CENTRE FOR ECONOMETRIC ANALYSIS
CEA@Cass

http://www.cass.city.ac.uk/cea/index.html

Cass Business School
Faculty of Finance
106 Bunhill Row
London EC1Y 8TZ

Quantiles, Expectiles and Splines

Giuliano De Rossi and Andrew Harvey

CEA@Cass Working Paper Series

WP–CEA–04-2007

Quantiles, Expectiles and Splines

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

February 19, 2007

Abstract

A time-varying quantile can be fitted to a sequence of observations by formulating a time series model for the corresponding population quantile and iteratively applying a suitably modified state space signal extraction algorithm. It is shown that such time-varying quantiles satisfy the defining property of fixed quantiles in having the appropriate number of observations above and below. Expectiles are similar to quantiles except that they are defined by tail expectations. Like quantiles, time-varying expectiles can be estimated by a state space signal extraction algorithm and they satisfy properties that generalize the moment conditions associated with fixed expectiles. Time-varying quantiles and expectiles provide information on various aspects of a time series, such as dispersion and asymmetry, while estimates at the end of the series provide the basis for forecasting. Because the state space form can handle irregularly spaced observations, the proposed algorithms can be easily adapted to provide a viable means of computing spline-based non-parametric quantile and expectile regressions.

KEYWORDS: Asymmetric least squares; cubic splines; dispersion; non-parametric regression; quantile regression; signal extraction; state space smoother.

JEL Classification: C14, C22

*Corresponding author: ach34@econ.cam.ac.uk

1 Introduction

The movements in a time series may be described by time-varying quantiles. These may be estimated non-parametrically by fitting a simple moving average or a more elaborate kernel. An alternative approach is to formulate a partial model, the role of which is to focus attention on some particular feature - here a quantile - so as to provide a (usually nonlinear) weighting of the observations that will extract that feature by taking account of the dynamic properties of the series. The model is not intended to be taken as a full description of the distribution of the observations. Indeed models for different features, for example different quantiles, may not be consistent with each other.

In an earlier paper, we showed how time-varying quantiles could be fitted to a sequence of observations by setting up a state space model and iteratively applying a suitably modified signal extraction algorithm; see De Rossi and Harvey (2006). Here we determine the conditions under which a linear time series model for the quantile will satisfy the defining property of fixed quantiles in having the appropriate number of observations above and below.

Expectiles are similar to quantiles except that they are defined by tail expectations. Newey and Powell (1987) discuss the theory underlying expectiles and show how they can be applied in a regression context using asymmetric least squares. Here we show how time-varying expectiles can be estimated by a state space signal extraction algorithm. This is similar to the algorithm used for quantiles, but, because the criterion function is everywhere differentiable, estimation is more straightforward and much quicker. We then show that the conditions needed for a time-varying expectile to generalize the moment conditions associated with fixed expectiles are similar to those needed for a time-varying quantile to satisfy the defining property of fixed quantiles.

With Gaussian observations the mean is more efficient than the median and this remains true when they are time-varying. More generally expectiles are likely to be more efficient than quantiles for many distributions, but this then raises the question of the interpretation of expectiles. What exactly do they mean and are they of any practical value? A fixed expectile will correspond to a particular fixed quantile but this is not necessarily the case when they are time-varying. Issues of interpretation aside, there is also the question of robustness: expectiles are more sensitive to outliers. Breckling and Chambers (1988) try to combine robustness and efficiency by adapting Huber's M-estimation method to produce what they call $M-$quantiles. This idea can also be extended to a dynamic setting.

Section 2 reviews the ideas underlying fixed quantiles and expectiles. Section 3 then describes the signal extraction algorithms for estimating them when they are time-varying and establishes some basic properties. The use of cross-validation to estimate key parameters is considered in section 4. Section 5 illustrates how time-varying quantiles and expectiles provide information on various aspects of a time series, such as dispersion, asymmetry and kurtosis. Section 6 notes that estimates at the end of the series are the basis for forecasting. Just as time-varying quantiles provide a different approach to the one adopted by Engle and Manganelli (2004) for assessing value at risk, so time-varying expectiles offer an alternative to the methods proposed by Granger and Sin (2000).

The final part of the paper is concerned with non-parametric estimation of regression models using splines. It has long been known that cubic splines can be fitted by signal extraction procedures because the state space form can handle irregularly spaced observations from a continuous time model; see, for example, Wahba (1978) and Kohn, Ansley and Wong (1992). The proposed algorithms for time-varying quantiles and expectiles are easily adapted so as to provide a viable means of computing spline-based non-parametric quantile and expectile regressions. As well as illustrating the technique, we give a general proof of the equivalence between splines and the continuous time models underlying our signal extraction procedures for quantiles, expectiles and $M-$quantiles.

2 Quantiles and expectiles

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, $y_t, t = 1,..,T,$ (which may be from a cross-section or a time series), 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

$$S_\tau = \sum_{t=1}^{T} \rho_\tau(y_t - \xi) = \sum_{y_t<\xi} (\tau - 1)(y_t - \xi) + \sum_{y_t \geq \xi} \tau(y_t - \xi) \qquad (1)$$

with respect to ξ, where $\rho_\tau(.)$ is the *check function,* defined for quantiles as

$$\rho_\tau(y_t - \xi) = (\tau - I(y_t - \xi < 0))\,(y_t - \xi) \qquad (2)$$

and $I(.)$ is the indicator function. Differentiating (minus) S_τ at all points where this is possible gives

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

where

$$IQ(y_t - \xi_t(\tau)) = \begin{cases} \tau - 1, & if \quad y_t < \xi_t(\tau) \\ \tau, & if \quad y_t > \xi_t(\tau) \end{cases} \tag{3}$$

defines the *quantile indicator function*[1] 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.

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).

If it is assumed that the observations are independently drawn from an asymmetric double exponential distribution,

$$p(y_t|\xi_t) = \tau(1-\tau)\lambda^{-1}\exp(-\lambda^{-1}\rho_\tau(y_t - \xi_t)), \tag{4}$$

where λ is a scale parameter, maximising the log-likelihood function is equivalent to minimising the criterion function S_τ in (1). Thus the model, (4), 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

[1]Note that when viewed as a function of x, $IQ(x)$, the derivative of $\rho_\tau(x)$, is known as the influence function.

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 (4) 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 estimator.

Just as the idea of the median may be extended to quantiles, so the concept of the mean may be extended to *expectiles*. The population expectiles, $\mu(\omega)$, are similar to quantiles but they are determined by tail expectations rather than tail probabilities. For a given value of ω the sample expectile, $\widetilde{\mu}(\omega)$, is obtained by minimizing a function of the form S_ω with

$$\rho_\omega(y_t - \mu(\omega)) = |\omega - I(y_t - \mu(\omega) < 0)| \, (y_t - \mu(\omega))^2, \qquad 0 < \omega < 1. \quad (5)$$

This criterion function is obtained if the observations are assumed to be independently drawn from an asymmetric normal distribution, that is, that is

$$p(y_t|\mu_t) = \frac{\sqrt{1-\omega} + \sqrt{\omega}}{4\sigma\sqrt{\pi\omega(1-\omega)}} \exp(-\sigma^{-2}\rho_\omega(y_t - \mu_t)). \quad (6)$$

Differentiating S_ω and dividing by minus two gives

$$\sum_{t=1}^{T} IE(y_t - \mu\,(\omega)) \quad (7)$$

where

$$IE(y_t - \mu\,(\omega)) = |\omega - I(y_t - \mu(\omega) < 0)| \ \ (y_t - \mu\,(\omega)), \qquad t = 1, ..., T. \quad (8)$$

There is no problem with defining $IE(0):$ since (8) is continuous, $IE(0) = 0$. The sample expectile, $\widetilde{\mu}(\omega)$, is the value of $\mu(\omega)$ that makes (7) equal to zero. Setting $\omega = 0.5$ gives the mean, that is $\widetilde{\mu}(0.5) = \overline{y}$. For other ω it is necessary to iterate to find $\widetilde{\mu}(\omega)$. In a regression context $\mu(\omega) = \mathbf{x}_t'\boldsymbol{\beta}$ and the estimates are obtained by asymmetric least squares.

The population expression corresponding to (7) is

$$(1-\omega) \int_{-\infty}^{\mu(\omega)} (y - \mu(\omega))\, dF\,(y) + \omega \int_{\mu(\omega)}^{\infty} (y - \mu(\omega))\, dF\,(y)\,, \quad (9)$$

where $F(y)$ is the cdf of y. Setting it to zero gives the expectile. Newey and Powell (1987, theorem 1) show that a unique solution exists if $E(y) = \mu(0.5) = \mu$ exists. A simple re-arrangement of the equation with $\widetilde{\mu}(\omega)$ is set to $\widetilde{\xi}(\tau)$ gives

$$\omega = \frac{\int_{-\infty}^{\xi(\tau)}(y-\xi(\tau))dF(y)}{\int_{-\infty}^{\xi(\tau)}(y-\xi(\tau))dF(y) - \int_{\xi(\tau)}^{\infty}(y-\xi(\tau))dF(y)} = \frac{\left[\int_{-\infty}^{\xi(\tau)} ydF(y)\right] - \tau\xi(\tau)}{2\left[\int_{-\infty}^{\xi(\tau)} ydF(y)\right] - \mu + (1-2\tau)\xi(\tau)} \tag{10}$$

For a Gaussian distribution

$$\omega = \frac{(2\pi)^{-1/2}\exp(-\xi(\tau)^2/2) + \tau\xi(\tau)}{(2/\pi)^{1/2}\exp(-\xi(\tau)^2/2) + (2\tau-1)\xi(\tau)} \tag{11}$$

where $\xi(\tau)$ is the $\tau-$quantile from $N(0,1)$. Thus for $\tau = 0.01, 0.05,\ 0.10, 0.25,$ 0.331 the corresponding values of ω are 0.00146, 0.0124, 0.0344, 0.153 and 0.250. For a uniform distribution we can find an expression for $\xi(\tau)$ in terms of τ with the result that $\omega = \tau^2/(2\tau^2 - 2\tau + 1)$. Similarly, if the observations are assumed to come from a double exponential (Laplace) distribution,

$$\omega = \frac{(\ln(2\tau)-1)\exp(-|\ln(2\tau)|) - \tau 2\ln(2\tau)}{2(\ln(2\tau)-1)\exp(-|\ln(2\tau)|) + 2(1-2\tau)\ln(2\tau)}, \qquad \tau \le 0.5; \tag{12}$$

see appendix B. For all these distributions, $\mu(\omega) = \xi(\tau)$ implies $\omega < \tau$ for $\tau < 0.5$ (and $\omega > \tau$ for $\tau > 0.5$). Conversely $|\mu(\tau)| < |\xi(\tau)|$.

The asymptotic distributions of the sample quantiles and expectiles can be found in Koenker (2005, p 71-2) and Newey and Powell (1987, theorem 3) respectively. Thus, for a standard Gaussian distribution, the variances of the limiting distribution for the quantiles at $\tau = 0.5, 0.25, 0.10, 0.05$ and 0.01 are $1.57, 1.86, 2.92, 4.47$ and 13.94. The corresponding figures for expectiles, using the $\omega's$ obtained from (11), are $1, 1.29, 2.44, 4.23$ and 16.28. The expectiles are more efficient, except in the extreme tails. For a Laplace distribution, on the other hand, quantiles are more efficient, while with a heavy-tailed distribution, such as the Cauchy, the expectiles may not even have a limiting distibution.

Breckling and Chambers (1988) try to combine robustness and efficiency by considering $M-$quantiles, denoted $M(\tau)$. Let

$$IM(y_t - M(\tau)) = \begin{cases} (\tau-1)\sigma(\tau_L), & if \quad y_t < M(\tau) - \sigma(\tau_L) \\ (1-\tau)(y_t - M(\tau)), & if \quad M(\tau) - \sigma(\tau_L) < y_t < M(\tau) \\ \tau(y_t - M(\tau)), & if \quad M(\tau) \leq y_t < M(\tau) + \sigma(\tau_U) \\ \tau\sigma(\tau_U), & if \quad y_t \geq M(\tau) + \sigma(\tau_U) \end{cases} \tag{13}$$

where $\sigma(\tau_L)$ and $\sigma(\tau_U)$ depend on some measure of scale. The $M-$quantile is estimated by solving $\sum IM(y_t - \widetilde{M}(\tau)) = 0$. Note that, as with expectiles, it is not generally true that $M(\tau)$ is the same as the $\tau-$quantile.

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

$$y_t = \xi_t(\tau) + \varepsilon_t(\tau), \qquad t = 1, ..., T, \tag{14}$$

where $\Pr(y_t - \xi_t < 0) = \Pr(\varepsilon_t < 0) = \tau$ 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 (4), and is independent of the disturbances driving ξ_t, we end up choosing the estimated quantiles so as to minimise $\sum_t \rho_\tau(y_t - \xi_t)$ subject to a set of constraints imposed by the time series model for the quantile. The model for expectiles, $y_t = \mu_t(\omega) + \varepsilon_t(\omega)$, is similar, except that the distribution of $\varepsilon_t(\omega)$ is asymmetric normal.

We will focus attention on three time series models, all of which are able to produce quantiles and expectiles that change relatively slowly over time with varying degrees of smoothness.

3.1 Models for evolving quantiles and expectiles

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

$$\xi_t(\tau) = (1-\phi_\tau)\xi_\tau^\dagger + \phi_\tau \xi_{t-1}(\tau) + \eta_t(\tau), \qquad |\phi_\tau| < 1, \qquad t = 1, ..., T, \tag{15}$$

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 $\xi^\dagger_\tau$ is the unconditional mean of $\xi_t(\tau)$. In what follows the τ appendage will be dropped where there is no ambiguity. The models for expectiles are of exactly the same form.

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

$$\begin{aligned} \xi_t &= \xi_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned} \tag{16}$$

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

$$Var \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} = \sigma^2_\zeta \begin{bmatrix} 1/3 & 1/2 \\ 1/2 & 1 \end{bmatrix} \tag{17}$$

results in the smoothed estimates being a cubic spline; see the penultimate section.

3.2 Theory and computation

The theory underlying the signal extraction approach to time-varying quantiles and expectiles can be stated generally for a model in state space form (SSF). This then leads on to a general algorithm based on the Kalman filter and associated smoother (KFS).

The state space model for a univariate time series is:

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\alpha}_t + \varepsilon_t, \qquad Var(\varepsilon_t) = h_t, \qquad t = 1, ..., T \\ \boldsymbol{\alpha}_t &= \mathbf{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad Var(\boldsymbol{\eta}_t) = \mathbf{Q}_t \end{aligned} \tag{18}$$

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 scalar, $\mathbf{T}_t$ is an $m \times m$ non-stochastic transition matrix and $\mathbf{Q}_t$ is an $m \times m$ covariance matrix. The specification is completed by assuming that $\boldsymbol{\alpha}_1$ has mean $\mathbf{a}_{1|0}$ and covariance matrix $\mathbf{P}_{1|0}$ and that the disturbances ε_t and $\boldsymbol{\eta}_t$ are independent of each other and of the initial state. In what follows we will assume that the initial state and the $\boldsymbol{\eta}_t' s$ are normally distributed. We will also assume that h_t is positive and $\mathbf{Q}_t$ positive definite for all $t = 1, ..., T$.

The joint density of the observations and the states is, ignoring irrelevant terms,

$$J = -\sum_{t=1}^{T} h_t^{-1} \rho_\tau(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t) - \frac{1}{2}\sum_{t=2}^{T} \boldsymbol{\eta}_t'\mathbf{Q}_t^{-1}\boldsymbol{\eta}_t - \frac{1}{2}(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})'\mathbf{P}_{1|0}^{-1}(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}), \tag{19}$$

where $\rho_\tau(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t)$ is as in (2) or (5) and $h_t = \lambda$ or σ^2 in a time-invariant model. For expectiles, differentiating J with respect to each element of $\boldsymbol{\alpha}_t$ gives

$$\begin{aligned}
\frac{\partial J}{\partial \boldsymbol{\alpha}_1} &= \mathbf{z}_1(2/h_1)IE(y_1 - \mathbf{z}_1'\boldsymbol{\alpha}_1) - \mathbf{P}_{1|0}^{-1}(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}) + \mathbf{T}_2'\mathbf{Q}_2^{-1}(\boldsymbol{\alpha}_2 - \mathbf{T}_2\boldsymbol{\alpha}_1) \\
\frac{\partial J}{\partial \boldsymbol{\alpha}_t} &= \mathbf{z}_t(2/h_t)IE(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t) - \mathbf{Q}_t^{-1}(\boldsymbol{\alpha}_t - \mathbf{T}_t\boldsymbol{\alpha}_{t-1}) + \mathbf{T}_{t+1}'\mathbf{Q}_{t+1}^{-1}(\boldsymbol{\alpha}_{t+1} - \mathbf{T}_{t+1}\boldsymbol{\alpha}_t), \\
&\quad t=2, \ldots, T-1, \\
\frac{\partial J}{\partial \boldsymbol{\alpha}_T} &= \mathbf{z}_T(2/h_T)IE(y_T - \mathbf{z}_T'\boldsymbol{\alpha}_T) - \mathbf{Q}_T^{-1}(\boldsymbol{\alpha}_T - \mathbf{T}_T\boldsymbol{\alpha}_{T-1}).
\end{aligned} \tag{20}$$

The smoothed estimates, $\widetilde{\boldsymbol{\alpha}}_t$, satisfy the equations obtained by setting these derivatives equal to zero. When $\omega = 0.5$, 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 J 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 Koopman (1997). An algorithm is available as a subroutine in the SsfPack set of programs within Ox; see Koopman *et al* (1999).

More generally, for any expectile, adding and subtracting $\mathbf{z}_t h_t^{-1}\mathbf{z}_t'\boldsymbol{\alpha}_t$ to the equations in (20) yields

$$\mathbf{z}_t h_t^{-1}[\mathbf{z}_t'\boldsymbol{\alpha}_t + 2.IE\,(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t)] - \mathbf{z}_t h_t^{-1}\mathbf{z}_t'\boldsymbol{\alpha}_t, \qquad t = 1, \ldots, T. \tag{21}$$

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

$$\widehat{y}_t^{(i-1)} = \mathbf{z}'\widehat{\boldsymbol{\alpha}}_t^{(i-1)} + 2.IE\left(y_t - \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(i-1)}\right). \qquad (22)$$

The iterations are carried out until the $\widehat{\boldsymbol{\alpha}}_t^{(i)\prime}s$ converge whereupon $\widetilde{\mu}_t(\omega) = \mathbf{z}_t'\widetilde{\boldsymbol{\alpha}}_t$. The algorithm can be easily adapted to estimate $M-$quantiles.

For quantiles, the first term in each of the three equations of (20) is given by $\mathbf{z}_t h_t^{-1} IQ(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t)$ and the synthetic observations in the KFS are

$$\widehat{y}_t^{(j-1)} = \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(j-1)} + IQ\left(y_t - \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(j-1)}\right), \qquad t = 1, ..., T \qquad (23)$$

However, the possibility of a solution where the estimated quantile passes through an observation means that the algorithm has to be modified somewhat; see De Rossi and Harvey (2006).

For the random walk quantile, the crucial parameter is the quasi-signal noise ratio, $q_{\xi(\tau)} = \sigma^2_{\eta(\tau)}/\lambda$. (The τ subscript will be dropped if there is no ambiguity). For a doubly infinite sample

$$\widetilde{\xi}_t = \frac{1-\theta}{1+\theta}\sum_{j=-\infty}^{\infty}\theta^{|j|}[\widetilde{\xi}_{t+j} + IQ(y_{t+j} - \widetilde{\xi}_{t+j})], \qquad (24)$$

where $\theta = (q_\xi + 2)/2 - \left(q_\xi^2 + 4q_\xi\right)^{1/2}/2$; see De Rossi and Harvey (2006, section 3.3). The corresponding expression for expectiles is similar:

$$\widetilde{\mu}_t = \frac{1-\theta}{1+\theta}\sum_{j=-\infty}^{\infty}\theta^{|j|}[\widetilde{\mu}_{t+j} + 2IE(y_{t+j} - \widetilde{\mu}_{t+j})] \qquad (25)$$

with θ defined with $q_\mu = \sigma^2_\eta/\sigma^2$ replacing q_ξ. However, because IE (times two) replaces IQ, it is easy to see that θ, and hence q is not affected by a change in scale. Note that when $\omega = 0.5$ expression (25) reduces to the classic Wiener-Komogorov formula.

3.3 Properties

Estimates of time-varying quantiles and expectiles obtained from the smoothing equations of the previous sub-section can be shown to satisfy properties that generalize the defining characteristics of fixed quantiles and expectiles.

Definition 1 *The fundamental property of sample time-varying quantiles is that the number of observations that are less than the corresponding quantile, that is* $y_t < \widetilde{\xi}_t(\tau)$, *is no more than* $[T\tau]$ *while the number greater is no more than* $[T(1-\tau)]$.

Definition 2 *The basic moment condition for time-varying expectiles is that the weighted residuals sum to zero, that is*

$$\sum_{t=1}^{T} \left|\omega - I\left(y_t - \widetilde{\mu}_t(\omega)\right)\right| .(y_t - \widetilde{\mu}_t(\omega)) = 0 \qquad (26)$$

In order to establish the conditions under which these properties holds, we first prove a preliminary result for any time series model in SSF, (18). It is assumed that the state has been arranged so that the first element represents the level and that (without loss of generality) the first element in $\mathbf{z}_t$ has been set to unity. Let the first derivative, with respect to $\boldsymbol{\alpha}_t$, of the second term of J be written $J_2' = \sum_{t=1}^{T} \mathbf{A}_t\boldsymbol{\alpha}_t$, where the $\mathbf{A}_t's$ are $m \times m$ matrices.

Lemma 3 *For a model in SSF with a diffuse prior on the initial state, a sufficient condition for the first element in the vector* J_2' *to be zero is that the first column of* $\mathbf{T}_t - \mathbf{I}$ *consists of zeroes for all* $t = 2, ..., T$.

Proof - Summing the terms in derivatives in question gives

$$\sum \mathbf{A}_t\boldsymbol{\alpha}_t = (\mathbf{Q}_2^{-1}\mathbf{T}_2 - \mathbf{T}_2'\mathbf{Q}_2^{-1}\mathbf{T}_2)\boldsymbol{\alpha}_1$$
$$+\sum_{t=2}^{T-1}(\mathbf{Q}_{t+1}^{-1}\mathbf{T}_{t+1} - \mathbf{T}_{t+1}'\mathbf{Q}_{t+1}^{-1}\mathbf{T}_{t+1} + \mathbf{T}_t'\mathbf{Q}_t^{-1} - \mathbf{Q}_t^{-1})\boldsymbol{\alpha}_t + \left(\mathbf{T}_T'\mathbf{Q}_T^{-1} - \mathbf{Q}_T^{-1}\right)\boldsymbol{\alpha}_T \qquad (27)$$

The matrix associated with $\boldsymbol{\alpha}_1$ is $\mathbf{A}_1 = \mathbf{Q}_2^{-1}\mathbf{T}_2 - \mathbf{T}_2'\mathbf{Q}_2^{-1}\mathbf{T}_2 = (\mathbf{I} - \mathbf{T}_2')\mathbf{Q}_2^{-1}\mathbf{T}_2$. A sufficient condition for it to have a null first row is that $\mathbf{I} - \mathbf{T}_2'$ has a null first row. The matrix associated with $\boldsymbol{\alpha}_T$ is $(\mathbf{T}_T' - \mathbf{I})\mathbf{Q}_T^{-1}$ and the condition for it to have a null first row is that $\mathbf{T}_T' - \mathbf{I}$ has a null first row. On examining the matrices, $\mathbf{A}_t, t = 2, .., T-1$, associated with the remaining state vectors we see that an analogous condition is sufficient for each to have a null first row.

Remark 4 *Letting some of the states have proper priors does not affect the result as long as they are uncorrelated with the diffuse prior on the first element in the state.*

Proposition 5 *If the condition of the Lemma holds, then, for expectiles, the generalized moment condition*

$$\sum_{t=1}^{T} |\omega - I(y_t - \mathbf{z}_t'\widetilde{\boldsymbol{\alpha}}_t)| (y_t - \mathbf{z}_t'\widetilde{\boldsymbol{\alpha}}_t)/h_t = 0,$$

holds. If h_t is time-invariant, the basic moment condition, (26), is satisfied.

The result follows because, when the first element in the vector $\sum \mathbf{A}_t\boldsymbol{\alpha}_t$ is zero, differentiating the first term in J gives

$$\sum_{t=1}^{T} h_t^{-1} IE(y_t - \mathbf{z}_t'\widetilde{\boldsymbol{\alpha}}_t) = 0.$$

The results for quantiles require a little more work.

Proposition 6 *If h_t is time-invariant and the conditions of Lemma 1 hold, the estimated quantiles satisfy the fundamental property.*

Proof - Suppose that the only one point at which the quantile passes through an observation is at $t = s$, so $\widetilde{\xi}_s = y_s$. All the derivatives of J, defined in (19) with $\rho_\tau(.)$ is as in (2), can be set to zero apart from this one. However, 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. Thus to have

$$IQ(y_s - \widetilde{\xi}_s) + \sum_{t \neq s} IQ(y_t - \widetilde{\xi}_t) = 0$$

implies

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

When the quantile passes through k observations, a similar argument leads to

$$-k\tau \leq \sum_{t \notin C} IQ(y_t - \widetilde{\xi}_t) \leq k\,(1 - \tau) \qquad (28)$$

where C is the set of all points such that $\widetilde{\xi}_s = y_s$. 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 $[(1-\tau)T]$.

Proposition 7 *If h_t is not time-invariant, the estimated quantiles satisfy a generalization of the fundamental property, which is that*

$$\sum_{t \in B} 1/h_t \leq \tau \sum_{t=1}^{T} 1/h_t \qquad \text{and} \qquad \sum_{t \in A} 1/h_t \leq (1-\tau) \sum_{t=1}^{T} 1/h_t$$

where $t \in B$ denotes the set of observations below the corresponding quantile and $t \in A$ denotes the set above.

The result follows because corresponding to (28) we have

$$-\tau \sum_{t \in C} 1/h_t \leq \sum_{t \notin C} (1/h_t) IQ(y_t - \mathbf{z}_t' \widetilde{\boldsymbol{\alpha}}_t) \leq (1-\tau) \sum_{t \in C} 1/h_t \qquad (29)$$

The condition of the lemma is obviously satisfied by the random walk. It is also satisfied by the local linear trend with (pd) covariance matrix as in (17); the SSF for the local linear trend has $\boldsymbol{\alpha}_t = (\xi_t, \beta_t)'$ and $\mathbf{z}' = (1\ 0)$. When a model contains a fixed level, as with the AR(1) process of (15), it may be put in the state vector and (without any loss in generality) assigned to the first position. Thus if the level is denoted $\mu^\dagger$, we put $\mu_t^\dagger = \mu_{t-1}^\dagger$ in the transition equation. If $\mu_t^\dagger$ is treated as stochastic, the condition in the lemma holds.[2]Finally, in a model with fixed explanatory variables, $\mathbf{x}_t$, the

[2]From the computational point of view including $\mu^\dagger$ in the state may not be efficient. If it is taken out, it can be estimated by a single closed form equation expressed in terms of the current estimates of the state. For example with an AR(1) model for expectiles, with the stationary zero mean component, $\mu_t - \mu^\dagger$, initiated with mean zero and variance $\sigma_\eta^2/(1-\phi^2)$,

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

The estimator, $\widetilde{\mu}^\dagger$, has an appealing symmetry.

first equation in (18) becomes

$$y_t = \mathbf{x}_t'\boldsymbol{\delta} + \mathbf{z}_t'\boldsymbol{\alpha}_t + \varepsilon_t, \qquad t = 1, ..., T$$

and if the coefficient vector is put in the state vector as $\boldsymbol{\delta}_t$ and given a diffuse prior, the conditions apply to the transition equation for $\boldsymbol{\alpha}_t$ as before.

Finally we turn to the conditions under which quantiles and expectiles match up when they are time-varying.

Proposition 8 *If the distribution of y is time invariant when adjusted for changes in location and scale, and is continuous with finite mean, the population $\tau-$quantiles and $\omega-$expectiles coincide for ω satisfying (10). Assuming this to be the case, $\widetilde{\mu}_t(\omega)$ is an estimator of the $\widetilde{\tau}-$quantile, $\xi_t(\widetilde{\tau})$, where $\widetilde{\tau}$ is defined as the proportion of observations for which $y_t < \widetilde{\mu}_t(\omega), t = 1, ..., T$.*

When used in this way we will denote the estimator $\widetilde{\mu}_t(\omega)$ as $\widetilde{\mu}_t(\widetilde{\tau})$. Note that it satisfies the fundamental property of definition 1 by construction. However, it will not, in general, coincide with the time-varying $\widetilde{\tau}-$quantile estimated directly since it weights the observations differently. In particular, it is unlikely to pass through any observations.

3.4 Non-parametric smoothing

When the quantiles or expectiles change over time they may be estimated non-parametrically. The simplest option is to compute them from a moving window. More generally estimation at any point in time is carried out by minimising a local check function, that is $\sum_{j=-h}^{h} K(j/h)\rho_\tau(y_{t+j} - \xi_t)$ 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. For quantiles $\widehat{\xi}_t$ must satisfy

$$\sum_{j=-h}^{h} K(j/h) IQ(y_{t+j} - \widehat{\xi}_t) = 0 \qquad (30)$$

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

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

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 signal-noise ratio plays the same role as the bandwidth. Of course, the model-based approach not only provides the basis for forecasting but it also 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.

Given the connection with non-parametric smoothing, the following proposition is relevant later when we go on to discuss setting parameters for extracting quantiles and expectiles. The only reason it doesn't apply directly is that in the model-based formula, $\widehat{\xi}_{t+j}$ is used instead of $\widehat{\xi}_t$ when j is not zero.

Proposition 9 *If the same kernel and bandwidth are used for different quantiles and expectiles they cannot cross (though they may touch).*

The result follows for quantiles on noting that, if $\widehat{\xi}_t$ does not coincide with an observation, (30) is

$$(\tau - 1)\sum_{j \in B} k_j + \tau \sum_{j \in A} k_j = 0$$

where $k_j = K(j/h)/\sum_{j=-h}^{h} K(j/h) \geq 0$ and $B(A)$ denotes the set of $k_j's, j = -h, .., h$, such that y_{t+j} is below (above) $\widehat{\xi}_t$. The result should now be apparent from inspection though it becomes even clearer if the equation is re-arranged to give $\tau = \sum_{j \in B} k_j$; if τ increases then $\widehat{\xi}_t$ must increase to (at least) the observation immediately above for $\sum_{j \in B} k_j$ to increase. Remember that when $\widehat{\xi}_t$ is equal to an observation, IQ is set to a value, τ^*, such that $\tau - 1 < \tau^* < \tau$. If (the original) $\widehat{\xi}_t$ is equal to an observation then

$$(\tau - 1)\sum_{j \in B} k_j + \tau \sum_{j \in A} k_j + \tau^* k_C = 0$$

where k_C denotes the weight for the observation equal to $\widehat{\xi}_t$. On re-arranging

$$\tau = \sum_{j \in B} k_j + (\tau - \tau^*) k_C$$

Since $\tau > \tau^*$, $\widehat{\xi}_t$ cannot decrease when τ increases, since if it did, $\tau^* = \tau$ and the right hand side decreases. Note that the weights don't have to be symmetric for the above result to hold, so it applies to points at the end of the sample, or near to it.

An expectile will not normally be equal to an observation so

$$\omega \sum_{j \in A} k_j (y_{t+j} - \widehat{\mu}_t) = (1 - \omega) \sum_{j \in B} k_j (y_{t+j} - \widehat{\mu}_t)$$

and it is clear that if ω increases then $\widehat{\mu}_t$ must also increase to compensate.

4 Parameter estimation

As noted at the end of sub-section 3.2, the decay in the weighting pattern is determined by parameters q_ξ and q_μ for quantiles and expectiles respectively. Note that q_ξ is expressed relative to λ which is not a variance although it is treated as such when the KFS is run. Hence q_ξ is not a true signal-noise ratio and, as was noted in sub-section 3.2, it is not scale invariant. De Rossi and Harvey (2006) discusse some of the practical implications for estimation.

4.1 Maximum likelihood

Maximum likelihood (ML) estimation of the unknown parameters is easily carried out for Gaussian unobserved components models using the prediction errors from the Kalman filter. Figure 1 shows 288 monthly figures on US inflation[3] from 1981(7) to 2005(6). Although fitting a local level model (random walk plus noise) gives some residual serial correlation the model is not unreasonable. The ML estimate of the signal-noise ratio, q_μ, is 0.011. If this value is used for all expectiles, the result is as shown in the figure[4].

Figure 2 shows cubic spline expectiles, that is from a local linear trend with disturbance covariance matrix as in (17), all with a signal-noise ratio, $q_\mu = \sigma_\zeta^2/\sigma_\varepsilon^2$, estimated by ML for the mean as 0.00007. They are somewhat smoother but quite close to the RW expectiles of figure 1.

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

[4] Busetti and Harvey (2006) propose tests for determining whether such movements represent significant departures from a time invariant distribution around the mean.

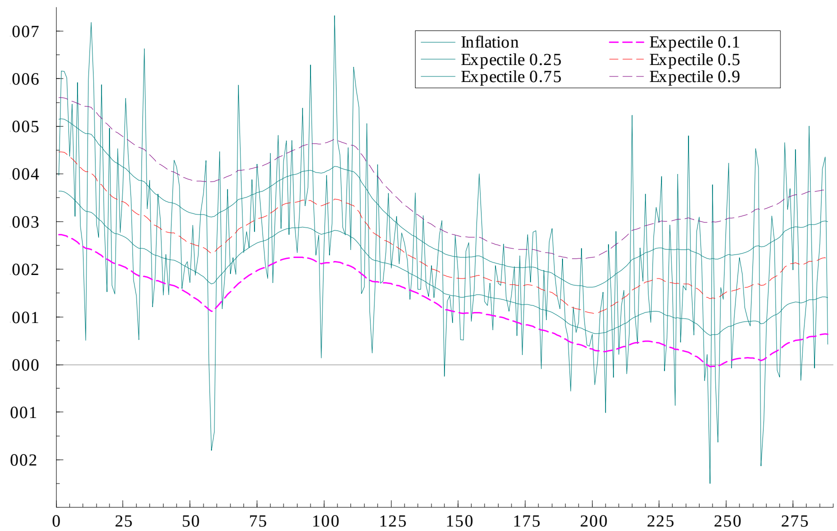

Figure 1: Time-varying RW expectiles fitted to monthly US inflation

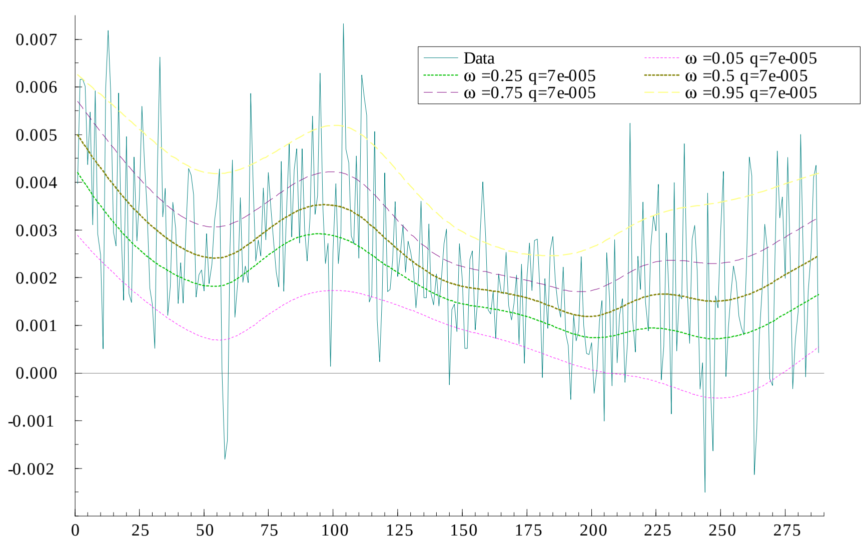

Figure 2: Time-varying cubic spline expectiles

4.2 Cross -validation

Setting the $q_\mu' s$ for all expectiles equal to the ML estimate for the mean is not always desirable. The same is true for quantiles. Furthermore ML estimation of q_ξ for the median is far more difficult because the prediction error decomposition only applies to Gaussian models.

The cross validation (CV) criterion for the mean 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. In the Gaussian case it can be computed from a single pass of the 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, cross validation represents a viable alternative.

For time-varying quantiles, the appropriate cross validation function is

$$CV(\tau) = \sum_{t=1}^{T} \rho_\tau (y_t - \widetilde{\xi}_t^{(-t)}) \tag{31}$$

where $\widetilde{\xi}_t^{(-t)}$ is the smoothed value at time t when y_t is dropped. A similar criterion, $CV(\omega)$, may be used for expectiles. Unfortunately, there appears to be no simple way of computing $CV(\tau)$ and $CV(\omega)$, $\omega \neq 0.5$, from the KFS and it would appear necessary to resort to a 'brute force' approach in which all T observations are dropped one at a time. However, if this is done, the number of iterations for each t may not be large as the starting values of the quantiles, obtained from dropping an adjacent observation will usually be close to the solution.

4.3 Choice of parameters for different quantiles

How might the $q's$ vary with τ? Figure 3 shows the smoothed estimates of the median and other quantiles, all modelled as random walks, for the US inflation data of figure 1. The $q_\xi^{1/2\prime} s$, were estimated by CV as 0.004, 0.008, 0.004, 0.006 and 0.008 for the 10%, 25 %, 50%, 75% and 90% respectively. In small samples there is a danger in estimating these parameters separately

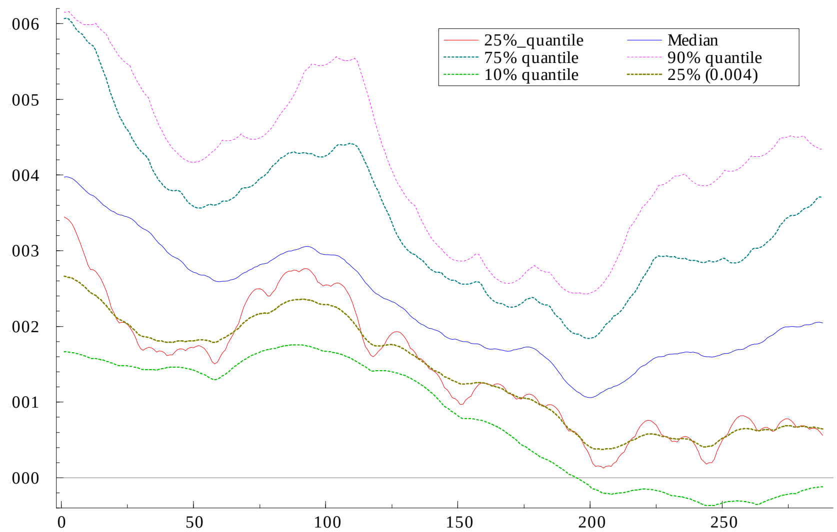

Figure 3: Inflation with two 25% quantiles

for different quantiles. For example $\xi(0.25)$ is very variable as $q^{1/2}_{\xi(0.25)} = 0.008$. When it is fitted with $q^{1/2} = 0.004$, as for the median, the changes are much less abrupt and more in keeping with the other quantiles.

The above example illustrates that one reason for having the $q's$ the same is that they are less vulnerable to small sample variation. We also saw in sub-section 3.4 that because the smoothed estimators of quantiles are closely related to non-parametric estimators, they are less likely to cross. Nevertheless, we may well have quantiles changing at different rates. For example with stock prices the median might be time-invariant. A compromise is perhaps to let q be a simple function of τ. These considerations also apply to expectiles.

5 Dispersion, Asymmetry and Heavy Tails

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. For example, 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, asymmetry or kurtosis. Evidence that changing asymmetry and kurtosis are a feature of some financial time series can be found in, for example, Jondeau and Rockinger (2003).

As noted in proposition 9, time-varying quantiles and expectiles will coincide if the shape of the distribution is constant over time. Its location and scale can vary, but not skewness and kurtosis. Thus $\mu_t(\omega) = \xi_t(\tau)$, $t = 1, .., T$, where ω depends on the standardized population distribution through (10). When there are changes in the shape of the distribution, (population) expectiles will not match up with quantiles.

5.1 Dispersion

The contrasts between complementary quantiles, that is

$$D(\tau) = \xi_t(1-\tau) - \xi_t(\tau), \qquad \tau < 0.5, \qquad t = 1, .., T \tag{32}$$

yield measures of dispersion. A corresponding measure from expectiles is

$$D_\mu(\tau) = \mu_t(1-\tau) - \mu_t(\tau), \qquad \tau < 0.5, \qquad t = 1, .., T$$

We can track its evolution, though it cannot be related to standard measures of dispersion, such as the standard deviation, without an assumption about the distribution. Figure 4 shows the interquartile range for US inflation (based on a RW with $q = 0.004$) and two IE ranges (based on IRWs).

When the distribution can be assumed symmetric about zero, the asymptotic results for fixed quantiles suggest that there is no gain from estimating IQ ranges by multiplying the $(1-2\tau) - th$ quantile for absolute values by two[5]. However, this may not carry over to time-varying quantiles, because the estimates are based on a localized weighting of observations. There is also the issue of estimating q which is easier with absolute values (as it need only be done once and it ensures the $q's$ are the same).

[5] The asymptotic variance of a fixed quantile contrast is easily obtained as $Avar(\widetilde{D}(\tau)) = 2Avar(\widetilde{\xi}(\tau)) - 2Acov(\widetilde{\xi}(\tau)\widetilde{\xi}(1-\tau))$
$= 2\tau(1-\tau)/f^2 - 2\tau^2/f^2 = 2\tau(1-2\tau)/f^2$, where $f = f(\xi(\tau))$. This is the same as the asymptotic variance of two times $\xi(1-2\tau)$ for $|y_t|$. In contrast to quantiles, it is not possible to find a general formula relating the expectiles of a symmetric random variable to the expectiles of its absolute value; the relationship depends on the distribution of y.

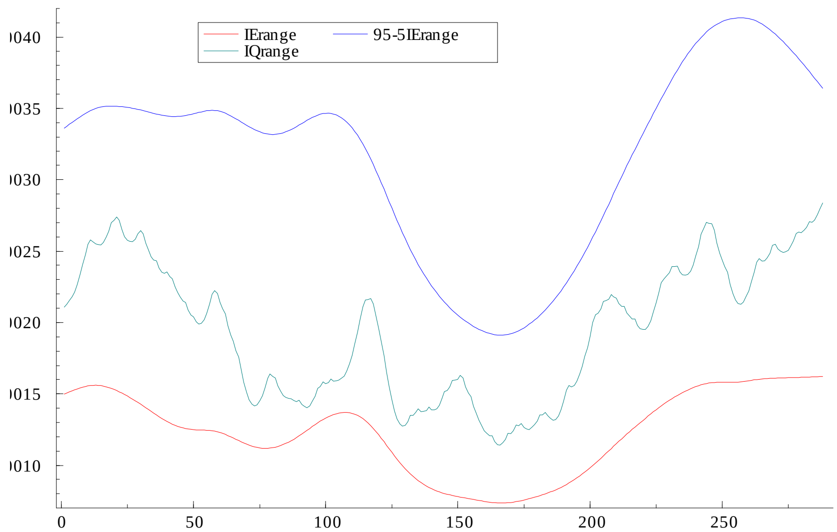

Figure 4: Inter-expectile ranges for US inflation (IRW) and IQ range (RW)

5.2 Tail dispersion and changing kurtosis

Some notion of the way in which tail dispersion changes can be obtained by plotting the ratio of the estimated interdecile range (0.1 to 0.9) or the 0.05 to 0.95 range to the interquartile range, that is

$$\widetilde{R}(0.05/0.25) = \frac{\widetilde{D}(0.05)}{\widetilde{D}(0.25)} = \frac{\widetilde{\xi}_t(0.95) - \widetilde{\xi}_t(0.05)}{\widetilde{\xi}_t(0.75) - \widetilde{\xi}_t(0.25)}. \tag{33}$$

For a normal distribution this ratio is 2.44, for t_3 it is 3.08 and for a Cauchy 6.31. Figure 5 shows the 5% and 25 % quantiles and the plot of (33) for the GM series.The calculations could also be done with symmetry imposed.

Similar plots to those shown in figure 5 can be carried out with expectiles. To assess the movements in excess kurtosis relative to a normal distribution ω is set according to (11), that is 0.0124 and 0.153 for $\tau = 0.05$ and 0.25 respectively.

There may be a variety of movements in different parts of the tail. To illustrate, we consider 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

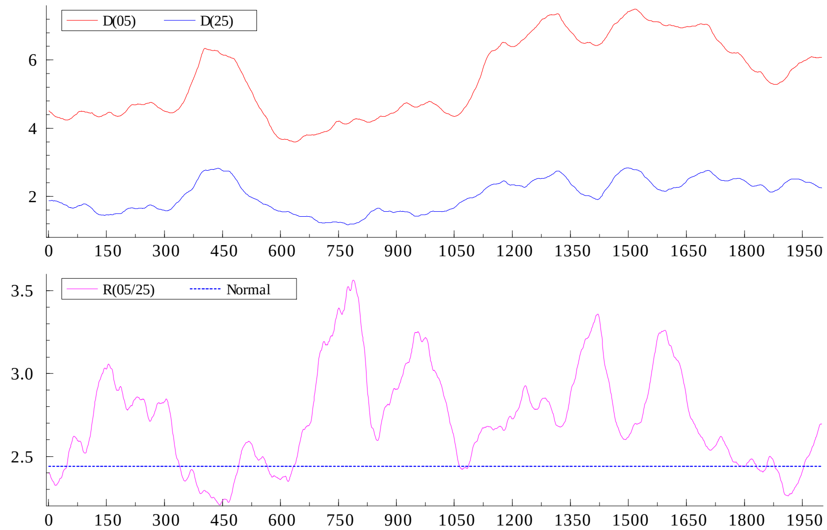

Figure 5: Interquartile, 5%-95% range and (lower graph) their ratio for GM

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[6], at a value of ten. As a result, the shape of the distribution changes because of changes in the tail dispersion. (As might be expected, the 25% and 75% quantiles are almost constant). What is particularly interesting is that the 5% and 1% quantiles move in different directions. As can be seen from figure 6, the (theoretical) 5% quantile decreases while the 1% quantile increases. It can be shown that kurtosis is a linear function of time, increasing during the period from 5.7 to 21.9. Plotting dispersion for different τ provides a more comprehensive picture.

The quantiles were estimated from the absolute values of 500 observations generated from the above model. Cross validation was used to select q_ξ for a RW model for the 50%, 90% and 98% quantiles of the absolute values. These correspond to the 25%, 5% and 1% (and 75%, 95% and 99%) quantiles of the raw data. The estimate of $q_{\xi(.05)}$ was 0.1. As can be seen from figure 6,

[6] The variance of the second component is $\sigma_t^2 = 20 + (60/500)t$, $t = 1, .., 500$, and the probability of the first component is $(1 - 9/(\sigma_t^2 - 1))$.

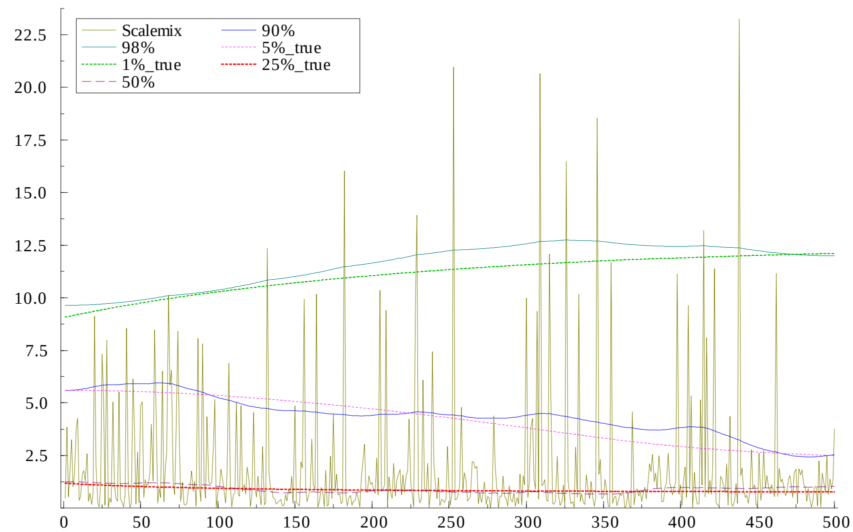

Figure 6: Absolute values of simulated series, true quantiles and random walk 98%, 90% and 50% quantiles fitted to simulated data by CV.

the estimated quantiles track the true quantiles quite well.

In order to compare expectiles and quantiles in this case, we used the $\omega's$ corresponding to $\tau = .25, .10, .05$ and $.01$ for a normal distribution[7]. In contrast to the quantiles, the two upper theoretical expectiles, shown in figure 7, both indicate increased dispersion over time. The lower expectiles are essentially constant.

5.3 Asymmetry

For a symmetric distribution

$$S(\tau) = \xi_t(\tau) + \xi_t(1-\tau) - 2\xi_t(0.5), \qquad \tau < 0.5 \tag{34}$$

[7]Other possibilities would be to work out the $\omega's$ from a heavy tailed distribution or to calculate them from estimated time-invariant quantiles via (10). However, all that is needed is a reasonable spread of expectiles in order to give a general impression of the overall pattern.

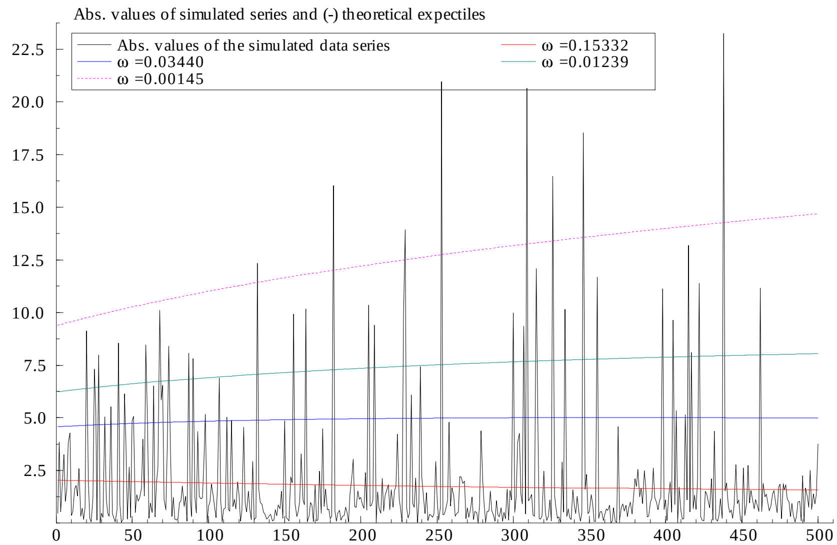

Figure 7: Absolute values of simulated series and theoretical expectiles

is zero for all $t = 1, ..., T$. Hence a plot of this contrast, or of the standardized measure[8] $S(\tau)/D(\tau)$, shows how the asymmetry captured by the complementary quantiles, $\xi_t(\tau)$ and $\xi_t(1-\tau)$ changes over time. A similar contrast for expectiles will also show the evolution of asymmetry, but not in the same way.

6 Prediction and filtering

The smoothed estimate of a quantile at the end of the sample, $\widetilde{\xi}_{T|T}$, is the filtered estimate. Predictions, $\widetilde{\xi}_{T+j|T}$, $j = 1, 2, ...$, are made by straightforwardly extending these estimates according to the time series model for the quantile. For a random walk the predictions are $\widetilde{\xi}_{T|T}$ for all lead times, while for a more general model in SSF, $\widetilde{\xi}_{T+j|T} = \mathbf{z}'\mathbf{T}^j\widetilde{\boldsymbol{\alpha}}_T$. 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

[8]See, for example, Stuart and Ord (1987, p343-4).

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; see (24). For a random walk quantile and a semi-infinite sample the filtered estimator must satisfy

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

where θ is the parameter defined below (24) and $\widetilde{\xi}_{t-j|t}$ is the smoothed estimator of ξ_{t-j} based on information at time t; see, for example, Whittle (1983, p69). Thus $\widetilde{\xi}_{t|t}$ is an exponentially weighted moving average (EWMA) of the synthetic observations, $\widetilde{\xi}_{t-j|t} + IQ(y_{t-j} - \widetilde{\xi}_{t-j|t})$. Similarly for an expectile

$$\widetilde{\mu}_{t|t}(\tau) = (1-\theta)\sum_{j=0}^{\infty}\theta^j[\widetilde{\mu}_{t-j|t} + 2IE(y_{t-j} - \widetilde{\mu}_{t-j|t})].$$

7 Nonparametric regression with cubic splines

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

$$\min\sum_{t=1}^{T}\rho_\tau\{y_t - \xi(x_t)\} + \lambda_2\left(\int\{\xi''(x)\}^2dx\right) \tag{35}$$

where $\xi(x)$ is a continuous function with square integrable second derivative, $0 \leq x \leq T$ and $x_t = t$. The parameter λ_2 controls the smoothness of the spline. We show in appendix A that the same cubic spline is obtained by quantile signal extraction of (16) and (17) with $\lambda_2 = \lambda/2\sigma_\zeta^2$. A random walk corresponds to $\xi'(x)$ rather than $\xi''(x)$ in the above formula; compare Kohn, Ansley and Wong (1992). Our 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 and shows that the result holds for expectiles. We furthermore establish the existence and uniqueness of the solution.

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 apart imply a variance for the discrete random walk of $\delta_t \sigma_\eta^2$, while for the continuous time IRW, (17) becomes

$$Var \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix} = \sigma_\zeta^2 \begin{bmatrix} (1/3)\delta_t^3 & (1/2)\delta_t^2 \\ (1/2)\delta_t^2 & \delta_t \end{bmatrix} \tag{36}$$

while the second element in the first row of the transition matrix is δ_t. The importance of this generalisation is that it allows the handling of nonparametric quantile and expectile regression by cubic splines when there is only one explanatory variable.[9] The observations, which may be from a cross-section, 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[10]. Unfortunately this necessitates the repeated inversion of large matrices of dimension up to $4T \times 4T$. This is very time consuming[11]. Our signal extraction appears to be much faster (and more general) and makes estimation of λ a feasible proposition. Bosch *et al.* had to resort to setting λ as small as possible without the quantiles crossing.

The fundamental property of quantiles continues to hold with irregularly spaced observations. All that happens is that the SSF becomes time-varying. If there are multiple observations at some points then n, the total number of observations, replaces T, number of distinct points, in the summation. The proof follows by adding more $\rho(.)$ terms at times where there are multiple observations.

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

[10] 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.

[11] Bosch et al (1995) report that it takes almost a minute on a Sun workstation for sample size less that 100.

Proposition 10 *If n denotes the total number of observations while T is the number of distinct points at which observations occur, the fundamental property of quantiles is stated in terms of n rather than T.*

The only difference in the proof is in the summation involving $IQ(.)'s$ which now becomes

$$\sum_{j=1}^{n} h_t^{-1} IQ(y_{j(t)} - \mathbf{z}_t' \widetilde{\boldsymbol{\alpha}}_t)$$

where $y_{j(t)}$ denotes that the $j - th$ observation is observed at point t for $t = 1, ..., T$.

Proposition 11 *If there are multiple observations at some points, the generalized moment condition for expectiles is*

$$\sum_{j=1}^{n} h_t^{-1} \left| \tau - I(y_{j(t)} - \mathbf{z}_t' \widetilde{\boldsymbol{\alpha}}_t) \right| (y_{j(t)} - \mathbf{z}_t' \widetilde{\boldsymbol{\alpha}}_t) = 0$$

The generalized cross-validation algorithm is regarded as being more appropriate for irregularly spaced observations; see Kohn et al (1992). However, it is not clear how to apply it with the brute force algorithm. Here we use the basic CV criterion, (31), for both quantiles and expectiles.

An example of cubic spline regression is provided by the "motorcycle data", which records measurements of the acceleration, in milliseconds, of the head of a dummy in motorcycle crash tests. The data set was originally analysed by Silverman (1985) and has been used in a number of textbooks, including Koenker (2005, p 222-6). The observations are irregularly spaced and at some time points there are multiple observations. Harvey and Koopman (2000) highlight the stochastic trend connection.

Figure 8 shows the IRW time-varying expectiles obtained using the value of $q_\mu = 0.07$ computed by CV for the mean. The ML estimate of q_μ reported by Harvey and Koopman (2000, p 17) is $q_\mu = 0.03$; using this gives slightly less movement but there is little difference in the overall pattern. Although the expectiles lack the nice interpretation of quantiles, the graph gives a clear visual impression of the movements in level and dispersion. (Of course if we count the number of observations below each expectile, they can be interpreted as quantiles if we are prepared to assume that the shape of the distribution is time invariant.)

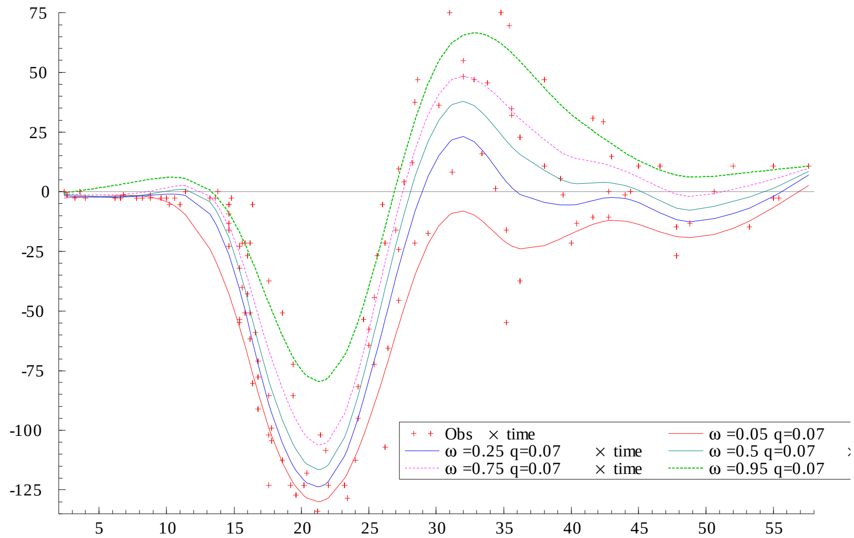

Figure 8: Cubic spline expectiles fitted to the motorcycle data. The parameter q_μ is estimated by cross validation.

Figure 9 shows the 75-25 and the 95-5 IE ranges. These measures track movements in dispersion very clearly.

The random walk specification is not sufficiently flexible to adapt to the sharp change that occurs just before 15 seconds. This is particularly apparent for the lowest expectile as can be seen from figure 11. Figure 10 shows the estimated time varying quartiles and median for the CV value of $q_\xi = 0.0625$, and a similar pattern is seen.

Harvey and Koopman (2000) re-estimate the spline with a correction for heteroscedasticity constructed from the residuals. The SSF is easily amended to allow for heteroscedasticity. A heteroscedasticity correction could also be made for the expectiles using a range measure as in figure 9.

8 Conclusions

Time-varying quantiles can be fitted iteratively applying a suitably modified state space signal extraction algorithm. Here we show that time-varying expectiles can also be estimated by a state space signal extraction algorithm.

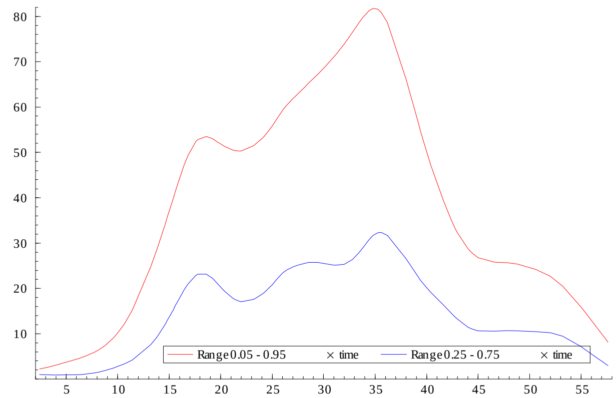

Figure 9: Motorcycle data: 75-25 and the 95-5 inter-expectile ranges.

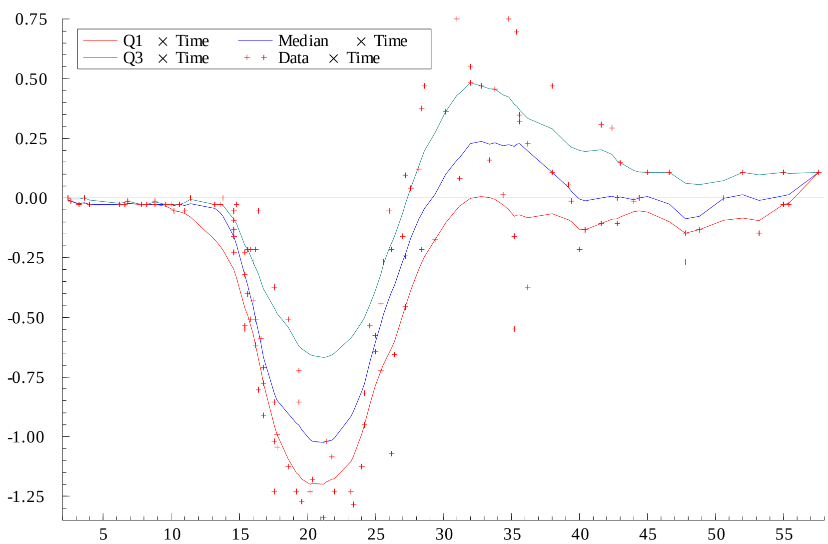

Figure 10: Quartiles and median fitted to the motorcycle data for $q_\xi = 0.0625$.

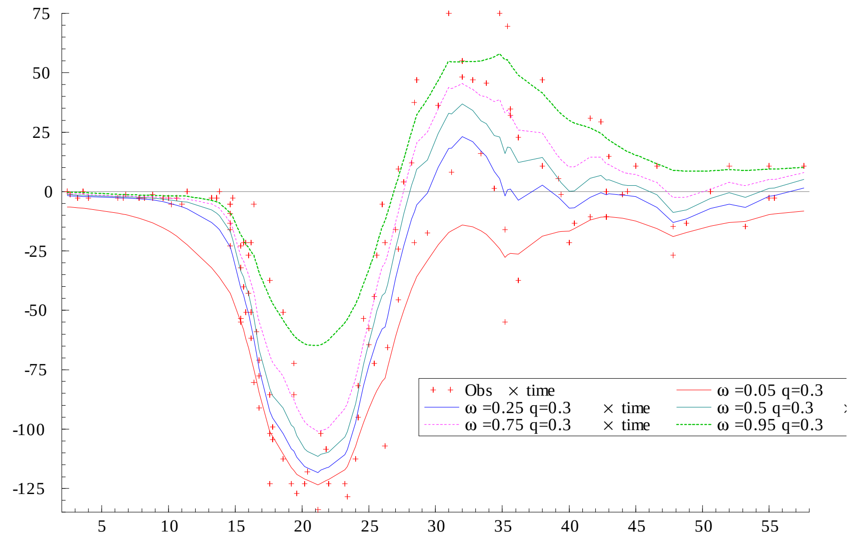

Figure 11: Time varying RW expectiles fitted to the motorcycle data. The parameter q_μ is estimated by cross validation.

The algorithm is much faster than the one for quantiles as there is no need to take account of corner solutions. We derive the conditions under which time-varying quantiles satisfy the defining property of fixed quantiles in having the appropriate number of observations above and below it, while expectiles satisfy properties that generalize the moment conditions associated with fixed expectiles. Practical ways in which expectiles can be made robust are also proposed.

Time-varying quantiles and expectiles provide information on various aspects of a time series, such as dispersion and asymmetry, and we investigate how contrasts between them can be informative in pointing to departures from the assumption of a distribution that is time invariant apart from location and scale. A simulated example illustrates how tail dispersion, as measured by different inter-quantile ranges, is a richer concept than kurtosis and tracking the movements in different $\tau-$quantiles can be very informative.

Our model-based approach means that time-varying quantiles and expectiles can be used for forecasting. As such they offer an alternative to methods such as those in Engle and Manganelli (2004) and Granger and Sin (2000), that are based on conditional autoregressive models.

Finally we prove that if the underlying time series model is a Wiener process or an integrated Wiener process, then the solution for quantiles, expectiles and M-quantiles is equivalent to fitting a spline; for an integrated Wiener process this is a cubic spline. We furthermore establish the existence and uniqueness of the solution. Because the state space form can handle irregularly spaced observations, the proposed algorithms are easily adapted to provide a viable means of computing spline-based non-parametric quantile and expectile regressions. We demonstrated how this worked for the 'motorcycle' data and showed, in that case, that fitting cubic spline expectiles gave a clear visual impression of the changing distribution.

Further work remains to be done. In particular, there are the issues of inference on parameters and the MSEs for smoothed estimates. It may also be worth investigating whether joint estimation of a set of quantiles (or expectiles) holds any advantages. Finally we note that Busetti and Harvey (2006) have proposed tests for the null hypothesis that quantiles are time-invariant.

A State space representation of quantile, expectile and M-quantile regression with smoothing splines

Consider a set of n observations $(y_1, \ldots, y_n)$ obtained at times $(t_1, \ldots, t_n)$, where $0 \leq t_1 < \ldots < t_n \leq b$. Moreover, consider a loss function $\rho_\tau(x) \geq 0$ such that $\int_{-\infty}^{\infty} \exp[-\rho_\tau(x)] \, \mathrm{d}x < \infty$. We will deal with the problem of finding the function $f : [0, b] \to \mathbb{R}$ that minimises

$$\lambda_m \int_0^b \left[f^{(m)}(t)\right]^2 \mathrm{d}t + \sum_{i=1}^n \rho_\tau\left(y_i - f(t_i)\right) \qquad (37)$$

for given $\tau \in (0, 1)$ and m, over all functions f having $m - 1$ absolutely continuous derivatives and square integrable $m-$th derivative.

Now consider the time series representation obtained by assuming that:

- $\left[f(0), f'(0), \ldots, f^{(m-1)}(0)\right] \sim \mathcal{N}(\mathbf{0}, \kappa \mathbf{I}_m)$;

- $$f(t) = \sum_{j=1}^{m} \frac{t^{j-1}}{(j-1)!} f^{(j-1)}(0) + \sigma_w \int_0^t \frac{(t-s)^{m-1}}{(m-1)!} \mathrm{d}W_s \tag{38}$$

  where W_t is a Wiener process (in terms of the notation of sub-section 3.1, $\sigma_w = \sigma_\eta$ for $m = 1$ and σ_ζ for $m = 2$;

- the distribution of $y_i | f(t_i)$ is characterized by a pdf defined as

  $$p(y_i | f(t_i)) \propto \exp\left[-\overline{\lambda}^{-1} \rho_\tau (y_i - f(t_i))\right], \tag{39}$$

  where $\overline{\lambda}$ is a constant.

Define $\mathbf{y} = (y_1, \ldots, y_n)'$ and $\mathbf{f} = (f(t_1), \ldots, f(t_n))$. We will show that, if $\lambda_m = \overline{\lambda}/(2\sigma_w^2)$, as $\kappa \to \infty$ the mode of the smoothing distribution $p(\mathbf{f}|\mathbf{y})$ converges to the point $(f(t_1), \ldots, f(t_n))$ obtained by evaluating the solution of the problem (37) at $(t_1, \ldots, t_n)$.

Remark 12 *Quantile regression can be obtained as a special case in which* $\rho_\tau(x) = (\tau - I(x < 0))x$ *and the distribution of* $y_i|f(t_i)$ *is asymmetric double exponential with* $\lambda = \overline{\lambda}$. *The quantile regression with smoothing splines problem described by Bosch et al. (1995) is a special case with* $m = 2$. *Similarly, expectile regression corresponds to the assumption* $\rho_\omega(x) = |\omega - I(x < 0)| x^2$, *which results in an asymmetric Gaussian distribution for the observations conditional on the signal; here* $\overline{\lambda} = \sigma^2$ *in (6). The* M*-quantile regression as described in Breckling and Chambers (1988) corresponds to setting* $\rho_\tau(x)$ *proportional to*$|\tau - I(x < 0)| \rho(x)$, *where* $\rho(x)$ *is such that* $d\rho(x)/dx = \psi(x)$ *and* $\psi(x)$ *is the* influence function.

Remark 13 *If the density in the measurement equation were Gaussian our argument would provide an alternative proof of the result of Wahba (1978) for the special cases* $m = 1, 2$. *This follows on noting that in a Gaussian model conditional means and conditional modes coincide. Wahba's proof requires the explicit solution of the spline smoothing problem (derived in Kimeldorf and Wahba (1971)), which is shown to be equal to the conditional mean. Our proof simply shows that the two optimisation problems, i.e. finding the mode and finding the optimal spline, are equivalent.*

Remark 14 *A sufficient condition for existence and uniqueness of the solution to problem (37) is convexity of* $\rho_\tau(x)$. *This follows immediately from the fact that if* ρ_τ *is convex then the log-likelihood of the time series representation is strictly concave in* $\mathbb{R}^n$.

Proof

The mode of $p(\mathbf{f}|\mathbf{y})$ is found by solving $\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{y})$.This is equivalent to solving $\max_{\mathbf{f}} p(\mathbf{y},\mathbf{f})$ and we proceed by first noting that

$$p(\mathbf{y},\mathbf{f}) = p\left(\mathbf{y}|\mathbf{f}\right)p\left(\mathbf{f}\right). \tag{40}$$

Consider the joint distribution $p\left(\mathbf{f}\right)$. It is a multivariate normal distribution with mean zero (because $f(0), f'(0), \ldots$ have zero mean) and covariance matrix $\sigma^2\mathbf{W}_n + \kappa\mathbf{T}\mathbf{T}'$, where

$$\mathbf{T}' = \begin{bmatrix} 1 & \ldots & 1 \\ t_1 & \ldots & t_n \\ \vdots & & \vdots \\ t_1^{m-1}/\left(m-1\right)! & \ldots & t_n^{m-1}/\left(m-1\right)! \end{bmatrix}$$

and

$$\mathbf{W}_n = \mathrm{Cov}\left[\left(\int_0^{t_1}\frac{\left(t_1-s\right)^{m-1}}{\left(m-1\right)!}dW_s,\ldots,\int_0^{t_n}\frac{\left(t_n-s\right)^{m-1}}{\left(m-1\right)!}dW_s\right)'\right].$$

Wahba (1978) shows that

$$\lim_{\kappa\to\infty}\left(\sigma_w^2\mathbf{W}_n+\kappa\mathbf{T}\mathbf{T}'\right)^{-1} = \sigma^{-2}\mathbf{W}_n^{-1}\left(\mathbf{I}-\mathbf{T}\left(\mathbf{T}'\mathbf{W}_n^{-1}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{W}_n^{-1}\right).$$

As a result

$$\lim_{\kappa\to\infty}p\left(\mathbf{f}\right)\propto\exp\left(-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{W}_n^{-1}\mathbf{f}+\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{W}_n^{-1}\mathbf{T}\left(\mathbf{T}'\mathbf{W}_n^{-1}\mathbf{T}\right)^{-1}\mathbf{T}'\mathbf{W}_n^{-1}\mathbf{f}\right).$$

Define

$$\mathbf{C} = \begin{bmatrix} \mathbf{T}'\mathbf{W}_n^{-1} \\ \mathbf{U}' \end{bmatrix}$$

where $\mathbf{U}'$ is a $(n-m)\times n$ matrix whose rows are orthogonal to the rows of $\mathbf{T}'$. The first term in the exponent can be written

$$\begin{aligned} -\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{W}_n^{-1}\mathbf{f} &= -\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{C}'\mathbf{C}^{-1\prime}\mathbf{W}_n^{-1}\mathbf{C}^{-1}\mathbf{C}\mathbf{f} \\ &= -\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{C}'\left(\mathbf{C}\mathbf{W}_n\mathbf{C}'\right)^{-1}\mathbf{C}\mathbf{f} \\ &= -\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{C}'\left(\begin{bmatrix} \mathbf{T}'\mathbf{W}_n^{-1}\mathbf{T} & \mathbf{T}'\mathbf{U} \\ \mathbf{U}'\mathbf{T} & \mathbf{U}'\mathbf{W}_n\mathbf{U} \end{bmatrix}\right)^{-1}\mathbf{C}\mathbf{f} \\ &= -\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{W}_n^{-1}\mathbf{T}\left(\mathbf{T}'\mathbf{W}_n^{-1}T\right)^{-1}\mathbf{T}'\mathbf{W}_n^{-1}\mathbf{f}-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f}. \end{aligned}$$

Note that $\mathbf{U}'\mathbf{T} = \mathbf{0}$ by construction and since $\mathbf{C}\mathbf{W}_n^{-1}\mathbf{C}'$ is block diagonal, so is its inverse. As a result, the density of $\mathbf{f}$ becomes

$$\lim_{\kappa\to\infty} p(\mathbf{f}) \propto \exp\left(-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f}\right).$$

From (40) and (39) we have

$$\lim_{\kappa\to\infty} p(\mathbf{f}|\mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f} - \frac{1}{\omega}\sum_{i=1}^{n}\rho_\tau\left(y_i - f(t_i)\right)\right).$$

For $m = 1$ we can set

$$\begin{aligned} \mathbf{U}' &= \begin{bmatrix} \mathbf{u}_1' \\ \mathbf{u}_2' \\ \vdots \\ \mathbf{u}_{n-1}' \end{bmatrix}, \quad \mathbf{u}_i' = \left(\underbrace{0,\ldots,0}_{\text{i-1 zeros}}, \alpha_0^i, \alpha_1^i, \underbrace{0,\ldots,0}_{\text{n-1-i zeros}} \right), \\ \alpha_0^i &= \frac{1}{t_{i+1}-t_i} \\ \alpha_1^i &= -\frac{1}{t_{i+1}-t_i}. \end{aligned}$$

It is easy to show that in this case $\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}$ is a diagonal matrix with entries $t_2 - t_1, t_3 - t_2, \ldots, t_n - t_{n-1}$. Thus

$$-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f} = -\frac{1}{2\sigma_w^2}\sum_{i=2}^{n}(t_i - t_{i-1})\left(\frac{f(t_i) - f(t_{i-1})}{t_i - t_{i-1}}\right)^2.$$

Well known results on spline interpolation (summarized, for example, in Schoenberg (1964)) imply that the solution to (37), $f(t)$, is a piecewise linear function with knots at $t_1, \ldots, t_n$. Thus we obtain

$$-\frac{1}{2\sigma_w^2}\mathbf{f}'\mathbf{U}\left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1}\mathbf{U}'\mathbf{f} = -\frac{1}{2\sigma_w^2}\int_0^b \left[f'(t)\right]^2 \mathrm{d}t.$$

If we set $\lambda_1 = \overline{\lambda}/2\sigma_w^2$ the maximisation problem is equivalent to minimising (37) with respect to $\mathbf{f}$ for $m = 1$.

The proof for $m = 2$ proceeds along the same lines. Here we set

$$\begin{aligned}
\mathbf{U}' &= \begin{bmatrix} \mathbf{u}'_1 \\ \mathbf{u}'_2 \\ \vdots \\ \mathbf{u}'_{n-2} \end{bmatrix}, \quad \mathbf{u}'_i = \Bigg(\underbrace{0, \ldots, 0}_{\text{i-1 zeros}}, \alpha^i_0, \alpha^i_1, \alpha^i_2, \underbrace{0, \ldots, 0}_{\text{n-2-i zeros}} \Bigg), \\
\alpha^i_0 &= \frac{1}{t_{i+1} - t_i} \\
\alpha^i_1 &= -\frac{1}{t_{i+2} - t_{i+1}} - \frac{1}{t_{i+1} - t_i} \\
\alpha^i_2 &= \frac{1}{t_{i+2} - t_{i+1}}.
\end{aligned}$$

It can be shown that $\mathbf{W}_n$ has entries

$$[\mathbf{W}_n]_{ij} = \frac{1}{3} \left[\min(t_i, t_j)\right]^3 + \frac{1}{2} \left|t_i - t_j\right| \left[\max(t_i, t_j)\right]^2.$$

Bosch *et al.* (1995) showed that

$$\lambda_m \int_0^b \left[f''(t)\right]^2 \mathrm{d}t = \lambda \mathbf{f}'\mathbf{U} \left(\mathbf{U}'\mathbf{W}_n\mathbf{U}\right)^{-1} \mathbf{U}'\mathbf{f}$$

where $f(t)$ is the solution to problem (37), a cubic spline with knots at $t_1, \ldots, t_n$. Thus finding the mode is equivalent to minimising (37) with respect to $\mathbf{f}$ for $m = 2$ and $\lambda_2 = \overline{\lambda}/2\sigma^2_w$.

B Calculations for Laplace distribution

If the observations are assumed to come from a double exponential (Laplace) distribution with mean zero, $p(y) = (1/4)\exp(-|y|/2)$. Since

$$\frac{1}{4}\int_{-\infty}^{\xi(\tau)} y \exp(y/2) dy = \frac{1}{2}(\xi(\tau) - 2)\exp(-|\xi(\tau)|/2), \qquad \xi(\tau) \leq 0$$

$$\omega = \frac{\frac{1}{2}(\xi(\tau) - 2)\exp(-|\xi(\tau)|/2) + \tau\xi(\tau)}{(\xi(\tau) - 2)\exp(-|\xi(\tau)|/2) + (2\tau - 1)\xi(\tau)} \tag{41}$$

Furthermore

$$\frac{1}{4}\int_{-\infty}^{\xi(\tau)} \exp(y/2)dy = \frac{1}{2}\exp(\xi(\tau)/2) = \tau, \qquad \xi(\tau) \leq 0$$

and so $\xi(\tau) = 2\ln(2\tau), \tau \leq 0.5$. Substituting in (41) gives (12).

Acknowledgements Early versions of this paper were presented at the Econometrics in Rio conference in July 2006 and at the 4th Oxmetrics Users Conference, CASS Business School, London in September, 2006. We would like to thank David Hendry, James Davidson, Siem Jan Koopman and Roger Koenker for helpful comments. We would like to thank the ESRC for financial support under the grant Time-Varying Quantiles, RES-062-23-0129.

References

Bosch, R.J., Ye, Y., Woodworth, G.G., 1995. A convergent algorithm for quantile regression with smoothing splines. Computational Statistics and Data Analysis 19, 613-630.

Breckling, J., Chambers, R., 1988. M-quantiles. Biometrika 75, 761-771.

Busetti, F., Harvey, A.C., 2006. Tests of time-invariance. Mimeo.

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 Rossi, G., Harvey, A.C., 2006. Time-varying quantiles. Faculty of Economics, Cambridge, CWPE 0649.

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

Efron, B., 1991. Regression percentiles using asymmetric squared error loss. Statistica Sinica 1, 93-125.

Engle, R.F., Manganelli, S., 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics 22, 367–381.

Granger, C.W.J., Sin C-Y., 2000. Modelling the absolute returns of different stock indices: exploring the forecastability of an alternative measure of risk. Journal of Forecasting 19, 277-298.

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

Harvey, A.C., Koopman, S.J., 2000. Signal extraction and the formulation of unobserved components models. Econometrics Journal 3, 84-107.

Huber, P.J., 1981. Robust statistics. Wiley, New York.

Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: existence, persistence and co-movements. Journal of Economic Dynamics and Control, 27, 1699-1737.

Kimeldorf, G., Wahba, G., 1971. Some results on Tchebycheffian spline functions. Journal of Mathematical Analysis and Applications 33, 82–95.

Koenker, R., 2005. Quantile regression. Cambridge University Press, Cambridge.

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

Kohn, R., Ansley, C.F., Wong, C-H., 1992. Nonparametric spline regression with autoregressive moving average errors. Biometrika 79, 335-346.

Koopman, S.J., Shephard, N., Doornik, J., 1999. Statistical algorithms for models in state space using SsfPack 2.2. Econometrics Journal 2, 113-166.

Newey, W.K., Powell, J.L., 1987. Asymmetric least squares estimation and testing. Econometrica 55, 819-847.

Schoenberg, I.J., 1964. Spline functions and the problem of graduation. Proceedings of the National Academy of Science 52, 947-950.

Silverman, B.W., 1985. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. Journal of the Royal Statistical Society B 47, 1-52.

Stock, J.H., Watson, M.W., 2005. Has inflation become harder to forecast? Mimeo. www.wws.princeton.edu/mwatson/wp.html

Stuart, A., Ord, J.K., 1987. Kendall's Advanced Theory of Statistics. Volume 1. Charles Griffin and Co., London.

Wahba, G., 1978. Improper priors, spline smoothing and the problem of guarding against model errors in regression. Journal of the Royal Statistical Society B 40 (3), 364–372.

Whittle, P., 1983. Prediction and Regulation. 2nd ed. Blackwell, Oxford.