

# Modelling the Dynamics of Exchange Rate Volatility: A Nonlinear Markov-Modelling Approach

Berndt Pilgram<sup>a</sup>, Peter Verhoeven<sup>b</sup>, Alistair Mees<sup>c</sup>, and Michael McAleer<sup>d</sup>

<sup>a,c</sup>Department of Mathematics and Statistics, University of Western Australia, Nedlands, WA 6907

<sup>b,d</sup>Department of Economics, University of Western Australia, Nedlands, WA 6907

## Abstract

We report on a novel forecasting method based on nonlinear Markov modelling and canonical variate analysis, and investigate the use of a prediction algorithm to forecast conditional volatility. In particular, we assess the dynamic behaviour of the model by forecasting exchange rate volatility. It is found that the nonlinear Markov model can forecast exchange rate volatility significantly better than the GARCH(1,1) model due to its flexibility in accommodating nonlinear dynamic patterns in volatility, which are not well captured by the linear GARCH(1,1) model.

## 1 Introduction

In finance, volatility is a key measure of risk and of the relative change in the price of a security, such as stock, stock index, or derivative, over time. Thus, the greater is the price variation, the greater is volatility. As the true underlying volatility of a security is unobservable, it must be estimated. Although there are different expressions for volatility, the definition used in finance is typically the standard deviation of the returns of a security over a given period.

Volatility is an essential input to the optimisation of financial models describing the expected risk-return trade-off. For example, it is a crucial input to mean-variance portfolio optimisation models and for the pricing of both primary and secondary derivative securities. In general, the higher is the volatility, the greater is the value of an option. Thus, it is essential for practitioners to be able to model the volatility dynamics of financial securities adequately.

Any model that attempts to predict volatility will need to incorporate the following important dynamics in returns:

1. Financial markets frequently experience large and sudden price movements. A recent example of extreme price movements is the October 1997 stock market crash originating in Asia.

On 28 October 1997, the Hang Seng Stock Index (HSI) dropped by 14.7%, the German Stock Index (DAX) by 7.2%, the Standard & Poor's 500 Composite Index (S&P500) by 5.0%, and the Japanese Stock Index (Nikkei 225) by 4.4%. A consequence of these extreme observations is the fat-tailed distribution of returns.

2. There is overwhelming evidence that the tail behaviour of equity returns evolves over time. In particular, absolute returns have significant positive serial correlation over long lags, implying that they have long term memory. This is known as volatility clustering, whereby large (small) absolute returns are more likely to be followed by large (small) absolute returns than by small (large) absolute returns. In other words, volatility is positively correlated over time.
3. Equity returns are highly asymmetric. In particular, negative shocks to returns (bad news) lead to larger volatility than equivalent positive shocks to returns (good news) (Koutmos [1998]). This is the so-called "leverage effect".
4. The persistence of shocks to volatility is asymmetrically related to the size of the shocks. When shocks to returns are high (low), trends persist for shorter (longer) periods (Engle and Lee [1993]), which means that the market reverses itself.

Hence, the implication for practitioners is that financial market volatility is predictable.

The most commonly used model to forecast volatility is the generalised autoregressive conditional heteroskedastic GARCH(1,1) model of Engle [1982] and Bollerslev [1986]. Its popularity is due to the fact that it: (i) captures the persistence of volatility; (ii) accommodates the fat-tails of the returns distribution; and (iii) is simple, and also mathematically and computationally straightforward. However, the GARCH model imposes a symmetrical influence of lagged residuals on current volatility, thereby failing to accommodate sign asymmetries. Moreover, high and low volatility

shocks have the same rate of persistence. Considering these shortcomings, more flexible methods of modelling volatility are in demand.

In this paper a nonlinear approach is taken which provides flexibility in its ability to model temporal asymmetries as well as persistence. Although it has been argued that improved in-sample fit does not necessarily lead to improved out-of-sample forecasting ability, unless the non-linearities are realised in the latter period (Terasvirta and Anderson [1992]), we argue that non-linear models will, on average, yield improved forecasts.

This paper is organised as follows. Section 2 describes the nonlinear Markov modelling approach. In Section 3, we give a detailed outline of the implementation of the nonlinear Markov modelling and forecasting algorithm. Section 4 describes the GARCH(1,1) model. Section 5 presents the data analysis while Section 6 analyses the empirical results. Some concluding remarks are given in Section 7.

## 2 Nonlinear Markov modelling

We introduce a nonlinear Markov modelling approach based on canonical variate analysis (CVA), which was first developed by Hotelling [1936]. The method we use for constructing models from time series with non-trivial dynamics is an extension of the work published by Larimore [1991], and involves the analysis of canonical correlations and variates from the *past* and *future* of a process. Although CVA theory was originally developed for independent and identically distributed (i.i.d.) random variables, we apply CVA to time series exhibiting nonlinear correlations.

Consider a nonlinear, time invariant, strict sense, discrete-time Markov process with no deterministic input to the system. Let this purely stochastic process be observed at equal "sampling intervals" to yield a time series given by

$$y_t |_{t=1,2,\dots,N} \quad (1)$$

where  $y_t$  has zero mean and constant variance  $\sigma^2$ . Associated with each time  $t$ , define a *past* vector  $p_t$ , given as an  $m$ -dimensional uniform embedding of the scalar time series  $y_t$ . However, there exist more sophisticated embedding procedures (Judd and Mees [1998]). Thus, consider a non-uniform embedding introduced by the lag vector  $l = (l_1, l_2, \dots, l_m)$ , a vector of positive integers, and obtain the past vector  $p_t$  as

$$p_t = (y_{t-l_1}, y_{t-l_2}, \dots, y_{t-l_m}). \quad (2)$$

The *future* vector  $f_t$  of finite window length  $n$  is given by

$$f_t = (y_t, y_{t+1}, \dots, y_{t+n-1}). \quad (3)$$

Given the two vectors  $p_t$  and  $f_t$ ,  $p_t$  is the set of predictor variables and  $f_t$  is the set of variables to be predicted.

The fundamental characteristic of a nonlinear, time invariant, strict sense discrete-time Markov process of finite state order is its finite dimensional state  $s_t$ . This is approximated by an  $r$ -dimensional reduced memory vector  $m_t$ , given as a nonlinear function  $\phi$  of the past, that is,

$$s_t \approx m_t = \phi(p_t). \quad (4)$$

State  $s_t$  has the property that the conditional probability of the future  $f_t$  given the past is identical to the conditional probability of  $f_t$  given  $s_t$ , that is,

$$P(f_t | p_t) = P(f_t | s_t). \quad (5)$$

Thus, only a finite number  $r$  of nonlinear combinations of the past is relevant to the future.

The primary effort in calculating an optimal nonlinear prediction  $\hat{f}_t$  of the future  $f_t$  involves the determination of  $r$  nonlinear combinations of the past  $p_t$ . The optimal prediction  $\hat{f}_t$  is a linear combination of the  $r$ -dimensional reduced memory vector  $m_t$ , where the nonlinear function  $\phi$  of the past  $p_t$  is chosen such that the optimal linear predictor  $\hat{f}_t(m_t)$  minimizes the prediction error.

Now select a class of nonlinear functions  $f_i, i = 1, 2, \dots, k$ , of the past  $p_t$  to obtain a set of basis functions  $\pi_t$  to approximate the future; that is,

$$\pi_t = (f_1(p_t), f_2(p_t), \dots, f_k(p_t)) \quad (6)$$

where  $k$  is the number of nonlinear basis functions. We use radial basis functions as basis functions  $f_i$  of the past  $p_t$  to approximate the future  $f_t$  for CVA. The standard radial basis function is defined as

$$f_i(p_t) = \Phi \left( \frac{|p_t - c_i|}{r_i} \right) \quad (7)$$

for suitably chosen centres  $c_i$ , radii  $r_i$ , and radial basis function  $\Phi$ .

The predominant effort in estimating the optimal basis functions  $f_i$  which are nonlinear functions of centres  $c_i$  and radii  $r_i$ , now involves the application of a selection algorithm (Judd and Mees [1995]). Construct a class of parameterised nonlinear autoregressive models called *pseudo-linear models* from the embedding  $p_t$ ; that is,

$$y_t = \sum_{i=1}^k \lambda_i f_i(p_t) + \epsilon_t = \sum_{i=1}^k \lambda_i \Phi \left( \frac{|p_t - c_i|}{r_i} \right) + \epsilon_t \quad (8)$$

for some selection of nonlinear functions  $f_i$ , unknown parameters  $\lambda_i$ , unknown i.i.d. random variates  $\epsilon_t$ , and a given number  $k$ . The choice of  $k$

is arbitrary, and  $k$  has to be large enough to describe the data sufficiently well from the measured system.

The basis set, the functions  $f_i|_{[i=1,2,\dots,k]}$ , is obtained as a set of basis functions that approximates the data  $y_t$ . In the following, we use the set of functions  $f_i|_{[i=1,2,\dots,k]}$  as a set of basis functions  $\pi_t = (f_1, f_2, \dots, f_k)$ , given in Eq. 6, to predict the future  $f_t$ .

The optimal prediction problem is solved by a near maximum likelihood system identification procedure (Larimore [1991]), as follows. Assuming a linear relationship describing the optimal prediction of  $f_t$  from  $\pi_t$ , consider the following model

$$\begin{aligned} f_t &= Bm_t + e_t \\ m_t &= A\pi_t(p_t) = \phi(p_t) \end{aligned} \quad (9)$$

where memory  $m_t$  is an intermediate set of  $r$  variables that may be fewer in number than  $\pi_t$ . Term  $e_t$  with covariance matrix  $\Sigma_{ee}$  is the error in the linear prediction of  $f_t$  from  $\pi_t$  given by matrices  $A$  and  $B$ . For simplicity, denote the matrices  $M$ , containing the intermediate set of  $r$  variables  $m_t$ ,  $E$  containing the prediction-error variables  $e_t$ ,  $F$  the future vectors  $f_t$ , and  $\Pi$  the basis set  $\pi_t$ . Furthermore, define the covariance matrices of the basis set, the future, and the prediction error by  $\Sigma_{\pi\pi} = \frac{1}{N}\Pi\Pi^T$ ,  $\Sigma_{ff} = \frac{1}{N}FF^T$ , and  $\Sigma_{ee} = \frac{1}{N}EE^T$ , respectively. The cross-covariance matrix of the basis set and the future is given by  $\Sigma_{\pi f} = \frac{1}{N}\Pi F^T$ .

A maximum likelihood estimator of  $A$ ,  $B$ , and  $\Sigma_{ee}$  is naturally defined by the conditional likelihood function  $p(F|\Pi; A, B, \Sigma_{ee})$  of the future  $F$  given the basis set  $\Pi$ . Maximum likelihood estimation (MLE) involves substituting  $\Sigma_{ee}$ , and estimating  $A$  and  $B$  as the matrices that maximize the likelihood for the given basis set and future of the observed process.

**CVA Theorem.** Let  $\Sigma_{\pi\pi}$  ( $m \times m$ ) and  $\Sigma_{ff}$  ( $n \times n$ ), the covariance matrices of the basis set and the future, respectively, be nonnegative definite (satisfied by covariance matrices). Then there exist matrices  $J$  ( $m \times m$ ) and  $L$  ( $n \times n$ ) such that

$$\begin{aligned} J\Sigma_{\pi\pi}J^T &= I_{r_{\pi\pi}} \\ L\Sigma_{ff}L^T &= I_{r_{ff}} \\ J\Sigma_{\pi f}L^T &= D = \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_r, 0, \dots, 0) \end{aligned} \quad (10)$$

where  $r_{\pi\pi} = \text{rank}(\Sigma_{\pi\pi})$ ,  $r_{ff} = \text{rank}(\Sigma_{ff})$ , and  $\gamma_i$  are the canonical correlations. Matrix  $I_r$  denotes the  $r \times r$  identity matrix.

CVA is a generalised singular value decomposition which transforms basis set  $\pi_t$  and future  $f_t$  to pairwise correlated i.i.d. random variables. Matrices  $J$  and  $L$  are obtained via a singular value decomposition (SVD) of the cross-covariance matrix  $\Sigma_{\pi f}$ .

After substitution of the CVA into the log of the likelihood function  $p(F|\Pi; A, B, \Sigma_{ee})$ , substitution

of  $\Sigma_{ee}$ , and maximisation over  $A$  and  $B$ , we obtain the following estimates for  $A$ :

$$\hat{A} = (I_r \ 0)J \quad (11)$$

with  $\hat{A}$  the first  $r$  rows of  $J$ , and for  $B$ :

$$\hat{B} = (I_r \ 0)L \quad (12)$$

with  $\hat{B}$  the first  $r$  rows of  $L$ . Subsequently, we obtain for  $M$ :

$$M = (I_r \ 0)J\Pi \quad (13)$$

or for instant time  $t$ :

$$m_t = (I_r \ 0)J\pi_t. \quad (14)$$

The critical problem now is to determine the rank  $r$  of memory  $M$ , *i.e.* the optimal dimension  $r$  of  $M$  to predict  $F$ . Matrix  $M$  contains the optimal rank  $r$  predictors which are the first  $r$  canonical variables  $c_1, c_2, \dots, c_r$ , where the optimal rank  $r$  is obtained from the number of dominant canonical correlations  $\gamma_i$  (Larimore [1991]). The number of dominant canonical correlations, *i.e.* the optimal rank  $r$ , is chosen as the one which gives the best prediction.

### 3 Implementation of forecasting

In practice, given the time series  $y_t|_{[t=1,2,\dots,N]}$  sampled at equal "sampling intervals", the standard problem is to construct a model and then to predict one-step ahead to obtain the future  $\hat{y}_t|_{[t=N+1}$ . The modelling problem is solved by a near maximum likelihood system identification procedure (Larimore [1991]) of the system, given in Eq. 9. Thus, one obtains matrix  $A$ , matrix  $B$ , and a nonlinear function  $\phi$  which is a nonlinear embedding  $\pi_t(p_t)$  of the past. Assume the past embedding  $p_N$  simply given as

$$p_N = (y_N, y_{N-1}, \dots, y_{N+1-m}) \quad (15)$$

where  $m$  is the embedding dimension. Substituting the past embedding  $p_N$  into Eq. 9, we obtain the future time series

$$\hat{f}_N = (\hat{y}_{N+1}, \hat{y}_{N+2}, \dots, \hat{y}_{N+n}) \quad (16)$$

as

$$\hat{f}_N = BA\pi_N(p_N) = B\phi(p_N). \quad (17)$$

Hence, future  $\hat{y}_t|_{[t=N+1}$  is the first element  $\hat{y}_{N+1}$  of the future vector  $\hat{f}_N$ .

In the following, we outline the implementation of the CVA prediction algorithm in detail.

1. Given the time series  $y_t|_{t=1,2,\dots,N}$ , determine the optimal embedding of the past  $p_t$ , *i.e.* embedding dimension  $m$  and lag vector  $l = (l_1, l_2, \dots, l_m)$ , construct the embedding, and obtain the embedding matrix  $P$ .
2. Select the  $k$  best fitting functions  $f_i$  from randomly generated radial basis functions to obtain an optimal nonlinear embedding. To ensure a good selection of basis functions, this procedure is repeated  $\iota$ -times and we obtain centres  $c_i$  and radii  $r_i$  of the selected basis functions which form the nonlinear embedding matrices  $\Pi_1, \Pi_2, \dots, \Pi_\iota$ . Finally, embedding matrix  $\Pi$  of size  $\nu$  is obtained from the nonlinear embedding matrices  $\Pi_i$ , a constant term  $c$  and linear embedding matrix  $P$ , *i.e.*

$$\Pi = (c \ P \ \Pi_1 \ \Pi_2 \ \dots \ \Pi_\iota). \quad (18)$$

3. Given a future window length  $n$ , generate the future matrix  $F$ .
4. Solve the following system

$$\begin{aligned} F &= BM + E \\ M &= A\Pi \end{aligned} \quad (19)$$

using CVA. Matrices  $J$  and  $L$  are obtained via an SVD of cross-covariance matrix  $\Sigma_{\pi f}$ . Then, calculate estimates of  $\hat{A} = (I_r \ 0)J$ ,  $\hat{B} = (I_r \ 0)L$ , and subsequently  $M = (I_r \ 0)J\Pi$ . The rank  $r$  of memory  $M$ , *i.e.* the optimal dimension  $r$  of  $M$  to predict  $F$ , is chosen as the one which gives the best one-step ahead predictions.

5. Build the embedding vector  $\pi_N$  from the past, using parameters  $r_i$ ,  $c_i$ , and  $c$ . Then, estimate the future vector  $\hat{f}_N$  using model parameters  $A$  and  $B$ . Subsequently, predict one-step ahead and obtain the estimated future  $\hat{y}_{N+1}$ .

## 4 The GARCH(1,1) model

The AR(1)-GARCH(1,1) model has the following structure:

$$r_t = \mu + \phi_1 r_{t-1} + \epsilon_t \quad (20)$$

and the conditional variance of  $\epsilon_t$  is generated by

$$\epsilon_t = \eta \sqrt{h_t} \quad (21)$$

$$h_t = \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1}, \quad (22)$$

where  $\eta_t$  is a sequence of normal i.i.d. random variables (shocks), with zero mean and unit variance. Sufficient conditions for positivity of the conditional variance and the GARCH(1,1) process to exist are that  $\omega > 0$ ,  $\alpha > 0$  and  $\beta \geq 0$ .

Several statistical properties have been established for the GARCH(1,1) process in order to define the unconditional moments of  $\{\epsilon_t\}$  (Bollerslev [1986]). First, the second moment of  $\{\epsilon_t\}$  exists if  $(\alpha + \beta) < 1$ . This condition must be met in order for the GARCH(1,1) process to be strictly stationary and ergodic, and  $E(\epsilon_t^2) < \infty$ .

Second, a necessary and sufficient condition for the existence of the fourth moment of  $\epsilon_t$  is  $(3\alpha^2 + 2\alpha\beta + \beta^2) < 1$  (Bollerslev [1986]). The normality condition is made in order to define the likelihood function, but is not required for asymptotic results.

## 5 Data analysis

This paper considers the nonlinear Markov modelling approach and the AR(1)-GARCH(1,1) model for returns. The models are evaluated using the noon (Pacific time) British Pound-U.S. Dollar (GBP/USD) spot exchange rates, obtained from the Pacific Exchange Rate Service.

A rolling window of 500 trading days was used to estimate the parameters of the model and generate the one-step ahead forecasts. The forecasts were compared with the realised volatility using the following definition:

$$\sigma_t = |y_t - \bar{y}|. \quad (23)$$

We applied the nonlinear Markov modelling approach to the volatility sequence  $\sigma_t$ . To reduce the additive noise component, we pre-filtered the volatility series by using a linear filter with exponentially decreasing filter coefficients, that is,

$$y_t = \sum_{j=t-f_t+1}^{j=t} \sigma_j w_{j-t+f_t} \quad (24)$$

where  $f_t$  is the filter length and  $w_j|_{[j=1,2,\dots,f_t]}$  are the filter coefficients obtained as follows:

$$w_j = \frac{1}{\sum_j w_j} \exp(-j/\xi) \quad (25)$$

with filter parameter  $\xi = 5$  and filter length  $f_t = 20$ . Then we build the nonlinear Markov model on  $N = 500$  trading days. The parameters for modelling and prediction were set as follows:

- lag vector  $l = (1, 2, \dots, 10)$ , so that  $p_t = (y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}, y_{t-8}, y_{t-12}, y_{t-16}, y_{t-20}, y_{t-26}, y_{t-32}, y_{t-40})$ ;
- number of best fitting functions  $k = 50$ ;
- size of embedding matrix  $\nu = 180$ ;
- future window length  $n = 90$ .

## 6 Empirical results

Table 1 provides a summary of the descriptive statistics for the unconditional distribution of the GBP/USD spot exchange rates.

Table 1. Summary statistics of the GBP/USD Spot Exchange Rates (1/6/88 to 13/5/92).

Mean	-1.375e-5
Median	1.804e-4
$\sigma$	6.990e-3
Maximum( $\sigma$ )	4.093
Minimum( $\sigma$ )	-3.954
SR( $\sigma$ )	8.047
# Obs. > 2/3/4/5 $\sigma$	52/12/2/0
Skewness	-0.301*
Kurtosis	4.608*
JB	123.93*

\*Significant at the 5% level. SR is the Studentised Range and is calculated as (max-min)/ $\sigma$ . JB is the Jarque-Bera test statistic for normality of the returns, which follows a  $\chi^2$  distribution with two degrees of freedom.

The Jarque-Bera Lagrange Multiplier (LM) statistic indicates that the time series is not normally distributed. While the skewness of the returns distributions is small, the kurtosis is large, implying that much of the departure from normality is due to leptokurtosis.

Table 2 reports for the various time series the mean values of the parameter estimates of the AR(1)-GARCH(1,1) model, their standard deviations and their mean t-ratios.

Table 2. Mean (of 500 estimations) of the parameters estimates of the AR(1)-GARCH(1,1) model for GBP/USD Spot Exchange Rates (1/6/88 to 13/5/92).

Parameter	Estimate	(t-ratio)
$\mu$	3.566e-4	(1.254)
$\varphi$	0.112*	(2.315)
$\omega$	1.515e-6*	(1.818)
$\alpha$	0.082*	(3.187)
$\beta$	0.889*	(29.397)
<i>Diagn. (<math>\eta_t</math>):</i>		
Skewness	-0.374*	
Kurtosis	4.283*	
JB	47.87*	
Q(12)	12.37	
Q(12) <sup>2</sup>	13.35	

\*Significant at the 5% level. The robust t-ratios are those of Bollerslev and Wooldridge [1992], and are designed to be insensitive to non-normality, especially the presence of outliers. JB is the Jarque-Bera LM test statistics for normality of  $\eta_t^2$ , which is asymptotically  $\chi^2$  distributed with two degrees of freedom under the null hypothesis of normality. Q(12) is the Ljung-Box test statistic for serial correlation in  $\eta_t$  with 12 lags. Q(12)<sup>2</sup> is the Ljung-Box test statistic for an ARCH process based on  $\eta_t^2$ . Under the null hypothesis of uncorrelated and conditional homoscedasticity, respectively, both test statistics are asymptotically  $\chi^2$  distributed with 12 degrees of freedom.

The diagnostic tests indicate that no serious model misspecification is observed, but that the GARCH(1,1) model cannot account for all the leptokurtosis in the returns. Also, none of the parameter estimates violates the second and fourth moment regularity conditions. Hence, the model provides an adequate description of the data. The parameter estimates imply that the GBD/USD returns are significantly positively correlated and that, on average, there is a rather weak reaction of the conditional volatility to shocks (ARCH effect) but with a long-term memory (GARCH effect).

Table 3 reports the various forecast errors of the models.

Table 3. Mean of the one-step ahead forecast errors of the CVA and AR(1)-GARCH(1,1) model for GBP/USD Spot Exchange Rates (1/6/88 to 13/5/92)

	CVA		GARCH(1,1)	
ME	-1.31e-4	(4.83e-3)	1.58e-3	(4.88e-3)
MAE	3.63e-3	(2.98e-3)	4.17e-3	(3.19e-3)
MSE	2.33e-5	(4.97e-5)	2.63e-5	(4.32e-5)
MSE(+)	1.27e-5	(1.53e-5)	2.09e-5	(2.07e-5)
MSE(-)	3.83e-5	(7.24e-5)	4.04e-5	(7.33e-5)
RMSE	4.83e-3		5.12e-3	
MAPE	74.40	(52.04)	77.32	(54.78)
MWAPE	52.04	(94.76)	54.78	(76.52)
PTTEST	-8.09*		-1.29	
Over(%)	58.4		72.4	
R <sup>2</sup> (%)	3.77		2.10	

\* Significant at the 5% level. Values in parentheses are standard errors. Adjusted R<sup>2</sup> is the regression coefficient of determination, adjusted for the number of independent variables. Over(%) is the percentage of forecasts that overpredict realised volatility. MSE(+) and MSE(-) are the MSE measures for the positive and negative forecast errors, respectively. PTTEST is the Pesaran and Timmermann test statistic, which is asymptotically normally distributed.

Based on the MAE, MSE and RMSE measures, the CVA model provides significantly improved (up to 10%) forecasts relative to GARCH(1,1). Unlike GARCH(1,1), the CVA model is not highly biased. In particular, the CVA model overpredicts volatility less than 60% of the time, compared to more than 70% for GARCH(1,1).

The Pesaran and Timmermann test statistic, which computes a non-parametric association between the forecasted and realised volatility, implies that there is a greater association for CVA between forecasts and realised volatility. No significant association is found for the GARCH(1,1) forecasts.

The advantages of the greater flexibility of the CVA model is particularly evident when trying to model both large and small volatility shocks. When GARCH(1,1) is applied to data that include sudden and large shocks to volatility, the predicted conditional variance persists strongly and inaccurately. In contrast, the CVA model accurately models the much smaller persistence of large shocks to volatility. This is evident from the MSE measure for positive forecast errors, which is substantially smaller (more than 40%) for the CVA model than for GARCH(1,1).

A simple gauge of the predictive power of the model is the adjusted  $R^2$ , which is obtained by regressing the ex-post volatility on the forecasts. These regressions are based on observed absolute returns as a measure of the realised volatility. Table 3 shows that the adjusted  $R^2$  is substantially larger for the CVA model than for GARCH(1,1).

## 7 Conclusion

The focus of this paper has been to obtain models that accurately reflect the dynamics of the system. Thus, a model should not only fit the sample data and forecast well, but it should also have dynamical behaviour similar to that of the measured system. As applied to financial exchange rate time series, the algorithm presented captures the dynamics of a complex system and also gives reliable one-step ahead predictions for short data sets.

## 8 Acknowledgements

The authors wish to acknowledge the financial support of the Australian Research Council.

## 9 References

Blair, B., Poon, S.H. and S.J. Taylor, Asymmetry and crash effects in stock volatility for the S&P 100 index and its constituents, Working Paper, Lancaster University, 1998.

Bollerslev, T., Generalized Autoregressive Conditional Heteroskedasticity, *J. Econometrics*, 31, 307-327, 1986.

Bollerslev, T. and J.M. Wooldridge, Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances, *Econometric Review*, 11, 143-172, 1992.

Engle, R.F., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007, 1982.

Engle, R.F. and G.G.J. Lee, A permanent and transitory component model of stock return volatility, Discussion Paper 92-44R, University of California, San Diego, 1993.

Ghose, D. and K. F. Kroner, The relationship between GARCH and symmetric stable processes: finding the source of fat tails in financial data, *J. Empirical Finance*, 2(3), 225-251, 1995.

Hillmer, S.C, Bell, W.R. and G.C. Tiao, Modeling considerations in the seasonal adjustment of economic time series, in Applied time series analysis of economic data, Ed. A. Zellner, Washington, D.C., U.S. Bureau of the Census, 74-100, 1983.

Hotelling, H., Relations between two sets of variates, *Biometrika*, 28, 321-377, 1936.

Judd, K. and A.I. Mees, A model selection algorithm for nonlinear time series, *Physica D*, 82, 426-444, 1995.

Judd, K. and A.I. Mees, Embedding as a modelling problem, *Physica D*, 120, 273-286, 1998.

Koutmos, G., Asymmetries in the conditional mean and the conditional variance: evidence from nine stock markets, *J. Business & Economic Statistics*, 50, 277-290, 1998.

Larimore, W.E., Identification and filtering of nonlinear systems using canonical variate analysis, in Nonlinear Modeling and Forecasting, Eds. M. Casdagli and S. Eubank, Addison-Wesley, Reading, MA, 283-303, 1991.

Terasvirta, T. and H.M. Anderson, Characterising nonlinearities in business cycles using smooth transition autoregressive models, *J. Applied Econometrics*, 7, 119-139, 1992.