Semiparametric Modeling of Multiple Quantiles

Leopoldo Catania[a], Alessandra Luati[b]

[a] *Aarhus BSS and CREATES*
[b] *University of Bologna*

Abstract

We develop a semiparametric model to track a large number of quantiles of a time series. The model satisfies the condition of non crossing quantiles and the defining property of fixed quantiles. A key feature of the specification is that the updating scheme for time varying quantiles at each probability level is based on the gradient of the check loss function, that forms a martingale difference sequence. Theoretical properties of the proposed model are derived, such as weak stationarity of the quantile process and consistency and asymptotic normality of the estimators of the fixed parameters. The model can be applied for filtering and prediction. We also illustrate a number of possible applications such as: i) semiparametric estimation of dynamic moments of the observables, *ii*) density prediction, and *iii*) quantile predictions.

Keywords: Dynamic quantiles, Score Driven models, Risk Management

1. Introduction

Modeling quantiles has traditionally been of much interest in econometrics. Since the seminal contribution of Koenker and Bassett (1978), quantile regression has been successfully employed to study, for example, the changes in the US wage structure, as in Buchinsky et al. (1994), the household electricity demand in the Chicago metropolitan area, see Hendricks and Koenker (1992), and the public decision making regarding hazardous waste cleanup, by Viscusi and Hamilton (1999). In a time series context, the original quantile regression needs to be modified to account for the dependence induced by the ordering of the observations (time). Quantile autoregression models can be then estimated, as in Koenker and Xiao (2006). One of the most successful quantile autoregression model has undoubtedly been the CAViaR specification by Engle and Manganelli (2004). The acronym CAViaR stands for Conditional Autoregressive Value at Risk, where "Value at Risk" is the name given to an extreme left quantile (usually associated with probability levels 1% or 5%) of financial returns, used as a measure of risk in financial econometrics, see Jorion (1996). A quantile autoregressive model delivers a sequence of filtered quantiles at a pre-specified probability confidence level, from which predictions are usually computed. Differently from quantile regression, where the quantile of the distribution of a random variable conditional on a set of exogenous explanatory variables is

Preprint submitted to *November 28, 2019*

studied, in quantile autoregression the interest lies in the quantile of the distribution of Y_t conditional on $\mathcal{F}_{t-1}$, say $p(Y_t|\mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1}$ incorporates all the past history of Y_t, available up to time $t-1$. Since no distributional assumption on $p(Y_t|\mathcal{F}_{t-1})$ is made, but only on the dynamic evolution of the quantile of interest, then quantile autoregressions are semiparametric models. The specification of the dynamic evolution of the quantile of interest is thus very important, as it characterizes the whole model. For instance, Engle and Manganelli (2004) assume that the quantile at time t depends on transformations of past realizations of Y_t such as $|y_{t-s}|^d$, for $d = 1, 2, 3, \dots$ and $s > 1$. As an example, if q_t^τ denotes the quantile of the distribution of Y_t conditional on $\mathcal{F}_{t-1}$ at probability level τ, the CAViaR specification defines $q_t^\tau = f(q_{t-1}^\tau, |y_{t-1}|, y_{t-1}^2, |y_{t-1}|^3)$ for an $\mathcal{F}_{t-1}$–measurable function $f(\cdot)$ to which we refer to as "the filter". Note that the choice of which transformation of past observations is to be included in $f(\cdot)$ is evidently a personal choice of the modeler and may be selected ad hoc for the application at hand.

Beside the formulation of the filter and the choice of the forcing variables, a second main issue of quantile autoregression models is that they are mainly designed for estimating one single quantile. When turning the attention to joint modelling of multiple quantiles (i.e. associated with different probability levels), the well-known quantile crossing problem arises, especially when predictions are considered. The multivariate multiple quantile VAR for VaR models by White et al. (2015) satisfy the non crossing condition only asymptotically and under the assumption of a correctly specified model. A common solution is to rearrange the estimated quantiles based on monotonization methods as in Dette and Volgushev (2008), Chernozhukov et al. (2009) and Chernozhukov et al. (2010) or by constrained non linear optimisation methods as in Bondell et al. (2010). Multiple quantile models that ensure monotonicity by construction are the dynamic additive quantile model by Gouriéroux and Jasiak (2008), where quantile curves are modelled as mixtures of baseline quantile functions, and the quantile regression by Schmidt and Zhu (2016), which is not developed in the times series context. The relevance of estimating well a group of quantiles, with the aim of recovering the underlying time varying distribution of the data, is emphasised by Granger (2010) and earlier discussed in Granger and Sin (2000).

In this paper, we develop a semiparametric model to track a large number of quantiles of a time series. Differently from available quantile autoregressive specifications, the model satisfies the condition of non crossing quantiles by construction, and not as a by-product of constrained nonlinear optimization procedures. The model also satisfies the defining property of fixed quantiles, which means that the limiting distribution ensures that the empirical frequency of observations below the unconditional τ–level quantile is τ. A key feature of our specification is that the updating scheme for time varying quantiles at each probability level is based on the gradient of the check loss function, that forms a martingale difference sequence. The check loss function used in quantile estimation is the negative likelihood function of a density related to the family of asymmetric Laplace distributions, see Poiraud-Casanova and Thomas-Agnan (2000) and Kotz et al. (2001), or, more generally, to the family of tick exponential functions introduced by Komunjer (2005). Without a specific distribution assumption on $Y_t|\mathcal{F}_{t-1}$, the check loss function plays the role of a quasi-likelihood. Hence, our updating scheme provides a quasi-score driven model or a generalisation of the score driven approach to a wider set of loss functions. Score driven models, developed by Creal et al. (2013) and Harvey (2013), are observation driven models where the dynamics of time varying parameters depend on the score of the conditional likelihood function of the parameter of interest. This class of models strongly relies on the distributional assumption made for the observables, which restricts their application to a parametric framework. One of the first attempts to go beyond the parametric framework of score driven models is the recent model by Patton et al. (2019) for joint filtering of VaR and Expected Shortfall. However, this approach might not be optimal for the filtering of extreme quantiles due to the little amount of information contained in the score in such cases. Our methodology overcomes this limitation by exploiting the information coming from other regions of the distribution to update extreme quantiles.

The novelty of the paper does not only lie on the gradient-based update of the quantiles dynamics, but also in a specification that ensures that the set of estimated quantiles do not cross. Based upon an idea of Granger (2010), that suggested to model differences, rather than, directly, quantiles, we specify the dynamics of a reference quantile $q_t^{\tau*}$ (typically, the median, $\tau^* = 0.5$) and then model the differences $q_t^{\tau_j} - q_t^{\tau_{j-1}}$ or $q_t^{\tau_j} - q_t^{\tau_{j+1}}$ as a positive or negative process, according to the sign of $\tau_j - \tau^*$ being positive or negative. With this respect, the paper is related with the work of Schmidt and Zhu (2016), who refer to their method as "quantile spacings". The dynamics of the reference quantile, as well as those of the increments, are driven by the gradient of the multiple check loss function. The paper is also related to the contribution by White et al. (2015) in that we share the same quasi maximum likelihood estimation method. Besides, the very general multi-quantile specification discussed in their paper nests our model for the median as a particular case. However, neither the score driven dynamics nor the non crossing condition are dealt with in the work of White et al. (2015).

Once specified the model, we derive its theoretical properties, such as: weak stationarity of the quantile process, limiting quantile values, consistency and asymptotic normality of the estimators of the fixed parameters. Along with asymptotic results, we also investigate the finite sample properties of the proposed estimators. The model can be applied for filtering and prediction of quantiles of a time series. We report an empirical illustration employing the time series of financial returns of Microsoft corporations. Besides filtering of many quantiles, we also detail how conditional moments can be recovered. A forecasting analysis illustrates the performance of the model in an out of sample context. The model is proven to outperform competing parametric and semiparametric alternatives.

The paper is structured in the following manner. Section 2 details the Dynamic Multiple Quantile model and Section 2.1 reports a general discussion on the forcing variable for updating the quantiles in the time series context. Section 3 details the estimation procedure as well as the consistency and asymptotic normality of the estimator. Finite sample properties are studied in Section 4. An empirical illustration is reported in Section 5. Conclusions and directions for future research are in Section 6.

2. The dynamic multiple quantile model

Let $Y = \{Y_t\}_{t\in T}$ be a stationary stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$ where $\mathcal{F} = \{\mathcal{F}_t\}_{t\in T}$ and $\mathcal{F}_t = \sigma(Y_{t-s}, s \geq 0)$ is the sigma-algebra generated by the random variables Y_s, $s \leq t$. The process Y is adapted to the filtration $\mathcal{F}$ and $\mathrm{E}(|Y_t|) < \infty$ for all $t \in T$. Let $F_{t|t-1}(y_t) = P(Y_t \leq y_t|\mathcal{F}_{t-1})$ be the cumulative distribution function of Y_t given $\mathcal{F}_{t-1}$ and, for a fixed j, let $\tau_j \in (0,1)$ be a probability level such that $P(Y_t \leq q_t^{\tau_j}|\mathcal{F}_{t-1}) = \tau_j$ where $|q_t^{\tau_j}| < \infty$ is the quantile level associated with τ_j at time t,

$$q_t^{\tau_j} = \inf\{F_{t|t-1}(y_t) \geq \tau_j\}$$

and, if $F_{t|t-1}(\cdot)$ is strictly increasing,

$$q_t^{\tau_j} = F_{t|t-1}^{-1}(\tau_j). \tag{1}$$

As a matter of fact,

$$\frac{\partial}{\partial q_t^{\tau_j}}\mathbb{E}\left[\rho_{\tau_j}(Y_t - q_t^{\tau_j})|\mathcal{F}_{t-1}\right] = 0, \quad \forall t \in T \iff F_{t|t-1}(q_t^{\tau_j}) = \tau_j, \tag{2}$$

where $\rho_\tau(x) = x(\tau - \mathbb{1}(x < 0))$ is the quantile check function, and $\mathbb{1}(\cdot)$ is the indicator function. If one considers the sequence of ordered probability levels $\tau = (\tau_1, \ldots, \tau_J)$, $\tau_i > \tau_j$ if $i > j$, and the sequence of associated quantiles at time t, $q_t = (q_t^{\tau_1}, \ldots, q_t^{\tau_J})'$, then (2) can be generalized to the multiple case:

$$\frac{\partial}{\partial q_t}\mathbb{E}\left[\sum_{j=1}^{J}\rho_{\tau_j}(Y_t - q_t^{\tau_j})|\mathcal{F}_{t-1}\right] = 0, \quad \forall t \in T \iff F_{t|t-1}(q_t^{\tau_j}) = \tau_j, \quad \forall j = 1, \ldots, J. \tag{3}$$

The sample analogue of (3) can then be used to build a filter for the vector of time–varying quantiles q_t. Specifically, the update $q_t \to q_{t+1}$, after observing y_t, can be driven by:

$$\frac{\partial}{\partial q_t}\sum_{j=1}^{J}\rho_{\tau_j}(y_t - q_t^{\tau_j}) = z_t, \tag{4}$$

that is $q_{t+1} = f(q_t, z_t)$, where $z_t = (z_{i,t}, i = 1, \ldots, J)'$ and

$$z_{i,t} = \mathbb{1}(y_t \leq q_t^{\tau_i}) - \tau_i \tag{5}$$

is the hit variable at time t for quantile $q_t^{\tau_i}$. Note that the quantile check function $\rho(y_t - q_t)$ is not differentiable in zero, i.e. when $y_t = q_t$, which holds with zero probability when Y_t is a continuous random variable. We next specify the filter $f(q_t, z_t)$ and introduce the model.

The Dynamic Multiple Quantile (DMQ) model is defined as follows,

$$q_t^{\tau_j} = \begin{cases} q_t^{\tau_{j+1}} - \eta_{j,t}, & \text{if } \tau_j < \tau_{j^*} \\ q_t^{\tau_{j^*}}, & \text{if } \tau_j = \tau_{j^*} \\ q_t^{\tau_{j-1}} + \eta_{j,t}, & \text{if } \tau_j > \tau_{j^*} \end{cases} \tag{6}$$

where

$$q_t^{\tau_{j^*}} = \bar{q}^{\tau_{j^*}}(1-\beta) + \alpha u_{t-1}^{\tau_{j^*}} + \beta q_{t-1}^{\tau_{j^*}}, \tag{7}$$

is the reference quantile, according to which the other quantiles are defined,

$$\eta_{j,t} = \exp(\xi_{j,t})$$

is a positive stochastic process,

$$\xi_{j,t} = \bar{\xi}_j(1-\phi) + \gamma u_{t-1}^{\tau_j} + \phi \xi_{j,t-1}, \tag{8}$$

and

$$u_t^{\tau_j} \propto \frac{\partial}{\partial q_t} \sum_{j=1}^{J} \rho_{\tau_j}(y_t - q_t^{\tau_j})$$

is the the martingale difference sequence which drives the dynamics of the time varying quantiles; finally, $\theta = (\alpha, \beta, \phi, \gamma)'$ are static parameters to be estimated, with $|\beta| < 1$, $|\phi| < 1$; $\bar{q}^{\tau_{j^*}}$ and $\bar{\xi}_j$ are defined such that $E[q_t^{\tau_{j^*}}] = \bar{q}^{\tau_{j^*}}$ and $E[\xi_{j,t}] = \bar{\xi}_j$, i.e. they determine the unconditional levels of $q^{\tau_{j^*}}$ and $\xi_{j,t}$, respectively, see Section 2.2.1.

The definition of the forcing variables $u_t^{\tau_j}$ is crucial for the specification of the model and depends on its parameterization. In the score driven framework, the forcing variable is set proportional to the score of the likelihood of the observables conditional on the past, see Creal et al. (2013) and Harvey (2013). In this paper, we adopt a similar approach and set $u_t^{\tau_j}$ as the derivative of the negative sample analogue of (3) with respect to $q_t^{\tau_{j^*}}$ or $\xi_{j,t}$, normalised by a positive scale constant a_j, such as it has unit variance. In the case $j \neq j^*$, according to (6), (8) and (7) the model has been reparameterized from $f(q_t, z_t)$ to $f((q_t^{\tau_{j^*}}, \xi_t)), z_t)$, where $\xi_t = (\xi_{j,t}, j = 1, \ldots, J, \quad j \neq j^*)$, this quantity is proportional to:

$$\frac{\partial}{\partial \xi_{j,t}} \sum_{j=1}^{J} \rho_{\tau_j}(y_t - q_t^{\tau_j}) \propto a_j^{-1} \left(\frac{\partial q_t}{\partial \xi_{j,t}} \right)' z_t,$$

the same reasoning applies to the case $j = j^*$ which yelds:

$$u_t^{\tau_j} = \begin{cases} b_j a_j^{-1} \sum_{i=1}^{j} z_{i,t}, & \text{if } j < j^* \\ a_j^{-1} \sum_{i=1}^{J} z_{i,t}, & \text{if } j = j^* \\ b_j a_j^{-1} \sum_{i=j}^{J} z_{i,t}, & \text{if } j > j^* \end{cases}, \tag{9}$$

where $b_j = \mathbb{1}(\tau_j < \tau_{j^*}) - \mathbb{1}(\tau_j > \tau_{j^*})$. Note that the sequence of hit variables $\{z_t\}_{t\in T}$ is independent and identically distributed (iid) since $\mathbb{E}_{t-1}[\mathbb{1}(y_t < q_t^{\tau_j})] = F_{t|t-1}(q_t^{\tau_j}) = \tau_j$, see also Christoffersen (1998). As a consequence $\{u_t^{\tau_j}\}_{t\in T}$ is also an iid sequence of zero mean random variables with constant variance

$$E_{t-1}[(u_t^{\tau_j})^2] = \varpi_j / a_j^2,$$

where

$$\varpi_j = \begin{cases} \sum_{i=1}^{j} \tau_i(1-\tau_i) + \sum_{l_1=j}^{J} \sum_{l_2 \neq l_1} m(\tau_{l_1}, \tau_{l_2}), & \text{if } j < j^* \\ \sum_{i=1}^{J} \tau_i(1-\tau_i) + \sum_{l_1=j}^{J} \sum_{l_2 \neq l_1} m(\tau_{l_1}, \tau_{l_2}), & \text{if } j = j^* \\ \sum_{i=j}^{J} \tau_i(1-\tau_i) + \sum_{l_1=j}^{J} \sum_{l_2 \neq l_1} m(\tau_{l_1}, \tau_{l_2}), & \text{if } j > j^* \end{cases},$$

with $m(a.b) = \min(a, b)(1 - \max(a, b))$. Thus, $E_{t-1}[(u_t^{\tau_j})^2] = 1$ is achieved by setting $a_j = \sqrt{\varpi_j}$.[1] For the rest of the paper we will employ this normalizing mechanism.

2.1. The shape of the forcing variables

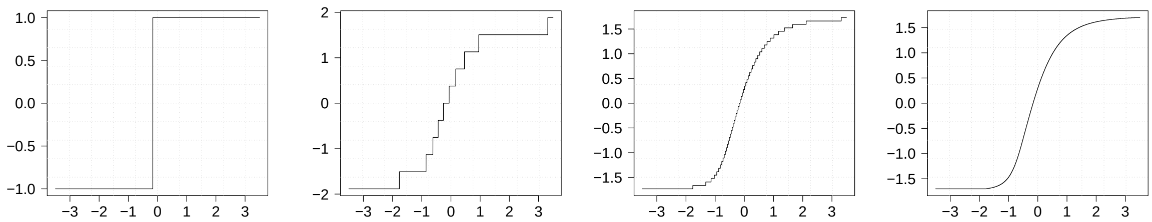
In a parametric framework, score driven filters provide updates for the time–varying parameters which are consistent with the shape of the conditional distribution of the data. This is also true in our case, where a discretization of the conditional cumulative density function (cdf) of $Y_t|\mathcal{F}_{t-1}$ is used. To see this, consider the driving force of the reference quantile, set to the median, i.e. $u_t^{\tau_{j^*}}$, with $\tau_{j^*} = 0.5$. This quantity is proportional to

$$u_t^{\tau_{j^*}} \propto \sum_{j=1}^{J} \mathbb{1}(y_t < q_t^{\tau_j}) \propto \widehat{F}_{t|t-1}^{J}(y_t),$$

where $\widehat{F}_{t|t-1}^{J}(y_t)$ is the discretized cdf computed using J quantiles. It follows that the number J of chosen quantiles plays an important role in the DMQ model, because it determines the strength of the signal delivered by the score driven type filter. In the case when $J = 1$, the forcing variable is proportional to the step function, $u_t^{\tau_1} \propto \mathbb{1}(y_t < q_t^{0.5}) - 0.5$, and we obtain a model which is similar to the specifications detailed in De Rossi and Harvey (2006) and Patton et al. (2019). As a matter of fact, in such case, the signal provided by the score driven filter is very weak and, as shown in our empirical application (Section 5), performs in an unsatisfactory way with real data. As long as J increases, more structure is stored in the model and the amount of information used in the forcing variable increases accordingly.

Figure 1 reports the values of the forcing variable $u_t^{0.5}$ for different choices of J, where quantiles have been computed according to a standardized skew Student's t distribution with skewness parameter equal to 1.5 and shape parameter equal to 3.5. The forcing variable for other quantiles is proportional to

[1]Other choices such as $a_j = \varpi_j$ and $a_j = 1$ are plausible and report good results. In principle, one can set $a_j = (\varpi_j)^g$ and estimate g in order to find the optimal way of scaling the signal using the g–power of its variance.



(a) $J = 1$ (b) $J = 10$ (c) $J = 50$ (d) $J = 1000$

Figure 1: Values of $u_t^{0.5}$ for different values of J against quantiles of a skew Student t distribution

$$u_t^{\tau_j} \propto \begin{cases} \mathbb{1}(q_t^{\tau_j} > y_t)\widehat{F}_{t|t-1}^J(y_t), & \text{if } j < j^* \\ \mathbb{1}(q_t^{\tau_j} < y_t)\widehat{F}_{t|t-1}^J(y_t), & \text{if } j > j^*, \end{cases} \tag{10}$$

which is the discretization of a truncated cdf updating the quantile differences in one direction only.

2.2. Statistical properties

Prediction

The one-step-ahead prediction of all quantiles $E_t[q_{t+1}] = E[q_{t+1}|\mathcal{F}_t]$ is immediately available from the filter since q_{t+1} is $\mathcal{F}_t$–measurable.

The multi-step prediction for the reference quantile, $\widehat{q}_{t+h|t}^{\tau_{j^*}} = E_t[q_{t+h}^{\tau_{j^*}}]$ for $h > 1$, is given by

$$\widehat{q}_{t+h|t}^{\tau_{j^*}} = \bar{q}^{\tau_{j^*}}(1-\beta)\sum_{s=0}^{h-2}\beta^s + \beta^{h-1}q_{t+1}^{\tau_{j^*}}$$

and the predictive variance of $\widehat{q}_{t+h|t}^{\tau_{j^*}}$ is given by

$$E_t[(q_{t+h}^{\tau_{j^*}} - \widehat{q}_{t+h|t}^{\tau_{j^*}})^2] = \alpha^2\sum_{s=0}^{h-2}\beta^{2s}.$$

For $h \to \infty$ we recover the unconditional reference quantile

$$\lim_{h\to\infty} E_t[q_{t+h}^{\tau_{j^*}}] = \bar{q}^{\tau_{j^*}}$$

with variance

$$\lim_{h\to\infty} E_t[(q_{t+h}^{\tau_{j^*}} - E_t[q_{t+h}^{\tau_{j^*}}])^2] = \frac{\alpha^2}{1-\beta^2}.$$

The multistep ahead prediction for other quantiles is given by

$$\widehat{q}^{\tau_j}_{t+h|t} = \begin{cases} \widehat{q}^{\tau_{j+1}}_{t+h|t} - E_t[\eta_{j,t+h}], & \text{if } \tau_j < \tau_{j^*} \\ \widehat{q}^{\tau_{j-1}}_{t+h|t} + E_t[\eta_{j,t+h}] & \text{if } \tau_j > \tau_{j^*} \end{cases}$$

where

$$E_t[\eta_{j,t+h}] = \begin{cases} \omega_{j,t+h} \prod_{s=0}^{h-2} \exp\{-a_j^{-1}\gamma\phi^s \sum_{l=1}^{j} \tau_l\} \sum_{l=0}^{j} h(\tau_l) \exp\{a_j^{-1}\gamma\phi^s(j-l)\}, & \text{if } \tau_j < \tau_{j^*} \\ \omega_{j,t+h} \prod_{s=0}^{h-2} \exp\{a_j^{-1}\gamma\phi^s \sum_{l=j}^{J} \tau_l\} \sum_{l=j-1}^{J} g(\tau_l) \exp\{-a_j^{-1}\gamma\phi^s(J-l)\}, & \text{if } \tau_j > \tau_{j^*} \end{cases} \tag{11}$$

with $\omega_{j,t+h} = \exp\{\xi_j(1-\phi)\sum_{s=0}^{h-2}\phi^s\}\exp\{\phi^{h-2}\xi_{j,t+1}\}$ and

$$h(\tau_l) = \begin{cases} \tau_1, & \text{if } l = 0 \\ \tau_{l+1} - \tau_l, & \text{if } 0 < l < j \\ 1 - \tau_j, & \text{if } l = j \end{cases}, \quad g(\tau_l) = \begin{cases} \tau_j, & \text{if } l = j-1 \\ \tau_{j+1} - \tau_j, & \text{if } j-1 < l < J \\ 1 - \tau_J, & \text{if } l = J. \end{cases}$$

The closed form expression of the unconditional moments of $\eta_{j,t+h}$ is rather complicate to derive. However, we are able to prove that the limiting quantile differences have bounded moments, i.e.

$$\lim_{h\to\infty} E_t[\eta_{j,t+h}] < \infty$$

and so are their higher order moments

$$\lim_{h\to\infty} E_t[\eta^m_{j,t+h}] < \infty.$$

The above results are obtained by direct calculations, that we outline here in the following. First observe that, for $j < j^*$, $\eta_{j,t+h} = \omega_{j,t+h} \ \exp\{\gamma\sum_{s=0}^{h-1}\phi^s u_{j,t+(h-1)+s}\} = \omega_{j,t+h} \prod_{s=0}^{h-1}\exp\{\gamma\phi^{h-1-s}u_{j,t+s}\}$, i.e. $\eta_{j,t+h}$ is a transformation of sums of j independent variables $z_{j,t+s}$, see equation (9). Taking the expectation conditional to $\mathcal{F}_t$ then gives the first line of equation (14). An analogous reasoning, applied to the case $j > j^*$, holds for the second line.

The proof of boundedness of the first unconditional moment of $\eta_{j,t+h}$ is carried out by induction on j. For $j = 1$, $j < j^*$, the limit of equation (14) reduces to $\lim_{h\to\infty} E_t[\eta_{1,t+h}] \propto \prod_{s=0}^{\infty}\exp\{-\gamma\phi^{h-1-s}\tau_1\}(\exp\{\gamma\phi^{h-1-s}\}\tau_1 + (1-\tau_1)) < \infty$ if $\sum_{s=0}^{\infty}\log((1-\tau_1)+\tau_1\exp\{\gamma\phi^{h-1-s}\}) < \infty$. One has that $\sum_{s=0}^{\infty}\log((1-\tau_1)+\tau_1\exp\{\gamma\phi^{h-1-s}\}) \leq \sum_{s=0}^{\infty}|\log((1-\tau_1)+\tau_1\exp\{\gamma\phi^{h-1-s}\})| \leq \sum_{s=0}^{\infty}|\log\exp\{\gamma\phi^{h-1-s}\}| < \infty$, as follows for $|\phi| < 1$ from the limit comparison test. For $j > 1$, repeating the same steps leads to expressions where the term of interest is dominated by $\sum_{s=0}^{\infty}|\log\exp\{j\gamma\phi^{h-1-s}\}|$, which is a convergent series. As $\eta_{j,t+h}$ is the product of exponential functions, the proof of boundedness of its unconditional moments of order $m > 1$ is identical to the proof for $m = 1$.

2.2.1. Quantile targeting

Quantile targeting can be used to select the intercept parameters $\bar{\xi}_j$ and $\bar{q}_t^{\tau_j^*}$ in order to target a reference distribution. Specifically, we can estimate the unconditional quantiles $\widehat{q}^{\tau_j}$ from the time series $(y_1, \dots, y_T)'$, and set $\bar{q}^{\tau_{j^*}} = \widehat{q}^{\tau_{j^*}}$ and:

$$\bar{\xi}_j = \log(\widehat{\Delta}_j) + \frac{\tau_j b_j \gamma}{1-\phi} + \kappa_j$$

where

$$\widehat{\Delta}_j = \begin{cases} \widehat{q}^{\tau_j} - \widehat{q}^{\tau_{j-1}}, & \text{if } \tau_j > \tau_{j^*} \\ \widehat{q}^{\tau_{j+1}} - \widehat{q}^{\tau_j}, & \text{if } \tau_j < \tau_{j^*}, \end{cases}$$

and

$$\kappa_j = \begin{cases} \sum_{s=0}^{\infty} \log\left(\sum_{l=0}^{j} g(\tau_l) \exp\{-a_j^{-1}\gamma\phi^s(J-l)\}\right), & \text{if } \tau_j > \tau_{j^*} \\ \sum_{s=0}^{\infty} \log\left(\sum_{l=0}^{j} h(\tau_l) \exp\{a_j^{-1}\gamma\phi^s(j-l)\}\right), & \text{if } \tau_j < \tau_{j^*}. \end{cases}$$

3. Estimation and inference

Let $\theta = (\alpha, \beta, \phi, \gamma)'$ and $\{y_t\}_{t=1,\dots,T}$ be an observed time series. Parameter estimates are obtained by minimizing the function,

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min} \sum_{t=1}^{T} \sum_{j=1}^{J} \rho_{\tau_j}(y_t - q_t^{\tau_j}(\theta)). \tag{12}$$

where $\Theta \subseteq \mathbb{R}^d$, $d = 4$. We may refer to the multiple check loss function $\sum_{j=1}^{J} \rho_{\tau_j}(y_t - q_t^{\tau_j}(\theta))$ as to the Hogg function, mentioned in Koenker (2005), section 5.5, referring to a private correspondence between Robert Hogg and himself in 1979. We now establish consistency (Theorem 1) and asymptotic normality (Theorem 2) of $\hat{\theta}$, for a correctly specified model where the true parameter $\theta_0 = (\alpha_0, \beta_0, \phi_0, \gamma_0)'$ belongs to the parameter space Θ and it is such that $q_t^{\tau_j}(\theta_0)$ is equal to $q_t^{\tau_j}$ in equation (1).

Theorem 1. *Let (i) $Y = \{Y_t\}_{t \in T}$ be a stationary and ergodic stochastic process on the complete probability space $(\Omega, \mathcal{F}, P)$ and (ii) $q_t^{\tau_j}(\theta_0)$ be specified as in equations (6), (7), (8) and (9), with $\theta_0 \in \Theta$, a compact subset of $\mathbb{R}^d$. (iii) Let $\hat{\theta}$ be the estimator of θ_0 defined in equation (12). (iv) Let $E|Y_t| < \infty$ and $|\alpha| < \infty, |\gamma| < \infty, |\beta| < 1, |\phi| < 1$. (v) Let the conditional density of Y_t, $f_{t|t-1}$, be continuous and bounded away from zero. For every $\upsilon > 0$, there exists $\delta_\upsilon > 0$ such that $\forall \theta \in \Theta$ with $\|\theta - \theta_0\| > \upsilon$, $P\left(\cup_{j=1,\dots,J} |q_t^{\tau_j}(\theta) - q_t^{\tau_j}(\theta_0)| > \delta_\upsilon\right) > 0$. Then $\hat{\theta} \to_{a.s.} \theta_0$.*

The proof is in the Appendix and consists in verifying that the assumptions of Corollary 5.11 of White (1994) are satisfied. The corollary establishes the strong consistency of quasi maximum likelihood estimators (QMLE) in correctly specified models. Indeed, the (negative Hogg) function $\varphi(y_t, q_t^{\tau_j}(\theta)) = -\sum_{j=1}^{J} \rho_{\tau_j}(y_t - q_t^{\tau_j}(\theta))$ is the log-likelihood of J independent variables whose density is related to the family of asymmetric Laplace distributions, see Poiraud-Casanova and Thomas-Agnan (2000) and Kotz et al. (2001), or, more generally, to the family of tick exponential functions introduced by Komunjer (2005), as is also recognised by White et al. (2015). The estimator $\hat{\theta}$ in equation (12) can therefore be interpreted as a QMLE. As detailed in the Appendix, the five assumptions for Corollary 5.11 in White (1994) to be satisfied consist of three standard regularity conditions on (i) the process, (ii) the model and (iii) the objective function, plus a uniform law of large number, which requires a dominance condition implied by (iv) and a condition for unique identifiability (v) that follows from Powell (1984). The proof is actually very similar to the proof of consistency of the VAR for VaR estimator by White et al. (2015), as the log-likelihood that we maximise is the same for the two models, which on the other hand differ in the specification of the quantiles dynamics.

As far as asymptotic normality is concerned, the proof is non standard, in that neither the likelihood function nor the filter satisfy the regularity conditions required for the usual first order Taylor expansions to apply. Indeed, the likelihood function is continuous in θ but a.e. differentiable and the forcing variable $u_t^{\tau_j}$, is a.e. continuous in θ. Almost everywhere differentiability of the objective function has been dealt with in several contributions, including Engle and Manganelli (2004), Komunjer (2005), White et al. (2015), Patton et al. (2019). All these papers rely on Powell (1984) and Weiss (1991) extension to the time series setting of Huber (1967) results for least absolute deviation estimators for iid data, see also Newey and McFadden (1994). The basic idea is to approximate the a.e. differentiable loss function with its smooth expectation. Strong consistency of the estimator is usually required. We apply this method as well. However, all the above mentioned papers (including Patton et al. (2019)) are grounded on the assumption that the estimator is a twice continuously differentiable function of the parameter, which is clearly not the case of our filter. Specifically, the derivative $\partial q_t^{\tau_j}(\theta)/\partial\theta_i$, $i = 1, \ldots, d$ is not defined when $y_t = q_t^{\tau_j}(\theta)$, although both the left and right derivatives exist. One could resort, as in Chan and Tsay (1998) in the context of threshold models, to write symbolically the gradient which results by adding the left and right derivatives in their domain of existence. Alternatively, as in Patton et al. (2019), one can write symbolically the gradient of the a.s. twice differentiable function whose points of discontinuity of $q_t^{\tau_j}(\theta)$ are of zero P-measure for all $j = 1, \ldots, J$. We prefer to adopt a solution which belongs to a different stream of methods for dealing with the non smoothness of the objective function of an estimator, considered in Amemiya (1982), Pollard (1985) and Phillips (1991), following Daniels (1961). The idea there consists in approximating the discontinuous function of interest with a continuous one (or with a generalised function) and then proving that the size of the approximation is negligible. The same approach is used by Engle and Manganelli (2004) to derive the asymptotic distribution of the dynamic quantile test statistic defined in their Theorem 4.

In the following, we shall consider the twice continuously differentiable approximation $\tilde{q}_t^{\tau_j}(\theta)$ of $q_t^{\tau_j}(\theta)$ and its gradient $\nabla\tilde{q}_t^{\tau_j}(\theta)$, both defined in the Appendix, whose expectation, in the limit, is equal to the expectation of their a.e. or symbolic equivalent counterparts. We shall also introduce a deterministic sequence $\{c_T\}$ and a sequence of probabilities $\{\pi_T\}$, better detailed in the Appendix, who play a role in the smoothing of the filter and in its invertibility, respectively.

Theorem 2. *Let assumptions (i)-(v) of Theorem 1 hold. In addition, (vi) let the conditional density of* y_t *be Lipschitz continuous, (vii)* $D_{1t} = \max_{j=1,\dots,J} \max_{i=1,\dots,d} \sup_{\theta\in\Theta} |(\partial/\partial\theta_i)\tilde{q}_t^{\tau_j}(\theta)|$ *and* $D_{2t} = \max_{j=1,\dots,J} \max_{i,l=1,\dots,d} \sup_{\theta\in\Theta} |(\partial^2/\partial\theta_i\partial\theta_l)\tilde{q}_t^{\tau_j}(\theta)|$ *with* $E(D_{kt}^h) < \infty$ *for* $h, k = 1, 2$ *(viii)* θ_0 *is an interior point of* Θ *and (ix) the matrices* Q_0 *and* V_0 *defined below are positive definite. Also let (x)* $c_T = o(T^{\frac{1}{2}})$ *and (xi)* $\pi_T = o_p(T^2)$*. Then*

$$\sqrt{T}(\hat{\theta} - \theta_0) \to_d N(0, Q_0^{-1} V Q_0^{-1})$$

where

$$Q_0 = \sum_{j=1}^{J} E\left[f_{t|t-1}(q_t^{\tau_j}(\theta_0))\nabla\tilde{q}_t^{\tau_j}(\theta_0)\nabla\tilde{q}_t^{\tau_j}(\theta_0)'\right] \tag{13}$$

and

$$V_0 = E(\eta_t(\theta_0)\eta_t'(\theta_0))$$

with

$$\eta_t(\theta_0) = \sum_{j=1}^{J} \nabla\tilde{q}_t^{\tau_j}(\theta_0)(\tau_j - \mathbb{1}(y_t \leq q_t^{\tau_j}(\theta_0))). \tag{14}$$

The proof, in the Appendix, consists in applying Theorem 2 in White et al. (2015) to the QMLE that maximises $\varphi(y_t, \tilde{q}_t^{\tau_j}(\theta))$ and then showing that $\sup_{\theta\in\Theta} |\tilde{q}_t^{\tau_j}(\theta) - q_t^{\tau_j}(\theta)| = o_p(1)$, i.e. the invertibility of the filter, which is implied by the strong condition *(xi)*.

Consistent estimators of V_0 and Q_0 are obtained as in Theorems 3 and 4 White et al. (2010), respectively, to which we refer for the proofs. Specifically, a consistent estimator of V_0 is $\hat{V}_0 = \frac{1}{T}\sum_{t=1}^{T} \eta_t(\hat{\theta})\eta_t(\hat{\theta})'$, where $\eta_t(\theta)$ is defined in equation (14). A consistent estimator of Q_0 is obtained under the additional assumption that $E(D_{1t}^3) < \infty$ as

$$\hat{Q}_0 = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}\left[\hat{f}_{t|t-1}(q_t^{\tau_j}(\hat{\theta}))\nabla\tilde{q}_t^{\tau_j}(\hat{\theta})\nabla\tilde{q}_t^{\tau_j}(\hat{\theta})'\right]$$

where the conditional density is directly recovered from the set of quantiles,

$$\hat{f}_{t|t-1}(q_t^{\tau_{j-1}}(\hat{\theta})) = \frac{\tau_j - \tau_{j-1}}{q_t^{\tau_j}(\hat{\theta}) - q_t^{\tau_{j-1}}(\hat{\theta})}, \quad j = 1, \dots, J$$

with $\tau_1 = 0$ and the gradients are in Appendix. Note that, in our setting: estimation of the conditional density is direct, calculation of the gradient is straightforward and non crossing of quantiles is ensured even in finite samples.

4. Simulation study

In this section we report a simulation study to asses the finite sample properties of the estimator detailed in Section 3. The analysis proceeds by iterating $B = 500$ times the following procedure: i) simulate a time series of T observations from the true model, and ii) estimate the model parameters using equation (12). Concerning i), simulation from our Dynamic Multiple Quantile (DMQ) model is not straightforward, since the conditional distribution of the data is not known. However, if the number of available quantiles is large, the conditional cumulative distribution function can be well approximated such that random draws can be obtained by the inverse cdf method. In this experiment we choose $J = 99$ quantile levels $\tau_j = \{0.01, 0.02, \ldots, 0.99\}$. Clearly, this choice will have an impact on the results of our analysis since it affects the simulation from the true data generating process. If the conditional distribution if wrongly approximated we expect to observe bias in the estimation of model parameters. We consider five sample sizes: small $T = 250$, medium-small $T = 500$, medium $T = 1000$, medium-large $T = 2500$, and large $T = 5000$. True parameters are fixed at $\beta = 0.2$, $\alpha = 0.05$, $\gamma = 0.1$, and $\phi = 0.95$. Results are robust to different values of the parameters.

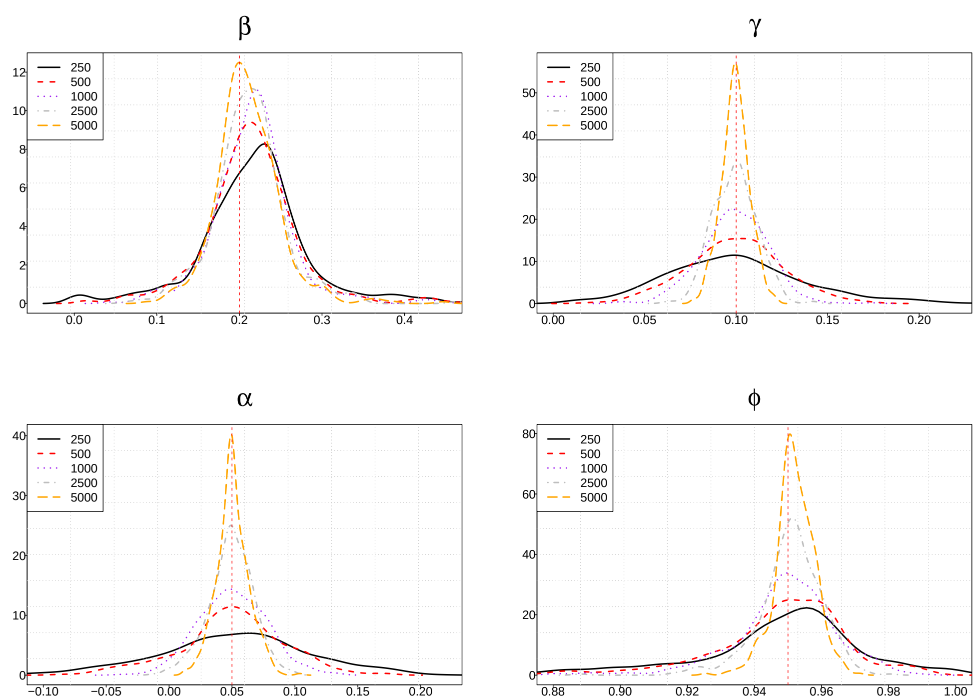


Figure 2: Gaussian kernel estimates based on $B = 500$ estimates from the DMQ model. True values are $\beta = 0.2$, $\alpha = 0.05$, $\gamma = 0.1$, and $\phi = 0.95$. Results are reported for five sample sizes: small $T = 250$ (black), medium-small $T = 500$ (red), medium $T = 1000$ (purple), medium-large $T = 2500$ (gray), and large $T = 5000$ (orange)

Figure 2 depicts the estimates of the parameters with the use of a Gaussian kernel with bandwidth selected according to the Silverman's rule of thumb. Results are very good and indicate that the Hogg estimator for the DMQ model has good finite sample properties. When the sample size is small ($T = 250$), we see that there is some bias in the estimate of β and that the estimates of γ are quite dispersed around the true value. However, for the medium-small and larger sample sizes ($T \geq 500$), estimates are good. The β parameter seems to the most difficult to estimate with a (small) bias that vanishes only when the sample size is large ($T = 5000$).

	$T = 250$			$T = 500$			$T = 1000$			$T = 2500$			$T = 5000$		
	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
β	0.224	0.024	0.090	0.215	0.015	0.062	0.218	0.018	0.058	0.213	0.013	0.050	0.210	0.010	0.041
α	0.049	−0.001	0.063	0.050	0.000	0.043	0.051	0.001	0.027	0.049	−0.001	0.018	0.051	0.001	0.013
γ	0.100	0.000	0.038	0.099	−0.001	0.026	0.099	−0.001	0.018	0.099	−0.001	0.012	0.100	0.000	0.008
ϕ	0.938	−0.012	0.063	0.946	−0.004	0.027	0.949	−0.001	0.017	0.951	0.001	0.010	0.952	0.002	0.006

Table 1: This table reports the mean of the estimated parameters, and the bias and root mean squared error (RMSE) between the estimated parameters and their true value. True values are $\beta = 0.2$, $\alpha = 0.05$, $\gamma = 0.1$, and $\phi = 0.95$. Results are reported for five sample sizes: small $T = 250$, medium-small $T = 500$, medium $T = 1000$, medium-large $T = 2500$, and large $T = 5000$.

To further investigate the finite sample properties of the estimator we compute the mean of the estimated parameters as well as their bias and root mean squared error (RMSE) with respect to their true values. Results are summarized in Table 1. We see that for all parameters the bias and the RMSE reduce when the sample size increases. Overall, we conclude that the Hogg estimator reports remarkable results in finite sample.

5. Empirical Illustration

We illustrate the DMQ model using the series of daily logarithmic returns, in percentage points, of Microsoft Corporation spanning from 8 December 2010 to 15 November 2018 for a total of $T = 2000$ observations. Observations are reported in Figure 5a from which we clearly observe the typical stylized facts that characterize financial returns such as heteroscedasticity and presence of extreme observations, see for example McNeil et al. (2015) for a textbook treatment of financial time series. Presence of heteroscedasticity is also supported by the ARCH-LM test of Engle (1982) computed with 12 lags that reports a statistic of 184.48 which is far from the critical value of 3.57 at the 5% confidence level. Gaussianity of the unconditional distribution is rejected according to the Jarque and Bera test which reports a statistics of 200. The empirical skewness and kurtosis coefficients are -0.42 and 4.31, respectively, indicating that the empirical distribution is left skewed and heavy tailed.

The illustration is divided in two parts. The first part reports full sample results regarding the estimation and goodness of fit of the DMQ model. The second part reports a forecasting exercise. For both parts we concentrate on $J = 99$ quantiles according to the series of equally spaced probability levels from $\tau_1 = 0.01$ to $\tau_J = 0.99$. The reference quantile is set to the median and it is assumed to be constant over time ($\alpha = \beta = 0$). Throughout the analysis we will employ two benchmark models:

i) the Conditional Autoregressive Value–at–Risk (CAViaR) model detailed in Engle and Manganelli (2004), and ii) the ARMA(P,Q)–GARCH(p,q) model of Bollerslev (1986) with skew Student's t distributed errors.

We implement the "asymmetric slope" CAViaR specification defined as:

$$q_t^{\tau_j} = \beta_{j,1} + \beta_{j,2} q_{t-1}^{\tau_j} + \beta_{j,3} y_{t-1} \mathbb{1}(y_{t-1} < 0) + \beta_{j,4} y_{t-1} \mathbb{1}(y_{t-1} > 0), \tag{15}$$

for $j = 1, \ldots, J$. CAViaR is thus a quantile autoregression where each quantile at time t linearly depends on its previous value at time $t-1$ and on past observations. Given its linear structure, CAViaR cannot guarantee the monotonicity of the filtered quantiles. Indeed, as long as the number of quantiles increases, the frequency of crossing quantiles will increase as well. The parameters $\beta_{1,j}$, $\beta_{2,j}$, $\beta_{3,j}$, and $\beta_{4,j}$ for $j = 1, \ldots, J$ are estimated by minimizing the loss function defined in equation (12). Since there are no constraints among the parameters for different quantile levels, estimation is divided in J independent minimization problems.

The ARMA(P,Q)–GARCH(p,q) is instead a location scale parametric model which assumes that:

$$y_t = \mu_t + \varepsilon_t, \tag{16}$$

where

$$\mu_t = \varpi + \sum_{i=1}^{P} \phi_i y_{t-j} + \sum_{l=1}^{Q} \theta_i \varepsilon_{t-l} \tag{17}$$

and $\varepsilon_t = \sigma_t z_t$ with

$$\sigma_t^2 = \omega + \sum_{m=1}^{p} \alpha_m \varepsilon_{t-m}^2 + \sum_{n=1}^{q} \zeta_n \sigma_{t-n}^2, \tag{18}$$

and z_t is assumed to be independently and identically distributed according to the standardized skew Student's t distribution built with the two–pieces method of Fernández and Steel (1998), with skewness and shape parameters $\upsilon > 0$ and $\nu > 2$, respectively. In the following we set $P = Q = p = q = 1$ which are the values that minimize the Bayesian information criterion for our sample of observations. The quantile at level τ_j and time t is then given by $q_t^{\tau_j} = \mu_t + \sigma_t q_z(\tau_j)$, where $q_z(\tau_j)$ is the time invariant τ_j–level quantile of z_t. The ARMA–GARCH specification ensures monotonicity of the quantiles at each point in time, however quantiles are constrained to follow a location-scale dynamic. Estimation of the ARMA-GARCH model is done by standard Maximum Likelihood, see Francq and Zakoian (2004).

We also consider the Dynamic Quantile (DQ) model defined as:

$$q_t^{\tau_j} = \bar{q}^{\tau_j}(1 - \beta_j) + \beta_j q_{t-1}^{\tau_j} + \alpha_j a_j^{-1} z_{j,t-1}, \tag{19}$$

for $j = 1, \ldots, J$, which is obtained by modelling each quantile separately within the DMQ model. In this case, all quantiles are set to the reference quantile and the forcing variable reduces to the step function $z_{j,t} = \mathbb{1}(y_t < q_t^{\tau_j}) - \tau_j$, normalized by its standard deviation $a_j = \sqrt{\tau_j(1-\tau_j)}$. Note that the DQ model is an EWMA in $z_{j,t}$, and resembles the specification adopted by De Rossi and Harvey (2006). Similar to CAViaR, the DQ model cannot guarantee the monotonicity of the filtered quantiles. Also, note that the DQ model provides an updating mechanism which is proportional to that used by Patton et al. (2019) in their joint model for the Value at Risk and Expected Shortfall.

5.1. Full sample results

We estimate the DMQ, CAViaR, ARMA-GARCH, and DQ models on the full sample of observations.[2] Filtered quantiles for the subset of probability levels $5\%, 10\%, \ldots, 95\%$ are reported in Figure 3. The graphical investigation suggests that DMQ quantiles are smoother than those reported by CAViaR and ARMA-GARCH. Indeed, it is evident from the picture that ARMA-GARCH filtered quantiles incorporate the rigidities of the underlying location-scale representation. Due to the very low signal in the conditional mean, all quantiles are basically driven by the filtered volatility. CAViaR filtered quantiles often cross: for the subset of quantiles $\{q_{5,t}, q_{10,t}, \ldots, q_{95,t}\}$, we observe a frequency of crossing of 11.5%. This frequency increases to 83.1% if the whole set of 99 quantiles is considered. Interestingly, the ARMA–GARCH quantiles at levels τ_{50} and τ_{55} are very similar. Filtered quantiles obtained by the DQ model (panel, d) are not satisfactory from a graphical point of view. Indeed, since in the DQ model each quantile update is only based on the hit variable $z_{i,t}$, the signal turns out to be rather weak.

To asses the goodness of fit of DMQ we compute the unconditional coverage (UC) and conditional coverage (CC) tests of Kupiec (1995) and Christoffersen (1998), respectively. Tests are computed independently for each of the 99 probability levels. Both tests are based on the requirement that, under the correct model specification, the j–th quantile violations are iid Bernoulli distributed with success rate equal to τ_j. Specifically, the null hypothesis of the UC test assumes the correct coverage of the unconditional distribution, which means that the observed relative frequency of the j–th quantile violations is equal to τ_j. The null hypothesis of the CC test assumes unconditional converage and independence of the quantile violations. See Kupiec (1995) and Christoffersen (1998) for further details on the UC and CC tests, respectively. Figure 4 reports the p-values of the UC (panel a) and CC (panel b) tests for each quantile level of the four models. It is not surprising that CAViaR reports the best results according to UC. Indeed, the independent estimation of each quantile level is made explicitly to target this result. Results for DMQ and GARCH are similar. The only case when the null hypothesis is rejected are the median in the ARMA-GARCH case, and the 88% quantile for the DQ model. Looking at the CC results in panel (b) we observe that p–values are similar across the

[2] Estimation of DMQ, CAViaR, and DQ requires the minimization of the Hogg function (12) which is not differentiable when $y_t = q_t^{\tau_j}$. We minimize the objective function employing the derivative free global optimization by differential evolution algorithm available in the DEoptim R package of Mullen et al. (2011).

four specifications. Interestingly, we note that all models fail to provide correct conditional coverage of the right tail of the conditional distribution above the 85% probability level. We find that ARMA-GARCH and DMQ reject the null hypothesis for the levels 86%–90%, and ARMA-GARCH rejects it for the 85% level. Overall, we find that quantiles of the right tail of the conditional distribution of Microsoft returns are more difficult to model compared to those from the left tail.

5.2. Conditional moments

A nice consequence of having many quantiles is that the conditional cumulative distribution function and associated moments can be well approximated. Indeed, we apply the following approximation:

$$\begin{aligned}\mathbb{E}[Y_t^p|\mathcal{F}_{t-1}] &= \int_{\mathbb{R}} y^p \mathrm{d}F_{t|t-1}(y) \\ &\approx \sum_{j=1}^{J} q_{j,t}^p(\tau_j - \tau_{j-1})\end{aligned}$$

where $\tau_0 = 0$. We compute the first four moments at each point in time and, from these, we recover the conditional mean and variance of the data, as well as the coefficients of skewness and kurtosis. Results are reported in Figure 5. Specifically, Figure 5a reports the filtered mean along with the observations. As expected, the mean is fairly constant around zero throughout the whole sample. Lower values of the time varying mean are generally associated with higher volatility levels, such as those registred during the periods of market instability associated with the European sovereign debt crisis of 2010–2011. The filtered variance is reported in Figure 5b and it is compared with the absolute returns, a common proxy of volatility. We observe that the filtered variance closely resembles the dynamic of the absolute returns indicating that the model is well adapting to periods of higher volatility. Figures 5c and 5d report the filtered skewness and kurtosis coefficients along with their unconditional value, respectively. Results indicate that their value oscillates around the empirical one. Skewness ranges between -1.5 and 0.5 indicating that the distribution has moved from negative to positive skewness over the sample period. Again, negative skewness is usually associated with periods of higher volatility. Kurtosis moves, on average, in the range 4–5 with the lowest and highest values close to 3 and 8, respectively. Overall, these results indicate that, beside the conditional mean, the shape of the conditional distribution is remarkably changing over the sample period.

5.3. Forecasting results

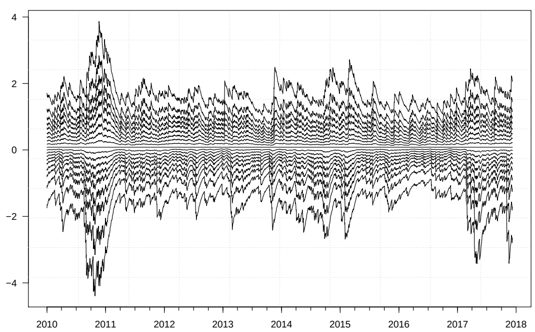
We now illustrate the performance of DMQ in an out of sample analysis. For this purpose, we divide our sample of 2000 observations in two parts of 1000 observations each. The first 1000 observations belong to the in sample period, which ranges from the beginning of the sample until 26 November 2014. Models parameters are estimated in sample and quantiles predictions are made over the following 1000 observations, that belong to the out of sample period. Predictions are made in a rolling-window way, i.e. each time a new observation becomes available, it is incorporated into the models before making new predictions. Predictions are made for one day ahead ($h = 1$) up to two weeks ($h = 10$) ahead. Among the four specifications considered, only DMQ and DQ provide closed form solutions for the h-step ahead predictions, with $h > 1$, see Section 2. The parametric formulation of ARMA–GARCH allows us to employ simulation techniques to draw observations from the predictive distribution, and thus to approximate the associated quantiles. In order to have a good coverage of the extreme left and right tails we employ 10'000 simulations. In the CAViaR model defined in equation (15), there are no ways for producing multistep quantiles predictions if a model for $|Y_t|$ is not in place. Thus, we resort to direct prediction which means that the model is reformulated as follows:

$$q_{t+h}^{\tau_j} = \beta_{j,1}^{(h)} + \beta_{j,2}^{(h)} q_{t+h-1}^{\tau_j} + \beta_{j,3}^{(h)} y_t \mathbb{1}(y_t < 0) + \beta_{j,4}^{(h)} y_t \mathbb{1}(y_t > 0), \tag{20}$$

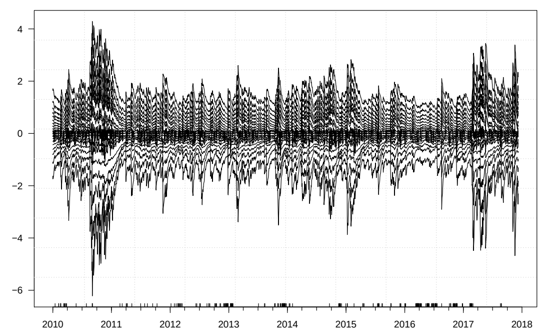
for $h = 1, \ldots, 10$ and $j = 1, \ldots, J$. Thus, direct prediction allows us to specify the h-step ahead quantile as a function of observations which are available at the end of the sample, see Marcellino et al. (2006) for a comparison of direct and iterate prediction in the linear autoregressive context. This solution comes at the cost of estimating $J \times 10$ single quantile CAViaR specifications, which is even more computational expensive than approximating the multistep ahead distribution of the ARMA–GARCH specification.

We report our results in the forms of aggregated quantile losses over the out of sample period and for different quantile levels. Table 2 reports the aggregated losses for subsets of quantile levels for $h = 1$, $h = 5$, and $h = 10$. Results are reported relative to the ARMA–GARCH specification which acts as a benchmark. Values lower than one indicate outperformance of the benchmark and viceversa. The results indicate that DMQ is the top performer among the considered specifications. Indeed, DMQ average losses are on average lower than the CAViaR, ARMA–GARCH, and DQ ones, irrespectively on the area of the predictive distribution and the forecast horizon. We statistically assess the magnitude of the differences between the considered specifications by employing the Model Confidence Set (MCS) procedure of Hansen et al. (2011) on each series of aggregated loss differentials. Results indicate that DMQ always belongs to the superior set of models at the 75% confidence level. Results are stronger for $h = 5$. Table 3 leads to similar considerations, but for all the forecast horizons and for the three subsets of the predictive distribution, namely: left tail ($\tau_j \leq 0.5$), right tail ($\tau_j \geq 0.5$), center ($0.25 \leq \tau_j \leq 0.75$), as well as for the whole distribution ($0.01 \leq \tau_j \leq 0.99$). Also in this case, results indicate that DMQ losses are on average lower than CAViaR, ARMA–GARCH, and DQ, irrespectively of the distribution region and the forecast horizon. The MCS procedure also supports these findings.

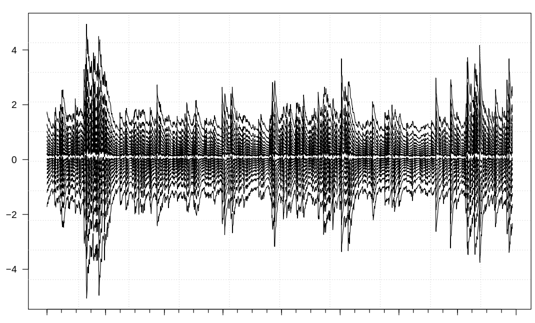
To conclude our illustration, we compute the UC and CC tests as in the full sample analysis, but considering one step ahead predictions. P-values for all quantile levels are reported in Figure 6. Interestingly, p-values associated with CAViaR, ARMA–GARCH, and DQ predictions are lower than in the full sample analysis, while those of DMQ are higher. This is remarkably true when looking at the right tail of the predictive distribution. Indeed, while DMQ is always able to provide correct conditional and unconditional coverage, p–values of the ARMA–GARCH, CAViaR, and DQ indicate rejection of the null hypothesis of both tests in the range $0.55 \leq \tau_j \leq 0.90$. Note that DQ provides the worst results, thus suggesting that a joint modelling of multiple quantiles in a score driven framework is required.



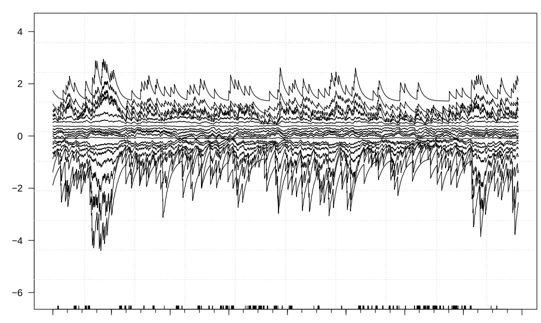
(a) DMQ



(b) CaViaR

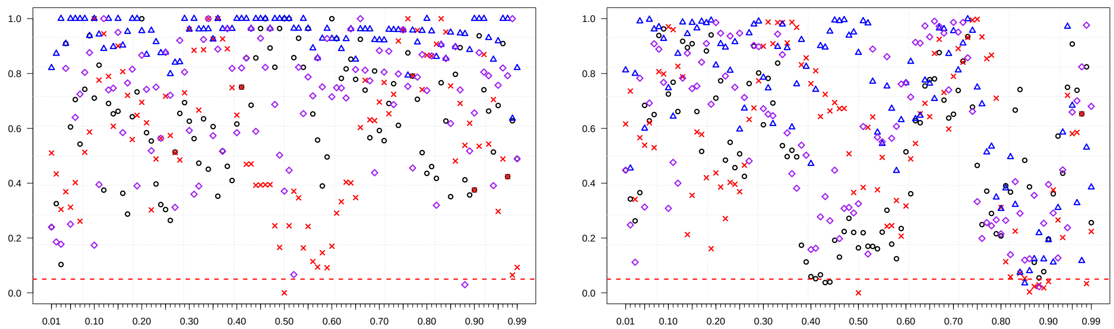


(c) ARMA-GARCH



(d) DQ

Figure 3: Filtered quantiles from the DMQ (a), CAViaR (b), ARMA-GARCH (c), and DQ (d) models at levels 5%, 10%,. . . , 95% for Microsoft Corporation from 8 December 2010 to 15 November 2018 for a total of $T = 2000$ observations. Rugs in panel (b) indicate points in time when CAViaR quantiles cross.



(a) UC p-values

(b) CC p-values

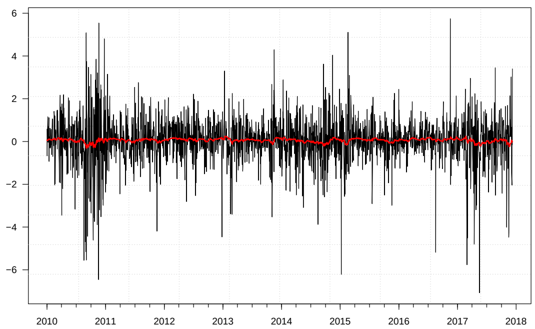
Figure 4: P-values of the UC (panel a) and CC (panel b) tests over the in sample period for $\tau_1 = 1\%$, $\tau_2 = 2\%$, ..., $\tau_{99} = 99\%$ quantile levels of DMQ (black circles), CAViaR (blue triangles), ARMA-GARCH (red crosses), and DQ (purple diamonds). The horizontal red dashed line represents the 5% probability level.

	$h=1$				$h=5$				$h=10$			
	DMQ	DQ	CAViaR	A–G	DMQ	DQ	CAViaR	A–G	DMQ	DQ	CAViaR	A–G
$0.01 \leq \tau_j < 0.10$	1.003	1.059	1.029	1.000	0.987	1.021	1.039	1.000	0.994	1.034	1.030	1.000
$0.10 \leq \tau_j < 0.20$	1.000	1.036	1.023	1.000	0.990	1.018	1.014	1.000	1.001	1.024	1.015	1.000
$0.20 \leq \tau_j < 0.30$	0.997	1.014	0.995	1.000	0.993	1.006	0.992	1.000	1.004	1.012	0.998	1.000
$0.30 \leq \tau_j < 0.40$	0.999	0.998	0.996	1.000	0.996	0.996	0.996	1.000	1.002	1.002	0.999	1.000
$0.40 \leq \tau_j < 0.50$	0.997	0.995	1.001	1.000	0.998	0.997	1.002	1.000	1.000	0.999	0.999	1.000
$0.50 \leq \tau_j < 0.60$	0.997	0.997	0.999	1.000	0.999	0.999	1.005	1.000	0.998	0.998	1.001	1.000
$0.60 \leq \tau_j < 0.70$	1.000	1.000	1.001	1.000	1.001	1.002	1.011	1.000	0.998	0.999	1.015	1.000
$0.70 \leq \tau_j < 0.80$	1.004	1.001	1.005	1.000	1.007	1.004	1.017	1.000	0.998	0.997	1.022	1.000
$0.80 \leq \tau_j < 0.90$	1.001	0.994	1.005	1.000	1.004	0.995	1.018	1.000	0.996	0.990	1.022	1.000
$0.90 \leq \tau_j \leq 0.99$	0.992	1.025	0.986	1.000	0.988	1.016	0.986	1.000	0.981	1.008	0.977	1.000
$0.01 \leq \tau_j \leq 0.99$	0.999	1.007	1.002	1.000	0.998	1.003	1.007	1.000	1.000	1.005	1.008	1.000

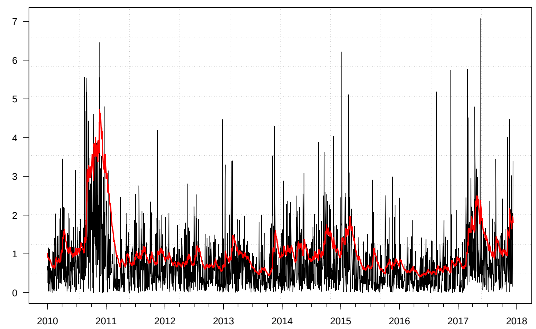
Table 2: Average quantile loss over the out of sample period. Losses are aggregated over subset of quantiles. For example, the first row reports the aggregated losses for quantiles associated with probability levels in the range $[0.01, 0.1)$. Results are reported for one ($h = 1$), five ($h = 5$), and ten ($h = 10$) steps ahead predictions and are relative to the ARMA–GARCH model indicated with "A–G". Values lower than one indicate outperformance with respect to ARMA–GARCH and viceversa. Gray cells indicate models that belongs to the Superior Set of Models computed according to the Model Confidence Set procedure of Hansen et al. (2011) at the 75% confidence level.

5.4. The choice of J

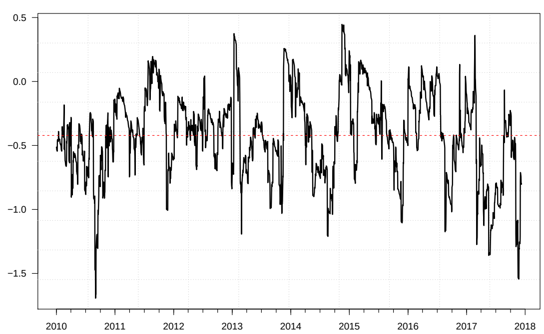
We conclude the empirical analysis by investigating how the choice of J, the dimension of the multiple quantile, affects quantile prediction. To this end, we make predictions of the nine



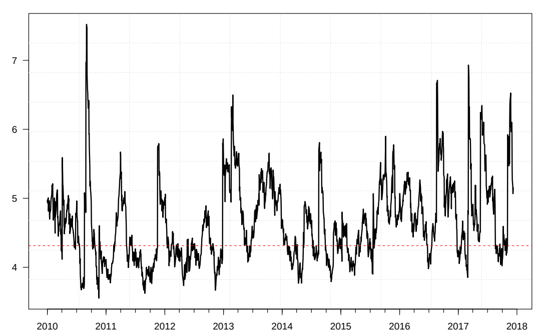
(a) Returns and filtered mean



(b) Absolute returns and filtered variance



(c) Filtered skewness coefficient



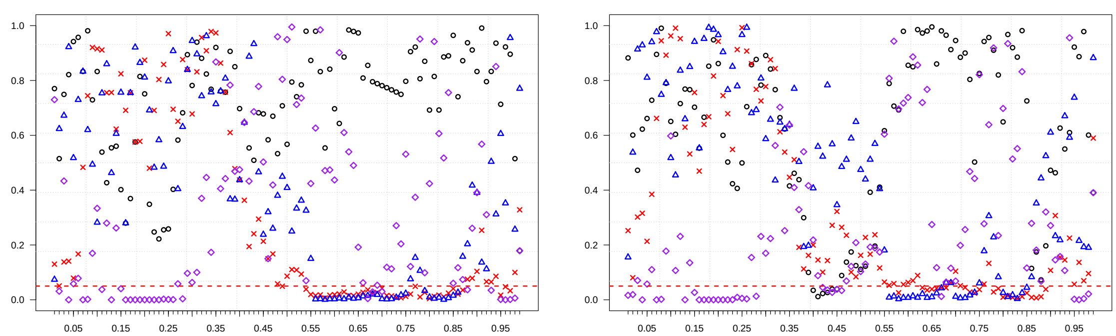
(d) Filtered kurtosis coefficient

Figure 5: The figure reports the filtered conditional mean (red line) along with the values of the Microsoft Corporation returns in panel (a). Panel (b) reports the filtered variance (red line) along with the absolute value of the returns. Panels (c) and (d) report the filtered skewness and kurtosis coefficients, respectively. The horizontal dashed red lines in panels (c) and (d) indicate the empirical values of the coefficients over the sample.

deciles $q_t^{10\%}, \ldots, q_t^{90\%}$ with the DMQ model estimated with different choices of J. Specifically, we consider (a) the case when $J = 1$, which corresponds to the DQ model where all quantiles are independently filtered, (b) the case when $J = 9$, with $\tau_1 = 10\%, \ldots, \tau_9 = 90\%$, and, in general, (c) the case when $J = 10S - 1$, for $S = 1, \ldots, 10$, with associated probability levels $\tau_1 = \xi_S, \tau_2 = 2\xi_S, \ldots, \tau_{10S-1} = (10S-1)\xi_S$, where $\xi_S = \frac{10\%}{S}$. Note that, for all the choices of J, the probability levels corresponding to the nine deciles of interest are always included. The experiment proceeds like the forecasting analysis reported in the previous section: the first half of the sample is used for model estimation, the second half is used for predictions. Quantile losses are subsequently

	Left tail				Right tail				Center			
	DMQ	DQ	CAViaR	A–G	DMQ	DQ	CAViaR	A–G	DMQ	DQ	CAViaR	A–G
$h = 1$	0.997	1.015	1.001	1.000	1.001	1.002	1.001	1.000	0.999	1.000	1.000	1.000
$h = 2$	0.993	1.010	0.997	1.000	0.999	1.001	1.003	1.000	0.998	0.998	0.998	1.000
$h = 3$	0.993	1.008	1.001	1.000	1.000	1.000	1.006	1.000	0.998	0.998	1.003	1.000
$h = 4$	0.993	1.007	0.997	1.000	1.003	1.002	1.001	1.000	0.999	0.999	0.998	1.000
$h = 5$	0.994	1.007	1.001	1.000	1.004	1.003	1.009	1.000	0.999	1.000	1.004	1.000
$h = 6$	0.994	1.007	1.000	1.000	1.004	1.003	1.009	1.000	0.999	0.999	1.003	1.000
$h = 7$	0.995	1.008	1.001	1.000	1.002	1.001	1.011	1.000	1.000	1.001	1.007	1.000
$h = 8$	0.996	1.009	1.002	1.000	1.002	1.000	1.009	1.000	0.999	1.000	1.006	1.000
$h = 9$	1.000	1.013	1.004	1.000	1.001	1.000	1.010	1.000	1.001	1.001	1.007	1.000
$h = 10$	0.999	1.010	1.001	1.000	1.000	1.000	1.012	1.000	1.000	1.001	1.005	1.000

Table 3: Average quantile loss over the out of sample period for different forecast horizons h. Losses are aggregated over the left tail ($\tau_j \leq 0.5$), right tail ($\tau_j \geq 0.5$), center ($0.25 \leq \tau_j \leq 0.75$), and all ($0.01 \leq \tau_j \leq 0.99$) distribution. Results are reported relative to the ARMA–GARCH model indicated with "A–G". Values lower than one indicate outperformance with respect to ARMA–GARCH and viceversa. Gray cells indicate models that belongs to the Superior Set of Models computed according to the Model Confidence Set procedure of Hansen et al. (2011) at the 75% confidence level.



(a) UC p-values

(b) CC p-values

Figure 6: P-values of the UC (panel a) and CC (panel b) tests over the out of sample period for $\tau_1 = 1\%$, $\tau_2 = 2\%$, ..., $\tau_{99} = 99\%$ quantile levels of DMQ (black circles), CAViaR (blue triangles), ARMA-GARCH (red crosses), and DQ (purple diamonds). The horizontal red dashed line represents the 5% probability level.

computed and averaged over the evaluation period. Results are reported in Table 4 relative to the case $J = 1$. As expected, we see that the jointly modelling of multiple quantiles is a better strategy compared to individual quantile filtering. Indeed, as detailed in Section 2.1, the choice of J determines the amount of discretization of the cdf used in the quantiles forcing variable. Interestingly, we find

	$J=1$	$J=9$	$J=19$	$J=29$	$J=39$	$J=49$	$J=59$	$J=69$	$J=79$	$J=89$	$J=99$
$\tau=0.1$	1.000	0.986	0.984	0.989	0.985	0.986	0.986	0.986	0.986	0.986	0.986
$\tau=0.2$	1.000	0.998	0.995	0.995	0.996	0.995	0.996	0.996	0.995	0.995	0.996
$\tau=0.3$	1.000	1.000	0.999	0.998	0.999	0.999	0.999	0.999	0.999	0.999	0.999
$\tau=0.4$	1.000	1.003	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002	1.002
$\tau=0.6$	1.000	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
$\tau=0.7$	1.000	0.999	0.998	0.997	0.997	0.998	0.998	0.997	0.997	0.997	0.997
$\tau=0.8$	1.000	0.998	0.996	0.994	0.996	0.996	0.996	0.996	0.995	0.995	0.995
$\tau=0.9$	1.000	1.011	1.003	1.001	1.003	1.003	1.004	1.003	1.002	1.002	1.003

Table 4: Average quantile loss over the out of sample period for one step ahead quantile predictions from the DMQ model with different choices of J. Results are reported relative to the case $J=1$, which is the DQ model. Values lower than one indicate outperformance with respect to DQ and viceversa. Gray cells indicate models that belongs to the Superior Set of Models computed according to the Model Confidence Set procedure of Hansen et al. (2011) at the 75% confidence level.

that the results are not very much affected by the choice of J when this is greater or equal than 9. Indeed, in all the cases, the results are very similar and do not seem to improve by increasing J. However, we must stress that if the goal of the analysis is to recover some quantity that involves the cdf, like the predictive moments of the predictive density, better results are most likely obtained by increasing J.

6. Conclusions

The paper poses the basis for a new way of modelling the conditional distribution of time–series. The specification of a semiparametric dynamic multiple quantile model has been shown to be convenient for a number of reasons such as: i) simple estimation, *ii*) closed form solutions for quantile predictions and their limiting values, *iii*) asymptotic results, *iv*) good finite sample properties of the estimators, and *v*) very promising empirical results. Many aspects have still to be further investigated, some of these are: i) the changes in the fixed parameters according to the number of selected quantiles, *ii*) different specifications of the forcing variable $u_{j,t}$, and mostly , *iii*) the case when $u_{j,t}$ is not $\mathcal{F}_{t-1}$–measurable.

In the empirical illustration, we analysed the series of financial returns of Microsoft Corporation which is characterized by very low persistence in the conditional mean. During the preparation of the paper we have also investigated the performance of DMQ in modelling US inflation, which is known to be a very persistent process. Also in that case the results have been found very promising. Future research can consider studying the performance of DMQ also in economic time series.

Appendix

Proof of Theorem 1 (Consistency) We apply Corollary 5.11 of White (1994). Assumptions *(i)-(v)* of Theorem 1 satisfy assumptions 2.1, 5.1, 5.4, 3.1, 3.2 of White (1994), which are required for the corollary to hold, as follows. Assumptions 2.1, 5.1, 5.4 are standard regularity conditions on the the process, on the model and on the objective function, that correspond to our assumptions *(i)*, *(ii)*, and *(iii)*, respectively. White's condition 3.1 requires that $E[\varphi(y_t, q_t^{\tau_j}(\theta))]$ is finite, continuous on Θ and that $\varphi(y_t, q_t^{\tau_j}(\theta))$ obeys a uniform law of large numbers (ULLN). Here we apply Theorem A.2.2 of White (1994), that is a ULLN for stationary and ergodic processes, as is our process by assumption *(i)*, which follows from Ranga Rao (1962). Condition *(iv)* implies the dominance condition

$$\max_{j=1,\dots,J} \sup_{\theta\in\Theta} |q_t^{\tau_j}(\theta)| = D < \infty \tag{21}$$

with $E(D) < \infty$, as follows. Let us consider the reference quantile. From equation (7), combined with (9) and with $|\beta| < 1$,

$$|q_t^{\tau_{j^*}}| \leq |\bar{q}^{\tau_{j^*}}(1-\beta)| + |\alpha| \sum_{s=0}^{\infty}\sum_{j=1}^{J} |\beta^s||z_{j,t-1-s}| \leq |\bar{q}^{\tau_{j^*}}(1-\beta)| + |\alpha| J\upsilon \sum_{s=0}^{\infty} |\beta^s| = |\bar{q}^{\tau_{j^*}}(1-\beta)| + \frac{|\alpha| J\upsilon}{1-|\beta|} < \infty$$

where we have set $\max_{j=1,\dots,J}\sup_{\theta\in\Theta} z_{j,t} = \max_{j=1,\dots,J}(|\tau_j|, |1-\tau_j|) = \upsilon$. The same reasoning extends to the other quantiles provided that $|\gamma| < \infty$ and $|\phi| < 1$. Condition (21) allows one to prove that $|\varphi(y_t, q_t^{\tau_j}(\theta))| < \infty$, as follows:

$$|\varphi(y_t, q_t^{\tau_j}(\theta))| = |\sum_{j=1}^{J} \rho_{\tau_j}(y_t, q_t^{\tau_j}(\theta))| \leq \sum_{j=1}^{J} |(y_t - q_t^{\tau_j}(\theta))(\tau_j - \mathbb{1}(y_t \leq q_t^{\tau_j}(\theta)))| < \sum_{j=1}^{J} |y_t - q_t^{\tau_j}(\theta)|.$$

As

$$\sum_{j=1}^{J} |y_t - q_t^{\tau_j}(\theta)| \leq \sum_{j=1}^{J} |y_t| + \sum_{j=1}^{J} |q_t^{\tau_j}(\theta)| \leq J|y_t| + J|D|$$

then, by *(iv)*, $E(|\varphi(y_t, q_t^{\tau_j}(\theta))|) < \infty$ and the conditions for theorem A.2.2 of White (1994) to hold are satisfied. Clearly, $E(|\varphi(y_t, q_t^{\tau_j}(\theta))|) < \infty \Rightarrow E(\varphi(y_t, q_t^{\tau_j}(\theta))) < \infty$. Also, continuity of $\varphi(y_t, q_t^{\tau_j}(\theta))$ in Θ follows by continuity of the check function in $q_t^{\tau_j}(\theta)$ and by almost everywhere continuity (with respect to P) of $q_t(\theta)$ on Θ. We have then verified assumption 3.1 of White (1994). To prove the validity of assumption 3.2 of Corollary 5.11 by White (1994) (unique identifiability), it is sufficient to prove that the following strict inequality holds, $E(\varphi(y_t, q_t^{\tau_j}(\theta))) < E(\varphi(y_t, q_t^{\tau_j}(\theta_0)))$ for $\theta \neq \theta_0$. Under assumption (v), the proof follows exactly the same steps as the proof by White et al. (2015), which, on its turn, follows closely by Powell (1984), in the multivariate (n variables) case, except that we are here in the univariate case ($n = 1$), and thus we omit it $\square$

Proof of Theorem 2 (Asymptotic normality) The proof of theorem 2 is divided in two parts. The first part consists in approximating the a.e. continuous function $q_t^{\tau_j}(\theta)$ by a twice continuously differentiable function denoted as $\tilde{q}_t^{\tau_j}(\theta)$ and then applying Theorem 2 in White et al. (2015) to prove the asymptotic normality of the QMLE obtained by maximising $\varphi(y_t, \tilde{q}_t^{\tau_j}(\theta))$. The second part consists in showing that the difference between the approximating function $\tilde{q}_t^{\tau_j}$ and the estimator $q_t^{\tau_j}$ is negligible, i.e. that $\sup_{\theta\in\Theta} |\tilde{q}_t^{\tau_j}(\theta) - q_t^{\tau_j}(\theta)| = o_p(1)$, an invertibility condition for our filter.

Let us approximate the hit variable $z_{j,t}$, defined in equation (5), with a twice continuously differentiable function, also used in Engle and Manganelli (2004). Based on y_1 (first observation) and $q_1^{\tau_j}$ (given initial condition), we construct the sequence of variables

$$\tilde{z}_{j,t} = \left(1 + \exp\left\{\frac{y_t - \tilde{q}_t^{\tau_j}(\theta)}{c_T}\right\}\right)^{-1} - \tau_j \tag{22}$$

where $\{c_T\}$ is a sequence of deterministic variables satisfying $\lim_{T\to\infty} c_T = 0$, see assumption *(x)*. For $t \geq 1$, $\tilde{q}_{t+1}^{\tau_j}$ and $\tilde{u}_{t_j}$ are smooth counterparts of $q_{t+1}^{\tau_j}$ and u_{tj} in equations (6) and (9), respectively, constructed, accordingly, as functions of $\tilde{z}_{jt}$. Note that

$$\lim_{T\to\infty} \tilde{z}_t^{\tau_j} = \mathbb{1}(y_t \leq \tilde{q}_t^{\tau_j}(\theta)) - \tau_j. \tag{23}$$

Under assumptions *(i)*-*(ix)*, the conditions for applying Theorem 2 in White et al. (2015) hold and

$$\sqrt{T}(\tilde{\theta} - \theta_0) \to_d N(0, \tilde{Q}_0^{-1}\tilde{V}\tilde{Q}_0^{-1})$$

where $\tilde{\theta} = \arg\max_{\theta\in\Theta} \varphi(y_t, \tilde{q}_t^{\tau_j}(\theta))$,

$$\tilde{V}_0 = E(\tilde{\eta}_t(\theta_0)\tilde{\eta}_t'(\theta_0))$$

with

$$\tilde{\eta}_t(\theta) = \sum_{j=1}^{J} \nabla\tilde{q}_t^{\tau_j}(\theta)(\tau_j - \mathbb{1}(y_t \leq \tilde{q}_t^{\tau_j}(\theta))) \tag{24}$$

and

$$\tilde{Q}_0 = \sum_{j=1}^{J} E[f_{t|t-1}(\tilde{q}_t^{\tau_j}(\theta_0)\nabla\tilde{q}_t^{\tau_j}(\theta_0)\nabla'\tilde{q}_t^{\tau_j}(\theta_0)]. \tag{25}$$

In essence, the proof of this first part is developed by applying the mean value theorem around θ_0 to the smooth expectation of the first derivative of the a.e. differentiable quasi-likelihood of $\tilde{q}_t^{\tau_j}(\theta)$and applying Theorem 5.24 of White (2001), a CLT for martingale difference sequences (mds), to $\tilde{\eta}_t(\theta_0)$. We summarise the main steps here in the following and refer the reader to White et al. (2015) for further details.

Assumption *(vii)*, along with twice differentiability of $\tilde{q}_t^{\tau_j}(\theta)$, ensures the existence of $\lambda(\theta) = E[\tilde{\eta}_t(\theta)]$, where $\tilde{\eta}_t = \nabla\varphi(y_t, \tilde{q}_t^{\tau_j}(\theta))$ is computed in equation (24) and its expansion as

$$\lambda(\theta) = \lambda(\theta_0) + Q_s(\theta - \theta_0), \tag{26}$$

where $Q_s = \nabla\lambda(\theta)|_{\theta=\theta_s}$ and θ_s lies between θ and θ_0 (assumption *(viii)*). By assumptions *(v)*-*(vii)* one has $Q_s = -\tilde{Q}_0 + O(||\theta - \theta_0||)$ where $\tilde{Q}_0$ is as in equation (25). Correct specification and the fact that the filter is measurable imply that $\lambda(\theta_0) = 0$. Assumptions *(v)*, *(vii)* and *(ix)* allow one to verify Weiss (1991) conditions for getting $\sqrt{T}\lambda(\tilde{\theta}) = -\frac{1}{T}\sum_{t=1}^{T}\tilde{\eta}(\theta_0) + o_p(1)$. Combining these results with the expansion in equation (26) gives

$$-\frac{1}{T}\sum_{t=1}^{T}\tilde{\eta}(\theta_0) + o_p(1) = -\tilde{Q}_0\sqrt{T}(\tilde{\theta} - \theta_0) + O(||\tilde{\theta} - \theta_0||^2).$$

As $\tilde{\eta}_t(\theta_0)$ is mds and $\tilde{V}_0$ is finite by *(vii)* and positive definite by *(ix)* and as $\frac{1}{T}\sum_{t=1}^{T}\tilde{\eta}_t(\theta_0)\tilde{\eta}_t'(\theta_0) + o_p(1)$ (by the ergodic theorem), Theorem 5.24 in White (2001) can be applied and one has

$$\tilde{V}_0^{-\frac{1}{2}}\tilde{Q}_0\sqrt{T}(\tilde{\theta} - \theta_0) \to_d N(0, I_d).$$

The crucial condition $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta(\tilde{\theta}) = o_p(1)$ follows from twice continuous differentiability of $\tilde{q}_t$ and from assumptions *(v)* and *(vii)*.

The second part of the proof consists in showing that condition (xi) implies that the filter is invertible, and that the difference between $\tilde{\theta}$ and $\hat{\theta}$ is asymptotically negligible. When this is the case, the limit in equation (23) is almost surely equal to z_{jt} as, in the limit, the variables $\tilde{z}_{jt}$ and z_{jt} differ only for a set of points of zero measure. Analogously, $\tilde{q}_t^{\tau_j} \to_{as} q_t^{\tau_j}$ and the claim of Theorem 2 is valid. For clarity of exposition, we shall consider the case when $J = 1$. Generalisations to different and multiple quantiles follow with similar arguments.

Let us consider a time point T_0 after which $c_T \to 0$. We shall reset the origin at $t = T_0 = 0$, where the filters q_t and $\tilde{q}_t$ become the same, i.e. belong to the same model and finite sample misspecification vanishes (ensuring consistency of $\tilde{\theta}$ for θ_0)

$$\begin{aligned} q_{t+1} \propto &\quad \alpha(\mathbb{1}(y_t \leq q_t) - \tau) + \beta q_t \\ \tilde{q}_{t+1} \propto &\quad \alpha(\mathbb{1}(y_t \leq \tilde{q}_t) - \tau) + \beta\tilde{q}_t. \end{aligned}$$

Define $\delta_t = q_t - \tilde{q}_t$, with $\delta_0 = q_0 - \tilde{q}_0$ and assume that $\delta_0 > 0$ (the case $\delta_0 < 0$ is analog). One has

$$\begin{aligned} \delta_{t+1} &= \alpha X_t + \beta\delta_t \\ &= \alpha\sum_{s=0}^{t}\beta^s X_{t-s} + \beta^{t+1}\delta_0 \end{aligned}$$

where $X_t = \mathbb{1}(y_t \leq q_t) - \mathbb{1}(y_t \leq \tilde{q}_t)$, a Bernoulli r.v. with success probability

$$\pi_t = P(\tilde{q}_t \leq y_t \leq q_t) = F_{t|t-1}(q_t) - F_{t|t-1}(\tilde{q}_t) < \tau.$$

Let us consider the sequence of events $y_t \in e_t$, where $e_t = [\tilde{q}_t, q_t]$. One can envisage the following cases: $P(y_t \in e_t) = \pi_t$ if $e_t > 0$ and $P(y_t \in e_t) = 0$ if $e_t = 0$; conversely, $P(y_t \notin e_t) = 1 - \pi_t$ if $e_t > 0$

and $P(y_t \notin e_t) = 1$ if $e_t = 0$. By assumption *(xi)*, $\sum_{t=1}^{\infty} P(y_t \in e_t) = \sum_{t=1}^{\infty} \pi_t < \infty$. By the first Borel-Cantelli lemma, as $\sum_{t=1}^{\infty} P(y_t \in e_t) < \infty$, then $P(y_t \in e_t \ \ i.o.) = 0$ almost surely and this may occur only if e_t is a (zero P-measure) point, i.e. if $q_t = \tilde{q}_t$ □

Gradients The variables $\tilde{z}_{j,t}$, considered also in Engle and Manganelli (2004), are bounded between $-\tau_j$ and $1-\tau_j$ and

$$\frac{\partial}{\partial\theta_i}\tilde{z}_{j,t} = k_{t,T}\frac{\partial}{\partial\theta_i}\tilde{q}_t^{\tau_j}(\theta)$$

where

$$k_{t,T} = \frac{\frac{1}{c_T}\exp\left\{\frac{y_t - \tilde{q}_t^{\tau_i}(\theta)}{c_T}\right\}}{\left(1+\exp\left\{\frac{y_t - \tilde{q}_t^{\tau_i}(\theta)}{c_T}\right\}\right)^2}$$

is the pdf of a logistic r.v. y_t with mean equal to $\tilde{q}_t^{\tau_j}$ and variance equal to $(c_T^2\pi^2)/3$.

One can write

$$\tilde{q}_t^{\tau_j} = \tilde{q}_t^{\tau_j^*} - \sum_{i=l}^{r} b_i\tilde{\eta}_{it}$$

where the pair $\{l, r\}$ is equal to $\{1, j\}$ for $j < j^*$, to $\{j, J\}$ for $j > j^*$ and is empty for $j = j^*$, while b_i was previously defined as $b_i = \mathbb{1}(\tau_i < \tau_{i^*}) - \mathbb{1}(\tau_i > \tau_{i^*})$. Then

$$\frac{\partial}{\partial\theta_i}\tilde{q}_t^{\tau_j}(\theta) = \frac{\partial}{\partial\theta_i}\tilde{q}_t^{\tau_j^*}(\theta) - \sum_{i=l}^{r} b_i\tilde{\eta}_{it}(\theta)\frac{\partial}{\partial\theta_i}\xi_{it}(\theta),$$

where the $i-$th derivative of the reference quantile is

$$\frac{\partial}{\partial\theta_i}\tilde{q}_t^{\tau_j^*}(\theta) = \frac{\partial}{\partial\theta_i}(\bar{q}^{\tau_{j^*}}(1-\beta)) + \frac{\partial}{\partial\theta_i}(\alpha\tilde{u}_{t-1}^{\tau_{j^*}}) + \frac{\partial}{\partial\theta_i}(\beta\tilde{q}_{t-1}^{\tau_{j^*}})$$

that is, writing $\tilde{u}_t^{\tau_j^*}$ as in equation (9) as a function of $\tilde{z}_{tj}$ and then differentiating,

$$\frac{\partial}{\partial\theta_i}\tilde{q}_t^{\tau_j^*}(\theta) = \frac{\partial}{\partial\theta_i}(\bar{q}^{\tau_{j^*}}(1-\beta)) + \left(\frac{\partial}{\partial\theta_i}\alpha\right)\sum_{j=1}^{J}\tilde{z}_{t-1,i} + \alpha\sum_{i=1}^{J}k_{t,T}(\theta)\frac{\partial}{\partial\theta_i}\tilde{q}_t^{\tau_j}(\theta) + \left(\frac{\partial}{\partial\theta_i}\beta\right)q_{t-1}^{\tau_{j^*}} + \beta\frac{\partial}{\partial\theta_i}q_{t-1}^{\tau_{j^*}}.$$

The derivative of $\tilde{\xi}_{j,t}$ is obtained analogously, except that the summations that involve the variables $z_{t,j}$ are truncated,

$$\frac{\partial}{\partial\theta_i}\tilde{\xi}_{tj}(\theta) = \frac{\partial}{\partial\theta_i}(\bar{\xi}^{\tau_{j^*}}(1-\phi)) + \left(\frac{\partial}{\partial\theta_i}\gamma\right)\sum_{l}\tilde{z}_{t-1,l} + \gamma\sum_{l}k_{t,T}(\theta)\frac{\partial}{\partial\theta_i}\tilde{q}_t^{\tau_l}(\theta) + \left(\frac{\partial}{\partial\theta_i}\phi\right)\xi_{t-1,j} + \phi\frac{\partial}{\partial\theta_i}\xi_{t-1,j}.$$

Taking, for instance, $\theta_i = \alpha$, one has, for the reference quantile

$$\frac{\partial}{\partial\alpha}\tilde{q}_t^{\tau_j^*} = \tilde{u}_{t-1}^{\tau_j*} + \alpha\frac{\partial}{\partial\alpha}\tilde{u}_{t-1}^{\tau_j*} + \beta\frac{\partial}{\partial\alpha}\tilde{q}_{t-1}^{\tau_j*}$$

that is

$$\frac{\partial}{\partial\alpha}\tilde{q}_t^{\tau_j^*} = \sum_{j=1}^{J}\tilde{z}_{j,t-1} + \alpha\sum_{j=1}^{J}\frac{\partial}{\partial\alpha}\tilde{z}_{j,t-1} + \beta\frac{\partial}{\partial\alpha}\tilde{q}_{t-1}^{\tau_j*}$$

and replacing the derivative of $\tilde{z}_{it}$

$$\frac{\partial}{\partial\alpha}\tilde{q}_t^{\tau_j^*} = \sum_{j=i}^{J}\tilde{z}_{j,t-1} + \alpha\sum_{j=1}^{J}k_{j,t-1}\frac{\partial}{\partial\alpha}\tilde{q}_{t-1}^{\tau_j} + \beta\frac{\partial}{\partial\alpha}\tilde{q}_{t-1}^{\tau_j*}$$

and for the remaining quantiles

$$\frac{\partial}{\partial\alpha}\tilde{\xi}_{tj} = \gamma\sum_{l}k_{t,T}\frac{\partial}{\partial\alpha}\tilde{q}_t^{\tau_l} + \phi\frac{\partial}{\partial\theta_i}\xi_{t-1,j}.$$

References

Amemiya, T. (1982). Two stage least absolute deviations estimators. *Econometrica*, 50(3):689–712.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bondell, H., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838.

Buchinsky, M. et al. (1994). Changes in the us wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62:405–405.

Chan, K. and Tsay, R. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85(2):413–426.

Chernozhukov, V., Fernandez Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):1093–1125.

Chernozhukov, V., Fernandez Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica: journal of the Econometric Society*, 78(3):1093–1125.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Daniels, H. E. (1961). The asymptotic efficiency of a maximum likelihood estimator. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 151–163, Berkeley, Calif. University of California Press.

De Rossi, G. and Harvey, A. C. (2006). Time-varying quantiles.

Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society*, 70(3):609 – 627.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Francq, C. and Zakoian, J.-M. (2004). Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637.

Gouriéroux, C. and Jasiak, J. (2008). Dynamic quantile models. *Journal of econometrics*, 147(1):198–205.

Granger, C. W. (2010). Some thoughts on the development of cointegration. *Journal of Econometrics*, 158(1):3–6.

Granger, C. W. and 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.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2):453–497.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press.

Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American statistical Association*, 87(417):58–68.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 221–233, Berkeley, Calif. University of California Press.

Jorion, P. (1996). Value at risk: a new benchmark for measuring derivatives risk. *Irwin Professional Pub*.

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

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101(475):980–990.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, 128(1):137–164.

Kotz, S., Kozubowski, T. J., and Krzysztof, P. (2001). *The Laplace Distribution and Generalizations*. Springer.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2):73–84.

Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series. *Journal of econometrics*, 135(1-2):499–526.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition*. Princeton university press.

Mullen, K., Ardia, D., Gil, D. L., Windover, D., and Cline, J. (2011). Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26.

Newey, W. K. and McFadden, D. (1994). Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*.

Phillips, P. C. B. (1991). A shortcut to lad estimator asymptotics. *Econometric Theory*, 7:450–463.

Poiraud-Casanova, S. and Thomas-Agnan, C. (2000). About monotone regression quantiles. *Statistics and Probability Letters*, 48(1):101–104.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory*, 1(3):295–313.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25:303–325.

Ranga Rao, R. (1962). Relation between weak and uniform convergence of measures with applications. *Annals of Mathematical Statistics*, 33:659–680.

Schmidt, L. D. and Zhu, Y. (2016). Quantile spacings: A simple method for the joint estimation of multiple quantiles without crossing. Technical report, SSRN.

Viscusi, W. K. and Hamilton, J. T. (1999). Are risk regulators rational? evidence from hazardous waste cleanup decisions. *American Economic Review*, 89(4):1010–1027.

Weiss, A. A. (1991). Estimating nonlinear dynamic models using least absolute erorr estimation. *Econometric Theory*, 7:46–68.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press.

White, H., Kim, T.-H., and Manganelli, S. (2010). Modeling autoregressive conditional skewness and kurtosis with multi?quantile caviar. volume 4 of *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, pages 1 – 32. Oxford University Press.

White, H., Kim, T.-H., and Manganelli, S. (2015). Var for var: Measuring tail dependence using multivariate regression quantiles. *Journal of econometrics*, 187(3):169–188.