Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall

James W. Taylor

Saïd Business School

University of Oxford

Journal of Financial Econometrics, 2008, Vol. 6, pp. 382-406.

Address for Correspondence:

James W. Taylor
Saïd Business School
University of Oxford
Park End Street
Oxford OX1 1HP, UK

Tel: +44 (0)1865 288927
Fax: +44 (0)1865 288805
Email: james.taylor@sbs.ox.ac.uk

Using Exponentially Weighted Quantile Regression to Estimate Value at Risk and Expected Shortfall

Abstract

We propose exponentially weighted quantile regression (EWQR) for estimating time-varying quantiles. The EWQR cost function can be used as the basis for estimating the time-varying expected shortfall associated with the EWQR quantile forecast. We express EWQR in a kernel estimation framework, and then modify it by adapting a previously proposed double kernel estimator in order to provide greater accuracy for tail quantiles that are changing relatively quickly over time. We introduce double kernel quantile regression, which extends the double kernel idea to the modelling of quantiles in terms of regressors. In our empirical study of 10 stock returns series, the versions of the new methods that do not accommodate the leverage effect were able to outperform GARCH-based methods and CAViaR models.

Key words: Financial Risk; Exponential Weighting; Quantile Regression; Kernel Smoothing; Kernel Density Estimation.

JEL: C22, C53, G10

The accurate assessment of the exposure to market risk of a financial institution is of great importance for internal risk control and financial regulation. Value at risk (VaR) has become the standard approach to quantifying market risk. It measures the maximum potential loss of a given portfolio over a prescribed holding period at a given confidence level, which is typically chosen to be 1% or 5%. Therefore, estimating the VaR amounts to forecasting, conditional on current information, the tail quantiles of the distribution of a series of financial returns. Although a variety of approaches have been proposed for forecasting conditional tail quantiles, there is no established method. GARCH methods are popular but they can be criticised because there may be error in the specification of the variance model and in the choice of distribution used (Manganelli and Engle 2004).

Exponential smoothing is a simple and pragmatic approach to forecasting whereby the forecast is constructed from an exponentially weighted average of past observations. The common use of exponential smoothing for volatility prediction motivates the development of the approach for quantile forecasting. In this paper, we consider the forecasting of quantiles using methods based on an exponential weighting of past data. We introduce exponentially weighted quantile regression (EWQR), which we show amounts to exponential smoothing of the cumulative distribution function (cdf). We point out that this nonparametric method can be viewed in a kernel framework. If the distribution of returns is changing relatively quickly over time, a relatively fast exponential decay is needed to ensure swift adapting. However, a fast decay in the EWQR method is analogous to the use of a low number of observations to construct a histogram. When few observations are available, kernel density estimation can offer an improvement on the density estimate provided by a histogram. We incorporate kernel density estimation within the EWQR method through the use of an exponentially weighted double kernel method adapted from the double kernel estimator of Yu and Jones (1998).

As a measure of financial risk, VaR has the disadvantage that it reports only a quantile, and thus disregards outcomes beyond the quantile. An alternative measure of risk that overcomes this weakness is expected shortfall (ES), which is defined as the expectation of the return given that it exceeds the VaR. ES also has the appeal of being a coherent measure of risk (see Artzner, Delbaen, Eber and Heath 1999). We show that the EWQR cost function can be used as the basis for estimating the time-varying ES associated with a EWQR quantile forecast.

Section 1 provides a brief overview of the literature on VaR estimation. Section 2 presents the new EWQR approach to estimating VaR, and describes how it can also be used for estimating ES. Section 3 introduces the double kernel version of EWQR. Section 4 uses 10 series of stock returns to illustrate implementation of the new methods, and to compare their VaR and ES estimation accuracy to established methods. Section 5 provides a summary and concluding comments.

1. Methods for Estimating VaR and ES

Recent reviews of the VaR literature are provided by Manganelli and Engle (2004) and Kuester, Mittnik and Paolella (2006). Manganelli and Engle divide VaR methods into three different categories: parametric, semiparametric and nonparametric. *Parametric* approaches involve a parameterisation of the time-varying stochastic behaviour of financial prices. Conditional quantile forecasts are constructed from a conditional volatility forecast and a distributional assumption. Typically, exponential smoothing or a GARCH model is used to forecast the volatility (see Poon and Granger 2003), and a Gaussian or Student-*t* distribution is assumed. For these distributions, analytical formulae exist for the calculation of the ES (see McNeil, Frey and Embrechts 2005, Section 2.2.4).

Semiparametric VaR approaches include those based on extreme value analysis, such as the method of McNeil and Frey (2000), which involves the peaks over threshold EVT method being applied to residuals standardised by GARCH conditional volatility estimates. McNeil, Frey and Embrechts (2005, p. 283) provide the analytical formula for the associated ES estimation. Also included in the semiparametric category of VaR methods are those based on the use of quantile regression, such as the conditional autoregressive value at risk (CAViaR) models of Engle and Manganelli (2004). Their four CAViaR models are presented in the following expressions:

Adaptive CAViaR: $$Q_t(\theta) = Q_{t-1}(\theta) + \alpha\left(\theta - I\left(y_{t-1} < Q_{t-1}(\theta)\right)\right)$$

Symmetric Absolute Value CAViaR: $$Q_t(\theta) = \omega + \alpha\, Q_{t-1}(\theta) + \beta\left|y_{t-1}\right|$$

Asymmetric Slope CAViaR: $$Q_t(\theta) = \omega + \alpha\, Q_{t-1}(\theta) + \beta_1 (y_{t-1})^{+} + \beta_2 (y_{t-1})^{-}$$

Indirect GARCH(1,1) CAViaR: $$Q_t(\theta) = \left(1 - 2\, I(\theta < 0.5)\right)\left(\omega + \alpha\, Q_{t-1}(\theta)^2 + \beta\, y_{t-1}^2\right)^{\frac{1}{2}}$$

where $Q_t(\theta)$ is the conditional θ quantile; ω, α, β and β_i are parameters; and $(x)^+ = \max(x,0)$ and $(x)^- = -\min(x,0)$. Note that we are modelling here a residual term, y_t, defined as $y_t = r_t - E(r_t|I_{t-1})$, where r_t is the return and $E(r_t|I_{t-1})$ is a conditional mean term, which is often assumed to be zero or a constant. CAViaR model parameters are estimated using the quantile regression minimisation, which is presented in expression (3) of Section 2.2. Although direct quantile modelling is an appealing feature of CAViaR models, it leads to the disadvantage that it is not clear how to calculate the corresponding ES. This prompts Taylor (2008) to present conditional autoregressive expectile models, which can be used as the basis for VaR and ES estimation.

In expressions (1) and (2), we show the adaptive method of Gorr and Hsu (1985), which was developed for a variety of applications. The formulation is similar to the adaptive CAViaR method, except that in expression (1), the Gorr and Hsu method softens the impact of the indicator function through the use of exponential smoothing. Note that if β=1, the two methods are identical.

$$Q_t(\theta) = Q_{t-1}(\theta) + \alpha\left(\theta - \hat{\theta}_{t-1}\right) \qquad (1)$$

where
$$\hat{\theta}_{t-1} = \beta I\left(y_{t-1} < Q_{t-1}(\theta)\right) + (1-\beta)\hat{\theta}_{t-2} \qquad (2)$$

Turning to *nonparametric* VaR methods, the most widely used is historical simulation, which estimates the VaR as the quantile of the empirical distribution of returns in a moving window of the most recent periods. For this method, it seems natural to estimate the ES as the mean of the returns, in the moving window, that exceed the VaR estimate. A problem is that it is not obvious how many past periods to include in the moving window. A small number would enable swift reaction to changes in the true distribution, but too few observations will lead to large sampling error. To overcome this problem, Boudoukh, Richardson and Whitelaw (1998) propose, for quantile estimation, the analogy of the exponentially weighted moving average volatility forecasting method. We term this the BRW method. It involves allocating to the sample of returns, exponentially decreasing weights, which sum to one. The returns are then ordered in ascending order and, starting at the lowest return, the weights are summed until θ is reached. The forecast of the θ quantile is set as the return that corresponds to the final weight used in the previous summation. The authors give no consideration as to how to derive ES forecasts from the BRW method. We consider the method again in Section 2.3.

2. Estimating VaR Using Exponentially Weighted Quantile Regression

2.1. Exponentially Weighted Least Squares Regression

Before introducing exponentially weighted quantile regression, let us consider the established practice of exponentially weighted least squares (EWLS) regression, which is also known as discounted least squares. For a model with intercept, *m*, but no regressors, EWLS is written as:

$$\min_{m} \sum_{t=1}^{T} \lambda^{T-t} \left(y_t - m\right)^2$$

where T is the sample size, and $\lambda \in [0,1]$ is a weighting parameter. It is well known that differentiation with respect to m leads to the following estimate, which is an exponentially weighted average.

$$m_T = \frac{\sum_{t=1}^{T} \lambda^{T-t} \; y_t}{\sum_{t=1}^{T} \lambda^{T-t}}$$

If T is large, this can be written in the recursive form of simple exponential smoothing:

$$m_T = (1-\lambda)\, y_T + \lambda\, m_{T-1}$$

2.2. Quantile Regression

Koenker and Bassett (1978) introduce quantile regression for the estimation of linear quantile models, $Q_t(\theta) = \boldsymbol{x}_t'\boldsymbol{\beta}$, where $\boldsymbol{x}_t$ is a vector of regressors and $\boldsymbol{\beta}$ is a parameter vector. The quantile regression minimisation is shown in expression (3). It is conveniently solved as a linear program.

$$\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \left(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}\right)\left(\theta - I\left(y_t < \boldsymbol{x}_t'\boldsymbol{\beta}\right)\right) \tag{3}$$

Koenker and Bassett show that, if the model includes an intercept term, the resulting quantile estimator, $\hat{Q}_t(\theta) = \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}$, obeys the partitioning inequalities in expression (4). The inequalities indicate that the estimator essentially partitions the y_t observations so that the proportion less than the corresponding quantile estimate is close to θ. (The proportion is not exactly θ because, in general, quantile regression has p residuals equal to zero, where p is the dimension of the parameter vector $\boldsymbol{\beta}$.)

$$\frac{1}{T}\sum_{t=1}^{T} I\left(y_t < \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}\right) \le \theta \qquad \text{and} \qquad \frac{1}{T}\sum_{t=1}^{T} I\left(y_t > \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}\right) \le (1-\theta) \tag{4}$$

2.3. Exponentially Weighted Quantile Regression

Our development of quantile regression is to propose exponentially weighted quantile regression (EWQR), which could also be referred to as discounted quantile regression. For a specified value of the weighting parameter, λ, the EWQR minimisation has the form:

$$\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \lambda^{T-t} \left(y_t - \mathbf{x}_t'\boldsymbol{\beta}\right)\left(\theta - I\left(y_t < \mathbf{x}_t'\boldsymbol{\beta}\right)\right) \tag{5}$$

For a linear quantile model, this minimisation can be formulated as a linear program. Koenker and Bassett (1978) derive the partitioning inequalities of expression (4) in Theorem 3.4 of their paper. We now present a new analogous theorem for EWQR.

Theorem 1. If the quantile model, $Q_t(\theta) = \mathbf{x}_t'\boldsymbol{\beta}$, includes an intercept term, the solution $\hat{\boldsymbol{\beta}}$ to the EWQR minimisation of expression (5) will satisfy the following inequalities:

$$\frac{\sum_{t=1}^{T} \lambda^{T-t} I\left(y_t < \mathbf{x}_t'\hat{\boldsymbol{\beta}}\right)}{\sum_{t=1}^{T} \lambda^{T-t}} \leq \theta \qquad \text{and} \qquad \frac{\sum_{t=1}^{T} \lambda^{T-t} I\left(y_t > \mathbf{x}_t'\hat{\boldsymbol{\beta}}\right)}{\sum_{t=1}^{T} \lambda^{T-t}} \leq (1-\theta) \tag{6}$$

Proof. See Appendix.

These partitioning inequalities show that the EWQR quantile estimator, $\hat{Q}_t(\theta) = \mathbf{x}_t'\hat{\boldsymbol{\beta}}$, partitions the y_t observations so that the sum of the weights on those observations less than the corresponding quantile estimator, as a proportion of the sum of all the weights, is close to θ. (The proportion is not exactly θ because, in general, EWQR has p residuals equal to zero, where p is the dimension of $\boldsymbol{\beta}$.) Therefore, EWQR delivers a quantile estimator for which the exponentially weighted average of the indicator function is close to θ. This is shown in expression (7) for the simple case of EWQR with an intercept, q, and no regressors.

$$\frac{\sum_{t=1}^{T} \lambda^{T-t} I\left(y_t < \hat{q}\right)}{\sum_{t=1}^{T} \lambda^{T-t}} \leq \theta \qquad \text{and} \qquad \frac{\sum_{t=1}^{T} \lambda^{T-t} I\left(y_t > \hat{q}\right)}{\sum_{t=1}^{T} \lambda^{T-t}} \leq (1-\theta) \tag{7}$$

Viewing θ as the target value of the cdf, expression (7) suggests the following cdf estimator for a specified value, y, in period T:

$$\hat{F}_T(y) = \frac{\sum_{t=1}^{T} \lambda^{T-t}\, I(y_t < y)}{\sum_{t=1}^{T} \lambda^{T-t}} \tag{8}$$

In a similar way to that of Section 2.1, if T is large, we can write expression (8) in recursive form as:

$$\hat{F}_T(y) = (1-\lambda)I(y_T < y) + \lambda\, \hat{F}_{T-1}(y) \tag{9}$$

This expression is simple exponential smoothing of the cdf for a given y value.

Although expression (8) is a cdf estimator, it can be used to estimate quantiles by iteratively evaluating the right hand side of the expression for different values of y until the desired value for the cdf estimator, $\hat{F}_T(y)$, is achieved to a required degree of tolerance. The same quantile estimate is derived by using EWQR with an intercept and no regressors. This shows that EWQR encompasses simple exponential smoothing of the cdf.

For univariate estimation of the time-varying quantiles of a time series, the use of EWQR with an intercept and no regressors is one possibility. If a time series is trending, or exhibiting substantial increase or decrease in its volatility, a trend term could be included in the EWQR. This is analogous to the inclusion of a trend term in the EWLS regression, which leads to Brown's (1963) double exponential smoothing. Using EWLS to fit models that are certain functions of time is termed 'general exponential smoothing' (GES) (see Gardner 1985). The inclusion of functions of time in EWQR is, therefore, the extension of GES to time-varying quantile forecasting. Following practice in GES, if the data is seasonal, sinusoidal terms or dummy variables can be included in the EWQR. For VaR estimation, the inclusion in EWQR of an intercept with no regressors seems reasonable, but a regressor could certainly be included in order to attempt to capture the leverage effect.

Forecasting the time-varying quantile using expression (8) is equivalent to the BRW VaR method described in Section 1. The BRW method is, therefore, equivalent to EWQR with an intercept and no regressors. The benefit in recognising this is threefold. Firstly, it gives the BRW method a sounder theoretical basis. Secondly, the regression framework enables the inclusion of regressors in the BRW method. Thirdly, the formal statistical framework of EWQR allows the possibility of statistical testing of parameters. Encouragingly for the EWQR method, Manganelli and Engle (2004) describe the BRW method as being a significant improvement over other simple VaR methods since it drastically simplifies the assumptions needed in the parametric models and it incorporate a more flexible specification than the historical simulation approach. They go on to conclude that it is a special case of their CAViaR class of methods. However, in our view, exponentially weighting the quantile regression minimisation is a fundamentally different approach to CAViaR, and so it is more reasonable to say simply that both approaches are special cases of quantile regression.

In Section 1, we presented the Gorr and Hsu (1985) method. Interestingly, expression (2) in their method is equivalent to expression (9). Note also that expression (8) is used by Dunsmuir, Scott and Qiu (1996) within a robust approach to point forecasting.

2.4. Expected Shortfall Using EWQR

Theorem 8.3 of Koenker (2005) shows that, for a variable *y*, the ES for the θ quantile in the lower tail of the distribution is given by the following expression:

$$ES = E(y) - \frac{1}{\theta} E\left(\left(y - Q(\theta) \right)\left(\theta - I\left(y < Q(\theta) \right) \right) \right) \tag{10}$$

where $Q(\theta)$ is the θ quantile of *y*. As with y_t in Section 1, let us define *y* to be a zero mean residual term. This enables us to rewrite expression (10) as the following:

$$ES = -\frac{1}{\theta} E\left(\left(y - Q(\theta) \right)\left(\theta - I\left(y < Q(\theta) \right) \right) \right) \tag{11}$$

This expression involves the expectation of the asymmetric ‘tick’ function used in the quantile regression minimisation of expression (3). The tick function of expression (11) is evaluated at the quantile, $Q(\theta)$, which can be estimated by the quantile regression minimisation. Therefore, a sample estimator of the expectation is the optimised value of the quantile regression objective function divided by the sample size. Using this in expression (11) gives the following ES estimator:

$$\hat{ES} = -\frac{1}{\theta T} \sum_{t=1}^{T} \left(y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}} \right)\left(\theta - I\left(y_t < \mathbf{x}_t' \hat{\boldsymbol{\beta}} \right) \right) \tag{12}$$

Quantile regression would, therefore, seem to deliver not only an estimator for the quantile, but also, via expression (12), an ES estimator. Unfortunately, this estimator is of limited use because, in view of the time-varying nature of the distribution of financial returns, we require an ES estimator that is time-varying. In other words, expression (12) provides an unconditional estimator, but what we need is an estimator conditional on information up to the current period. We address this by proposing an exponentially weighted time-varying ES estimator that puts a larger weight on the contribution of more recent observations. To achieve this, in expression (12), we simply replace the quantile regression objective function with the EWQR objective function of expression (5). Therefore, for quantiles in the lower tail, we propose that for quantile models, $Q_t(\theta) = \boldsymbol{x}_t'\boldsymbol{\beta}$, estimated using EWQR, the corresponding ES be estimated using the following new adaptation of expression (12):

$$\hat{ES}_T = -\frac{1}{\theta\sum_{t=1}^{T}\lambda^{T-t}}\sum_{t=1}^{T}\lambda^{T-t}\left(y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}\right)\left(\theta - I\left(y_t < \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}\right)\right) \tag{13}$$

For a θ quantile in the upper tail of the distribution, the analogous expression is:

$$\hat{ES}_T = \frac{1}{(1-\theta)\sum_{t=1}^{T}\lambda^{T-t}}\sum_{t=1}^{T}\lambda^{T-t}\left(y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}\right)\left(\theta - I\left(y_t < \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}\right)\right) \tag{14}$$

3. Exponentially Weighted Double Kernel Quantile Regression

3.1. Viewing EWQR as a form of Kernel Quantile Regression

Jones and Hall (1990) consider a kernel weighting scheme, as in expression (15), for the nonparametric estimation of quantiles of y_t conditional upon a scalar variable x_t.

$$\min_q \sum_{t=1}^{T} K_h(x - x_t)(y_t - q)(\theta - I(y_t < q)) \tag{15}$$

where $K_h(x\text{-}x_t)$ is a conveniently defined kernel weighting function. The use of an intercept, q, with no regressors in expression (15) implies local constant fitting. Local linear fitting is another possibility that is considered by Yu and Jones (1998). Following similar steps to those in the proof of Theorem 1, it can be shown that the quantile estimated by expression (15) satisfies the following partitioning inequalities.

$$\frac{1}{\sum_{t=1}^{T}K_h(x-x_t)}\sum_{t=1}^{T}K_h(x-x_t)\,I(y_t<\hat{q}) \le \theta \qquad \text{and} \qquad \frac{1}{\sum_{t=1}^{T}K_h(x-x_t)}\sum_{t=1}^{T}K_h(x-x_t)\,I(y_t<\hat{q}) \le (1-\theta)$$

These inequalities suggest the following cdf estimator for a specified value, y, in period T:

$$\hat{F}_T(y) = \frac{1}{\sum_{t=1}^{T} K_h(x-x_t)} \sum_{t=1}^{T} K_h(x-x_t)\, I(y_t < y) \tag{16}$$

This is the standard kernel estimator for the cdf, which has been the focus of many studies (e.g. Abberger 1997; Hall, Wolff and Yao 1999; Cai 2002).

Consider the application of the kernel estimator of expressions (15) and (16) to the univariate time series context. Let $x_t = t$, and the location $x=T$, which is the most recent period. If we then define the kernel function to be one-sided with exponentially declining weight on data to the left of the location, T, we have $K_h(x\text{-}x_t)=\lambda^{T-t}$. Substituting this into expression (15) delivers EWQR with an intercept and no regressors, and substitution of the same term into (16) gives the EWQR expression (8). This shows that EWQR can be viewed as a form of kernel quantile estimation. This is consistent with the work of Gijbels, Pope and Wand (1999) who show that GES can be viewed in a kernel (least squares) regression framework. Therefore, our work can be viewed as extending the study of Gijbels, Pope and Wand to quantile forecasting.

3.2. An Exponentially Weighted Double Kernel CDF Estimator

The choice of bandwidth is a crucial issue for kernel estimators. For the exponentially weighted kernel, this problem translates into the choice of λ. In Section 4.1, we describe the procedure that we used to optimise λ. If the distribution of returns is changing relatively quickly over time, a relatively low value of λ is needed to ensure swift adapting. However, it seems intuitive that for tail quantiles the value of λ in the EWQR method must be relatively large in order that a relatively sizeable weight is given to many observations. The use of a low value of λ in the EWQR method is analogous to the use of a low number of observations to construct a histogram. When few observations are available, kernel density estimation often provides an improvement on the density estimate given by the histogram. Butler and Schachter (1998) extend the historical simulation VaR approach by applying kernel density estimation to a histogram of past returns. This method is the focus of Chen and Tang (2005) who consider standard errors for the resulting VaR estimates.

Kernel density estimation can be incorporated within the EWQR cdf estimator through the use of the following double kernel cdf estimator of Yu and Jones (1998):

$$\hat{F}_T(y) = \frac{1}{\sum_{t=1}^{T} K_{h_1}(x - x_t)} \sum_{t=1}^{T} K_{h_1}(x - x_t)\,\Omega_{h_2}(y - y_t) \tag{17}$$

where $\Omega_{h_2}(y - y_t) = \int_{-\infty}^{y} W_{h_2}(u - y_t)\,du$

This cdf estimator replaces the indicator function of the more standard estimator in expression (16) with a continuous distribution function, Ω_{h_2}. Yu and Jones describe the kernels as having two distinct bandwidths, one in the *y* direction and one in the *x* direction. They select W_{h_2} and K_{h_1} to be uniform and local linear kernels, respectively. Note that the standard kernel density estimator, which amounts to the smoothing of a histogram, corresponds to $K_{h_1} = 1$. In this paper, we propose the following new exponentially weighted double kernel cdf estimator:

$$\hat{F}_T(y) = \frac{1}{\sum_{t=1}^{T} \lambda^{T-t}} \sum_{t=1}^{T} \lambda^{T-t}\,\Omega_{h_2}(y - y_t) \tag{18}$$

where $\Omega_{h_2}(y - y_t) = \int_{-\infty}^{y} W_{h_2}(u - y_t)\,du$

The kernel W_{h_2} could be defined as being uniform, as in the work of Yu and Jones, with Gaussian and Epanechnikov being two other obvious possibilities. We selected the value of λ and the bandwidth h_2 for the kernel W_{h_2} using a procedure described in Section 4.1. As we mentioned in Section 2.3 with regard to expression (8), expression (18) can be used to estimate quantiles by repeatedly evaluating the right hand side of the expression for different values of *y*. This double kernel approach addresses the suggestion of Fan and Gu (2003) that a combination of both time-domain and state-domain smoothing of volatility is an interesting direction for future research. An interesting extension of the theoretical work of Chen and Tang (2005) would be to derive standard errors for the VaR estimates resulting from this new double kernel method.

3.3. Double Kernel Quantile Regression

In Section 2.3, we mentioned that, in the context of univariate VaR estimation, regressors could be included in EWQR in order to try to model the leverage effect. Although the double kernel estimators of expressions (17) and (18) have some appeal, they do not allow quantile modelling in terms of regressors. The double kernel least squares regression of Yu and Jones (1998) does provide a modelling framework, but only for the cdf, and not for a quantile. In this section, we present the analogous regression framework for quantiles by introducing double kernel quantile regression.

In kernel density estimation, the observations, y_t, are essentially replaced by a kernel function, which we write as W_{h_2}, centred at each observation. If we do the same with the local constant fitting quantile regression in expression (15), we get:

$$\min_{q} \sum_{t=1}^{T} K_{h_1}(x - x_t)\left(\int_{-\infty}^{\infty} (y - q)(\theta - I(y < q)) W_{h_2}(y - y_t)\, dy \right) \tag{19}$$

If we differentiate the objective function of this minimisation with respect to q, we get the Yu and Jones double kernel estimator of expression (17). This result gives insight into the double kernel cdf estimator, and allows direct estimation of a quantile, rather than iterative derivation of the quantile from the cdf estimators of expressions (17) and (18). Furthermore, the result enables us to broaden the double kernel cdf estimator to one that models the quantile as a function of regressors. We do this in expression (20), which is a generalisation of expression (19) for the case of estimating a quantile model, $Q_t(\theta) = \boldsymbol{x}_t'\boldsymbol{\beta}$:

$$\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} K_{h_1}(x - x_t)\left(\int_{-\infty}^{\infty} (y - \boldsymbol{x}_t'\boldsymbol{\beta})(\theta - I(y < \boldsymbol{x}_t'\boldsymbol{\beta})) W_{h_2}(y - y_t)\, dy \right) \tag{20}$$

As an example, if we select K_{h_1} to be the same exponentially weighted kernel considered in Sections 3.1 and 3.2, and we specify W_{h_2} to be Gaussian, this minimisation becomes the following:

$$\min_{\boldsymbol{\beta}} \sum_{t=1}^{T} \lambda^{T-t} \left(\theta(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}) + (\boldsymbol{x}_t'\boldsymbol{\beta} - y_t)\Phi((\boldsymbol{x}_t'\boldsymbol{\beta} - y_t)/h_2) + h_2\phi((\boldsymbol{x}_t'\boldsymbol{\beta} - y_t)/h_2) \right) \tag{21}$$

where Φ and ϕ are the standard Gaussian cdf and probability density function, respectively. A non-linear optimisation algorithm is needed to solve this minimisation. We term this new approach exponentially weighted double kernel quantile regression (EWDKQR).

In this paper, we set the kernel K_{h_1} to be an exponential weighting scheme. In many other quantile estimation applications, this kernel is unlikely to be needed, and setting $K_{h_1}=1$ would be appropriate. However, the same may not be true for the other kernel W_{h_2}. Indeed, the inclusion of this kernel would seem to have relevance in other quantile regression applications where there is a lack of data, which is often the case when extreme tail quantiles are being estimated.

In Section 2.4, we described how the EWQR cost function can be used as the basis for estimating the time-varying ES. We can adapt the EWQR ES expressions (13) and (14), from Section 2.4, for the EWDKQR estimator of expression (21). The EWDKQR ES expressions for quantiles in the lower and upper tail are given by expressions (22) and (23), respectively.

$$\hat{ES} = -\frac{1}{\theta\sum_t \lambda^{T-t}} \sum_t \lambda^{T-t} \left(\theta(y_t - \mathbf{x}_t'\boldsymbol{\beta}) + (\mathbf{x}_t'\boldsymbol{\beta} - y_t)\Phi((\mathbf{x}_t'\boldsymbol{\beta} - y_t)/h_2) + h_2\phi((\mathbf{x}_t'\boldsymbol{\beta} - y_t)/h_2) \right) \qquad (22)$$

$$\hat{ES} = \frac{1}{(1-\theta)\sum_t \lambda^{T-t}} \sum_t \lambda^{T-t} \left(\theta(y_t - \mathbf{x}_t'\boldsymbol{\beta}) + (\mathbf{x}_t'\boldsymbol{\beta} - y_t)\Phi((\mathbf{x}_t'\boldsymbol{\beta} - y_t)/h_2) + h_2\phi((\mathbf{x}_t'\boldsymbol{\beta} - y_t)/h_2) \right) \qquad (23)$$

4. Empirical Study

In this section, we describe the implementation of the new exponentially weighted methods within a study that compared their accuracy to that of a variety of established methods. The study considered day-ahead forecasting of the 1%, 5%, 95% and 99% conditional quantiles and their associated ES. We chose these quantiles because they are widely considered in practice. The focus on day-ahead estimation is consistent with the holding period considered for internal risk control by most financial firms. We used daily log returns for the 10 individual S&P500 stocks that had highest market capitalisation at the end of April 2005. The stocks are listed in Table 1, in descending order of market capitalisation, along with values of skewness and excess kurtosis. The Procter and Gamble returns series contains a large outlier, and this is reflected in the large values for the skewness and excess kurtosis. Although multivariate quantile models are being developed (e.g. De Gooijer, Gannoun and Zerom 2006), in this study, we followed the common practice of treating each series independently.

---------- Table 1 ----------

The sample period used in our study consisted of 13 years of daily data, from 29 April 1992 to 29 April 2005. This period delivered 3393 log returns. We used the first 2893 returns to estimate method parameters and the remaining data to evaluate 500 post-sample day-ahead quantile estimates. Our use of 13 years of data, with 500 periods for post-sample evaluation, follows the procedure of Engle and Managanelli (2004) in their VaR study. Following common practice, we did not estimate models for the conditional mean of each series (see Poon and Granger 2003). For all 10 series, we subtracted from each return, r_t, the mean, μ, of the 2893 in-sample returns. The quantile estimation methods were applied to the resultant residuals, $y_t = r_t - \mu$.

4.1. Methods Used for Estimating VaR and ES

EWQR and EWDKQR Methods

We found that, with no regressors in the EWQR, the cdf estimator in expression (8) provided a faster means of estimating the quantile, for a given value of θ, than using linear programming to solve the EWQR minimisation in expression (5). We iteratively evaluated the right hand side of expression (8) for different values of y until the required value for the cdf, $\hat{F}_T(y)$, was delivered to a specified degree of tolerance. Of course, for quantile models with regressors, this approach is of no use, and the EWQR minimisation of expression (5) must be performed. In order to allow for the leverage effect, we considered the use of EWQR with an indicator variable defined to take a value of one if the value of y_t was negative in the previous period, and a value of zero otherwise. Other regressors could certainly be considered to model the leverage effect, but in this initial study of EWQR, we opted for simplicity. When presenting the results in Sections 4.2 to 4.4, we refer to the method with the regressor as "EWQR Leverage" and the method with no regressors simply as "EWQR". We use analogous terms to label our two EWDKQR methods, which are discussed below.

Each EWQR was performed using a moving window of just the most recent 250 observations. We experimented with more observations in each moving window, but performance of the method was not improved. Optimisation of the weighting parameter λ proceeded by the use of a rolling window of 250 observations to produce day-ahead quantile forecasts for each of the remaining observations in the estimation sample of 2893 observations. The value of λ deemed to be optimal was the value that produced day-ahead quantile forecasts leading to the minimum QR Sum, where QR Sum is defined as the summation in the standard form of quantile regression presented in expression (3). We computed the QR Sum over a grid of values for λ between 0.80 and 1, with a step size of 0.005. We performed the optimisation separately for each value of θ (i.e. for each different quantile). The resulting values for EWQR with no regressors are reported in Table 2. A relatively large value of λ implies that the older observations in the moving window of 250 are given a larger weighting than they would have received if a smaller value of λ had been used. Giving a sizeable weight to all 250 observations would seem to be more important for the more extreme tail quantiles because these quantiles require more observations for their estimation. It is, therefore, intuitive that the values of λ that we derived were generally greater for the 1% and 99% quantiles than the 5% and 95% quantiles.

---------- Tables 2 and 3 ----------

We implemented the EWDKQR methods of Section 3. We considered three choices for the kernel W_{h_2}: Gaussian, uniform and Epanechnikov. The results for the three were similar and so, for simplicity, we report only the results for the Gaussian kernel in Sections 4.2 to 4.4. For the most basic form of EWDKQR, which involves no regressors, we found that, similarly to EWQR, iterative derivation of the quantile from the cdf estimator of expression (18) was faster than solving the EWDKQR minimisation in expression (21) with no regressors. To allow for the leverage effect, we considered the use of EWDKQR with the same regressor that we had used in the EWQR method. For this method, we performed the EWDKQR minimisation of expression (21).

For the EWDKQR method, we derived the values of λ and the bandwidth h_2 for the kernel W_{h_2} using the same procedure used for the EWQR method. We computed the QR Sum over a grid of values for λ between 0.80 and 1, with a step size of 0.005, and for h_2 between zero and 0.02, with a step size of 0.0005. A value of zero for h_2 corresponds to the EWQR method. The resulting λ values are presented in Table 3 for the EWDKQR method with no regressors. It is interesting to see that all but two of the entries in Table 3 are less than or equal to the corresponding entries for the EWQR method in Table 2. Note also that, by contrast with the values for the EWQR method in Table 2, the values of λ for the EWDKQR method in Table 3 tend to be noticeably smaller for the 1% quantile than for the 5% and 95% quantiles. We infer from this that the 1% quantile changes more radically over time. This was not captured by the EWQR method, but the inclusion of the kernel density estimation allows it to be accommodated within the EWDKQR estimator. Figure 1 is a plot of the derived values for λ and h_2. The negative relationship between the two parameters is intuitive because a lower value of λ implies faster exponential decay, and hence less historical information is captured, and so there is a need for a greater degree of kernel density smoothing, which is manifested in a larger value of h_2.

---------- Figure 1 ----------

For the EWQR and EWDKQR methods, ES predictions were produced using, respectively expressions (13) and (14) from Section 2.4, and expressions (22) and (23) from Section 3.3.

Historical Simulation

We included in our study the historical simulation approach using a moving window of 250 periods. We also implemented the EWDKQR method with no regressors and λ=1. This amounts to Butler and Schachter's (1998) inclusion of kernel density estimation within the historical simulation approach. We refer to this method as "Kernel Historical Simulation".

Methods Based on Volatility Forecasts

We generated volatility forecasts by applying exponential smoothing to the squared residuals with parameter optimised by minimising the sum of squared day-ahead variance forecast errors. Conditional quantile and ES forecasts were produced using first a Gaussian distribution and then the method of McNeil and Frey (2000), which involves applying EVT to the standardised residuals.

We implemented the GARCH(1,1) model and the asymmetric GARCH(1,1) model of Glosten, Jagannathan and Runkle (1993), which we term GJRGARCH. Our choice of the (1,1) specification was based on our analysis of the initial in-sample period of 2893 returns and on the general popularity of this order for GARCH models. We derived the model parameters using maximum likelihood based on a Student-*t* distribution with optimised degrees of freedom. We produced quantile and ES forecasts using the Student-*t* distribution and the EVT method of McNeil and Frey (2000).

CAViaR Models

We estimated the four CAViaR models presented in Section 1 using a procedure similar to that described by Engle and Manganelli (2004). For each model, we first generated 10^5 vectors of parameters from a uniform random number generator between 0 and 1, or between -1 and 0, depending on the appropriate sign of the parameter. For each of the vectors, we then evaluated the QR Sum. The 10 vectors that produced the lowest values for the function were used as initial values in a quasi-Newton algorithm. The QR Sum was then calculated for each of the 10 resulting vectors, and the vector producing the lowest value of the QR Sum was chosen as the final parameter vector.

4.2. VaR Results

We evaluated the post-sample conditional quantile forecasts using the two measures employed by Engle and Manganelli (2004): the hit percentage and the dynamic quantile (DQ) test statistic. The hit percentage assesses the unconditional coverage of a θ conditional quantile estimator. It is the percentage of observations falling below the estimator. Ideally, the percentage should be θ. We examined significant difference from this ideal using a test based on the binomial distribution. The Engle and Manganelli DQ test evaluates the dynamic properties of a conditional quantile estimator. It involves the joint test of whether the hit variable, defined as $Hit_t = I\left(y_t \leq \hat{Q}_t(\theta)\right) - \theta$, is distributed i.i.d. Bernoulli with probability θ, and is independent of the conditional quantile estimator, $\hat{Q}_t(\theta)$. Ideally, Hit_t will have zero unconditional and conditional expectations. As in the empirical study of Engle and Manganelli, we included four lags of Hit_t in the test's regression to deliver a DQ test statistic, which, under the null hypothesis of perfect unconditional and conditional coverage, is distributed $\chi^2(6)$.

Table 4 presents the values of the hit percentage measure for each method applied to each of the 10 stock returns series for estimation of the 5% quantiles. The final column presents the number of stocks for which the hit percentage is significantly different from the ideal of 5% when testing at the 5% significance level. The best results were achieved using the EWQR method with no reressors, the adaptive CAViaR model and the two methods based on exponential smoothing volatility forecasts. Table 5 reports the values of the DQ test statistic for the 5% quantiles. As in Table 4, the final column summarises the number of significant entries for each of the methods. The methods performing particularly well in Table 5 are the EWQR and EWDKQR methods with no regressors, and the exponential smoothing volatility forecasting method with Gaussian assumption. In Section 4.4, we summarise the VaR results for all four quantiles.

---------- Tables 4, 5 and 6 ----------

4.3. ES Results

We employed a similar approach to that of McNeil and Frey (2000) to evaluate the conditional ES estimates. The procedure considers the discrepancy between an observation and the conditional ES estimate for only those periods for which the observation exceeds the conditional quantile estimate. When standardised by the conditional volatility, these discrepancies should be i.i.d. with a mean of zero. Because the EWQR and EWDKQR methods do not involve the estimation of the conditional volatility, instead of standardising with the volatility, we standardised using the conditional quantile estimate for each method. In order to avoid distributional assumptions, McNeil and Frey use a bootstrap test to test the standardised discrepancies for a zero mean (see page 224 of Efron and Tibshirani 1993). Table 6 reports p-values for the bootstrap test for the post-sample conditional 5% ES estimates. The table reports no results for the CAViaR models because, as noted in Section 1, it is not clear how ES forecasts can be produced for this class of models. As in Tables 4 and 5, the final column in Table 6 presents a count for the number of series for which the null is rejected at the 5% level. The results are impressive for all the methods except exponential smoothing of the volatility with a Gaussian assumption. The results for the other three quantiles were more varied, and we see this in Section 4.4, where we summarise all of the ES results.

To evaluate the dynamic properties of the ES estimator, we need to test whether the standardised discrepancies are i.i.d. However, the test has low power because, with a post-sample period of 500 observations, there are a very low number of discrepancies, and this is particularly so for the 1% and 99% estimation. We tested for zero autocorrelation in each series of discrepancies corresponding to 5% and 95% estimation. Using a 5% significance level, we found that the total number of rejections of the null hypothesis across these two quantiles and the 10 series was zero or one. These results provide little insight, so we do not present them in further detail here.

4.4. Summary of VaR and ES Results

Table 7 is a summary of the VaR and ES results for all four quantiles. The table presents the final columns from Tables 4, 5 and 6, along with the corresponding summary measures for the 1%, 95% and 99% quantiles. As our study involved 10 stocks, for a given quantile, the maximum number of test rejections for any single test is 10. The columns labelled “Total” contain the total number of rejections across the four quantiles.

---------- Table 7 ----------

Let us first consider the hit percentage results in Table 7. These show that, overall, the best performing methods were EWQR with no regressors and exponential smoothing of the volatility with EVT. The method with the leverage effect regressor also performed well, along with the method based on exponential smoothing of the volatility with a Gaussian assumption. For the EWDKQR method, the results for the more extreme quantiles, 1% and 99%, were more competitive than for the 5% and 95% quantiles. Turning to the results in Table 7 for the DQ test statistic, we can see that the EWDKQR method with no regressors performed very well. Other methods that performed well in terms of the DQ test were the EWQR method with no regressors and the exponential smoothing and GARCH volatility forecasting methods with EVT.

With regard to ES estimation, Table 7 shows that the best performance was achieved by the EWQR and EWQR leverage methods and by exponential smoothing of the volatility with EVT. The poorest results were produced by exponential smoothing of the volatility with a Gaussian assumption. The performance of the two EWDKQR methods was similar to that of the GARCH models.

Our overall conclusion from the study is that our new methods were competitive. The EWQR method performed better than the EWDKQR method in terms of VaR hit percentage and the ES test, while the DQ test statistic results suggest that the dynamic properties of the quantile are better explained by the EWDKQR method. The inclusion of the leverage term in the EWQR and EWDKQR methods provided no improvement in the results, which was consistent with the results for the asymmetric GARCH and CAViaR models.

5. Summary and Concluding Comments

In this paper, we have introduced EWQR as a means of using exponential smoothing to estimate the time-varying quantiles of the conditional returns distribution. The approach can be viewed as exponential smoothing of the cdf. Expressing the approach as kernel estimation of the cdf prompted us to adapt the double kernel estimator of Yu and Jones (1998) to give a new exponentially weighted double kernel estimator. This estimator has the appeal of incorporating kernel density estimation within the exponentially weighted estimator. Although this estimator can be used to deliver quantile estimates, our new exponentially weighted double kernel quantile regression (EWDKQR) enables this to be done directly, and also has the appeal of allowing quantile modelling in terms of regressors. We showed that an appealing feature of the EWQR and EWDKQR approaches is that the cost function can be used in a very simple way as the basis for a new predictor of the time-varying ES associated with the respective quantile forecasts. An empirical comparison of the new methods with a selection of widely used approaches gave encouraging results in terms of both quantile and ES forecast performance.

In terms of future research, it would be interesting to see further empirical evaluation of the methods proposed in this paper, using perhaps different data and a different set of benchmark methods. Drawing on the results from this paper, Taylor (2007) considers the use of EWQR for the substantially different application of forecasting supermarket sales. Throughout the paper, we have considered linear quantile models. However, De Gooijer and Zerom (2003) suggest that in many practical situations a non-linear model is needed to capture the underlying structure in a quantile. EWQR could certainly be used in future work to estimate such models, although the minimisation could not be solved using linear programming. Another potential research area is the development, for other applications, of the double kernel quantile regression method, which was introduced in expression (20) and which incorporates kernel density estimation within standard quantile regression.

Acknowledgements

We acknowledge the helpful comments of Jan De Gooijer, Ev Gardner, Patrick McSharry and Keming Yu on an earlier version of this paper. We are also grateful for the useful comments of the editor, Eric Renault, and two referees.

Appendix

Proof of Theorem 1: In this proof, we draw heavily on Section 2.2 of Koenker (2005). Our development here adapts Koenker's analysis of quantile regression for the case of EWQR. The EWQR objective function $R(\boldsymbol{\beta})$ is presented in the following expression:

$$R(\boldsymbol{\beta}) = \sum_{t=1}^{T} \lambda^{T-t} \left(y_t - \mathbf{x}_t'\boldsymbol{\beta}\right)\left(\theta - I\left(y_t < \mathbf{x}_t'\boldsymbol{\beta}\right)\right)$$

The function $R(\boldsymbol{\beta})$ is not differentiable at the points at which any of the residuals, $\left(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}\right)$, are equal to zero. For this reason, when considering the minimisation of $R(\boldsymbol{\beta})$, we consider directional derivatives. The directional derivative of R in direction $\boldsymbol{w}$ is given by

$$\begin{aligned}
\nabla R(\boldsymbol{\beta}, \boldsymbol{w}) &\equiv \frac{d}{db} R(\boldsymbol{\beta} + b\boldsymbol{w}) \,\Big|_{b=0} \\
&= \sum_{t=1}^{T} \lambda^{T-t}\left(y_t - \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{x}_t' b\mathbf{w}\right)\left(\theta - I\left(y_t < \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{x}_t' b\mathbf{w}\right)\right) \,\Big|_{b=0} \\
&= -\sum_{t=1}^{T} \lambda^{T-t}\psi\left(y_t - \mathbf{x}_t'\boldsymbol{\beta}, -\mathbf{x}_t'\mathbf{w}\right)\boldsymbol{x}_t'\boldsymbol{w}
\end{aligned}$$

where

$$\psi(u,v) = \begin{cases} \theta - I(u<0) & if \;\; u \neq 0 \\ \theta - I(v<0) & if \;\; u = 0 \end{cases}$$

The parameter vector $\hat{\boldsymbol{\beta}}$ minimises $R(\boldsymbol{\beta})$ if and only if the directional derivatives, $\nabla R\left(\hat{\boldsymbol{\beta}}, \boldsymbol{w}\right)$, are nonnegative for all directions $\boldsymbol{w}$. We present this condition in the following expression:

$$-\sum_{t=1}^{T} \lambda^{T-t}\psi\left(y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}, -\mathbf{x}_t'\mathbf{w}\right)\mathbf{x}_t'\mathbf{w} \;\geq\; 0 \tag{24}$$

A requirement of Theorem 1 is that the model includes an intercept term. This implies that there exists a vector $\boldsymbol{\alpha}$ such that $\boldsymbol{x}_t'\boldsymbol{\alpha} = 1$ for all t. If we let $\boldsymbol{w}$=-$\boldsymbol{\alpha}$ in expression (24), we get

$$\sum_{t=1}^{T} \lambda^{T-t}\psi\left(y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}, 1\right) \;\geq\; 0$$

This can be expressed as

$$\theta\sum_{t=1}^{T}\lambda^{T-t}\, I\left(y_t > \mathbf{x}'\hat{\boldsymbol{\beta}}\right) - (1-\theta)\sum_{t=1}^{T}\lambda^{T-t}\, I\left(y_t < \mathbf{x}_t'\hat{\boldsymbol{\beta}}\right) + \theta\sum_{t=1}^{T}\lambda^{T-t}\, I\left(y_t = \mathbf{x}_t'\hat{\boldsymbol{\beta}}\right) \geq 0 \tag{25}$$

In a similar way, we can let $\boldsymbol{w}=\boldsymbol{\alpha}$ in expression (24) to deliver the following

$$\theta\sum_{t=1}^{T}\lambda^{T-t}\, I\left(y_t > \mathbf{x}_t'\hat{\boldsymbol{\beta}}\right) - (1-\theta)\sum_{t=1}^{T}\lambda^{T-t}\, I\left(y_t < \mathbf{x}_t'\hat{\boldsymbol{\beta}}\right) - (1-\theta)\sum_{t=1}^{T}\lambda^{T-t}\, I\left(y_t = \mathbf{x}_t'\hat{\boldsymbol{\beta}}\right) \leq 0 \tag{26}$$

The inequalities of expressions (25) and (26) can be rewritten as the inequalities of expression (6).

Figure Legends

Figure 1 **Plot of Gaussian kernel bandwidth, h_2, and exponential weight, λ, derived for EWDKQR with an intercept and no regressors. Pearson correlation is –0.80. Values derived using the estimation sample of 2893 periods.**

References

Abberger, K. 1997. Quantile Smoothing in Financial Time Series. *Statistical Papers* 38, 125-148.

Artzner, P., F. Delbaen, J.M. Eber, D. Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* 9, 203-228.

Boudoukh, J., M. Richardson, R.F. Whitelaw. 1998. The Best of Both Worlds. *Risk* 11, May 64-67.

Brown, R.G. 1963. *Smoothing, Forecasting and Prediction of Discrete Time-Series*. Englewood Cliffs, NJ: Prentice-Hall.

Butler, J.S., B. Schachter, B. 1998. Estimating Value-at-Risk with a Precision Measure by Combining Kernel Estimation with Historical Simulation. *Review of Derivatives Research* 1, 371-390.

Cai, Z. 2002. Regression Quantiles for Time Series. *Econometric Theory* 18, 169-192.

Chen, S.X., C.Y. Tang. 2005. Nonparametric Inference of Value at Risk for Dependent Financial Returns. *Journal of Financial Econometrics* 3, 227-255.

De Gooijer, J.G., D. Zerom. 2003. On Additive Conditional Quantiles with High-Dimensional Covariates. *Journal of the American Statistical Association* 98, 135-146.

De Gooijer, J.G., A. Gannoun, D. Zerom. 2006. Multivariate Conditional Quantile Prediction. *Communications in Statistics: Theory and Methods* 35, 133-147.

Dunsmuir, W.T.M., D.J. Scott, W. Qiu. 1996. The Distribution of the Weighted Moving Median of a Sequence of IID Observations. *Communications in Statistics: Simulation and Computation* 25, 1015-1029.

Engle, R.F., S. Manganelli. 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics* 22, 367-381.

Fan, J., J. Gu. 2003. Semiparametric Estimation of Value at Risk. *Econometrics Journal* 6, 261-290.

Gardner, E.S. Jr. 1985. Exponential Smoothing: The State of the Art. *Journal of Forecasting* 4, 1-28.

Gijbels, I., A. Pope, M.P. Wand. 1999. Understanding Exponential Smoothing via Kernel Regression. *Journal of the Royal Statistical Society Series B* 61, 39-50.

Glosten, L.R., R. Jagannathan, D.E. Runkle. 1993. On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance* 48, 1779-1801.

Gorr, W.L., C. Hsu. 1985. An Adaptive Filtering Procedure for Estimating Regression Quantiles. *Management Science* 31, 1019-1029.

Hall, P., Wolff, R.C., Q. Yao. 1999. Methods for Estimating a Conditional Distribution Function. *Journal of the American Statistical Association* 94, 154-163.

Koenker, R.W. 2005. *Quantile Regression*. Cambridge, UK: Cambridge University Press.

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

Jones, M.C., P. Hall. 1990. Mean Squared Error Properties of Kernel Estimates of Regression Quantiles. *Statistics and Probability Letters* 10, 283-289.

Kuester, K., S. Mittnik, M.S. Paolella. 2006. Value-at-Risk Prediction: A Comparison of Alternative Strategies. *Journal of Financial Econometrics* 4, 53-89.

Manganelli S., R.F. Engle. 2004. A Comparison of Value-at-Risk Models in Finance. in *Risk Measures for the 21st Century*, ed. G. Szegö, Chichester: Wiley, pp. 123-144.

McNeil, A.J., R. Frey. 2000. Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance* 7, 271-300.

McNeil, A.J., R. Frey, P. Embrechts. 2005. *Quantitative Risk Management*. New Jersey: Princeton University Press.

Poon, S., C.J.W. Granger. 2003. Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature* 41, 478-639.

Taylor, J.W. 2007. Forecasting Supermarket Sales Using Exponentially Weighted Quantile Regression. *European Journal of Operational Research* 178, 154-167.

Taylor, J.W. 2008. Estimating Value at Risk and Expected Shortfall Using Expectiles. *Journal of Financial Econometrics* 6, 231-252.

Yu, K., M.C. Jones. 1998. Local Linear Quantile Regression. *Journal of the American Statistical Association* 93, 228-237.

Table 1 Skewness and excess kurtosis for the 10 stocks.

	Skewness	Excess Kurtosis
General Electric	0.04	4.17**
Exxon Mobil	0.08*	3.29**
Microsoft	-0.09*	4.70**
Citigroup	0.07	4.77**
Johnson and Johnson	-0.36**	6.46**
Pfizer	-0.17**	2.44**
Bank of America	-0.16**	3.11**
Wal Mart Stores	0.01	3.58**
Intel	-0.39**	5.63**
Procter and Gamble	-3.47**	76.89**

Note: Significance at 5% and 1% levels is indicated by * and **, respectively. Values calculated using the entire sample of 3393 periods.

Table 2 Exponential weight λ for EWQR with an intercept and no regressors.

θ	GE	Exxon	MS	Citigrp	J&J	Pfizer	Bank of Am	Wal Mart	Intel	P&G	Mean
1%	1.000	0.990	0.980	0.990	0.995	1.000	0.995	0.995	0.995	1.000	0.994
5%	0.985	0.985	0.985	0.985	0.980	0.985	0.980	0.985	0.970	0.985	0.983
95%	0.985	0.990	0.980	0.980	0.985	0.980	0.975	0.990	0.995	0.990	0.985
99%	0.990	0.985	0.995	0.995	0.990	0.990	0.990	1.000	0.995	0.995	0.993

Note: Values derived using the estimation sample of 2893 periods.

Table 3 Exponential weight λ for EWDKQR with an intercept and no regressors.

θ	GE	Exxon	MS	Citigrp	J&J	Pfizer	Bank of Am	Wal Mart	Intel	P&G	Mean
1%	0.835	0.940	0.815	0.905	0.860	0.850	0.835	0.875	0.840	0.995	0.875
5%	0.975	0.950	0.970	0.980	0.975	0.980	0.960	0.985	0.955	0.985	0.972
95%	0.985	0.985	0.975	0.975	0.980	0.975	0.975	0.985	0.995	0.990	0.982
99%	0.990	0.990	0.995	0.990	0.885	0.980	0.960	1.000	1.000	0.990	0.978

Note: Values derived using the estimation sample of 2893 periods.

Table 4 Evaluation of forecasts of 5% quantiles. Hit percentage for 500 post-sample forecasts.

	GE	Exxon	MS	Citigrp	J&J	Pfizer	Bank of Am	Wal Mart	Intel	P&G	Number significant at 5% level
EWQR	4.0	5.8	4.2	3.8	4.2	5.4	4.8	4.4	4.6	5.0	0
EWQR Leverage	3.8	6.0	3.6	4.2	3.0*	5.2	5.0	3.8	3.8	5.0	1
EWDKQR	3.8	4.6	3.4	2.8*	2.0**	4.8	3.0*	3.0*	3.6	4.2	4
EWDKQR Leverage	1.4**	5.0	2.2**	2.6*	2.4**	3.0*	2.4**	2.6*	1.8**	3.0*	9
Historical Simulation	2.6*	6.2	3.0*	2.2**	2.0**	4.8	3.6	3.8	3.8	5.2	4
Kernel Historical Simulation	1.8**	4.4	1.4**	2.0**	1.6**	4.2	2.6*	3.0*	2.8*	3.2	7
Exp Sm Variance Gaussian	5.8	5.4	4.6	5.0	4.6	5.0	6.2	4.6	3.4	3.8	0
Exp Sm Variance EVT	6.0	5.2	5.4	5.4	4.6	5.4	6.0	5.2	3.6	5.2	0
GARCH Variance Student-*t*	2.0**	3.2	1.2**	0.4**	1.2**	2.2**	1.2**	2.0**	1.6**	0.8**	9
GARCH Variance EVT	5.0	5.0	2.2**	1.8**	2.6*	4.8	3.2	3.6	3.2	3.0*	4
GJRGARCH Variance Student-*t*	1.4**	3.2	1.2**	0.2**	1.2**	2.0**	1.2**	1.4**	1.6**	0.8**	9
GJRGARCH Variance EVT	4.4	5.0	2.2**	2.0**	2.2**	4.2	2.6*	3.2	3.0*	2.6*	6
Adaptive CAViaR	4.2	4.6	4.8	4.8	4.6	4.8	4.8	4.8	3.2	5.6	0
Sym Abs Value CAViaR	5.2	4.2	3.0*	1.8**	1.8**	5.6	5.2	3.0*	4.2	1.2**	5
Asym Slope CAViaR	3.8	4.0	2.4**	2.4**	1.8**	4.0	3.6	3.6	3.0*	4.8	4
Indirect GARCH CAViaR	5.0	4.0	2.8*	2.2**	1.8**	5.0	4.4	2.8*	4.6	1.0**	5

Note: Significance at 5% and 1% levels is indicated by * and **, respectively.

Table 5 **Evaluation of forecasts of 5% quantiles. DQ test p-values for 500 post-sample forecasts.**

	GE	Exxon	MS	Citigrp	J&J	Pfizer	Bank of Am	Wal Mart	Intel	P&G	Number significant at 5% level
EWQR	0.435	0.686	0.274	0.068	0.612	0.323	0.316	0.204	0.426	0.424	0
EWQR Leverage	0.032	0.114	0.467	0.074	0.198	0.006	0.246	0.061	0.309	0.181	2
EWDKQR	0.295	0.481	0.323	0.076	0.057	0.586	0.198	0.454	0.511	0.755	0
EWDKQR Leverage	0.010	0.735	0.090	0.035	0.071	0.091	0.122	0.153	0.030	0.459	3
Historical Simulation	0.030	0.118	0.440	0.005	0.042	0.077	0.456	0.113	0.277	0.081	3
Kernel Historical Simulation	0.017	0.476	0.011	0.002	0.010	0.233	0.206	0.417	0.207	0.157	4
Exp Sm Variance Gaussian	0.249	0.659	0.413	0.125	0.177	0.558	0.145	0.223	0.653	0.137	0
Exp Sm Variance EVT	0.248	0.502	0.527	0.007	0.177	0.505	0.159	0.100	0.700	0.274	1
GARCH Variance Student-*t*	0.089	0.524	0.010	0.001	0.007	0.021	0.016	0.126	0.025	0.005	7
GARCH Variance EVT	0.858	0.614	0.012	0.002	0.178	0.690	0.433	0.707	0.579	0.440	2
GJRGARCH Variance Student-*t*	0.034	0.541	0.010	0.000	0.007	0.060	0.017	0.031	0.025	0.005	8
GJRGARCH Variance EVT	0.999	0.572	0.013	0.006	0.090	0.735	0.196	0.542	0.492	0.285	2
Adaptive CAViaR	0.075	0.422	0.034	0.062	0.050	0.205	0.019	0.010	0.514	0.238	4
Sym Abs Value CAViaR	0.533	0.609	0.373	0.002	0.058	0.615	0.273	0.271	0.719	0.018	2
Asym Slope CAViaR	0.307	0.782	0.113	0.025	0.058	0.600	0.458	0.415	0.107	0.839	1
Indirect GARCH CAViaR	0.858	0.769	0.098	0.015	0.057	0.595	0.540	0.416	0.770	0.010	2

Note: Higher p-values are better.

Table 6 Evaluation of forecasts of 5% ES. Bootstrap test p-values for zero mean standardised discrepancies based on 500 post-sample forecasts of 5% ES.

	GE	Exxon	MS	Citigrp	J&J	Pfizer	Bank of Am	Wal Mart	Intel	P&G	Number significant at 5% level
EWQR	0.779	0.411	0.343	0.301	0.372	0.554	0.598	0.338	0.908	0.204	0
EWQR Leverage	0.012	0.423	0.843	0.908	0.910	0.647	0.754	0.975	0.799	0.359	1
EWDKQR	0.532	0.790	0.266	0.405	0.680	0.779	0.570	0.643	0.802	0.699	0
EWDKQR Leverage	0.567	0.465	0.325	0.598	0.678	0.248	0.675	0.822	0.335	0.376	0
Historical Simulation	0.813	0.416	0.526	0.131	0.538	0.461	0.840	0.944	0.834	0.170	0
Kernel Historical Simulation	0.687	0.554	0.140	0.084	0.686	0.323	0.958	0.719	0.784	0.920	0
Exp Sm Variance Gaussian	0.654	0.007	0.084	0.060	0.090	0.021	0.044	0.030	0.120	0.181	4
Exp Sm Variance EVT	0.096	0.253	0.452	0.929	0.949	0.348	0.325	0.719	0.554	0.395	0
GARCH Variance Student-*t*	0.478	0.954	0.345	0.507	0.546	0.183	0.145	0.498	0.550	0.636	0
GARCH Variance EVT	0.080	0.609	0.322	0.127	0.337	0.736	0.530	0.805	0.795	0.212	0
GJRGARCH Variance Student-*t*	0.987	0.765	0.368	0.000	0.366	0.154	0.178	0.870	0.727	0.636	1
GJRGARCH Variance EVT	0.178	0.655	0.320	0.152	0.459	0.560	0.450	0.768	0.976	0.171	0

Note: Higher p-values are better.

Table 7 **Summary of VaR and ES results. Number of test rejections at 5% significance level for each of the four θ quantiles. Note that CAViaR models produce only VaR estimates.**

	VaR Hit % Test					VaR DQ Test					ES Bootstrap Test				
	θ (×100)					θ (×100)					θ (×100)				
	1	5	95	99	Total	1	5	95	99	Total	1	5	95	99	Total
EWQR	0	0	1	0	1	2	0	2	1	5	1	0	1	1	3
EWQR Leverage	0	1	2	0	3	2	2	3	2	9	0	1	2	0	3
EWDKQR	0	4	4	0	8	0	0	0	0	0	3	0	0	5	8
EWDKQR Leverage	0	9	7	3	19	1	3	1	4	9	5	0	1	1	7
Historical Simulation	0	4	4	0	8	3	3	4	2	12	0	0	1	4	5
Kernel Historical Simulation	0	7	6	1	14	1	4	5	1	11	3	0	3	4	10
Exp Sm Variance Gaussian	1	0	0	2	3	3	0	0	4	7	2	4	4	2	12
Exp Sm Variance EVT	0	0	1	0	1	2	1	0	2	5	0	0	1	2	3
GARCH Variance Student-*t*	1	9	10	4	24	1	7	5	4	17	3	0	2	4	9
GARCH Variance EVT	0	4	7	2	13	0	2	1	2	5	2	0	3	3	8
GJRGARCH Variance Student-*t*	2	9	9	4	24	2	8	6	4	20	2	1	3	3	9
GJRGARCH Variance EVT	1	6	8	3	18	1	2	4	4	11	1	0	3	3	7
Adaptive CAViaR	2	0	3	2	7	3	4	2	2	11	-	-	-	-	-
Sym Abs Value CAViaR	1	5	7	1	14	2	2	2	1	7	-	-	-	-	-
Asym Slope CAViaR	0	4	8	3	15	1	1	4	5	11	-	-	-	-	-
Indirect GARCH CAViaR	2	5	6	2	15	3	2	3	2	10	-	-	-	-	-

NOTE: Evaluation is for 500 post-sample forecasts. Smaller values are better.

Figure 1 **Plot of Gaussian kernel bandwidth, h_2, and exponential weight, λ, derived for EWDKQR with an intercept and no regressors. Pearson correlation is –0.80. Values derived using the estimation sample of 2893 periods.**