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 structures are not to be a common feature of data generated by efficient financial markets has led to increased interest in detecting and modelling nonlinear structures in financial data. Numerous tests have been developed to detect the existence of nonlinearities (Ramsay, 1969; McLeod and Li, 1983; Koenen, 1985; Tsay, 1986; White, 1987; 1989; 1992a; Brock et al., 1990; Hurnich, 1992; Ashley et al., 1986; Liu et al., 1993; Pahal et al., 1993) and their empirical performance is well documented (Lee et al., 1993; Teresvita et al., 1993; Brooks, 1996; Becker and Hurn, 1997). 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 (Teresvita et al., 1994) and there is 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 it retains 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 modeled. 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, $x$. It is thus the error term which accounts for the stochastic nature of $y$, hence

$$ y_t = \mu(x_t) + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma^2) \quad (1) $$

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

Assume the two random variables $y$ and $\mu(x)$ are multivariate normal. Let

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

then the optimal forecast of $\mu(x)$ using information on $y$ is distributed as follows (Hamilton, 1994)

$$ \hat{\mu}(x_t) | y \sim N \left( \Omega_{22} \Omega_{12}^{-1} (y - \mu_1), \Omega_{22} - \Omega_{22} \Omega_{12}^{-1} \Omega_{12} \right) \quad (2) $$

Conditional on $y$ therefore the expectation of the conditional mean may be either linear or nonlinear dependent on the form of the covariance between $y$ and $\mu(x)$, the matrix $\Omega_{12}$. Accordingly, let $\mu(x_t)$ be given by

$$ \mu(x_t) = \alpha_0 + \alpha_1 x_t + \lambda m(g, x_t) \quad (3) $$

with $\alpha_0$ and $\lambda$ scalar parameters, $\alpha_1$ and $g$ are $k \times 1$ parameter vectors, $\lambda$ represents element by element multiplication and $m(\cdot)$ is a random variable. The parameters $\alpha_0$ and $\alpha_1$ govern the linear contribution of $x_t$ to $\mu(x_t)$, while the scale parameter $\lambda$ determines the weight of the nonlinear contribution to $\mu(x)$ and $g$ will be shown to be a crucial element in the covariance calculation.

For any $y$ and $x$ it is possible to find a linear mapping $f$ such that $E(y - f(x)) = 0$. This mapping is represented by $\alpha_0 + \alpha_1^* x$ in (3). It is therefore obvious that the $E(m(\cdot)) = 0$ is required in order to achieve unbiased estimates of $\mu(x)$. 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 (1998c) 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:

$$
m(z_i) \sim N(0, 1),
$$

$$
E(m(z_i) m(z_j)) = \begin{cases} H_0(h) & \text{if } h \leq 1 \\
0 & \text{otherwise} \
\end{cases} h = \frac{1}{2} \sqrt{(z_i - z_o)^T(z_i - z_o)},
$$

where closed form expressions for $H_0(h)$ are derived in Hamilton ([1999]) for $h = 1, \ldots, 5$. The important thing to note is that the covariance is unity when $h = 0$. The basic idea is that $m(z_i)$ will be correlated with $m(z_o)$ whenever $z_i$ and $z_o$ are sufficiently close, or more specifically when $|z_i - z_o| \leq 2$. When making inference taking $y$ 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: $y = (y_1, y_2, \ldots, y_T)^T$, $\beta = (\alpha_0, \alpha_T)^T$, $\epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_T)^T$ and

$$
X = \begin{bmatrix}
1 & x_1^T \\
1 & x_2^T \\
\vdots & \vdots \\
1 & x_T^T
\end{bmatrix}
$$

Note that:

1. Since $E(m(\cdot)) = 0$ it follows that

$$
\mu(X) = N(X\beta, P_0)
$$

2. Since $y = \mu + \epsilon$

$$
\sim N(X\beta, P_0 + \sigma^2 I_T)
$$

3. $E(y|x) = X\beta$ the unconditional joint distribution of $y$ and $\mu(X)$ is

$$
\begin{bmatrix} y \\
\mu(X) \end{bmatrix} \sim N\left(\begin{bmatrix} X\beta \\
P_0 + \sigma^2 I_T \\
P_0 \end{bmatrix}, \begin{bmatrix} P_0 & 0 & 0 \\
0 & P_0 & P_0 \\
0 & P_0 & P_0 \end{bmatrix}\right)
$$

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

$$
\mu(X|y) \sim N(X\beta + P_0 (P_0 + \sigma^2 I_T)^{-1}(y - X\beta), P_0 - P_0 (P_0 + \sigma^2 I_T)^{-1} P_0)
$$

for the conditional distribution. The crucial role of $g$ in the covariance calculation is now apparent. When $g \rightarrow \infty$ the distance $h_{1a}$ between $g \cdot x_1$ and $g \cdot x_o$ becomes arbitrarily large (for $x_t \neq x_o$) and therefore the nonlinear component of the model $P_0$ becomes $\lambda^2 I_T$ which is indistinguishable from the disturbance covariance $\sigma^2 I_T$. Likewise when $g \rightarrow 0$, the nonlinear component $P_0$ becomes $\lambda^2 11^T$, 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 nonlinear in the data.

Furthermore from (4) the unconditional log likelihood function for $y$ may be written as

$$
ln L(y) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |P_0 + I_T \sigma^2| - \frac{1}{2} (y - X\beta)^T(P_0 + I_T \sigma^2)^{-1}(y - X\beta)
$$

Although $y|X \sim N(X\beta, P_0 + I_T \sigma^2)$ is not the exact distribution for stochastic $X$, it turns out that (9) remains valid in these situations.

It is now useful to write

$$
P_0 = \lambda^2 H_0(h_{1a}(X, g)) = P_0(X, \lambda, g)
$$

and to reformulate $P_0 + I_T \sigma^2$ to

$$
P_0 + I_T \sigma^2 = \sigma^2 \left(\frac{\lambda^2}{\sigma^2} H_0 + I_T\right) = \sigma^2 (\zeta H_0 + I_T)/\sigma^2
$$

$$
\zeta = \sigma^2 W(X, \lambda, g) = \sigma^2 W(X, \theta)
$$

using $\zeta$ as the ratio of the standard deviation of the nonlinear component to the standard deviation of the residual $\varepsilon_n$. Furthermore, let $\theta = (\lambda, g)^T$, then by substituting (8) into (6) the log likelihood can be rewritten as

$$
ln L(y) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2} \ln |W(X, \theta)| - \frac{1}{2} (y - X\beta)^T W(X, \theta)^{-1} (y - X\beta)
$$

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

$$
\beta(\theta) = (X'W(X, \theta)^{-1}X)^{-1}X'W(X, \theta)^{-1}y
$$

$$
\sigma^2(\theta) = (y - X\beta(\theta))'W(X, \theta)^{-1}(y - X\beta(\theta))/T
$$

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, $y$ and $X$, and the set of nonlinear parameters, $\theta$, only

$$
ln L(\theta | y, X ) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2(\theta) - \frac{1}{2} \ln |W(X, \theta)| - \frac{T}{2}
$$

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

Hamilton proves that this algorithm will provide a consistent estimator of $\mu(x)$ for a very general class of models, both linear and nonlinear. Dahl [1998] scrutinized these
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 recognizing the usefulness 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 heteroskedastic 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, particularly 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. 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]). In order to capture the changing nature of the data generating process these tests were applied to a moving time window (Nt = 250). The results are summarised in Figure 1 which indicates the existence of significant nonlinearities when two consecutive time windows have p-values below 5%.

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.

HNLL to HNLL3 represent varying lag specifications.
The results clearly illustrate the temporary nature of nonlinearities. 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 2. The estimates for ζ are significant in all four selected sub-periods. 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 sub-period with the highest nonlinear component content. This is also reflected in highly significant estimates for g1 and g2.

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.

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

One obvious choice to model this nonlinearity is a cubic term in the first lag, \((y_{t-1} = (0.11)^2)\). This model was rejected by the data. Alternatively a threshold type model was tested, \(y_t = \alpha_0 + \alpha_1 y_{t-1}^{+ \text{h}}\), where \(y_{t-1}^{+ \text{h}} = y_{t-1}\) if \(y_{t-1} > 0.1\) and \(y_{t-1}^{+ \text{h}} = 0\) otherwise. This specification proved to be satisfactory. The estimation yields a significantly negative estimate for \(\alpha_1\). 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 USD/AED 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 \(y_{t-1}^{+ \text{h}}(y_{t-1}^{+ \text{h}} = y_{t-1}\) if \(y_{t-1} > 0.4\) and \(y_{t-1}^{+ \text{h}} = 0\) 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.

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

Estimates for the linear parameter vector \(\alpha\) are not presented but available from the authors.

---

Table 2

<table>
<thead>
<tr>
<th>ζ</th>
<th>g1</th>
<th>g2</th>
<th>g3</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>0.368</td>
<td>0.155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.157)</td>
<td>(0.157)</td>
<td>(0.157)</td>
<td>(0.157)</td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>0.251</td>
<td>0.251</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.126)</td>
<td>(0.126)</td>
<td>(0.126)</td>
<td>(0.126)</td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td>0.251</td>
<td>0.251</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.126)</td>
<td>(0.126)</td>
<td>(0.126)</td>
<td>(0.126)</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>0.251</td>
<td>0.251</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.126)</td>
<td>(0.126)</td>
<td>(0.126)</td>
<td>(0.126)</td>
<td></td>
</tr>
</tbody>
</table>

---

In addition to linearity the residuals were tested on serial correlation (Ljung-Box and heteroskedastic consistent test for serial correlation), heteroskedasticity (testing on exogenous variables and ARCH) and normality. Normality is typically rejected. The residuals appear to be leptokurtic.

It is interesting to note that a modification in the mean generating process removed significant nonlinearities in the variance.
The fourth period\(^2\) 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 \(\lambda\) and \(g_1\) and \(g_2\) all statistically significant.

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.

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

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

where \(D_t\) is the distance from the identified maximum at \((-1, 1)\). 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\(^3\).

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 cause 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., [1996]) and hence the \(R^2\), as a measure of

\(^2\)Results for the third period are not presented but available upon request.

\(^3\)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. 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, 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


The forecasting period was the next 250 observations.

Results are available from the authors.