A Quantile Regression Neural Network Approach to Estimating

the Conditional Density of Multiperiod Returns

James W. Taylor

Saïd Business School

University of Oxford

Address for Correspondence:

James W. Taylor
Saïd Business School
University of Oxford
59 George Street
Oxford OX1 2BE, UK

Tel: +44 (0)1865 288678
Fax: +44 (0)1865 288651
Email: james.taylor@sbs.ox.ac.uk

A Quantile Regression Neural Network Approach to Estimating the Conditional Density of Multiperiod Returns

Abstract

This paper presents a new approach to estimating the conditional probability distribution of multiperiod financial returns. Estimation of the tails of the distribution is particularly important for risk management tools, such as value at risk models. A popular approach is to assume a Gaussian distribution, and to use a theoretically derived variance expression which is a nonlinear function of the holding period, k, and the 1-step-ahead volatility forecast, $\hat{\sigma}_{t+1}$. The new method avoids the need for a distributional assumption by applying quantile regression to the historical returns from a range of different holding periods to produce quantile models which are functions of k and $\hat{\sigma}_{t+1}$. A neural network is used to estimate the potentially nonlinear quantile models. Using daily exchange rates, the approach is compared to GARCH-based quantile estimates. The results suggest that the new method offers a useful alternative for estimating the conditional density.

Key words: quantile regression; neural networks; multiperiod returns; conditional density

1. INTRODUCTION

This paper aims to improve the estimation of the conditional probability distribution of financial returns. Accurate estimation of the tails of the distribution are of particular importance for risk management tools, such as value at risk models, which have received considerable attention in recent years (see Duffie and Pan, 1997). Value at risk calculations aim to measure the worst expected loss over a given time interval under normal market conditions at a given confidence level (see Jorion, 1997). The conditional density is not only needed for risk management. Hansen (1994) notes that the density is often important for option pricing, and Baillie and Bollerslev (1992) highlight the necessity for accurate confidence interval estimation to accompany forecasts of the conditional mean.

A popular procedure for estimating the distribution of 1-period returns is to forecast the volatility and then to make a Gaussian assumption. However, returns are not always normally distributed which has prompted alternatives, such as the use of a *t*-distribution and nonparametric methods. Boudoukh *et al.* (1997) observe that, even if the 1-period return is Gaussian, the distribution may well be much more complicated for the multiperiod return. This provides additional motivation for the use of a nonparametric method.

In this paper, we propose a quantile regression approach to the estimation of the distribution of multiperiod returns. This nonparametric approach uses historical returns from a range of different holding periods and produces quantile models which are functions of the length, k, of the holding period and the 1-step-ahead volatility forecast, $\hat{\sigma}_{t+1}$, as suggested by theoretically derived variance expressions. The θth quantile of a variable y is the value, $Q(\theta)$, for which $P(y<Q(\theta))=\theta$. An approximation of the full probability distribution can be produced from the quantile estimates corresponding to a range of values of θ ($0<\theta<1$).

The functional form of multiperiod volatility forecasts varies greatly depending upon the choice of forecasting method. For example, with moving average methods, the k-period

volatility forecast is usually calculated as the 1-step-ahead forecast, $\hat{\sigma}_{t+1}$, inflated by $\sqrt{k}$. By contrast, the GARCH(1,1) volatility forecast is a much more complex nonlinear function of k and $\hat{\sigma}_{t+1}$. Whilst the motivation for using k and $\hat{\sigma}_{t+1}$ as explanatory variables in the quantile regression models is apparent, the appropriate nonlinear specification is much less clear. In this paper, we overcome this problem by using a neural network to perform the quantile regression. This computationally intensive approach to modelling enables the estimation of potentially nonlinear models, without the need to specify a precise functional form.

The next section of the paper considers the existing approaches to estimating the quantiles of the multiperiod returns. In the section that follows, we describe the theory of quantile regression, and discuss how an artificial neural network can be used to perform quantile regression. We then present our new proposal. In the next section, we use daily exchange rate data to compare the quantile estimates of our new approach with those of two commonly used GARCH-based methods. The final section provides a summary and conclusion.

2. TRADITIONAL APPROACHES TO MULTIPERIOD QUANTILE ESTIMATION

The traditional procedure for estimating quantiles of multiperiod returns consists of two stages. Firstly, the volatility is estimated for the periods under consideration, and secondly, a probability distribution is assumed. This procedure is used by Alexander and Leigh (1997) and proposed by Kroner *et al.* (1995). In this section, we discuss the two stages, and highlight potential improvements.

2.1. Distributional Assumption

The quantiles of 1-step-ahead returns are usually constructed using a Gaussian distribution (see Duffie and Pan, 1997). This is consistent with an assumption of Gaussian log returns in the finance literature. It is also consistent with ARCH models, provided the parameters have been

estimated using maximum likelihood based on Gaussian disturbances. However, market returns are frequently found to have excess kurtosis relative to a normal distribution (see, for example, the analysis of Hull and White (1998)). Baillie and Bollerslev (1989) suggest the use of a *t*-distribution for the estimation of ARCH models in order to accommodate fat tails. If a *t*-distribution is used for parameter estimation, it would be consistent to use a *t*-distribution to estimate the quantiles. However, log returns do not always have a normal or *t*-distribution, and so a nonparametric approach to estimating quantiles has strong appeal.

A normal distribution is also often used for estimating the quantiles of multiperiod returns. If the returns at different lead times are assumed to be uncorrelated and normally distributed, then perhaps it seems reasonable to assume normality for the distribution of the multiperiod returns, since they are the sum of the returns over a holding period. However, this assumption is inappropriate because, although the returns at different lead times are uncorrelated, they are not necessarily independent. This interdependence is apparent from the variance which would be unpredictable by moving average or ARCH methods, if it did not depend on its past values. Indeed, the hypothesis underlying ARCH processes is that the variance is autoregressive. Furthermore, Baillie and Bollerslev (1992) discuss how the higher order moments (such as the skewness and kurtosis) of the returns distribution are also interdependent for ARCH models. Recognising this, Hansen (1994) presents a nonparametric autoregressive conditional density (ARCD) model which aims to model the full probability density in an autoregressive framework. Hansen's work produces *k*-step-ahead density estimates but his work has not been extended to multiperiod estimation (which focuses on the sum of the returns over a holding period). In summary, as Boudoukh *et al.* (1997) note, even if the 1-period innovation is normally distributed, the multiperiod innovation will have a much more complicated distribution. There is thus strong motivation for using a nonparametric approach to estimating the quantiles of multiperiod returns.

2.2. Functional Form of the Volatility Forecasts

The functional form of multiperiod volatility forecasts varies greatly depending upon the choice of forecasting method. Moving average methods predict volatility using simple or weighted moving averages of past volatility. One of the most popular approaches is the exponentially weighted moving average estimator, which is used in the *RiskMetrics*[TM] *Technical Document* (1996). With moving average methods, the usual approach for calculating the volatility of the *k*-period return (i.e. the return over a holding period of length *k*) is to assume that the returns are uncorrelated with the same variance. The forecast for the *k*-period variance is then just the 1-step-ahead forecast, $\hat{\sigma}_{t+1}^2$, multiplied by *k*. A traditional estimate of the θth quantile would then be

$$Q_{t,k}(\theta) = Z_\theta \, \hat{\sigma}_{t,k} = Z_\theta \, k^{\frac{1}{2}} \hat{\sigma}_{t+1} \tag{1}$$

where Z_θ is the θth quantile of the standard normal distribution. Here the quantile is a simple function of *k* and $\hat{\sigma}_{t+1}$.

Autoregressive Conditional Heteroskedasticity (ARCH) models provide estimates of the variance of the return, r_t, at time *t* conditional upon I_{t-1}, the information set of all observed returns up to time *t*-1 (see Engle, (1982)). This can be viewed as the variance of the error term, $e_t = r_t - E(r_t/I_{t-1})$. Bollerslev (1986) extended the ARCH class of models to Generalised Autoregressive Conditional Heteroskedastic (GARCH) models which enables a more parsimonious representation in many applications. GARCH models express the conditional variance as a linear function of lagged squared error terms and also lagged conditional variance terms. For example, the 1-step-ahead GARCH(1,1) variance forecast is given by

$$\hat{\sigma}_{t+1}^2 = \alpha_0 + \alpha_1 e_t^2 + \beta_1 \hat{\sigma}_t^2$$

The *s*-step-ahead forecast is given by the following recursive expression for *s*>1

$$\hat{\sigma}_{t+s}^2 = \alpha_0 + (\alpha_1 + \beta_1)\hat{\sigma}_{t+s-1}^2$$

Using this, we can write the k-period GARCH(1,1) variance forecast as

$$\hat{\sigma}_{t,k}^2 = \sum_{i=1}^{k} \hat{\sigma}_{t+i}^2$$

$$= \frac{\alpha_0 k}{1 - \alpha_1 - \beta_1} + \left(\hat{\sigma}_{t+1}^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \right) \left(\frac{1 - (\alpha_1 + \beta_1)^k}{1 - \alpha_1 - \beta_1} \right)$$

If we assume the conditional mean of the returns is zero, a traditional estimate of the θth quantile is then

$$Q_{t,k}(\theta) = Z_\theta \left(\frac{\alpha_0 k}{1 - \alpha_1 - \beta_1} + \left(\hat{\sigma}_{t+1}^2 - \frac{\alpha_0}{1 - \alpha_1 - \beta_1} \right) \left(\frac{1 - (\alpha_1 + \beta_1)^k}{1 - \alpha_1 - \beta_1} \right) \right)^{\frac{1}{2}} \qquad (2)$$

The quantile is a complicated nonlinear function of k and $\hat{\sigma}_{t+1}$. Indeed, the same is true for other GARCH volatility models.

Expressions (1) and (2) suggest that the functional form of the volatility, and hence the quantile, is open to debate. Indeed, as Kroner *et al.* (1995) point out, the volatility modelling literature indicates that volatility is mean reverting at a hyperbolic rate which is slower than GARCH models permit. In addition, Diebold *et al.* (1998) show that scaling 1-step-ahead volatility forecasts by $k^{\frac{1}{2}}$ is inappropriate and overestimates the volatility at long horizons. Furthermore, as we discussed earlier, the Gaussian assumption is often inappropriate. There is thus strong potential for a nonparametric approach that allows more flexible modelling of the multiperiod quantile as a function of the holding period, k. In this paper, we present a new nonparametric approach which uses quantile regression. Before describing the procedure, we first introduce the quantile regression theory of Koenker and Bassett (1978, 1982).

3. QUANTILE REGRESSION

This section consists of two parts. Firstly, we present the linear quantile regression theory of Koenker and Bassett (1978, 1982). Secondly, we describe White's (1992) proposal for the use of quantile regression within an artificial neural network for nonlinear quantile modelling.

3.1. Linear Quantile Regression

Koenker and Bassett (1978, 1982) developed theory for the estimation of the quantiles of a variable y_t which is assumed to be a linear function of other variables. In order to provide some intuition, let us first consider the simple case of the constant model $y_t = \beta_0 + e_t$, where β_0 is a constant parameter and e_t is an i.i.d. random error term. Koenker and Bassett began by noting that the θth quantile of y_t can be derived, from a sample of observations, as the solution $\beta_0(\theta)$ to the following minimisation problem:

$$\min_{\beta_0} \left(\sum_{t \mid y_t \geq \beta_0} \theta \left| y_t - \beta_0 \right| + \sum_{t \mid y_t < \beta_0} (1 - \theta) \left| y_t - \beta_0 \right| \right)$$

The case of the median ($\theta = \frac{1}{2}$) is well known, but the general result is not. The minimisation problem, as a means for finding the θth sample quantile, readily extends to the more general case where y_t is a linear function of explanatory variables. Consider the following rather general model of systematic heteroscedasticity,

$$y_t = \mu_t(\boldsymbol{x}_t) + \sigma_t(\boldsymbol{x}_t) e_t$$

where $\boldsymbol{x}_t$ is a row vector of explanatory variables, $\mu_t(\boldsymbol{x}_t)$ may be thought of as the conditional mean of the regression process, $\sigma_t(\boldsymbol{x}_t)$ as the conditional scale, and e_t as an error term independent of vector $\boldsymbol{x}_t$. The θth quantile of e_t is defined as the value, $Q_e(\theta)$, for which $P(e_t < Q_e(\theta)) = \theta$. Note that having μ_t and σ_t depend on the same vector $\boldsymbol{x}_t$ is solely for notational convenience. The conditional quantile functions of y_t are then

$$Q_{yt}(\theta \mid \boldsymbol{x}_t) = \mu_t(\boldsymbol{x}_t) + \sigma_t(\boldsymbol{x}_t)Q_e(\theta)$$

Consider the case where μ_t and σ_t are linear functions of $\boldsymbol{x}_t$ which has 1 as first element,

$$Q_{yt}(\theta \mid \boldsymbol{x}_t) = \boldsymbol{x}_t\,\boldsymbol{\beta} + (1 + \boldsymbol{x}_t\,\boldsymbol{\gamma})Q_e(\theta) \tag{3}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of parameters. (Setting all the elements of $\boldsymbol{\gamma}$ to zero is equivalent to assuming that the error term of y_t is i.i.d.) (3) can be rewritten as

$$Q_{yt}(\theta \mid \boldsymbol{x}_t) = \boldsymbol{x}_t\,\boldsymbol{\beta}(\theta) \tag{4}$$

where $\boldsymbol{\beta}(\theta)$ is a vector of parameters dependent on θ. Koenker and Bassett (1978) defined the *θth regression quantile* ($0<\theta<1$) as any solution, $\boldsymbol{\beta}(\theta)$, to the *quantile regression* minimisation problem

$$\min_{\boldsymbol{\beta}} \left(\sum_{t \mid y_t \geq \boldsymbol{x}_t\boldsymbol{\beta}} \theta \left| y_t - \boldsymbol{x}_t\,\boldsymbol{\beta} \right| + \sum_{t \mid y_t < \boldsymbol{x}_t\boldsymbol{\beta}} (1-\theta)\left| y_t - \boldsymbol{x}_t\,\boldsymbol{\beta} \right| \right) \tag{5}$$

Koenker and Bassett (1982) showed that if y_t and $\boldsymbol{x}_t$ are selected as dependent and independent variables respectively, then quantile regression delivers parameters that asymptotically approach the parameters, $\boldsymbol{\beta}(\theta)$, in (4) as the number of observations increases.

The common procedure for building an explanatory model for a variable is to look for a relationship between past observations of that variable and past observations of potential explanatory variables. This is not a feasible procedure for building a model for the quantiles of a variable because past observations of the quantiles will not be available, as they are unobservable. The appeal of quantile regression is that past observations of the quantiles are not required. Instead, the variable itself is regressed on explanatory variables to produce a model for the quantile.

3.2. A Quantile Regression Artificial Neural Network

Artificial neural networks allow the estimation of possibly nonlinear models without the need to specify a precise functional form. The most widely-used neural network for forecasting is the single hidden layer feedforward network (Zhang *et al.*, 1998). It consists of a set of n inputs, which are connected to each of m units in a single hidden layer, which, in turn, are connected to an output. In regression terminology, the inputs are explanatory variables, x_{it}, and the output is the dependent variable, y_t. The resultant model can be written as

$$f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w}) = g_2\left(\sum_{j=0}^{m} v_j\, g_1\left(\sum_{i=0}^{n} w_{ji}\, x_{it}\right)\right)$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are activation functions, which are frequently chosen as sigmoidal and linear respectively, and w_{ji} and v_j are the weights (parameters) to be estimated.

White (1992) presents theoretical support for the use of quantile regression within an artificial neural network for the estimation of potentially nonlinear quantile models. The only other work that we are aware of, that considers quantile regression neural networks, is that of Burgess (1995) who briefly discusses the appeal of the procedure. Instead of fitting a linear quantile function using the expression in (5), a quantile regression neural network model, $f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w})$, of the θth quantile can be estimated using the following minimisation

$$\min_{\boldsymbol{v}, \boldsymbol{w}} \left(\sum_{t\,|\,y_t \geq f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w})} \theta \left| y_t - f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w}) \right| + \sum_{t\,|\,y_t < f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w})} (1-\theta) \left| y_t - f(\boldsymbol{x}_t, \boldsymbol{v}, \boldsymbol{w}) \right| + \lambda_1 \sum_{j,i} w_{ji}^2 + \lambda_2 \sum_{i} v_i^2 \right)$$

$$(6)$$

where λ_1 and λ_2 are regularisation parameters which penalise the complexity of the network and thus avoid overfitting (see Bishop, 1997, §9.2). The optimal values of λ_1 and λ_2 and the number, m, of units in the hidden layer can be established by cross-validation (see Donaldson and Kamstra, 1996; Bishop, 1997, §9.8). In the next section, we describe how a quantile regression neural network can be used to estimate the probability distribution of multiperiod returns.

4.ESTIMATING THE MULTIPERIOD DISTRIBUTION USING QUANTILE REGRESSION

Our proposal is to use quantile regression to construct quantile functions for the multiperiod returns. As dependent variable, we use a series of multiperiod returns corresponding to various holding periods, k. In view of the quantile expressions in (1) and (2), candidates for the explanatory variables could be simple, linear and nonlinear, functions of k and $\hat{\sigma}_{t+1}$. However, selecting appropriate explanatory variables is not straightforward and so, in this paper, we use an artificial neural network to estimate the nonlinear quantile models. It is important to note that our proposal is very different from standard quantile regression described in the previous section. The standard approach involves the estimation of a model for the quantile of a variable in period t, conditional upon information available up to period t. Our procedure aims to estimate a conditional quantile model which describes the evolution of the quantile over the next k periods.

4.1. Implementation of the Quantile Regression Approach

The method proceeds by collecting together, as a single series, multiperiod returns corresponding to various holding periods, k. For illustrative purposes, consider the case where we are interested in constructing the returns distributions for holding periods of length 1, 3, 5, 7, 10, 12 and 15 periods. Suppose we have a thousand returns available for each of these holding periods. The single series would have the thousand 1-period returns first, followed by the thousand 3-period returns, then the thousand 5-period returns, etc. This is the *returns series*. We then define the elements of the *k series* as taking a value of k when the corresponding element of the forecast error series is a k-period return. If estimating quantiles for holding periods of length 1, 3, 5, 7, 10, 12 and 15, the k series will consist of a thousand 1's, followed by a thousand 3's, then a thousand 5's, etc. The third series to be constructed is the *volatility series*. This series contains 1-step-ahead volatility forecasts, $\hat{\sigma}_{t+1}$, which have been estimated by any method, such

as a GARCH model. The forecast origin of these forecasts is set at the same origin as the multiperiod return in the corresponding element of the multiperiod returns series. If the first 1-period return has the same origin as the first 3-period return, the first 5-period return, etc., then the volatility series will have the same forecast in the 1st entry as in the 1001st entry, and the 2001st entry, etc.

We then carry out quantile regressions with the multiperiod returns series as dependent variable. Earlier work investigated the use of simple functions of the k vector and the vector of volatility forecasts, $\hat{\sigma}_{t+1}$, as explanatory variables (Taylor, 1999a). For example, the 95th quantile, $Q_{t,k}(0.95)$, was estimated by using $\theta=0.95$ in the quantile regression minimisation in (5) with the returns series as dependent variable and k, $k\hat{\sigma}_{t+1}$ and $k^{\frac{1}{2}}\hat{\sigma}_{t+1}$ as regressors. (The vector $k\hat{\sigma}_{t+1}$ was constructed with ith element equal to the product of the ith element of the k vector and the ith element of the volatility vector.) The result was a model of the form

$$Q_{t,k}(0.95) = a + b\,k + c\,k\,\hat{\sigma}_{t+1} + d\,k^{\frac{1}{2}}\,\hat{\sigma}_{t+1}$$

where a, b, c and d are parameters estimated by quantile regression. The choice of explanatory variables was based on an inspection of coefficient bootstrapped standard errors and a pseudo R^2 statistic. A reasonable approximation to the correct quantile expression should be obtained if the explanatory variables are well chosen. In this paper, we take the view that a more efficient approach to quantile model specification is to use an artificial neural network. Neural networks avoid the laborious, and potentially inefficient, procedure of selecting transformed nonlinear variables for the linear regression. We use a quantile regression neural network to fit a nonlinear quantile model as a function of k and $\hat{\sigma}_{t+1}$.

4.2. Additional Features of the Quantile Regression Approach

The moving average and ARCH multiperiod volatility forecasts are estimated solely from 1-period returns. Therefore, the corresponding quantile forecasts in (1) and (2) are also based on just 1-period returns. Our new proposal has the appeal of using multiperiod returns from several different holding periods in constructing the multiperiod quantiles.

If the return is correlated with other market returns, then, ideally, this should be accommodated in a quantile estimate. This is an important issue for value at risk applications. Our proposal can be adapted to allow for correlation by simply including an estimate of the correlation as an extra input variable to the neural network. Gibson and Boyer (1998) review the various statistical and market-based approaches to estimating the correlation between returns.

Interestingly, our approach enables quantile models with completely different specifications to be estimated for different quantiles. For example, whilst the model for the 95th quantile might be a complex nonlinear function of k and $\hat{\sigma}_{t+1}$, the model for the 5th quantile might be a simple linear function. Yar and Chatfield (1990) note that one advantage of a theoretical method for estimating prediction intervals over an empirical procedure is that the theoretical formulae give insight into how the forecast error variance varies with k. A similar point can be made for our new proposal.

With modern computing power, we feel that the computational intensity of the approach is not a significant constraint. From a theoretical perspective, there may be inefficiencies due to the likely correlation between dependent variable observations. It is not clear how to handle this in quantile regression, and it is clearly an area for further research if the basic appeal of the method is accepted.

5. A COMPARISON OF EXCHANGE RATE QUANTILE ESTIMATES

In this section, we describe a study that compared our new proposal with traditional quantile estimation methods. Our analysis used 2028 daily observations of the exchange rates for the German deutsche mark and the Japanese yen quoted against the U.S. dollar from 4 July 1988 to 5 July 1996.

We estimated quantiles of the multiperiod log returns for holding periods of length 1, 3, 5, 7, 10, 12 and 15 days. Although these periods are chosen arbitrarily, for daily returns, it seems quite reasonable to consider a range of holding periods from one day to three weeks. For example, Duffie and Pan (1997) and Jorion (1997) report that two weeks has been proposed by various organisations as a standard for value at risk calculations. We compared the quantile estimates with the corresponding actual multiperiod returns to reveal post-sample performance. We carried out this procedure for 1000 moving windows, each consisting of 1014 observations, to give 1000 post-sample quantile estimates for each of the holding periods.

We compared three different quantile estimators of the 1st, 5th, 25th, 75th, 95th and 99th quantiles. We chose several quantiles in the tails of the distribution as estimation of the tail is often considered of greatest importance. The quantile estimators that we used are based on ARCH estimates of the volatility. We acknowledge that better estimators may exist, however, we felt that it was sensible to employ straightforward and commonly used estimators in our study. In the next subsection, we describe the quantile estimators, and in the subsection that follows we present the results.

5.1. Quantile Estimation Methods

We fitted an ARMA-GARCH model to an initial data set of 1014 log returns using the common practice of maximising a Gaussian likelihood function. We did not find any significant ARMA terms. We concluded that the most suitable model was GARCH(1,1) with the

conditional mean assumed to be a constant. This is consistent with numerous analyses of exchange rate data in the literature (e.g. Jorion, 1995; Xu and Taylor, 1995; Andersen and Bollerslev, 1997). We produced quantile estimates based on the GARCH variance forecasts and a Gaussian distribution. A second GARCH quantile estimator was constructed from the GARCH variance forecasts and the empirical distribution of in-sample standardised residuals. A standardised 1-period return is calculated by dividing the 1-period residual by the 1-step-ahead volatility forecast, $\hat{\sigma}_{t+1}$. A standardised multiperiod residual is the k-period residual divided by the multiperiod volatility forecast, $\hat{\sigma}_{t,k}$.

We also estimated quantiles using the quantile regression neural network procedure described in the previous section. The *returns vector*, used as dependent variable, consisted of returns from holding periods of length 1, 3, 5, 7, 10, 12 and 15. We used GARCH(1,1) 1-step-ahead volatility forecasts, $\hat{\sigma}_{t+1}$, and the length of the holding period, k, as explanatory variables. We acknowledge that if the GARCH model is misspecified, then the quantile regression neural network approach will suffer. In view of this, there is a strong argument for using another method, such as a moving average, to produce the 1-step-ahead forecasts for the new approach. However, since GARCH(1,1) forecasts are the most obvious benchmark to use with exchange rate data, we felt that the simplest and least controversial option was to use GARCH(1,1) 1-step-ahead forecasts in our method. Furthermore, if we were to use 1-step-ahead forecasts from another method in the new approach, and the multiperiod forecasting results were found to be notably different to those of the GARCH(1,1) benchmark, we would wonder whether the difference was due largely to the choice of 1-step-ahead forecast used.

We applied separate tenfold cross-validation for each of the six quantiles to determine the optimal number of hidden units and values of the regularisation parameters, λ_1 and λ_2. This led to either one or two hidden units being used in the neural networks. This is consistent with several rules-of-thumb for the appropriate number of units; Kang (1991) suggested "$n/2$", and

Tang and Fishwick (1993) proposed "*n*", where *n* is the number of inputs. We did not re-estimate the neural network weights for each of the 1000 moving windows used in the study. Instead, we simply used the first window of 1014 daily observations. We felt that if the neural network quantile model was to be re-estimated, then best practice would dictate that the penalty function parameters and network architecture should also be re-estimated. This is clearly not practical in this kind of study. For consistency, we also did not re-estimate the GARCH parameters for each moving window. The quantile regression neural network minimisation in expression (6) was carried out in Gauss for UNIX. We used the quasi-Newton optimisation algorithm of Broyden, Fletcher, Goldfarb and Shanno (see Luenberger, 1984, page 269).

5.2. Post-Sample Results

As with volatility forecasting, the unobservable nature of quantiles implies that the conventional measures of forecast accuracy, such as MSE, are not directly applicable. The most popular measure of quantile estimator accuracy is the percentage of observations falling below the quantile estimator. For an unbiased estimator of the θth quantile, this will be $\theta\%$. This criterion is employed by Alexander and Leigh (1997) in a value at risk study, and is used extensively for evaluating quantile estimators and prediction intervals (e.g. Granger *et al.*, 1989; Taylor and Bunn, 1999; Taylor, 1999b). In this paper, we use this measure as a basis for comparison of the three estimators.

Table I compares estimation of the 1st, 5th, 25th, 75th, 95th and 99th quantiles for the 1000 post-sample German deutsche mark returns at the different holding periods. The table shows the percentage of post-sample returns falling below the quantile estimators. The asterisks indicate the entries that are significantly different from the ideal value at the 5% significance level. The acceptance region for the hypothesis test is constructed using a Gaussian distribution and the standard error formula for a proportion. We have highlighted in bold the best measures

for each quantile at each holding period. Table II compares post-sample estimation for the Japanese yen.

<center>---------- **TABLES I & II** ----------</center>

Table I shows that for the deutsche mark data, the GARCH volatility estimator with empirical distribution performs very well for the 5th, 25th 75th and 95th quantiles. The quantile regression method matches the GARCH model with empirical distribution for the 5th and 95th quantiles, and outperforms it for the 99th. The GARCH model with Gaussian distribution appears to be the overall winner only for the 1st quantile. Interestingly, all three methods severely underestimate the 25th quantiles.

Table II shows that for the yen data the quantile regression method outperforms the other two methods for four of the six quantiles. The GARCH model with empirical distribution is the most successful for the 75th quantile and the GARCH model with Gaussian distribution performs the best for the 99th quantile.

In order to give some indication of the relative overall performance for the three methods at the different holding periods, we calculated chi-squared goodness of fit statistics (see Hull and White, 1998). For each method, at each holding period, we calculated the statistic for the total number of post-sample German deutsche mark and Japanese yen returns falling within the following seven categories: below the 1st quantile estimator, between the 1st and 5th estimators, between the 5th and 25th, between the 25th and 75th, between the 75th and 95th, between the 95th and 99th, and above the 99th. Table III shows the resulting chi-squared statistics. The table shows that the GARCH model with empirical distribution performs the worst. The GARCH model with Gaussian distribution is better than the quantile regression approach for four of the seven holding periods, but overall there is little to choose between the two. Unfortunately, we cannot sum the chi-squared statistics across holding periods to give a single summary measure for each of the three methods because these

<center>15</center>

statistics are not independent. To provide another indication of the relative overall performance, we summed the number of times that each method outperformed the other two in Tables I and II. These results, which are reported in Table IV, confirm our conclusion from Table III that the quantile regression method is very competitive.

---------- TABLE III & IV ----------

In our initial empirical work with the deutsche mark returns, we used just 436 observations to estimate the neural network quantile regression model. However, the results for 1-step-ahead quantile estimation were poor so we decided to expand the estimation data set to 1014. The need for a large estimation data set is, perhaps, not surprising since both neural networks and empirical approaches to density estimation generally require sizeable amounts of data.

6. SUMMARY AND CONCLUSIONS

We have presented a nonparametric approach to estimating the conditional density of multiperiod returns. The method uses historical returns from a range of different holding periods and produces quantile models which are functions of the holding period, k, and the 1-step-ahead volatility forecast, $\hat{\sigma}_{t+1}$, as suggested by theoretically derived variance expressions. We avoided the need to specify appropriate explanatory variables by using an artificial neural network to estimate the nonlinear quantile models. Using exchange rate data, we performed comparative analysis which gave encouraging results. Research investigating the usefulness of the method for other exchange rates is currently in progress. Although in this paper we used 1-step-ahead volatility forecasts from a GARCH(1,1) model as input to the quantile regression method, other 1-step-ahead volatility forecasting methods could certainly be used for exchange rates and for other applications.

ACKNOWLEDGEMENTS

REFERENCES

Alexander, C.O. and Leigh, C.T., 'On the Covariances Matrices Used in Value at Risk Models', *Journal of Derivatives*, **4** Spring (1997), 50-62.

Andersen, T.G. and Bollerslev, T., 'Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts', Working Paper 6023, National Bureau of Economic Research, 1997.

Baillie, R.T. and Bollerslev, T., 'The Message in Daily Exchange Rates: A Conditional Variance Tale', *Journal of Business and Economic Statistics*, **7** (1989), 297-305.

Baillie, R.T. and Bollerslev, T., 'Prediction in Dynamic Models with Time-Dependent Conditional Variances', *Journal of Econometrics*, **52** (1992), 91-114.

Bishop, C.M., *Neural Networks for Pattern Recognition*, Oxford, Oxford University Press. 1997.

Bollerslev, T., 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics*, **31** (1986), 307-327.

Boudoukh, J., Richardson, M. and Whitelaw, R.F., 'Investigation of a Class of Volatility Estimators', *Journal of Derivatives*, **4** Spring (1997), 63-71.

Burgess, A.N., 'Robust Financial Modelling by Combining Neural Network Estimators of Mean and Median', *Proceedings of Applied Decision Technologies*, UNICOM Seminars, Brunel University, London, UK, 1995.

Diebold, F.X., Hickman, A., Inoue, A. and Schuermann, T., 'Scale Models', *Risk*, January (1998), 104-107.

Donaldson, R.G. and Kamstra, M., 'Forecast Combining with Neural Networks', *Journal of Forecasting*, **15** (1996), 49-61.

Duffie, D., and Pan, J., 'An Overview of Value at Risk', *Journal of Derivatives*, **4** Spring (1997), 7-49.

Engle, R.F., 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica*, **50** (1982), 987-1008.

Gibson, M.S. and Boyer, B.H., 'Evaluating Forecasts of Correlation Using Option Pricing', *Journal of Derivatives*, **5** Winter (1998), 18-38.

Granger, C.W.J., White, H. and Kamstra, M., 'Interval Forecasting: An Analysis Based Upon ARCH-Quantile Estimators', *Journal of Econometrics*, **40**, 87-96, 1989.

Hansen, B.E., 'Autoregressive Conditional Density Estimation', *International Economic Review*, **35** (1994), 705-730.

Hull, J., and White, A., 'Value at Risk When Daily Changes in Market Variables Are Not Normally Distributed', *Journal of Derivatives*, **5** Spring (1998), 9-19.
Jorion, P., 'Predicting Volatility in the Foreign Exchange Market', *Journal of Finance*, 50 (1995), 507-528.

Jorion, P., *Value at Risk*, Illinois: Irwin, 1997.

Kang, S., *An Investigation of the Use of Feedforward Neural Networks for Forecasting*, Unpublished Doctoral Thesis, Kent State University, 1991.

Koenker, R.W. and Bassett, G.W., 'Regression Quantiles', *Econometrica*, **46** (1978), 33-50.

Koenker, R.W. and Bassett, G.W., 'Robust Tests for Heteroscedasticity Based on Regression Quantiles', *Econometrica*, **50** (1982), 43-62.

Kroner, K.F., Kneafsey, K.P. and Claessens, S., 'Forecasting Volatility in Commodity Markets', *Journal of Forecasting*, **14** (1995), 77-95.

Luenberger, D.G., *Linear and Nonlinear Programming*, Reading, MA: Addison-Wesley, 1984.

RiskMetrics^{TM} *Technical Document*, fourth edition. J.P. Morgan/Reuters, 1996.

Tang, Z. and P.A. Fishwick, 'Feedforward Neural Nets as Models for Time Series Forecasting', *ORSA Journal on Computing*, **5** (1993), 374-385.

Taylor, J.W., 'A Quantile Regression Approach to Estimating the Distribution of Multiperiod Returns', *Journal of Derivatives*, **7** Fall (1999a), 64-78.

Taylor, J.W., 'Evaluating Volatility and Interval Forecasts', *Journal of Forecasting*, **18** (1999b), 111-128.

Taylor, J.W. and Bunn, D.W., 'A Quantile Regression Approach to Generating Prediction Intervals', *Management Science*, **45** (1999).

White, H., 'Nonparametric Estimation of Conditional Quantiles Using Neural Networks', in *Artificial Neural Networks: Approximation and Learning Theory*, A.R. Gallant *et al.*, Cambridge and Oxford, Blackwell (1992), 191-205.

Xu, X., and Taylor, S.J., 'Conditional Volatility and the Informational Efficiency of the PHLX Currency Options Market', *Journal of Banking and Finance*, **19** (1995), 803-821.

Yar, M., and Chatfield, C., 'Prediction Intervals for the Holt-Winters Forecasting Procedure', *International Journal of Forecasting*, **6** (1990), 127-137.

Zhang, G., Patuwo, B.E. and Hu, M.Y., 'Forecasting with Artificial Neural Networks: The State of the Art', *International Journal of Forecasting*, **14** (1998), 35-62.

Table I. Percentage of post-sample German deutsche mark returns falling below quantile estimates

		Holding Period						
		1	3	5	7	10	12	15
1st	GARCH(1,1) & Gaussian	1.3	**1.3**	**0.8**	**0.9**	**0.3***	0.4	0.4
	GARCH(1,1) & Empirical	**1.0**	1.6	2.2*	3.0*	3.2*	2.3*	1.7*
	Quantile Regression	0.6	2.0*	2.3*	2.9*	2.1*	**1.3**	**0.9**
5th	GARCH(1,1) & Gaussian	4.0	4.0	3.9	4.1	4.3	3.1*	2.3*
	GARCH(1,1) & Empirical	**4.1**	**4.9**	5.3	5.2	5.9	5.7	**4.6**
	Quantile Regression	3.3*	4.6	**5.2**	**5.0**	**5.6**	**5.0**	4.3
25th	GARCH(1,1) & Gaussian	19.0*	19.1*	17.9*	18.4*	16.7*	16.8*	17.7*
	GARCH(1,1) & Empirical	**22.3***	**21.0***	19.9*	**20.0***	**17.8***	**17.6***	**18.4***
	Quantile Regression	19.2*	**21.0***	**20.5***	19.2*	17.1*	17.4*	17.0*
75th	GARCH(1,1) & Gaussian	77.6	**76.5**	76.7	74.1	73.8	72.7	72.8
	GARCH(1,1) & Empirical	**74.4**	**76.5**	**75.8**	**74.8**	**74.0**	**75.0**	**74.4**
	Quantile Regression	76.2	**73.5**	73.5	72.4	73.8	74.3	76.2
95th	GARCH(1,1) & Gaussian	96.0	95.4	**94.8**	**94.9**	93.7	94.2	94.4
	GARCH(1,1) & Empirical	95.7	**95.3**	94.7	**94.9**	**94.7**	**94.9**	94.6
	Quantile Regression	**95.4**	94.1	93.6*	**94.9**	94.2	94.6	**95.3**
99th	GARCH(1,1) & Gaussian	98.5	98.7	98.6	98.4	98.5	98.5	**98.6**
	GARCH(1,1) & Empirical	**98.9**	**98.8**	98.6	**98.5**	98.4	98.3*	98.5
	Quantile Regression	99.2	98.7	**98.7**	**98.5**	99.4	99.1	99.4

* significant at 5% level
bold indicates best performing method for each quantile at each holding period

Table II. Percentage of post-sample Japanese yen returns falling below quantile estimates

		Holding Period						
		1	3	5	7	10	12	15
1st	GARCH(1,1) & Gaussian	2.0*	2.5*	2.3*	2.4*	1.8*	2.2*	2.9*
	GARCH(1,1) & Empirical	**1.0**	2.5*	2.7*	2.8*	3.3*	3.1*	4.2*
	Quantile Regression	0.7	**1.7***	**1.4**	**1.2**	**1.3**	**1.8***	**2.3***
5th	GARCH(1,1) & Gaussian	**4.9**	5.7	**6.2**	**6.8***	6.5*	6.8*	7.3*
	GARCH(1,1) & Empirical	4.7	5.8	6.8*	7.2*	7.6*	7.9*	8.7*
	Quantile Regression	3.2*	**5.4**	6.5*	**6.8***	**5.5**	**4.7**	**5.1**
25th	GARCH(1,1) & Gaussian	21.4*	21.6*	22.6	22.6	**25.0**	**25.3**	26.4
	GARCH(1,1) & Empirical	**24.8**	23.0	23.7	22.4	24.8	23.9	**24.4**
	Quantile Regression	21.0*	**24.9**	**25.9**	**26.0**	27.0	26.9	26.8
75th	GARCH(1,1) & Gaussian	78.2*	78.1*	79.1*	79.4*	77.4	**76.2**	**76.1**
	GARCH(1,1) & Empirical	**73.8**	75.9	**74.7**	**74.8**	**73.4**	73.2	72.1*
	Quantile Regression	77.7*	**75.4**	77.1	77.8*	77.7*	77.5	77.3
95th	GARCH(1,1) & Gaussian	95.8	96.9*	96.1	95.7	95.4	96.5*	97.1**
	GARCH(1,1) & Empirical	**94.6**	93.5*	93.5*	92.9*	92.6*	93.1*	**93.5***
	Quantile Regression	95.8	**94.9**	**95.3**	**95.1**	**95.1**	**95.5**	96.5*
99th	GARCH(1,1) & Gaussian	98.2*	**98.9**	**98.8**	**99.1**	**98.8**	**99.3**	**99.2**
	GARCH(1,1) & Empirical	**98.7**	98.1*	97.4*	96.6*	97.0*	97.1*	97.5*
	Quantile Regression	99.5	98.5	98.0*	97.5*	98.0*	98.2*	98.2*

* significant at 5% level
bold indicates best performing method for each quantile at each holding period

20

Table III. Chi-squared goodness of fit statistics summarising the overall performance of the three estimators for the post-sample German deutsche mark and Japanese yen returns

	Holding Period						
	1	3	5	7	10	12	15
GARCH(1,1) & Gaussian	85.9^*	82.6^*	80.4^*	$\mathbf{74.9^*}$	$\mathbf{97.0^*}$	$\mathbf{59.6^*}$	$\mathbf{80.2^*}$
GARCH(1,1) & Empirical	$\mathbf{8.7}$	61.4^*	100.5^*	174.2^*	195.4^*	157.3^*	173.7^*
Quantile Regression	47.3^*	$\mathbf{32.9^*}$	$\mathbf{57.6^*}$	110.3^*	127.3^*	64.8^*	88.5^*

* significant at 5% level
bold indicates best performing method for each holding period

Table IV. Number of times an estimator outperformed the others for the post-sample German deutsche mark and Japanese yen returns

	Deutsche Mark (Table I summary)	Japanese Yen (Table II summary)	Total
GARCH(1,1) & Gaussian	8	13	21
GARCH(1,1) & Empirical	**24**	10	34
Quantile Regression	17	**21**	**38**

bold indicates best performing method