Dealing with Markov-Switching Parameters in Quantile Regression Models[1]

Yunmi Kim[a], Lijuan Huo[b], and Tae-Hwan Kim[c]

[a]Department of Economics, University of Seoul, Republic of Korea

[b]School of Humanities and Social Sciences, Beijing Institute of Technology, China

[c]School of Economics, Yonsei University, Republic of Korea

September 19, 2019

Abstract: Quantile regression has become a standard modern econometric method because of its capability to investigate the relationship between economic variables at various quantiles. The econometric method of Markov-switching regression is also considered important because it can deal with structural models or time-varying parameter models flexibly. A combination of these two methods, known as "Markov-switching quantile regression (MSQR)," has recently been proposed. Liu (2016) and Liu and Luger (2017) propose MSQR models using the Bayesian approach whereas Ye et al.'s (2016) proposal for MSQR models is based on the classical approach. In our study, we extend the results of Ye et al. (2016). First, we propose an efficient estimation method based on the expectation-maximization algorithm. In our second extension, we adopt the quasi-maximum likelihood approach to estimate the proposed MSQR models unlike the maximum likelihood approach that Ye et al. (2016) use. Our simulation results confirm that the proposed expectation-maximization estimation method for MSQR models works quite well at all quantiles, even with sample sizes as small as 200.

Keywords: Quantile regression; Markov-switching; Structural breaks; Quasi-maximum likelihood estimation; EM algorithm.

JEL classifications: C21, C24.

Corresponding author: Tae-Hwan Kim, Yonsei University, School of Economics,134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea; Tel.: +82-2-2123-5461; Fax: +82-2-2123-8638; E-mail address: tae-hwan.kim@yonsei.ac.kr. Tae-Hwan Kim is grateful for financial support from the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A2A01025435). Lijuan Huo's research is supported by the National Natural Science Foundation of China, grant 71803009.

1. Introduction

Quantile regression originally developed by Koenker and Bassett (1978) has become a popular econometric method for better understanding the relationship among economic/financial variables. Its popularity is mainly because of its flexibility which allows researchers to investigate the relationship between economic variables not only at the center but also over many different quantiles and, therefore, the entire conditional distribution of the dependent variable. During the early development stages of the quantile regression method, the main focus was on how to analyze cross-section data. However, its development has subsequently been extended to panel as well as time-series data. Some selected papers in the time-series domain are Koenker and Xiao (2004, 2006), Galvao (2009), Xiao (2009), Galvao et al. (2009, 2011), Cho et al. (2015), and White et al. (2015).

In the time-series domain, there has been another important development for understanding the instable or dynamic relationship between economic/financial variables. A growing number of studies have presented empirical evidence of widespread instability (in the form of regime-switching or structural breaks) in univariate and multivariate macroeconomic time series relations; for example, Stock and Watson (1998), Diebold (1998), and Qu and Perron (2007). In response to this empirical evidence, researchers have worked intensively to develop Markov-switching models since the seminal work of Hamilton (1989). This approach allows the relationship between economic/financial variables to move from one regime to another and the dynamics are regulated by a discrete Markov chain with transition probabilities. If the model parameters are set in such a manner that the model moves to another regime and stays there forever, then the Markov-switching model is specialized to the well-known structural break model pioneered by Chow (1960), Andrews (1993), Bai (1997), and Bai and Perron (1998). However, the main development of Markov-switching models has taken place only in the area of the conditional mean function only.

Recently, some studies such as Liu (2016), Ye et al. (2016), and Liu and Luger (2017) have attempted to combine these two important methods (quantile regression and Markov-switching models) in a unified framework called "Markov-switching quantile regression (MSQR)” models. In these models, the coefficients in the conditional quantile function are allowed to move from one regime to another. On the one hand, Liu (2016) and Liu and Luger (2017) originally develop MSQR models in the framework of quantile autoregression proposed by Koenker and Xiao (2006) based on the Bayesian approach. Specifically, Liu (2016) allows all the quantile coefficients to be subject to Markov-switching, whereas only the quantile location (equivalent to the unconditional mean in AR models) is subject to Markov-switching in Liu and Luger (2017). The Bayesian sampling methods used in the two studies for inference are the Metropolis-Hastings sampling and Gibbs sampling approaches, respectively. Meanwhile, Ye et al. (2016) independently develop MSQR models in the standard linear quantile regression framework; however, they employ the classical approach rather than the Bayesian approach, unlike the previous two studies. The choice of which of the two approaches (classical vs. Bayesian) to use in analyzing data depends mainly on both researchers’ preferences and the complexity involved in data processing. To the best of our knowledge, if a researcher wishes to stay within the classical approach and to employ MSQR models, Ye et al. (2016) is the only study to rely on.

In our study, we have attempted to extend the results in Ye et al. (2016) in multiple ways. First, we propose an improved estimation method. It is not obviously clear in Ye et al. (2016) how to maximize their likelihood function. They state “We can then apply an optimization method to find the estimates $\hat{\theta}$ which maximize $L(\theta)$.” Because of the non-differentiability of the quantile objective function, standard optimization methods such as the Newton-Raphson method cannot be applied directly. Moreover, it is not possible to use the typical linear programming method used in the standard linear quantile regression model because of the nonlinearity introduced by Markov-switching. A possible way out is to use the “minimax” method advanced by Komunjer (2005). Although this is feasible, it is fairly time-consuming according to our preliminary Monte Carlo simulations. In this paper, we propose a computationally efficient estimation method based on the expectation maximization (EM) algorithm. The proposed estimation is intuitively described as an iterative procedure between (i) the expectation stage in which some estimates of unknown state variables indicating different regimes are produced, and (ii) the maximization stage which implements the standard linear quantile regression fitting with some dummy variables constructed from the state variables estimated in the expectation stage. The iterative procedure between the two stages stops when a desirable precision level is achieved. Hence, our proposed MSQR models can be easily estimated, for example, by the increasingly frequently used R program or by a typical linear programming method such as “qreg” in GAUSS.

In our second extension, we adopt the quasi-maximum likelihood (QML) approach to estimate the proposed MSQR models unlike the ML approach used in Ye et al. (2016). In their paper, Ye et al. (2016) assume that the quantile error term (u) follows an asymmetric Laplace distribution $\alpha(1-\alpha)\exp(-\rho_{\alpha}(u)/\sigma)/\sigma$ where α is a given quantile index, $\rho_{\alpha}(\bullet)$ is the usual quantile check function, and σ is a scale parameter. In reality, the underlying unknown distribution for the quantile error is not necessarily given by such an asymmetric Laplace distribution, and model misspecification is more abundant than the correction specification of the error. Following the spirit of Kim and White (2003) and Komunjer (2005), which embrace possible conditional quantile model misspecification, we do not assume that the quantile error follows an asymmetric Laplace distribution. Instead, we employ the new family of densities called “tick-exponential,” proposed by Komunjer (2005) for constructing the QML function to estimate the conditional quantile function. The resulting QML estimator is consistent and asymptotically normal. The tick-exponential density does not depend on any scale parameter and therefore, our proposed method can be applied to the situation in which the underlying unknown density does not have any finite variance, which is not possible in the MSQR models in Ye et al. (2016). In the third extension, we allow the intercept parameter in the conditional quantile function to be subject to regime switching, unlike Ye et al. (2016), where it is fixed.

The rest of the paper is structured as follows. Section 2 presents the model specification. We develop a QML estimation method in Section 3. Section 4 presents a computationally efficient and practical EM estimation method to obtain the QML estimator for the proposed MSQR models. Section 5 presents Monte Carlo experiments and Section 6 concludes.

2. Model Specification

We consider the following quantile regression model in which the conditional quantile parameters are subject to Markov-switching:

$$q_{y_t}(x_t, \beta_{\alpha,S_t}) = x_t'\beta_{\alpha,S_t} \tag{1}$$

$$\beta_{\alpha,S_t} = \beta_{\alpha,0}(1-S_t) + \beta_{\alpha,1}S_t \tag{2}$$

$$\Pr[S_t = 1 \mid S_{t-1} = 1] = p \text{ } and \text{ } \Pr[S_t = 0 \mid S_{t-1} = 0] = q, \tag{3}$$

where y_t is a real-valued random variable at time t; x_t is a $k \times 1$ random vector of exogenous or predetermined variables; $q_{y_t}(x_t, \beta_{\alpha,S_t})$ is the α-quantile of y_t conditional on x_t for a given value of quantile index $\alpha \in (0,1)$; β_{α,S_t} are the unknown parameters that are quantile (α) - and regime (S_t) - dependent; and the regime variable (or the state variable) S_t takes a discrete value of zero or one, evolving according to the transition probabilities of equation (3), as in Hamilton (1989). Hence, the conditional quantile function in equation (1) is a two-state, first-order Markov-switching[2] model, and the proposed model in (1) through (3) is called the two-state, first-order MSQR model.

As briefly mentioned in Section 1, the Markov-Switching model can be specialized to the well-known structural break model pioneered by Chow (1960). Hence, the MSQR model can also be used to describe unknown structural breaks in the conditional quantile function as analyzed in Qu (2008) and Oka and Qu (2011) if we allow for a one-time change between regimes and staying in the regime ($p = 1$) during the whole sample period. If β_{α,S_t}

[2] Our proposed modeling approach is sufficiently general to allow (i) the state variable S_t to take more than two states (i.e., n states so that S_t = 1, 2, …, n) and (ii) the state variable S_t to depend on S_{t-1} as well as more lags such as S_{t-2}, S_{t-3}, …, S_{t-r}, which will produce an n-state and r^{th}-order MSQR model.

are not regime-dependent (i.e., $\beta_{\alpha,S_t=0}=\beta_{\alpha,S_t=1}$), the MSQR model specializes to the standard linear quantile regression model originally proposed by Koenker and Bassett (1978).

3. Estimation

In the absence of regime switching in the conditional quantile parameters (i.e., $\beta_\alpha \equiv \beta_{\alpha,0}=\beta_{\alpha,1}$), model (1) can be estimated by solving the following minimization programme:

$$\min_{\beta_\alpha}\sum_{t=1}^{T}\rho_\alpha(y_t - x_t'\beta_\alpha) \tag{4}$$

where ρ_α is the usual "check function" defined as $\rho_\alpha(z)=\left(\alpha-1_{[z\le 0]}\right)z$. Alternatively, Komunjer (2005) estimates the model by employing the QML estimation method by solving the following maximization programme:

$$\max_{\beta_\alpha} L_T(\beta_\alpha)\equiv T^{-1}\sum_{t=1}^{T}\ln l_t(\beta_\alpha) \tag{5}$$

where $l_t(\cdot)$ is a period-*t* conditional quasi-likelihood function. Komunjer (2005) shows that if the so-called "tick-exponential" density is used for constructing the likelihood function $l_t(\cdot)$, it will deliver consistent and asymptotically normal QML estimators. The tick-exponential density is defined as follows:

$$\begin{aligned} l_t(\beta_\alpha) &= f(y_t \mid I_{t-1};\beta_\alpha) \\ &= \exp\left(\frac{1}{\alpha}\left(y_t - x_t'\beta_\alpha\right)1_{\left[y_t - x_t'\beta_\alpha \le 0\right]}\right)\times \exp\left(-\frac{1}{(1-\alpha)}\left(y_t - x_t'\beta_\alpha\right)1_{\left[y_t - x_t'\beta_\alpha > 0\right]}\right) \end{aligned} \tag{6}$$

where I_{t-1} refers to information up to time $t-1$.

To estimate the MSQR model in equations (1)-(3) consistently, we maximize the likelihood function constructed by using the tick-exponential density in (6) as follows:

$$\max_{\beta_{\alpha,0},\beta_{\alpha,1},p,q} L_T(\beta_{\alpha,0},\beta_{\alpha,1},p,q) \equiv T^{-1}\sum_{t=1}^{T}\ln l_t(\beta_{\alpha,0},\beta_{\alpha,1},p,q)\,, \qquad (7)$$

where $l_t(\beta_{\alpha,0},\beta_{\alpha,1},p,q)$ is the tick-exponential density for y_t conditional on I_{t-1}. The consistency and asymptotic normality of the resulting QML estimator follows from Komunjer (2005). However, how to construct $l_t(\beta_{\alpha,0},\beta_{\alpha,1},p,q)$ is not straightforward because of the presence of the transition probabilities denoted by $\theta_2 \equiv (p\ \ q)'$ in addition to the conditional quantile parameters denoted by $\theta_1 \equiv (\beta_{\alpha,0}\ \ \beta_{\alpha,1})'$. In what follows, we explain how to appropriately derive the quasi-likelihood function $l_t(\beta_{\alpha,0},\beta_{\alpha,1},p,q)$ = $f(y_t \mid I_{t-1};\beta_{\alpha,0},\beta_{\alpha,1},p,q)$.

Consider that our model includes the unobserved state variables S_t, $t=1,\cdots,T$. Similar to the conventional Markov-switching models at the conditional mean function, we use the following two steps to determine the likelihood function:[3]

Step 1: We construct the joint density of y_t and S_t, which can be expressed as the product of the conditional and marginal densities as follows:

$$f\left(y_t,S_t \mid I_{t-1}\right) = f\left(y_t \mid S_t,I_{t-1}\right) f\left(S_t \mid I_{t-1}\right) \qquad (8)$$

where I_{t-1} is the information set up to time $t-1$.

Step 2: We then obtain the marginal density of y_t by integrating S_t out of the joint density in equation (8) by summing over all possible values of S_t as follows:

[3] For more details of the conventional Markov-switching model at the conditional mean function, readers are referred to Hamilton (1989) and Kim and Nelson (1999, pp. 59-96).

$$
\begin{aligned}
&f\left(y_t \mid I_{t-1}\right) \\
&=\sum_{S_t=0}^{1} f\left(y_t, S_t \mid I_{t-1}\right) \\
&=\sum_{S_t=0}^{1} f\left(y_t \mid S_t, I_{t-1}\right) f\left(S_t \mid I_{t-1}\right) \\
&= f\left(y_t \mid S_t=0, I_{t-1}\right)\times \Pr\left[S_t=0 \mid I_{t-1}\right]+ f\left(y_t \mid S_t=1, I_{t-1}\right)\times \Pr\left[S_t=1 \mid I_{t-1}\right]
\end{aligned} \tag{9}
$$

where

$$
\begin{aligned}
&f\left(y_t \mid S_t=0, I_{t-1}\right) \\
&\quad=\exp\left(\frac{1}{\alpha}\left(y_t-x_t'\beta_{\alpha,0}\right)1_{\left[y_t-x_t'\beta_{\alpha,0}\le 0\right]}\right)\times \exp\left(-\frac{1}{(1-\alpha)}\left(y_t-x_t'\beta_{\alpha,0}\right)1_{\left[y_t-x_t'\beta_{\alpha,0}>0\right]}\right)
\end{aligned} \tag{10}
$$

and

$$
\begin{aligned}
&f\left(y_t \mid S_t=1, I_{t-1}\right) \\
&\quad=\exp\left(\frac{1}{\alpha}\left(y_t-x_t'\beta_{\alpha,1}\right)1_{\left[y_t-x_t'\beta_{\alpha,1}\le 0\right]}\right)\times \exp\left(-\frac{1}{(1-\alpha)}\left(y_t-x_t'\beta_{\alpha,1}\right)1_{\left[y_t-x_t'\beta_{\alpha,1}>0\right]}\right).
\end{aligned} \tag{11}
$$

Therefore, the likelihood function is given by

$$
L_T(\beta_{\alpha,0},\beta_{\alpha,1},p,q)\equiv T^{-1}\sum_{t=1}^{T}\ln\left[\sum_{i=0}^{1} f\left(y_t \mid S_t=i, I_{t-1}\right)\Pr\left[S_t=i \mid I_{t-1}\right]\right]. \tag{12}
$$

As is clear from equation (12), the likelihood function can be interpreted as the sum of the weighted average of the marginal densities of y_t in which the weight factors are given by the probabilities governing the behavior of the state variable S_t. However, the derivation of the likelihood function is not complete as shown in equation (12) because the likelihood function is not yet expressed properly in terms of the transition probabilities p and q. To do so, we need to calculate the weighting factors, $\Pr\left[S_t=0 \mid I_{t-1}\right]$ and $\Pr\left[S_t=1 \mid I_{t-1}\right]$ appropriately. We employ the following filter to calculate these weighting factors:

Step 1: At the beginning of time t or the t-th iteration, the probabilities $\Pr[S_{t-1}=i\,|\,I_{t-1}]$ for $i=0,1$ are known. Hence, the weighting factors $\Pr[S_t=j\,|\,I_{t-1}]$, sometimes called "the filtered probabilities," are calculated as follows:

$$\begin{aligned}\Pr[S_t=j\,|\,I_{t-1}] &= \sum_{i=0}^{1}\Pr[S_t=j,S_{t-1}=i\,|\,I_{t-1}]\\ &= \sum_{i=0}^{1}\Pr[S_t=j\,|S_{t-1}=i]\Pr[S_{t-1}=i\,|\,I_{t-1}],\end{aligned} \tag{13}$$

where $\Pr[S_t=j\,|S_{t-1}=i]$ are the transition probabilities expressed in terms of p and q. The information set I_{t-1} drops out of the conditional set because of the first-order Markovian property.

Step 2: At the end of time t, or at the end of the t-th iteration, y_t is observed. Hence, we can then update the probability term in the following way:

$$\begin{aligned}\Pr[S_t=j\,|\,I_t]=\Pr[S_t=j\,|\,I_{t-1},y_t] &= \frac{f(S_t=j,y_t\,|\,I_{t-1})}{f(y_t\,|\,I_{t-1})}\\ &= \frac{f(y_t\,|\,S_t=j,I_{t-1})\Pr[S_t=j\,|\,I_{t-1}]}{\sum_{j=0}^{1}f(y_t\,|\,S_t=j,I_{t-1})\Pr[S_t=j\,|\,I_{t-1}]},\end{aligned} \tag{14}$$

where $I_t=\{I_{t-1},y_t\}$.

The two steps above can be iterated to obtain $\Pr[S_t=j\,|\,I_{t-1}]$ for $t=1,\cdots,T$. To start the filter above at time $t=1$, however, we need to know the initial probability, $\Pr[S_0\,|\,I_0]$. We employ the following steady-state or unconditional probabilities of S_t:

$$\Pr[S_0=0\,|\,I_0]=\frac{1-p}{2-p-q}, \tag{15}$$

$$\Pr\left[S_0 = 1 \mid I_0\right] = \frac{1-q}{2-p-q}. \tag{16}$$

Once $\Pr\left[S_t = j \mid I_{t-1}\right]$ for $t = 1,\cdots,T$ are obtained as above, these are substituted into the likelihood function in equation (12), which completes the derivation of the desirable likelihood function. Now, it is clear that the likelihood function in equation (12) is a function of all the relevant parameters $\beta_{\alpha,0}, \beta_{\alpha,1}, p,$ and $q.$[4]

It is well known that it is not sufficient to make inferences on the model parameters in Markov-switching models, because researchers also wish to make inferences on the unobservable state variable S_t. In principle, inferences on S_t can be carried out using the filter provided above because all required information can be gathered. Because the required information needed for inferences on S_t is recursively obtained by iterations through the sample period, not all information in the sample is used unlike inferences on the true model parameters. However, once the QML estimates for all model parameters are obtained through the procedure explained above, it is now possible to make inferences on S_t using all the information in the sample. Specifically, instead of estimating $\Pr\left[S_t = j \mid I_{t-1}\right]$, it is possible to estimate $\Pr\left[S_t = j \mid I_T\right]$ (for $t = 1,\cdots,T$). This probability conditional on all information in the sample I_T is called the "smoothed probability." We explain how to compute the smoothed probability below.

Based on full information, the joint probability that $S_t = j$ and $S_{t+1} = k$ can be calculated as follows:

$$\begin{aligned}
\Pr\left[S_t = j, S_{t+1} = k \mid I_T\right] &= \Pr\left[S_t = j \mid S_{t+1} = k, I_T\right] \times \Pr\left[S_{t+1} = k \mid I_T\right] \\
&= \Pr\left[S_{t+1} = k \mid I_T\right] \times \Pr\left[S_t = j \mid S_{t+1} = k, I_t\right] \\
&= \Pr\left[S_{t+1} = k \mid I_T\right] \times \frac{\Pr\left[S_t = j, S_{t+1} = k \mid I_t\right]}{\Pr\left[S_{t+1} = k \mid I_t\right]} \\
&= \Pr\left[S_{t+1} = k \mid I_T\right] \times \frac{\Pr\left[S_{t+1} = k \mid S_t = j\right] \times \Pr\left[S_t = j \mid I_t\right]}{\Pr\left[S_{t+1} = k \mid I_t\right]}.
\end{aligned} \tag{17}$$

[4]The proposed MSQR model can be easily extended to account for serially correlated data. For this, we need to modify the filter above using the conventional Hamilton (1989) filter. For details of the Hamilton filter, readers are referred to Hamilton (1989).

Therefore, the smoothed probability is obtained by integrating S_{t+1} out from the joint distribution of S_t and S_{t+1} in (17) as follows:

$$\Pr\left[S_t = j \mid I_T\right] = \sum_{k=0}^{1} \Pr\left[S_t = j, S_{t+1} = k \mid I_T\right]. \tag{18}$$

Once we obtain $\Pr\left[S_T \mid I_T\right]$ from the last iteration of the basic filter in equation (14), equations (17) and (18) can be iterated for $t = T-1, T-2, \cdots, 1$ in a backward manner.[5]

4. EM Algorithm

The previous section provides the theoretical derivation of the likelihood function for the QML approach based on the tick-exponential family of densities. The QML estimators for conditional quantiles are consistent and asymptotically normal which follows from Komunjer (2005). In practice, however, solving the maximization problem in equation (7) along with equations (8)-(16) is not straightforward because the likelihood function is not everywhere differentiable.

The practical optimization algorithm that we propose in this paper is based on some modification of the EM algorithm. Originally pioneered by Dempster, Laird, and Rubin (1977), the EM algorithm was introduced to compute maximum likelihood estimates for models with missing observations or unobserved variables. Hamilton (1990) then proposes a variant of the EM algorithm for Markov-switching models and it was subsequently applied in Turner, Startz, and Nelson (1989) and Engel and Hamilton (1990).

In this paper, we attempted to modify the EM algorithm for our Markov-switching quantile models. The standard EM algorithm consists of two steps: the "expectation" and "maximization" steps. The algorithm is carried out by an iterative process that oscillates

[5] The smoothing algorithm derived above can be generalized to a general AR(k) model with Markov-switching as considered by Hamilton (1989). For more details of the generalization, see Hamilton (1989) and Kim and Nelson (1999, pp. 59-96).

between these two steps until a desirable level of precision is achieved. Suppose that θ is the vector containing all the model parameters; that is, $\theta = (\theta_1', \theta_2')'$ where $\theta_1 \equiv (\beta_{\alpha,0} \;\; \beta_{\alpha,1})'$ and $\theta_2 \equiv (p \;\; q)'$. At each iteration (say, at the k^{th} iteration), the following operations are carried out:

Expectation Step: The expected value of the unobserved variables S_t (denoted by $p_{t,j} = \Pr\left[S_t = j \mid I_T\right]$) is calculated, provided that parameter estimates (θ^{k-1}) are obtained from the $(k-1)^{th}$ maximization step,

Maximization Step: The likelihood function is maximized with respect to θ, conditional on the expected values $p_{t,j}$ from the k^{th} expectation step, which results in θ^k.

It is well documented in the literature that, regardless of any arbitrary initial values of the parameter vector (denoted by θ^0), each iteration results in a higher value of the likelihood function. The iterative procedure stops when either the increment of the likelihood function is negligible or the distance between θ^k and θ^{k-1} is sufficiently close to zero. We note that the expectation step is essentially to compute the smoothed probabilities of the unobserved Markov-switching variables, which has already been explained in the previous section. Hence, we focus on the maximization step of the EM algorithm below.

Let us define $\tilde{y}_T = (y_1 \;\; y_2 \;\; \cdots \;\; y_T)'$ and $\tilde{S}_T = (S_1 \;\; S_2 \;\; \cdots \;\; S_T)'$. Then, the joint density of $\tilde{y}_T$ and $\tilde{S}_T$ and the log likelihood function can be written as follows:

$$\begin{aligned} f(\tilde{y}_T, \tilde{S}_T; \theta) &= f(\tilde{y}_T \mid \tilde{S}_T; \theta_1) \times f(\tilde{S}_T; \theta_2) \\ &= \prod_{t=1}^{T} f(y_t \mid S_t; \theta_1) \times \prod_{t=1}^{T} f(S_t \mid S_{t-1}; \theta_2) \end{aligned} \tag{19}$$

$$\ln f(\tilde{y}_T, \tilde{S}_T; \theta) = \sum_{t=1}^{T} \ln f(y_t \mid S_t; \theta_1) + \sum_{t=1}^{T} \ln f(S_t \mid S_{t-1}; \theta_2) \tag{20}$$

If $\tilde{S}_T$ are observed, the parameter vector θ_2 is not relevant and the log likelihood function is simply maximized with respect to θ_1 only, as follows:

$$\frac{\partial \ln f(\tilde{y}_T, \tilde{S}_T; \theta)}{\partial \theta_1} = \sum_{t=1}^{T} \frac{\partial \ln f(y_t \,|\, S_t; \theta_1)}{\partial \theta_1} = 0 \tag{21}$$

In reality, $\tilde{S}_T$ is not observed. Hence, as an alternative to the approach discussed in the previous section, one can maximize the following expected log likelihood function:

$$\begin{aligned} Q(\theta; \tilde{y}_T, \theta^{k-1}) &= \int_{\tilde{S}_T} \ln f(\tilde{y}_T, \tilde{S}_T; \theta)\, f(\tilde{y}_T, \tilde{S}_T; \theta^{k-1}) \\ &= \int_{\tilde{S}_T} \ln\left[f(\tilde{y}_T \,|\, \tilde{S}_T; \theta_1)\, f(\tilde{S}_T; \theta_2) \right] f(\tilde{y}_T, \tilde{S}_T; \theta^{k-1}) \end{aligned} \tag{22}$$

where the expectation is formed conditional on θ^{k-1} and the integral is taken over all possible values of $\tilde{S}_T = (S_1 \; S_2 \; \cdots \; S_T)'$; that is, $\int_{\tilde{S}_T} = \sum_{S_1} \sum_{S_2} \cdots \sum_{S_T}$.

The first order condition for maximizing the expected log likelihood function in equation (22) with respect to θ_1 is given by

$$\frac{\partial Q(\theta; \tilde{y}_T, \theta^{k-1})}{\partial \theta_1} = \int_{\tilde{S}_T} \frac{\partial \ln\left[f(\tilde{y}_T \,|\, \tilde{S}_T; \theta_1) \right]}{\partial \theta_1} f(\tilde{y}_T, \tilde{S}_T; \theta^{k-1}) = 0\,. \tag{23}$$

Dividing both sides of equation (23) by $f(\tilde{y}_T; \theta^{k-1})$ results in the following:

$$\int_{\tilde{S}_T} \frac{\partial \ln\left[f(\tilde{y}_T \,|\, \tilde{S}_T; \theta_1) \right]}{\partial \theta_1} \frac{f(\tilde{y}_T, \tilde{S}_T; \theta^{k-1})}{f(\tilde{y}_T; \theta^{k-1})} = 0 \tag{24}$$

which in turns implies the following three equations in sequence:

$$\int_{\tilde{S}_T} \frac{\partial \ln\left[f(\tilde{y}_T \mid \tilde{S}_T;\theta_1)\right]}{\partial \theta_1} f(\tilde{S}_T \mid \tilde{y}_T;\theta^{k-1}) = 0, \tag{25}$$

$$\int_{\tilde{S}_T} \sum_{t=1}^{T} \frac{\partial \ln\left[f(y_t \mid S_t;\theta_1)\right]}{\partial \theta_1} f(\tilde{S}_T \mid \tilde{y}_T;\theta^{k-1}) = 0, \tag{26}$$

$$\sum_{S_t=0}^{1} \sum_{t=1}^{T} \frac{\partial \ln\left[f(y_t \mid S_t;\theta_1)\right]}{\partial \theta_1} f(S_t \mid \tilde{y}_T;\theta^{k-1}) = 0, \tag{27}$$

where $f(S_t \mid \tilde{y}_T;\theta^{k-1})$ is the smoothed probability $\Pr\left[S_t = j \mid I_T\right]$ discussed in the previous section. When comparing equations (27) and (21), one can easily notice that the left-hand term of equation (27) is a weighted average of the score function, which appears on the left-hand side of equation (21). The weights are given by $f(S_t \mid \tilde{y}_T;\theta^{k-1})$ which is the smoothed probabilities of S_t obtained from the expectation step, conditional on θ^{k-1} obtained from the previous iteration. Therefore, it can be clearly seen that the EM algorithm computes θ_1^k (i.e., the estimate of θ_1 at the k^{th} iteration) by equating the weighted average of the score function to 0.

Conditional on, $S_t = j$, we have

$$\ln f\left(y_t \mid S_t = j;\theta_1\right) = \ln \varphi^{\alpha}_{t,S_t=j}\left(y_t, q_\alpha(x_t,\beta_{\alpha,j})\right) \tag{28}$$

where

$$\begin{aligned}\varphi^{\alpha}_{t,S_t=j}&\left(y_t, q_\alpha(x_t,\beta_{\alpha,j})\right) \\ &\equiv \exp\left(\frac{1}{\alpha}\left(y_t - x_t{}'\beta_{\alpha,j}\right)1_{\left[y_t - x_t{}'\beta_{\alpha,j} \le 0\right]}\right) \times \exp\left(-\frac{1}{(1-\alpha)}\left(y_t - x_t{}'\beta_{\alpha,j}\right)1_{\left[y_t - x_t{}'\beta_{\alpha,j} > 0\right]}\right).\end{aligned} \tag{29}$$

The first-order conditions in equation (27) with respect to $\beta_{\alpha,j}$, $j = 0,1$, are given by

$$\sum_{t=1}^{T}\sum_{S_t=0}^{1}\frac{\partial \ln\left[f(y_t|S_t)\right]}{\partial \beta_{\alpha,j}} f(S_t \mid \tilde{y}_T;\theta^{k-1}) = 0\,. \tag{30}$$

which in turn implies that

$$\sum_{t=1}^{T}\left(\alpha - 1_{\{y_t \le q_\alpha(x_t,\beta_{\alpha,j})\}}\right) x_t\, p(S_t = j \mid \tilde{y}_T;\theta^{k-1}) = 0\,. \tag{31}$$

Defining $p_{t,j} = p(S_t = j \mid \tilde{y}_T;\theta^{k-1})$, the condition in equation (31) can be equivalently written as follows:

$$\sum_{t\in\{t:p_{t,j}>0\}}\left(\alpha - 1_{\{p_{t,j}y_t \le p_{t,j}q_\alpha(x_t,\beta_{\alpha,j})\}}\right) x_t\, p_{t,j} = 0\,. \tag{32}$$

It can be easily recognized that the equivalent condition in equation (32) is exactly the first-order condition for the quantile regression model where transformed variables $y_t^* \equiv p_{t,j} \times y_t$ and $x_t^* \equiv p_{t,j} \times x_t$ for $t \in \{t : p_{t,j} > 0\}$ are used as the dependent and independent variables, respectively. Hence, $\beta_{\alpha,j}^k$, the solution in equation (32), can alternatively be obtained in the standard quantile regression by regressing y_t^* on x_t^*. Hence, our EM algorithm can be carried out on any software in which the standard linear quantile regression can be implemented, such as R or GAUSS.

We make one last comment about how to calculate the transition probabilities within the EM algorithm. Let us define the transition probability at the k^{th} iteration (denoted by p_{jj}^k) as follows: $p_{jj}^k = P^{(k)}[S_t = j \mid S_{t-1} = j]$. We can easily obtain p_{jj}^k by differentiating the expected log likelihood function in equation (22):

$$p_{jj}^k = \frac{\sum_t p(S_t = j, S_{t-1} = j \mid \tilde{y}_T;\theta^{k-1})}{\sum_t p(S_{t-1} = j \mid \tilde{y}_T;\theta^{k-1})},\ j = 0,1. \tag{33}$$

We refer readers to Hamilton (1990) and Kim and Nelson (1999, p.59-96) for more detail on exactly how to derive equation (33), and on the EM algorithm in general,

5. Monte Carlo Simulations

To investigate the finite sample properties of the proposed estimation procedures, we carry out Monte Carlo experiments in this section. We generate 1,000 datasets with a sample size $T = 200$ using the following data generating process:

$$y_t = \beta_{0,S_t} + x_t \beta_{1,S_t} + e_t,\ e_t \sim IIDN(0, \sigma_e^2)\,, \tag{34}$$

$$\beta_{0,S_t} = \beta_{0,0}(1 - S_t) + \beta_{0,1} S_t;\ \beta_{1,S_t} = \beta_{1,0}(1 - S_t) + \beta_{1,1} S_t\,, \tag{35}$$

$$\Pr[S_t = 1 \mid S_{t-1} = 1] = 0.95 \textit{ and } \Pr[S_t = 0 \mid S_{t-1} = 0] = 0.95\,, \tag{36}$$

$$x_t \sim IIDN(0, 2)\,, \tag{37}$$

$$\beta_{0,0} = -1;\ \beta_{1,0} = -1;\ \beta_{0,1} = 1;\ \beta_{1,1} = 1;\ \sigma_e^2 = 1\,. \tag{38}$$

For each dataset generated, we estimate the MSQR model in equations (34) to (38) by employing the proposed EM estimation method presented in Section 4. Our experiments are performed for different quantiles ($\alpha = 0.1, 0.3, 0.5, 0.7, 0.9$).

The simulation results are shown in Table 1, where sample means and sample standard deviations for each parameter, based on the generated 1,000 replications, are reported. Let us take a close look at the median case; $\alpha = 0.5$. The true values for the low and high regime intercept parameters are given by $\beta_{0,S_t=0} = -1$ and $\beta_{0,S_t=1} = 1$, respectively; whereas their point estimates are -1.009 and 0.990 with precision measures (standard deviations) 0.163 and 0.141, respectively. We also have very similar results for the slope parameter β_{1,S_t} for the median case; the true values are -1 and 1 as before, whereas the point estimates are -1.016 and 0.990 with precision measures given by 0.098 and 0.105. All the point estimates are fairly close to their corresponding true values and all precision measures are reasonably small given that the sample size is just 200.

Moving to a different quantile index, say $\alpha = 0.1$, one can notice that the true values for the low and high regime intercept parameters are given by $\beta^*_{0,S_t=0} = -2.282$ and $\beta^*_{0,S_t=1} = -0.282$, respectively, not the imposed values of -1 and 1. Such a change in the intercept parameter is typical in any standard quantile regression because the true value of the intercept parameter is adjusted to reflect the varying quantile effect of the quantile error term. As a result, the true value of the intercept, denoted by β^*_{0,S_t}, is calculated by $\beta^*_{0,S_t} = \beta_{0,S_t} + \Phi^{-1}(\alpha)$, where $\Phi^{-1}(\alpha)$ is the inverse of the cumulative distribution function of the standard normal random variable at quantile α, and β_{0,S_t} is the imposed intercept parameter in the data generation process. For example, when $\alpha = 0.1$, we have $\Phi^{-1}(0.1) = -1.282$. Hence, with $\beta_{0,S_t=0} = -1$, we obtain $\beta^*_{0,S_t=0} = -2.282$. When comparing the true values for $\alpha = 0.1$ with their point estimates, they are again fairly close to each other, although precision measures are a little bit larger than the median case, which is also typical in quantile regression; precision becomes worse in either low or high quantiles compared with the central quantiles.

We have qualitatively similar results for the other quantile indexes; $\alpha = 0.3, 0.7, 0.9$. Hence, we can conclude that the proposed estimation procedure performs well and produces reasonably accurate estimates of the MSQR model even with a small sample size as small as $T = 200$. To examine the effect of larger sample sizes, we generated 1,000 replications again, using the same model in equations (34) to (38), but with $T = 500$. The results are reported in Table 2. As expected, all point estimates are closer to their corresponding true values and all precision measures are smaller when the sample size becomes larger.

Overall, our simulation results confirm that the proposed EM estimation method for MSQR models works quite well at all quantiles, even with sample sizes as small as 200.

7. Conclusion

Quantile regression has become a standard modern econometric method because of its capability to investigate the relationship between economic variables at various quantiles. Another important econometric method is Markov-switching regression because it can deal with structural models or time-varying parameter models flexibly. A combination of these two methods, known as "MSQR," has been recently proposed. Liu (2016) and Liu and Luger (2017) propose MSQR models using the Bayesian approach whereas Ye et al.'s (2016) proposal for MSQR models is based on the classical approach. In our study, we have extended the results in Ye et al. (2016) in many ways. First, we proposed an improved estimation method based on the EM algorithm. Consequently, our proposed MSQR models can be easily estimated by any software program that can estimate the standard linear quantile regression models, such as R or GAUSS. In our second extension, we adopted the QML approach to estimate the proposed MSQR models unlike the ML approach used in Ye et al. (2016). Hence, our framework is more general and encompasses wider conditional quantile functions. Finally, we allowed the intercept parameter in the conditional quantile function to be subject to regime switching, unlike Ye et al. (2016).

References

Andrews, D.W.K. (1993), "Tests for parameter instability and structural change with unknown change point," *Econometrica*, 61, 821-856.

Bai, J. (1997), "Estimating Multiple Breaks One at a Time," *Econometric Theory*, 13, 315-352.

Bai, J. and P. Perron (1998), "Estimating and testing linear models with multiple structural changes," *Econometrica*, 66, 47-78.

Cho, J., T.-H. Kim, and Y. Shin (2015), "Quantile cointegration in the autoregressive distributed-lag modeling framework," *Journal of Econometrics*, 188, 281-300.

Chow, G.C. (1960), "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591-605.

Dempster, A.P., N.M. Laird, and D.B. Rubin (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)*, 39, 1-38.

Diebold, F. X. (1998), "The past, present, and future of macroeconomic forecasting," *Journal of Economic Perspectives*, 12, 175-192.

Engel, C. and J. Hamilton (1990), "Long swings in the dollar: are they in the data and do markets know it?" *American Economic Review*, 80, 689-713.

Galvao, A. (2009), "Unit root quantile autoregression testing using covariates," *Journal of Econometrics*, 152, 165-178.

Galvao, A., G. Montes-Rojas, and S. Park (2009), "Quantile autoregressive distributed lag model with an application to house price returns," *Oxford Bulletin of Econometrics and Statistics*, 75, 307-321.

Galvao, A., G. Montes-Rojas, and J. Olmo (2011), "Threshold quantile autoregressive models," *Journal of Time Series Analysis*, 32, 253-267.

Hamilton, J. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357-384.

Hamilton, J. (1990), "Analysis of time series subject to changes in regime," *Journal of Econometrics*, 45, 39-70.

Kim, C-J, and C.R. Nelson (1999), *State-space models with regime switching*, London, England, The MIT press.

Kim, C-J, J. Piger, and R. Startz (2008), "Estimation of Markov regime-switching regression models with endogenous switching," *Journal of Econometrics*, 143, 263-273.

Kim, T.-H., and H. White (2003), "Estimation, inference, and specification testing for possibly misspecified quantile regression," *Advances in Econometrics*, 17, 107-132.

Koenker, R., and G. Bassett (1978) " Regression quantiles," *Econometrica*, 46, 33-50.

Koenker, R., and Z. Xiao (2004) "Unit root quantile autoregression inference," *Journal of the American Statistical Association*, 99, 775-787.

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

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

Liu, X. (2016), "Markov switching quantile autoregression," *Statistica Neerlandica*, 2016, 70, 356-395.

Liu, X. and R. Luger (2017), "Markov-switching quantile autoregression: a Gibbs sampling approach," *Studies in Nonlinear Dynamics & Econometrics*, 22, 1-36.

Oka, T. and Z. Qu (2011), "Estimating structural changes in regression quantiles," *Journal of Econometrics*, 162, 248–267.

Qu, Z. (2008), "Testing for structural change in regression quantiles," *Journal of Econometrics*, 146, 170–184.

Qu, Z., and P. Perron (2007), "Estimating and testing structural changes in multivariate regressions," *Econometrica*, 75, 459-502.

Turner, C.M., R. Startz, and C.R. Nelson (1989), "A Markov model of heteroscedasticity, risk, and learning in the stock market," *Journal of Financial Economics*, 25, 3-22.

White, H., T.-H. Kim, and S. Manganelli (2015), "VAR for VaR: measuring tail dependence using multivariate regression quantiles," *Journal of Econometrics*, 187, 169-188.

Ye, W. Y. Zhu, Y. Wu, and B. Miao, (2016), “Markov regime-switching quantile regression models and financial contagion detection,” *Insurance: Mathematics and Economics*, 67, 21-26.

Table 1. Simulation results based on the EM algorithm: (i) model parameters: ($\beta_{0,0} = -1$; $\beta_{1,0} = -1$; $\beta_{0,1} = 1$; $\beta_{1,1} = 1$; $\sigma_e^2 = 1$), (ii) $T = 200$, (iii) number of replications = 1,000.

Parameters	True values	Mean	Standard Deviation
$\alpha = 0.1$			
$\beta^*_{0,S_t=0}$	-2.282	-2.201	0.196
$\beta_{1,S_t=0}$	-1	-1.049	0.151
$\beta^*_{0,S_t=1}$	-0.282	-0.051	0.217
$\beta_{1,S_t=1}$	1	1.050	0.134
$\alpha = 0.3$			
$\beta^*_{0,S_t=0}$	-1.524	-1.511	0.149
$\beta_{1,S_t=0}$	-1	-1.013	0.109
$\beta^*_{0,S_t=1}$	0.476	0.511	0.144
$\beta_{1,S_t=1}$	1	1.002	0.100
$\alpha = 0.5$			
$\beta^*_{0,S_t=0}$	-1	-1.009	0.163
$\beta_{1,S_t=0}$	-1	-1.016	0.098
$\beta^*_{0,S_t=1}$	1	1.009	0.141
$\beta_{1,S_t=1}$	1	0.990	0.105
$\alpha = 0.7$			
$\beta^*_{0,S_t=0}$	-0.476	-0.515	0.152
$\beta_{1,S_t=0}$	-1	-1.006	0.106
$\beta^*_{0,S_t=1}$	1.524	1.508	0.142
$\beta_{1,S_t=1}$	1	1.011	0.102

$\alpha = 0.9$			
$\beta^*_{0,S_t=0}$	0.2816	0.052	0.219
$\beta_{1,S_t=0}$	-1	-1.052	0.132
$\beta^*_{0,S_t=1}$	2.2816	2.209	0.200
$\beta_{1,S_t=1}$	1	1.054	0.155

Table 2. Simulation results based on the EM algorithm: (i) model parameters: ($\beta_{0,0} = -1$; $\beta_{1,0} = -1$; $\beta_{0,1} = 1$; $\beta_{1,1} = 1$; $\sigma_e^2 = 1$), (ii) $T = 500$, (iii) number of replications = 1,000.

Parameters	True values	Mean	Standard Deviation
$\alpha = 0.1$			
$\beta^*_{0,S_t=0}$	-2.282	-2.245	0.119
$\beta_{1,S_t=0}$	-1	-1.034	0.090
$\beta^*_{0,S_t=1}$	-0.282	-0.120	0.139
$\beta_{1,S_t=1}$	1	1.039	0.083
$\alpha = 0.3$			
$\beta^*_{0,S_t=0}$	-1.524	-1.524	0.092
$\beta_{1,S_t=0}$	-1	-1.006	0.062
$\beta^*_{0,S_t=1}$	0.476	0.500	0.090
$\beta_{1,S_t=1}$	1	1.004	0.062
$\alpha = 0.5$			
$\beta^*_{0,S_t=0}$	-1	-1.009	0.083
$\beta_{1,S_t=0}$	-1	-1.002	0.057
$\beta^*_{0,S_t=1}$	1	1.003	0.084
$\beta_{1,S_t=1}$	1	1.001	0.060
$\alpha = 0.7$			
$\beta^*_{0,S_t=0}$	-0.476	-0.505	0.092
$\beta_{1,S_t=0}$	-1	-1.005	0.061
$\beta^*_{0,S_t=1}$	1.524	1.517	0.089
$\beta_{1,S_t=1}$	1	1.006	0.064

$\alpha = 0.9$			
$\beta^{*}_{0,S_t=0}$	0.2816	0.121	0.132
$\beta_{1,S_t=0}$	-1	-1.039	0.083
$\beta^{*}_{0,S_t=1}$	2.2816	2.235	0.119
$\beta_{1,S_t=1}$	1	1.036	0.092