Nonlinearities in Australian FX data: modelling with flexible nonlinear inference

Ralf Becker and Stan Hurn
School of Economics and Finance
Queensland University of Technology

Abstract — **This paper employs the method of flexible nonlinear inference to analyse nonlinearities in time-series data for the US$ / A$ exchange rate. By treating the functional form of the relationship between present and lagged exchange rate returns as a random variable itself, it is possible to identify features of a nonlinear relationship which facilitate the search for an adequate underlying nonlinear model. The application to daily exchange rate data illustrates the usefulness and limitations of this method and reveals that the US$ / A$ exchange rate switches between very complex processes.**

I. Introduction

The recognition that easily detectable linear structure ought not to be a common feature of data generated by efficient financial markets has led to increased interest in detecting and modelling nonlinear structure in financial data. Numerous tests have been developed to detect the existence of nonlinearities (Ramsey, [1969], McLeod and Li, [1983], Keenan, [1985], Tsay, [1986], White, [1987][1989][1992a], Brock *et al.*, [1996], Hinich, [1982], Ashley *et al.*, [1986], Liu *et al.*, [1993], Paluš *et al.*, [1995]) and their empirical performance is well documented (Lee *et al.*, [1993], Teresvirta *et al.*, [1993], Brooks, [1996], Becker and Hurn, [1999]). It is clear, however, that detection is just a first step; the final goal must be to model the data-generating process. The parametric approach to the modelling exercise offers a bewildering array of potential models, but little or no indication as to which should be used (Teräsvirta *et al.*, [1994]) and this has led to an increasing reliance on nonparametric techniques (Campbell *et al.*, [1997]).

A recent advance, pioneered by Hamilton [1999], seems to offer a valuable addition to the methods for dealing with nonlinearities in time-series data. The technique, known as flexible nonlinear inference, is designed to have the same flexibility as nonparametric estimation methods, but retain the inference capabilities of parametric methods. It has been shown that this approach not only has the ability to model simulated nonlinearities, but also that the test for nonlinearity, which is obtained as a by-product, yields encouraging empirical results (Dahl, [1998]).

The contribution of this paper is to examine the usefulness of the method using relatively high frequency Australian exchange rate data. After a brief exposition of the method, it is applied to daily foreign exchange rate data (US$/A$). Four distinct periods of nonlinearity in the behaviour of the data are identified and each of these is modelled. The nature of the appropriate model in each case appears to be markedly different, thus negating any presumption that these episodes exhibit similar kinds of nonlinearity. In addition the usefulness of the chosen models for forecasting purposes appears limited.

II. Flexible nonlinear inference

In a conventional econometric formulation, the function representing the conditional expectation of a dependent variable, y, is assumed to be deterministic function of explanatory variables, $\mathbf{x}$. It is thus the error term which accounts for the stochastic nature of y, hence

$$y_t = \mu(\mathbf{x_t}) + \varepsilon_t \qquad \varepsilon_t \sim \mathbf{iid}(0, \sigma^2) \tag{1}$$

Hamilton [1999] suggests viewing $\mu(\mathbf{x_t})$ itself as a random variable and shows that this allows a flexible and powerful representation of both linear and nonlinear functional forms. As $\mu(\mathbf{x})$ is stochastic, it cannot be observed directly. The objective is to draw inference about the model parameters by observing the realisations of y_t and $\mathbf{x_t}$ only.

Assume the two random variables $\mathbf{y}$ and $\boldsymbol{\mu}(\mathbf{x})$ are multivariate normal[1]

$$\begin{pmatrix} \mathbf{y} \\ \mu(\mathbf{x}) \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} \right)$$

then the optimal forecast of $\boldsymbol{\mu}(\mathbf{x})$ using information on $\mathbf{y}$ is distributed as follows (Hamilton, [1994])

$$\widehat{\mu(\mathbf{x})}\,|\mathbf{y} \sim N\left(\boldsymbol{\mu}_2 + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\mathbf{y}-\boldsymbol{\mu}_1), \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12}\right). \tag{2}$$

Conditional on $\mathbf{y}$ therefore the expectation of the conditional mean may be either linear or nonlinear dependent on the form of the covariance between $\mathbf{y}$ and $\boldsymbol{\mu}(\mathbf{x})$, the matrix $\boldsymbol{\Omega}_{21}$.

Accordingly, let $\mu(\mathbf{x_t})$ be given by

$$\mu(\mathbf{x_t}) = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1'\mathbf{x_t} + \boldsymbol{\lambda}\mathbf{m}(\mathbf{g} \odot \mathbf{x_t}) \tag{3}$$

with α_0 and λ scalar parameters, $\boldsymbol{\alpha}_1$ and $\mathbf{g}$ are $(k \times 1)$ parameter vectors, $\odot$ represents element by element multiplication and $m(\cdot)$ is a random variable. The parameters α_0 and $\boldsymbol{\alpha}_1$ govern the linear contribution of $\mathbf{x_t}$ to $\mu(\mathbf{x_t})$, while the scale parameter λ determines the weight of the nonlinear contribution to $\mu(\mathbf{x})$ and $\mathbf{g}$ will be shown to be a crucial element in the covariance calculation.

For any $\mathbf{y}$ and $\mathbf{x}$ it is possible to find a linear mapping f such that $E(\mathbf{y}-\mathbf{f}(\mathbf{x})) = \mathbf{0}$. This mapping is represented by $\alpha_0 + \boldsymbol{\alpha}_1'\mathbf{x}$ in (3). It is therefore obvious that the $E(m(\cdot)) = 0$ is required in order to achieve unbiased estimates of $\boldsymbol{\mu}(\mathbf{x})$. Consequently any nonlinear pattern has to be captured by the covariance matrix of $m(\cdot)$, consistent with (2) above.

[1] The validity of this assumption must be an empirical question. As the method has been shown by Dahl ([1998]) to capture a wide class of nonlinear models accurately there is at least an a priori empirical case to support this assumption.

Specifically, the properties of $m(\cdot)$ may be enumerated as follows

$$\begin{aligned} m(\mathbf{z_t}) &\sim \mathrm{N}(0,1). \\ E\left[m(\mathbf{z_t})'\mathbf{m}(\mathbf{z_s})\right] &= \begin{cases} H_k(h) & \text{if } h \leq 1 \\ 0 & \text{otherwise} \end{cases} \\ h &= \tfrac{1}{2}\sqrt{(\mathbf{z_t}-\mathbf{z_s})'(\mathbf{z_t}-\mathbf{z_s})}, \end{aligned}$$

where closed form expressions for $H_k(h)$ are derived in Hamilton ([1999]) for $k = 1,\ldots,5$. The important thing to note is that the covariance is unity when $h = 0$. The basic idea is that $m(\mathbf{z_t})$ will be correlated with $m(\mathbf{z_s})$ whenever $\mathbf{z_t}$ and $\mathbf{z_s}$ are sufficiently close, or more specifically when $|\mathbf{z_t}-\mathbf{z_s}| \leq 2$. When making inference taking $\mathbf{y}$ as given, this covariance structure can force a departure from linearity.

At this stage it is useful to introduce a slightly different notation for the model. Let the following matrices and vectors be defined: $\mathbf{y} = (\mathbf{y}_1,\ \mathbf{y}_2,\ldots \mathbf{y}_T)'$, $\boldsymbol{\beta} = (\alpha_0,\ \alpha)'$, $\varepsilon = (\varepsilon_1,\ \varepsilon_2,\ldots \varepsilon_T)'$ and

$$\mathbf{X} = \begin{bmatrix} 1 & x_1' \\ 1 & x_2' \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_T' \end{bmatrix}$$

Note that:
1. Since $E\left(m(\cdot)\right) = 0$ it follows that

$$\mu(\mathbf{X}) \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta},\ \mathbf{P}_0)$$

$$p_0(t,s) = \begin{cases} \lambda^2 H_k(h_{ts}) & \text{if } h_{ts} \leq 1 \\ 0 & \text{if } h_{ts} > 1 \end{cases}$$

$$h_{ts} = \tfrac{1}{2}\sqrt{[\mathbf{g}\odot(\mathbf{x_t}-\mathbf{x_s})]'[\mathbf{g}\odot(\mathbf{x_t}-\mathbf{x_s})]}$$

2. Since $\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$

$$\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta},\ \mathbf{P}_0 + \sigma^2\mathbf{I_T}) \tag{4}$$

where $\mathbf{I_T}$ is the T-dimensional identity matrix.
3. Since $E(\boldsymbol{\mu}(\mathbf{X})\,|\,\mathbf{X}) = \mathbf{E}(\mathbf{y}\,|\,\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$ the unconditional joint distribution of $\mathbf{y}$ and $\boldsymbol{\mu}(\mathbf{X})$ is

$$\begin{pmatrix} \mathbf{y} \\ \boldsymbol{\mu}(\mathbf{X}) \end{pmatrix} \sim N\left(\begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{X}\boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_0 + \sigma^2\mathbf{I_T} & \mathbf{P}_0 \\ \mathbf{P}_0 & \mathbf{P}_0 \end{bmatrix}\right). \tag{5}$$

4. Using the result in (2), we have

$$\begin{aligned} \widehat{\boldsymbol{\mu}(\mathbf{X})}\,|\mathbf{y} \sim\ & N(\mathbf{X}\boldsymbol{\beta} + \mathbf{P}_0(\mathbf{P}_0 + \sigma^2\mathbf{I_T})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}), \\ & \mathbf{P}_0 - \mathbf{P}_0(\mathbf{P}_0 + \sigma^2\mathbf{I_T})^{-1}\mathbf{P}_0) \end{aligned}$$

for the conditional distribution. The crucial role of $\mathbf{g}$ in the covariance calculation is now apparent. When $\mathbf{g} \to \infty$ the distance h_{ts} between $\mathbf{g}\odot\mathbf{x_t}$ and $\mathbf{g}\odot\mathbf{x_s}$ becomes arbitrarily large (for $\mathbf{x_t} \neq \mathbf{x_s}$) and therefore the nonlinear component of the model $\mathbf{P}_0$ becomes $\lambda^2\mathbf{I_T}$ which is indistinguishable from the disturbance covariace $\sigma^2\mathbf{I_T}$. Likewise when $\mathbf{g} \to \mathbf{0}$, the nonlinear component $\mathbf{P}_0$ becomes $\lambda^2\mathbf{1}\mathbf{1}'$, which makes it indistinguishable from the constant term. Neither of these two extreme cases is an improvement to the linear model. In theory the estimated parameters should avoid these extremes and values should be chosen that are most appropriate to the particular nonlinearity in the data.
5. Furthermore from (4) the unconditional log likelihood function for $\mathbf{y}$ may be written as

$$\begin{aligned} \ln L(\mathbf{y}) &= -\frac{\mathbf{T}}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\mathbf{P}_0 + \mathbf{I_T}\sigma^2\right| \\ &\quad -\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\left(\mathbf{P}_0 + \mathbf{I_T}\sigma^2\right)^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \end{aligned} \tag{6}$$

Although $\mathbf{y}\,|\mathbf{X} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{P}_0 + \mathbf{I_T}\sigma^2)$ is not the exact distribution for stochastic $\mathbf{X}$, it turns out that (6) remains valid in these situations.

It is now useful to write

$$\mathbf{P}_0 = \lambda^2\mathbf{H_k}(\mathbf{h_{ts}}(\mathbf{X},\mathbf{g})) = \mathbf{P}_0(\mathbf{X},\lambda,\mathbf{g}). \tag{7}$$

and to reformulate $\mathbf{P}_0 + \mathbf{I_T}\sigma^2$ to

$$\begin{aligned} \mathbf{P}_0 + \mathbf{I_T}\sigma^2 &= \sigma^2\left(\frac{\lambda^2}{\sigma^2}\mathbf{H}_0 + \mathbf{I_T}\right) = \sigma^2(\zeta^2\mathbf{H}_0 + \mathbf{I_T}) && (8) \\ &= \sigma^2\mathbf{W}(\mathbf{X};\lambda,\mathbf{g}) = \sigma^2\mathbf{W}(\mathbf{X};\boldsymbol{\theta}) && (9) \end{aligned}$$

using ζ as the ratio of the standard deviation of the nonlinear component to the standard deviation of the residual ε_t. Furthermore, let $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \mathbf{g}')'$, then by substituting (8) into (6) the log likelihood can be rewritten as

$$\begin{aligned} \ln L(\mathbf{y}) &= -\frac{\mathbf{T}}{2}\ln(2\pi) - \frac{\mathbf{T}}{2}\ln\sigma^2 - \frac{1}{2}\ln|\mathbf{W}(\mathbf{X};\boldsymbol{\theta})| \\ &\quad -\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{X};\boldsymbol{\theta})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \end{aligned} \tag{10}$$

For given $\boldsymbol{\theta}$ the GLS estimates $\boldsymbol{\beta}(\boldsymbol{\theta})$ and $\sigma^2(\boldsymbol{\theta})$ can be computed as

$$\begin{aligned} \widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) &= (\mathbf{X}'\mathbf{W}(\mathbf{X};\boldsymbol{\theta})^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}(\mathbf{X};\boldsymbol{\theta})^{-1}\mathbf{y} && (11) \\ \widetilde{\sigma}^2(\boldsymbol{\theta}) &= (\mathbf{y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}))'\mathbf{W}(\mathbf{X};\boldsymbol{\theta})^{-1}(\mathbf{y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}))/\mathbf{T} \end{aligned}$$

Substituting these analytical expressions into (10) yields a concentrated likelihood function that allows a formulation of the log likelihood function, which is a function of the observed dependent and independent variables, $\mathbf{y}$ and $\mathbf{X}$, and the set of nonlinear parameters, $\boldsymbol{\theta}$, only

$$\begin{aligned} \ln L(\boldsymbol{\theta}\,|\mathbf{y},\mathbf{X}) &= -\frac{\mathbf{T}}{2}\ln(2\pi) - \frac{\mathbf{T}}{2}\ln\widetilde{\sigma}^2(\boldsymbol{\theta}) \\ &\quad -\frac{1}{2}\ln|\mathbf{W}(\mathbf{X};\boldsymbol{\theta})| - \frac{T}{2}. \end{aligned} \tag{12}$$

which may be maximised to yield the ML estimate $\widehat{\boldsymbol{\theta}}$. The corresponding ML estimates of the linear parameters are $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})$ and $\widehat{\sigma}^2(\widehat{\boldsymbol{\theta}})$ and can be calculated by plugging $\widehat{\boldsymbol{\theta}}$ into (11).

Hamilton proves that this algorithm will provide a consistent estimator of $\boldsymbol{\mu}(\mathbf{x})$ for a very general class of models, both linear and nonlinear. Dahl [1998] scrutinised these theoretical propositions in a Monte Carlo experiment and demonstrated that the methodology does indeed have the ability to approximate a variety of different simulated nonlinear processes.

Finally, it may be noted that testing for nonlinearity in this framework amounts to testing whether or not $\lambda = 0$. Hamilton [1999] proposes a LM test for this hypothesis.

III. Modelling nonlinear structure

After identifying the presence of nonlinear structure in a data-set, the next natural step is to attempt to model the process. This is no mean task as there exists a wide variety of nonlinear models to choose from, but no coherent framework within which to search for the appropriate specification[2]. Current practice, as summarised by Teräsvirta *et al.* [1994] for example, suggests the following procedure. Once the hypothesis of linearity is rejected, the data set is divided into a training set and a test set. A selection of different nonlinear models is estimated using a training data-set and the model which performs best in explaining the values of the test set is selected and reestimated on the entire sample. This procedure has obvious shortcomings; for example, models may be included or excluded *a priori* without justification and different partitioning of the sample might yield different results.

It will now be demonstrated how the flexible functional form approach, outlined in the previous section, can be utilised to help identify an underlying data-generating process. The analysis starts with a maximum likelihood estimation of (12). If the parameter λ is significantly different from zero, the nonlinear term makes an important contribution to the conditional expectation function and the exogenous variable x_i causing the nonlinearity can be identified by a significant parameter g_i. In order to find out how the expected mean $\mu(\cdot)$ depends on the values of x_i, an $(N \times (k+1))$ matrix $\mathbf{X}^*$ is created.

$$\underset{(N\times(k+1))}{\mathbf{X}^*} = \begin{pmatrix} \vdots & \vdots & & \vdots & & \vdots \\ 1 & \overline{x}_1 & & \overline{x}_i - 2s & & \overline{x}_k \\ 1 & \overline{x}_1 & & \overline{x}_i - s & & \overline{x}_k \\ 1 & \overline{x}_1 & \cdots & \overline{x}_i & \cdots & \overline{x}_k \\ 1 & \overline{x}_1 & & \overline{x}_i + s & & \overline{x}_k \\ 1 & \overline{x}_1 & & \overline{x}_i + 2s & & \overline{x}_k \\ \vdots & \vdots & & \vdots & & \vdots \end{pmatrix}$$

The values for x_i are varied around the mean value on as fine a grid as required, while all other variables remain fixed at their mean values. Using equation (13)

$$\widehat{\mu}(\mathbf{X}^*) = \mathbf{X}^*\beta + \mathbf{Q}'(\mathbf{P}_0 + \sigma^2\mathbf{I}_\mathrm{T})^{-1}(\mathbf{y} - \mathbf{X}\beta) \qquad (13)$$

an $(N \times 1)$ vector of estimated conditional mean values is generated. Plotting $\widehat{\mu}(\mathbf{X}^*)$ against the values of x_i, fed into equation (13) produces a pseudo functional plot, which illustrates the dependence of the conditional mean on the variable x_i. If it is suspected that two variables excert their nonlinear influence jointly both variables are varied across a certain range and a three dimensional functional plot is created using the result in (13). As demonstrated in the following paragraph these functional plots facilitate the identification of possible parametric nonlinear models.

Hamilton provides a number of examples to illustrate the use of pseudo-functional plots. Whilst recognising the usefullness of the method, a primary remaining concern is the robustness of the method in the presence of signals with a high noise content. To investigate this issue a little further Hamilton's examples were repeated with different signal to noise ratios. The results of these experiments reveal that increasing the noise content conceals the presence of nonlinearities. As the noise increases the estimation of the parameter λ becomes less significant. As heteroscedastic noise is a widespread feature of financial data, the same experiment was repeated with ARCH and GARCH noise components. Qualitatively the same results hold.

These results are generally encouraging, particulary when compared to nonparametric estimation methods. Methods such as neural networks are known to have the power to essentially fit functions to noise (Weigend, [1996]). It appears that this type of overfitting will not be a problem with flexible nonlinear inference, indicating that the method has a sufficient degree of robustness to tackle noisy financial data.

IV. Modelling the Australian dollar

Exchange rate data have been very closely examined on nonlinearities with mixed results (Hsieh, [1989], Brooks, [1996]). Recent evidence suggests that nonlinearities in exchange rate data are episodic in nature (Guarda and Salmon, [1996], Becker and Hurn, [1999]). The exchange rate examined in this paper is the US$ / Australian $ rate[3]. Daily data were available from 1 January 1975 to 31 July 1998 (5921 observations). The log returns of these data were tested for the existence of nonlinearities using the Neural Network Test - NNT (Lee *et al.*, [1993]), the redundancy based surrogate data test - R (Becker and Hurn, [1999]) and Hamilton's test for nonlinearities - HNL (Hamilton, [1999])[4]. In order to capture the changing nature of the data generating process these tests were applied to a moving time window ($N_w = 256$). The results are summarised in Figure 1 which indicates the existence of significant nonlinearities when two consecutive time windows have p-values below 5%.

[2]Of course in the linear case Wold's representation theorem provides such a framework.

[3]The rates are spot buying rates (12.00 a.m. New York time) downloaded from the Website of the Board of Governors of the US Federal Reserve System.

[4]HNL1 to HNL3 represent varying lag specifications.

	ζ	g_1	g_2	g_3	σ
P1	-0.358	6.953	0.209		0.155
	(0.157)	(3.197)	(1.219)		(0.007)
P2	0.974	1.285	0.150		0.262
	(0.293)	(0.126)	(0.227)		(0.013)
P3	0.761	1.466	0.821	2.296	0.453
	(0.266)	(0.369)	(0.395)	(0.496)	(0.0356)
P4	1.782	1.730	2.844		0.679
	(0.357)	(0.120)	(0.237)		(0.065)

TABLE I

ESTIMATION RESULTS. PERIODS 1-4.

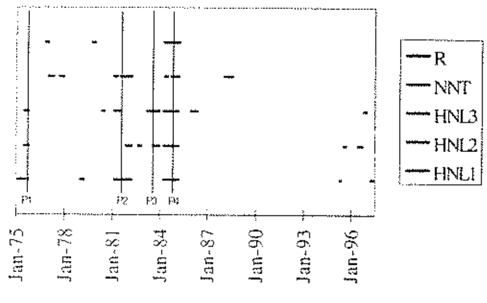

Figure 1: Nonlinearity test results for A$ / US$.

The results clearly illustrate the temporary nature of nonlinearities[5]. Four periods were singled out for further examination, namely P1 (7-Jul-75 to 13-Jul-76), P2 (8-Jul-81 to 15-Jul-82), P3 (6-Sep-83 to 12-Sep-84) and P4 (13-Jun-84 to 21-Jun-85.) for which the nonlinearity tests are significant.

The estimation results for the four periods are reported in Table 1[6]. The estimates for ζ are significant in all four selected subperiods. This corresponds to significant values of λ in (3). Since ζ is the ratio of the nonlinear component's variance to the variance of the additive noise period 4 appears to be the subperiod with the highest nonlinear component content. This is also reflected in highly significant estimates for g_1 and g_2.

Estimation of the flexible functional form (3) serves as an auxiliary step to identify a specific parametric model. The shape of the pseudo-functional plots will facilitate the identification of a particular parametric model, which subsequently will be estimated using conventional estimation techniques. The residuals resulting from this estimation should be subject to the usual battery of diagnostic tests. A model captures the nonlinearities if nonlinearity tests applied to the residuals remain insignificant.

Figures 2 contains the pseudo-functional plots for the one perios lagged return, which is primarily responsible for the nonlinearity.

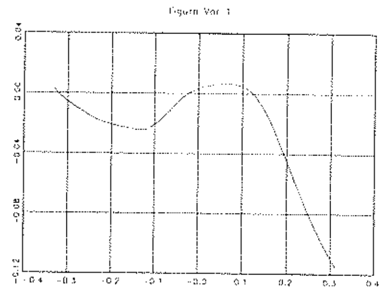

Figure 2: Pseudo-functional plot (period 1). Estimated conditional return (vertical axis) as a function of lagged exchange rate returns, y_{t-1}(horizontal axis).

One obvious choice to model this nonlinearity is a cubic term in the first lag, $(y_{t-1} - (-0.1))^3$. This model was rejected by the data. Alternatively a threshold type model was tested, $y_t = \alpha_0 + \alpha_1 y^+_{t-1}$, where $y^+_{t-1} = y_{t-1}$ if $y_{t-1} > 0.1$ and $y^+_{t-1} = 0$ otherwise. This specification proved to be satisfactory[7]. The estimation yields a significantly negative estimate for α_1. This result indicates that a negative autocorrelation between subsequent returns is significant only in cases where previous returns were of substantial positive size. Negative autocorrelation in exchange rate returns can be interpreted as a form of mean reversion. It seems reasonable to argue that mean reversion effects should become increasingly important with significant departures from a perceived equilibrium value. In the absence of any fundamental news, which might change the equilibrium value for exchange rates, a large return can be interpreted as a significant departure from a previously established equilibrium value. Since during period 1 the US$/A$ was not a freely floating exchange rate, but rather a very dirty float, mean reversion accords with economic intuition.

The second specified period displays significant nonlinearities for the one period lag, as illustrated in Figure 3. A model, including a constant term and a variable, proportional to the distance of y_{t-1} from -0.05 is sufficient to yield an insignificant result for Hamilton's nonlinearity test. However, the neural network and ARCH tests remained significant. The inclusion of a threshold variable y^+_{t-1}($y^+_{t-1} = y_{t-1}$ if $y_{t-1} > 0.4$ and $y^+_{t-1} = 0$ otherwise) addressed this problem. Both variables and the constant term are estimated highly significant. The residual diagnostics do not indicate any further problems with this specification[8].

[5]Further testing showed the presence of conditional heteroskedasticity throughout most of the sample period. The results are not reported but are available.

[6]Estimates for the linear parameter vector α are not presented but available from the authors.

[7]In addition to linearity the residuals were tested on seriel correlation (Ljung-Box and heteroscedastic consistent test for seriel correlation), heteroscedasticity (basing on exogeneous variables and ARCH) and normality. Normality is typically rejected. The residuals appear to be leptokurtic.

[8]It is interetsing to note that a modification in the mean generating process removed significant nonlinearities in the variance.

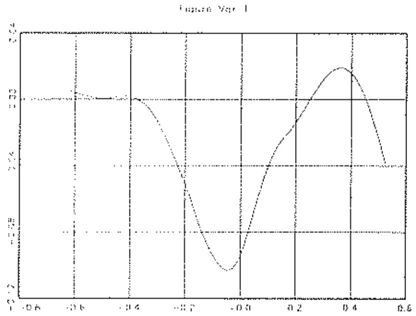

Figure 3: Pseudo-functional plot (period 2). Estimated conditional return (vertical axis) as a function of lagged exchange rate returns, y_{t-1}(horizontal axis).

The fourth period[9] is a period for which all three applied tests for nonlinearity indicate a significant departure from the null hypothesis of linearity. The estimation results reveal that the one- and two-period lagged returns contribute to the nonlinear pattern in the data, with the parameters λ and g_1 and g_2 all statistically significant.

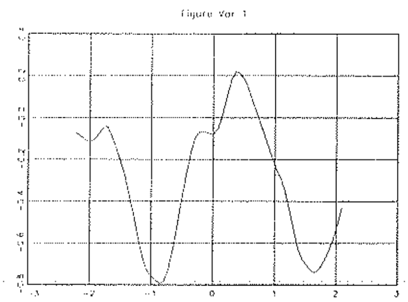

Figure 4: Pseudo-functional plot (period 4). Estimated conditional return (vertical axis) as a function of lagged exchange rate returns, y_{t-1}(horizontal axis).

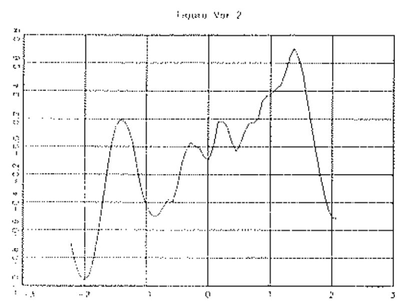

Figure 5: Pseudo-functional plot (period). Estimated conditional return (vertical axis) as a function of lagged exchange rate returns, y_{t-2}(horizontal axis).

The pseudo functional plots for $y_t = f(y_{t-1})$ and $y_t = f(y_{t-2})$ in Figure 4 and 5 confirm this and further reveal that the nonlinearity appears to be more complex than in the previous example. Further information can be drawn from the pseudo functional plot $y_t = f(y_{t-1}, y_{t-2})$, Figure 6.

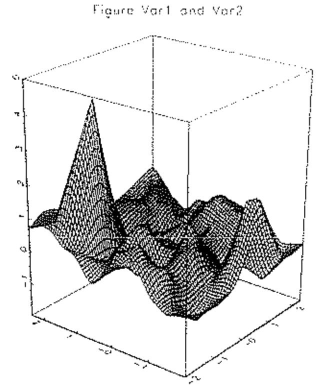

Figure 6: Pseudo-functional (surface) plot (period 4). Estimated conditional return (inverted vertical axis) as a function of lagged exchange rate returns, $y_{t-1}$and y_{t-2} (horizontal axis).

The complexity of the nonlinearities is confirmed; indeed the surface is very irregular and the only pattern which can possibly be modelled parametrically is the peak around (-1,1). The following candidate model was estimated

$$y_t = \alpha_0 + \alpha_1 D_t$$

$$D_t = \begin{cases} 1/\sqrt{(y_{t-1} - (-1))^2 + (y_{t-2} - 1)^2} & \text{if } D_t \leq 0.5 \\ 0 & \text{if } D_t > 0.5. \end{cases} \tag{14}$$

where D_t is the distance from the identified maximum at (-1,1)[10]. While the Hamilton test on nonlinearities remains insignificant when applied to this model's residuals, the redundancy and the neural network test still indicate the presence of additional nonlinear structure. ARCH effects are also present[11].

To summarise the results thus far, it must be recognised that whilst there is a reasonably sound theoretical explanation for the nonlinear relationship identified in period 1, it appears that no such economic rationale exists for periods 2 to 4. The explanation of the particular relationships which are identified remains a puzzle and it must be concluded, rather lamely, that market dynamics appear to generate very diverse and complex nonlinear patterns in the data.

Naturally the question should be asked whether the obtained estimations might be the result of overfitting. As the experiment in the previous section suggests, it is unlikely that the presence of noise causes the identification of spurious nonlinearities. One way of detecting a statistical overfit is to check whether improved in-sample fit translates into superior out-of-sample forecasting performance (Weigend *et al.*, [1995]) and hence the R^2, as a measure of

[9]Results for the third period are not presented but available upon request.

[10]The values of this maximum are approximately equivalent to exchange rate returns of -1% and +1% respectively.

[11]This methodology does not aim to remove true ARCH effects, which represent nonlinearity in variance.

in-sample fit, was compared against two measures of predictive performance, the RMSE and the MAD[12]. The forecasting performance measures were applied to the one-step ahead forecasts, based on the parameter estimates from the in-sample estimation. Summarising these experiments[13] it can be said that the improved in-sample fit is not indicative for improved forecasting performance.

V. Conclusion

The research presented in this paper allows several conclusions. It appears that the method of nonlinear flexible inference is powerful enough to identify very diverse nonlinear relationships. A tentative remark here is that threshold type models seem to be the easiest to identify, but further research is needed to clarify the range of nonlinear models which flexible inference can deal with in practice.

In terms of modelling the US$/A$ four episodes of nonlinear structure in the data were identified. The nature of the nonlinearities was distinctively different for each of the four identified episodes. This implies that simple regime switching models might be ill-fated. However, none of these relationships generalises very well into the post-sample periods which is arguably the most important feature of the modelling process. Consequently the identified models did not improve the quality of point forecasts when compared to linear models. Future research might examine whether the incorporation of nonlinearities can improve interval or density forecasts.

References

[1986] Ashley, R., Patterson, D. and Hinich, M (1986) "A diagnostic test for nonlinear serial dependence in time series fitting errors", *Journal of Time Series Analysis*, **7**, 165-178.

[1999] Becker R. and Hurn A.S. (1999) "Testing for fundamental nonlinearities in time-series data using the method of surrogates", Discussion Paper No.52, School of Economics and Finance, Queensland University of Technology.

[1996] Brock, W. A., Dechert, W. D., Scheinkman, J. A. and LeBaron, B. (1996) "A Test for Independence Based on the Correlation Dimension" *Econometric Reviews*, **15**, 3, 197-235.

[1996] Brooks, S. (1996) "Testing for non-linearity in daily sterling exchange rates" *Applied Financial Economics*, **6**, 307-317.

[1997] Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997) *The Econometrics of Financial Markets*, Princeton University Press: Princeton.

[1998] Dahl C.M. (1998) "An investigation of tests for linearity and the accuracy of flexible nonlinear inference", mimeo, University of Aarhus.

[1986] Fraser A.M. and Swinney H.L. (1986) "Independent coordinates for strange attractors from mutual information" *Physical Review A*, 33, 1134-1140.

[1996] Guarda P. and Salmon M. (1996) "Detection of nonlinearity in foreign-exchange data" in: Barnett W.A., Kirman A.P. and Salmon M. *Nonlinear Dynamics in Economics*, Cambridge University Press: Cambridge.

[1994] Hamilton J.D. (1994) *Time Series Analysis*, Princeton University Press: Princeton.

[1999] Hamilton J.D. (1999) "A parametric approach to flexible nonlinear inference", mimeo, University of California, San Diego.

[1982] Hinich, M.J. (1982) "Testing for gaussianity and linearity of stationary time series", *Journal of Time Series Analysis*, **3**, 169-176.

[1989] Hsieh D.A. (1989) "Testing for nonlinear dependence in daily foreign exchange rates" *Journal of Business*, **62**, 3, 239-368.

[1985] Keenan, D.M. (1985) "A Tukey nonadditivity type test for time series nonlinearity", *Biometrika*, **72**, 39-44.

[1999] Kim H.S., Eykholt R. and Salas J.D. (1999) "Nonlinear dynamics, delay times and embedding windows" *Physica D*, 127, 48-60.

[1993] Lee, T., White, H. and Granger, C.W.J., (1993) "Testing for neglected nonlinearity in time series models", *Journal of Econometrics*, **56**, 269-290.

[1993] Liu, T., Granger, C.W.J. and Heller, W.P., (1993) "Using the correlation exponent to decide whether an economic series is chaotic", in: Pesaran, M.H., Potter, S.M. *(eds.)*, *Nonlinear Dynamics, Chaos and Econometrics*, J. Wiley: New York.

[1983] McLeod, A. I. and Li, W. K. (1983) "Diagnostic checking ARMA time series models using squared-residual autocorrelations" *Journal of Time Series Analysis*, **4**, 4, 269-273.

[1995] Paluš M., Pecan L. and Pivka D. (1995) "Estimating predictability: redundancy and surrogate data method" *Neural Network World*, **5**, 4, 537-552.

[1969] Ramsey, J.B. (1969) "Tests for specification errors in classical linear least squares regression analysis", *Journal of the Royal Statistical Society B*, **31**, 350-371.

[1993] Teresvirta T., Lin C.F. and Granger C.W.J. (1993) "Power of the neural network linearity test", *Journal of Time Series Analysis*, **14**, 209-220.

[1994] Teresvirta T., Tjøstheim D. and Granger C.W.J. (1994) "Aspects of modelling nonlinear time series", in: Engle R.F. and McFadden D.L., Handbook of Econometrics, Vol.IV, Elsevier Science.

[1986] Tsay, R. S. (1986) "Nonlinearity test for time series" *Biometrika*, **73**, 2, 461-466.

[1996] Weigend A.S. (1996) "Time Series Analysis and Prediction" in: Smolensky P., Mozer M.C. and Rumelhart D.E., *Mathematical Perspectives on Neural Networks*, Lawrence Erlbaum, Mahwah.

[1995] Weigend A.S., Mangeas M. and Srivastava A.N. (1995) "Nonlinear gated experts for time series: discovering regimes and avoiding overfitting", *International Journal of Neural Systems*, **6**, 373-399.

[1987] White H. (1987) "Specification testing in dynamic models" in Bewley T.F. (*ed.*) *Advances in Econometrics, Fifth World Congress*, Cambridge University Press: Cambridge.

[1989] White H. (1989) "An additional hidden unit test for neglected nonlinearity in multilayer feedworward networks" in *Proceedings of the International Joint Conference of Neural Networks, Washington, DC*, IEEE Press: New York.

[1992a] White H. (1992a) *Estimation, Inference and Specification Analysis*, Cambridge University Press: Cambridge.

[12] The forecasting period was the next 256 observations.

[13] Results are available from the authors.