yields the lowest value of the AIC. The selected variables will compose the vector $\mathbf{z}_t$.

b) Testing Linearity: In practical nonlinear time series modeling, testing linearity plays an important role. In the context of model (4), testing linearity has two objectives. The first one is to verify if a linear model is able to adequately describe the data generating process. The second one refers to the variable selection problem. The linearity test is used to determine the elements of $\mathbf{x}_t$. After selecting the elements of $\mathbf{z}_t$ with the procedure described above, we choose the elements of $\mathbf{x}_t$ by running the linearity test described below setting $\mathbf{x}_t$ equal to each possible subset of the elements of $\mathbf{z}_t$ and choosing the one that minimize the p-value of the test.

In order to test for linearity, (4) is rewritten as

$$y_t = \boldsymbol{\alpha}'\mathbf{z}_t + \sum_{i=1}^{h} \boldsymbol{\lambda}_i'\mathbf{z}_t F(\gamma_i(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - c_i)) + \varepsilon_t \tag{6}$$

where $\gamma_i = \|\boldsymbol{\omega}_i\|$, $\tilde{\boldsymbol{\omega}}_i = \boldsymbol{\omega}_i/\gamma_i$, and $c_i = \beta_i/\gamma_i$, $i = 1, \ldots, h$. The transition function $F(\gamma_i(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - c_i))$ is redefined as

$$F(\gamma_i(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - c_i)) = \frac{1}{1+\exp(-\gamma_i(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - c_i))} - \frac{1}{2}. \tag{7}$$

Subtracting one-half from the logistic function is useful just in deriving linearity tests where it simplifies notation but not affect the argument. The models estimated in this paper do not contain that term.

Consider (6) with (7) and the testing of the hypothesis that y_t is a linear process. Note that $F(0) = 0$, implying the null hypothesis of linearity

$$\mathrm{H}_0 : \gamma_i = 0, \qquad i = 1, \ldots, h. \tag{8}$$

Hypothesis (8) offers a convenient starting point for studying the linearity problem in the LM (score) testing framework.

Note that model (6) is not identified under the null. A consequence of this complication is that the standard asymptotic distribution theory for the likelihood ratio or other classical test statistics for testing (8) is not available. We solve the problem by replacing $F(\gamma_i(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - c_i))$ by a third-order Taylor expansion approximation about $\gamma_i = 0$, $i = 1, \ldots, h$.

After rearranging terms, we get

$$\begin{aligned} y_t = {} & \boldsymbol{\pi}'\mathbf{z}_t + \sum_{i=1}^{q}\sum_{j=i}^{q} \theta_{ij} x_{i,t}^* x_{j,t} + \sum_{i=1}^{p-q}\sum_{j=1}^{q} \beta_{ij} z_{i,t}^* x_{j,t} \\ & + \sum_{i=1}^{q}\sum_{j=i}^{q}\sum_{k=j}^{q} \theta_{ijk} x_{i,t}^* x_{j,t} x_{k,t} \\ & + \sum_{i=1}^{p-q}\sum_{j=1}^{q}\sum_{k=j}^{q} \beta_{ijk} z_{i,t}^* x_{j,t} x_{k,t} \\ & + \sum_{i=1}^{q}\sum_{j=i}^{q}\sum_{k=j}^{q}\sum_{l=k}^{q} \theta_{ijkl} x_{i,t}^* x_{j,t} x_{k,t} x_{l,t} \\ & + \sum_{i=1}^{p-q}\sum_{j=1}^{q}\sum_{k=j}^{q}\sum_{l=k}^{q} \beta_{ijkl} z_{i,t}^* x_{j,t} x_{k,t} x_{l,t} + \varepsilon_t^* \end{aligned} \tag{9}$$

where $\varepsilon_t^* = \varepsilon_t + R(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi})$. $R(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi})$ is the combined remainder of the third-order Taylor expansion of the logistic functions. The vector $\mathbf{z}_t^* \in \mathbb{R}^{p-q}$ is formed by the elements of $\mathbf{z}_t$ that are not in $\mathbf{x}_t$. The null hypothesis is defined as H_0 : $\theta_{ij} = 0$, $\beta_{ij} = 0$, $\theta_{ijk} = 0$, $\beta_{ijk} = 0$, $\theta_{ijkl} = 0$, and $\beta_{ijkl} = 0$.

From (9) it is seen that the test is just a test of a linear hypothesis in a linear model, so that standard asymptotic inference is available.

It is important to stress that the linearity test against a STAR model [28] and the neural-network linearity test [29] are special cases of the test discussed here.

c) Determining the Number of Hidden Units: In a practical situation one wants to be able to test for the number of hidden units of the neural network. The basic idea is to start using the linearity test described above and test the linear model against the nonlinear alternative with only one hidden unit. If the linearity is rejected, then fit a model with one hidden unit and test for the second one. Proceed in that way until the first acceptance of the null hypothesis. The individual tests are based on linearizing the nonlinear contribution of the additional hidden neuron. Consider the general case in which the model contains h hidden units, and we want to know whether an additional unit is required or not. Write the model as

$$\begin{aligned} y_t = {} & \boldsymbol{\alpha}'\mathbf{z}_t + \sum_{i=1}^{h} \boldsymbol{\lambda}_i'\mathbf{z}_t F(\gamma_i(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - c_i)) \\ & + F(\gamma_{h+1}(\tilde{\boldsymbol{\omega}}_{h+1}'\mathbf{x}_t - c_{h+1})) + \varepsilon_t. \end{aligned} \tag{10}$$

An appropriate null hypothesis is

$$\mathrm{H}_0 : \gamma_{h+1} = 0. \tag{11}$$

Note that (10) is only identified under the alternative. Using a third-order expansion and after rearranging terms, the resulting model is

$$\begin{aligned} y_t = {} & \boldsymbol{\pi}'\mathbf{z}_t + \sum_{i=1}^{h} \boldsymbol{\lambda}_i'\mathbf{z}_t F(\gamma_i(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - c_i)) \\ & + \sum_{i=1}^{q}\sum_{j=i}^{q} \theta_{ij} x_{i,t}^* x_{j,t} + \sum_{i=1}^{p-q}\sum_{j=1}^{q} \beta_{ij} z_{i,t}^* x_{j,t} \\ & + \sum_{i=1}^{q}\sum_{j=i}^{q}\sum_{k=j}^{q} \theta_{ijk} x_{i,t}^* x_{j,t} x_{k,t} \\ & + \sum_{i=1}^{p-q}\sum_{j=1}^{q}\sum_{k=j}^{q} \beta_{ijk} z_{i,t}^* x_{j,t} x_{k,t} \\ & + \sum_{i=1}^{q}\sum_{j=i}^{q}\sum_{k=j}^{q}\sum_{l=k}^{q} \theta_{ijkl} x_{i,t}^* x_{j,t} x_{k,t} x_{l,t} \\ & + \sum_{i=1}^{p-q}\sum_{j=1}^{q}\sum_{k=j}^{q}\sum_{l=k}^{q} \beta_{ijkl} z_{i,t}^* x_{j,t} x_{k,t} x_{l,t} + \varepsilon_t^*. \end{aligned} \tag{12}$$

The null hypothesis is defined as H_0 : $\theta_{ij} = 0$, $\beta_{ij} = 0$, $\rho_{ij} = 0$. Again, standard asymptotic inference is available.

2) Parameter Estimation: After specifying the model, the parameters should be estimated by nonlinear least squares (NLS). Hence the parameter vector $\boldsymbol{\Psi}$ of (4) is estimated as

$$\begin{aligned} \hat{\boldsymbol{\Psi}} &= \underset{\boldsymbol{\Psi}}{\operatorname{argmin}}\, Q_T(\boldsymbol{\Psi}) \\ &= \underset{\boldsymbol{\Psi}}{\operatorname{argmin}} \sum_{t=1}^{T} (y_t - G(\mathbf{z}_t, \mathbf{x}_t; \boldsymbol{\Psi}))^2. \end{aligned} \tag{13}$$

Under some regularity conditions the estimates are consistent and asymptotically normal [5].

The estimation procedure is carried together with the test for the number of hidden units. First we test for linearity against a model given by (4) with $h = 1$. If linearity is rejected we estimate the parameters of the nonlinear model and test for the second hidden unit. If the null hypothesis is rejected, we use the estimated values for the first hidden unit as starting values and use the procedure described in [19] to compute initial values for

the second hidden unit. We proceed in that way until the first acceptance of the null hypothesis.

3) Model Evaluation: After the NCSTAR model has been estimated it has to be evaluated. This means that the assumptions under which the model has been estimated have to be checked. These assumptions include the hypothesis of no serial correlation, parameter constancy, and homoscedasticity. Testing for normality is also a common practice in econometrics. In this paper we use the tests discussed in [20]. They are Lagrange multiplier (LM) type tests of parameter constancy against the alternative of smoothly changing ones, of serial independence of the error term, and homoscedasticity against the hypothesis that the variance smoothly changes between regimes. To test for normality we use the Jarque–Bera test [13].

III. Artificial Neural Networks and Bayesian Regularization

A feedforward artificial neural network (ANN) time series model can be defined as

$$y_t = G(\mathbf{z}_t;\boldsymbol{\Psi}) = \lambda_0 + \sum_{i=1}^{h} \lambda_i F(\boldsymbol{\omega}_i'\mathbf{z}_t - \beta_i) + \varepsilon_t \tag{14}$$

where $\boldsymbol{\lambda} = [\lambda_0, \ldots, \lambda_q]'$ and $\boldsymbol{\omega}_i = [\omega_{1i}, \ldots, \omega_{pi}]'$ are vectors of real parameters, $\mathbf{z}_t = [1, \tilde{\mathbf{z}}_t']'$, $\tilde{\mathbf{z}}_t \in \mathbb{R}^p$ is a vector of lagged values of y_t, $\{\varepsilon_t\}$ is assumed to be a sequence of independent, normally distributed random variables with zero mean and finite variance, and $F(\boldsymbol{\omega}_i'\mathbf{z}_t - \beta_i)$ is the logistic function. Note that model (14) is just a special case of the NCSTAR model.

In this paper we adopt the regularization approach to estimate the ANN models. The fundamental idea is to find a balance between the number of parameters and goodness of fit by penalizing large models. The objective function is modified in such a way that the estimation algorithm effectively prunes the network by driving irrelevant parameter estimates to zero during the estimation process. The parameter vector $\boldsymbol{\Psi}$ is estimated as

$$\begin{aligned}\hat{\boldsymbol{\Psi}} &= \underset{\boldsymbol{\Psi}}{\operatorname{argmin}}\, \tilde{Q}_T(\boldsymbol{\Psi}) \\ &= \underset{\boldsymbol{\Psi}}{\operatorname{argmin}} (\eta Q_T(\boldsymbol{\Psi}) + (1-\eta) Q_T^*(\boldsymbol{\Psi}))\end{aligned} \tag{15}$$

where $Q_T(\boldsymbol{\Psi}) = \sum_{t=1}^{T} (y_t - G(\mathbf{z}_t;\boldsymbol{\Psi}))^2$, $Q_T^*(\boldsymbol{\Psi})$ is the *regularization* or *penalty term*, and $\eta > 0$ is often called the *decay constant*. The usual penalty is the sum of squared parameters

$$Q_T^*(\boldsymbol{\Psi}) = \sum_{j=0}^{p} \lambda_j^2 + \sum_{i=1}^{h} \beta_i^2 + \sum_{j=1}^{p}\sum_{i=1}^{h} \omega_{ji}^2. \tag{16}$$

The forecasting ability of the ANN model can depend crucially on the decay constant η, especially with small in-sample periods. If η is too large, the network may still overfit, and if it is too small, the ANN model does not have an adequate fit in the estimation period. Usually, different types of parameters in the ANN model will usually require different decay constants for good forecasting ability.

One approach to determine the optimal regularization parameter η is the Bayesian framework of [17], where the parameters of the network are assumed to be random variables with well-specified distributions. The regularization parameters are related to the unknown variances associated with these distributions and can be estimated with statistical techniques. Reference [8] give a detailed discussion of the use of Bayesian regularization in combination with the Levenberg–Marquardt optimization algorithm. The main advantage of this method is that even if the ANN model is over-parametrized, the irrelevant parameter estimates are likely to be close to zero and the model behaves like a small network.

All the ANN models in this paper are estimated with Bayesian regularization in combination with the Levenberg-Marquardt algorithm. The starting-values for the parameters are selected by the Nguyen–Widrow rule [23].

IV. Benchmark Models

In this section we outline two simple linear models that are often used as benchmark formulations in the financial time series literature.

A. The Random Walk (RW) Model

Consider the following RW model for the level of the exchange rate series

$$p_t = \alpha + p_{t-1} + u_t \tag{17}$$

where p_t is the price at the time instant t, α is a constant, and u_t is a random term identically distributed.

Taking the first difference of the logarithms, the resulting model becomes

$$y_t = \ln(p_t) - \ln(p_{t-1}) = c + \varepsilon_t \tag{18}$$

where y_t is the return at time t and c is a constant. Usually ε_t is assumed to be a normally distributed random variable with zero mean and finite variance.

B. The Linear Autoregressive Model

A linear autoregressive (AR) model of order p for the returns is defined as

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_p y_{t-p} + \varepsilon_t \tag{19}$$

where $\alpha_0, \alpha_1, \ldots, \alpha_p$ are real coefficients and ε_t is a identically normally distributed random variable with zero mean and finite variance. The order of the model is determined by inspection of the autocorrelation and partial autocorrelation functions (ACF and PACF).

V. The Experiment

To assess the practical usefulness of the NCSTAR and ANN models in comparison with the linear AR and RW models and to address the questions proposed in the Introduction, an experiment with 14 different monthly exchange rates time series is conducted. We have decided to work with monthly time series just to avoid to model any ARCH effect in the conditional variance of the series. It is well known that daily exchange rates are more volatile than the monthly ones and, of course, will have more nonlinearity to model. The data

TABLE I
DATA SETS

Series	Description	Period	T	t_0
1 - Austria	Austrian Schillings to one US Dollar	Jan/1971–Jul/2000	354	298
2 - Belgium	Belgian Francs to one US Dollar	Jan/1971–Jul/2000	354	298
3 - Denmark	Danish Krones to one US Dollar	Jan/1971–Jul/2000	354	298
4 - Finland	Finnish Markkas to one US Dollar	Jan/1971–Jul/2000	354	298
5 - France	French Francs to one US Dollar	Jan/1971–Jul/2000	354	298
6 - Germany	German Marks to one US Dollar	Jan/1971–Jul/2000	354	298
7 - The Netherlands	Dutch Guilders to one US Dollar	Jan/1971–Jul/2000	354	298
8 - Norway	Norwegian Krones to one US Dollar	Jan/1971–Jul/2000	354	298
9 - Sweden	Swedish Krones to one US Dollar	Jan/1971–Jul/2000	354	298
10 - Spain	Spanish Pesetas to one US Dollar	Jan/1973–Jul/2000	330	274
11 - India	Indian Rupees to one US Dollar	Jan/1973–Jul/2000	330	274
12 - Sri Lanka	Sri Lanka Rupees to one US Dollar	Jan/1973–Jul/2000	330	274
13 - Australia	US Dollars to one Australian Dollar	Jan/1971–Jul/2000	354	298
14 - United Kingdom	US Dollars to one British Pound	Jan/1971–Jul/2000	354	298

are summarized in Table I. The series were obtained from *Economagic* (www.economagic.com).

Both in-sample and out-of-sample performance are considered. The first step is to test linearity in all series. We discard all the series that do not have evidence of nonlinearity according to the test described in Section II-B. For the series that turn out to be nonlinear we proceed estimating linear and nonlinear models, each of which are evaluated according to their in-sample explanatory power and out-of-sample forecasting ability. The forecasts made by each estimated model are compared according to the following statistics (described in Appendix A: nRMSE, MAE, MAD, and SIGN.

The forecasting experiment can be viewed of consisting of the following steps.

1) Split the sample into two subsamples: the estimation set ($t = 1, \ldots, t_0$) and the forecasting set ($t = t_0+1, \ldots, T$).
2) Estimate the parameters of each model using only the estimation set and analyze the in-sample performance of the estimated models.
3) For $t = t_0, \ldots, T-4$, compute the out-of-sample forecasts of 1- to 4-step-ahead, $\hat{y}_{t+k|t}$, and the associated forecast errors denoted by $\hat{\varepsilon}_{t+k|t}$, where k is the forecasting horizon. Multistep forecasts for the nonlinear models are obtained by Monte Carlo simulation as described in Appendix B
4) For each forecasting horizon, compute different performance measures.

VI. RESULTS

A. Specification and Estimation

Using the variables selected by the AIC and the linearity test described in Section II-B, evidence of nonlinearity was found in ten series: Austria, Belgium, Finland, France, Germany, Sweden, Spain, India, Sri Lanka, and Australia. However, linearity was strongly rejected only in Finland, Australia, Sri Lanka, and India. The results are summarized in Table II.

To check if the nonlinearity is uniformly spread over the in-sample period, we fix the specification of $\mathbf{z}_t$ and $\mathbf{x}_t$ and test linearity in a rolling window with 100 observations. The p-value of the linearity test for each sub-sample is shown in Fig. 2. With the exception of the Sri Lanka, nonlinearity is only significant in a few number of periods, specially in the beginning or in the end of the series. This is an interesting result and explains why linearity is not strongly rejected for most of the series considered here.

TABLE II
LINEARITY TEST

Series	Lags in $\mathbf{z}_t$	Lags in $\mathbf{x}_t$	p-value
1 - Austria	1, 4	1	0.0103
2 - Belgium	1, 2, 3	1, 3	0.0068
3 - Denmark	1	1	0.4470
4 - Finland	1, 2, 3, 4	2, 4	2.0191×10^{-5}
5 - France	1, 3, 6	3, 6	0.0019
6 - Germany	1, 4	1	0.0079
7 - The Netherlands	1	1	0.5078
8 - Norway	1	1	0.1300
9 - Sweden	1, 6	1	0.0059
10 - Spain	1, 3	1	0.0085
11 - India	1, 3, 5, 8	1, 3, 5, 8	2.5690×10^{-10}
12 - Sri Lanka	1, 2, 6, 8	8	0
13 - Australia	1, 4	1, 4	0.0001
14 - United Kingdom	1, 2	1	0.1961

For those series that turned out to be nonlinear, we continue estimating the models.

The specification and estimation results for the NCSTAR models can be found in Table III. This table shows, for each series, the estimated number of hidden units (h), the p-value of the LM test of serial correlation of order r in the residuals ($LM_F^{iid}(r)$), the p-value of the LM test of parameter constancy (LM_F^{pc}), the p-value of the LM test of homocedasticity (LM_F^{σ}), the p-value of the Jarque-Bera (JB) test of normality of the residuals, and, finally, the estimated residual standard deviation.

Analyzing the results in Table III, we observe that all the estimated models have uncorrelated errors at 0.01 level. Only Sweden and Australia have evidence of serial correlated errors (of order 1) at 0.05 level. The hypothesis of parameter constancy is rejected at 0.05 level but not at 0.01 level for France

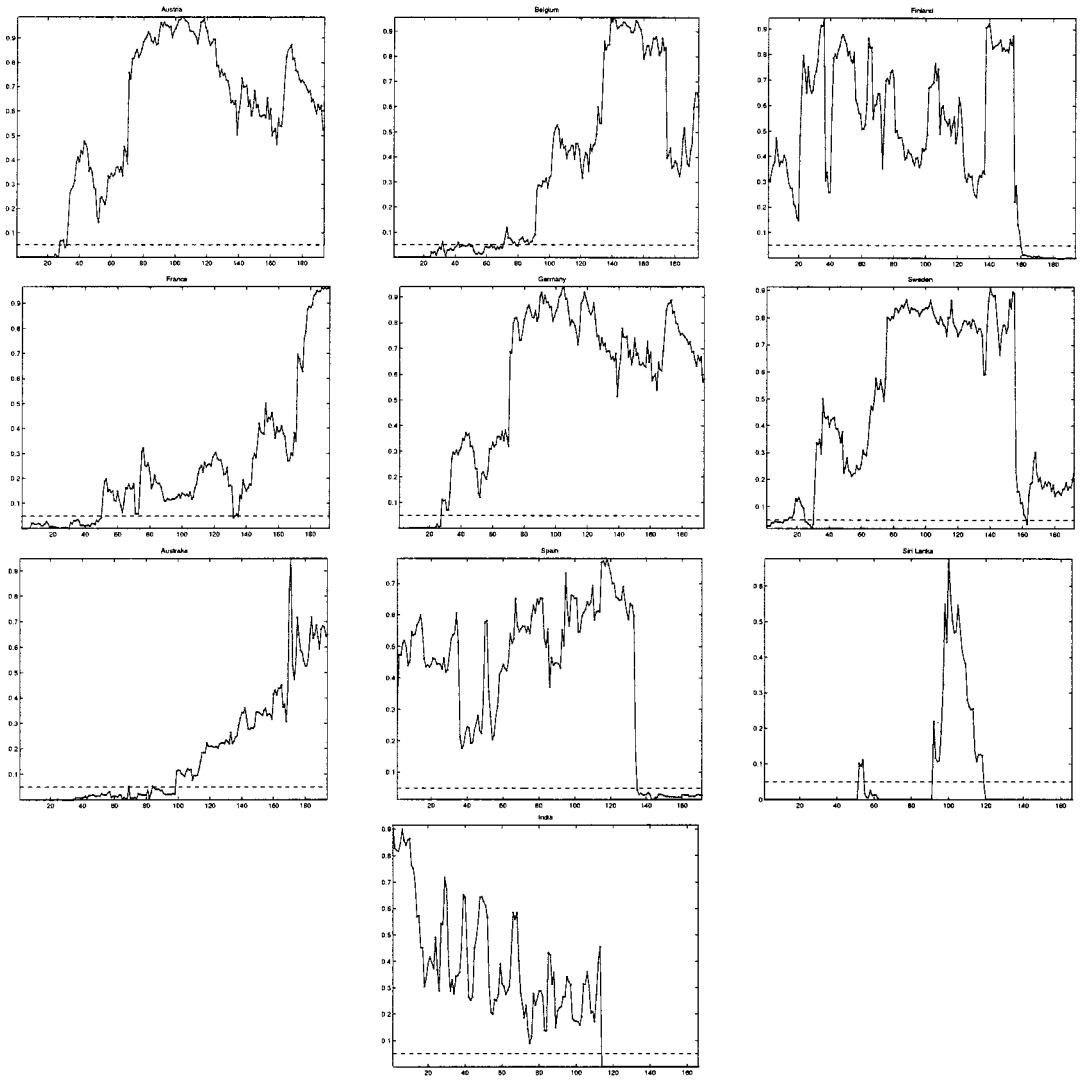

Fig. 2. p-value of the linearity test.

TABLE III
SPECIFICATION AND ESTIMATION RESULTS—NCSTAR MODEL

Series	h	$LM_F^{iid}(1)$	$LM_F^{iid}(2)$	$LM_F^{iid}(6)$	LM_F^{pc}	LM_F^{σ}	JB	$\hat{\sigma}$
Austria	1	0.7720	0.2933	0.6973	0.6270	0.8084	0.2314	0.0259
Belgium	1	0.4365	0.7502	0.2342	0.4537	0.1963	0.2208	0.0248
Finland	1	0.4330	0.2944	0.7634	0.4206	0.6865	0.0000	0.0216
France	2	0.9072	0.4074	0.1935	0.0262	0.6150	0.0078	0.0239
Germany	1	0.4232	0.3812	0.5026	0.6417	0.6425	0.1050	0.0258
Sweden	1	0.0293	0.0715	0.0240	0.5776	0.3946	0.0000	0.0216
Australia	1	0.0439	0.1281	0.3585	0.0144	0.0631	0.0000	0.0204
Spain	1	0.1706	0.2590	0.5155	0.3407	0.9207	0.0000	0.0242
Sri Lanka	2	0.6268	0.0758	0.0547	0.5389	0.8013	0.0000	0.0151
India	2	0.7260	0.4370	0.3675	0.1755	0.4691	0.0000	0.0130

and Australia. The only case where the null hypothesis of homoscedasticity is rejected (at 0.05 level) is France. Due the fact that the null hypothesis of serial independence, parameter constancy, and homoscedasticity are not strongly rejected (see the p-values of the tests) we do not take this into account and we accept the estimated models as our final specifications. Fig. 3 shows, for each model, the scatter plot of the transition function versus the linear combination of transition variables. With few exceptions, the transitions between regimes are rather smooth.

The specification and estimation results for the AR models are shown in Table IV. The columns show, respectively, the selected lags, the p-value of the Ljung-Box test of order 1, 2, and 6, the p-value of the Jarque-Bera (JB) test of normality of the residuals, and the estimated residual standard deviation.

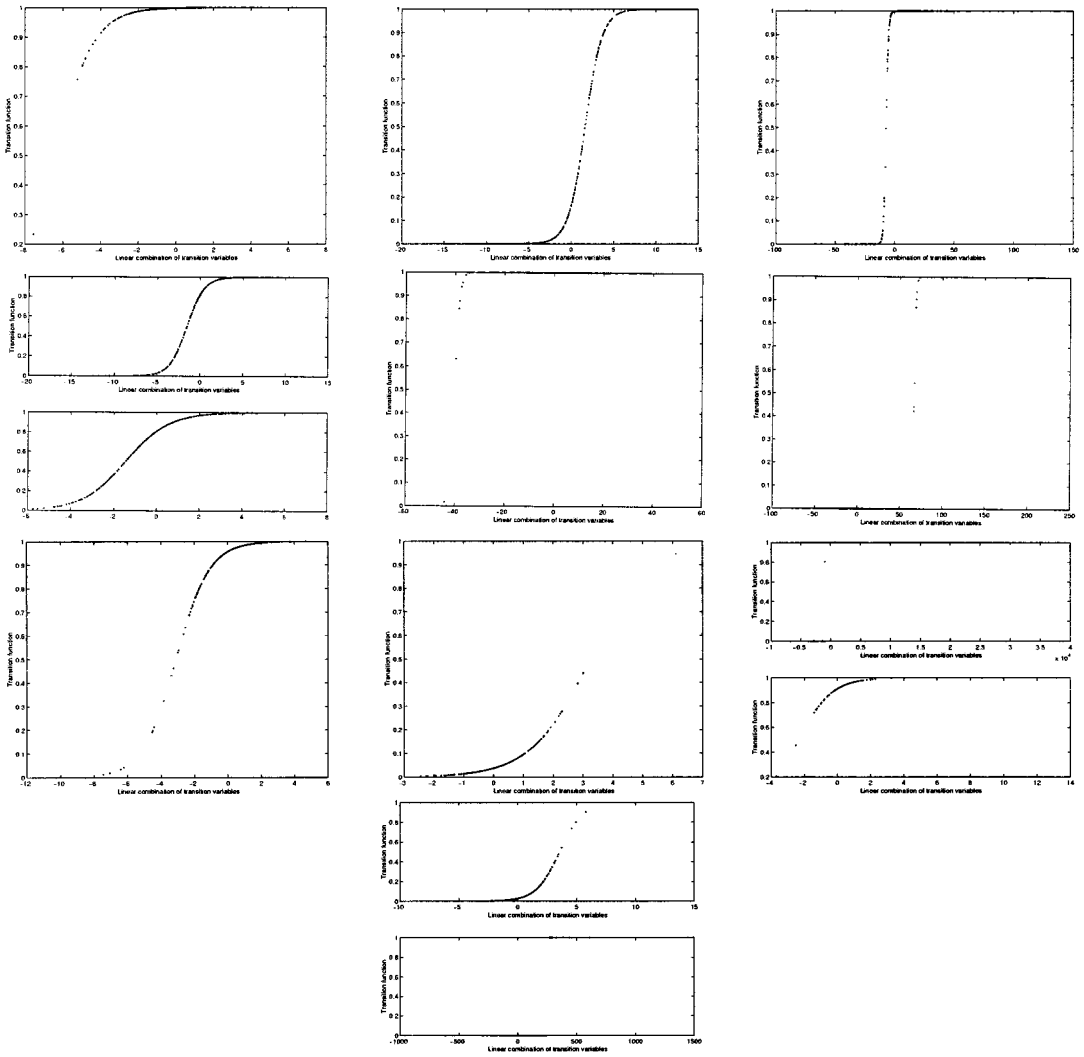

Fig. 3. Scatter plot of transition function versus the linear combination of transition variables.

TABLE IV
SPECIFICATION AND ESTIMATION RESULTS—AR MODEL

Series	Lags	$LB(1)$	$LB(2)$	$LB(6)$	JB	$\hat{\sigma}$
Austria	1	0.8040	0.5350	0.6140	0.0522	0.0264
Belgium	1	0.8170	0.3461	0.2820	0.0921	0.0257
Finland	1,2	0.8020	0.8832	0.8210	0.0000	0.0224
France	1,3	0.6774	0.3760	0.6700	0.0004	0.0254
Germany	1	0.7630	0.4420	0.6110	0.0630	0.0266
Sweden	1,2	0.8862	0.7781	0.4851	0.0000	0.0223
Australia	1,4	0.8100	0.8781	0.9823	0.0000	0.0213
Spain	1	0.8540	0.3660	0.3290	0.0000	0.0250
Sri Lanka	1,8	0.8720	0.8810	0.9360	0.0000	0.0179
India	1,8	0.4390	0.7100	0.5550	0.0000	0.0281

Observing Table IV, we note that all the linear models have uncorrelated errors.

In Table V we show the lags in the ANN models, the number of hidden units, the minimum value of the absolute correlation between the outputs of the hidden units of the estimated ANN models, and the residual standard deviation.

By inspection of Table V, we observe that, with exception of Australia and Sri Lanka, the hidden units of the ANN models are heavily correlated pointing to the fact that a model with only one hidden unit will be enough to model the data. Although not shown here, the plots of the outputs of the hidden units of most of the estimated models indicate that the hidden neurons are almost linear. This can be also checked by comparing the standard deviation of the residuals from the ANN and AR models.

Additionally, we should stress that the standard deviation of residuals from the NCSTAR model is smaller than the ones from the linear AR and the ANN for all the series.

B. Forecasting Experiment

The forecasting results are shown in Tables VI–VIII. Table I shows the number of series where each model is the best model according to the performance measures used here. Note that more than one model can be the "winner" for each series, specially according to the *SIGN* criterion.

For one-step-ahead forecasts the linear AR model has the best performance in 50% of the cases when the nRMSE, the MAE, and the MAD are used as performance measures. According to SIGN, the results are mixed, with the RW model having a small advantage. For two-, three-, and four-step-ahead forecasts the results are not very clear and there is no evidence of a "winner"

TABLE V
ESTIMATION RESULTS—ANN MODEL

Series	Lags	Hidden Units	$min(\lvert\hat{\Sigma}\rvert)$	$\hat{\sigma}$
Austria	6	5	0.9680	0.0264
Belgium	6	5	0.9098	0.0256
Finland	6	5	0.9603	0.0224
France	6	5	0.9997	0.0254
Germany	6	5	0.9603	0.0266
Sweden	6	5	0.9835	0.0223
Australia	6	5	0.2395	0.0213
Spain	6	5	0.8734	0.0248
Sri Lanka	10	12	0.0156	0.0051
India	10	12	1.0000	0.0177

TABLE VI
NUMBER OF SERIES WHERE EACH MODEL IS THE BEST MODEL

	1-step-ahead				2-step-ahead			
	NCSTAR	ANN	AR	RW	NCSTAR	ANN	AR	RW
nRMSE	1	2	5	2	2	1	3	4
MAE	2	2	5	1	2	0	5	3
MAD	2	1	5	2	3	5	1	1
SIGN	3	2	3	4	3	3	4	6
	3-step-ahead				4-step-ahead			
	NCSTAR	ANN	AR	RW	NCSTAR	ANN	AR	RW
nRMSE	2	2	4	2	4	2	3	1
MAE	3	1	2	4	2	1	3	4
MAD	6	2	0	2	4	3	1	2
SIGN	3	4	2	5	1	4	3	6

TABLE VII
NUMBER OF SERIES WHERE MODEL A (COLUMN) IS BETTER THAN MODEL B (ROW) ACCORDING TO THE MODIFIED DIEBOLD-MARIANO TEST AT A 0.05 LEVEL (nRMSE TEST)

	1-step-ahead				2-step-ahead			
	NCSTAR	ANN	AR	RW	NCSTAR	ANN	AR	RW
NCSTAR	–	0	0	0	–	0	3	0
ANN	1	–	1	1	1	–	1	1
AR	1	0	–	0	0	0	–	2
RW	0	2	3	–	0	0	0	–
	3-step-ahead				4-step-ahead			
	NCSTAR	ANN	AR	RW	NCSTAR	ANN	AR	RW
NCSTAR	–	0	1	0	–	0	1	1
ANN	1	–	1	2	1	–	1	0
AR	0	0	–	0	0	1	–	0
RW	0	0	0	–	0	0	0	–

TABLE VIII
NUMBER OF SERIES WHERE MODEL A (COLUMN) IS BETTER THAN MODEL B (ROW) ACCORDING TO THE MODIFIED DIEBOLD-MARIANO TEST AT A 0.05 LEVEL (MAE TEST)

	1-step-ahead				2-step-ahead			
	NCSTAR	ANN	AR	RW	NCSTAR	ANN	AR	RW
NCSTAR	–	0	0	0	–	1	2	1
ANN	1	–	1	1	1	–	1	2
AR	1	0	–	0	0	0	–	1
RW	2	5	4	–	0	0	0	–
	3-step-ahead				4-step-ahead			
	NCSTAR	ANN	AR	RW	NCSTAR	ANN	AR	RW
NCSTAR	–	0	1	1	–	0	0	0
ANN	1	–	1	1	1	–	1	1
AR	0	0	–	1	0	0	–	0
RW	0	0	0	–	0	0	0	–

model when the nRMSE and the MAE are considered. However, the RW model seems to be the best predictor of the SIGN. It is also important to notice that according to the MAD, that is a measure robust to outliers, the nonlinear models outperform the concurrent linear specifications in most of the cases.

To check if the forecasts produced by different models are statistically different or not, the number of series where model A (column) is better than model B (row) according to the modified Diebold-Mariano test [6], [11] at a 0.05 level are shown in Tables VII and VIII (see Appendix C for details). The results in Table VII concern to the nRMSE test and the ones in Table VIII concern to the MAE test.

Observing Tables VII and VIII, we conclude that, in most of the series, the differences in the forecast performance between NCSTAR, ANN, AR, and RW models are not significant according to the Diebold-Mariano test. It is important to notice that the NCSTAR model is better than the AR and ANN specifications when the Sri Lanka series is considered. This is not surprising, because that is the only series where the nonlinearity is uniformly spread. It is also important to observe that for 1-step-ahead forecasts and specially when the MAE is used as a comparison criterion, the results are quite supportive in favor of the linear and nonlinear specifications against the naïve RW model.

VII. CONCLUSION

This paper has presented and compared different alternatives to model and forecast monthly exchange rates time series. The models that have been used are the neuro-coefficient smooth transition autoregressive (NCSTAR) model, artificial neural networks (ANN), linear autoregression (AR), and, the random walk (RW) formulation.

In conclusion, we can now answer the questions raised earlier. How relevant is nonlinearity in the series? Nonlinearity is only relevant in some periods of the series, specially in the beginning or in the end of the sample. Is the nonlinearity uniformly spread? No, with the exception of the Sri Lanka, the nonlinearity is concentrated in only a small subsample of the data. Are nonlinear models better predictors? Nonlinear models stand a better chance only in the cases where nonlinearity is uniformly spread. Otherwise, there is no significant differences in the forecasts made by a concurrent linear model. What is the lost (if any) of applying a nonlinear alternative when there is no evidence of nonlinearity? If a statistical procedure to build nonlinear models is used, probably the final estimated model will be close to a linear specification and the forecasting ability will be close to the one from a linear specification. In this paper we have used a statistical approach to build the NCSTAR model and the ANN formulations have been estimated with Bayesian regularization that tries to build a parsimonious model based on Bayesian fundamentals. Concerning the predictability of exchange rates, we conclude that for one-step-ahead forecasts and when the MAE is used as a performance metric, there are some supportive results in favor of linear and nonlinear models against the simple random walk.

APPENDIX A
EVALUATING FORECASTS

The performance measures used in this paper are the following.

1. Normalized root mean squared error (nRMSE):

$$nRMSE(k) = \sqrt{\frac{1}{T-t_0-3}\frac{\sum_{t=t_0}^{T-4}\hat{\varepsilon}^2_{t+k|t}}{\hat{\sigma}^2_y}} \tag{20}$$

where $\hat{\sigma}^2_y$ is the estimated in-sample unconditional variance of the series.

2. Mean absolute error (MAE):

$$MAE(k) = \frac{1}{T-t_0-3}\sum_{t=t_0}^{T-4}|\hat{\varepsilon}_{t+k|t}| \tag{21}$$

3. Median absolute deviation (MAD):

$$MAD(k) = \text{median}(|\hat{\varepsilon}_{t+k|t} - \text{median}(\hat{\varepsilon}_{t+k|t})|). \tag{22}$$

The MAD is as a measure that is robust to outliers.

4. The proportion of times the sign of excess returns is correctly forecasted (SIGN):

$$SIGN(k) = \frac{1}{T-t_0-3}\sum_{t=t_0}^{T-4}\delta_t \tag{23}$$

where

$$\delta_t = \begin{cases} 1, & \text{if } y_{t+k}\hat{y}_{t+k|t} \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

APPENDIX B
FORECASTING WITH NONLINEAR MODELS

Multistep forecasting with nonlinear models is more challenging than forecasting with linear models. See, for example, [10, Sec. 8.1] for a general discussion.

Consider the simple nonlinear model defined as

$$y_t = G(y_{t-1};\boldsymbol{\Psi}) + \varepsilon_t \tag{24}$$

where $G(\cdot)$ is a nonlinear function with parameter vector $\boldsymbol{\Psi}$. The term ε_t is an independent identically distributed random variable with zero mean and finite variance. The history of the process up to time t is called $\Im_t$.

Due the fact that $E(\varepsilon_{t+1}|\Im_t) = 0$, the optimal one-step-ahead forecast of y_{t+1} is given by

$$\hat{y}_{t+1|t} = E(y_{t+1}|\Im_t) = G(y_t;\boldsymbol{\Psi}) \tag{25}$$

which is equivalent to the optimal one-step-ahead forecast when $G(\cdot)$ is linear.

For multistep forecasts, the problem is much more complicated. For two-step-ahead the optimal forecast is given by

$$\begin{aligned}\hat{y}_{t+2|t} &= E(y_{t+2}|\Im_t) = E(G(y_{t+1};\boldsymbol{\Psi})|\Im_t)) \\ &= \int_{-\infty}^{\infty} G(y_{t+1};\boldsymbol{\Psi}) f(\varepsilon_{t+1})\, d\varepsilon_{t+1}\end{aligned} \tag{26}$$

where $f(\varepsilon_{t+1})$ is the density of ε_{t+1}. Usually the expression (26) is approximated by numerical techniques, such as, for example, Monte Carlo or bootstrap.

The Monte Carlo method is a simple simulation technique for obtaining multistep forecasts. For model (24), the k-step-ahead forecast is defined as

$$\hat{y}_{t+k|t} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}^{(i)}_{t+k|t} \tag{27}$$

where N is the number of replications and

$$\hat{y}^{(i)}_{t+k|t} = G(\hat{y}_{t+k-1|t};\boldsymbol{\Psi}) + \xi^{(i)}_{t+k|t}. \tag{28}$$

$\xi^{(i)}_{t+k|t}$ is a random number drawn from a normal distribution with the same mean and standard deviation as the in-sample estimated residuals.

In this paper we adopt the Monte Carlo method with 2000 replications to compute the multistep forecasts.

APPENDIX C
THE DIEBOLD–MARIANO TEST

In order to test if the forecasts produced by two concurrent methods are statistically different or not, we use the Diebold-Mariano statistic [6] with the correction proposed by [11]. Suppose that a pair of k-steps-ahead, forecasts have produced the errors $(\hat{\varepsilon}^{(1)}_{t+k|t}, \hat{\varepsilon}^{(2)}_{t+k|t}), t = t_0, \ldots, T-k$. The quality of the forecasts is measured based on a specified loss function $g(\hat{\varepsilon}_{t+k|t})$ of the forecast error. Defining

$$d_t = g(\hat{\varepsilon}^{(1)}_{t+k|t}) - g(\hat{\varepsilon}^{(2)}_{t+k|t}) \tag{29}$$

and

$$\overline{d} = \frac{1}{T-t_0-3}\sum_{t=t_0}^{T-4} d_t \tag{30}$$

the Diebold–Mariano statistic is

$$S = [\hat{V}(\overline{d})]^{-1/2}\overline{d} \tag{31}$$

where

$$\hat{V}(\overline{d}) = \frac{1}{T-t_0-3}\left[\hat{\gamma}_0 + 2\sum_{i=1}^{k-1}\hat{\gamma}_i\right] \tag{32}$$

and

$$\hat{\gamma}_i = \frac{1}{T-t_0-3}\sum_{t=t_0+i}^{T-4}(d_t - \overline{d})(d_{t-i} - \overline{d}). \tag{33}$$

Under the null hypothesis, S is asymptotic normally distributed with zero mean and unit variance. However, the test is oversized even in moderate samples. To circumvent this problem, [11] proposed the following statistic:

$$S^* = \left[\frac{n+1-2h+n^{-1}k(k-1)}{n}\right]^{1/2} S \tag{34}$$

where $n = T - t_0 - k + 1$.

Under the null, S^* is assumed to have a Student's t distribution with $(n-1)$ degrees of freedom.

In this paper we adopt the following loss functions:

$$g(\hat{\varepsilon}_{t+k|t}) = \hat{\varepsilon}^2_{t+k|t}, \quad \text{for the } nRMSE \text{ test; and}$$
$$g(\hat{\varepsilon}_{t+k|t}) = |\hat{\varepsilon}_{t+k|t}|, \quad \text{for the } MAE \text{ test.}$$

ACKNOWLEDGMENT

The authors would like to thank T. Teräsvirta, R. Flores, and D. van Dijk for valuable comments, and the CNPq for the financial support. This work is part of the first author's Ph.D. dissertation at the Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Brazil.

REFERENCES

[1] U. Anders and O. Korn, "Model selection in neural networks," *Neural Networks*, vol. 12, pp. 309–323, 1999.
[2] K. S. Chan and H. Tong, "On estimating thresholds in autoregressive models," *J. Time Series Anal.*, vol. 7, pp. 179–190, 1986.
[3] P. H. K. Chang and C. L. Osler, "Methodical madness: Technical analysis and the irrationality of exchange-rate forecasts," *Economic Journal*, vol. 10, pp. 636–661, 1999.
[4] R. Chen and R. S. Tsay, "Functional coefficient autoregressive models," *J. Amer. Statist. Assoc.*, vol. 88, pp. 298–308, 1993.
[5] R. Davidson and J. G. MacKinnon, *Estimation and Inference in Econometrics*. New York: Oxford Univ. Press, 1993.
[6] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Business .Economic Statist.*, vol. 13, pp. 253–263, 1995.
[7] Ø. Eitrheim and T. Teräsvirta, "Testing the adequacy of smooth transition autoregressive models," *J. Econometrics*, vol. 74, pp. 59–75, 1996.
[8] F. D. Foresee and M. T. Hagan, "Gauss-newton approximation to Bayesian regularization," in *IEEE Int. Conf. Neural Networks*, vol. 3. New York, 1997, pp. 1930–1935.
[9] R. Gencay, "Linear, nonlinear and essential foreign exchange rate prediction with simple technical trading rules," *J. Int. Econ.*, vol. 47, pp. 91–107, 1999.
[10] C. W. J. Granger and T. Teräsvirta, *Modeling Nonlinear Economic Relationships*. New York: Oxford Univ. Press, 1993.
[11] D. Harvey, S. Leybourne, and P. Newbold, "Testing the equality of prediction mean squared errors," *Int. J. Forecasting*, vol. 13, pp. 281–291, 1997.
[12] J. T. G. Hwang and A. A. Ding, "Prediction intervals for artificial neural networks," *J. Amer. Statist. Assoc.*, vol. 92, no. 438, pp. 109–125, 1997.
[13] C. M. Jarque and A. K. Bera, "A test of normality of observations and regression residuals," *Int. Statist. Rev.*, vol. 55, pp. 163–172, 1987.
[14] C. M. Kuan and H. White, "Artificial neural networks: An econometric perspective," *Econ. Rev.*, vol. 13, pp. 1–91, 1994.
[15] V. Kurková and P. C. Kainen, "Functionally equivalent feedforward neural networks," *Neural Comput.*, vol. 6, pp. 543–558, 1994.
[16] R. Luukkonen, P. Saikkonen, and T. Teräsvirta, "Testing linearity against smooth transition autoregressive models," *Biometrika*, vol. 75, pp. 491–499, 1988.
[17] D. J. C. MacKay, "Bayesian interpolation," *Neural Comput.*, vol. 4, pp. 415–447, 1992.
[18] M. C. Medeiros and A. Veiga, "A hybrid linear-neural model for time series forecasting," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1402–14 012, 2000.
[19] ——, "A flexible coefficient smooth transition time series model," in *Working Paper Series in Economics and Finance 361*: Stockholm School of Economics, 2000.
[20] ——, "Diagnostic checking in a flexible nonlinear time series model," J. Time Series Anal., to be published.
[21] R. A. Meese and A. K. Rose, "Nonlinear, nonparametric, nonessential exchange rate estimation," *Amer. Econ. Rev. Papers Proc.*, vol. 80, pp. 192–196, 1990.
[22] ——, "An empirical assessment of nonlinearities in models of exchange rate determination," *Rev. Econ. Studies*, vol. 58, pp. 603–619, 1991.
[23] D. Nguyen and B. Widrow, "Improving the learning speed of two-layer neural networks by choosing initial values of the adaptive weights," in *Proc. Int. Joint Conf. Neural Networks*, vol. 3, 1990, pp. 21–26.
[24] M. Qi and Y. Wu, "Exchange rates and fundamentals: Evidence from ou-of-sample forecasting using neural networks," in Computational Finance 1999: MIT Press, to be published.
[25] G. Rech, T. Teräsvirta, and R. Tschernig, "A simple variable selection technique for nonlinear models," in *Working Paper Series in Economics and Finance 296*: Stockholm School of Economics, 1999.
[26] N. Sarantis, "Modeling nonlinearities in real effective exchange rates," *J. Int. Money Finance*, vol. 18, no. 1, pp. 27–45, 1999.
[27] H. J. Sussman, "Uniqueness of the weights for minimal feedforward nets with a given input-output map," *Neural Networks*, vol. 5, pp. 589–593, 1992.
[28] T. Teräsvirta, "Specification, estimation, and evaluation of smooth transition autoregressive models," *J. Amer. Statist. Assoc.*, vol. 89, no. 425, pp. 208–218, 1994.
[29] T. Teräsvirta, C. F. Lin, and C. W. J. Granger, "Power of the neural-network linearity test," *J Time Series Anal.*, vol. 14, no. 2, pp. 309–323, 1993.
[30] H. Tong, "On a threshold model," in *Pattern Recognition and Signal Processing*, C. H. Chen, Ed. Amsterdam, The Netherlands: Sijthoff and Noordhoff, 1978.
[31] H. Tong and K. Lim, "Threshold autoregression, limit cycles and cyclical data (with discussion)," *J. Roy. Statist. Soc.*, ser. B, vol. 42, pp. 245–292, 1980.
[32] A. Trapletti, F. Leisch, and K. Hornik, "Stationary and integrated autoregressive neural-network processes," *Neural Comput.*, vol. 12, pp. 2427–2450, 2000.
[33] R. Tsay, "Testing and modeling threshold autoregressive processes," *J. Amer. Statist. Assoc.*, vol. 84, pp. 431–452, 1989.
[34] Y. Xia and W. K. Li, "On single-index coefficient regression models," *J. Amer. Statist. Assoc.*, vol. 94, no. 448, pp. 1275–1285, 1999.

Marcelo C. Medeiros was born in Rio de Janeiro, RJ, Brazil, in 1974. He received the B.S. degree in electrical engineering (systems) and the M.S. and the Ph.D. degrees from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 1996, 1998, and 2000, respectively.

Currently he is with the Department of Economics of the same university. His main research interest is nonlinear time series modeling.

Álvaro Veiga was born in Florianópolis, SC, Brazil, in 1955. He received the B.S. degree in electrical engineering (systems) from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 1978 and the M.S. degree from COPPE-UFRJ, Rio de Janeiro, in 1982, and the Ph.D. degree from École Nationale Supé rieure des Télécommunications, Paris, France, in 1989.

Since 1990, he has been at the Pontificial Catholic University of Rio de Janeiro (PUC-Rio) as an Assistant Professor. His main research interest is nonlinear modeling.

Carlos Eduardo Pedreira was born on 11 April 1956. He received the Bachelor's degree in electrical engineering (Control Systems) and the M.Sc. degree in systems engineering from Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, in 1979 and 1981, respectively, and the Ph.D. degree from Imperial College of Science Technology and Medicine, University of London, U.K., in 1987.

He is currently Associate Professor at the Electrical Engineering Department of (PUC-RIO.

Dr. Pedreira became Founding President of the Brazilian Neural Networks Council in 1992. He was a member of the International Conference on Neural Networks for Capital Markets program committee in 1995 and 1996 and a member of the International Conference on Computational Finance program committee 1997 and 1999.