Semiparametric modeling of multiple quantiles

Published Version:
Catania, L., Luati, A. (2023). Semiparametric modeling of multiple quantiles. JOURNAL OF ECONOMETRICS, 237(2, Part B, December 2023), 1-16 [10.1016/j.jeconom.2022.11.002].

Availability:
This version is available at: https://hdl.handle.net/11585/919799 since: 2023-03-02

Published:
DOI: http://doi.org/10.1016/j.jeconom.2022.11.002

Semiparametric Modeling of Multiple Quantiles

Leopoldo Catania[a], Alessandra Luati[b]

[a] *Aarhus University and CREATES*
[b] *Imperial College London and 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. Theoretical properties of the proposed model are derived such as weak stationarity of the quantile process and consistency 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 then be estimated as in Koenker and Xiao (2006). One of the most successful dynamic quantile models 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 dynamic quantile 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 studied, the interest in dynamic quantile models 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 dynamic quantile models are semiparametric. 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 $q_t^\tau = g(q_{t-1}^\tau, |y_{t-1}|, y_{t-1}^2, |y_{t-1}|^3)$ for an $\mathcal{F}_{t-1}$–measurable function $g(\cdot)$, which we refer to as "the filter". Note that the choice of which transformation of past observations is to be included in $g(\cdot)$ is evidently a personal choice of the modeler and may be selected ad hoc for the application at hand.

Besides the formulation of the filter and the choice of the forcing variables, a second main issue of dynamic quantile models is that they are mainly designed to estimate one single quantile. Turning attention to joint modeling 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) satisfies 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 optimization 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 modeled as mixtures of baseline quantile functions and the quantile regression by Schmidt and Zhu (2016), which is not developed in the times series context.

In this paper, we develop a semiparametric model to track a large number of quantiles of a time series. Differently from available dynamic quantile 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 Koenker and Machado (1999), Poiraud-Casanova and Thomas-Agnan (2000) and Kotz et al. (2001) or in general 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 generalization 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 depends 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 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 lies not only in the gradient-based update of the quantiles dynamics, but also in a specification that ensures that estimated quantiles do not cross. Based upon an idea of Granger (2010), we model differences between adjacent quantiles, rather than quantiles. This is equivalent to specify the dynamics of a reference quantile, namely the median, and then model the distances from the median as positive processes. Such approach is also used by Schmidt and Zhu (2016), in a quantile regression context, who refer to their method as "quantile spacings". The paper is also related to the contribution by White et al. (2015), as we share the same 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 the model is specified, we derive its theoretical properties such as: weak stationarity of the quantile process, limiting quantile values, and consistency of the estimators of the fixed parameters. Finite sample properties of the proposed estimators are also investigated. 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 reports a general discussion on the forcing variable for updating the quantiles in the time series context. Section 3 details the estimation procedure, derives consistency and discusses the challenges related to the asymptotic distribution 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. Proofs are deferred to the Appendix.

2. The dynamic multiple quantile model

Let $Y = \{Y_t\}_{t\in\mathbb{N}}$ be a stationary stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$ where $\mathcal{F} = \{\mathcal{F}_t\}_{t\in\mathbb{N}}$ and $\mathcal{F}_t = \sigma(Y_s, s \leq t)$ is the sigma-algebra generated by the random variables Y_s, $s \leq t$. 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,

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

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,

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

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 set of ordered probability levels $\tau = (\tau_1, \ldots, \tau_J)$, $\tau_i > \tau_j$ if $i > j$, and the vector 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:

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

The empirical counterpart 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 $q_{t+1} = g(q_t, z_t)$, where $z_t = (z_{i,t}, i = 1, \ldots, J)'$ and

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

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 $g(q_t, z_t)$ 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 with

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

and

$$u_t^{\tau_j} \propto \frac{\partial}{\partial \xi_{j,t}} \sum_{k=1}^{J} \rho_{\tau_k}(y_t - q_t^{\tau_k}), \qquad u_t^{\tau_{j^*}} \propto \frac{\partial}{\partial q_t^{\tau_{j^*}}} \sum_{k=1}^{J} \rho_{\tau_k}(y_t - q_t^{\tau_k}),$$

are the martingale difference sequences which drive the dynamics of the time-varying quantiles; finally, $\theta = (\alpha, \beta, \phi, \gamma)'$ are static parameters to be estimated with $|\beta| < 1$ and $|\phi| < 1$. The intercept terms $\bar{q}^{\tau_{j^*}}$ 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 $q_t^{\tau_{j^*}}$ and $\xi_{j,t}$, respectively, see Section 2.2.1. Throughout the paper, we assume that the initial values y_0 and $q_0 = (q_0^{\tau_1}, \dots, q_0^{\tau_J})'$ are known constants.

In practice, each quantile is defined as the reference quantile plus or minus a number of positive processes. For example, for $J = 2$, fixing $q_t^{\tau_1^*}$ as the reference quantile with dynamics described by equation (7), then $q_t^{\tau_2} = q_t^{\tau_1^*} + \eta_{t2}$. On the other hand, for $J = 2m + 1$, by setting the reference quantile as $q_t^{\tau^*_{m+1}}$, then $q_t^{\tau_1} = q_t^{\tau^*_{m+1}} - \sum_{i=1}^{m} \eta_{ti}$, $q_t^{\tau_2} = q_t^{\tau^*_{m+1}} - \sum_{i=1}^{m-1} \eta_{ti}$ and so on and so forth. In the latter example, the case $m = 49$ corresponds to the case of modeling the percentiles with the median as the reference quantile. Although quantiles that lie on the left and right of the reference quantile follow a dynamic that is based upon the same parameters (γ and ϕ), the dynamic equation for each quantile turns out to be quantile-specific and incorporates the information coming from all the regions of the distribution.

The definition of the forcing variables $u_t^{\tau_j}$ is crucial for the specification of the model and depends on its parameterization. We note that in our case $q_t^{\tau_j} = g((q_t^{\tau_{j^*}}, \xi_t), z_t)$, where $\xi_t = (\xi_{j,t}, j = 1, \dots, J, \quad j \neq j^*)'$, such that, by an application of the chain rule, we obtain:

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

and analogously for $j = j^*$, which yields:

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

where $b_j = \mathbb{1}(\tau_j < \tau_{j^*}) - \mathbb{1}(\tau_j > \tau_{j^*})$. Note that the sequence of hit variables $\{z_t\}_{t \in \mathbb{N}}$ is independent and identically distributed (iid).[1] As a consequence $\{u_t^{\tau_j}\}_{t \in \mathbb{N}}$ is also an iid sequence of zero mean

[1] The random variable $z_t | \mathcal{F}_{t-1}$ takes 2^J possible outcomes but only $J + 1$ of them have positive probability. Indeed

$$P(z_t = z | \mathcal{F}_{t-1}) = \begin{cases} \tau_1, & \text{if } z = [(1 - \tau_i), i = 1, \dots, J] \\ \tau_{j+1} - \tau_j, & \text{if } z = [-\tau_1, \dots, -\tau_j, (1 - \tau_j), \dots, (1 - \tau_J)] \\ 1 - \tau_J, & \text{if } z = [-\tau_i, i = 1, \dots, J] \\ 0, & \text{otherwise,} \end{cases}$$

random variables with constant variance

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

where $E_{t-1}[\cdot] = E[\cdot|\mathcal{F}_{t-1}]$ and

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

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 $a_j = \sqrt{\varpi_j}$.[2] 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 which we now set to the median, i.e. $u_t^{\tau_{j^*}}$, with $\tau_{j^*} = 0.5$. This quantity is proportional to

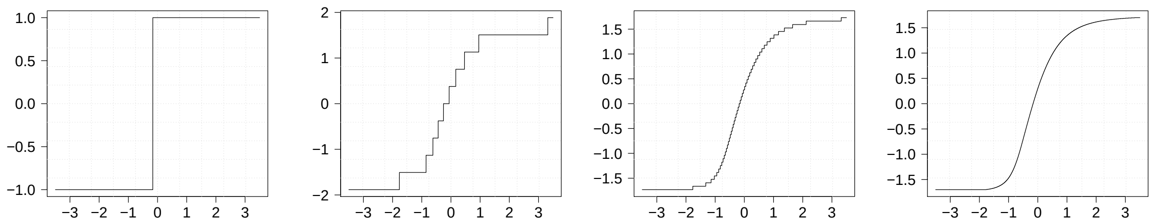
$$u_t^{\tau_{j^*}} \propto \sum_{j=1}^{J} \mathbb{1}(y_t < q_t^{\tau_j}) \propto \widehat{F}_{t|t-1}^{J}(y_t),$$

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 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

so that $P(z_t = z|\mathcal{F}_{t-1}) = g(z)$ for all t. The fact that this conditional probability mass function does not depend on time implies both independence over time of z_t and $P(z_t = z) = g(z)$ for all t, such that $\{z_t\}_{t\in\mathbb{N}}$ is iid.

[2]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

parameter equal to 1.5 and shape parameter equal to 3.5. The forcing variable for other quantiles is proportional to

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

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 $E_t[q_{t+1}]$ 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 multi-step ahead prediction for other quantiles is given by

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

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

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

By letting $h \to \infty$ we obtain the unconditional mean of $\eta_{j,t}$ as

$$E[\eta_{j,t}] = \begin{cases} \exp(\bar{\xi}_j) \prod_{s=0}^{\infty} \exp\{-a_j^{-1}\gamma\phi^s \sum_{l=1}^{j} \tau_l\} \sum_{l=0}^{j} h(\tau_l) \exp\{a_j^{-1}\gamma\phi^s (j-l)\}, & \text{if } \tau_j < \tau_{j^*} \\ \exp(\bar{\xi}_j) \prod_{s=0}^{\infty} \exp\{a_j^{-1}\gamma\phi^s \sum_{l=j}^{J} \tau_l\} \sum_{l=j-1}^{J} g(\tau_l) \exp\{-a_j^{-1}\gamma\phi^s (J-l)\}, & \text{if } \tau_j > \tau_{j^*} \end{cases}$$

Furthermore, we have that quantile differences have finite moments, i.e.

$$E_t[\eta_{j,t}^m] < \infty,$$

for all t since $\eta_{j,t}$ has bounded support, such that $\sup_{\omega\in\Omega} |q_t^{\tau_j}| \le M < \infty$, a.s. for all t, j.

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 for Y_t. Specifically, we can estimate the unconditional quantiles $\hat{\bar{q}}^{\tau_{j^*}}$ from the time series $(y_1, \ldots, y_T)'$, and set $\bar{q}^{\tau_{j^*}} = \hat{\bar{q}}^{\tau_{j^*}}$ and:

$$\bar{\xi}_j = \log(\widehat{\Delta}_j) + \frac{b_j\gamma}{1-\phi}\sum_{i=1}^{j}\frac{\tau_i}{a_j} + \kappa_j,$$

where

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

and

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

It is important to note that in the following section it is implicitly assumed that $\bar{q}^{\tau_{j^*}}$ and $\bar{\xi}_j$ for all $j \neq j^*$ are known. In empirical applications, these quantities are of course estimated from the data as a function of the $\hat{\bar{q}}^{\tau_j}$.

3. Estimation

Let $\{y_t\}_{t=1,\dots,T}$ be an observed time series. Parameter estimates are obtained as the solution of the following minimization problem,

$$\widehat{\theta}_T = \arg\min_{\theta\in\Theta} \sum_{t=1}^{T}\sum_{j=1}^{J} \rho_{\tau_j}(y_t - q_t^{\tau_j}(\theta)), \tag{12}$$

where $\Theta \subset \mathbb{R}^m$, $m = 4$ and $q_t^{\tau_j}(\theta)$ is the j-th quantile process at the parameter hypothesis θ. 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), in reference to a private correspondence between Robert Hogg and himself in 1979.

Consistency (Theorem 1) of $\widehat{\theta}_T$ for estimating $\theta_0 = \arg\min_{\theta\in\Theta} \sum_{j=1}^{J} E[\rho_{\tau_j}(y_t - q_t^{\tau_j}(\theta))]$ is established based on the following assumptions.

Assumption 1. $\{Y_t\}_{t\in\mathbb{N}}$ is a stationary and ergodic (s.e.) stochastic process and $\exists\theta_0 \in \Theta$, $\theta_0 = (\alpha_0, \beta_0, \gamma_0, \phi_0)'$ with $\alpha_0 \neq 0$, $\gamma_0 \neq 0$, $|\phi_0| < 1$ and $|\beta_0| < 1$ such that $F_{t|t-1}(q_t^{\tau_j}(\theta_0)) = \tau_j$ for $\tau_j \in (0,1)$, $j = 1,\dots,J$, and $\forall t \in \mathbb{N}$, where $F_{t|t-1}(y) = P(Y_t \leq y|\mathcal{F}_{t-1})$.

Assumption 2. The initial values Y_0 and q_0 are fixed and known. The parameter space $\Theta \subset \mathbb{R}^m$, $m = 4$, is a compact set such that $\theta_0 \in \Theta$.

As the initial values y_0 and q_0 are known, $q_t^{\tau_j}(\theta_0) = q_t^{\tau_j}$, where $q_t^{\tau_j}$ is defined in equation (1). Stationarity and ergodicity of $\{Y_t\}_{t\in\mathbb{Z}}$ (Assumption 1) or similar heterogeneity restrictions are common assumptions in the literature on dynamic quantile models, see Engle and Manganelli (2004), Gouriéroux and Jasiak (2008), White et al. (2015), and Patton et al. (2019). Assumption 2 (known initial conditions) is also common in the dynamic quantile literature, though often implicit, as in Engle and Manganelli (2004), White et al. (2015), and Patton et al. (2019). This is in contrast with the score driven literature, where asymptotic negligibility of initial conditions in general is proved under high level theoretical conditions on the Lipschitz coefficient associated with the filter, see Blasques et al. (2018). In our context, such assumptions cannot be easily relaxed due to the non continuity of the filter, which does not allow us to directly apply the theory developed by Bougerol (1993). The latter constitutes a basic building block for many invertibility results in dynamic non-linear models.

Assumption 3. $E|Y_t| < \infty$.

Assumption 4. $\{Y_t \mathbb{1}(Y_t \leq q_t^{\tau_j}(\theta)\}_{t\in\mathbb{N}}$ and $\{q_t^{\tau_j}(\theta) z_{j,t}(\theta)\}_{t\in\mathbb{N}}$ obey a uniform law of large numbers (ULLN) $\forall j = 1, \ldots, J$. In addition, $\mathbb{E}[\rho_{\tau_j}(y_t - q_t^{\tau_j}(\theta))]$ is a continuous function of θ for all $j = 1, \ldots, J$.

Assumption 5. $f_{t|t-1}(x)$ is bounded and Lipschitz continuous, i.e. $f_{t|t-1}(x) \leq L$ and $|f_{t|t-1}(x) - f_{t|t-1}(y)| \leq M|x - y|$, with $L, M < \infty$, for all $t \in \mathbb{N}$.

Theorem 1. *Let Assumptions 1-5 hold. Then* $\widehat{\theta}_T \to_{a.s.} \theta_0$.

The proof is in the Appendix and is a canonical proof of consistency for M-estimators in correctly specified non-linear dynamic models with parameter defined on a compact parameter space, see e.g. Pötscher and Prucha (2001). Besides Assumptions 1 and 2, the proof requires Assumption 4, a ULLN, that we assume due to the non-smoothness of the quantile model; it is, however, quite common in quantile models, see Engle and Manganelli (2004) and Patton et al. (2019). Absolute integrability (Assumption 3) can be relaxed as in Amemiya (1982), Powell (1984), and Weiss (1991), based on a modified objective function which attains the same minimum as the original for each T. Assumption 5 is a standard regularity condition on the conditional density of Y_t.

As far as the asymptotic distribution of the estimator is concerned, the setting is highly non-standard, in that neither the criterion function nor the filter satisfy the regularity conditions required for the usual first order Taylor expansions to apply. The non-smoothness 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 valid for least absolute deviation estimators with iid data, see also Newey and McFadden (1994). The basic idea is to approximate the almost everywhere differentiable loss function with its smooth expectation. Strong consistency of the estimator is usually required, a condition that is satisfied in our context by Theorem 1. However, all the above mentioned papers (including Patton et al. (2019)) are grounded on the assumption that the estimator is (almost everywhere) a twice continuously differentiable function of the parameter, which is clearly not the case of our filter.

4. Simulation study

We report the summary results of a simulation study aimed 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 is 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 parameter choices.

	$T = 250$			$T = 500$			$T = 1000$			$T = 2500$			$T = 5000$		
	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: Mean of the estimated parameters, bias, and root mean squared error (RMSE) with respect to the true values, that 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 $T = 5000$.

Table 1 reports the mean of the estimated parameters as well as their bias and root mean squared error (RMSE) with respect to the true values. When the sample size is small ($T = 250$), there is some bias in the estimate of β and 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 be the most difficult to estimate with a (small) bias that vanishes only when the sample size is large ($T = 5000$). For all parameters the bias and the RMSE decreases when the sample size increases. Overall, we conclude that the Hogg estimator in equation (12) reports remarkable results in finite samples.

5. Empirical illustration

We illustrate the DMQ model using the series of daily logarithmic returns, in percentage points, of Microsoft Corporation spanning from December 8, 2010 to November 15, 2018 for a total of $T = 2000$ observations. Observations are reported in Figure 4a 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% significance level. Gaussianity of the unconditional distribution is rejected according to the Jarque and Bera test which reports a statistic of 200. The empirical skewness and kurtosis coefficients are -0.42 and 4.31, respectively, indicating that the empirical distribution is left skewed and with tails which are heavier than the Gaussian ones.

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 ($\alpha = \beta = 0$). Thus, in this case only two parameters (γ and ϕ) are estimated in the DMQ model. Throughout the analysis we will employ two benchmark models: *i*) the Multiple Quantile Conditional Autoregressive Value–at–Risk (MQCAViaR) model detailed in White et al. (2015, Eq. 4), and *ii*) the ARMA(P,Q)–GARCH(p,q) model of Bollerslev (1986) with skew Student's t distributed errors.

The Multiple Quantile CAViaR (MQCAViaR) model of White et al. (2015, Eq. 4) is defined as:

$$q_t^{\tau_j} = \Psi_{t-1}' \boldsymbol{\beta}_j + q_{t-1}{\prime}\gamma_j, \qquad j = 1, \dots, J, \tag{13}$$

where, as before, $q_t = (q_t^{\tau_1}, \dots, q_t^{\tau_J})'$, and we set $\Psi_t = (1, y_{t-1}\mathbb{1}(y_{t-1} < 0), y_{t-1}\mathbb{1}(y_{t-1} > 0))'$ following the "asymmetric slope" specification as in Engle and Manganelli (2004). Model parameters in the MQCaViaR are $\beta_j = (\beta_{1,j}, \beta_{2,j}, \beta_{3,j})'$ and $\gamma_j = (\gamma_{j,1}, \dots, \gamma_{j,J})'$, for a total of $3J + J^2$ free parameters. In our case of $J = 99$ probability levels, the MQCaViaR cannot be estimated unless some parameter constraint is imposed. We set $\gamma_{j,l} = 0$ for $l \neq j$ such that (13) can be written as $q_t^{\tau_j} = \Psi_t' \beta_j + \gamma_j q_{t-1}^{\tau_j}$ where $\gamma_j = \gamma_{j,j}$ for $j = 1, \dots, J$. This constraint rules out cross quantiles dependence in the MQCaViaR model and allows for separate estimation of the parameters of each quantile level. Essentially, MQCaViaR reduces to J independent CaViaR models as in Engle and Manganelli (2004). Given its linear structure, MQCAViaR 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. In an unreported analysis we have also considered the unconstrained MQCaViaR specification (i.e. with cross quantile dependence) setting $J = 4$ with $\tau = (0.05, 0.25, 0.75, 0.95)'$. This model has 28 parameters that must be jointly estimated by minimizing the Hogg loss function (12), giving rise to a complex optimization problem. Our analysis indicates that the constrained specification performs better in quantile prediction, and it is thus retained in our empirical analysis.

The ARMA(P,Q)–GARCH(p,q) is instead a location scale parametric model which assumes that $y_t = \mu_t + \varepsilon_t, \varepsilon_t = \sigma_t z_t$, where

$$\mu_t = \varpi + \sum_{i=1}^{P} \phi_i y_{t-j} + \sum_{l=1}^{Q} \theta_i \varepsilon_{t-l}, \qquad \sigma_t^2 = \omega + \sum_{m=1}^{p} \alpha_m \varepsilon_{t-m}^2 + \sum_{n=1}^{q} \zeta_n \sigma_{t-n}^2,$$

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 Maximum Likelihood, see Bollerslev (1987).

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

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

for $j = 1, \ldots, J$, which is obtained by modeling 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 MQCAViaR, the DQ model cannot guarantee the monotonicity of the filtered quantiles. Also, note that the DQ model provides an updating mechanism which is related 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, MQCAViaR, ARMA-GARCH, and DQ models on the full sample of observations.[3] Filtered quantiles for the subset of probability levels $5\%, 10\%, \ldots, 95\%$ are reported in Figure 2. The graphical investigation suggests that DMQ quantiles are smoother than those reported by MQCAViaR 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. MQCAViaR 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 a 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 coverage and independence of the quantile violations. See Kupiec (1995) and Christoffersen (1998)

[3] Estimation of DMQ, MQCAViaR, 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).

for further details on the UC and CC tests, respectively. Figure 3 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 MQCAViaR 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 are only two cases when the null hypothesis is rejected: 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 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 positive 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:

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

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 4. Specifically, Figure 4a 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 registered during the periods of market instability associated with the European sovereign debt crisis of 2010–2011. The filtered variance is reported in Figure 4b 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 adapting well to periods of higher volatility. Figures 4c and 4d 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 November 26, 2014. Model 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. 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 10000 simulations. In the MQCAViaR model defined in equation (13), there are no ways to produce multi-step 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:

$$q_{t+h}^{\tau_j} = \Psi_t' \boldsymbol{\beta}_j + \gamma_j q_{t+h-1}^{\tau_j}, \tag{14}$$

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 10 MQCAViaR specifications, which is even more computational expensive than approximating the multi-step ahead distribution of the ARMA–GARCH specification.

We report our results in the forms of aggregated quantile losses (AQL) computed over the out of sample period for different regions of the h-step ahead predictive distribution and defined as

$$AQL(h, \underline{j}, \overline{j}) = \sum_{t=S+h}^{T} \sum_{j=\underline{j}}^{\overline{j}} \rho_{\tau_j}(y_t - \widehat{q}_{t|t-h}^{\tau_j}),$$

where $\underline{j} \leq \overline{j}$, $S = 1000$ is the length of the in sample period, and $\widehat{q}_{t|t-h}^{\tau_j} = E_{t-h}[q_t^{\tau_j}]$ is the h-step ahead prediction of the τ_j level quantile. The values of $\underline{j}$ and $\overline{j}$ allow us to focus on different regions of the predictive distribution. Table 2 reports the aggregated losses for subsets of quantile levels for $h = 1$, $h = 5$, and $h = 10$. The first ten rows of Table 2 report results for $(\underline{j}, \overline{j})$ equals to $(1, 9), (10, 19), \ldots, (90, 99)$, while in the last row we consider the full support $(1, 99)$. 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 MQCAViaR, 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

	$h=1$				$h=5$				$h=10$			
	DMQ	DQ	MQCAViaR	A–G	DMQ	DQ	MQCAViaR	A–G	DMQ	DQ	MQCAViaR	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

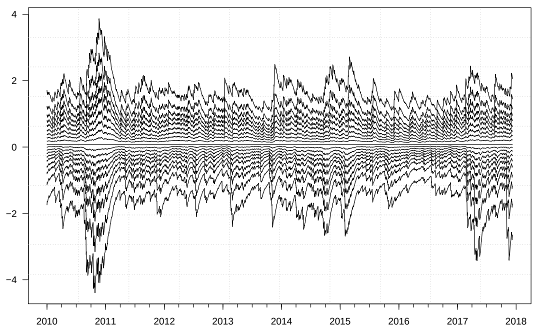
Table 2: Average quantile loss over the out of sample period. Losses are aggregated over subset of quantiles. For example, the first row reports the aggregated losses for quantiles associated with probability levels in the range $[0.01, 0.1)$. Results are reported for one ($h = 1$), five ($h = 5$), and 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 belong to the Superior Set of Models computed according to the Model Confidence Set procedure of Hansen et al. (2011) at the 75% confidence level.

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 $\tau_j \leq 0.5$ ($\underline{j} = 1$, $\overline{j} = 50$), right tail $\tau_j \geq 0.5$ ($\underline{j} = 50$, $\overline{j} = 99$), center $0.25 \leq \tau_j \leq 0.75$ ($\underline{j} = 25$, $\overline{j} = 75$), as well as for the whole distribution $0.01 \leq \tau_j \leq 0.99$ ($\underline{j} = 1$, $\overline{j} = 99$). Also in this case, results indicate that DMQ losses are on average lower than MQCAViaR, ARMA–GARCH, and DQ, irrespectively of the distribution region and the forecast horizon. The MCS procedure also supports these findings.

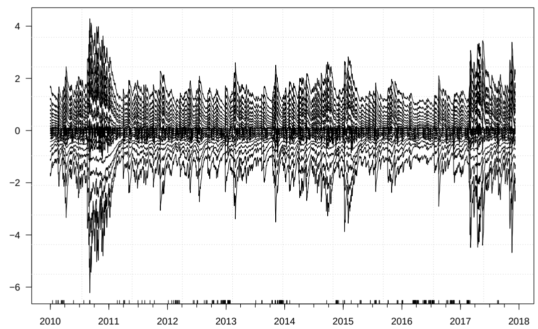
To conclude our illustration, we compute the UC and CC tests as in the full sample analysis, but considering one-step ahead predictions. P-values for all quantile levels are reported in Figure 5. Interestingly, p-values associated with MQCAViaR, ARMA–GARCH, and DQ predictions are lower than in the full 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, MQCAViaR, 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 modeling of multiple quantiles in a score driven framework is required.

5.4. The choice of J

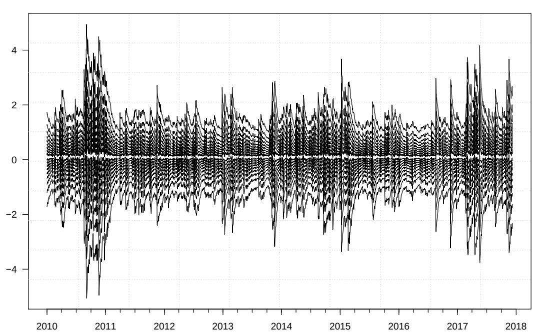
We conclude the empirical analysis by investigating how the choice of J, the number of quantiles, affects quantile prediction. To this end, we make predictions of the nine deciles $q_t^{10\%}, \ldots, q_t^{90\%}$ with



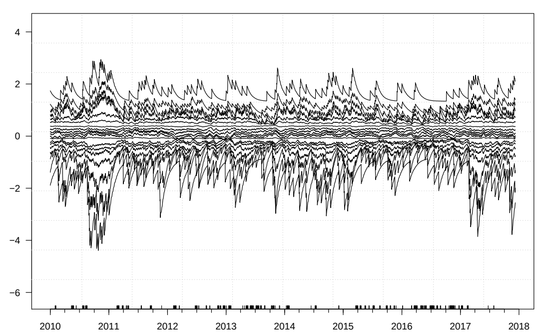
(a) DMQ



(b) MQCAViaR



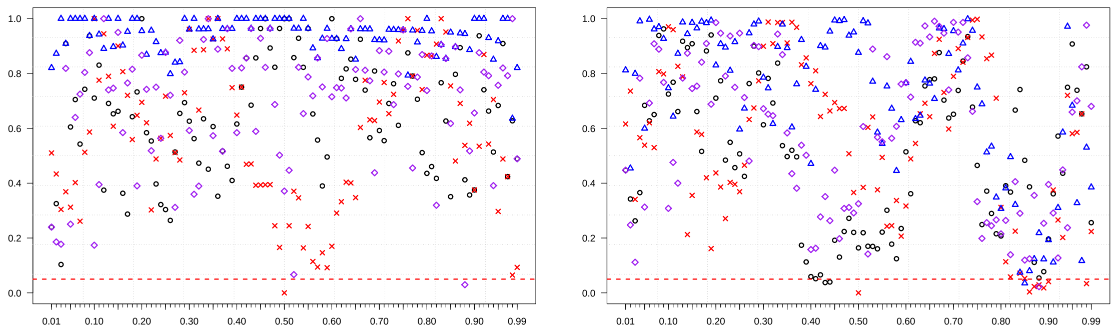
(c) ARMA-GARCH



(d) DQ

Figure 2: Filtered quantiles from the DMQ (a), MQCAViaR (b), ARMA-GARCH (c), and DQ (d) models at levels 5%, 10%,..., 95% for Microsoft Corporation from December 8, 2010 to November 15, 2018 for a total of $T = 2000$ observations. Rugs in panel (b) indicate points in time when MQCAViaR quantiles cross.

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 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 joint



(a) UC p-values

(b) CC p-values

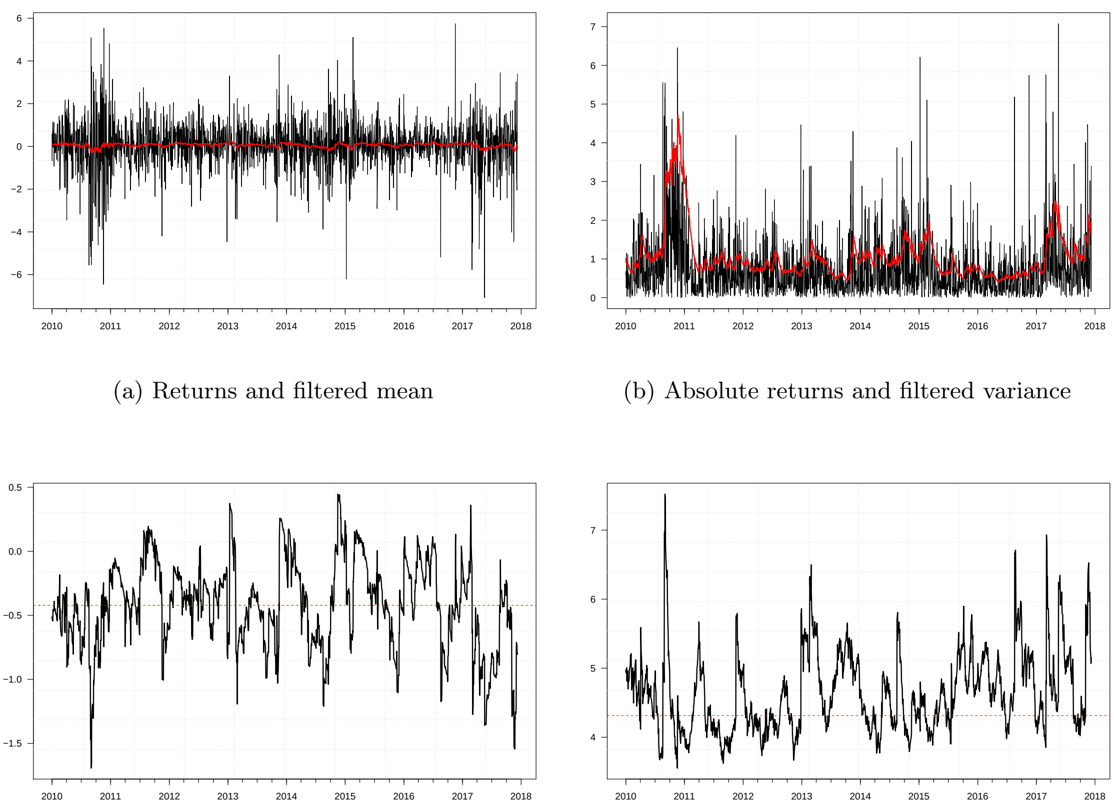
Figure 3: P-values of the UC (panel a) and CC (panel b) tests over the in sample period for $\tau_1 = 1\%$, $\tau_2 = 2\%$, ..., $\tau_{99} = 99\%$ quantile levels of DMQ (black circles), MQCAViaR-C (blue triangles), ARMA-GARCH (red crosses), and DQ (purple diamonds). The horizontal red dashed line represents the 5% probability level.

modeling 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 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, better results are most likely obtained by increasing J.

5.5. The choice of τ_{j^*}

Together with J, in the DMQ model the reference quantile must be chosen. In our empirical analysis with financial returns we have set $\tau^* = 0.5$ (i.e. the median) and further imposed that $\alpha = \beta = 0$ such that $q_t^{\tau_{j^*}} = \bar{q}^{\tau_{j^*}}$. We now consider the unrestricted DMQ specification where α and β are estimated, and investigate the effect of selecting the reference quantile among the $J = 99$ possible choices. We do so by estimating DMQ with $j^* = 1, \ldots, J$ and compare the value of the objective function evaluated at its minimum. Figure 6 reports the values of objective function (12) at its minimum divided by T for different choices of τ_{j^*}. The figure shows that our choice of the reference quantile is indeed appropriate, resulting in lower values of the objective function. More generally, one might consider the QMLE estimator of j^* by solving

$$\widehat{j^*} = \underset{j^*=1,\ldots,J}{\arg\min} \left\{ \underset{\theta \in \Theta}{\arg\min} \sum_{t=1}^{T} \sum_{j=1}^{J} \rho_{\tau_j}(y_t - q_t^{\tau_j}(\theta, j^*)) \right\},$$



(a) Returns and filtered mean

(b) Absolute returns and filtered variance

(c) Filtered skewness coefficient

(d) Filtered kurtosis coefficient

Figure 4: 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.

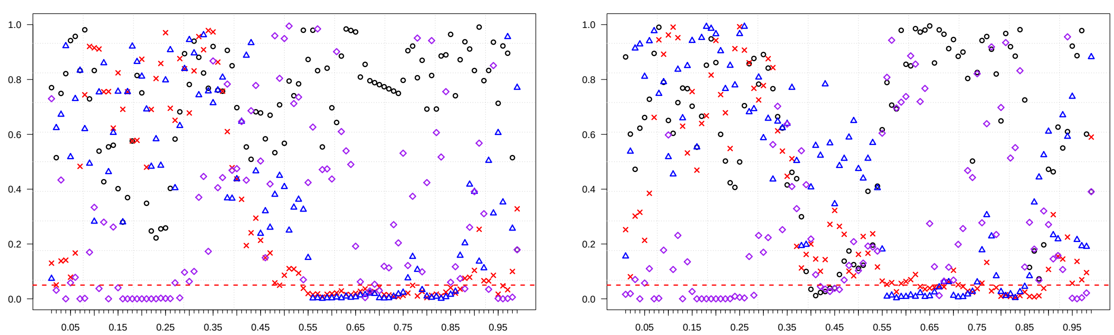
where $q_t^{\tau_j}(\theta, j^*)$ now refers to the filtered quantile according to the parameter hypothesis θ and reference quantile τ_{j^*}. In our case, we find $\tau_{\widehat{j^*}} = 0.54$, which is close to what we used in our analysis.

6. Conclusions

The paper poses the basis for a new way of modeling 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

	Left tail				Right tail				Center			
	DMQ	DQ	MQCAViaR	A–G	DMQ	DQ	MQCAViaR	A–G	DMQ	DQ	MQCAViaR	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 ($0.01 \leq \tau_j \leq 0.99$) parts of the conditional 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 belong 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 5: 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), MQCAViaR (blue triangles), ARMA-GARCH (red crosses), and DQ (purple diamonds). The horizontal red dashed line represents the 5% probability level.

predictions and their limiting values, *iii*) consistency 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 $u_{j,t}$, *iii*) the case when $u_{j,t}$ is not $\mathcal{F}_{t-1}$–measurable, and *iv*) the asymptotic distribution of the estimator.

	$J=1$	$J=9$	$J=19$	$J=29$	$J=39$	$J=49$	$J=59$	$J=69$	$J=79$	$J=89$	$J=99$
$\tau=0.1$	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
$\tau=0.5$	1.000	1.002	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 belong to the Superior Set of Models computed according to the Model Confidence Set procedure of Hansen et al. (2011) at the 75% confidence level.

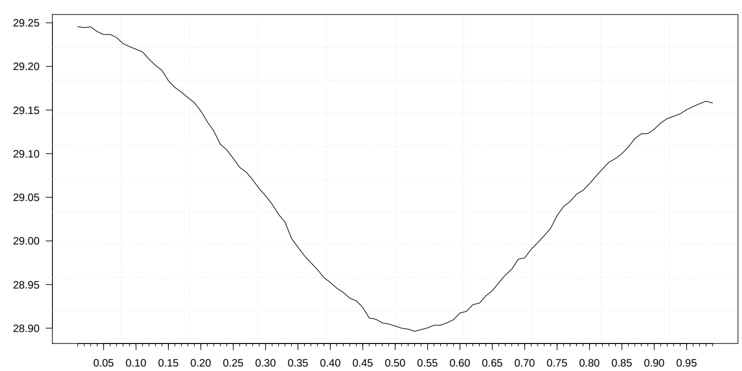


Figure 6: Values of the objective function (12) at its minimum divided by T for different choices of τ_{j^*}. The horizontal axis reports the value of τ_{j^*}.

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 modeling US inflation, which is known to be a very persistent process. Also in that case, the results have been found very promising. Future research would be to study the performance of DMQ also in economic time series.

Appendix

Let us define $m_t = (\xi_{1,t}, \ldots, \xi_{j^*-1,t}, q_t^{\tau_{j^*}}, \xi_{j^*+1,t}, \ldots, \xi_{J,t})'$ and the vector-valued function $h(\cdot)$, such that $q_t = h(m_t)$, with generic element $h_j(\cdot)$ equal to the map induced by (6)–(8), i.e. $q_t^{\tau_j} = h_j(m_t)$. One has:

$$m_{t+1} = (\mathbf{I} - \mathbf{B}_0)\,\bar{m} + \mathbf{A}_0\mathbf{N}z_t + \mathbf{B}_0 m_t, \tag{15}$$

where $\mathbf{I}$ is the $J \times J$ identity matrix, $\bar{m} = (\bar{\xi}_1, \ldots, \bar{\xi}_{j^*-1}, \bar{q}^{\tau_{j^*}}, \bar{\xi}_{j^*+1}, \ldots, \bar{\xi}_J)'$, $\mathbf{N}$ is a selection matrix, $\mathbf{A}_0$ and $\mathbf{B}_0$ are $J \times J$ diagonal matrices with $\mathbf{A}_{0,j^*j^*} = \alpha_0$, $\mathbf{B}_{0,j^*j^*} = \beta_0$, $\mathbf{A}_{0,jj} = \gamma_0$, $\mathbf{B}_{0,jj} = \phi_0$ for $j \neq j^*$, and $z_t = (z_{1,t}, \ldots, z_{J,t})'$.

Let us denote the loss functions and corresponding minimizers as $Q_T(\theta) = \sum_{t=1}^T \sum_{j=1}^J \rho_{\tau_j}(Y_t - q_t^{\tau_j}(\theta))$, with $\widehat{\theta}_T = \underset{\theta\in\Theta}{\arg\min}\ Q_T(\theta)$ and $Q_0(\theta) = \sum_{j=1}^J \mathbb{E}\left[\rho_{\tau_j}(Y_t - q_t^{\tau_j}(\theta))\right]$, with $\theta_0 = \underset{\theta\in\Theta}{\arg\min}\ Q_0(\theta)$.

Proof of Theorem 1

The proof is divided in two parts which consist in showing that: *i*) $\sup_{\theta\in\Theta} |\frac{1}{T}Q_T(\theta) - Q_0(\theta)| \to 0$ a.s. for $T \to \infty$ and *ii*) $Q_0(\theta)$ is uniquely minimized at θ_0.

Let us consider part *i*). The sequence $\{\rho_{\tau_j}(Y_t - q_t^{\tau_j}(\theta))\}_{t\in\mathbb{N}}$ is s.e. since so is $\{(Y_t, q_t^{\tau_j}(\theta))\}_{t\in\mathbb{N}}$ by Assumption 1. Furthermore $E[|\rho_{\tau_j}(Y_t - q_t^{\tau_j}(\theta))|] < \infty$ as $E[|Y_t|] < \infty$ (Assumption 3) and $q_t^{\tau_j}(\theta)$ is bounded (Assumption 2). One has

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}Q_T(\theta) - Q_0(\theta)\right| \leq \sum_{j=1}^J \left(A_{1,j,T} + A_{2,j,T} + A_{3,j,T}\right),$$

where

$$\begin{aligned}
A_{1,j,T} &= \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^T \mathbb{1}(Y_t \leq q_t^{\tau_j}(\theta))Y_t - E[\mathbb{1}(Y_t \leq q_t^{\tau_j}(\theta))Y_t]\right| \\
A_{2,j,T} &= \tau_j\left|\sum_{t=1}^T \frac{1}{T}Y_t - E[Y_t]\right| \\
A_{3,j,T} &= \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^T q_t^{\tau_j}(\theta)z_{j,t}(\theta) - E[q_t^{\tau_j}(\theta)z_{j,t}(\theta)]\right|.
\end{aligned}$$

We have that $A_{1,j,T} \to 0$ and $A_{3,j,T} \to 0$ a.s., due to (A4) and $A_{2,j,T} \to 0$ by (A3) and an application of the law of large numbers.

The proof of part *ii*) is analogous to the last point of the proof of Theorem 1 of White et al. (2015) and consists in showing that, for $\theta \neq \theta_0$ and $|\theta - \theta_0| > \varepsilon$, $\Delta(\theta) > 0$ where $\Delta(\theta) = \sum_{j=1}^J E[\Delta_{j,t}(\theta)]$ and $\Delta_{j,t} = \rho_{\tau_j}(Y_t - q_t^{\tau_j}(\theta)) - \rho_{\tau_j}(Y_t - q_t^{\tau_j}(\theta_0))$, and $\varepsilon_{j,t} = Y_t - q_t^{\tau_j}(\theta_0)$. Let $\delta_{j,t}(\theta, \theta_0) = q_t^{\tau_j}(\theta) - q_t^{\tau_j}(\theta_0)$,

$\varepsilon_{j,t} = Y_t - q_t^{\tau_j}(\theta_0)$ and let $f^{\varepsilon}_{j,t|t-1}(\cdot)$ denote the conditional density of $\varepsilon_{j,t}$ given $\mathcal{F}_{t-1}$. By Assumption 5, using the same arguments of White et al. (2015), we obtain the following inequality:

$$E[\Delta_{j,t}(\theta)] \geq \frac{1}{2}\delta_\varepsilon^2 E[\mathbb{1}(|\delta_{j,t}(\theta,\theta_0)| > \delta_\varepsilon)],$$

where $\delta_\varepsilon > 0$ comes from $f^{\varepsilon}_{j,t|t-1}(s) > \delta_\varepsilon$ for $|s| < \delta_\varepsilon$ (A5). Then:

$$\Delta(\theta) = \sum_{j=1}^{J} E[\Delta_{j,t}(\theta)] \geq \frac{1}{2}\delta_\varepsilon^2 \sum_{j=1}^{J} E[\mathbb{1}(|\delta_{j,t}(\theta,\theta_0)| > \delta_\varepsilon)] = \frac{1}{2}\delta_\varepsilon^2 \sum_{j=1}^{J} P(|\delta_{j,t}(\theta,\theta_0)| > \delta_\varepsilon),$$

such that if $P(|q_t^{\tau_j}(\theta) - q_t^{\tau_j}(\theta_0)| > \delta_\varepsilon) > 0$ for all $\delta_\varepsilon > 0$ then $\Delta(\theta) > 0$. A sufficient condition for $\Delta(\theta) > 0$ is $q_t(\theta) - q_t(\theta_0) = 0$ a.s. only if $\theta = \theta_0$. Note that $q_t(\theta) = q_t(\theta_0)$ a.s. is equivalent to $m_t(\theta) = m_t(\theta_0)$ a.s., since $h(\cdot)$ is continuous and measurable and $q_t(\theta)$ is s.e. for all $\theta \in \Theta$. We proceed by contradiction, as in Section 5.1 of Straumann and Mikosch (2006). Let $\theta \neq \theta_0$ and assume that $m_t(\theta) = m_t(\theta_0)$ a.s., then:

$$0 = m_t(\theta) - m_t(\theta_0) = (\mathbf{B}(\theta) - \mathbf{B}(\theta_0))\bar{m} + (\mathbf{A}(\theta) - \mathbf{A}(\theta_0))\mathbf{N}z_{t-1} + (\mathbf{B}(\theta) - \mathbf{B}(\theta_0))m_{t-1},$$

where $\mathbf{B}(\theta_0) = \mathbf{B}_0$ and $\mathbf{A}(\theta_0) = \mathbf{A}_0$ in Equation (15), and we have exploited the fact that $m_{t-1}(\theta) = m_{t-1}(\theta_0) = m_{t-1}$. Now, if $\mathbf{B}(\theta) \neq \mathbf{B}(\theta_0)$, then the random variable m_{t-1} would be at the same time a measurable function of z_{t-1} and independent of z_{t-1}. By Lemma 5.4.2 of Straumann (2005) we conclude that m_{t-1} must be deterministic. However, simple evaluations show that, since the diagonal elements of $\mathbf{A}(\theta_0)$ are non-zero (Assumption 2), the variance of m_{t-1} has non-zero diagonal elements. So it must be that $\mathbf{B}(\theta) = \mathbf{B}(\theta_0)$. Furthermore, if $\mathbf{A}(\theta) \neq \mathbf{A}(\theta_0)$ we have $(\mathbf{A}(\theta) - \mathbf{A}(\theta_0))\mathbf{N}z_{t-1} = 0$ a.s. Let $l_{j,t-1}$ be the generic element of $(\mathbf{A}(\theta) - \mathbf{A}(\theta_0))\mathbf{N}z_{t-1}$, we have from equation (9) that, up to a non-zero constant:

$$l_{j,t-1} \propto \begin{cases} \sum_{i=1}^{j} \delta^a_{j,i}\mathbb{1}(Y_{t-1} \leq q_{t-1}^{\tau_j}) - \sum_{i=1}^{j} \delta^a_{j,i}\tau_i, & \text{if } j < j^* \\ \sum_{i=1}^{J} \delta^a_{j,i}\mathbb{1}(Y_{t-1} \leq q_{t-1}^{\tau_j}) - \sum_{i=1}^{J} \delta^a_{j,i}\tau_i, & \text{if } j = j^* \\ \sum_{i=j}^{J} \delta^a_{j,i}\mathbb{1}(Y_{t-1} \leq q_{t-1}^{\tau_j}) - \sum_{i=j}^{J} \delta^a_{j,i}\tau_i, & \text{if } j > j^*, \end{cases}$$

such that $l_{j,t-1} = 0$ a.s. for all j only if $\sum_{i=1}^{j} \delta^a_{j,i}\mathbb{1}(Y_{t-1} \leq q_{t-1}^{\tau_j}) = \sum_{i=1}^{j} \delta^a_{j,i}\tau_i$, $\sum_{i=1}^{J} \delta^a_{j,i}\mathbb{1}(Y_{t-1} \leq q_{t-1}^{\tau_j}) = \sum_{i=1}^{J} \delta^a_{j,i}\tau_i$, and $\sum_{i=j}^{J} \delta^a_{j,i}\mathbb{1}(Y_{t-1} \leq q_{t-1}^{\tau_j}) = \sum_{i=j}^{J} \delta^a_{j,i}\tau_i$ a.s. Since Y_{t-1} is continuously distributed (Assumption 5) and $q_{t-1}^{\tau_j} = q_{t-1}^{\tau_i}$ only if $i = j$, the variance of the quantities on the left hand side of the previously reported equalities are all positive if at least one of the associated $\delta^a_{j,i}$ is different from zero, so that necessarily $\mathbf{A}(\theta) = \mathbf{A}(\theta_0)$. We then conclude that $q_t(\theta) - q_t(\theta_0) = 0$ a.s. only if $\theta = \theta_0$, such that $\Delta(\theta) > 0$ for all $\theta \neq \theta_0$. As θ is arbitrary, it follows from parts *i)*–*ii)* and continuity of $Q_0(\theta)$ (Assumption 4) that $\widehat{\theta}_T \to \theta_0$ a.s. for $T \to \infty$.

Acknowledgments

We would like to thank the Editor, Serena Ng, the Associate Editors and the referees for their insightful comments. The paper has been presented at the Workshop on score driven time series models at the University of Cambridge, the NBER-NSF Time Series Conference at the Chinese University of Hong Kong, Econometrics in the Arena at the University of Verona and the Econometric Society World Congress at the Bocconi University in Milan. We would like to thank Siem Jan Koopman, Andrew Harvey, and participants of seminars at Giessen-Margburg, CREST, Exeter Business School, Cardiff Business School, Tor Vergata, LUISS. Special thanks for useful discussions to Enzo D'Innocenzo, Iliyan Georgiev, Christian Gourieroux, Simone Manganelli, Federico Martellosio, Paolo Paruolo, Pietro Rigo, Berndt Scwhaab, and Pierluigi Vallarino.

References

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

Blasques, F., Gorgi, P., Koopman, S. J., and Winterberger, O. (2018). Feasible invertibility conditions and maximum likelihood estimation for observation–driven models. Electronic Journal of Statistics, 12(18):1019–1052.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics, 69(3):542–547.

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

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. SIAM Journal on Control and Optimization, 31(4):942–959.

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

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, 78(3):1093–1125.

Christoffersen, P. F. (1998). Evaluating interval forecasts. International Economic Review, 39(4):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.

De Rossi, G. and Harvey, A. C. (2006). Time-varying quantiles. Cambridge Working Papers in Economics, CWPE0649.

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.

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.

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. University of California Press.

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

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

Koenker, R. and Bassett, G. (1978). Regression quantiles. Econometrica, 46(1):33–50.

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association, 94(448):1296–1310.

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, 211:388–413.

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

Pötscher, B. and Prucha, I. (2001). Dynamic nonlinear econometric models. Springer.

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

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

Straumann, D. (2005). Estimation in Conditionally Heteroscedastic Time Series Models. Springer, Berlin, Heidelberg.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. The Annals of Statistics, 34(5):2449–2495.

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 error estimation. Econometric Theory, 7:46–68.

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.