Testing for ARCH in the presence of nonlinearity of unknown form in the conditional mean: The case of the exchange rate[1]

Andrew P. Blake[2]
CCBS, Bank of England

George Kapetanios[3]
Queen Mary, University of London

June 2005

[1]This paper contains the opinions and analysis of the authors and should not be thought to reflect the views of the Bank of England or of the Monetary Policy Committee. We are grateful to an editor and two referees for comments which have both improved the focus and expanded the scope of the paper.
[2]Email: Andrew.Blake@bankofengland.co.uk.
[3]Email: G.Kapetanios@qmul.ac.uk

Abstract

Tests of ARCH are a routine diagnostic in empirical econometric and financial analysis. However, it is well known that misspecification of the conditional mean may lead to spurious rejection of the null hypothesis of no ARCH. Nonlinearity is a prime example of this phenomenon. There is little work on the extent of the effect of neglected nonlinearity on the properties of ARCH tests. We investigate this using new ARCH testing procedures that are robust to the presence of neglected nonlinearity. Monte Carlo evidence shows that the problem is serious and that the new methods alleviate this problem to a very large extent. We apply the new tests to exchange rate data and find substantial evidence of spurious rejection of the null hypothesis of no ARCH. This is further evidence that exchange rates exhibit complicated dynamic behaviour, with important nonlinearity and volatility effects.

Keywords: Nonlinearity; ARCH; Neural networks; Exchange rate

JEL Classification: C12; C22; C45

1 Introduction

Since the introduction of the autoregressive conditional heteroscedasticity (ARCH) model by Engle (1982) testing for the presence of ARCH has become a routine diagnostic in the econometric analysis of macroeconomic and especially financial time series. Although a variety of different tests have been developed the most widely applied one is based on an autoregression of the squared residual on a constant and its p lags whereby the joint significance of all the lags included is tested. A rejection leads to the conclusion that ARCH is present.

As observed by a number of authors, tests for ARCH may also reject the null if misspecification of the conditional mean of the model is present. Notable cases include work by Bera, Higgins, and Lee (1992), Bera and Higgins (1997) and Lumsdaine and Ng (1999). Lumsdaine and Ng (1999) suggest procedures based on recursive residuals that may alleviate misspecifications in the conditional mean and thereby reduce the potential for falsely rejecting the null of no ARCH when other forms of misspecification are present in the model. Bera and Higgins (1997) observe that bilinear and ARCH models have a similar moment structure raising the possibility that bilinear and ARCH processes may be confused in practical applications.[1]

The current paper follows on from this literature. In particular we suggest that other forms of nonlinearity in the conditional mean may be causing false rejection in testing for ARCH. Examples include nonlinearities of the smooth transition autoregressive (STAR) form or the self-exciting threshold autoregressive (SETAR) form. This observation in itself is not novel. However, there is little work that investigates the interplay of the degree of nonlinearity with the degree of rejection of the no-ARCH null hypothesis.

[1] It should be noted that although ARCH and Bilinear models may have similarities, their unconditional moment structures are potentially quite different. This makes the relationship between these classes of models more complex than usually acknowledged in the literature. Note that the results of Bera and Higgins (1997) mainly relate to bilinear models with one lag and do not readily generalise to higher lag orders.

Taking this as a starting point we suggest new testing procedures that enable valid detection of ARCH in the presence of such nonlinearity. The approach we take is based on neural networks. Neural networks are a flexible form of nonlinear model that are able to approximate nonlinearities of unknown form (up to the choice of variables that enter the nonlinear conditional mean function) arbitrarily well and thereby produce well behaved residuals under the null of no ARCH on which standard ARCH tests may then be carried out. We use a neural network to purge the residuals of the effects of nonlinearity before applying an ARCH test. Thus, we intend to correctly size the ARCH test while retaining good power for the ARCH test in the case where ARCH effects are actually present.

However, obtaining the residuals from what are in effect neural network models is not straightforward. Estimation using a standard neural network specification based on the logistic function involves nonlinear least squares. There are additional identification problems if the true DGP happens to be linear. Lee, White, and Granger (1993) circumvented the identification problem by randomly generating the coefficients of an arbitrary neural network. In this paper we avoid the need for complicated estimation techniques by using two alternative strategies that also resolve potential identification issues. The first uses an alternative to the logistic neural network model, the radial basis function (RBF) artificial neural network model. This type of neural network has been used in the econometric literature to test for neglected nonlinearity (see Blake and Kapetanios (2003b)). Estimation of such models can be carried out using ordinary least squares. The second strategy uses polynomial approximations of the logistic neural network specification following the work of Teräsvirta, Lin, and Granger (1993). Again, this leads to a model that can be estimated using ordinary least squares.

This is not the first time neural networks have been used for ARCH testing. Peguin-Feissolle (1999) and Blake and Kapetanios (2000) develop ARCH tests where the form of ARCH is unknown and allowed to be quite general.

Thus the potentially nonlinear process driving the ARCH part of the model is detected using the neural network with the intention to increase the power of the ARCH test. For the present paper, the problem addressed is different. The nonlinearity that matters is (potentially) in the rest of the model, and in particular in the conditional mean. This nonlinearity is removed from the conditional mean using neural networks prior to applying an ARCH test on the residuals. The test for ARCH applied as part of our procedure can be of any form. Indeed, we use the Peguin-Feissolle (1999) test as one possible procedure. Given the different possible neural network specifications and applicable ARCH tests we investigate a number of potential test procedures using Monte Carlo experiments.

The plan of the paper is as follows. Section 2 gives more details about the nature of the problem we consider by drawing on previous work in the area. Section 3 discusses the new testing procedures we suggest. Section 4 provides a Monte Carlo investigation of the new and existing testing procedures that reveals the extent of the problem and the ability of the new tests to deal with it. Although using the new tests is at the cost of slightly reduced power for the detection of ARCH when it actually is present, it is offset by the enormous gains in producing properly sized tests. In Section 5 we apply the new tests to exchange-rate data and conclude that ARCH is absent for a substantial number of cases where standard tests find otherwise. This has important implications for the determination of exchange-rate dynamics. It would be easy to conclude that pure volatility effects accounted for a substantial part of changes in exchange rates, whereas it may be that nonlinear, threshold effects are the determining factor. Section 6 concludes.

2 Testing for ARCH under conditional mean nonlinearity

We concentrate on the following univariate model for the series y_t, $t = 1, \ldots, T$:

$$y_t = f(x_{1,t}, \ldots, x_{p,t};\ \alpha) + \epsilon_t \tag{1}$$

where $f(\cdot;\cdot)$ is an unknown function. We make standard assumptions about ϵ_t: It is mean zero, with variance conditional on a Borel measurable σ-field[2] with respect to the exogenous (or predetermined at time t) variables $x_{1,t}, x_{2,t} \ldots, x_{p,t}$ denoted by $\mathcal{I}_{t-1}$, given by:

$$h_t = \beta_0 + \sum_{j=1}^{q} \beta_j \epsilon_{t-j}^2 \tag{2}$$

and with unconditional variance given by σ^2. The variables $x_{j,t}$ are assumed to be stationary and ergodic with finite second moments. This setup encompasses both linear and nonlinear autoregressive stationary models. The null hypothesis of interest is:

$$H_0 : \beta_1 = \ldots = \beta_q = 0. \tag{3}$$

If one assumes linearity then the conditional mean model becomes:

$$y_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{j,t} + \epsilon_t. \tag{4}$$

The estimated OLS residual is given by:

$$\hat{\epsilon}_t = \epsilon_t + \left(f(x_{1,t}, \ldots, x_{p,t}; \alpha) - \hat{\alpha}_0 - \sum_{j=1}^{p} \hat{\alpha}_j x_{j,t} \right) = \epsilon_t + \chi_t \tag{5}$$

[2]Details on σ-fields and their use may be found in (Davidson, 1994, Ch. 8).

where $\chi_t = f(x_{1,t}, \ldots, x_{p,t}; \alpha) - \hat{\alpha}_0 - \sum_{j=1}^{p} \hat{\alpha}_j x_{j,t}$, with the squared residual given by:

$$\epsilon_t^2 + \left(f(x_{1,t}, \ldots, x_{p,t}; \alpha) - \hat{\alpha}_0 - \sum_{j=1}^{p} \hat{\alpha}_j x_{j,t} \right)^2 + 2\epsilon_t \left(f(x_{1,t}, \ldots, x_{p,t}; \alpha) - \hat{\alpha}_0 - \sum_{j=1}^{p} \hat{\alpha}_j x_{j,t} \right) = \epsilon_t^2 + \psi_t \quad (6)$$

where we similarly define $\psi_t = \chi_t^2 + 2\epsilon_t \chi_t$. The expectation of both (5) and (6) conditional on $\mathcal{I}_{t-1}$ is clearly not constant even if the null hypothesis holds. Of course, this analysis holds for any misspecified conditional mean function and the root of the false rejection in ARCH tests, under the null hypothesis, lies in the presence of serial correlation in ψ_t as Lumsdaine and Ng (1999) have observed. Our conjecture is that nonlinearity in χ_t may induce large probabilities of rejection of the null hypothesis of ARCH tests, under the null hypothesis, compared with linear misspecification arising by, say, the omission of an extra lag for the conditional mean model. Clearly, this probability will depend on many things, chief among which is the actual values of the parameters of the model and the nonlinearity considered. Therefore, any rigorous theoretical demonstration is bound to be of limited use.

A heuristic argument for our conjecture may be given as follows: Assume two cases of conditional mean misspecification. In the first case the conditional mean is assumed to be constant when the true conditional mean model is an AR(1) model. In the second case the assumed model is an AR(1) model but the true model is a SETAR model of the form:

$$y_t = \gamma_0 + \gamma_1 I(y_{t-1} < r) y_{t-1} + \gamma_2 I(y_{t-1} \geq r) y_{t-1} + \epsilon_t \quad (7)$$

where $I(\cdot)$ is the indicator function. In the first case:

$$\chi_t = (\hat{\alpha}_0 - \alpha_0) + \alpha_1 y_{t-1}. \quad (8)$$

In the second case:

$$\begin{aligned}\chi_t &= (\hat{\gamma}_0 - \gamma_0) + \tilde{\gamma}_1 I(y_{t-1} < r) y_{t-1} + \tilde{\gamma}_2 I(y_{t-1} \geq r) y_{t-1} \\ &= (\hat{\gamma}_0 - \gamma_0) + \gamma_{t-1} y_{t-1} \end{aligned} \tag{9}$$

where $\tilde{\gamma}_i = \hat{\alpha}_1 - \gamma_i$, $i = 1, 2$ and:

$$\gamma_{t-1} = \tilde{\gamma}_1 I(y_{t-1} < r) + \tilde{\gamma}_2 I(y_{t-1} \geq r). \tag{10}$$

In the first case variation in the conditional variance of the residual, under the null of no ARCH, comes from y_{t-1} only. In the second case both γ_{t-1} and y_{t-1} contribute to the variation.

This example serves to illustrate that neglected nonlinearity may induce quite complicated variation in the conditional variance from very simple models. This, and other forms of nonlinearity, may possibly lead to acute problems of overrejection for standard ARCH tests. We indeed find this to be the case with a variety of nonlinear models in Section 4. It is then reasonable to suggest that methods for accounting for general forms of nonlinearity prior to applying ARCH tests may be useful. The next section suggests such methods.

3 Nonlinearity robust ARCH tests

Following on from the previous section it is clear that nonlinearity has the potential to introduce problems in the detection of ARCH for dynamic models. As the form of nonlinearity is usually unknown we require a test for ARCH that is robust to as large a class of nonlinearities as possible. Artificial neural network models provide an ideal framework for this analysis. This is due to the useful property that they can approximate continuous functions arbitrarily well. More specifically, a continuous function $f(z)$ can be arbitrarily well approximated in the supremum norm by $\sum_{i=1}^{q} b_i g(z_i')$ for $z_i' = d_{0,i} + d_i' z$ if either (i) $g(\cdot)$ is sigmoidal, i.e. $g(\cdot)$ is non-decreasing with

$\lim_{z \to -\infty} g(z) = 0$ and $\lim_{z \to \infty} g(z) = 1$ or (ii) $g(\cdot)$ has non-zero expectation and is L_p bounded for some $p \geq 1$. For more details see Hornik, Stinchcombe, and White (1989), Stinchcombe and White (1989) and Cybenko (1989). A good general reference on the theoretical properties of neural networks is White (1992).[3] Note that the above results are only a small subset of the available results on the approximation capabilities of neural networks. The universal approximator properties of neural networks have been put to effect in the econometrics literature by, among others, Lee, White, and Granger (1993) and Blake and Kapetanios (2003b) to construct tests for neglected nonlinearity in stationary models, and by Blake and Kapetanios (2003a) to test for unit roots against stationary nonlinear processes.

In the present context we wish to robustify ARCH tests by fitting a neural network model to (1), obtaining the residuals and carrying out a standard ARCH test. By the universal approximation property of the neural networks for continuous functions discussed above we know that the model (1) may be written as:

$$y_t = a_0 + a_1' x_t + \sum_{i=1}^{q} b_i g(d_{0,i} + d_i' x_t) + \epsilon_t \tag{11}$$

where $x_t = (x_{1,t}, \ldots, x_{p,t})'$ with $g(\cdot)$ now a known function. This specification can then be estimated consistently and the residuals then tested for ARCH using standard tests since $\hat{\epsilon}_t$ from the estimation of (11) will converge in probability to ϵ_t. However estimation of (11), although feasible, may not be easy as it requires nonlinear least squares for the function used for $g(\cdot)$. For testing, however, this may not be necessary. There are neural network specifications or approximations of neural networks where such iterative schemes are not needed for estimation. We consider two.

The first is a radial basis function (RBF) artificial neural network, which is often referred to as a linear network. These are simple to define. Let

[3] See also Campbell, Lo, and MacKinlay (1997) for an excellent introduction to artificial neural networks, which covers the RBF networks that we use below. Bishop (1995) gives a thorough and very readable account.

there be q centers, denoted c_j, and a radius vector τ. Utilise a function that is monotonically decreasing about c_j. A natural choice is the Gaussian function, and the appropriate RBF is:

$$g([c_j'\ \tau']',\ x_t) = e^{-\|(x_t - c_j)/\tau\|^2} \tag{12}$$

where the division notation denotes element by element division. By the monotonicity property, each RBF has maximum activation (of unity) when the input vector coincides with the jth center independent of τ. Conversely, if the input vector is far enough away from the center (controlled by τ) the activation is almost zero.[4] The linearity of this network derives from the property that if the centers (c_j) and radii (τ) are determined by some procedure then the RBFs can be straightforwardly used as additional regressors. The trick is to use data-based procedures to determine both. It is convenient to use a simple function of the data, such as a multiple of the maximum change from period t to period $t+1$ for all t of each input as the radius for that input. In fact, we follow the suggestion by Orr (1995) and set this multiple to two. Note that in Blake and Kapetanios (2003b), we considered the alternative of a unit variance for the normalised data and found that it worked comparably to our choice in this paper. Thus, the performance of the method does not seem to be very sensitive to the choice of τ. Following Orr (1995), we then allow there to be T potential centers (the c_j) for the RBFs which is each of the vector of observations, i.e., x_t, $t = 1, \dots, T$. The T RBFs thus obtained are ranked according to their ability to reduce the unexplained variance of (11) when entered individually. Then, we successively add the sorted RBFs into (11) until we minimize an information criterion, chosen to be AIC. This procedure is known as forward selection. Note that we sort the RBFs only once at the beginning of the procedure. An alternative might be

[4]Other functional forms (such as the multiquadratic) have the same property and can be used instead, although in experiments Blake and Kapetanios (2003b) found very little difference in performance between various functions for the related problem of nonlinearity testing.

to sort the RBFs every time a new RBF is added to (11). For every added RBF, the remaining RBFs would be sorted according to their ability to reduce the unexplained variance of (11) when all chosen RBFs are included in the specification. However, our simulations, presented in the next section, suggest that our chosen method performs well and avoids the increased computational cost of repeated sortings.

The pair of data-based procedures (normalisation to set the radii and center choice from the data points by forward selection) yield a linear estimation problem for b as all the terms in (12) are defined by the process. In summary, the algorithm for the RBF specification is as follows:

1. We construct T initial RBF terms given by:
$$g([x_1' \ \tau']', \ x_t), \ldots, g([x_T' \ \tau']', \ x_t).$$
2. These are ranked according to their ability to reduce the residual variance, when entered individually in (11).
3. The AIC information criterion is used to determine how many of the T sorted RBF terms will eventually enter (11).

A second approach that is amenable to linear estimation is motivated by a test of neglected nonlinearity that approximates the logistic neural network developed by Teräsvirta, Lin, and Granger (1993). The logistic network uses:

$$g(d, \ x_t) = 1/(1 + e^{-d'x_t}) \tag{13}$$

for the activation function. As remarked above, to estimate the complete nonlinear model (all the parameters a, b and d) is costly. Teräsvirta, Lin, and Granger (1993) suggest a third-order Taylor expansion of the logistic neural network is used instead. This choice of expansion is arbitrary, and can clearly be replaced by the n-th order Taylor expansion of the logistic neural network. This turns out to be an n-th order polynomial in $(x_{1,t}, \ldots, x_{p,t})$, where the operational order of the Taylor expansion needs to be chosen. In common with our treatment of the RBF network we use an information criterion, again AIC. This is chosen from expansions of order two, three and four where for simplicity for the third and fourth order Taylor expansions only cross products of (powers of) up to two variables are considered. This approach is based on an approximation of a neural network specification rather than on a neural network specification. As Teräsvirta, Lin, and Granger (1993) and Blake and Kapetanios (2003b) show through simulations, this method is very effective in practice.

Note that there is a problem with the above methods since the model we estimate is not necessarily appropriate for a linear specification. Indeed, for a linear model the additional terms should be absent. This identification problem was solved by Lee, White, and Granger (1993) by adding an arbitrary fixed number of randomly generated logistic functions as regressors.[5] Our two procedures solve it by the use of the information criterion, which is restricted to pick a minimum of one additional regressor. Our test for ARCH, of course, uses the residuals from this model. Note, that unlike nonlinearity tests, identification is not as important in our context where the neural network construction is only used as a first step to remove nonlinearities in the conditional mean. However, the above approaches are computationally cheaper compared to the application of nonlinear least squares to estimate (11).[6]

[5] Considering the complete arbitrariness of this as a procedure, this is extraordinarily effective. It does lead to the possibility that the 'neurons' have zero or unit activation over the entire sample and hence are colinear, so this is avoided by using selected principle components of the output of the neural network. Note that an advantage to RBF approach is that centers would have to be the same for a similar problem to arise, which for our implementation would mean that regressors had the same value for two different periods. This is both unlikely (never having happened in our simulation experiments) and easily dealt with by deleting one of the repeated centers.

[6] The RBF procedure we suggest, takes less than a second for a sample of 100 observations on a 1Ghz personal computer compared to more than a minute for nonlinear least squares leaving aside problems of convergence of the algorithm, especially when no nonlinearity is present in the data.

After fitting the neural network model, or its approximation, the residuals obtained can be tested using any residual-based ARCH test. The asymptotic distribution of the usual LM ARCH test is still χ^2_q under the null hypothesis and assuming no nonlinearity. Further refinement of the new tests may be contemplated if the techniques suggested by Lumsdaine and Ng (1999) are combined with the neural network specifications. In particular if, as suggested by Lumsdaine and Ng (1999), (functions of) lags of recursive residuals from the original linear specification in (1), help in picking up misspecifications in the conditional mean then augmenting the variable set x_t by (functions of) this lagged recursive residual may further improve the performance of the tests under the null hypothesis. Additionally, OLS residuals may be used in place of recursive residuals. In small samples these will introduce biases since they depend on the whole of the sample via the parameter estimates but asymptotically these parameter estimates will converge to some limit and the lagged OLS residuals will not cause any further asymptotic biases.

4 Monte Carlo experiments

In this section we carry out Monte Carlo experiments to illustrate the problem arising out of pronounced nonlinearity in the context of ARCH testing and the extent to which the new tests alleviate the problem. The issue occurs under the null hypothesis and therefore the size performance of the existing tests is under scrutiny. Nevertheless, it is important that the new methods do not reduce the power of the ARCH tests unduly. Therefore, there is the usual tradeoff between power and size which needs to be explored.

We concentrate on nonlinear autoregressive models and add a linear AR(1) model given by $y_t = \alpha y_{t-1} + \epsilon_t$ for comparison. We consider three different classes of nonlinear models under the null hypothesis. These are

SETAR, STAR and bilinear models. The models are given by:

$$y_t = \gamma_1 I(y_{t-1} < r)y_{t-1} + \gamma_2 I(y_{t-1} \geq r)y_{t-1} + \epsilon_t \tag{14}$$

$$y_t = \delta_1 y_{t-1} + \delta_2(1 - e^{-\delta_3 y_{t-1}^2})y_{t-1} + \epsilon_t \tag{15}$$

and:

$$y_t = \zeta \epsilon_{t-1} y_{t-1} + \epsilon_t \tag{16}$$

respectively. Under the null hypothesis of no ARCH both the conditional and unconditional variance of ϵ_t is equal to σ^2 which is set to 1 for all size experiments. We consider one DGP from the AR class, four DGPs from SETAR models, four DGPs from STAR models and two DGPs from bilinear models. The coefficients for each class are:

- AR Model
  - Experiment 1: $\alpha = 0.5$.
- SETAR Models
  - Experiment 2: $\gamma_1 = 0.1$, $\gamma_2 = 0.5$, $r = 0$.
  - Experiment 3: $\gamma_1 = -0.1$, $\gamma_2 = 0.5$, $r = 0$.
  - Experiment 4: $\gamma_1 = -0.3$, $\gamma_2 = 0.9$, $r = 0$.
  - Experiment 5: $\gamma_1 = -0.8$, $\gamma_2 = 0.9$, $r = 0$.
- STAR Models
  - Experiment 7: $\delta_1 = 0.5$, $\delta_2 = -0.5$, $\delta_3 = 0.05$.
  - Experiment 8: $\delta_1 = 0.5$, $\delta_2 = -0.5$, $\delta_3 = 0.5$.
  - Experiment 9: $\delta_1 = 0.9$, $\delta_2 = -1.4$, $\delta_3 = 0.05$.
  - Experiment 10: $\delta_1 = 0.9$, $\delta_2 = -1.4$, $\delta_3 = 0.5$.
- Bilinear Models

- Experiment 12: $\zeta = 0.1$.
- Experiment 13: $\zeta = 0.3$.

Experiments 6 and 11 are described below. The above experiments are based entirely on artificial models. Instead of relying on models based on simple data generation processes, we have investigated additional empirically based SETAR and STAR data generation processes. This provides a further means for checking whether the proposed tests are empirically useful. More specifically, we have considered two well known nonlinearity studies. The first is Potter (1995) on SETAR models for the business cycle and the second is Sarantis (1999) on real exchange rates.

The SETAR model is taken from Table II of Potter (1995) and is parameterised as:

$$\begin{aligned} y_t = {} & (-0.705 + 0.510y_{t-1} - 0.849y_{t-2} - 0.048y_{t-3} \\ & - 0.123y_{t-4} + 0.398y_{t-5})I_{\{y_{t-2}\leq 0\}} \\ & + \ (0.545 + 0.312y_{t-1} + 0.245y_{t-2} - 0.104y_{t-3} \\ & - 0.057y_{t-4} - 0.940y_{t-5})I_{\{y_{t-2}>0\}} + \epsilon_t \end{aligned} \tag{17}$$

which we include as experiment 6. y_t is the first difference of the logarithm of quarterly US GNP at time t. The model is estimated during the period 1948Q3-1990Q4. The STAR model is taken from Table 4 of Sarantis (1999) and is a model for Canada. It is given by:

$$\begin{aligned} y_t = {} & 0.099y_{t-1} - 0.196y_{t-2} \\ & + \ (1 - e^{-0.324y_{t-1}^2})(1.058y_{t-1} - 0.629y_{t-2}) + \epsilon_t. \end{aligned} \tag{18}$$

where y_t is the first logarithmic difference of the Canadian real effective exchange rate. This is experiment 11 in the tables.

We also need some power experiments. We therefore consider the following ARCH models. An AR(1) model above is used for the conditional mean and an ARCH(1) model of $E(\epsilon_t^2|\epsilon_{t-1},\ldots) = h_t = \beta_0 + \beta_1\epsilon_{t-1}^2$ for the conditional variance. For each of three experiments the parameters are:

- ARCH Models
  - Experiment 14: $\alpha = 0.5$, $\beta_0 = 0.1$, $\beta_1 = 0.1$.
  - Experiment 15: $\alpha = 0.5$, $\beta_0 = 0.1$, $\beta_1 = 0.5$.

For each of the experiments all errors are obtained from the GAUSS normal pseudo-random number generator. For every sample, initial conditions are set to zero and 20 observations are dropped to minimise dependence on the choice of initial conditions. Throughout the lag order, p, is set to 1. The obvious exception is experiments 6 and 11 where the lag order used corresponds to the lag order of the data generation process which is based on the relevant empirical application taken from the literature. We consider samples of 100 and 200 observations. In Table 1 we give the simulation values of the unconditional means and variances for the Monte Carlo models, as an indication of the degree of nonlinearity in each model.

The testing framework is then in two parts. We use the procedure outlined in Section 3 to obtain residuals that are subsequently tested for ARCH. This test can be any of those that are available. In this paper we consider two: The first is the usual Engle (1982) LM-test (denoted LM in the tables) and the second is the ARCH test developed by Peguin-Feissolle (1999). This is a residual-based test using neural network ideas derived from the work of Lee, White, and Granger (1993) (denoted NN in the tables). The Peguin-Feissolle (1999) test assumes that the conditional variance of the error term of the regression model can be approximated by a neural network. The test is implemented by simply running the Lee, White, and Granger (1993) artificial neural network (ANN) nonlinearity test on the squared residuals of the regression model, where lagged residuals are used to construct the neural network used by the ANN test.[7] We choose this test because we wish

[7]In Blake and Kapetanios (2000) we used the RBF set-up proposed here for constructing a test of ARCH along the lines of the test proposed by Peguin-Feissolle (1999) and described in the text.

Table 1: Sample unconditional means and variances

		$\bar{y}$	σ_y^2	% time in regime 1	% time in regime 2
SETAR	2	0.263	1.136	40	60
	3	0.344	1.103	37	63
	4	1.475	2.619	18	82
	5	1.609	2.484	14	86
	6	0.769	1.373	24	76
STAR	7	0.012	1.232	—	—
	8	0.003	1.049	—	—
	9	0.006	1.708	—	—
	10	0.010	1.156	—	—
	11	0.005	1.708	—	—

to illustrate that radically different ARCH tests suffer from the problem of overrejection under the null to a similar extent and can benefit from the approach we suggest.

Given our discussion of appropriate ways to approximate any possible underlying nonlinear model we consider two nonlinear modelling choices. The subscript tlg denotes the use of the Taylor expansion method and the subscript rbf denotes use of the RBF neural network. Also we consider two extended ARCH tests. The superscript r denotes augmentation of the variable set, used to construct the neural network specification, by the lag of the recursive residual, as suggested by Lumsdaine and Ng (1999), whereas the superscript o denotes similar augmentation of the variable set using the lag of the OLS residual. Results in the form of rejection probabilities are presented in Tables 2–5. We do not report size corrected powers because it is not clear what are the proper empirical critical values to use to correct the rejection probabilities under the alternative. As we will see the rejection probabilities under the null of no ARCH vary greatly depending on the nonlinear model used. The columns headed simply LM and NN in Tables 2-5 apply the LM and NN ARCH tests on residuals obtained from fitting AR(1) models to the

Table 2: Monte Carlo Results for $T = 100$ and ARCH LM tests

		LM	LM^r	LM_{tlg}	LM_{rbf}	LM^r_{tlg}	LM^r_{rbf}	LM^o_{tlg}	LM^o_{rbf}
					Size				
AR	1	0.050	0.037	0.029	0.027	0.031	0.034	0.027	0.024
SETAR	2	0.043	0.026	0.027	0.026	0.023	0.027	0.029	0.025
	3	0.049	0.039	0.031	0.035	0.024	0.027	0.031	0.029
	4	0.144	0.081	0.029	0.028	0.023	0.031	0.031	0.027
	5	0.280	0.130	0.023	0.026	0.019	0.023	0.032	0.027
	6	0.081	0.068	0.043	0.038	0.028	0.027	0.032	0.037
STAR	7	0.048	0.040	0.037	0.036	0.027	0.031	0.026	0.031
	8	0.050	0.037	0.044	0.041	0.042	0.042	0.031	0.041
	9	0.137	0.088	0.042	0.046	0.026	0.034	0.036	0.027
	10	0.097	0.077	0.031	0.033	0.022	0.034	0.024	0.026
	11	0.100	0.050	0.020	0.044	0.020	0.020	0.018	0.030
Bilin	12	0.051	0.030	0.026	0.023	0.021	0.026	0.023	0.021
	13	0.458	0.127	0.046	0.046	0.037	0.033	0.018	0.032
					Power				
ARCH	14	0.112	0.085	0.074	0.074	0.036	0.044	0.039	0.057
	15	0.701	0.567	0.491	0.464	0.246	0.314	0.236	0.333

data[8].

The results make very interesting reading. The performance of the LM and NN tests depend markedly on the nonlinear model used. Rejection probabilities under the null reach 90% for the bilinear model, which accords with the findings of Bera and Higgins (1997) who note the difficulties encountered, for ARCH tests, with bilinear models, but even for SETAR and STAR models pronounced nonlinearity, measured by the difference between γ_1 and γ_2 for SETAR models and the magnitude of δ_2 for STAR models, induces rejection probabilities of up to 50%. Clearly nonlinearity in the mean and ARCH are difficult to distinguish using standard tests. The recursive residual tests of Lumsdaine and Ng (1999) are considerable improvements on the standard tests. Rejection probabilities fall substantially but still remain

[8]In the case of experiments 6 and 11 the lag order of the AR model corresponds to that used for generating the data.

Table 3: Monte Carlo Results for $T = 200$ and ARCH LM tests

		LM	LM^r	LM_{tlg}	LM_{rbf}	LM^r_{tlg}	LM^r_{rbf}	LM^o_{tlg}	LM^o_{rbf}
					Size				
AR	1	0.047	0.037	0.037	0.040	0.027	0.032	0.033	0.033
SETAR	2	0.055	0.047	0.044	0.041	0.042	0.045	0.046	0.043
	3	0.064	0.039	0.042	0.040	0.033	0.041	0.034	0.031
	4	0.258	0.124	0.036	0.037	0.035	0.038	0.047	0.041
	5	0.487	0.266	0.044	0.045	0.042	0.039	0.042	0.039
	6	0.113	0.097	0.045	0.041	0.035	0.039	0.036	0.034
STAR	7	0.044	0.039	0.036	0.036	0.039	0.047	0.034	0.035
	8	0.050	0.042	0.025	0.027	0.026	0.029	0.033	0.032
	9	0.208	0.150	0.041	0.041	0.041	0.041	0.046	0.042
	10	0.221	0.184	0.037	0.039	0.028	0.033	0.035	0.030
	11	0.192	0.143	0.023	0.067	0.030	0.044	0.027	0.043
Bilin	12	0.080	0.030	0.030	0.027	0.024	0.025	0.028	0.027
	13	0.777	0.173	0.071	0.066	0.043	0.051	0.034	0.047
					Power				
ARCH	14	0.255	0.204	0.175	0.176	0.114	0.132	0.121	0.141
	15	0.955	0.915	0.887	0.874	0.749	0.799	0.715	0.814

quite high for very nonlinear conditional mean processes. For example for the most extreme SETAR model, experiment 5, and 200 observations the rejection probability for the LM test is 26%.

Moving on to the new testing procedures we observe a dramatic improvement. Rejection probabilities are very close to 5%. None of the SETAR or STAR nonlinear models can induce even minor departures from the correct significance level. Bilinear models still produce overrejections. However this is to be expected: The variables involved in the bilinear model include the lagged error. This is not included in the set of variables used in the construction of the neural network. Therefore, the nonlinearity cannot be completely captured. However, even for these models noticeable and worthwhile improvement is observed. In particular adding recursive or OLS residuals produces satisfactory rejection probabilities apart from the final Bilinear model, experiment 14, where the nonlinearity is very pronounced and the model is,

Table 4: Monte Carlo Results for $T = 100$ and ARCH NN tests

		NN	NN^r	NN_{tlg}	NN_{rbf}	NN^r_{tlg}	NN^r_{rbf}	NN^o_{tlg}	NN^o_{rbf}
					Size				
AR	1	0.052	0.046	0.039	0.036	0.025	0.029	0.026	0.029
SETAR	2	0.057	0.033	0.023	0.024	0.020	0.020	0.021	0.019
	3	0.053	0.034	0.020	0.022	0.019	0.020	0.013	0.021
	4	0.163	0.085	0.027	0.029	0.016	0.021	0.022	0.023
	5	0.301	0.146	0.026	0.029	0.027	0.025	0.026	0.018
	6	0.082	0.067	0.046	0.038	0.026	0.030	0.023	0.032
STAR	7	0.063	0.036	0.035	0.042	0.023	0.025	0.023	0.031
	8	0.058	0.036	0.024	0.021	0.032	0.030	0.024	0.031
	9	0.156	0.088	0.023	0.025	0.024	0.022	0.026	0.027
	10	0.136	0.106	0.028	0.029	0.015	0.022	0.012	0.016
	11	0.122	0.056	0.018	0.059	0.017	0.027	0.017	0.040
Bilin	12	0.072	0.034	0.025	0.025	0.025	0.022	0.018	0.028
	13	0.451	0.128	0.054	0.051	0.023	0.040	0.021	0.036
					Power				
ARCH	14	0.130	0.096	0.087	0.089	0.055	0.062	0.050	0.070
	15	0.734	0.654	0.615	0.577	0.378	0.438	0.352	0.444

in fact, noninvertible.

When we consider the rejection probabilities under the alternative we see that the new procedures do not exhibit a very marked reduction in these probabilities, as desired. In particular we observe a reduction of about 20% for tests that do not include recursive or OLS residuals for samples of 100 observations and only a reduction of about 10% for samples of 200 observations. In some cases, as in Table 4, there is almost no loss of power for the NN_{tlg} and NN_{rbf} tests. The results are indeed very encouraging and clearly suggest that the new methods are useful.

5 An application to exchange rate data

Although the Monte Carlo experiments, including those on empirically based models, are persuasive, it is interesting to set up an application on actual

Table 5: Monte Carlo Results for $T = 200$ and ARCH NN tests

		NN	NN^r	NN_{tlg}	NN_{rbf}	NN^r_{tlg}	NN^r_{rbf}	NN^o_{tlg}	NN^o_{rbf}
					Size				
AR	1	0.048	0.038	0.027	0.029	0.029	0.024	0.028	0.028
SETAR	2	0.066	0.039	0.035	0.029	0.031	0.026	0.023	0.029
	3	0.077	0.038	0.026	0.028	0.026	0.023	0.021	0.023
	4	0.282	0.128	0.045	0.038	0.025	0.028	0.026	0.025
	5	0.502	0.264	0.036	0.038	0.024	0.030	0.021	0.033
	6	0.109	0.099	0.049	0.045	0.041	0.047	0.040	0.045
STAR	7	0.048	0.038	0.032	0.032	0.025	0.022	0.028	0.025
	8	0.065	0.047	0.014	0.014	0.027	0.017	0.023	0.016
	9	0.240	0.178	0.031	0.033	0.023	0.023	0.031	0.026
	10	0.275	0.213	0.035	0.038	0.031	0.043	0.023	0.037
	11	0.221	0.165	0.027	0.093	0.034	0.033	0.023	0.044
Bilin	12	0.105	0.030	0.025	0.024	0.024	0.021	0.021	0.023
	13	0.761	0.181	0.091	0.087	0.052	0.065	0.026	0.063
					Power				
ARCH	14	0.238	0.176	0.160	0.152	0.108	0.117	0.115	0.126
	15	0.969	0.949	0.944	0.938	0.850	0.885	0.842	0.893

data. We consider the modelling of bilateral exchange rates. These have often been modelled as a nonlinear process (see Sarantis (1999)). The interest here is to see how often models which accept the presence of ARCH reject for the new tests. The model we use is simple. We run an autoregressive model on the differenced real exchange rates, choosing the lag length using BIC, from a maximum lag order of 4 lags. We then performed ARCH tests on the residuals from these models.

We construct the data in the following way. Every bilateral real exchange rate, denoted q, for a given 'home' currency j against an alternative i-th 'foreign' currency at time t is

$$q_{i,t} = s_{i,t} + p_{j,t} - p_{i,t}$$

where $s_{i,t}$ is the corresponding nominal exchange rate (in i-th currency units per unit of the home currency), $p_{j,t}$ the price level in the home country, and

$p_{i,t}$ the price level of the i-th alternative country. That is, a rise in $q_{i,t}$ implies a real appreciation of the home country against the i-th currency.

The j-denoted home currencies that we used were the US dollar, the Deutsche Mark and the Yen with the foreign countries from 44 possible currencies.[9] All data are quarterly, spanning from 1957Q1 to 1998Q4 and the bilateral nominal exchange rates against the currencies other than the US dollar are cross-rates computed using the US dollar rates. As noted above we consider three different cases each one of which consists of up to 39 country pairs and corresponds to a different home currency (US dollar, DM, Yen). We use the average quarterly nominal exchange rates and the prices used are consumer price indices. All variables are in logs and are not seasonally adjusted. The data were obtained from IMF (2004) and OECD (2004) as appropriate.

When we consider the Japan bilateral exchange rates, the Pacific Basin countries of our sample represent approximately one fourth (24%) of Japan's total trade, the US represents another one fourth, and the remaining five G7 are 12%.[10] This data is quarterly, spanning from 1960Q1 to 2000Q4 and as before the bilateral nominal exchange rates against the currencies other than the US dollar are cross-rates computed using the US dollar rates.

We give the results in Tables 7–6. In roughly half the cases that we formerly rejected the hypothesis of no ARCH we now accept the null. Table 6 illustrates this point. Using the standard LM ARCH provides evidence against the null hypothesis of no ARCH in 18 out of 39 series of Yen bilateral real exchange rates. When the new tests are used, evidence against the null hypothesis of no ARCH appears in only a quarter of the series. As the

[9] The 44 currencies considered are those of Australia, Austria, Bangladesh, Belgium, Bolivia, Brazil, Canada, Chile, Colombia, Cyprus, Czech Republic, Denmark, Finland, France, Germany, Greece, Hong Kong, Hungary, Iceland, Indonesia, Italy, Japan, Korea, Luxembourg, Malaysia, Malta, Mexico, Netherlands, New Zealand, Norway, Philippines, Poland, Portugal, Singapore, South Africa, Spain, Sri Lanka, Sweden, Switzerland, Thailand, Turkey, United Kingdom, United States, and Venezuela.

[10] This data was obtained from OECD (2004).

Monte Carlo study, in the previous section, does not suggest a significant loss of power for samples sizes of 200 observations, we conclude that the standard LM ARCH test can lead to spurious rejection of the null hypothesis of no ARCH in a significant proportion of the data considered. Similar but less striking results are obtained for the US and DM real exchange rates. These results clearly demonstrate the utility of our approach.

6 Conclusions

It is well known that tests for ARCH are powerful against a wider variety of mispecifications. In particular, it is well known from the work of Lumsdaine and Ng (1999) and others that mispecification in the conditional mean may lead to spurious rejection of the no ARCH hypothesis. However, apart from the general heuristic methods of Lumsdaine and Ng (1999) there is little in terms of methods to avoid this problem.

This paper suggests a solution in the case where the conditional mean function suffers from neglected nonlinearity of unknown form. We initially show, via simulations, that the problem of spurious rejection of the no ARCH hypothesis is serious in the case of neglected conditional mean nonlinearity. We then suggest the use of neural networks to approximate to an arbitrary level of accuracy any unknown nonlinearity. Once nonlinearity has been removed, the residuals are tested for ARCH using standard tests. Monte Carlo evidence suggests that the new methods are able to remove the large distortions introduced by nonlinearity at a rather modest cost in terms of power loss. When we consider bilateral exchange rates we find that there is substantial evidence that processes could be misunderstood as containing ARCH while they are, in fact, processes with nonlinearity in the conditional mean.

Our work relates to the larger topic of distinguishing nonlinearity in the mean from ARCH. We deal with only one facet of this topic, namely spurious rejection of the no-ARCH hypothesis using standard ARCH tests. Of course, application of both linearity tests of the conditional mean and ARCH may provide a useful method of distinguishing between the two. In this context it is worth noting the well known result that the LM ARCH test has zero asymptotic relative efficiency compared to the typical LM linearity test, shown by Luukkonen, Saikkonen, and Teräsvirta (1988). An implication of this result is that the ARCH test is likely to have considerably less power against mispecification of the conditional mean than the LM linearity test. A related possibility, which is not the focus of our work, concerns the effect of the presence of ARCH for the selection of the conditional mean model for a particular process. It is hoped that our work will stimulate further interest in the research of the interplay of ARCH and nonlinearity in time series analysis.

References

BERA, A. K., AND M. L. HIGGINS (1997): "ARCH and Bilinearity as Competing Models for Conditional Dependence," *Journal of Business and Economic Statistics*, 15, 43–50.

BERA, A. K., M. L. HIGGINS, AND S. LEE (1992): "Interaction Between Autocorrelation and Conditional Heteroscedasticity: A Random-Coefficient Approach," *Journal of Business and Economic Statistics*, 10, 133–142.

BISHOP, C. M. (1995): *Neural Networks for Pattern Recognition*. Oxford University Press.

BLAKE, A. P., AND G. KAPETANIOS (2000): "A Radial Basis Function Artificial Neural Network Test for ARCH," *Economics Letters*, 69, 15–23.

——— (2003a): "Pure Significance Tests of the Unit Root Hypothesis Against Nonlinear Alternatives," *Journal of Time Series Analysis*, 24, 253–267.

——— (2003b): "A Radial Basis Function Artificial Neural Network Test for Neglected Nonlinearity," *Econometrics Journal*, 6, 357–373.

CAMPBELL, J. Y., A. W. LO, AND A. C. MACKINLAY (1997): *The Econometrics of Financial Markets*. Princeton University Press.

CYBENKO, G. (1989): "Approximation by Superpositions of a Sigmoidal Function," *Mathematics of Control, Signals and Systems*, 2, 304–314.

DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

HORNIK, K., M. STINCHCOMBE, AND H. WHITE (1989): "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, 2, 359–366.

IMF (2004): *International Financial Statistics*. International Monetary Fund, Washington, D.C., CD-ROM.

LEE, T. H., H. WHITE, AND C. W. J. GRANGER (1993): "Testing for Neglected Nonlinearity in Time Series Models: A Comparison of Neural Network Methods and Alternative Tests," *Journal of Econometrics*, 56, 269–290.

LUMSDAINE, R. L., AND S. NG (1999): "Testing for ARCH in the Presence of a Possibly Misspecified Conditional Mean," *Journal of Econometrics*, 93, 257–279.

LUUKKONEN, R., P. SAIKKONEN, AND T. TERÄSVIRTA (1988): "Testing Linearity in Univariate Time Series Models," *Scandinavian Journal of Statistics*, 15, 161–175.

OECD (2004): *Statistical Compendium*. Organisation for Economic Co-operation and Development, Paris, CD-ROM.

ORR, M. J. (1995): "Regularisation in the Selection of Radial Basis Function Centers," *Neural Computation*, 7, 606–623.

PEGUIN-FEISSOLLE, A. (1999): "A Comparison of the Power of Some Tests for Conditional Heteroscedasticity," *Economics Letters*, 63, 5–17.

POTTER, S. (1995): "A nonlinear approach to US GNP," *Journal of Applied Econometrics*, 10, 109–125.

SARANTIS, N. (1999): "Modelling Nonlinearities in Real Effective Exchange Rates," *Journal of International Money and Finance*, 18, 27–45.

STINCHCOMBE, M., AND H. WHITE (1989): "Universal Approximation Using Feedforward Networks with Non-Sigmoid Hidden Layer Activation Functions," *Proceedings of the International Joint Conference on Neural Networks*, 1, 359–366.

TERÄSVIRTA, T., C. F. LIN, AND C. W. J. GRANGER (1993): "Power of the Neural Network Linearity Test," *Journal of Time Series Analysis*, 14, 209–220.

WHITE, H. (1992): *Artificial Neural Networks: Approximation and Learning Theory*. Basil Blackwell.

Table 6: Yen bilateral real exchange rates

	LM	LM_{tlg}	LM_{rbf}		LM	LM_{tlg}	LM_{rbf}
	Probability value				Probability value		
Australia	**0.000**	0.145	0.142	Korea	**0.000**	**0.012**	**0.013**
Austria	**0.018**	0.115	0.114	Malaysia	0.060	0.061	0.059
Bangladesh	0.803	0.702	0.676	Mexico	**0.006**	**0.002**	**0.002**
Belgium	**0.001**	**0.001**	**0.001**	Netherlands	**0.019**	**0.017**	**0.017**
Bolivia	**0.000**	0.950	0.812	N. Zealand	**0.021**	**0.039**	**0.038**
Brazil	0.169	0.615	0.668	Norway	0.151	0.142	0.125
Canada	0.219	0.407	0.400	Philippines	0.924	0.904	0.902
Chile	0.513	0.467	0.628	Poland	**0.013**	0.133	0.130
Colombia	**0.015**	0.517	0.526	Portugal	**0.032**	0.093	0.089
Czech Rep.	0.773	0.786	0.795	Singapore	**0.020**	**0.006**	**0.013**
Denmark	**0.014**	**0.020**	**0.020**	Spain	0.898	0.746	0.751
Finland	0.293	0.671	0.744	Sri Lanka	0.905	0.892	0.883
France	0.261	0.291	0.290	Sweden	**0.004**	**0.024**	0.067
Germany	**0.007**	**0.006**	**0.006**	Switzerland	**0.002**	**0.003**	**0.003**
Greece	0.389	0.797	0.809	Thailand	0.743	0.610	0.698
Hong Kong	0.991	0.685	0.681	Turkey	0.993	0.994	0.994
Hungary	0.929	0.317	0.321	UK	0.769	0.997	0.997
Iceland	0.875	0.953	0.942	US	**0.044**	0.134	0.133
Indonesia	**0.001**	0.968	0.129	Venezuela	0.943	0.933	0.933
Italy	0.054	0.224	0.052				
Total rejections for all countries					18	10	9

Note: Rejections at the 5% significance level in bold.

Table 7: US bilateral real exchange rates

	LM	LM_{tlg}	LM_{rbf}
	Probability value		
Australia	**0.005**	**0.030**	**0.031**
Austria	**0.001**	**0.019**	**0.017**
Belgium	0.723	0.879	0.862
Canada	0.702	0.462	0.493
Cyprus	**0.025**	**0.017**	**0.017**
Finland	0.167	0.306	0.274
France	0.099	0.166	0.345
Germany	**0.000**	**0.000**	**0.000**
Greece	0.558	0.556	0.557
Italy	**0.009**	**0.031**	**0.035**
Japan	**0.005**	**0.011**	**0.011**
Luxembourg	**0.004**	**0.005**	**0.005**
Malta	**0.001**	**0.001**	**0.016**
Netherlands	**0.000**	**0.004**	**0.000**
New Zealand	0.338	0.445	0.445
Norway	**0.050**	0.059	0.059
Portugal	0.509	0.563	0.536
South Africa	0.246	0.192	0.188
Spain	0.770	0.790	0.789
Sweden	**0.045**	0.265	0.280
Switzerland	**0.006**	0.291	0.296
UK	0.160	0.150	0.168
No. of rejections	12	9	9

Note: Rejections at the 5% level in bold.

Table 8: DM real exchange rates

	LM	LM_{tlg}	LM_{rbf}
	Probability value		
Australia	0.145	0.428	0.423
Austria	**0.024**	0.625	0.626
Belgium	0.969	0.834	0.833
Canada	**0.000**	**0.003**	**0.000**
Cyprus	**0.000**	0.330	0.325
Finland	0.347	0.918	0.986
France	0.315	0.425	0.433
Greece	0.154	0.322	0.228
Italy	**0.001**	**0.038**	0.071
Japan	**0.002**	**0.002**	**0.002**
Luxembourg	0.154	0.536	0.558
Malta	**0.005**	0.079	0.075
Netherlands	0.657	0.855	0.854
New Zealand	0.318	0.587	0.598
Norway	**0.001**	0.160	0.156
Portugal	0.742	0.747	0.742
South Africa	**0.007**	**0.005**	**0.006**
Spain	0.901	0.834	0.834
Sweden	**0.003**	0.641	0.448
Switzerland	**0.000**	0.656	0.656
UK	**0.013**	0.152	0.156
US	**0.000**	**0.000**	**0.000**
No. of rejections	12	5	4

Note: Rejections at the 5% level in bold.