Quantiles, Expectiles and Splines

Andrew Harvey

University of Cambridge

December 2007

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.

Time-varying quantiles can be fitted to a sequence of observations by setting up a state space model and iteratively applying a suitably modified signal extraction algorithm.

Satisfy the defining property of fixed quantiles in having the appropriate number of observations above and below.

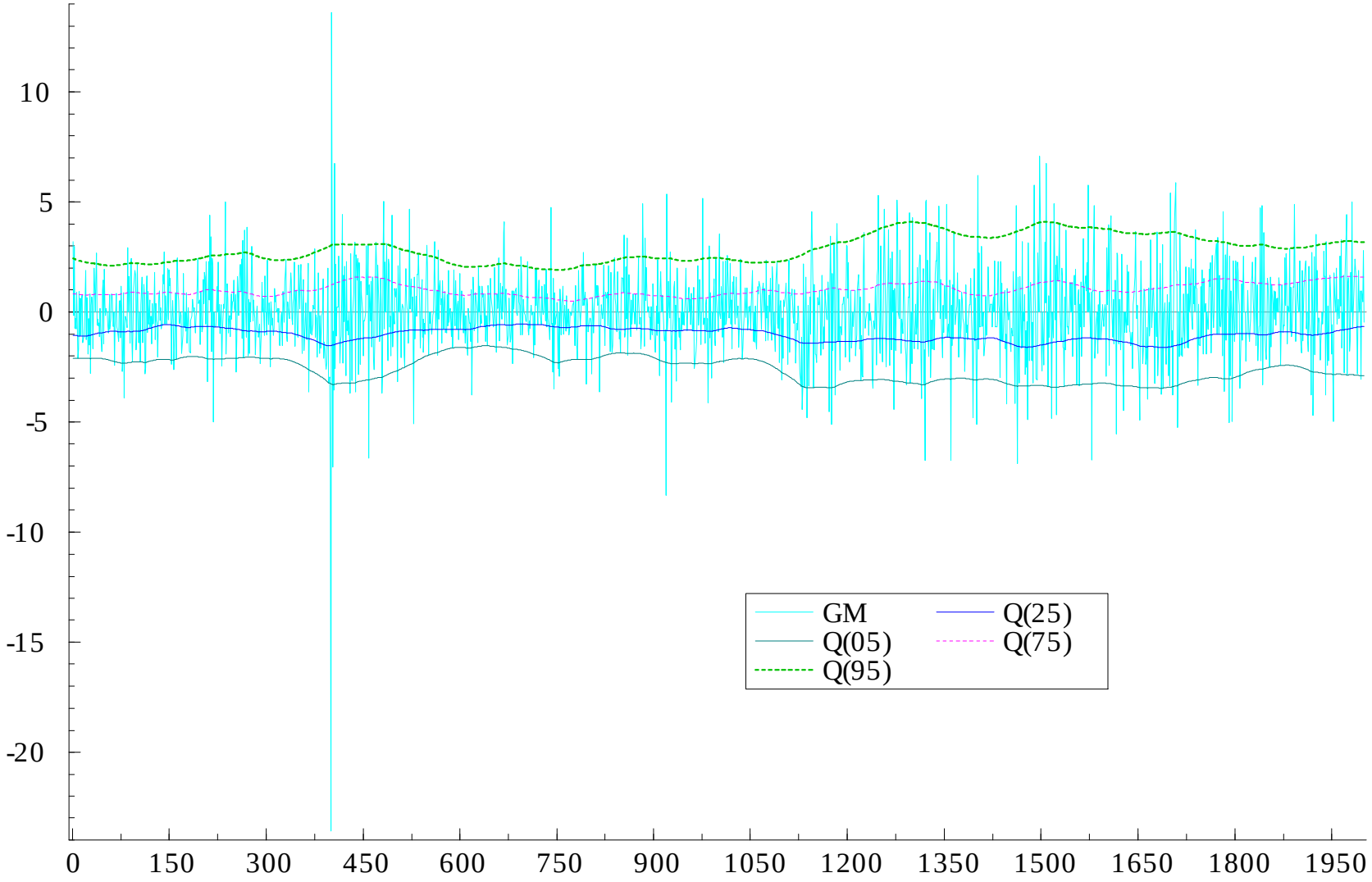


Figure: Quantiles fitted to GM returns

Expectiles are similar to quantiles except that they are defined by tail expectations; see Newey and Powell (1987). Time-varying expectiles can be estimated by a state space algorithm. This is similar to the algorithm used for quantiles, but estimation is more straightforward and much quicker.

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, $y_t, t = 1, .., T$, the sample quantile, $\widetilde{\xi}(\tau)$, can be obtained 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*.

Expectiles, denoted $\mu(\omega), 0 < \omega < 1$, 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 the asymmetric least squares function,

$$S_\omega = \sum \rho_\omega(y_t - \mu(\omega)) = \sum |\omega - I(y_t - \mu(\omega) < 0)| \, (y_t - \mu(\omega))^2, \quad (3)$$

with respect to $\mu(\omega)$. Differentiating S_ω and dividing by minus two gives

$$\sum_{t=1}^{T} |\omega - I(y_t - \mu(\omega) < 0)| \;\; (y_t - \mu\,(\omega)) \quad (4)$$

The sample expectile, $\widetilde{\mu}(\omega)$, is the value of $\mu(\omega)$ that makes (4) 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.

Signal extraction

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

$$y_t = \xi_t(\tau) + \varepsilon_t(\tau), \qquad t = 1, ..., T, \tag{5}$$

where $\Pr(\varepsilon_t(\tau) < 0) = \tau$ with $0 < \tau < 1$. The disturbances, $\varepsilon_t(\tau)$, are assumed to be serially independent and independent of $\xi_t(\tau)$. The problem is then one of signal extraction. The assumption that the quantile or expectile follows a stochastic process can be regarded as a device for inducing local weighting of the observations. One possibility is a random walk,

$$\xi_t(\tau) = \xi_{t-1}(\tau) + \eta_t(\tau), \qquad \eta_t(\tau) \backsim IID(0, \sigma^2_{\eta(\tau)}), \tag{6}$$

A smoother quantile can be extracted by a local linear trend

$$\begin{aligned} \xi_t &= \xi_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned} \tag{7}$$

where β_t is the slope and ζ_t is $IID(0, \sigma_\zeta^2)$. It is well known that in a Gaussian model setting $Var(\eta_t) = \sigma_\zeta^2/3$ and $Cov(\eta_t, \zeta_t) = \sigma_\zeta^2/2$ results in the smoothed estimates being a cubic spline.
The model for expectiles is set up in a similar way with (5) replaced by $y_t = \mu_t(\omega) + \varepsilon_t(\omega)$.

Derivation and properties

The proposed solution minimizes $\sum_t \rho_\tau(y_t - \xi_t)$ subject to a set of constraints imposed by the time series model for the quantile. Suppose this model is a first-order autoregression,. The criterion function is then

$$J = -\sum_{t=1}^{T} \frac{\rho_\tau(y_t - \xi_t)}{\omega} - \frac{1}{2}\frac{(1-\phi^2)(\xi_1 - \xi^\dagger)^2}{\sigma_\eta^2} - \frac{1}{2}\sum_{t=2}^{T}\frac{\eta_t^2}{\sigma_\eta^2}. \qquad (8)$$

where ω is a scaling constant. Given the observations, the estimated time-varying quantiles, $\widetilde{\xi}_1, .., \widetilde{\xi}_T$, are the values that maximise J.

In a classical signal extraction framework,

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

where μ_t is a Gaussian stochastic process and ε_t is $NID(0, \sigma^2)$. If J is redefined with μ_t in place of $\tilde{\zeta}_t$ and $\rho_\tau(y_t - \tilde{\zeta}_t)/\omega$ replaced by $(y_t - \mu_t)^2/2\sigma^2$, it can be interpreted as the logarithm of the joint density of the observations and $\mu_t's$. Differentiating 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.

Returning to the quantiles and differentiating J with respect to ξ_t gives

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

for $t = 2, \ldots, T-1$, (modified at the endpoints), where $IQ\left(y_t - \xi_t\right)$ is defined as in (24). 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 $IQ\left(y_t - \xi_t\right)$ is not defined as the check function is not differentiable at zero.

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) of μ_t, L is the lag operator, and $g_y = g + \sigma_\varepsilon^2$. The WK formula has the attraction that, for simple models, g_y can be written in terms of the reduced form and g/g_y can be expanded to give an explicit expression for the weights. Here the reduced form of y_t is an $ARMA(1,1)$ model and

$$\widetilde{\mu}_t = \mu + \frac{q_\mu \theta}{\phi(1-\theta^2)} \sum_{j=-\infty}^{\infty} \theta^{|j|}(y_{t+j} - \mu) \tag{11}$$

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 $IQ_t's$ as the values that give equality of the associated derivative of J. Then we can set $\partial J/\partial \tilde{\xi}_t$ in (10) equal to zero to give

$$\frac{\tilde{\xi}_t - \xi^\dagger}{g} = \frac{1}{\omega} IQ(y_t - \tilde{\xi}_t) \tag{12}$$

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

$$\tilde{\xi}_t = \xi^\dagger + \frac{\sigma_\eta^2}{\omega} \sum_{j=-\infty}^{\infty} \frac{\phi^{|j|}}{1-\phi^2} IQ(y_{t+j} - \tilde{\xi}_{t+j}) \tag{13}$$

Note that if the observations are multiplied by a constant, then the quantile is multiplied by the same constant, as are ω and σ_η^2. By defining the quasi 'signal-noise' ratio as $q = \sigma_\eta^2/\omega^2$, it remains scale invariant.

An expression for extracting quantiles that has a similar form to (11) can be obtained by adding $(\widetilde{\xi}_t - \xi^\dagger)/\omega^2$ to both sides of (12) and re-arranging to give

$$\widetilde{\xi}_t = \xi^\dagger + \frac{g}{g_y^\omega}\left[\widetilde{\xi}_t - \xi^\dagger + \omega IQ(y_t - \widetilde{\xi}_t)\right] \tag{14}$$

where $g_y^\omega = g + \omega^2$.This is not an ACGF but it can be treated as though it were. Thus we obtain

$$\widetilde{\xi}_t = \xi^\dagger + \frac{q_\tau \theta}{\phi(1-\theta^2)} \sum_{j=-\infty}^{\infty} \theta^{|j|}[\widetilde{\xi}_{t+j} - \xi^\dagger + \omega IQ(y_{t+j} - \widetilde{\xi}_{t+j})] \tag{15}$$

where θ is as defined for (11) but with q_μ replaced by $q = \sigma_\eta^2/\omega^2$. For the random walk, the weights sum to one, so $\xi^\dagger$ drops out giving

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

Non-parametric smoothing

Changing quantiles may be estimated non-parametrically, as in Yu and Jones (1998), by minimising a local check function to give an estimator, $\widehat{\xi}_t$, that satisfies

$$\sum_{j=-h}^{h} K(j/h) IQ(y_{t+j} - \widehat{\xi}_t) = 0 \tag{17}$$

where $K(.)$ is a weighting kernel, h is a bandwidth.with $IQ(y_{t+j} - \widehat{\xi}_t)$ defined appropriately if $y_{t+j} = \widehat{\xi}_t$.

Proposition

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

It is interesting to compare the above weighting scheme with the one implied by the random walk model. If $\widetilde{\xi}_{t+j}$ in (16) is constant, it satisfies (17) with $K(j/h)$ replaced by $\theta^{|j|}$ so giving an (infinite) exponential decay. The time series model determines the shape of the kernel while the q parameter plays the same role as the bandwidth.

An important advantage of the more model-based approach for forecasting is that it automatically determines a weighting pattern at the end of the sample that is consistent with the one in the middle.

State Space Form

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

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\alpha}_t + \varepsilon_t, \qquad Var(\varepsilon_t) = \sigma_t^2, \qquad t = 1, ..., T \\ \boldsymbol{\alpha}_t &= \mathbf{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t, \qquad Var(\boldsymbol{\eta}_t) = \mathbf{Q}_t \end{aligned} \tag{18}$$

where $\boldsymbol{\alpha}_t$ is an $m \times 1$ state vector, $\mathbf{z}_t$ is a non-stochastic $m \times 1$ vector, σ_t^2 is a non-negative 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 serially independent disturbances ε_t and $\boldsymbol{\eta}_t$ are independent of each other and of the initial state.

Consider the criterion function

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

where $\rho(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t)$ is as in (2) or (3), with $\mathbf{z}_t'\boldsymbol{\alpha}_t$ equal to $\xi_t(\tau)$ or $\mu_t(\omega)$, $\mathbf{Q}_t$ and $\mathbf{P}_{1|0}$ are are assumed positive definite matrices as in (18) and h_t is a non-stochastic sequence of postive scalars. Suppose that the initial state and the $\boldsymbol{\eta}_t's$ are normally distributed. For a Gaussian model of the form (18), logarithm of the joint density of the observations and the states is, ignoring irrelevant terms, given by J with $\rho(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t) = (y_t - \mu_t(0.5))^2$ and $h_t = 2\sigma_t^2$. Differentiating J with respect to to each element of $\boldsymbol{\alpha}_t$ gives a set of equations, which, when set to zero and solved gives the minimum mean square error estimates of $\boldsymbol{\alpha}_t$. These 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, the last term in J disappears. An algorithm is available as a subroutine in the SsfPack set of programs within Ox; see Koopman *et al* (1999).

We can think of (19) as a criterion function that provides the basis for computing a quantile or expectile subject to a set of constraints imposed by the time series model for the quantile or expectile. For expectiles differentiating J gives

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

where

$$IE(y_t - \mu_t(\omega)) = |\omega - I(y_t - \mu_t(\omega) < 0)|\ (y_t - \mu_t(\omega)), \qquad t = 1, \ldots, T. \tag{21}$$

The smoothed estimates, $\widetilde{\boldsymbol{\alpha}}_t$, satisfy the equations obtained by setting these derivatives equal to zero.

Let $h_t = g_t/\kappa$, where κ is a constant, the interpretation of which will become apparent. For any expectile, adding and subtracting $\mathbf{z}_t g_t^{-1}\mathbf{z}_t'\boldsymbol{\alpha}_t$ to the equations in (20) allows the first term to be written as

$$\mathbf{z}_t g_t^{-1}\left[\mathbf{z}_t'\boldsymbol{\alpha}_t + 2\kappa IE\left(y_t - \mathbf{z}_t'\boldsymbol{\alpha}_t\right)\right] - \mathbf{z}_t g_t^{-1}\mathbf{z}_t'\boldsymbol{\alpha}_t, \qquad t = 1, \ldots, T. \tag{22}$$

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

$$\widehat{y}_t^{(i-1)} = \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(i-1)} + 2\kappa IE\left(y_t - \mathbf{z}_t'\widehat{\boldsymbol{\alpha}}_t^{(i-1)}\right). \tag{23}$$

The iterations are carried out until the $\widehat{\boldsymbol{\alpha}}_t^{(i)\prime}s$ converge whereupon $\widetilde{\mu}_t(\omega) = \mathbf{z}_t'\widetilde{\boldsymbol{\alpha}}_t$.

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

$$IQ(y_t - \xi_t(\tau)) = \begin{cases} \tau - 1, & if \quad y_t < \xi_t(\tau) \\ \tau, & if \quad y_t > \xi_t(\tau) \end{cases} \qquad t = 1, ..., T, \tag{24}$$

and the synthetic observations in the KFS are

$$\widehat{y}_t^{(j-1)} = \mathbf{z}_t' \widehat{\boldsymbol{\alpha}}_t^{(j-1)} + \kappa IQ\left(y_t - \mathbf{z}_t' \widehat{\boldsymbol{\alpha}}_t^{(j-1)}\right), \qquad t = 1, ..., T \tag{25}$$

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).

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.

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 $\mathbf{j}_2' = \sum_{t=1}^{T} \mathbf{A}_t \boldsymbol{\alpha}_t$, where the $\mathbf{A}_t' s$ are $m \times m$ matrices.

Lemma *For a model in SSF with a diffuse prior on the initial state, a sufficient condition for the first element in the vector* $\mathbf{j}_2'$ *to be zero is that the first column of* $\mathbf{T}_t - \mathbf{I}$ *consists of zeroes for all* $t = 2, ..., T$.

If h_t is time-invariant and the conditions of the Lemma hold, the estimated quantiles satisfy the fundamental property of sample time-varying quantiles, namely 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* $[T(1-\tau)]$.

Proposition

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

$$\omega = \frac{\int_{-\infty}^{\xi(\tau)} (y - \xi(\tau)) dF(y)}{\int_{-\infty}^{\xi(\tau)} (y - \xi(\tau)) dF(y) - \int_{\xi(\tau)}^{\infty} (y - \xi(\tau)) dF(y)}$$

where $F(y)$ is the cdf of y. 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 $\widetilde{\mu}_t(\widetilde{\tau})$. 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.

Parameter estimation

The smoothing algorithms depend on parameters that can be estimated by cross validation. For time-varying quantiles, the function to be minimized is

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

where $\tilde{\xi}_t^{(-t)}$ is the smoothed value at time t when y_t is dropped; see De Rossi and Harvey (2006). A similar criterion, $CV(\omega)$, may be used for expectiles.

In a time invariant model with quantiles or expectiles following a random walk, $\mathbf{Q}_t$ is a scalar equal to $\sigma_\eta^2(\tau)$ or $\sigma_\eta^2(\omega)$. We would like a suitable parameterization in terms of a quasi signal-noise ratio that is scale invariant. For the mean it will be recalled that $h_t = 2\sigma_t^2$ and so in a time invariant model the usual signal-noise ratio, σ_η^2/σ^2, implies that $g_t = \sigma^2$ and $\kappa = 0.5$ in (22). A similar normalization can be applied for other expectiles so the quasi signal-noise ratios are $q_\omega = \sigma_\eta^2(\omega)/\sigma^2$. Hence the iterations are based on (22) with g_t set to one and $\kappa = 0.5$.

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

$$\min \sum_{t=1}^{T} \rho_\tau\{y_t - \xi(x_t)\} + \lambda_2 \left(\int \{\xi''(x)\}^2 dx \right) \tag{27}$$

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 the same cubic spline is obtained by quantile signal extraction.

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. This generalisation allows the handling of nonparametric quantile and expectile regression by cubic splines when there is only one explanatory variable. The observations, which may be from a cross-section, are arranged so that the values of the explanatory variable are in ascending order.

Bosch, Ye and Woodworth (1995) propose a solution to cubic spline quantile regression that uses quadratic programming. Unfortunately this necessitates the repeated inversion of large matrices of dimension up to $4T \times 4T$. This is very time consuming. Our signal extraction appears to be much faster (and more general) and makes estimation of the smoothing parameter (quasi signal-noise ratio) a feasible proposition.
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.

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 observations are irregularly spaced and at some time points there are multiple observations. Harvey and Koopman (2000) highlight the stochastic trend connection. Figure 2 shows the cubic spline expectiles obtained using the value of $\sigma_\zeta^2/\sigma^2 = 0.07$ computed by CV for the mean. The graph gives a clear visual impression of the movements in level and dispersion. 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.

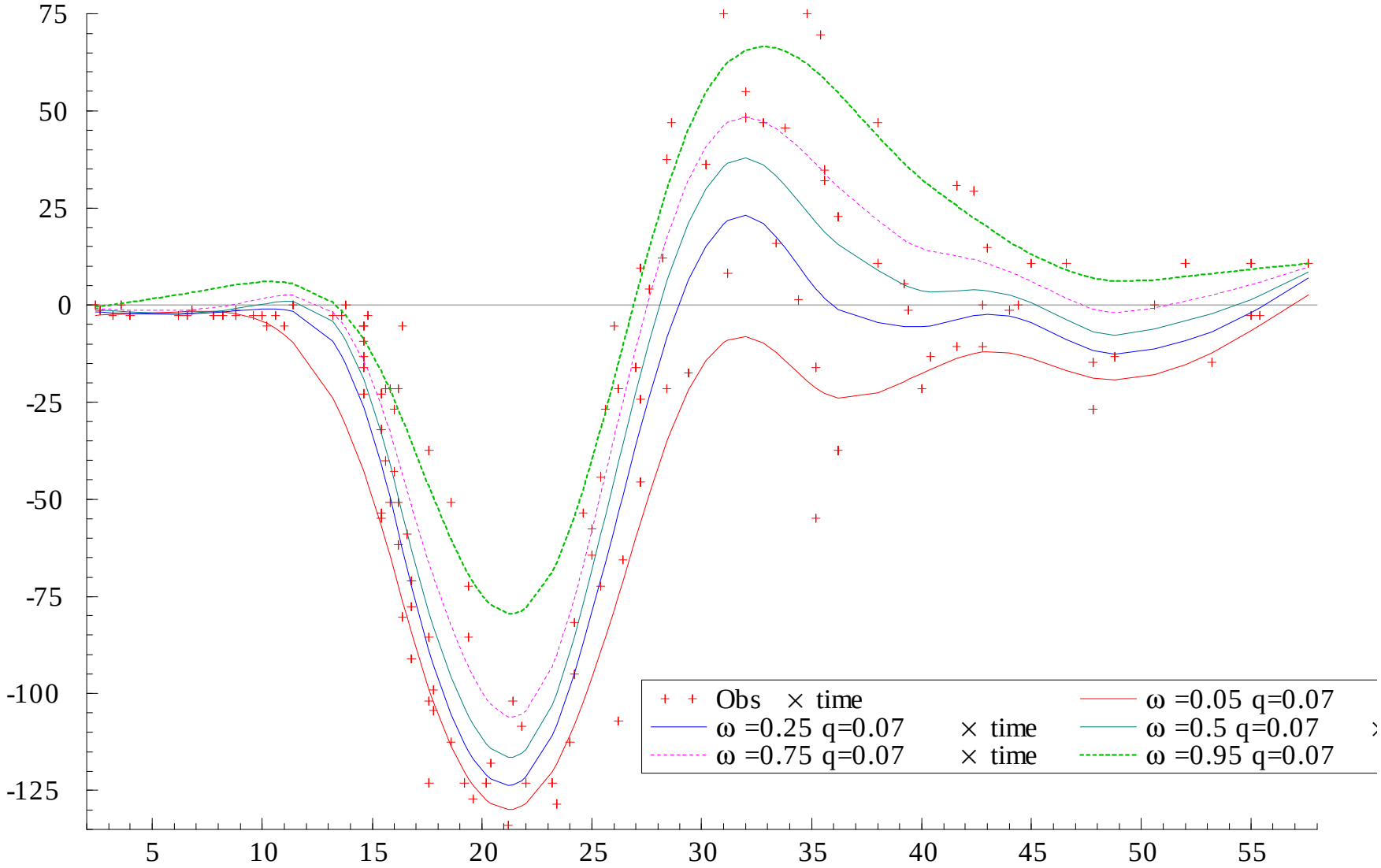


Figure: Cubic spline expectiles fitted to the motorcycle data. The parameter q_μ is estimated by cross validation.

Conclusions (so far)

Time-varying quantiles and expectiles are **GOOD THINGS** and provide information on various aspects of a time series, such as dispersion and asymmetry.

Time-varying quantiles can be fitted iteratively applying a suitably modified state space signal extraction algorithm. The algorithm for time-varying expectiles is much faster as there is no need to take account of corner solutions. 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.

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.

Equivalent to fitting a cubic spline. 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

But estimating the signal – noise ratio for quantiles takes a long time

The dual: tracking the distribution

A different, but complementary approach, is to pre-assign a value $\overline{\xi}$ and then construct a binary series, $I(y_t \leq \overline{\xi})$. At any point in time

$$E(I(y_t \leq \overline{\xi}) = \tau_t, \qquad t = 1, ..., T \tag{28}$$

If the quantiles are fixed and known, setting $\overline{\xi} = \xi(\tau)$ means that $E(I_t) = \tau$. If the distribution changes, a discount parameter, ω, leads to

$$\widetilde{\tau}_{t+1|t} = a_{t+1|t} / \left(a_{t+1|t} + b_{t+1|t}\right), \qquad t = 1, ..., T,$$

where

$$a_{t+1|t} = \omega a_{t|t-1} + \omega I_t, \qquad t = 1, ..., T \tag{29}$$

$$b_{t+1|t} = \omega b_{t|t-1} + \omega(1 - I_t) \tag{30}$$

with $a_{1|0} = b_{1|0} = 1$. Note that
$\widetilde{\tau}_{t+1|t} = \Pr\left(I_{t+1}(\tau) = 1 | I_j(\tau), j = 1, ..., t\right) = \widetilde{\tau}_{t|t}$.

The log-likelihood function is

$$\log L(\omega) = \sum_{t=1}^{T} \{ I_t \ln \widetilde{\tau}_{t|t-1} + (1 - I_t(\tau)) \ln(1 - \widetilde{\tau}_{t|t-1}) \}, \qquad (31)$$

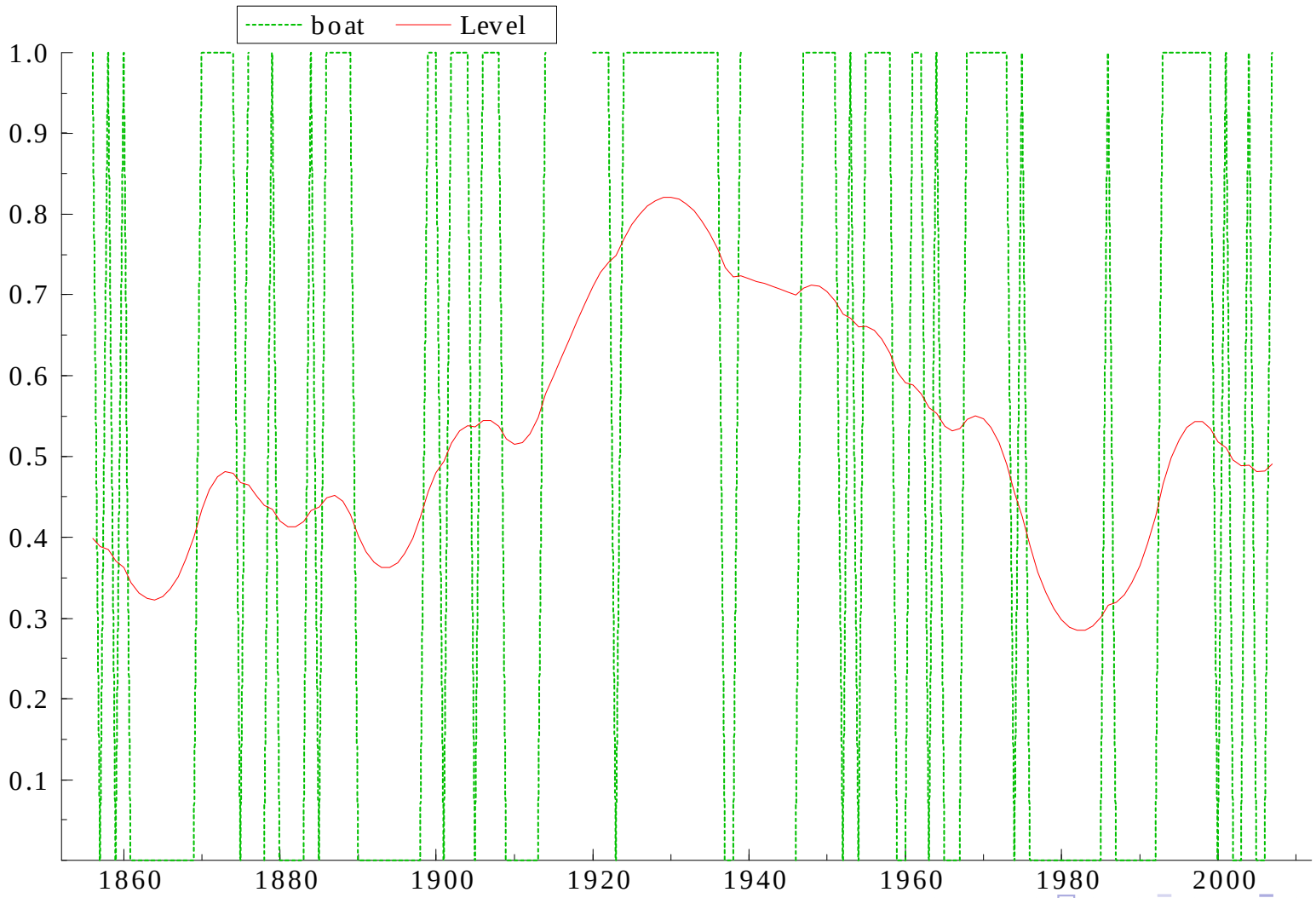
where $\widetilde{\tau}_{1|0} = 1/2$.
The filtered estimates are an EWMA in the indicators, that is

$$\widetilde{\tau}_{t+1|t} = \widetilde{\tau}_{t|t} = \frac{\sum_{j=0}^{t-1} \omega^j I_{t-j} + \omega^t}{\sum_{j=0}^{t-1} \omega^j + 2\omega^t} \simeq (1-\omega) \sum_{j=0}^{t-1} \omega^j I_{t-j}$$

with the terms in ω^t not present if $a_{1|0} = b_{1|0} = 0$. It is reasonable to construct a two-sided smoothed estimator of the same form. This may be done using the smoother for a Gaussian random walk plus noise with $q = (1-\omega)^2/\omega$. In the middle of a large sample

$$\widetilde{\tau}_{t|T} = \frac{1-\omega}{1+\omega} \sum_j \omega^j I_{t-j} \qquad (32)$$

This is also the same problem as forecasting the sign of a variable, or more generally $y_t - \xi$, where ξ is some pre-assigned value. Indeed it can be used for any binary series eg $I_t = 1$ if Cambridge win the boat race. Here $\omega = 0.877$ (and $q = 0.017$)



By forming a whole grid a changing distribution function can be estimated and tracked. There are N categories defined by the boundaries $-\infty, \xi_1, ..., \xi_{N-1}, \infty$. Let $I_{t,j} = 1$ if $y_t \leq \xi_j$, $j = 1, ..., N-1$ and zero otherwise, ie $I(y_t \leq \xi_j)$, and define $I_{N,t} = 1$ and $I_{0,t} = 0$. Then

$$a_{j,t+1|t} = \omega_j a_{j,t|t-1} + \omega_j I_{j,t}, \qquad j = 1, ..., N, \quad t = 1, ..., T \tag{33}$$

with $a_{j,1|0} = 1, j = 1, ..., N$, and $\widetilde{\tau}_{j,t+1|t} = \widetilde{\tau}_{j,t|t} = \Sigma_{i=1}^{j} a_{i,t+1|t} / \Sigma_{j=1}^{N} a_{j,t+1|t}$. The $\widetilde{\tau}'_{j,t|t}s$ lie in the range [0,1] by construction and, of course, $\widetilde{\tau}_{N,t|t} = 1$. When the $\omega's$ are the same, the estimated series of probabilities, $\widetilde{\tau}_{j,t|t}$, cannot cross.

The Dirichlet-multinomial or Polya predictive distribution leads to the log-likelihood function

$$\ln L(\omega_1, .., \omega_N) = \sum_{t=1}^{T} \ln \ell_t$$

where

$$\ln \ell_t = -\ln \Gamma(\Sigma_{j=1}^{N} a_{j,t|t-1}) - \Sigma_{j=1}^{N} \ln\{\Gamma(l_j - l_{j-1} + a_{j,t|t-1})/\Gamma(a_{j,t|t-1})\}, \tag{34}$$

The binary likelihood, (31), is obtained when $N = 2$.
Figure 4 shows a graph of estimates of the discount factors for GM with $\xi_1, ..., \xi_{N-1}$ set to the 5%, 10%,..., 95% quantiles from the raw (empirical) distribution. As expected there is very little discounting at and around the centre of the distribution. Quadratic fitted by OLS. This suggests that a general approach for estimating $\omega's$ as a slowly changing function of τ is to maximize the likelihood, (34), wrt a, b and c in the quadratic $\omega(\tau) = a + b + c\tau^2$. May set $\omega = 1$ and/or impose symmetry.

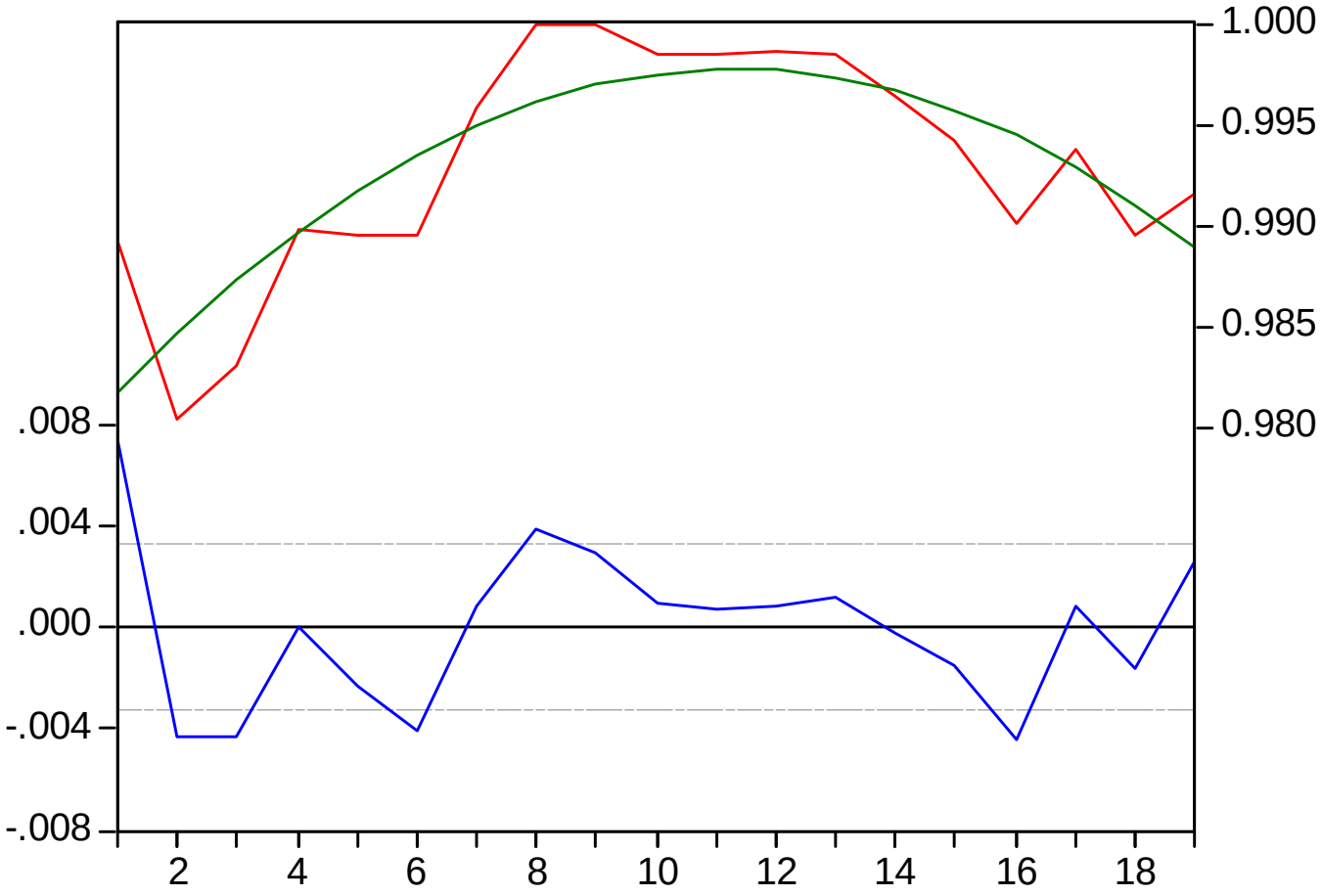


Figure: GM - quadratic fitted to estimated discount factors

Time-varying quantiles may be extracted from the $\widetilde{\tau}'_{j,t|t}s$ as follows. Suppose an estimate of $\xi_t(\tau)$ is required, and that $\widetilde{\tau}_{k-1,t|t} \leq \tau \leq \widetilde{\tau}_{k,t|t}$ for some $k = 1, .., N$. Linear interpolation then yields

$$\widehat{\xi}_t(\tau) = \frac{\tau - \widetilde{\tau}_{k-1,t|t}}{\widetilde{\tau}_{k,t|t} - \widetilde{\tau}_{k-1,t|t}}\{\xi_k - \xi_{k-1}\} + \xi_{k-1}, \qquad t = 1, ..., T$$

What is the relationship between these estimates and those obtained directly ? If $\widetilde{\xi}_{t+j}(\tau) = \xi(\tau)$ is constant, then $IQ(y_{t+j} - \xi(\tau)) = \tau - I_{t+j}(y_t \leq \xi(\tau))$, and so the condition $\sum_{j=-\infty}^{\infty} \theta^{|j|} IQ(y_{t+j} - \xi(\tau)) = 0$ implied by fitting the time-varying quantile implies in turn that $\sum_{j=-\infty}^{\infty} \theta^{|j|} I_{t+j}(y_t \leq \xi(\tau)) = \tau$. A comparison with (32) shows that $\omega = \theta$.

So tracking the distribution may be better - but really only viable when there is no trend in mean or variance.

********** THE END **********