Time-Varying Quantiles

Giuliano De Rossi and Andrew Harvey
Faculty of Economics, Cambridge University

February 16, 2007

Abstract

A time-varying quantile can be fitted to a sequence of observations by formulating a state space model and iteratively applying a suitably modified signal extraction algorithm. Quantiles estimated in this way provide information on various aspects of a time series, including dispersion, asymmetry and, for financial applications, value at risk. Estimates of the quantiles at the end of the series are the basis for forecasting. As such they offer an alternative to conditional quantile autoregressions and, at the same time, give some insight into their structure and potential drawbacks.

KEYWORDS: CAViaR; Dispersion; quantile regression; signal extraction; state space smoother; value at risk.

JEL Classification: C14, C22

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) appeal 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 (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 a 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

[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)

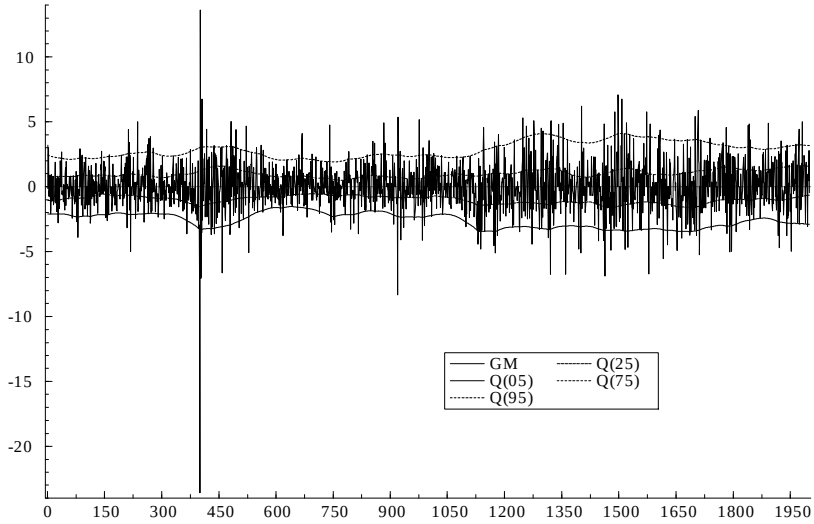

Figure 1: Quantiles fitted to GM returns

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 forecasting. Autoregressive and autoregressive-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

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.

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 present an algorithm for computing the quantiles and investigate how well it works. The algorithm iterates the Kalman filter and associated (fixed-interval) smoother until convergence, taking special care in the way it deals with corner solutions, that is when a quantile passes through an observation. The models for the quantiles usually depend on only one or two parameters. Section 4 suggests that these parameters be estimated by cross-validation and a set of simulation experiments examine the effectiveness of this approach.

Section 5 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[2]. 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 con-

[2]Standard GARCH and SV model specify the complete distribution of the observations. Our method for extracting time-varying quantiles is also based on a model, albeit one that only serves as a device to get a suitable criterion function. However, in certain special cases the same criterion function can be obtained by appealing to cubic spline methodology, which some would regard as nonparametric.

stant volatilities but changing error distributions..' The same is true of our time-varying quantiles for the lower tails.

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 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[3]. The stochastic trend models that we use 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.

2 Quantiles and quantile regression

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

$$S_\tau = \sum_t \rho_\tau(y_t - \xi) = \sum_t (\tau - I(y_t - \xi < 0))(y_t - \xi) \tag{1}$$

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

$$\sum_t IQ(y_t - \xi(\tau)),$$

[3]Specifically it requires the repeated inversion of matrices, the orders of which are up to four times the sample size. Bosch *et al* (1995) only report results for samples of size less than 100 and note that '..large samples could be numerically infeasible..' (p. 626).

where

$$IQ(y_t - \xi_t(\tau)) = \begin{cases} \tau - 1, & if \quad y_t < \xi_t(\tau) \\ \tau, & if \quad y_t > \xi_t(\tau) \end{cases} \tag{2}$$

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

The sample quantile, $\widetilde{\xi}(\tau)$, 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 $\widetilde{\xi}$ between the $T\tau - th$ smallest observation and the one immediately above will make $\sum IQ(y_t - \widetilde{\xi}) = 0$. If $T\tau$ is not an integer, $\widetilde{\xi}$ will coincide with one observation. This observation is the one for which $\sum IQ(y_t - \widetilde{\xi})$ changes sign. These statements need to be modified slightly if several observations take the same value and coincide with $\widetilde{\xi}$. 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 $y_t < \widetilde{\xi}$, is no more than $[T\tau]$ while the number greater is no more than $[T(1-\tau)]$.

In quantile regression, the quantile, $\xi_t(\tau)$, corresponding to the $t - th$ observation is a linear function of explanatory variables, $\mathbf{x}_t$, that is $\xi_t = \mathbf{x}_t'\boldsymbol{\beta}$. 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 ξ_t is a linear combination of past observations.

If the observations are assumed to be independently drawn from an asymmetric double exponential distribution

$$p(y_t|\xi_t) = \tau(1-\tau)\omega^{-1}\exp(-\omega^{-1}\rho_\tau(y_t - \xi_t)), \tag{3}$$

where ω is a scale parameter, maximising the log-likelihood function is equivalent to minimising the criterion function S_τ in (1). Thus the model defines ξ_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 for a sample quantile, when ξ 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 (3) 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.

3 Signal extraction

A model-based framework for estimating time-varying quantiles, $\xi_t(\tau)$, 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

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

where $\Pr(y_t < \xi_t) = \Pr(\varepsilon_t < 0) = \tau$ with $0 < \tau < 1$. The problem is then one of signal extraction with the model for $\xi_t(\tau)$ being treated as a transition equation. By assuming that the serially independent disturbance term, ε_t, has an asymmetric double exponential distribution, as in (3), and is independent of the disturbances driving ξ_t, we end up choosing the estimated quantiles so as to minimize $\sum_t \rho_\tau(y_t - \xi_t)$ 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. However, the theory can be applied to any linear time series model.

3.1 Models for evolving quantiles

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

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

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 $\xi_\tau^\dagger$ is the unconditional mean of $\xi_t(\tau)$. 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

$$\xi_t = \xi_{t-1} + \eta_t, \qquad t = 2, ..., T.$$

The initial value, ξ_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

$$\begin{aligned} \xi_t &= \xi_{t-1} + \beta_{t-1} + \eta_t, \\ \beta_t &= \beta_{t-1} + \zeta_t, \end{aligned} \tag{6}$$

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

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

results in the smoothed estimates being a cubic spline; see, for example, Kohn et al (1992).

3.2 Derivation and properties

If we let the quantile be a first-order autoregression, as in (5), the logarithm of the joint density for the observations in (4) and the quantiles is, ignoring terms independent of quantiles,

$$\begin{aligned} J &= \log p(y_1, .., y_T, \xi_1, .., \xi_T) \\ &= -\frac{1}{2}\frac{(1-\phi^2)(\xi_1 - \xi^\dagger)^2}{\sigma_\eta^2} - \frac{1}{2}\sum_{t=2}^{T}\frac{\eta_t^2}{\sigma_\eta^2} - \sum_{t=1}^{T}\frac{\rho_\tau(y_t - \xi_t)}{\omega}. \end{aligned} \tag{8}$$

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

When ε_t is $NID(0, \sigma^2)$ we may replace ξ_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 $\rho_\tau(y_t - \xi_t)/\omega$ 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; see Durbin and Koopman (2001).

Returning to the quantiles and differentiating with respect to ξ_t gives

$$\frac{\partial J}{\partial \xi_t} = \frac{\phi\xi_{t-1} - (1+\phi^2)\xi_t + \phi\xi_{t+1} + (1-\phi)^2\xi^\dagger}{\sigma_\eta^2} + \frac{1}{\omega}IQ\left(y_t - \xi_t\right), \tag{10}$$

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

$$\frac{\partial J}{\partial \xi_1} = -\frac{(1-\phi^2)(\xi_1 - \xi^\dagger)}{\sigma_\eta^2} + \frac{[\phi(\xi_2 - \phi\xi_1) - \phi(1-\phi)\xi^\dagger]}{\sigma_\eta^2} + \frac{1}{\omega} IQ(y_1 - \xi_1)$$

and

$$\frac{\partial J}{\partial \xi_T} = \frac{-(\xi_T - \phi\xi_{T-1}) + (1-\phi)\xi^\dagger}{\sigma_\eta^2} + \frac{1}{\omega} IQ(y_T - \xi_T)$$

where $IQ(y_t - \xi_t)$ is defined as in (2). For $t = 2, .., T-1$, setting $\partial J/\partial \xi_t$ to zero gives an equation that is satisfied by the estimated quantiles, $\widetilde{\xi}_t, \widetilde{\xi}_{t-1}$ and $\widetilde{\xi}_{t+1}$, and similarly for $t = 1$ and T. If a solution is located on an observation, that is $\widetilde{\xi}_t = y_t$, then $IQ(y_t - \xi_t)$ 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 $-{\xi_1}^2/2\kappa$ 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 $\partial J/\partial \xi_1$ dropping out. It is easy to see that the terms associated with the random walk, that is ignoring the $IQ's$, sum to zero. On the other hand, when ξ_t is stationary, summing the terms in the derivatives not involving the $IQ's$ yields

$$(T-2)(1-\phi)^2\xi^\dagger + 2(1-\phi)\xi^\dagger - (1-\phi)\xi_1 - (1-\phi)^2 \sum_{t=2}^{T-1} \xi_t - (1-\phi)\xi_T$$

divided by σ_η^2. However, setting

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

ensures that the sum is zero at the mode[4].

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[5] of time-varying quantiles, namely that *the number of observations that are less than the corresponding quantile, that is $y_t < \widetilde{\xi}_t$, is no more than $[T\tau]$ while the number greater is no more than $[T(1-\tau)]$*. The proof is in appendix A.

[4]The same result is obtained if $\xi^\dagger$ is treated as random and its estimator is obtained by solving $\partial J/\partial \xi^\dagger = 0$.

[5]A similar property holds for regression quantiles though, as Koenker (2005, p35-7) observes, the number of corner solutions rarely exceeds the number of explanatory variables.

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 corner solutions by defining the corresponding $I\xi_t s$ as the values that give equality of the associated derivative of J. Then we can set $\partial J/\partial \xi_t$ in (10) equal to zero to give

$$\frac{\widetilde{\xi}_t - \xi^\dagger}{g} = \frac{1}{\omega} IQ(y_t - \widetilde{\xi}_t) \tag{13}$$

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

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

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 $(\widetilde{\xi}_t - \xi^\dagger)/\omega$ to both sides of (13) and re-arranging to give

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

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

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

where θ is as defined for (12) but with q_μ replaced by 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. The argument leading to (15) carries over to any stationary linear model for the quantiles, with the ACGF, g, being appropriately modified. This formulation suggests a procedure for computing the $\widetilde{\xi}_t s$ in which synthetic 'observations', $\widehat{\xi}_t - \widehat{\xi}^\dagger + IQ(y_t - \widehat{\xi}_t)$, are constructed using current estimates, $\widehat{\xi}_t$, and inserted in a standard smoothing algorithm which is iterated to convergence.

For the random walk, expression (14) can no longer be obtained. In (16) the weights sum to one, so $\xi^\dagger$ drops out giving

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

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

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

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

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

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

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

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 q plays the same role as the bandwidth. Note also that in the model-based formula, $\widehat{\xi}_{t+j}$ is used instead of $\widehat{\xi}_t$ 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 $\widetilde{\xi}_t, t = 1, .., T$, that set each of the derivatives of J in (10) equal to zero when the corresponding $\widetilde{\xi}_t$ is not set equal to y_t. We first describe a very simple recursive procedure, focussing on the random walk and AR(1), and then set out a general recursion based on the Kalman filter and smoother. The two methods are compared in the last sub-section and a way of combining them is proposed. It is assumed that parameters such as q are known; estimation is considered in the next section.

3.4.1 Simple recursion

Suppose that we knew the value of all the $\widetilde{\xi}_t s$ except one, say at time s. Then the remaining value could be obtained by choosing it so that J is maximized. For a RW or AR(1), the solution would depend on $\widetilde{\xi}_{s-1}$ and $\widetilde{\xi}_{s+1}$. It could be an interior solution or a corner solution ($\widetilde{\xi}_s = y_s$), depending on the values of $\widetilde{\xi}_{s-1}$, $\widetilde{\xi}_{s+1}$ and y_s. The simple algorithm described below solves iteratively all the equations $\partial J/\partial \xi_t = 0, t = 1, ..., T$, starting from an appropriate initial condition and continuing until convergence to the conditional mode.

To obtain the estimated sequence, $\widehat{\xi}_1^{(i)}, \ldots, \widehat{\xi}_T^{(i)}$ from $\widehat{\xi}_1^{(i-1)}, \ldots, \widehat{\xi}_T^{(i-1)}$ we perform the following calculations for $t = 1, \ldots, T$, for the RW. Define $J_t(\xi_t)$ as the derivative $\partial J / \partial \xi_t$, calculated at $\xi_t = \widehat{\xi}_{t-1}^{(i)}$, $\xi_t = \widehat{\xi}_{t+1}^{(i-1)}$. If $J_t(\xi_t)$ is positive for $\xi_t < y_t$ and negative for $\xi_t > y_t$, then the solution is on a corner point, i.e. we set $\widehat{\xi}_t^{(i)} = y_t$. Otherwise, if both the left limit and the right limit of $J_t(\xi_t)$ for $\xi_t \to y_t$ are positive then the solution $\widehat{\xi}_t^{(i)}$ must satisfy $\widehat{\xi}_t^{(i)} > y_t$. Therefore $\widehat{\xi}_t^{(i)}$ is set equal to the value that solves

$$\widehat{\xi}_{t-1}^{(i)} - 2\xi_t + \widehat{\xi}_{t+1}^{(i-1)} + \frac{\sigma_\eta^2}{\omega}(\tau - 1) = 0, \tag{18}$$

which, by (10), ensures that the derivative $\partial J / \partial \xi_t$ is equal to zero. Finally, if both the left limit and the right limit of $J_t(\xi_t)$ for $\xi_t \to y_t$ are negative then the solution $\widehat{\xi}_t^{(i)}$ must satisfy $\widehat{\xi}_t^{(i)} < y_t$. Therefore $\widehat{\xi}_t^{(i)}$ is set equal to the value that solves

$$\widehat{\xi}_{t-1}^{(i)} - 2\xi_t + \widehat{\xi}_{t+1}^{(i-1)} + \frac{\sigma_\eta^2}{\omega}\tau = 0. \tag{19}$$

Equations (18) and (19) are modified in a straightforward way to deal with the endpoints $t = 1$ and $t = T$.

For the AR(1) model the above scheme is modified in an obvious way from (10) and $\widehat{\xi}^\dagger$ is re-calculated at the end of each iteration using (11). The algorithm can be generalised to handle other models using the state space form.

3.4.2 State space smoothing

The state space form (SSF) for a univariate Gaussian time series is:

$$\begin{aligned} y_t &= \mathbf{z}_t' \boldsymbol{\alpha}_t + \varepsilon_t, \qquad \varepsilon_t \sim NID(0, h_t), \qquad t = 1, ..., T \\ \boldsymbol{\alpha}_t &= \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim NID(\mathbf{0}, \mathbf{Q}_t) \end{aligned} \tag{20}$$

where $\boldsymbol{\alpha}_t$ is an $m \times 1$ state vector, $\mathbf{z}_t$ is a non-stochastic $m \times 1$ vector, h_t is a positive scalar, $\mathbf{T}_t$ is an $m \times m$ non-stochastic transition matrix and $\mathbf{Q}_t$ is an $m \times m$ positive definite covariance matrix. The specification is completed by assuming that $\boldsymbol{\alpha}_1 \sim N(\mathbf{a}_{1|0}, \mathbf{P}_{1|0})$ and that the disturbances ε_t and $\boldsymbol{\eta}_t$ are independent of each other and of the initial state.

The joint density of the observations and the states is, ignoring terms not involving the states,

$$J = -\frac{1}{2}\sum_{t=1}^{T}(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t)^2/h_t - \frac{1}{2}\sum_{t=2}^{T}\boldsymbol{\eta}_t'\mathbf{Q}_t^{-1}\boldsymbol{\eta}_t - \frac{1}{2}(\boldsymbol{\alpha}_1-\mathbf{a}_{1|0})'\mathbf{P}_{1|0}^{-1}(\boldsymbol{\alpha}_1-\mathbf{a}_{1|0}). \tag{21}$$

Differentiating J with respect to each element of $\boldsymbol{\alpha}_t$ gives

$$\begin{aligned}
\frac{\partial J}{\partial \boldsymbol{\alpha}_1} &= \mathbf{z}_1 h_1^{-1}(y_1 - \mathbf{z}_1'\boldsymbol{\alpha}_1) - \mathbf{P}_{1|0}^{-1}(\boldsymbol{\alpha}_1-\mathbf{a}_{1|0}) + \mathbf{T}_2'\mathbf{Q}_2^{-1}(\boldsymbol{\alpha}_2-\mathbf{T}_2\boldsymbol{\alpha}_1) \\
\frac{\partial J}{\partial \boldsymbol{\alpha}_t} &= \mathbf{z}_t h_t^{-1}(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t) - \mathbf{Q}_t^{-1}(\boldsymbol{\alpha}_t-\mathbf{T}_t\boldsymbol{\alpha}_{t-1}) + \mathbf{T}_{t+1}'\mathbf{Q}_{t+1}^{-1}(\boldsymbol{\alpha}_{t+1}-\mathbf{T}_{t+1}\boldsymbol{\alpha}_t), \\
&\quad t=2,\ldots,T-1, \\
\frac{\partial J}{\partial \boldsymbol{\alpha}_T} &= \mathbf{z}_T h_T^{-1}(y_T - \mathbf{z}_T'\boldsymbol{\alpha}_T) - \mathbf{Q}_T^{-1}(\boldsymbol{\alpha}_T-\mathbf{T}_T\boldsymbol{\alpha}_{T-1}).
\end{aligned} \tag{22}$$

The smoothed estimates, $\widetilde{\boldsymbol{\alpha}}_t$, satisfy the equations obtained by setting these derivatives equal to zero. They may be computed efficiently by the Kalman filter and associated smoother (KFS) as described in Durbin and Koopman (2001, pp. 70-73). If all the elements in the state are nonstationary and given a diffuse prior, that is $\boldsymbol{\alpha}_1 \sim N(\mathbf{0}, \kappa\mathbf{I})$, the last term in (21) disappears. The treatment of the diffuse prior in the KFS is not trivial but methods exist for dealing with it; see de Jong (1989) and Durbin and Koopman (2001). An algorithm is available as a subroutine in the SsfPack set of programs within Ox; see Koopman et al (1999).

In a time invariant model the subscripts on $\mathbf{z}_t, h_t, \mathbf{T}_t$ and $\mathbf{Q}_t$ can be dropped. Assuming, for simplicity, that this is the case, let us set $\mathbf{z}'\boldsymbol{\alpha}_t = \xi_t(\tau)$ and assume that ε_t has an asymmetric double exponential distribution, (3). Then

$$J = -\frac{1}{\omega}\sum_{t=1}^{T}\rho_\tau(y_t - \mathbf{z}'\boldsymbol{\alpha}_t) - \frac{1}{2}\sum_{t=2}^{T}\boldsymbol{\eta}_t'\mathbf{Q}^{-1}\boldsymbol{\eta}_t - \frac{1}{2}(\boldsymbol{\alpha}_1-\mathbf{a}_{1|0})'\mathbf{P}_{1|0}^{-1}(\boldsymbol{\alpha}_1-\mathbf{a}_{1|0}) \tag{23}$$

and the first term in each of the three equations of (22), that is $\mathbf{z}_t h_t^{-1}(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t)$, is replaced by $\mathbf{z}\omega^{-1}IQ(y_t - \mathbf{z}'\boldsymbol{\alpha}_t)$. Adding and subtracting $\mathbf{z}\omega^{-1}\mathbf{z}'\boldsymbol{\alpha}_t$ yields

$$\mathbf{z}\omega^{-1}[\mathbf{z}'\boldsymbol{\alpha}_t + IQ(y_t - \mathbf{z}'\boldsymbol{\alpha}_t)] - \mathbf{z}\omega^{-1}\mathbf{z}'\boldsymbol{\alpha}_t, \qquad t = 1,\ldots,T. \tag{24}$$

This suggests that we set up an iterative procedure in which the estimate of the state at the j-th iteration, $\widetilde{\boldsymbol{\alpha}}_t^{(j)}$, is computed from the KFS applied to a set of synthetic observations constructed as

$$\widehat{y}_t^{(j-1)} = \mathbf{z}'\widehat{\boldsymbol{\alpha}}_t^{(j-1)} + IQ(y_t - \mathbf{z}'\widehat{\boldsymbol{\alpha}}_t^{(j-1)}). \tag{25}$$

If, at a particular iteration, the solution to the $t-th$ equation in (22) is on a corner, h_t is set equal to zero rather than ω and the original observation y_t, rather than the transformed one, is fed into filtering and smoothing algorithm. This forces the KFS to set the estimated quantile equal to the corresponding observation. The iterations are carried out until the $\widehat{\boldsymbol{\alpha}}_t^{(j)\prime}s$ converge.

The SSF for a random walk quantile is trivial, while for the local linear trend, (6), $\boldsymbol{\alpha}_t = (\xi_t, \beta_t)'$ and $\mathbf{z}' = (1\ 0)$. For an AR(1), the stationary zero mean component, $\xi_t - \xi_t^{\dagger}$, is initiated with its unconditional distribution as in (8). In a Gaussian model, the mean, $\xi^{\dagger}$, would normally be included in the state as well by treating it as stochastic but time invariant (ie $\xi_t = \xi_{t-1}$) with a diffuse prior. However, since the smoother is being repeated, an alternative strategy is to update it using (11) at each iteration. Casting more general models in SSF is straightforward; see, for example, Harvey (1989, ch2).

For a model in SSF with a diffuse prior placed on the initial state, the fundamental property of dynamic quantiles, stated in sub-section 3.2, is satisfied if summing the derivatives with respect to the second term in (21) yields a term $\sum_{t=1}^{T}\mathbf{A}_t\boldsymbol{\alpha}_t$, where $\mathbf{A}_t, t = 1, .., T$ is an $m \times m$ matrix, the first row of which contains only zeroes. A sufficient condition is that the first column of $\mathbf{T} - \mathbf{I}$ consists of zeroes. This condition is satisfied by the local linear trend; note that the covariance matrix in (7) is positive definite.

3.4.3 Combined algorithm

Figure **??** shows a plot of the convergence criterion

$$C^{(j)} = \sum_t \left(\widehat{\xi}_t^{(j)} - \widehat{\xi}_{t-1}^{(j-1)}\right)^2 / \sum_t \left(\widehat{\xi}_t^{(j-1)}\right)^2, \qquad j = 1, 2, ...$$

for the simple and state space algorithms, obtained during the estimation of the 5% quantile of the returns of General Motors, as shown in figure 1. Typically, the KFS algorithm is faster while the simple one makes relatively small changes at each step. The jumps for the KFS algorithm occur when

there is a change in the points designated as corner solutions, but there is no corresponding increase in the objective function J. As can be seen from figure **??** the simple algorithm needs many more iterations to reach the same level of J as the KFS one. Results for other models and data sets show a similar pattern.

In the empirical sections of the paper we have combined the two algorithms by running the KFS first and then using the simple algorithm to check or fine tune the estimates. The procedure stops[6] when $C^{(j)}$ in the simple algorithm falls below a pre-assigned tolerance level.

Initial estimates for series with no trend, such as returns, are given by the fixed quantiles. More generally we might compute the median and then add the relevant fixed sample quantile estimated from the residuals.

Note that when q is zero, that is the quantile is constant, there is usually only one corner, while for a value of infinity, obtained when $\omega = 0$, there are T corners because all quantiles pass through all observations.

4 Parameter estimation

It is apparent from the previous discussion that random walk quantiles depend only on q. For an AR(1) there is an additional parameter, ϕ. The scale parameter ω is of no interest since the model is not intended to be a description of a data generating process. For the same reason, ML estimation is not an option. We therefore consider cross validation (CV). A set of Monte Carlo experiments provide evidence on how successful CV is at yielding estimated quantiles that closely match the true ones.

4.1 Cross -validation

ML estimation of unknown parameters is easily carried out for Gaussian UC models of the form (9) using the prediction errors from the Kalman filter.

[6] Alternatively, one can use as a convergence criterion the norm of the gradient of J (more precisely, of the subvector obtained by eliminating the components that correspond to corner solutions) or the largest derivative of J in absolute value, scaled by the value of J. An advantage of such criteria is that, unlike $C^{(j)}$, their values can be compared for the KFS and simple algorithms. However, we found no particular advantage in adopting these alternatives as convergence criteria.

The cross validation criterion is

$$CV = \sum_{t=1}^{T} (y_t - \widetilde{\mu}_t^{(-t)})^2,$$

where $\widetilde{\mu}_t^{(-t)}$ is the smoothed estimator of μ_t when y_t is dropped, but in the Gaussian case it can be computed from a single pass of the KFS; see de Jong (1988). Kohn, Ansley and Wong (1992) compare ML and CV estimators for models of this kind and conclude, on the basis of Monte Carlo experiments that, even though ML tends to perform better, CV represents a viable alternative.

In the present context, the appropriate cross validation function is

$$CV(\tau) = \sum_{t=1}^{T} \rho_\tau(y_t - \widetilde{\xi}_t^{(-t)}) \tag{26}$$

where $\widetilde{\xi}_t^{(-t)}$ is the smoothed value at time t when y_t is dropped. Unfortunately, there appears to be no simple way of computing this from the KFS except by a 'brute force' approach in which all T observations are dropped one at a time. Our experience indicates that the best way to proceed is by iterating the KFS to compute initial estimates of the quantiles, as in sub-section 3.4.3, and then using the simple algorithm to calculate the CV criterion by dropping observations one at a time. The number of iterations of the simple algorithm for each t is usually not large as the starting values of the quantiles obtained from dropping an adjacent observation tend to be close to the solution.[7]This process is repeated for different values of the parameters and the minimum of the CV function found by a search procedure. In order to guard against local minima we use a grid search (with respect to $q^{1/2}$). Some notion of a sensible range for $q^{1/2}$ can be obtained by using some of the methods outlined in appendix B.

Figure 1 shows random walk quantiles for the first 2000 observations of GM. The estimates of $q^{1/2}$ obtained by CV are 0.09, 0.06, 0.01, 0.06 and 0.08

[7]Computational savings might be made by only using a limited number of observations around the one that is dropped. In other words, one can choose an integer δ such that the difference between the estimate of $\widetilde{\xi}_t^{(-t)}$ computed using the full set of observations and the one computed using only y_{t+j}, $j = -\delta, \ldots, -2, -1, 1, 2, \ldots, \delta$ is negligible. Such a procedure requires a number of operations that is $O(T)$ for each signal-noise ratio rather than $O(T^2)$. The cut-off, δ, needs to increase as the signal-noise ratio goes down.

for the 5%, 25 %, 50%, 75% and 95% respectively. Generalized stationarity tests, as described in Busetti and Harvey (2006), reject the null hypothesis of time invariance at the 1% level of significance. We also estimated the 1% quantile. In this case, $q^{1/2}$ was estimated as 0.04 and a plot of the quantile shows relatively little variation. This may be a reflection of the difficulty of estimating a 1% quantile. Indeed the asymptotic standard error associated with the estimator of a fixed 1% quantile is about three times that of the 5% quantile for the normal distribution and five times for Laplace; the formula is given in Koenker (2005, p 71-2).

4.2 Choice of model for different quantiles

For modeling returns, the main concern is to have slowly changing quantiles and so persistent autoregressions, RW and local linear trend specifications are appropriate. We could, in principle, have different models for different quantiles. However, there is a strong case for having the same model (albeit with different parameters). The RW and local linear trend models are additive and so if two quantiles follow one of these models then so does their difference. A constant difference is obtained limiting case, so a set-up with the same model for different quantiles is able to capture a time invariant distribution around a non-stationary time-varying median.

How might the $q's$ vary with τ? The Monte Carlo experiments reported below are designed to throw some light on this question. First we give an example. Figure **??** shows 288 monthly figures on US inflation[8] from 1981(7) to 2005(6). Although fitting a random walk plus noise with STAMP gives some residual serial correlation, the model is a reasonable one. The figure shows the smoothed estimates of the level together with estimates of the median and other quantiles, also modelled as random walks. The CV estimates of $q^{1/2}$ for the 10%, 25%, 50%, 75% and 90% quantiles are 0.004, 0.008, 0.004, 0.006 and 0.008 respectively.

[8] To be precise, the first difference of the logarithm of the personal consumer expenditure deflator (all) as given by Stock and Watson (2005).

4.3 Monte Carlo experiments

Firstly, we generated 200 realizations of size 100 from a random walk plus Laplace noise model, that is

$$y_t = \xi_t + \varepsilon_t, \qquad \xi_t = \xi_{t-1} + \eta_t, \qquad t = 1, ..., 100$$

where $\xi_t = \xi_t(0.5)$ and ε_t has a Laplace (double exponential) distribution with parameter $\omega = 0.5$, while η_t is normally distributed with variance σ_η^2. Pseudo random numbers were generated by using the Modified Park and Miller generator as implemented in Ox. The cross validation procedure was used to estimate the median of each of the simulated samples. The results are presented in Table 1 for a range of values of $q^{1/2}$ as given by $\sigma_\eta/\sqrt{\omega}$. In all cases the median of the estimates is slightly lower than the true value. However, since the ultimate goal is to extract the median from an observed time series, we compare the CV estimator with infeasible estimators that require knowledge of the true quantiles (columns 6 and 7). For each simulated sample and each point on the grid of parameter values we computed the MSE and mean absolute deviation (MAD) associated with the estimated median, that is

$$MSE(\sigma_\eta) = T^{-1}\sum_{t=1}^{T}(\widehat{\xi}_t - \xi_t)^2, \qquad \text{and} \qquad MAD(\sigma_\eta) = T^{-1}\sum_{t=1}^{T}|\widehat{\xi}_t - \xi_t|.$$

The values of $q^{1/2}$ displayed in columns 5 and 6 are those that minimize the average of, respectively, $MSE(\sigma_\eta)$ and $MAD(\sigma_\eta)$ across all simulated samples. Hence they are in some sense optimal. We see that cross validation, even with a small sample size, results in an estimator that is very close to these infeasible estimators.

Table 2 illustrates a Monte Carlo experiment with the same design as in table 1 except that the sample size increased from 100 to 500. It is reassuring that the median of the CV estimates is even closer to the medians of the infeasible estimators (again displayed in columns 5 and 6) and that the distribution becomes more concentrated.

Next we examine what happens when CV is used to extract quantiles other than the median from these realizations. Table 3 shows that, for the 25% quantile, CV selects a value of q that is smaller than the one that minimizes the mean MSE or MAD across all realizations. However, the values of MSE and MAD are not very sensitive to q around the minimum, so that

in practice the indicators calculated at the median of the CV estimator are less than 10% larger than the ones computed at the optimum. Given that the optimum is obtained by computing an infeasible estimator, we regard this result as an indication that the CV estimator performs well. Similar conclusions can be drawn about the 10% quantile; see table 4. It is worth noting that, despite the fact that the optimal signal noise ratios - as judged by the infeasible estimates - are virtually the same for all quantiles of the distribution, our CV procedure seems to select lower ratios as we move from the median towards the tail.

We next consider an experiment where ε_t is Gaussian. Its variance is set to two so as to be the same as the variance of the Laplace measurement errors used in the previous experiments. Hence the values of q used in the previous experiments satisfy (36). The estimation results, shown in table 5, are comparable with those in table 3. The CV estimator performs even better when the distribution of ε_t is (scaled) Student t with three degrees of freedom, as table 6 indicates. These results show that our findings are not restricted to Laplace disturbances.

5 Dispersion and Asymmetry

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.

5.1 Dispersion

The contrasts between complementary quantiles, that is

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

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 nor-

mal model, scale provides a more natural measure of dispersion than variance.....and offers substantial advantages from the robustness[9] viewpoint'.

Figure **??** shows the interquartile range for the US inflation series of figure **??**. It is worth stressing that this is not dependent on any measure of location.

If the distribution is assumed to be symmetric around zero, estimates of $\xi_t(\tau)$ and $\xi_t(1-\tau)$ can be obtained by assuming that one is the negative of the other. This can be done simply by estimating the $(1-2\tau)th$ quantile for $|y_t|$. Thus to compute $\widetilde{\xi}_t(0.75) = -\widetilde{\xi}_t(0.25)$ we estimate the median for $|y_t|$. Doubling this series gives the interquartile range.

The outer quantiles can highlight interesting behaviour in the tails. As an illustration, 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 over time, with the tails becoming heavier. However, the 1% and 5% quantiles head off in different directions. Figure **??** shows the absolute values of the series and random walk 98%, 90% and 50% quantiles fitted by CV. As can be seen, the estimated quantiles track the true quantiles, corresponding to 1%, 5% and 25% in the original series, quite well.

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

$$\frac{\widetilde{D}(0.05)}{\widetilde{D}(0.25)} = \frac{\widetilde{\xi}_t(0.95) - \widetilde{\xi}_t(0.05)}{\widetilde{\xi}_t(0.75) - \widetilde{\xi}_t(0.25)}. \tag{27}$$

For a normal distribution this ratio is 2.44, for t_3 it is 3.08 and for a Cauchy 6.31. Figure **??** shows the 5% and 25 % quantiles and the plot of (27) for the GM series.

[9]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.

5.2 Asymmetry

For a symmetric distribution

$$S(\tau) = \xi_t(\tau) + \xi_t(1-\tau) - 2\xi_t(0.5), \qquad \tau < 0.5 \tag{28}$$

is zero for all $t = 1, .., T$. Hence a plot of this contrast shows how the asymmetry captured by the complementary quantiles, $\xi_t(\tau)$ and $\xi_t(1-\tau)$, changes over time. Evidence that changing skewness may be a feature of some financial time series can be found in Harvey and Siddique (1999).

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 at the end of the sample is the filtered estimate. For a model in SSF, predictions of the quantile can be made as $\widetilde{\xi}_{T+j|T} = \mathbf{z}'\mathbf{T}^j\widetilde{\boldsymbol{\alpha}}_T$, $j = 1, 2, ...$ 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

$$\widetilde{\xi}_{t|t} = (1-\theta)\sum_{j=0}^{\infty}\theta^j[\widetilde{\xi}_{t-j|t} + IQ(y_{t-j} - \widetilde{\xi}_{t-j|t})] \tag{29}$$

where $\widetilde{\xi}_{t-j|t}$ is the smoothed estimator of ξ_t based on information at time t; see, for example, Whittle (1983, p69). Thus $\widetilde{\xi}_{t|t}$ is an exponentially weighted

moving average of the synthetic observations, $\widetilde{\xi}_{t-j|t} + IQ(y_{t-j} - \widetilde{\xi}_{t-j|t})$. The predictions at time T are simply equal to $\widetilde{\xi}_{T|T}$ for all lead times.

6.2 Specification and diagnostic testing

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

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

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

$$L(\tau) = \frac{\sum_{t=T+1}^{T+L} IQ(y_t - \widetilde{\xi}_{t|t-1})}{\sqrt{L\tau(1-\tau)}} \tag{30}$$

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 **??** 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, $L(0.25)$, is -1.43.

6.3 Quantile autoregression

In quantile autoregression (QAR), the (conditional) quantile is assumed to be a linear combination of past observations; see, for example, Komunjer (2005), Koenker (2005, p.126-8, 260-5) and the references therein. Koenker and Xiao (2006) allow for different coefficients for each quantile, so that

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

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. 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 be adapted by letting the quantile in (4) be

$$\xi_t(\tau) = \alpha_{0,\tau} + \alpha_{1,\tau}\xi_{t-1}(\tau) + ... + \alpha_{p,\tau}\xi_{t-p}(\tau) + \eta_t(\tau),$$

where $\eta_t(\tau)$, like $\varepsilon_t(\tau)$, has an asymmetric double exponential distribution. 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

$$\xi_t(\tau) = \alpha_0 + \sum_{i=1}^{q} \beta_i \xi_{t-i}(\tau) + \sum_{j=1}^{r} \alpha_j f(y_{t-j}). \tag{31}$$

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

$$\xi_t(\tau) = \alpha_0 + \beta\xi_{t-1}(\tau) + \gamma\,|y_{t-1}| \tag{32}$$

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

$$\xi_t(\tau) = \xi_{t-1}(\tau) + \gamma\{[1 + \exp(G[y_{t-1} - \xi_{t-1}(\tau)])]^{-1} - \tau\}, \tag{33}$$

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

$$\xi_t(\tau) = (\alpha_0 + \beta\xi_{t-1}^2(\tau) + \alpha_1 y_{t-1}^2)^{1/2}. \tag{34}$$

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 checking 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, $\xi_t(\tau)$ in (31) is not the actual quantile so a change in notation to $\widehat{\xi}_{t|t-1}(\tau)$ is helpful. Similarly the lagged values, which are approximations to smoothed estimators calculated as though they were filtered estimators, are best written as $\widehat{\xi}_{t-j|t-1-j}(\tau),\ j = 1, ..., q$. The idea is then to compute the $\widehat{\xi}_{t|t-1}(\tau)'s$ - 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 (29) is

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

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

$$\widehat{\xi}_{t|t-1} = \widehat{\xi}_{t-1|t-2} + (1-\theta)\widehat{\nu}_{t-1}, \tag{35}$$

where

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

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 $\widehat{\xi}'_{t|t-1}s$ set equal to the fixed quantile computed from a small number of observations 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

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

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

The recursion in (35) has the same form as the limiting case ($G \to \infty$) of the adaptive CAViaR model, (33). 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 (32) and (34), 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 (31) 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 (34) 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 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, though there is clearly scope for improving the 'modified brute force' cross-validation procedure. There is also a need for investigating how, for example, bootstrap procedures might be able to provide a viable method of assessing precision and how the methods might be extended to jointly estimating several quantiles. Including explanatory variables, for example to pick up breaks, is quite straightforward in principle but needs to be investigated in practice.

The cross-validation criterion appears to work well. Monte Carlo experiments carried out for the Laplace distribution show that the CV estimates are close to the true values. There is a slight tendency to underestimate but

this appears to have no adverse implications for the accuracy with which the true quantiles are estimated.

The estimation of time-varying quantiles opens up a wealth of possibilities for capturing the evolving distribution of a time series. In particular the lower quantiles can be used to directly estimate the evolution of value at risk, VaR, while changes in dispersion and asymmetry can be highlighted by suitably constructed contrasts.

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.

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. Any errors are our responsibility. We would like to thank the ESRC for financial support under the grant Time-Varying Quantiles, RES-062-23-0129.*

Appendix A: Fundamental property of time-varying quantiles

Suppose that there is a single corner solution at $t = s$ so $\widetilde{\xi}_s = y_s$. All the derivatives of J can be set to zero apart from the one at s. Call G the sum of the terms in J that do not depend on the function $\rho_\tau(.)$. Then from (10)

$$\sum_{t \neq s} \frac{\partial G}{\partial \xi_t} = \sum_{t \neq s} IQ_t$$

and if $\sum_{t=1}^{T} \partial G / \partial \xi_t = 0$, it follows that

$$\sum_{t \neq s} IQ_t + \frac{\partial G}{\partial \xi_s} = 0$$

Now a small increase in $\widetilde{\xi}_s$ gives $IQ(y_s - \widetilde{\xi}_s)$ a value of $\tau - 1$, while a small decrease makes it equal to τ, so

$$-(1-\tau) \leq \frac{\partial G}{\partial \xi_s} \leq \tau$$

Hence

$$-\tau \leq \sum_{t \neq s} IQ(y_t - \widetilde{\xi}_t) \leq 1 - \tau.$$

When there are k corners a similar argument leads to

$$-k\tau \leq \sum_{t \notin C} IQ(y_t - \widetilde{\xi}_t) \leq k\,(1-\tau)$$

where C is the set of all points where the quantile passes through an observation. Now suppose that $\underline{n}$ denotes the number of observations (strictly) below the corresponding quantile while $\overline{n} = (T - \underline{n} - k)$ is the number (strictly) above. Then, abbreviating $IQ(y_t - \widetilde{\xi}_t)$ to IQ_t,

$$\sum_{t \notin C} IQ_t = \underline{n}\,(\tau - 1) + (T - \underline{n} - k)\,\tau = T\tau - \underline{n} - k\tau$$

Now $\sum_{t \notin C} IQ_t \geq -k\tau$ implies $\underline{n} \leq [\tau T]$ because $\sum_{t \notin C} IQ_t$ would be less than $-k\tau$ if $\underline{n}$ were greater than $[\tau T]$. Similarly, $\sum_{t \notin C} IQ_t \leq k(1-\tau)$ implies $\overline{n} \leq [(1-\tau)T]$ because $\sum_{t \notin C} IQ_t = \overline{n} - (1-\tau)\,T + k\,(1-\tau)$ would be greater than $k(1-\tau)$ if $\overline{n}$ were to exceed $[(1-\tau)T]$.

Appendix B: Initial values and scaling

The quasi signal-noise ratio, q, is expressed relative to ω which is not a variance but is treated as such when the KFS is run. As was noted in sub-section 3.3, q is not scale invariant. This creates difficulties for a search procedure because the user has no idea of what constitutes a plausible range of values and numerical instabilities may result by mixing estimated quantiles with scale invariant indicators in the synthetic observations. On the other hand the signal-noise ratio, $q_\mu = \sigma_\eta^2/\sigma_\varepsilon^2$, upon which the MMSLE of the time-varying mean, μ_t, in an AR(1) or random walk model depends, is scale invariant and easily estimated. For a symmetric noise distribution, the mean is the same as the median and so comparability with the calculations for the

mean could provide the basis for a strategy for estimating a series with a time-varying median.

If σ_η^2 and σ_ε^2 are given, the estimator of the median given by the mode of J satisfies (16) with

$$q = \sigma_\eta^2/\omega = \sqrt{8}\sigma_\varepsilon q_\mu \tag{36}$$

since it is contructed under the assumption that the noise has a double exponential (Laplace) distribution for which $\sigma_\varepsilon^2 = 8\omega^2$. If q, q_μ and σ_ε^2 have all been estimated, (36) should be approximately true. In the US inflation example of the previous sub-section $\widetilde{q}_\mu = 0.011$ and $\widetilde{\sigma}_\varepsilon = 0.0015$. Expression (36) gives $\widetilde{q}^{1/2} \simeq 0.007$ which is of a similar order of magnitude to the estimate of 0.004 obtained by CV.

The weighting patterns in the expressions for the mean and median, that is (12) and (16), decay at different rates as θ depends on q in one case and q_μ in the other. However, if we add $(\widetilde{\xi}_t - \xi^\dagger)/\sigma_\varepsilon^2$ rather than $(\widetilde{\xi}_t - \xi^\dagger)/\omega$ to both sides of (13) then q is redefined so as to be equal to q_μ. Furthermore the synthetic observation in (16) becomes

$$\widetilde{\xi}_{t+j} - \xi^\dagger + \sqrt{8}\sigma_\varepsilon . IQ(y_{t+j} - \widetilde{\xi}_{t+j})$$

so the part involving the indicator changes in proportion to the quantile when there is a change in scale. The attraction of proceeding in this way is the comparability with the calculations for the MMSLE and the fact that key quantities no longer depend on scale. The disadvantage is that it requires an estimate of σ_ε and making such an estimate an integral part of the calculations for a median is not appropriate. An alternative approach is based on noting that the same effect could be obtained by dividing the observations by $\sqrt{8}\sigma_\varepsilon$. This suggests the following strategy for estimating any quantile, not just the median.

(i) Estimate the model assuming Gaussianity to give $\widetilde{q}_\mu$ and $\widetilde{\sigma}_\varepsilon$; if heavy-tails are suspected trim the residuals, add to the trend to creat new observations and re-estimate the parameters.

(ii) Divide the observations by $\sqrt{8}\widetilde{\sigma}_\varepsilon$.

(iii) Estimate q using $\widetilde{q}_\mu$ as a starting value or as a guide to a suitable range.

(iv) Rescale q and the (original) observations (by multiplying by$\sqrt{8}\widetilde{\sigma}_\varepsilon$) for estimating the quantile.

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.

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 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.

Durbin, J., and S.J. Koopman (2000): Time series analysis of non Gaussian observations based on state space models from both classical and Bayesian perspectives, *Journal of the Royal Statistical Society B*, 62, 3–56.

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

Engle, R.F., and Manganelli, S. (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*, pp 327-412, 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 S.J Koopman (2000). Signal extraction and the formulation of unobserved components models, *Econometrics Journal,* 3**,** 84-107.

Harvey, C., and A. Siddique (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis,* 34, 465-87.

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

Koenker, R., Pin, N.G. and S. Portnoy, (1994), Quantile smoothing splines. *Biometrika*, 81, 673-80.

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

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.

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.

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.

		Cross validation			Infeasible	
q_μ	$q^{1/2}$	25%	Median	75%	MSE	MAD
.005	0.14	0.07	0.12	0.20	0.13	0.13
.125	0.71	0.40	0.55	0.90	0.60	0.65
.5	1.41	0.85	1.25	1.65	1.20	1.25

Table 1: Empirical distribution of the cross validation estimator of the square root of the quasi signal - noise ratio for the median. $T = 100$. The data generating process is Gaussian random walk plus Laplace noise with $\omega = 0.5$. Columns 5 and 6 contain the median of the empirical distribution of two infeasible estimators.

	Cross validation			Infeasible	
$q^{1/2}$	25%	Median	75%	min MSE	min MAD
0.14	0.11	0.13	0.19	0.13	0.13
0.71	0.50	0.65	0.75	0.65	0.65
1.41	1.00	1.20	1.40	1.30	1.30

Table 2: Empirical distribution of the cross validation estimator of the square root of the quasi signal - noise ratio for the median. The data generating process is as in table 1, but with $T = 500$.

	Cross validation			Infeasible	
$q^{1/2}$	25%	Median	75%	min MSE	min MAD
0.14	0.09	0.12	0.18	0.12	0.11
0.71	0.40	0.60	0.65	0.65	0.65
1.41	0.90	1.00	1.20	1.30	1.30

Table 3: Empirical distribution of the cross validation estimator of the square root of the quasi signal - noise ratio for the first quartile. The data generating process is as in table 1 with $T = 500$

	Cross validation				
$q^{1/2}$	25%	Median	75%	min MSE	min MAD
0.14	0.07	0.10	0.15	0.11	0.11
0.71	0.40	0.50	0.60	0.60	0.60
1.41	0.70	0.80	0.95	1.30	1.30

Table 4: Empirical distribution of the cross validation estimator of the pseudo signal - noise ratio for the 10% quantile. The data generating process is as in table 1 with $T = 500$

	Cross validation				
$\sigma_\eta/\sqrt{\omega}$	25%	Median	75%	min MSE	min MAD
0.14	0.09	0.11	0.15	0.10	0.10
0.71	0.40	0.50	0.60	0.55	0.55
1.41	0.90	1.00	1.20	1.20	1.20

Table 5: Empirical distribution of the cross validation estimator of the square root of the quasi signal - noise ratio for the first quartile. The data generating process is Gaussian random walk plus Gaussian noise with $\sigma_\varepsilon^2 = 2$ and $T = 500$.

	Cross validation			Infeasible	
$q^{1/2}$	25%	Median	75%	min MSE	min MAD
0.14	0.10	0.14	0.19	0.13	0.13
0.71	0.55	0.70	0.80	0.70	0.70
1.41	1.20	1.35	1.60	1.35	1.35

Table 6: Empirical distribution of the cross validation estimator of the square root of the quasi signal - noise ratio for the first quartile. The data generating process is Gaussian random walk plus scaled Student t noise with three degrees of freedom and variance equal to 2, $T = 500$.