Time-Varying Quantiles

Giuliano De Rossi and Andrew Harvey
Faculty of Economics, Cambridge University
e-mail: ach34@cam.ac.uk

January 31, 2007

1 Introduction

In modelling the evolution of a time series, we may wish to take account of changes in many aspects of the distribution over time rather than in just the mean or, as is often the case in finance, the variance. In a cross-section, the sample quantiles provide valuable and easily interpretable information; indeed if enough of them are calculated they effectively provide a comprehensive description of the whole distribution. It is not difficult to capture the evolution of such quantiles over time. For example, Harvey and Bernstein (2003) use standard unobserved component (UC) models to extract underlying trends from the time series of quantiles computed for the U.S. wage distribution.

Estimating time-varying quantiles for a single series is far more difficult. The problem is a fundamental one from the statistical point of view. Furthermore it is of considerable practical importance, particularly in areas like finance where questions of, for example, value at risk (VaR) appeals directly to a knowledge of certain quantiles of portfolio returns; see the RiskMetrics approach of J.P. Morgan (1996) and the discussion in Christoffersen, Hahn and Inoue (2001). Engle and Manganelli (2004) highlight this particular issue, as do Chernozhukov and Umantsev (2001). They propose the use of various nonlinear conditional autoregressive value at risk (CAViaR) models to make one-step ahead predictions of VaR in time series of stock returns. These models are based on quantile regression (QR); see the recent monograph by Koenker (2005). As with linear quantile autoregressions (QARs), the properties and applications of which have recently been investigated by Koenker and Xiao (2004, 2006), the conditioning on past observations enables estimation to be carried out by standard QR estimation procedures, usually based on linear programming, and the construction of forecasts is immediate. Engle and Manganelli have to resort to trying to capture the behavior of tail quantiles in returns by nonlinear functions of past observations because linear functions lack the necessary flexibility. But in doing so they encounter a fundamental problem, namely what functional form should be chosen? A good deal of their article is concerned with devising methods for dealing this issue.

Treating the problem as one of signal extraction provides an altogether different line of attack. Estimating quantiles in this way provides a description of the series, while the estimates at the end form the basis for predictions. This approach seems entirely natural and once it is adopted a much clearer indication is given as to the way in which past observations should be weighted for prediction in a nonlinear quantile autoregression. The motivation for what we are doing is provided by the simplest case, namely stock market returns. The base model is that returns are independently and identically distributed (IID). By allowing the quantiles to evolve over time, it becomes possible to capture a changing distribution. Movements may be stationary or non-stationary, but they will usually be slowly changing. Figure 1 shows 2000 daily returns for General Motors[1] together with smoothed estimates for quantiles obtained from a model that assumes that they are generated by random walks. The implications for forecasting are obvious.

The distinction between models motivated by description and those set up to deal directly with prediction is a fundamental one in time series. Structural time series models (STMs) are formulated in terms of unobserved components, such as trends and cycles, that have a direct interpretation; see Harvey (1989). Signal extraction, or smoothing, provides estimates of these components over the whole sample, while the (filtered) estimates at the end provide the basis for nowcasting and forecasting. Autoregressive and autoregressive-

[1] The stock returns data used as illustrations are taken from Engle and Manganelli (2004). Their sample runs from April 7th, 1986, to April 7th, 1999. The large (absolute) values near the beginning of figure 1 are associated with the great crash of 1987.

The histogram of the series from observation 501 to 2000 (avoiding the 1987 crash) shows heavy tails but no clear evidence of skewness. The excess kurtosis is 1.547 and the associated test statistic, distributed as χ^2_1 under normality, is 149.5. On the other hand, skewness is 0.039 with an associated test statistic of only 0.37.

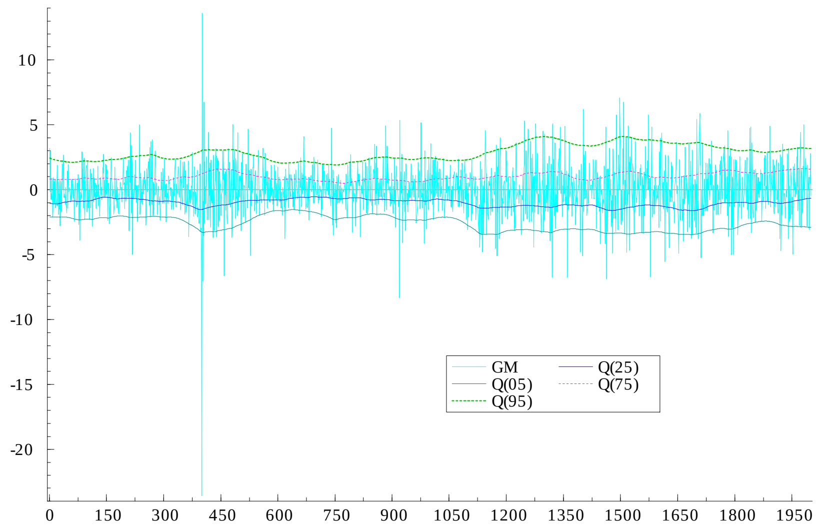

Figure 1: Quantiles fitted to GM returns

integrated-moving average (ARIMA) models, on the other hand, are constructed primarily with a view to forecasting. In a linear Gaussian world, the reduced form of an STM is an ARIMA model and questions regarding the merits of STMs for forecasting revolve round the gains, or losses, from the implied restrictions on the reduced form and the guidance, or lack of it, given to the selection of a suitable model; see the discussion in Harvey (2006) and Durbin and Koopman (2001). Once nonlinearity and non-Gaussianity enter the picture, the two approaches can be very different. For example, changing variance can be captured by a model from the generalized autoregressive conditional heteroscedasticity (GARCH) class, where conditional variance is a function of past observations, or by a stochastic volatility (SV) model in which the variance is a dynamic unobserved component; see the recent collection of readings by Shephard (2005).

Section 2 of the paper reviews the basic ideas of quantiles and QR and notes that the criterion of choosing the estimated quantile so as to minimize the sum of absolute values around it can be obtained from a model in which the observations are generated by an asymmetric double exponential distribution. In section 3 this distribution is combined with a time series model, such as a stationary first-order autoregressive process or a random walk, to produce the criterion function used as the basis for extracting time-varying quantiles. Differentiating this criterion function leads to a set of equations which, when solved, generalize the defining characteristic of a sample quantile to the dynamic setting. We outline an algorithm for computing the quantiles. The algorithm iterates the Kalman filter and associated (fixed-interval) smoother until convergence, taking special care in the way it deals with cusp solutions, that is when a quantile passes through an observation. The models for the quantiles usually depend on only one or two parameters and these parameters be estimated by cross-validation.

Section 4 details the various aspects of a distribution that can be captured by time–varying quantiles. We first note that the inter-quartile range provides an alternative to GARCH and SV models for estimating and predicting dispersion and that the assumptions needed to compute it are much less restrictive. We then go on to observe that different quantile-based ranges can provide constrasts between movements near the centre of the distribution and those near the tails. Other contrasts can be designed to capture asymmetries, while the 1% and 5% quantiles provide estimates of VaR. Engle and Manganelli (2004, p369) observe, in connection with CAViaR models, that one of their attractions is that they are useful '..for situations with constant volatilities but changing error distributions..' The same is true of our time-varying quantiles for the lower tails.

Section 5 sets out a stationarity test for the null hypothesis that a quantile is time-invariant against the alternative that it is slowly changing. This test is based on binomial quantile indicators - sometimes called 'quantile hits' - and is a generalization of a test proposed by DeJong, Amsler and Schmidt (2005) for the constancy of the median. Tests[2] for the constancy of contrasts between quantiles are illustrated with data on returns. A full invetigation into the properties of the tests can be found in Busetti and Harvey (2007).

QAR and CAViaR models are discussed in section 6. We suggest that conditional autoregressive specifications, especially linear ones, are influenced too much by Gaussian notions even though on the surface the emphasis on quantiles appears to escape from Gaussianity. One consequence is that these models may not be robust to additive outliers. The functional forms proposed by Engle and Manganelli (2004) are assessed with respect to robustness and compared with specifications implied by the signal extraction approach. A

[2] Linton and Whang (2006) have studied the properties of the quantilogram, the correlogram of the quantile hits, and an associated portmanteau test.

way of combining the two approaches then emerges.

So far as we know, the only work which is related to our approach to extracting time-varying quantiles is by Bosch *et al* (1995). Their paper concerns cubic spline quantile regression and since they do not apply it to time series the connections may not be apparent. Bosch *et al* (1995) propose a quadratic programming algorithm, but this appears to be very computationally intensive. We show in section 7 that the stochastic trend models that we used earlier lead to estimated quantiles that are, in fact, splines. Our state space smoothing algorithm can be applied to cubic spline quantile regression by modifying it to deal with irregular observations. This method of dealing with splines is well known in the time series literature and its extension to quantile regression may prove useful.

2 Quantiles and quantile regression

Let $Q(\tau)$ - or, when there is no risk of confusion, Q - denote the $\tau - th$ quantile. The probability that an observation is less than $Q(\tau)$ is τ, where $0 < \tau < 1$. Given a set of T observations, $y_t, t = 1, .., T$, (which may be from a cross-section or a time series), the sample quantile, $\widetilde{Q}(\tau)$, can be obtained by sorting the observations in ascending order. However, it is also given as the solution to minimizing

$$\begin{aligned} S(\tau) &= \sum_t \rho_\tau(y_t - Q) = \sum_{y_t<Q} (\tau - 1)(y_t - Q) + \sum_{y_t \geq Q} \tau(y_t - Q) \\ &= \sum_t (\tau - I(y_t - Q < 0))\,(y_t - Q) \end{aligned}$$

with respect to Q, where $\rho_\tau(.)$ is the *check function* and $I(.)$ is the indicator function. Differentiating (minus) $S(\tau)$ at all points where this is possible gives

$$\sum IQ(y_t - Q(\tau)),$$

where

$$IQ(y_t - Q_t(\tau)) = \begin{cases} \tau - 1, & if \quad y_t < Q_t(\tau) \\ \tau, & if \quad y_t > Q_t(\tau) \end{cases} \tag{1}$$

defines the *quantile indicator function* for the more general case where the quantile may be time-varying. Since $\rho_\tau(.)$ is not differentiable at zero, the quantile indicator function is not continuous at 0 and $IQ(0)$ is not determined.[3]

The sample quantile, $\widetilde{Q}(\tau)$, is such that, if $T\tau$ is an integer, there are $T\tau$ observations below the quantile and $T(1-\tau)$ above. In this case any value of $\widetilde{Q}$ between the $T\tau - th$ smallest observation and the one immediately above will make $\sum IQ(y_t - \widetilde{Q}) = 0$. If $T\tau$ is not an integer, $\widetilde{Q}$ will coincide with one observation. This observation is the one for which $\sum IQ(y_t - \widetilde{Q})$ changes sign. These statements need to be modified slightly if several observations take the same value and coincide with $\widetilde{Q}$. Taking this point on board, a general definition of a sample $\tau-$quantile is a point such that the number of observations smaller, that is $y_t < \widetilde{Q}$, is no more than $[T\tau]$ while the number greater is no more than $[T(1-\tau)]$.

When $y_t = \widetilde{Q}$ we will assign $IQ(y_t - \widetilde{Q})$ a value that makes the sum of the $IQ(y_t - \widetilde{Q})'s$ equal to zero. This has the additional attraction that it enables us to think of the condition $\sum IQ(y_t - \widetilde{Q}) = 0$ as characterizing a sample quantile in much the same way as $\sum(y_t - \widetilde{\mu}) = 0$ defines a method of moments estimator, $\widetilde{\mu}$, for a mean.

In quantile regression, the quantile, $Q_t(\tau)$, corresponding to the $t - th$ observation is a linear function of explanatory variables, $\mathbf{x}_t$, that is $Q_t = \mathbf{x}_t'\boldsymbol{\beta}$. The quantile regression estimates are obtained by minimizing $\sum_t \rho_\tau(y_t - \mathbf{x}_t'\boldsymbol{\beta})$ with respect to the parameter vector $\boldsymbol{\beta}$. Estimates may be computed by linear programming as described in Koenker (2005). In quantile autoregression Q_t is a linear combination of past observations.

Finally suppose that the observations are assumed to come from an asymmetric double exponential distribution

$$p(y_t|Q_t) = \tau(1-\tau)\omega^{-1}\exp(-\omega^{-1}\rho_\tau(y_t - Q_t)), \tag{2}$$

where ω is a scale parameter. Maximising the log-likelihood function is equivalent to minimising the criterion function

$$S(\tau) = \sum_t \rho_\tau(y_t - Q_t). \tag{3}$$

Thus the model, (2), defines Q_t as a (population) quantile by the condition that the probability of a value below is τ while the form of the distribution leads to the maximum likelihood (ML) estimator satisfying the conditions

[3] Hence we have not written $IQ_t = (\tau - I(y_t - Q < 0))$.

for a sample quantile, when Q is constant, or a quantile regression estimate. Since quantiles are fitted separately, there is no notion of an overall model for the whole distribution and assuming the distribution (2) for one quantile is not compatible with assuming it for another. Setting up this particular parametric model is simply a convenient device that leads to the appropriate criterion function for what is essentially a nonparametric estimate.

3 Signal extraction

A model-based framework for estimating time-varying quantiles, $Q_t(\tau)$, can be set up by assuming that they are generated by a Gaussian stochastic process and are connected to the observations through a measurement equation

$$y_t = Q_t(\tau) + \varepsilon_t(\tau), \qquad t = 1, ..., T, \tag{4}$$

where $\Pr(y_t - Q_t < 0) = \Pr(\varepsilon_t < 0) = \tau$ with $0 < \tau < 1$. The problem is then one of signal extraction with the model for $Q_t(\tau)$ being treated as a transition equation. By assuming that the disturbance term, ε_t, has an asymmetric double exponential distribution, as in (2), we end up choosing the estimated quantiles so as to minimise $\sum_t \rho_\tau(y_t - Q_t)$ subject to a set of constraints imposed by the time series model for the quantile.

We will focus attention on three time series models, all of which are able to produce quantiles that change relatively slowly over time with varying degrees of smoothness. The theory can be applied to any linear time series model, but it will often be the case that prior notions on the behaviour of the quantile play an important role in specification.

3.1 Models for evolving quantiles

The simplest model for a stationary time-varying quantile is a first-order autoregressive process

$$Q_t(\tau) = (1-\phi)Q_\tau^\dagger + \phi_\tau Q_{t-1}(\tau) + \eta_t(\tau), \qquad |\phi_\tau| < 1, \qquad t = 1, ..., T, \tag{5}$$

where $\eta_t(\tau)$ is normally and independently distributed with mean zero and variance $\sigma^2_{\eta(\tau)}$, that is $\eta_t(\tau) \sim NID(0, \sigma^2_{\eta(\tau)})$, ϕ_τ is the autoregressive parameter and $Q^{\dagger}_{\tau}$ is the unconditional mean of $Q_t(\tau)$. In what follows the τ appendage will be dropped where there is no ambiguity.

The random walk quantile is obtained by setting $\phi = 1$ so that

$$Q_t = Q_{t-1} + \eta_t, \qquad t = 2, ..., T.$$

The initial value, Q_1, is assumed to be drawn from a $N(0, \kappa)$ distribution. Letting $\kappa \to \infty$ gives a diffuse prior; see Durbin and Koopman (2001). A nonstationary quantile can also be modelled by a local linear trend

$$\begin{aligned} Q_t &= Q_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned} \tag{6}$$

where β_t is the slope and ζ_t is $NID(0,\ \sigma^2_{\zeta})$. It is well known that in a Gaussian model setting

$$Var\begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} = \sigma^2_{\zeta}\begin{bmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{bmatrix} \tag{7}$$

results in the smoothed estimates being a cubic spline. (In practice, the integrated random walk (IRW), obtained by suppressing η_t, gives very similar results). The spline connection will be taken up further in section 8.

3.2 Theory

If we assume an asymmetric double exponential distribution for the disturbance term in (4) and let the quantile be a first-order autoregression, as in (5), the logarithm of the joint density for the observations and the quantiles is, ignoring terms independent of quantiles,

$$\begin{aligned} J &= \log p(y_1, .., y_T, Q_1, .., Q_T) \\ &= -\frac{1}{2}\frac{(1-\phi^2)(Q_1 - Q^{\dagger})^2}{\sigma^2_{\eta}} - \frac{1}{2}\sum_{t=2}^{T}\frac{\eta_t^2}{\sigma^2_{\eta}} - \sum_{t=1}^{T}\frac{\rho_{\tau}(y_t - Q_t)}{\omega}. \end{aligned} \tag{8}$$

Given the observations, the estimated time-varying quantiles, $\widetilde{Q}_1, .., \widetilde{Q}_T$, are the values of the $Q_t's$ that maximise J. In other words they are the conditional modes.

When ε_t is $NID(0, \sigma^2)$ we may replace Q_t by the mean, μ_t, and write

$$y_t = \mu_t + \varepsilon_t, \qquad t = 1, ..., T \tag{9}$$

and J is redefined with $\rho_\tau(y_t - Q_t)/\omega$ in (8) replaced by $(y_t - \mu_t)^2/2\sigma^2$. Differentiating J with respect to $\mu_t, t = 1, ..., T$, setting to zero and solving gives the modes, $\widetilde{\mu}_t, t = 1, ..., T$, of the conditional distributions of the $\mu_t's$. For a multivariate Gaussian distribution these are the conditional expectations, which by definition are the smoothed (minimum mean square error) estimators. They may be efficiently computed by the Kalman filter and associated smoother; see de Jong (1989) and Durbin and Koopman (2001).

Returning to the quantiles and differentiating with respect to Q_t gives

$$\frac{\partial J}{\partial Q_t} = \frac{\phi Q_{t-1} - (1+\phi^2)Q_t + \phi Q_{t+1} + (1-\phi)^2 Q^\dagger}{\sigma_\eta^2} + \frac{1}{\omega} IQ\left(y_t - Q_t\right), \quad (10)$$

for $t = 2, \ldots, T-1$, and, at the endpoints,

$$\frac{\partial J}{\partial Q_1} = -\frac{(1-\phi^2)(Q_1 - Q^\dagger)}{\sigma_\eta^2} + \frac{\left[\phi\left(Q_2 - \phi Q_1\right) - \phi(1-\phi)Q^\dagger\right]}{\sigma_\eta^2} + \frac{1}{\omega} IQ\left(y_1 - Q_1\right)$$

and

$$\frac{\partial J}{\partial Q_T} = \frac{-\left(Q_T - \phi Q_{T-1}\right) + (1-\phi)Q^\dagger}{\sigma_\eta^2} + \frac{1}{\omega} IQ\left(y_T - Q_T\right)$$

where $IQ\left(y_t - Q_t\right)$ is defined as in (1). For $t = 2, .., T-1$, setting $\partial J/\partial Q_t$ to zero gives an equation that is satisfied by the estimated quantiles, $\widetilde{Q}_t, \widetilde{Q}_{t-1}$ and $\widetilde{Q}_{t+1}$, and similarly for $t = 1$ and T. If a solution is on a cusp, that is $\widetilde{Q}_t = y_t$, then $IQ\left(y_t - Q_t\right)$ is not defined as the check function is not differentiable at zero. Note that, when it does exist, the second derivative is negative.

For the random walk we can write the first term as $-{Q_1}^2/2\kappa$ and let $\kappa \to \infty$, noting that $\kappa = \sigma_\eta^2/(1-\phi^2)$. The derivatives of J are a simplified version of those in (10) with the first term in $\partial J/\partial Q_1$ dropping out. It is easy to see that the terms associated with the random walk, that is ignoring the $IQ_t's$, sum to zero. On the other hand, when Q_t is stationary, summing the terms in the derivatives not involving the $IQ_t's$ yields

$$-(1-\phi)Q_1 - (1-\phi)^2 \sum Q_t - (1-\phi)Q_T + (T-2)(1-\phi)^2 Q^\dagger + 2(1-\phi)Q^\dagger$$

divided by σ_η^2. However, setting

$$\widetilde{Q}^\dagger = \frac{(1-\phi)(\widetilde{Q}_1 + \widetilde{Q}_T) + (1-\phi)^2 \sum_{t=2}^{T-1} \widetilde{Q}_t}{(T-2)(1-\phi)^2 + 2(1-\phi)} \quad (11)$$

ensures that the sum is zero at the mode. The mean in the Gaussian model needs to be estimated in a similar way if the residuals, $y_t - \widetilde{\mu}_t$, $t = 1, ..., T$, are to sum to zero.

Establishing that the derivatives of that part of the criterion function associated with the time series model for the quantile sum to zero enables us to establish a fundamental property of time-varying quantiles, namely that *the number of observations that are less than the corresponding quantile, that is* $y_t < \widetilde{Q}_t$, *is no more than* $[T\tau]$ *while the number greater is no more than* $[T(1-\tau)]$. A general proof can be found in De Rossi and Harvey (2007).

A similar property can be obtained for regression quantiles. However, as Koenker (2005, p35-7) observes, the number of cusps rarely exceeds the number of explanatory variables.

Finally note that when the signal-noise ratio is zero, that is $\sigma_\eta^2 = 0$, there is usually only one cusp, while for a value of infinity, obtained when $\omega = 0$, there are T cusps as all quantiles pass through all observations.

3.3 The form of the solution

In a Gaussian model, (9), a little algebra leads to the classic Wiener-Kolmogorov (WK) formula for a doubly infinite sample. For the AR(1) model

$$\widetilde{\mu}_t = \mu + \frac{g}{g_y}(y_t - \mu)$$

where $\mu = E(\mu_t)$, $g = \sigma_\eta^2/((1-\phi L)(1-\phi L^{-1}))$ is the autocovariance generating function (ACGF), L is the lag operator, and $g_y = g + \sigma_\varepsilon^2$. The WK formula has the attraction that for simple models g/g_y can be expanded to give an explicit expression for the weights. Here

$$\widetilde{\mu}_t = \mu + \frac{q_\mu \theta}{\phi(1-\theta^2)} \sum_{j=-\infty}^{\infty} \theta^{|j|}(y_{t+j} - \mu) \tag{12}$$

where $q_\mu = \sigma_\eta^2/\sigma_\varepsilon^2$ and $\theta = (q_\mu + 1 + \phi^2)/2\phi - \left[\left(q_\mu + 1 + \phi^2\right)^2 - 4\phi^2\right]^{1/2}/2\phi$. This expression is still valid for the random walk except that μ disappears because the weights sum to one.

In order to proceed in a similar way with quantiles, we need to take account of cusp solutions by defining the corresponding $IQ_t's$ as the values that give equality of the associated derivative of J. Then we can set $\partial J/\partial Q_t$ in (10) equal to zero to give

$$\frac{\widetilde{Q}_t - Q^{\dagger}}{g} = \frac{1}{\omega} IQ(y_{t+j} - \widetilde{Q}_{t+j}) \tag{13}$$

for a doubly infinite sample with $Q^{\dagger}$ known. Using the lag operator yields

$$\widetilde{Q}_t = Q^{\dagger} + \frac{\sigma_\eta^2}{\omega} \sum_{j=-\infty}^{\infty} \frac{\phi^{|j|}}{1-\phi^2} IQ(y_{t+j} - \widetilde{Q}_{t+j}) \tag{14}$$

It is reassuring to note that a change in scale does not alter the form of the solution: if the observations are multiplied by a constant, then the quantile is multiplied by the same constant, as is the quasi 'signal-noise' ratio $q = \sigma_\eta^2/\omega$.

An expression for extracting quantiles that has a similar form to (12) can be obtained by adding $(\widetilde{Q}_t - Q^{\dagger})/\omega$ to both sides of (13) to give

$$\frac{\widetilde{Q}_t - Q^{\dagger}}{g} + \frac{\widetilde{Q}_t - Q^{\dagger}}{\omega} = \frac{\widetilde{Q}_t - Q^{\dagger}}{\omega} + \frac{IQ(y_t - \widetilde{Q}_t)}{\omega}$$

leading to

$$\widetilde{Q}_t = Q^{\dagger} + \frac{g}{g_y^{\omega}} \left[\widetilde{Q}_t - Q^{\dagger} + IQ(y_t - \widetilde{Q}_t) \right] \tag{15}$$

where $g_y^{\omega} = g + \omega$. (This is not an ACGF, but it can be treated as though it were). Thus, we obtain

$$\widetilde{Q}_t = Q^{\dagger} + \frac{q\theta}{\phi(1-\theta^2)} \sum_{j=-\infty}^{\infty} \theta^{|j|} [\widetilde{Q}_{t+j} - Q^{\dagger} + IQ(y_{t+j} - \widetilde{Q}_{t+j})] \tag{16}$$

where θ depends on ϕ and q. Since the solution satisfies (14), it must be the case that multiplying the observations by a constant means that the quantile is multiplied by the same constant, but that q, and hence θ, adapt accordingly. This argument carries over to any stationary linear model for the quantiles, with the ACGF, g, being appropriately modified in expression (15). This formulation suggests a procedure for computing the $\widetilde{Q}_t's$ in which synthetic 'observations', $\widehat{Q}_t - \widehat{Q}^{\dagger} + IQ(y_t - \widehat{Q}_t)$, are constructed using current estimates, $\widehat{Q}_t$, and inserted in a standard smoothing algorithm which is iterated to convergence. This is more convenient than basing an iterative algorithm on

(14). In the middle of a large sample, smoothing corresponds to the repeated application of (16).

For the random walk, (10) becomes

$$-\widetilde{Q}_{t-1}+2\widetilde{Q}_t-\widetilde{Q}_{t+1}=\Delta^2\widetilde{Q}_{t+1}=\frac{\sigma_\eta^2}{\omega}IQ(y_t-\widetilde{Q}_t)$$

and expression (14) can no longer be obtained. Since the weights in (16) sum to one, $Q^{\dagger}$ drops out giving

$$\widetilde{Q}_t=\frac{1-\theta}{1+\theta}\sum_{j=-\infty}^{\infty}\theta^{|j|}[\widetilde{Q}_{t+j}+IQ(y_{t+j}-\widetilde{Q}_{t+j})] \tag{17}$$

Thus standard smoothing algorithms can be applied to the synthetic observations, $\widehat{Q}_t+IQ(y_t-\widehat{Q}_t)$.

When the quantiles change over time they may be estimated non-parametrically. The simplest option is to compute them from a moving window; see, for example, Kuester et al (2006). More generally a quantile may be estimated at any point in time by minimising a local check function, that is

$$\min\sum_{j=-h}^{h}K(\frac{j}{h})\rho_\tau(y_{t+j}-Q_t)$$

where $K(.)$ is a weighting kernel and h is a bandwidth; see Yu and Jones (1998). Differentiating with respect to Q_t and setting to zero defines an estimator, $\widehat{Q}_t$, in the same way as was done in section 2. That is $\widehat{Q}_t$ must satisfy

$$\sum_{j=-h}^{h}K(\frac{j}{h})IQ(y_{t+j}-\widehat{Q}_t)=0$$

with $IQ(y_{t+j}-\widehat{Q}_t)$ defined appropriately if $y_{t+j}=\widehat{Q}_t$. Adding and subtracting $\widehat{Q}_t$ to each of the $IQ(y_{t+j}-\widehat{Q}_t)$ terms in the sum, leads to

$$\widehat{Q}_t=\frac{1}{\sum_{j=-h}^{h}K(j/h)}\sum_{j=-h}^{h}K(\frac{j}{h})[\widehat{Q}_t+IQ(y_{t+j}-\widehat{Q}_t)].$$

It is interesting to compare this with the weighting scheme implied by the random walk model where $K(j/h)$ is replaced by $\theta^{|j|}$ so giving an (infinite) exponential decay. An integrated random walk implies a kernel with a slower decline for the weights near the centre; see Harvey and Koopman (2000). The time series model determines the shape of the kernel while the signal-noise ratio plays the same role as the bandwidth. Note also that in the model-based formula, $\widehat{Q}_{t+j}$ is used instead of $\widehat{Q}_t$ when j is not zero.

Of course, the model-based approach has the advantage that it automatically determines a weighting pattern at the end of the sample that is consistent with the one in the middle.

3.4 Algorithms for computing time-varying quantiles

An algorithm for computing time-varying quantiles needs to find values of $\widetilde{Q}_t, t = 1, .., t$, that set each of the derivatives of J in (10) equal to zero when the corresponding $\widetilde{Q}_t$ is not set equal to y_t. The algorithm described in De Rossi and Harvey (2006) combines an iterated state space smoother with a simpler algorithm that takes account of cusp solutions. It is suggested that parameters be estimated by cross-validation.

4 Dispersion, Asymmetry and Value at Risk

The time-varying quantiles provide a comprehensive description of the distribution of the observations and the way it changes over time. The choice of quantiles will depend on what aspects of the distribution are to be highlighted. The lower quantiles, in particular 1% and 5%, are particularly important in characterizing value at risk over the period in question. A contrast between quantiles may be used to focus attention on changes in dispersion or asymmetry.

4.1 Dispersion

The contrasts between complementary quantiles, that is

$$D(\tau) = Q_t(1-\tau) - Q_t(\tau), \qquad \tau < 0.5, \qquad t = 1, .., T$$

yield measures of dispersion. A means of capturing an evolving interquartile range, $D(0.25)$, provides an alternative to GARCH and stochastic volatility models. As Bickel and Lehmann (1976) remark 'Once outside the normal model, scale provides a more natural measure of dispersion than vari-

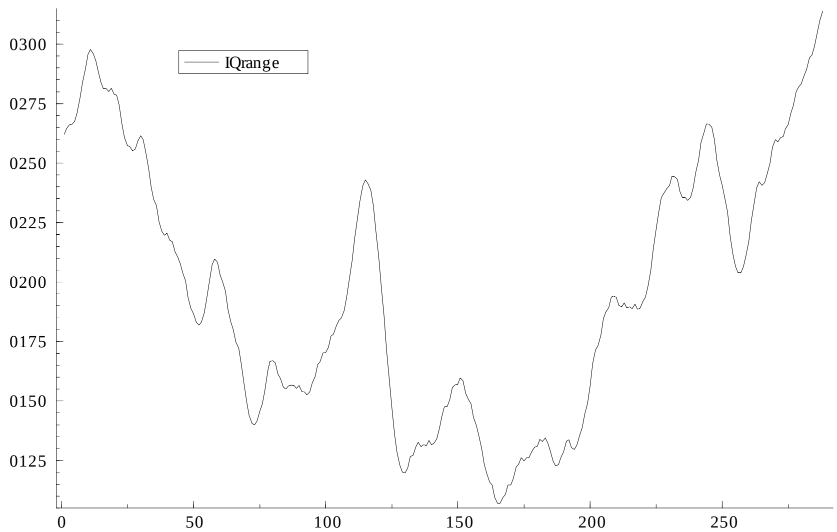

Figure 2: Interquartile range for US inflation

ance......and offers substantial advantages from the robustness[4] viewpoint'. Other contrasts may give additional information.

Figure 2 shows the interquartile range for US inflation.

4.1.1 Imposing symmetry

If the distribution is assumed to be symmetric around zero, better estimates of $Q_t(\tau)$ and $Q_t(1-\tau)$ can be obtained by assuming they are the same[5]. This can be done simply by estimating the $(1-2\tau)th$ quantile for $|y_t|$. Thus to compute $\widetilde{Q}_t(0.75) = -\widetilde{Q}_t(0.25)$ we estimate the median for $|y_t|$. Doubling this series gives the interquartile range.

[4]Koenker and Zhao (1996, p794) quote this observation and then go on to model scale (of a zero mean series) as a linear combination of past absolute values of the observations (rather than variance as a linear combination of past squares as in GARCH). They then fit the model by quantile regression. As we will argue later - in section 7 - this type of approach is not robust to additive outliers.

[5]A test of the null hypothesis that the observations are symmetric about zero could be carried out by noting that for IID observations $\widetilde{Q}(1-\tau) + \widetilde{Q}(\tau)$ is asymptotically normal with mean zero and variance $(2\tau - \tau^2)/f^2(Q(\tau))$.

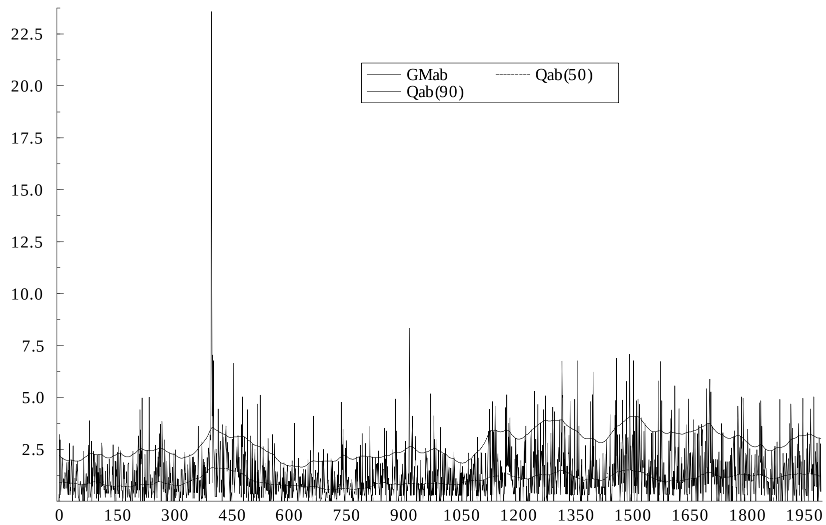

Figure 3: 25% and 5% quantiles estimated from absolute values of GM returns

Figure 3 shows absolute values and the resulting estimates of the symmetric 5%/95% and 25%/75% quantiles for the raw series. The square roots of q for the 5%/95% and 25%/75% quantiles were 0.11 and 0.06 respectively. (These are close to the values obtained for the individual corresponding quantiles, though it is not clear that they should necessarily be the same).

Figure 4 shows estimates of the standard deviation of GM obtained by fitting a random walk plus noise to the squared observations and then taking the square root of the smoothed estimates[6]. These can be compared with the estimates of the interquartile range and the 5%-95% range, obtained by doubling the quantiles 50% and 90%, respectively, of the absolute value of the series. Dividing the 5%-95% range by 3.25 is known to give a good estimate of the standard deviation for a wide range of distributions; see, for example, the recent study by Taylor (2005). Estimation was carried out using the observations from 1 to 2000, but the graph shows 300 to 600 in order to

[6]This approach can be regarded as a combination of Riskmetrics and UC. Estimating IGARCH parameters and using these to construct smoothed estimates (by calculating the implied signal-noise ratio) can be expected to give a similar result, as will fitting an SV model.

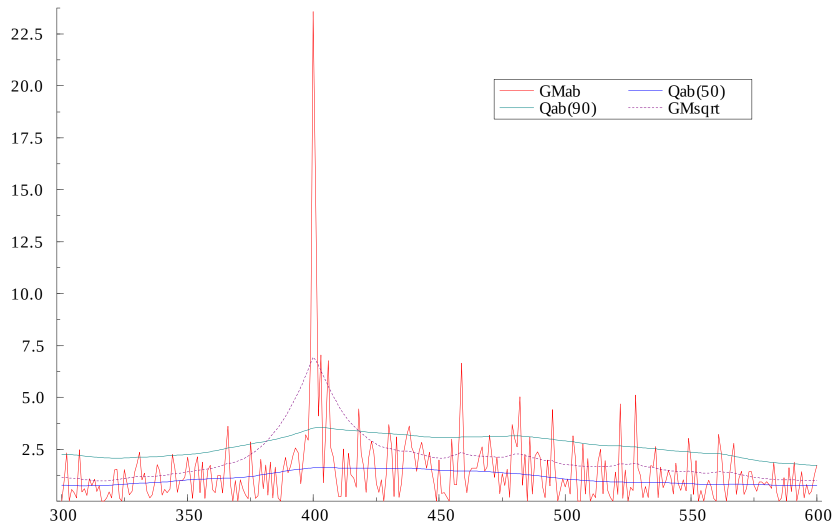

Figure 4: 25% and 5% quantiles estimated from absolute values of GM returns, together with SD from smoothed squared observations

highlight the differences emanating from the large outliers around 400. As can be seen the quantile ranges are less volatile than the variance estimates.

4.1.2 Tail dispersion

A comparison between the interquartile range and the interdecile range (0.1 to 0.9), or the 0.05 to 0.95 range, could be particularly informative in pointing to different behaviour in the tails. These contrasts can be constructed with or without symmetry imposed.

We generated 500 observations from a scale mixture of two Gaussian distributions with time-varying weights and variances chosen so as to keep the overall variance constant over time. Specifically, while the variance of the first component was set equal to one, the variance of the second component was allowed to increase, as a linear function of time, from 20 to 80. The weights used to produce the mixture were adjusted so as to keep the overall variance constant, at a value of ten. As a result, the shape of the distribution changes during the sample period, with the tail dispersion decreasing, as can be seen from the 5% quantile shown in figure 5. The cross validation

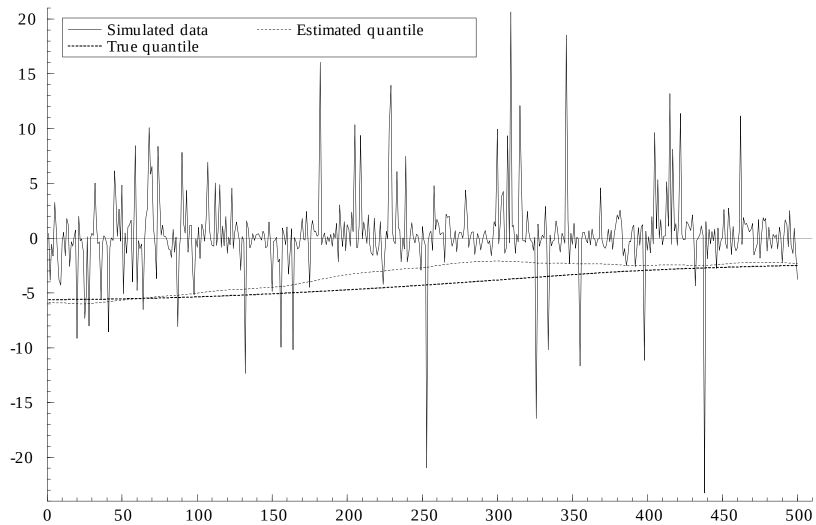

Figure 5: 5% quantile and estimate from simulated contaminated normal distribution

procedure was then used to select the signal-noise ratio for RW model and extract the corresponding 5% quantile. The square root of the estimated ratio was 0.1. As can be seen from figure 5, it tracks the true quantile quite well (and an estimate based on absolute values would be even better) . When a stochastic volatility model was estimated by quasi-maximum likelihood using STAMP, it produced a constant variance.

As a second example we gradually changed the mixture in a contaminated normal distribution to give the 5% quantile shown in figure 6. The estimated quantile is shown in figure 7.

Some notion of the way in which tail dispersion changes can be obtained by plotting the ratio of the 0.05 to 0.95 range to the interquartile range (without imposing symmetry), that is

$$\frac{\widetilde{D}(0.05)}{\widetilde{D}(0.25)} = \frac{\widetilde{Q}_t(0.95) - \widetilde{Q}_t(0.05)}{\widetilde{Q}_t(0.75) - \widetilde{Q}_t(0.25)}. \tag{18}$$

For a normal distribution this ratio is 2.44, for t_3 it is 3.08 and for a Cauchy 6.31. Figure 8 shows the 5% and 25 % quantiles and the plot of (18) for the GM series.

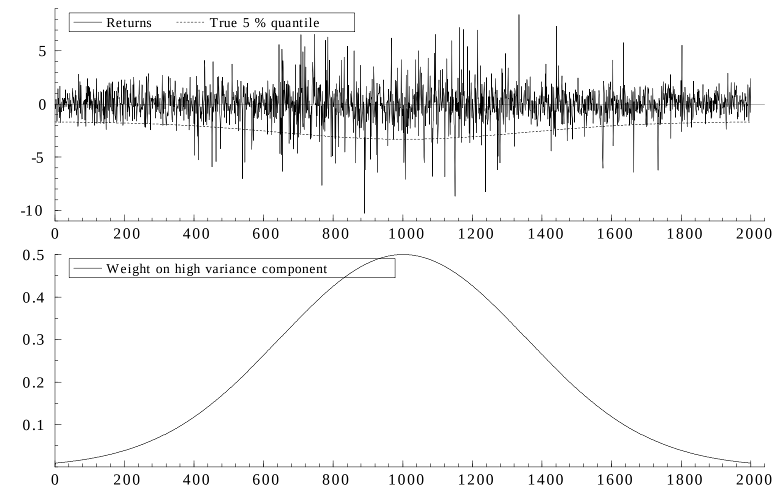

Figure 6: Contaminated normal, variances one and nine with weight on the high variance component given by $\exp(-((t-1000)/500)^2)/2$.

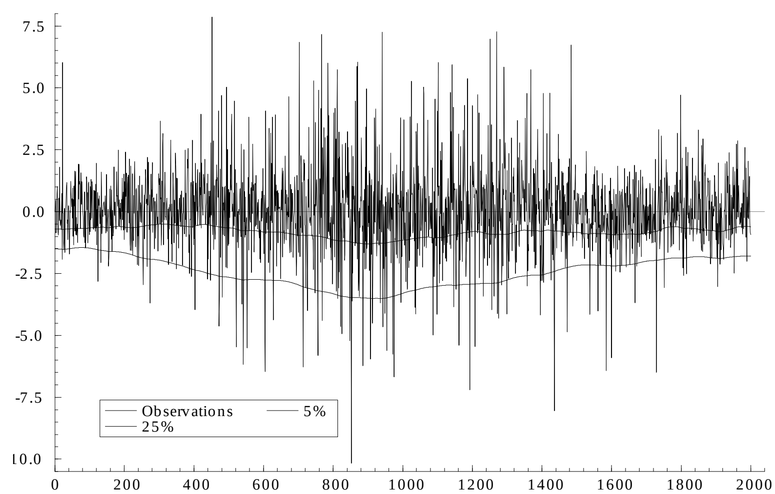

Figure 7: Estimates of 5% and 95% quantiles from data in previous figure.

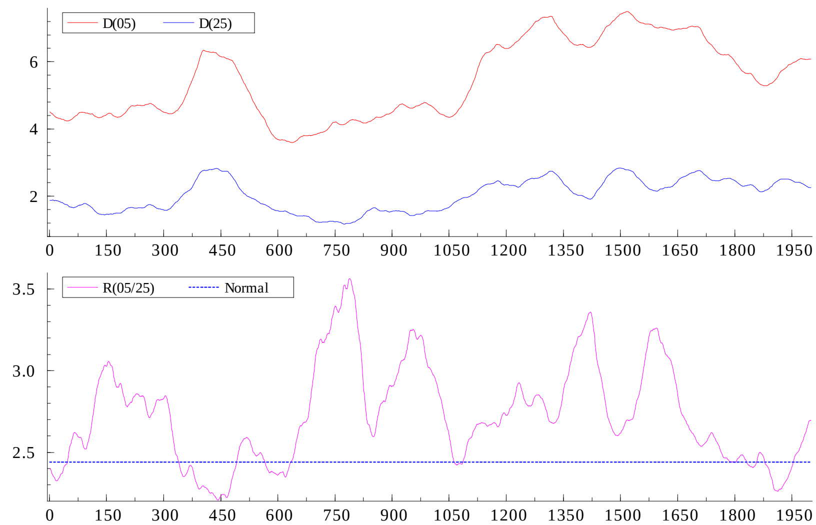

Figure 8: Interquartile, 5%-95% range and (lower graph) their ratio

4.2 Asymmetry

For a symmetric distribution

$$S(\tau) = Q_t(\tau) + Q_t(1-\tau) - 2Q_t(0.5), \qquad \tau < 0.5 \tag{19}$$

is zero for all $t = 1, .., T$. Hence a plot of this contrast shows how the asymmetry captured by the complementary quantiles, $Q_t(\tau)$ and $Q_t(1-\tau)$ changes over time.

If there is reason to suspect asymmetry, changes in tail dispersion can be investigated by plotting

$$S(0.05/0.25) = \frac{\widetilde{Q}_t(0.05) - \widetilde{Q}_t(0.5)}{\widetilde{Q}_t(0.25) - \widetilde{Q}_t(0.5)}$$

and similarly for $\widetilde{Q}_t(0.75)$ and $\widetilde{Q}_t(0.95)$.

5 Tests of time invariance (IQ tests)

Before estimating a time-varying quantile it may be prudent to test the null hypothesis that it is constant. Such a test may be based on the sample quantile indicator or *quantic*, $IQ(y_t - \widetilde{Q}(\tau))$. If the alternative hypothesis is that $Q_t(\tau)$ follows a random walk, a modified version of the basic stationarity test is appropriate; see Nyblom and Mäkeläinen (1983) and Nyblom (1989). This test is usual applied to the residuals from a sample mean and because the sum of the residuals is zero, the asymptotic distribution of the test statistic is the integral of squared Brownian bridges and this is known to be a Cramér-von Mises (CvM) distribution; the 1%, 5% and 10% critical values are 0.743, 0.461 and 0.347 respectively. Nyblom and Harvey (2001) show that the test has high power against an integrated random walk while Harvey and Streibel (1998) note that it is also has a locally best invariant (LBI) interpretation as a test of constancy against a highly persistent stationary AR(1) process. This makes it entirely appropriate for the kind of situation we have in mind for time-varying quantiles.

Assume that under the null hypothesis the observations are IID. The population quantile indicators, $IQ(y_t - Q(\tau))$, have a mean of zero and a variance of $\tau(1-\tau)$. The quantics sum to zero if, when an observation is equal to the sample quantile, the corresponding quantic is defined to ensure that this is the case; see section 2. Hence the stationarity test statistic

$$\eta_\tau = \frac{1}{T^2\tau(1-\tau)} \sum_{t=1}^{T} \left(\sum_{i=1}^{t} IQ(y_i - \widetilde{Q}(\tau)) \right)^2 \tag{20}$$

has the CvM distribution (asymptotically). De Jong, Amsler and Schmidt (2006) give a rigorous proof for the case of the median.[7] Carrying over their assumptions to the case of quantiles and IID observations we find that all that is required is that $Q(\tau)$ be the unique population $\tau-$quantile and that $y_t - Q(\tau)$ has a continuous positive density in the neighbourhood of zero.

Test statistics for GM are shown in the table below for a range of quantiles. All (including the median) reject the null hypothesis of time invariance at the 1% level of significance. However, the test statistic for $\tau = 0.01$ is only 0.289 which is not significant at the 10% level. This is consistent with the small estimate obtained for the signal-noise ratio. Similarly the statistic is 0.354 for $\tau = 0.99$ is only just significant at the 10% level.

[7] De Jong, Amsler and Schmidt (2005) are primarily concerned with a more general version of the median test when there is serial correlation under the null hypothesis. Assuming strict stationarity (and a mixing condition) they modify the test of Kwiatkowski et al (1992) - the KPSS test - so that it has the CvM distribution under the null. Similar modifications could be made to quantile tests. More general tests, for both stationary and nonstationary series are currently under investigation.

τ	.05	.25	.50	.75	.95
η_τ	1.823	1.026	2.526	1.544	2.962

De Jong et al (2005) also reject constancy of the median for certain weekly exchange rates. The stationarity test with no correction for serial correlation does not reject in these cases. The same is true here as its value is only 0.146. A plot of the estimated median shows that is it close to zero most of the time apart from a spell near the beginning of the series and a shorter one near the end.

5.1 Correlation between quantics

Tests involving more than one quantic need to take account of the correlation between them. To find the covariance of the τ_1 and τ_2 population quantics with $\tau_2 > \tau_1$, their product must be evaluated and weighted by (i) τ_1 when $y_t < Q(\tau_1)$, (ii) $\tau_2 - \tau_1$ when $y_t > Q(\tau_1)$ but $y_t < Q(\tau_2)$, (iii) $1 - \tau_2$ when $y_t > Q(\tau_2)$. This gives

$$(\tau_2 - 1)(\tau_1 - 1)\tau_1 + (\tau_2 - 1)\tau_1(\tau_2 - \tau_1) + \tau_2\tau_1(1 - \tau_2)$$

and on collecting terms we find that

$$cov(IQ_t(\tau_1), IQ_t(\tau_2)) = \tau_1(1 - \tau_2), \qquad \tau_2 > \tau_1. \tag{21}$$

where $IQ_t(\tau)$ denotes $IQ(y_t - Q(\tau))$. It follows that the correlation between the population[8] quantics for τ_1 and τ_2 is

$$\frac{\tau_1(1 - \tau_2)}{\sqrt{\tau_1(1 - \tau_1)\tau_2(1 - \tau_2)}}, \qquad \tau_2 > \tau_1$$

The correlation between the complementary quantics, $IQ_t(\tau)$ and $IQ_t(1-\tau)$, is simply $\tau/(1-\tau)$. This is 1/3 for the quartiles and 1/9 for the first and last deciles, that is 0.1 and 0.9.

5.2 Contrasts

A test based on a quantic contrast can be useful in pointing to specific departures from a time-invariant distribution. A quantic contrast of the form

$$aIQ_t(\tau_1) + bIQ_t(\tau_2), \qquad t = 1, ..., T,$$

[8] When both $T\tau_1$ and $T\tau_2$ are integers the sample covariance and correlation are exactly the same; see the earlier footnote on variances.

where a and b are constants, has a mean of zero and a variance that can be obtained directly from (21) as

$$a^2Var\left(IQ_t\left(\tau_1\right)\right) + b^2Var\left(IQ_t\left(\tau_2\right)\right) + 2abCov(IQ_t\left(\tau_1\right), IQ_t\left(\tau_2\right)). \qquad (22)$$

Tests statistics analogous to η_τ, constructed from the sample quantic constrasts, again have the CvM distribution when the observations are IID.

A test of constant dispersion can be based on the complementary quantic contrast

$$DIQ_t(\tau) = IQ_t(1-\tau) - IQ_t(\tau), \qquad \tau < 0.5. \qquad (23)$$

This mirrors the quantile contrast, $Q_t(1-\tau) - Q_t(\tau)$, used as a measure of dispersion. It follows from (22) that the variance of (23) is $2\tau(1-2\tau)$. For GM, the test statistics for the interquartile range and the 5%/95% range are 3.589 and 3.210 respectively. Thus both decisively reject.

A test of changing asymmetry may be based on

$$SIQ_t(\tau) = IQ_t(\tau) + IQ_t(1-\tau), \qquad \tau < 0.5. \qquad (24)$$

The variance of $SIQ_t(\tau)$ is 2τ and it is uncorrelated with $DIQ(\tau)$. The sample quantic contrast takes the value -1 when $y_t < \tilde{Q}\left(\tau\right)$, 1 when $y_t > \tilde{Q}\left(1-\tau\right)$ and zero otherwise.

For GM the asymmetry test statistics are 0.133 and 0.039 for $\tau = 0.25$ and $\tau = 0.05$ respectively. Thus neither rejects at the 10% level of significance.

6 Prediction, specification testing and conditional quantile autoregression

In this section we investigate the relationship between our methods for predicting quantiles and those based on conditional quantile autoregressive models.

6.1 Filtering and Prediction

The smoothed estimate of a quantile at the end of the sample, $\widetilde{Q}_{T|T}$, is the filtered estimate or 'nowcast'. Predictions, $\widetilde{Q}_{T+j|T}$, $j = 1, 2, \ldots$, are made by straightforwardly extending these estimates according to the time series model for the quantile. For a random walk the predictions are $\widetilde{Q}_{T|T}$ for all lead times, while for a more general model in SSF, $\widetilde{Q}_{T+j|T} = \mathbf{z}'\mathbf{T}^j\widetilde{\boldsymbol{\alpha}}_T$. As new observations become available, the full set of smoothed estimates should theoretically be calculated, though this should not be very time consuming given the starting value will normally be close to the final solution. Furthermore, it may be quite reasonable to drop the earlier observations by having a cut-off, δ, such that only observations from $t = T - \delta + 1$ to T are used.

Insight into the form of the filtered estimator can be obtained from the weighting pattern used in the filter from which it is computed by repeated applications; compare the weights used to compute the smoothed estimates in sub-section 3.3. For a random walk quantile and a semi-infinite sample the filtered estimator must satisfy

$$\widetilde{Q}_{t|t} = (1-\theta)\sum_{j=0}^{\infty}\theta^j[\widetilde{Q}_{t-j|t} + IQ(y_{t-j} - \widetilde{Q}_{t-j|t})] \tag{25}$$

where $\widetilde{Q}_{t-j|t}$ is the smoothed estimator of Q_{t-j} based on information at time t; see, for example, Whittle (1983, p69). Thus $\widetilde{Q}_{t|t}$ is an exponentially weighted moving average (EWMA) of the synthetic observations, $\widetilde{Q}_{t-j|t} + IQ(y_{t-j} - \widetilde{Q}_{t-j|t})$.

6.2 Specification and diagnostic testing

The one-step ahead prediction indicators in a post sample period are defined by

$$\widetilde{\nu}_t = IQ(y_t - \widetilde{Q}_{t|t-1}), \qquad t = T+1, ..., T+L$$

If these can be treated as being serially independent, the test statistic

$$\xi(\tau) = \frac{\sum_{t=T+1}^{T+L} IQ(y_t - \widetilde{Q}_{t|t-1})}{\sqrt{L\tau(1-\tau)}} \tag{26}$$

is asymptotically standard normal. (The negative of the numerator is L times the proportion of observations below the predicted quantile minus τ). The suggestion is that this be used to give an internal check on the model; see also Engle and Manganelli (2004, section 5).

Figure 9 shows the one-step ahead forecasts of the 25% quantile for GM from observation 2001 to 2500. As expected these filtered estimates are more variable than the smoothed estimates shown in figure 1. The proportion of observations below the predicted value is 0.27. The test statistic, $\xi(0.25)$, is -1.43.

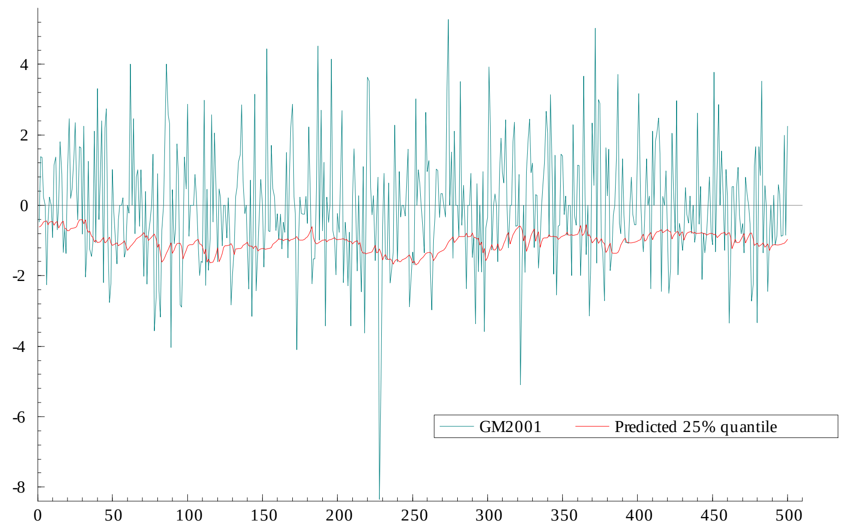

Figure 9: One-step ahead forecasts of the 25% quantile for GM from 2001 to 2500

6.3 Quantile autoregression

In quantile autoregression (QAR), the (conditional) quantile is assumed to be a linear combination of past observations; see, for example, Herce (1996), Jurekova and Hallin (1999), Koenker and Xiao (2004), Komunjer (2005), Koenker (2005, p. 126-8, 260-5) and the references therein. The parameters in a QAR are estimated by minimizing a criterion function as in (3). Koenker and Xiao (2005) allow for different coefficients for each quantile, so that

$$Q_t(\tau) = \alpha_{0,\tau} + \alpha_{1,\tau} y_{t-1} + ... + \alpha_{p,\tau} y_{t-p}, \quad t = p+1, ..., T.$$

If the sets of coefficients of lagged observations, that is $\alpha_{1,\tau}, .., \alpha_{p,\tau}$, are the same for all τ, the quantiles will only differ by a constant amount. Koenker and Xiao (2006) provide a test of this hypothesis and give some examples of fitting conditional quantiles to real data.

In a Gaussian signal plus noise model, the optimal (MMSE) forecast, the conditional mean, is a linear function of past observations. This implies an autoregressive representation, though, if the lag is infinite, an ARMA model might be more practical. When the Gaussian assumption is dropped there are two responses. The first is to stay within the autoregressive framework and assume that the disturbance has some non-Gaussian distribution. The second is to put a non-Gaussian distribution on the noise and possibly on the signal as well. If a model with non-Gaussian additive noise is appropriate the consequence is that the conditional mean is no longer a linear function of past observations. Hence the MMSE will, in general, be non-linear[9]. A potentially serious practical implication is that if the additive noise is drawn from a heavy-tailed distribution, the autoregressive forecasts will be sensitive to outliers induced in the lagged observations. Assuming a non-Gaussian distribution for the innovations driving the autoregression does not deal with this problem.

The above considerations are directly relevant to the formulation of dynamic quantile models. While the QAR model is useful in some situations, it is not appropriate for capturing slowly changing quantiles in series of returns. It could, however, be adapted by the introduction of a measurement equation so that

$$\begin{aligned} y_t &= Q_t(\tau) + \varepsilon_t(\tau), \qquad t = 1, ..., T \\ Q_t(\tau) &= \alpha_{0,\tau} + \alpha_{1,\tau} Q_{t-1}(\tau) + ... + \alpha_{p,\tau} Q_{t-p}(\tau) + \eta_t(\tau), \end{aligned} \tag{27}$$

where both ε_t and η_t are taken to follow asymmetric double exponential distributions. Indeed, the cubic spline LP algorithm of Koenker *et al* (1994) is essentially fitting a model of this form.

6.4 Nonlinear QAR and CaViaR

Engle and Manganelli (2004) suggest a general nonlinear dynamic quantile model in which the conditional quantile is

$$Q_t(\tau) = \alpha_0 + \sum_{i=1}^{q} \beta_i Q_{t-i}(\tau) + \sum_{j=1}^{r} \alpha_j f(y_{t-j}). \tag{28}$$

The information set in $f()$ can be expanded to include exogenous variables. This is their CAViaR specification. Suggested forms include the symmetric absolute value

$$Q_t(\tau) = \alpha_0 + \beta Q_{t-1}(\tau) + \gamma \left| y_{t-1} \right| \tag{29}$$

[9] Though the estimator constructed under the Gaussian assumption is still the MMSLE - ie best linear estimator.

and a more general specification that allows for different weights on positive and negative returns. Both are assumed to be mean reverting. They also propose an adaptive model

$$Q_t(\tau) = Q_{t-1}(\tau) + \gamma\{[1 + \exp(G[y_{t-1} - Q_{t-1}(\tau)])]^{-1} - \tau\}, \qquad (30)$$

where G is some positive number, and an indirect GARCH (1,1) model

$$Q_t(\tau) = (\alpha_0 + \beta Q_{t-1}^2(\tau) + \alpha_1 y_{t-1}^2)^{1/2}. \qquad (31)$$

There is no theoretical guidance as to suitable functional forms for CAViaR models and Engle and Manganelli (2004) place a good deal of emphasis on developing diagnostic cheching procedures. However, it might be possible to design CAViaR specifications based on the notion that they should provide a reasonable approximation to the filtered estimators of time-varying quantiles that come from signal plus noise models. Under this interpretation, $Q_t(\tau)$ in (28) is not the actual quantile so a change in notation to $\widehat{Q}_{t|t-1}(\tau)$ is helpful. Similarly the lagged values, which are approximations to smoothed estimators calculated as though they were filtered estimators, are best written as $\widehat{Q}_{t-j|t-1-j}(\tau)$, $j = 1, ..., q$. The idea is then to compute the $\widehat{Q}_{t|t-1}(\tau)'s$ - for given parameters - with a single recursion[10]. The parameters are estimated, as in CAViaR, by minimizing the check function formed from the one-step ahead prediction errors.

For a random walk quantile, a CAViaR approximation to the recursion that yields (25) is

$$\widehat{Q}_{t|t-1} = \widehat{Q}_{t-1|t-2} + (1-\theta)(\widehat{y}_{t-1} - \widehat{Q}_{t-1|t-2})$$

with $\widehat{y}_t = \widehat{Q}_{t|t-1} + IQ(y_{t-1} - \widehat{Q}_{t-1|t-2})$. This simplifies to

$$\widehat{Q}_{t|t-1} = \widehat{Q}_{t-1|t-2} + (1-\theta)\widehat{\nu}_{t-1}, \qquad (32)$$

where

$$\widehat{\nu}_t = IQ(y_t - \widehat{Q}_{t|t-1})$$

is an indicator that plays an analogous role to that of the innovation, or one-step ahead prediction error, in the standard Kalman filter. More generally, the CAViaR approximation can be obtained from the Kalman filter

[10] The recursion could perhaps be initialized with $\widehat{Q}'_{t|t-1}s$ set equal to the fixed quantile computed from a small number of observations at at the beginning of the sample.

for the underlying UC model with the innovations given by $\widehat{\nu}_t$. For the integrated random walk quantile, this filter can, if desired, be written as a single recursion

$$\widehat{Q}_{t|t-1} = 2\widehat{Q}_{t-1|t-2} - \widehat{Q}_{t-2|t-3} + k_1\widehat{\nu}_{t-1} + k_2\widehat{\nu}_{t-2},$$

where k_1 and k_2 depend on the signal-noise ratio.

The recursion in (32) has the same form as the limiting case ($G \to \infty$) of the adaptive CAViaR model, (30). Other CAViaR specifications are somewhat different from what might be expected from the filtered estimators given by UC specification. One consequence is that models like (29) and (31), which are based on actual values, rather than indicators, may suffer from a lack of robustness to additive outliers. That this is the case is clear from an examination of figure 1 in Engle and Manganelli (2004, p373). More generally, recent evidence on predictive performance in Kuester et al (2006, p 80-1) indicates a preference for the adaptive specification.

Some of the CAViaR models suffer from a lack of identifiability if the quantile in question is time invariant. This is apparent in (28) where a time invariant quantile is obtained if either $\alpha_j = \beta_i = 0$ for all i and non-zero j and $\alpha_0 \neq 0$, or $\beta_1 = 1$ and all the other coefficients are zero. The same thing happens with the indirect GARCH specification (31) and indeed with GARCH models in general; see Andrews (2001, p 711). These difficulties do not arise if we adopt functional forms suggested by signal extraction.

7 Nonparametric regression with cubic splines

A slowly changing quantile can be estimated by minimizing the criterion function $\sum \rho_\tau\{y_t - Q_t\}$ subject to smoothness constraints. The cubic spline solution seeks to do this by finding a solution to

$$\min \sum \rho_\tau\{y_t - Q(x_t)\} + \lambda\left(\int \{Q''(x)\}^2 dx\right) \tag{33}$$

where $Q(x)$ is a continuous function, $0 \leq x \leq T$ and $x_t = t$. The parameter λ controls the smoothness of the spline. It is shown in De Rossi and Harvey (2007) that the same cubic spline is obtained by quantile signal extraction of (6) and (7) with $\lambda = \omega/2\sigma_\zeta^2$. A random walk corresponds to $g'(x)$ rather than $g''(x)$ in the above formula; compare Kohn, Ansley and Wong (1992). The proof not only shows that the well-known connection between splines and stochastic trends in Gaussian models carries over to quantiles, but it does so in a way that yields a more compact proof for the Gaussian case.

The SSF allows irregularly spaced observations to be handled since it can deal with systems that are not time invariant. The form of such systems is the implied discrete time formulation of a continuous time model; see Harvey (1989, ch 9). For the random walk, observations δ_t time periods imply a variance on the discrete random walk of $\delta_t \sigma_\eta^2$, while for the continuous time IRW, (7) becomes

$$Var\begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} = \sigma_\zeta^2 \begin{bmatrix} (1/3)\delta_t^3 & (1/2)\delta_t^2 \\ (1/2)\delta_t^2 & \delta_t \end{bmatrix} \tag{34}$$

while the second element in the first row of the transition matrix is δ. The importance of this generalisation is that it allows the handling of nonparametric quantile regression by cubic splines when there is only one explanatory variable[11]. The observations are arranged so that the values of the explanatory variable are in ascending order. Then x_t is the $t-th$ value and $\delta_t = x_t - x_{t-1}$.

Bosch, Ye and Woodworth (1995) propose a solution to cubic spline quantile regression that uses quadratic programming[12]. Unfortunately this necessitates the repeated inversion of large matrices of dimension $4T \times 4T$. This is very time consuming[13]. Our signal extraction algorithm is far more efficient (and more general) and makes estimation of λ a feasible proposition. Bosch, Ye and Woodworth had to resort to making λ as small as possible without the quantiles crossing.

An example of cubic spline quantile regression is provided by the "motorcycle data", which records measurements of the acceleration of the head of a dummy in motorcycle crash tests; see De Rossi and Harvey (2007). These have been used in a number of textbooks, including Koenker (2005, p 222-6). Harvey and Koopman (2000) illustrate the stochastic trend connection.

[11] Other variables can be included if they enter linearly into the equation that specifies Q_t.

[12] Koenker et al (1994) study a general class of *quantile smoothing splines*, defined as solutions to

$$\min \sum \rho\{y_i - g(x_i)\} + \lambda \left(\int |g''(x)|^p \right)^{1/p}$$

and show that the $p = 1$ case can be handled by LP algorithm.

[13] Bosch et al report that it takes almost a minute on a Sun workstation for sample size less that 100.

8 Conclusion

Our approach to estimating time-varying quantiles is, we believe, conceptually very simple. Furthermore the algorithm based on the KFS appears to be quite efficient and cross-validation criterion appears to work well.

The time-varying quantiles up a wealth of possibilities for capturing the evolving distribution of a time series. In particular the lower quantiles can be use to directly estimate the evolution of value at risk, VaR, though one of our findings is that detecting changes in the 1% VaR for a given series may not be a feasible proposition. We have also illustrated how changes in dispersion and asymmetry can be highlighted by suitably constructed contrasts and how the disperson measures compare with the changing variances obtained from GARCH and SV models. Stationarity tests of whether these contrasts change over time can be constructed and their use illustrated with real data.

At the end of the series, a filtered estimate of the current state of a quantile (or quantiles) provides the basis for making forecasts. As new observations become available, updating can be carried out quite efficiently. The form of the filtering equations suggests ways of modifying the CaViaR specifications proposed by Engle and Manganelli (2004). Combining the two approaches could prove fruitful and merits further investigation.

Newey and Powell (1987) have proposed expectiles as a complement to quantiles. De Rossi and Harvey (2007) explain how the algorithm for computing time-varying quantiles can also be applied to expectiles. Indeed it is somewhat easier since there is no need to take account of cusp solutions and convergence is rapidly achieved by the iterative Kalman filter and smoother algorithm. The proof of the spline connection also applies to expectiles. Tests of time invariance of expectiles can also be constructed by adapting stationarity tests in much the same way as was done for quantiles. Possible gains from using expectiles and associated tests are investigated in Busetti and Harvey (2007).

Acknowledgement *Earlier versions of this paper were presented at the Econometrics and MFE workshops at Cambridge, at the Portuguese Mathematical Society meeting in Coimbra, at the Tinbergen Institute, Amsterdam, the LSE, Leeds, CREST-Paris and Ente Einaudi in Rome. We would like to thank participants, including Jan de Gooijer, Siem Koopman, Franco Peracchi and Chris Rogers for helpful comments. Special thanks go to Stepana Lazarova who was the discussant when the paper was presented at the CIMF workshop at Cambridge.*

References

Andrews, D.W.K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica,* 69, 683-734.

Andrews. D.F., Tukey, J. W., Bickel, P. J. (1972). *Robust estimates of location.* Princeton: Princeton University Press.

Bosch, R.J.,Ye, Y. and G.G. Woodworth (1995). A convergent algorithm for quantile regression with smoothing splines, *Computational Statistics and Data analysis*, 19, 613-30.

Busetti, F and A. C. Harvey (2007). Tests of time invariance. Faculty of Economics Discussion paper, Cambridge.

Chernozhukov, V. and Umantsev, L., (2001). Conditional value-at-risk: aspects of modeling and estimation. *Empirical Economics* 26, 271-292.

Christoffersen, P., Hahn, J. and A. Inoue (2001). Testing and comparing Value-at-Risk measures. *Journal of Empirical Finance,* 8, 325-42.

De Rossi, G. and A. C. Harvey (2007). Time-varying quantiles. Faculty of Economics, Cambridge, CWPE 0649.

De Rossi, G. and A. C. Harvey (2007). Quantiles, expectiles and splines. Faculty of Economics Discussion paper, Cambridge.

de Jong, P. (1988). A cross-validation filter for time series models. *Biometrika* 75, 594-600.

de Jong, P. (1989). Smoothing and interpolation with the state-space model. *Journal of the American Statistical Association* 84, 1085-1088.

de Jong, R.M., Amsler, C. and P. Schmidt (2006). A Robust Version of the KPSS Test based on Indicators, *Journal of Econometrics*, (to appear).

Durbin, J. and S.J. Koopman, (2001). *Time series analysis by state space methods.* Oxford University Press, Oxford.

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

Harvey A.C. (1989). *Forecasting, structural time series models and the Kalman filter* (Cambridge: Cambridge University Press).

Harvey A.C. (2006). Forecasting with Unobserved Components Time Series Models. *Handbook of Economic Forecasting*, edited by G Elliot, C Granger and A Timmermann. North Holland, Amsterdam.

Harvey, A.C. and J. Bernstein (2003). Measurement and testing of inequality from time series deciles with an application to US wages. *Review* *of Economics and Statistics,* 85, 141-52.

HARVEY, A. C., AND M. STREIBEL (1998). Testing for a slowly changing level with special reference to stochastic volatility. *Journal of Econometrics* 87, 167-89.

HARVEY, A.C., AND S.J KOOPMAN (2000). Signal extraction and the formulation of unobserved components models, *Econometrics Journal,* **3,** 84-107.

KIMELDORF, G., AND G. WAHBA (1971). Some results on Tchebycheffian spline functions. *Journal of mathematical analysis and applications* 33, 82–95.

KOENKER, R. (2005) *Quantile regression*. (Cambridge: Cambridge University Press).

KOENKER, R. AND Q. ZHAO (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* 12, 793-813.

KOENKER, R., PIN, N.G. AND S. PORTNOY, (1994), Quantile smoothing splines. *Biometrika*, 81, 673-80.

KOENKER, R AND Z. XIAO (2004). Unit root quantile autoregression inference, *Journal of the American Statistical Association,* 94, 775-87

KOENKER, R AND Z. XIAO (2006). *Quantile autoregression* (with discussion), *Journal of the American Statistical Association,* to appear.

KOHN, R, ANSLEY, C F AND C-H WONG (1992), Nonparametric spline regression with autoregressive moving average errors, *Biometrika* 79, 335-46.

KOMUNJER, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, 128, 137-64.

KOOPMAN, S.J., N. SHEPHARD AND J. DOORNIK (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal,* 2**,** 113-66**.**

KUESTER, K., MITTNIK, S. AND PAOLELLA, M.S. (2006). Value-at-risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics* 4: 53-89.

KWIATKOWSKI, D., PHILLIPS, P.C.B., SCHMIDT P., AND Y.SHIN (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 44, 159-78.

LINTON, O. AND Y-J WHANG (2006). The quantilogram: with an application to evaluating directional predictability. *Journal of Econometrics* (to appear).

MORGAN, J.P. (1996). *RiskMetrics.* Technical Document, 4th edition. New York.

NEWEY, W K AND J.L.POWELL (1987). Asymmetric Least Squares Estimation and Testing, *Econometrica*, 55, 819-847.

NYBLOM, J.(1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223-30.

NYBLOM, J., AND A.C.HARVEY (2000). Tests of common stochastic trends, *Econometric Theory,* 16: 176-99.

NYBLOM, J., AND A.C.HARVEY (2001). Testing against smooth stochastic trends, *Journal of Applied Econometrics,* 16, 415-29.

NYBLOM, J. AND T. MAKELAINEN (1983). Comparison of tests for the presence of random walk coefficients in a simple linear model, *Journal of the American Statistical Association,* 78, 856-64.

SHEPHARD, N (2005). *Stochastic Volatility.* (Oxford: Oxford University Press).

STOCK, J.H. AND M.W.WATSON (2005) Has inflation become harder to forecast? Mimeo. www.wws.princeton.edu/mwatson/wp.html

STUART, A. AND J.K. ORD (1987). *Kendall's Advanced Theory of Statistics.* Volume 1. Charles Griffin and Co., London.

TAYLOR, J (2005). Generating volatility forecasts from value at risk estimates. *Management Science*, 51, 712-25.

WHITTLE, P. (1983) *Prediction and Regulation*, 2nd ed. Oxford: Blackwell.

YU, K. AND M.C. JONES (1998). Local linear quantile regression, *Journal of the American Statistical Association*, 93, 228-37.