Non-Linear Markov Modelling Using Canonical Variate Analysis: Forecasting Exchange Rate Volatility

Berndt Pilgram[a], Peter Verhoeven[b], Alistair Mees[c], and Michael McAleer[d]

[a,c)]Department of Mathematics and Statistics, University of Western Australia, WA 6907

[b)]School of Business, Curtin University of Technology, WA 6845

[d)]Department of Economics, University of Western Australia, WA 6907

January 29, 2000

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 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 (Mandelbrot (1963)). In particular, absolute returns have significant positive serial correlation over long lags, implying that they have long term memory (Ding *et al*. (1993)). 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) (Black (1976), Christie (1982), Campbell and Hentschel (1990), Duffee (1995), and Blair *et al*. (1998), and Koutmos (1998)). This has commonly been called the 'leverage effect' because the decline in the firm's stock price will increase the debt to equity ratio.

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 empirical and theoretical appeal is due to the following: (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, its theoretical and empirical simplicity is also the main reason for its numerous limitations. For example, the GARCH model imposes a symmetrical influence of lagged squared residuals on current volatility, thereby failing to accommodate sign asymmetries. Moreover, high and low volatility shocks are imposed to have the same rate of persistence. Considering these shortcomings, numerous extensions have been suggested to the GARCH model in order to capture the many stylised facts of volatility. For example, the GARCH model has been extended and refined to include the asymmetric effects of positive and negative shocks to returns on volatility (such as the Exponential GARCH model (EGARCH) (Nelson (1990), the GJR-GARCH model (Glosten *et al*. (1993)), the threshold GARCH model (TGARCH) (Zakoian (1991)), the Asymmetric Power GARCH model (APGARCH) (Ding *et al*. (1993), and the Quadratic GARCH model (QGARCH) (Sentana (1995)). Also, regime switching GARCH models (Cai (1994); Hamilton and Susmel (1994); Kim and Kim (1996); and Susmel (1998)) have been developed that incorporate the different degrees of persistence of low-, moderate- and high-volatility regimes, and that does not attribute a large degree of persistence to the effects of extreme and outlying observations.

In this paper, we take a more general nonlinear non-parametric approach 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 gives 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. CVA theory was originally developed for independent and identically distributed (i.i.d.) random variables. However, we apply CVA to correlated vector time series which is discussed in detail by Larimore (1997).

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

$$y_t|_{[t=1,2,...N]}\ . \tag{1}$$

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 $l = (l_1, l_2, ..., l_m)$, a vector of positive integers, and obtain the past vector p_t as

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

Having obtained an embedding the dynamics the system can be described by

$$y_t = G(p_t) + \epsilon_t \tag{3}$$

with nonlinear function G and ϵ_t as the residual error. Judd and Mees (1995) describe an approach how nonlinear function G can be found. Once, the nonlinear function G is found, the future value $\hat{y}_{N+1}$ can be estimated. The Markov modelling approach extends this concept and we predict n steps ahead. Hence, the *future* vector f_t of finite window length n is introduced by

$$f_t = (y_t, y_{t+1}, ..., y_{t+n-1})^T. \tag{4}$$

Vector 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. Finite dimensional state s_t is approximated by an r-dimensional reduced memory vector m_t, given as a nonlinear function ϕ of the past, that is,

$$s_t \approx m_t = \phi(p_t). \tag{5}$$

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,

$$P(f_t|p_t) = P(f_t|s_t). \tag{6}$$

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 ϕ of the past p_t is chosen such that the optimal linear predictor $\hat{f}_t(m_t)$ minimizes the prediction error.

So far we defined the linear embedding of the time series y_t, *i.e.* the *past* vector p_t, but we have not yet introduced any nonlinear functions to approximate the future. Hence, we select a class of nonlinear functions $f_i|_{[i=1,2,...,k]}$, of the past p_t to obtain a set of basis functions π_t to approximate the future; that is,

$$\pi_t = (f_1(p_t), f_2(p_t), ..., f_k(p_t)) \tag{7}$$

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

$$f_i(p_t) = \Phi\left(\frac{|p_t - c_i|}{r_i}\right) \tag{8}$$

for suitably chosen centres c_i, radii r_i, and radial basis function Φ.

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,

$$y_t = \sum_{i=1}^{k} \lambda_i f_i(p_t) + \epsilon_t = \sum_{i=1}^{k} \lambda_i \Phi\left(\frac{|p_t - c_i|}{r_i}\right) + \epsilon_t \tag{9}$$

for some selection of nonlinear functions f_i, unknown parameters λ_i, unknown i.i.d. random variates ϵ_t, and a given number k. The choice of k is not critical. However, k has to be large enough to describe the data from the measured system sufficiently well, *i.e.* to guarantee a residual error ϵ_t lower than a pre-specified level. Then, the basis set, the functions $f_i|_{[i=1,2,...,k]}$, is obtained as a set of basis functions that approximates the data y_t. In the following, we use the set of functions $f_i|_{[i=1,2,...,k]}$ as a set of basis functions $\pi_t = (f_1, f_2, ..., f_k)$, given in Eq. **??**, to predict the future f_t.

Now we define the optimal prediction problem which is solved by a maximum likelihood system identification procedure (Larimore (1991)), as follows. We just give the results; details can be found in Larimore (1997). Assuming a linear relationship describing the optimal prediction of f_t from π_t, consider the following model

$$\begin{aligned} f_t &= Bm_t + e_t \\ m_t &= A\pi_t(p_t) = \phi(p_t) + e_t \end{aligned} \tag{10}$$

where memory m_t is an intermediate set of r variables that may be fewer in number than π_t. Vector e_t with covariance matrix Σ_{ee} is the error in the linear prediction of f_t from π_t given by matrices A and B. One may also predict the future f_t from the past p_t using Eqs. **??**; that is,

$$f_t = BA\pi_t(p_t) + e_t = C\pi_t(p_t) \tag{11}$$

where the rank of matrix $C = BA$ is given by $rank(C) \leq r$. Hence, when solving the prediction problem it is much easier to deal with matrices A and B with fixed dimension r than to deal with the constraint $rank(C) \leq r$.

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 Π the basis set π_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 $\Sigma_{\pi f} = \frac{1}{N}\Pi F^T$.

Matrices A and B will be determined by a maximum likelihood procedure and the CVA Theorem stated below provides the means of solving Eqs. **??** for the optimal A and B, given Σ_{ee}. We assume that p_t and f_t are normal random variables, jointly distributed with zero mean and covariance matrices $\Sigma_{\pi\pi}$, Σ_{ff}, and $\Sigma_{\pi f}$. A maximum likelihood estimator of A, B, and Σ_{ee} is naturally defined by the conditional likelihood function $p(F|\Pi; A, B, \Sigma_{ee})$ of the future F given the basis set Π. Maximum likelihood estimation (MLE) involves substituting Σ_{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

$$\begin{aligned} J\Sigma_{\pi\pi}J^T &= I_{r_{\pi\pi}} \\ L\Sigma_{ff}L^T &= I_{r_{ff}} \\ J\Sigma_{\pi f}L^T &= D = \mathrm{diag}(\gamma_1, \gamma_2, ..., \gamma_\mathrm{r}, 0, ..., 0) \end{aligned} \tag{12}$$

where $r_{\pi\pi} = \mathrm{rank}(\Sigma_{\pi\pi})$, $r_{ff} = \mathrm{rank}(\Sigma_{\mathrm{ff}})$, and γ_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 π_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 Σ_{ee}, and maximisation over A and B, we obtain the following estimates for A:

$$\hat{A} = (I_r \; 0)J \tag{13}$$

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

$$\hat{B} = (I_r \; 0)L \tag{14}$$

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

$$M = (I_r \; 0)J\Pi \tag{15}$$

or for instant time t:

$$m_t = (I_r \; 0)J\pi_t. \tag{16}$$

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 $c_1, c_2, ..., c_r$, where the optimal rank r is obtained from the number of dominant canonical correlations γ_i (Larimore (1991)). The number of dominant canonical correlations, *i.e.* the optimal rank r, is chosen as the one which gives the best in-sample one-step ahead predictions.

3 Implementation of forecasting

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

$$p_N = (y_N, y_{N-1}, ..., y_{N+1-m}) \tag{17}$$

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

$$\hat{f}_N = (\hat{y}_{N+1}, \hat{y}_{N+2}, ..., \hat{y}_{N+n}) \tag{18}$$

as

$$\hat{f}_N = BA\pi_N(p_N) = B\phi(p_N). \tag{19}$$

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 $y_t|_{[t=1,2,...N]}$, determine the optimal embedding of the past p_t, *i.e.* embedding dimension m and lag vector $l = (l_1, l_2, ..., l_m)$, 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 ι-times and we obtain centres c_i and radii r_i of the selected basis functions which form the nonlinear embedding matrices $\Pi_1, \Pi_2, \cdots, \Pi_\iota$. Finally, embedding matrix Π of size ν is obtained from the nonlinear embedding matrices Π_i, a constant term c and linear embedding matrix P, *i.e.*
$$\Pi = (c \; P \; \Pi_1 \; \Pi_2 \; \cdots \; \Pi_\iota). \tag{20}$$

3. Given a future window length n, generate the future matrix F.

4. Solve the following system
$$\begin{aligned} F &= BM + E \\ M &= A\Pi \end{aligned} \tag{21}$$
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 in-sample one-step ahead predictions.

5. Build the embedding vector π_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 AR(1)-GARCH(1,1) model

Consider the AR(1)-GARCH(1,1) model, where the conditional mean (or log-return) is given by

$$r_t = \mu + \varphi r_{t-1} + \varepsilon_t \tag{22}$$

where

$$\varepsilon_t = \eta_t \sqrt{h_t} \tag{23}$$

with $\varepsilon_t \sim N(0, h_t)$, $\eta_t \sim i.i.d.N(0,1)$, and the conditional variance of ε_t is given by

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}. \tag{24}$$

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 $\{\varepsilon_t\}$ (see Bollerslev (1986)). In general, the higher is the moment considered, the stronger is the condition and the less likely is it to be satisfied empirically. A sufficient condition for the second moment of $\{\varepsilon_t\}$ to exist is that $(\alpha + \beta) < 1$. If this condition is met, $\{\varepsilon_t, h_t\}$ is strictly stationary and ergodic.

Diebold (1988) showed that stationary models converge to normality, while non-stationary models do not converge to normality. Violation of the second-order stationarity condition does not necessarily imply non-stationarity of the process. If some weaker requirements (such as the log moment condition) are met, $\{\varepsilon_t, h_t\}$ may still be stationary even though $(\alpha + \beta)$ might be equal to or greater than unity, in which case $E(\varepsilon_t^2) = \infty$ (see Nelson (1990); Lee and Hansen (1994); Lumsdaine (1995)). For example, Nelson (1990) shows that when $\omega > 0$ and $h_t < \infty$, $\{\varepsilon_t, h_t\}$ is strictly stationary and ergodic if and only if $E[ln(\beta + \alpha \eta_t^2)] < 0$. A practical problem with this condition is that it is difficult to apply in practice because it is the mean value of a distribution of a random variable. A large number of simulations is typically required to obtain statistically significant values for η_t[1].

A sufficient condition for the existence of the fourth moment of $\{\varepsilon_t\}$ is $(k\alpha^2 + 2\alpha\beta + \beta^2) < 1$ (Bollerslev (1986)) [2], where k is the conditional fourth moment of η_t. Under the assumption of conditional normality, $k \equiv E(\eta_t^4) = 3$, so that the regularity condition becomes $(3\alpha^2 + 2\alpha\beta + \beta^2) < 1$. The assumption of normality is used to define the likelihood function, but is not necessary for the asymptotic results[3].

For estimation purposes, if normality is assumed when the true conditional density is not normal, the resulting maximum likelihood estimates (MLE) should be interpreted as quasi-maximum likelihood estimates (QMLE). Weiss (1986) and Bollerslev and Wooldridge (1992) show that, even in the presence of non-normality, the resulting QMLE are asymptotically normally distributed and consistent if the second and fourth moment conditions are satisfied. Ling and McAleer (1999c) show that efficient estimates for non-stationary ARMA models with GARCH errors can be constructed in the absence of knowledge of the conditional distribution through adaptive estimation.

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 for 1 June 1988 to 13 May 1992, obtained from the Pacific Exchange Rate Service.

[1] This holds because η_t is the true error rather than the estimated error for a given sample.

[2] He and Terasvirta (2000) provide a more detailed characterization of the fourth moment structure of the GARCH(p,q) process. Ling and McAleer (1999b) clarify the necessity and sufficiency of He and Terasvirta's fourth moment condition, and provide the necessary and sufficient conditions for all moments of the general GARCH process, as well as those of Ding *et al.'s* (1993) asymmetric power GARCH process.

[3] Terasvirta (1996) derived the unconditional fourth moment of GARCH(1,1) without the normality assumption.

Mean values of the parameter estimates, moment conditions and forecast errors were calculated using 500 one-day ahead volatility forecasts. The first 500 trading days were used to estimate the model, which yielded the one-day ahead forecasts for h_t. Then the estimation time interval was moved one-day ahead into the future by deleting the first trading day and adding an extra day at the end of the sample period. The parameters of the model were re-estimated and the one-day ahead forecasts re-generated. This procedure was repeated 500 times. In this paper, the following definition for realised volatility is used:

$$\sigma_t = \mid r_t - \bar{r} \mid \tag{25}$$

where the daily logarithmic returns are defined as $r_t = ln(\frac{P_t}{P_{t-1}})$, $\bar{r}$ is the conditional sample mean of r_t given the previous values $r_{t-k}, k \geq 1$, and P_t denotes price in period t.

We applied the nonlinear Markov modelling approach to the volatility sequence σ_t. To reduce the additive noise component[4], we pre-filtered the volatility series by using a linear filter with exponentially decreasing filter coefficients, that is,

$$y_t = \sum_{j=t-f_l+1}^{j=t} \sigma_j w_{j-t+f_l} \tag{26}$$

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

$$w_j = \frac{1}{\sum_j w_j} exp(-j/\xi) \tag{27}$$

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

- lag vector $l = (1, 2, \cdots, 10)$, 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 $k = 50$;

[4] Andersen and Bollerslev (1998) acknowledge that, while absolute (or squared) daily returns provide unbiased estimates of the underlying unobservable volatility, they are very noisy estimators of daily movements in volatility due to the large idiosyncratic error term.

- dimension ν of embedding matrix Π, *i.e.* $\nu = 180$;
- 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
σ	6.990e-3
Maximum(σ)	4.093
Minimum(σ)	-3.954
SR(σ)	8.047
# +ve observations $> 1/2/3/4/5\sigma$	125/17/4/0/0
# -ve observations $> 1/2/3/4/5\sigma$	141/35/8/2/0
Skewness	-0.301*
Kurtosis	4.608*
LM(N)	123.93*

*Significant at the 5% level. SR(σ) is the Studentised Range of (σ) and is calculated as (max(σ)-min(σ)). LM(N) is the Jarque-Bera Lagrange multiplier test statistic for normality of the returns, which is asymptotically χ^2 distribution with two degrees of freedom under the null hypothesis of normality.

The Jarque-Bera Lagrange multiplier (LM(N)) test 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 values of 500 estimates of the parameters estimates of the AR(1)-GARCH(1,1) model for GBP/USD Spot Exchange Rates (1 June 1988 to 13 May 1992).

Parameter	Estimate	(std)	[t-ratio]
μ	3.566e-4	(8.924e-5)	[1.254]
φ	0.112	(0.037)	[2.315]
ω	1.515e-6	(5.484e-7)	[1.818]
α	0.082	(0.017)	[3.187]
β	0.889	(0.025)	[29.397]
$(\alpha + \beta)$	0.971	(0.010)	
$(3\alpha^2 + 2\alpha\beta + \beta^2)$	0.956	(0.017)	
Diagnostics			
Mean	-0.032	(0.009)	
Std	1.000	(0.004)	
Skewness	-0.37	(0.13)	
Kurtosis	4.28	(0.18)	
LM(N)	47.87	(13.03)	
Q(12)	12.37	(2.16)	
$Q(12)^2$	13.35	(15.19)	

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 η_t^2, which is asymptotically χ^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 η_t with 12 lags. $Q(12)^2$ is the Ljung-Box test statistic for an ARCH process based on η_t^2. Under the null hypothesis of uncorrelated and conditionally homoskedastic errors, respectively, the test statistics are asymptotically χ^2 distributed with 12 degrees of freedom.

The diagnostic tests indicate that there are no serious model misspecifications, but that the GARCH(1,1) model cannot account for the skewness or all of the kurtosis 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 GBP/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. Forecast errors of the CVA and AR(1)-GARCH(1,1) model for GBP/USD Spot Exchange Rates (1 June 1988 to 13 May 1992)

	total (500)		low volatility (433)		high volatility (67)	
	CVA	GARCH	CVA	GARCH	CVA	GARCH
ME	-1.147e-4	1.60e-3	1.231e-3	2.992e-3	-8.815e-3	-7.400e-3
MAE	3.578e-3	4.117e-3	2.768e-3	7.400e-3	8.815e-3	7.400e-3
RMSE	4.788e-3	5.056e-3	3.368e-3	4.259e-3	9.887e-3	8.578e-3
RMedSE	2.782e-3	3.750e-3	2.411e-3	3.417e-3	8.050e-3	6.062e-3
RMSE(+)	3.512e-3	4.539e-3	3.512e-3	4.539e-3	0.00	0.00
RMSE(-)	6.197e-3	6.222e-3	3.028e-3	2.314e-3	9.887e-3	8.578e-3
SMAPE	72.87	76.98	70.99	78.87	84.98	64.75
SMWAPE	62.28	53.12	46.85	43.49	88.96	69.76
PTTEST	-6.50	-8.63	-4.99	-4.79	0.98	0.20
Over(%)	59.4	72.6	68.6	83.8	0.0	0.0
R^2(%)	4.76	3.61	2.65	3.15	5.35	6.66

R^2 is the coefficient of determination by regressing the ex-post volatility on the forecast volatility. Over(%) is the percentage of forecasts that overpredict realised volatility. RMSE(+) and RMSE(-) are the RMSE measures for the positive and negative forecast errors, respectively. PTTEST is the Pesaran and Timmermann test statistic, which is asymptotically normally distributed. The loss functions are defined as follows:

ME $= \frac{1}{N}\sum_{t=1}^{N}(\sqrt{h_t}-\sigma_t)$, RMSE $= \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\sqrt{h_t}-\sigma_t)^2}$, MAE $= \frac{1}{N}\sum_{t=1}^{N} \mid \sqrt{h_t}-\sigma_t \mid$,
SMAPE $= \frac{100}{N}\sum_{t=1}^{N}(\frac{|\sqrt{h_t}-\sigma_t|}{0.5(\sigma_t+\sqrt{h_t})})$, SMWAPE $= \frac{100}{N}\sum_{t=1}^{N}(\frac{\sigma_t}{\bar{\sigma}}\frac{|\sqrt{h_t}-\sigma_t|}{0.5(\sigma_t+\sqrt{h_t})})$.

Based on the MAE, RMSE, and RMedSE measures calculated over the entire sample, the CVA model provides significantly improved (up to 25%) forecasts relative to GARCH(1,1). Unlike the GARCH(1,1) model, 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 the GARCH(1,1) model.

The Pesaran and Timmermann test statistic (PTTEST), which computes a non-parametric association between the forecasted and realised volatility, implies that there is a strong association between both the CVA and GARCH(1,1) forecasted volatility and the realised volatility.

When the sample is split into low and high volatility periods, substantially reduced mean forecast errors are observed only for low volatility, whereas the mean forecast errors of the CVA model are substantially larger for high volatility. For example, base on MAE, RMSE and RMedSE measures, the CVA model provides up to 63% lower forecast errors for low volatility periods compared to up to 30% worse forecast errors for high volatility.

7 Discussion

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.

The CVA model might be advantageous 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 RMSE measure for positive forecast errors, which is substantially smaller (more than 40%) for the CVA model than for GARCH(1,1). However, the forecast ability of the CVA model is lower for periods of high volatility. This might be due to the effects of the filtering applied which substantially reduces the value of extreme and outlying observations. Furthermore, it is possible that there is some degree of overfitting with the current version of the method. This is because it is difficult to estimate the optimal model order for this new and relatively complex approach.

8 Acknowledgements

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

References

Andersen, T.G., and T. Bollerslev, Answering the skeptics: yes, standard volatility models do provide accurate forecasts, International Economic Review, 39, 885-905, 1998.

Black, F., Studies in stock price volatility changes, Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Association, American Statistical Association, 177-181, 1976.

Blair, B., S.H. Poon, 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, Journal of 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.

Cai, J., A Markov model of switching-regime ARCH, Journal of Business & Economic Statistics, 12, 309-316, 1994.

Campbell, J., and L. Hentschel, No news is good news: an asymmetric model of changing volatility in stock returns, International Conference on ARCH models, Paris, June 25-26, 1990.

Christie, A.A., The stochastic behaviour of common stock variances: value, leverage and interest rate effects, Journal of Financial Economics, 10, 407-432, 1982.

Diebold, F.X., Empirical modelling of exchange rate dynamics., Springer Verlag, New York, 1988.

Ding, Z., R.F. Engle, and C.W.J. Granger, A long memory property of stock market returns and a new model, Journal of Empirical Finance, 1, 83-106, 1993.

Duffee, G.R., Stock returns and volatility: a firm level analysis, Journal of Financial economics, 37, 399-420, 1995.

Engle, R.F., Autoregressive conditional heteroskedasticity 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, Journal of Finance, 2, 225-251, 1995.

Glosten, L.R., Jagannathan, R. and D. runkle, On the relation between the expected value and the volatility of the nominal excess return on stocks, Journal of Finance, 48, 1779-1801, 1993.

Hamilton, J.D., and R. Susmel, Autoregressive conditional heteroskedasticity and changes in regime, Journal of Econometrics, 64, 307-333, 1994.

He, C., and T. Terasvirta, Fourth moment structure of the GARCH(p,q) process, to appear in Econometric Theory, 16, 2000.

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.

Kim, C.J., and M.J. Kim, Transient fads and the crash of '87, Journal of Applied Econometrics, 11, 41-58, 1996.

Koutmos, G., Asymmetries in the conditional mean and the conditional variance: evidence from nine stock markets, Journal of Economics and Business, 50, 277-290, 1998.

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

Larimore, W.E., Canonical variate analysis in control and signal processing, Statistical Methods in Control and Signal Processing, Eds. Katayama, T., and S. Sugimoto, Marcel Dekker, New York, pp 83-120, 1997a.

Larimore, W.E., Optimal reduced rank modeling, prediction, monitoring, and control using canonical variate analysis, Proceedings of the IFAC 1997 International Symposium on Advanced Control of Chemical Processes, Banff, Canada, pp 61-66, 1997b.

Lee, S.W., and B.E. Hansen, Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator, Econometric Theory, 10, 29-52, 1994.

Ling, S., and M. McAleer, Non-nested tests of GARCH and E-GARCH, Working Paper, Department of Economics, University of Western Australia, 1999a.

Ling, S., and M. McAleer, Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models, 1999b. Submitted.

Ling, S., and M. McAleer, On adaptive estimation in nonstationary ARMA models with GARCH errors, 1999c. Submitted.

Lumsdaine, R.L., Finite-sample properties of the maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models: a Monte-Carlo investigation, Journal of Business & Economic Statistics, 13, 1-10, 1995.

Mandelbrot, B., The variation of certain speculative prices, Journal of Business, 36, 394-419, 1963.

Nelson, D.B., Stationarity and persistence in the GARCH(1,1) model, Journal of Econometrics, 45, 7-38, 1990.

Sentana, E.,Quadratic ARCH models: a potential reinterpretation of ARCH models as second-order Taylor approximations, Unpublished paper, London School of Economics, 1991.

Susmel, R., Switching volatility in Latin American emerging equity markets, Emerging Market Quarterly, 2, 44-56, 1998.

Terasvirta, T., Two stylized facts and the GARCH(1,1) model, Working Paper Series in Economics and Finance No. 96, Stockholm School of Economics, 1996.

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

Weiss, A.A., Asymptotic theory for ARCH models: estimation and testing, Econometric Theory, 2, 107-131, 1986.

Zakoian, J.M., Threshold heteroskedastic models, Journal of Economic Dynamics and Control, 18, 931-955, 1994.