Markov-Switching Quantile Autoregression

Liu, Xiaochun

Department of Economics and Finance, University of Central Arkansas

7 October 2013

Markov-Switching Quantile Autoregression

Xiaochun Liu*

October 16, 2015

Abstract

This paper considers the location-scale quantile autoregression in which the location and scale parameters are subject to regime shifts. The regime changes in lower and upper tails are determined by the outcome of a latent, discrete-state Markov process. The new method provides direct inference and estimate for different parts of a nonstationary time series distribution. Bayesian inference for switching regimes within a quantile, via a three-parameter asymmetric-Laplace distribution, is adapted and designed for parameter estimation. Using the Bayesian output, the marginal likelihood is readily available for testing the presence and the number of regimes. The simulation study shows that the predictability of regimes and conditional quantiles by using asymmetric Laplace distribution as the likelihood is fairly comparable with the true model distributions. However, ignoring that autoregressive coefficients might be quantile-dependent leads to substantial bias in both regime inference and quantile prediction. The potential of this new approach is illustrated in the empirical application to the U.S. financial market of different frequencies.

Keywords: Asymmetric-Laplace Distribution, Metropolis-Hastings, Block-at-a-Time, Transition Probability, Marginal Likelihood

JEL: C22, C51, C11

*Department of Economics, Finance and Insurance & Risk Management, College of Business, University of Central Arkansas, Conway AR 72035 USA. Email: xliu@uca.edu

1 Introduction

Koenker and Xiao (2006) study quantile autoregression models in which the autoregressive coefficients may take distinct values over different quantiles of the innovation process. Their models can capture systematic influences of conditioning variables on the location, scale and shape of the conditional distribution. Let $\{U_t\}$ be a sequence of i.i.d. standard uniform random variables. Consider the mth-order autoregressive process

$$y_t = \theta_0(U_t) + \theta_1(U_t)\, y_{t-1} + ... + \theta_m(U_t) y_{t-m} \tag{1.1}$$

where y_t is the time series observation at time t, and θ's are unknown functions $[0,1] \to \mathbb{R}$ to be estimated. Provided that the right side of (1.1) is monotone increasing in U_t, it follows that the τth conditional quantile function of y_t can be obtained as

$$Q_{y_t}\left(\tau | \boldsymbol{y}_{t-1}\right) = \theta_0(\tau) + \theta_1(\tau) y_{t-1} + ... + \theta_m(\tau) y_{t-m} \tag{1.2}$$

where $\boldsymbol{y}_{t-1} = (y_{t-1}, ..., y_{t-m})'$. The transition from (1.1) to (1.2) is an immediate consequence of equivariance to monotone transformations.[1] In (1.2), the quantile autoregressive coefficients may be τ-dependent and thus can vary over the quantiles. The conditioning variables not only shift the location of the distribution of y_t , but also may alter the scale and shape of the conditional distribution. Koenker and Xiao (2006) also show that quantile autoregressive models exhibit a form of asymmetric persistence and temporarily explosive behavior.

However, the linear quantile autoregressive models cannot accommodate many stylized facts such as structural breaks and nonlinearities in macroeconomic and financial time series. The aim of this article is to extend the quantile autoregression of Koenker and Xiao (2006) by modeling nonstationary quantile dynamics. Particularly, I consider the location-scale quantile autoregression in which the location and scale parameters are subject to regime shifts within a quantile. Switching quantile regimes is determined by the outcome of an unobserved state indicator variable that follows a Markov process with unknown transition probabilities. The proposed Markov-Switching Quantile Autoregression (MSQAR) nests the quantile autoregression of Koenker and Xiao (2006) as a special case when conditional distributions are stationary.

[1] See the theorem of equivariance to monotone transformations in Koenker (2005), page 39.

MSQAR is a convenient approach built on the vast literature of Markov-switching time series models.[2] Nonetheless, simply combining quantile autoregressive models with Markov-switching techniques is econometrically infeasible. The challenge is that the objective function of quantile autoregression is a non-likelihood based function generally estimated by nonlinear least square. The non-likelihood based function does not allow make inference on the latent state variable for switching regimes. To solve this problem, I assume that quantile error terms follow a three-parameter asymmetric-Laplace distribution (Yu and Zhang (2005)). This paper shows that maximizing this distribution is mathematically equivalent to minimizing quantile objective functions. Importantly, it also satisfies the restrictive conditions of quantile regression. With this distribution, the inference for switching quantile regimes can be made through the standard Hamilton filter approach (Hamilton (1994)).

This paper adopts Bayesian approach for model estimation. As discussed in Yu and Moyeed (2001), the use of an asymmetric Laplace distribution for error terms provides a natural way to deal with some serious computational challenges through Bayesian quantile regression. Also see Chernozhukov and Hong (2003). In the terminology of Chib and Greenberg (1995), this paper adopts a "block-at-a-time" Metropolis-Hastings sampling to reduce computational cost.[3] This algorithm groups highly correlated parameters as one block to be simultaneously updated at each Metropolis-Hasting step conditional on the remaining blocks, see e.g., Tierney (1994), Ausin and Lopes (2010), Geweke and Tanizaki (2001), among others. To further speed up convergence and to achieve desirable mixing properties in MCMC chains, I employ the adaptive scheme of Gerlach et al. (2011) and Chen et al. (2012), which combines a random walk and an independent kernel Metropolis-Hastings algorithm, each based on a mixture of multivariate normal distributions.

Importantly, in this paper the Bayesian output provides a convenient way to assess the presence and the number of regimes by constructing Bayesian factors through marginal likelihoods in the sense of Chib and Jeliazkov (2001). The empirical illustration in the U.S. financial market clearly supports the presence of regimes in quantile autoregression. In addition, this paper performs a simulation

[2]See e.g., Sims and Zha (2006), Gray (1996), Cheung and Erlandsson (2005), Hamilton and Susmel (1994), Kim et al. (2008), among many others. Guidolin (2012) provides a recent review for the applications of Markov-switching models in empirical finance.

[3]To the best of my knowledge, the work closely relevant to this paper is in parallel developed by Liu and Luger (2015) who have proposed Gibbs sampling approach to estimate Markov-Switching quantile autoregressive models. However, this paper differs from their work in several important ways. Unlike their work that restricts autoregressive coefficients independent of regime changes and only allows regime changes in quantile locations, this paper generalizes the MSQAR model to allow autoregressive coefficients dependent on both regimes and quantiles. However, once the autoregressive coefficients depend on regime changes, their Gibbs sampling algorithm fails to derive the full conditional densities due to the product between the autoregressive coefficients and the lagged quantile location parameters. Therefore, the Metropolis-Hastings within Gibbs sampling approach proposed in this paper is the appropriate approach to address this issue, and can also reduce computational cost.

study to examine the adequacy of using the asymmetric Laplace distribution as the likelihood in regime inference and quantile prediction when the true model errors follow different distributions. The potential effect of ignoring that autoregressive coefficients might be τ-dependent has also been investigated via the simulation study.

The rest of this paper is structured as follows. Section 2 introduces the connection of asymmetric-Laplace distributions to the solution of quantile regressions. Section 3 defines Markov-Switching quantile autoregression. In this section Markov-Switching autoregression is discussed as the benchmark model. Section 4 describes the Bayesian methods in this paper for model estimation. Section 5 introduces the computation method of marginal likelihoods for testing the presence and the number of regimes. Section 6 presents model simulations and results. Section 7 reports the results of the empirical application to the U.S. financial markets. Section 8 concludes this paper.

2 Asymmetric Laplace Distribution Connection

The QAR(m) model of (1.2) can be reformulated in a more conventional regression form as

$$y_t = \theta_0(\tau) + \sum_{l=1}^{m} \theta_l(\tau) y_{t-l} + \varepsilon_t(\tau) \tag{2.1}$$

where $\varepsilon_t(\tau)$ is quantile error terms which follow an asymmetric-Laplace (AL) distribution, denoted by $AL(0, \varsigma, \tau)$, with the density function as

$$f(\varepsilon_t; 0, \varsigma, \tau) = \frac{\tau(1-\tau)}{\varsigma} exp\left\{ -\frac{\varepsilon_t\left(\tau - I(\varepsilon_t \leq 0)\right)}{\varsigma} \right\} \tag{2.2}$$

where $I(\cdot)$ is an indicator function. τ determines the skewness of the distribution, $\varsigma > 0$ is a scale parameter. $AL(0, \varsigma, \tau)$ with the location parameter being zero provides that the τth quantile of the distribution is zero as $Pr\left(\varepsilon_t \leq 0\right) = \tau$, which satisfies the quantile regression condition $\int_{-\infty}^{0} f_\varepsilon(s) ds = \tau$. The asymmetric-Laplace distribution with the density function of (2.2) has the mean and variance, $E(\varepsilon_t) = \varsigma(1-2\tau)/[(1-\tau)\tau]$ and $Var(\varepsilon_t) = \varsigma^2(1-2\tau+2\tau^2)/[(1-\tau)^2\tau^2]$, respectively. See Yu and Zhang (2005) for details. With the assumption of i.i.d. $\varepsilon_t(\tau)$, the sample likelihood function is given by

$$
\begin{aligned}
L(\boldsymbol{\theta}, \tau) \quad = \quad & [\tau(1-\tau)/\varsigma]^{T-m} \\
& exp\left\{-\sum_{t=m+1}^{T} \frac{y_t - Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1}\right)}{\varsigma}\left[\tau - I\left(y_t \leq Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1}\right)\right)\right]\right\}
\end{aligned} \tag{2.3}
$$

In the literature the error density is often left unspecified, see e.g., Koenker and Bassett (1978), Koenker (2005), and Koenker and Xiao (2006), etc. Quantile autoregression is the solution to the following minimization problem

$$
\boldsymbol{\theta}(\tau) = arg \min_{\boldsymbol{\theta}(\tau)\in\mathbb{R}^{m+1}} E\left(\rho_\tau\left(y_t - Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1};\boldsymbol{\theta}(\tau)\right)\right)\right) \tag{2.4}
$$

where $\theta\left(\tau\right) = \left(\theta_0(\tau), ..., \theta_m\left(\tau\right)\right)$ is the parameter space to be estimated. The quantile criterion (check or loss) function $\rho_\tau(\cdot)$ is defined as $\rho_\tau(\varepsilon) = \varepsilon\left(\tau - I(\varepsilon < 0)\right)$ in Koenker and Bassett (1978). Solving the sample analog gives the estimator of $\theta(\tau)$

$$
\hat{\boldsymbol{\theta}}(\tau) = arg \min_{\boldsymbol{\theta}(\tau)\in\mathbb{R}^{m+1}} \sum_{t=1}^{T} \rho_\tau\left(y_t - Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1};\boldsymbol{\theta}(\tau)\right)\right) \tag{2.5}
$$

Recently, Yu and Moyeed (2001), Yu and Zhang (2005) and Gerlach et al. (2011), among others, have illustrated the link between the quantile estimation problem and asymmetric-Laplace distribution. Since the quantile loss function is contained in the exponent of the asymmetric-Laplace likelihood, maximizing the sample likelihood of (2.3) is mathematically equivalent to minimizing the quantile loss function of (2.5). It is important to emphasize that, in practice, the parameter τ is chosen by researchers as quantile levels of interest during parameter estimation and only a single quantile of the distribution of y_t is estimated. More importantly, the asymmetric Laplace distribution transforms the non-likelihood based quantile regression of (2.5) to a likelihood based approach (Chernozhukov and Hong (2003)), so that the inference for the probability of switching regimes is possibly made through Hamilton filter.

3 Markov-Switching Quantile Autoregression

For the τth conditional quantile of y_t, let $\{s_t\}$ be an ergodic homogeneous Markov chain on a finite set $S = \{1, ..., K\}$, with a transition matrix $P(\tau)$ defined by the following transition probabilities

$$\{p_{ij}(\tau) = Pr\left(s_t = j | s_{t-1} = i; \tau\right)\}$$

for $i, j \in S$ and assume s_t follows a first-order Markov chain. The transition probabilities satisfy $\sum_{j \in S} p_{ij}(\tau) = 1$. The stochastic process for s_t is strictly stationary if $p_{ij}(\tau)$ is less than unity and does not take on the value of 0 simultaneously. Specifically, this paper considers constant transition probabilities to keep the model tractable. However, this assumption can be relaxed to time-varying transition probabilities as discussed by e.g., Filardo (1994) and Diebold et al. (1994), among others. In their works, transition probabilities can vary with economic fundamentals.

Using transition probabilities above, this paper defines Markov-Switching quantile autoregressive models (MSQAR) as

$$\begin{aligned} y_t &= Q_{y_t}(\tau | \mathbf{y}_{t-1}; \boldsymbol{\theta}_{s_t}(\tau)) + \varepsilon_t(\tau) \\ &= \theta_{s_t,0}(\tau) + \sum_{l=1}^{m} \theta_{s_t,l}(\tau) y_{t-l} + \varepsilon_t(\tau) \end{aligned} \tag{3.1}$$

Note that if no regimes occur ($K = 1$), the MSQAR model in (3.1) collapses to the QAR model of Koenker and Xiao (2006). In this sense, the proposed MSQAR model nests the quantile autoregression of Koenker and Xiao (2006) as a special case.

Suppose that $\mathbf{y}_t$ can be observed directly but can only make an inference about the value of s_t based on the observations as of date t. With the asymmetric Laplace connection established in previous section, the latent discrete Markov process, $\{s_t\}$, can be inferred by applying the Hamilton filtering approach (Hamilton, 1994) to each quantile level. The likelihood function of the MSQAR model can also be obtained from the Hamilton filter algorithm. *Appendix A* details how to draw the probabilistic inferences about the unobserved states, $\{s_t\}$, given observations on y_t.

The connection to the solution of quantile regression can also be viewed as follows. Based on quantile loss functions, $\mathbf{\Theta}(\tau)$ is solved for the following minimization problem

$$\underset{\mathbf{\Theta}(\tau)}{min}\; E\left(\sum_{j \in S} \rho_\tau \left(y_t - Q_{y_t}\left(\tau | s_t = j,\ \mathbf{y}_{t-1}; \mathbf{\Theta}(\tau)\right)\right) I\left(s_t = j; \tau\right)\right) \tag{3.2}$$

where $\mathbf{y}_t = \{y_t, y_{t-1}, ..., y_1, y_0\}$. Apply the law of iterated expectation to rewrite (3.2) as

$$\underset{\mathbf{\Theta}(\tau)}{min} \sum_{j \in S} E\left[\rho_\tau\left(y_t - Q_{y_t}\left(\tau | s_{t,\tau} = j,\ \mathbf{y}_{t-1}; \mathbf{\Theta}(\tau)\right)\right) Pr\left(s_t = j | \mathbf{y}_t, \tau; \mathbf{\Theta}(\tau)\right)\right] \tag{3.3}$$

Provided that τ is chosen by researchers of interest, maximizing the likelihood of (A.3) is mathematically equivalent to the minimization of (3.2), since the likelihood function can be alternatively rewritten as $L(\mathbf{\Theta}; \tau) = \prod_{t=1}^{T} \sum_{j \in S} f\left(y_t | s_t = j, \mathbf{y}_{t-1}, \tau; \mathbf{\Theta}(\tau)\right) Pr\left(s_t = j | \mathbf{y}_t, \tau; \mathbf{\Theta}(\tau)\right)$ with $Pr(s_t = j | \mathbf{y}_t, \tau; \mathbf{\Theta}(\tau)) = \sum_{i \in S} p_{ij}(\tau) \xi_{i,t-1|t-1}(\tau)$ (see *Appendix A*). However, $Pr\left(s_t = j | \mathbf{y}_t, \tau; \mathbf{\Theta}(\tau)\right)$ cannot be filtered by using the nonlinear least square estimation of (3.3); therefore, the likelihood function of the asymmetric Laplace distribution must be used to infer transition probabilities.

Given the filtered probability for s_t and parameter estimates, it is then straightforward to forecast the one-step-ahead τth quantile of y_{t+1} at time t conditional on knowing $s_{t+1,\tau}$,

$$Q_{y_{t+1}}\left(\tau | s_{t+1} = j, \mathbf{y}_t; \boldsymbol{\theta}_j(\tau)\right) = \theta_{j,0}(\tau) + \sum_{l=0}^{m-1} \theta_{j,l+1}(\tau) y_{t-l} \tag{3.4}$$

Further, from (A.2), the quantile forecast for $t+1$ conditional on time t is obtained as

$$Q_{y_{t+1}}\left(\tau | \mathbf{y}_t; \boldsymbol{\theta}(\tau)\right) = \sum_{j=1}^{k} Q_{y_{t+1}}\left(\tau | s_{t+1} = j, \mathbf{y}_t; \boldsymbol{\theta}_j(\tau)\right) Pr\left(s_{t+1} = j | \mathbf{y}_t, \tau; \mathbf{\Theta}(\tau)\right) \tag{3.5}$$

which is to multiply the appropriate forecast of the quantile in the jth regime given by (3.4) with the probability that the process will be in that regime given by (A.2), and to sum those products for every regime together. Note that h-step-ahead forecasts for $h \geq 2$ require different approaches since it involves forecasts of y_{t+h-1} in (3.4) for $Q_{y_{t+h}}\left(\tau | s_{t+h-1,\tau} = j, \mathbf{y}_{t+h-1}; \boldsymbol{\theta}_j(\tau)\right)$, as shown in Cai (2010).

An alternative method to obtain Markov-Switching conditional quantiles is to consider an autoregressive error structure while the mean and variance are subject to regime changes

$$y_t = \beta_{s_t,0} + \sum_{l=1}^{m} \beta_{s_t,l} y_{t-l} + \sigma_{s_t} \varepsilon_t \tag{3.6}$$

where β_{s_t} and σ_{s_t} are conditional mean coefficients and standard deviation functions whose values depend on the outcome of s_t. This model in (3.6) corresponds to the Markov-switching autoregressive (MSAR) model used by many studies, e.g., Garcia and Perron (1996), Kim and Nelson (1999), Engel and Kim (1999), Engel (1994), Cheung and Erlandsson (2005), Kim et al. (2008), etc., to analyze macroeconomic and financial dynamics under the assumption that ε_t follows either a standard normal or a standardized student-t distribution. The τth conditional quantile function implied by (3.6) is

$$Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1}\right) = \beta_{s_t,0}(\tau) + \sum_{l=1}^{m} \beta_{s_t,l} y_{t-l} \tag{3.7}$$

where $\beta_{s_t,0}(\tau) = \beta_{s_t,0} + \sigma_{s_t} Q_{\varepsilon_t}(\tau)$ and $Q_{\varepsilon_t}(\tau)$ is the τth theoretical quantile of a distribution assumed for ε_t. Observe that $\beta_{s_t,0}$, σ_{s_t} and $\beta_{s_t,1}, ..., \beta_{s_t,m}$ are τ-independent, meaning that they determine the time-series behavior of the conditional quantiles of y_t regardless of the probability level $\tau \in (0,1)$. The conditional quantile in (3.7) implied by a MSAR model is a location quantile model in which the difference across conditional quantiles is determined completely by the quantile intercept, such as the theoretical quantiles ($Q_{\varepsilon_t}(\tau)$) of the assumed distribution of autoregressive error terms. Notably, this implied conditional quantile function is simply a special case of the proposed MSQAR model if the autoregressive coefficients $\theta_{s_t,l}(\tau)$ in (3.1) are independent of τ. According to Corollary 1 of Koenker and Xiao (2006) a quantile autoregressive process may allow for some (transient) forms of explosive behavior while maintaining stationary in the long run. This explosive behavior, nonetheless, is not allowed in MSAR models.

For Markov-Switching time series models, one must use some identification restrictions for both MSQAR and MSAR models to avoid the label switching issue. See Bauwens et al. (2010) and Hamilton et al. (2007) for a discussion. Similarly in this paper, regimes for MSQAR models are labeled by the restrictions on quantile intercepts, for example, $\theta_{1,0}(\tau) > ... > \theta_{K,0}(\tau)$. In the empirical application of this paper, the transition probabilities are allowed but not imposed dependent on τ. The intuition is that even though economic states are common across quantiles implying the same unconditional probabilities, no theories show that regime persistence should be the same across quantiles.[4] The potential difference in regime persistence might be explained by asymmetric risk preference in that markets in downturn periods (lower tails) are more likely intervened by central governments or appear to have higher level of fluctuations due to market uncertainties than in upturn periods (upper tails). To obtain some insights on this empirical question, regime persistence in this paper is allowed to be driven by data across quantiles.

[4]Unconditional probabilities are defined as $\pi_j(\tau) = Pr\left(s_t = j; \tau\right)$ for $j \in S$. For example, if $S = \{1, 2\}$, unconditional probabilities of π_1 and π_2 can be obtained as $(1 - p_{22})/(2 - p_{11} - p_{22})$ and $(1 - p_{11})/(2 - p_{11} - p_{22})$, respectively. The regime persistence for regime 1 and 2 is directly given by p_{11} and p_{22}, respectively. To illustrate the intuition, assume $P_{MSAR} = \{p_{11} = 0.95, p_{22} = 0.9\}$ and $P_{MSQAR}(\tau) = \{p_{11}(\tau) = 0.80, p_{22}(\tau) = 0.60\}$, both give the same unconditional probabilities $\pi_1 = \pi_1(\tau) = 0.667$ and $\pi_2 = \pi_2(\tau) = 0.333$ but different regime persistence.

4 Bayesian Inference

MSQAR models are non-linear and involve indicator functions, which introduce kinks and discontinuities into the sample likelihood function in (A.3). In addition, less observations fall in more extreme quantiles, which leads to the potential small sample issue. These issues make classical methods such as MLE very difficult for model estimation. In this paper, I instead prefer to use Bayesian MCMC methods to learn about the model parameters.

Given the sample realizations, y_t for $t = 1, ..., T$, the posterior distribution of $\mathbf{\Theta}(\tau)$ takes the usual form: $p(\mathbf{\Theta}(\tau)|\mathbf{y}_t) \propto L(\mathbf{y}_t|\mathbf{\Theta}(\tau))\pi(\mathbf{\Theta}(\tau))$, where $L(\mathbf{y}_t|\mathbf{\Theta}(\tau))$ is the sample likelihood function and $\pi(\mathbf{\Theta}(\tau))$ is the prior distribution. Yu and Moyeed (2001) and Cai and Stander (2008) prove that the posterior distribution is proper under the improper prior for general quantile regression models. In this paper, the prior distribution is taken as uniform over $\Xi(\tau)$, the admissible parameter space of $\mathbf{\Theta}(\tau)$, i.e., satisfying the label switching restrictions. The prior for the scale parameter is $\pi(\varsigma(\tau)) \propto \varsigma(\tau)^{-1}$ also used in Gerlach et al. (2011).

Just like Vrontos et al. (2002) and Ausin and Lopes (2010), I also find that MCMC mixing can be improved and the computational cost reduced by using simultaneous updating of the highly correlated parameter groups at each Metropolis-Hastings (MH) step. In the terminology of Chib and Greenberg (1995), this approach is therefore based on a "block-at-a-time" MH sampler which is carried out by cycling repeatedly through draws of each parameter block conditional on the remaining parameter blocks. Let $\mathbf{\Theta}(\tau) = (\boldsymbol{P}(\tau), \boldsymbol{\theta}_1(\tau), ..., \boldsymbol{\theta}_K(\tau))$ represent the blocks of the population parameters. $\boldsymbol{P}(\tau) = (p_{ij}(\tau))$ contains all transition probability parameters and $\boldsymbol{\theta}_j(\tau)$ includes all parameters in the jth regime for $j = 1, ..., K$. Hence, the parameters in $\mathbf{\Theta}(\tau)$ are grouped in $K+1$ blocks and the parameters of each block are simultaneously updated conditional on the remaining blocks.

This paper implements the MH sampler according to the adaptive scheme of Gerlach et al. (2011) and Chen et al. (2012) which combines the random walk MH (RW-MH) and the independent kernel MH (IK-MH) algorithms, each based on a mixture of multivariate normal distributions. The random walk part of this scheme is designed to allow occasional large jumps, perhaps away from local modes, thereby improving the chances that the Markov chain will explore the posterior distribution space. Hence, this adaptive scheme allows for further speeding convergence and achieving desirable mixing properties in MCMC chains.

To illustrate this adaptive algorithm in the block-at-a-time MH sampler, I rewrite the notation of the parameter blocks as $\boldsymbol{\Theta}(\tau) = (\boldsymbol{\theta}_1(\tau), \boldsymbol{\theta}_2(\tau), ..., \boldsymbol{\theta}_{K+1}(\tau))$, where $\boldsymbol{\theta}_1(\tau) = \boldsymbol{P}(\tau)$ and $\boldsymbol{\theta}_j(\tau)$ denotes the parameters in the $(j-1)$th regime for $j = 2, ..., K+1$. And, let $\boldsymbol{\Theta}_{-j}(\tau)$ denote the vector $\boldsymbol{\Theta}(\tau)$ excluding the block $\boldsymbol{\theta}_j(\tau)$ for $j = 1, ..., K+1$. Starting at $g = 1$ with $\boldsymbol{\Theta}^{[1]}(\tau) = \left(\boldsymbol{\theta}_1^{[1]}(\tau), ..., \boldsymbol{\theta}_{K+1}^{[1]}(\tau)\right)$, the G_1 random walk MH iterations for $\boldsymbol{\Theta}(\tau)$ proceed as follows

Step1. Increment g by 1 and set $\boldsymbol{\Theta}^{[g]}(\tau)$ equal to $\boldsymbol{\Theta}^{[g-1]}(\tau)$.

Step2. For $i = 1, ..., k+1$ in turn, generate $\boldsymbol{\theta}_i^c(\tau)$ as

$$\boldsymbol{\theta}_i^c(\tau) = \boldsymbol{\theta}_i^{[g]}(\tau) + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \rho N\left(\mathbf{0}, diag\left\{b_i\right\}\right) + (1-\rho)\, N\left(\mathbf{0}, \omega diag\left\{b_i\right\}\right)$$

and replace $\boldsymbol{\theta}_i^{[g]}(\tau)$ in $\boldsymbol{\Theta}^{[g]}(\tau)$ by $\boldsymbol{\theta}_i^c(\tau)$ with the probability $min\left(\zeta_i, 1\right)$, where

$$\zeta_i = \frac{L\left(\boldsymbol{y}_t | \boldsymbol{\theta}_i^c(\tau), \boldsymbol{\Theta}_{-i}^{[g]}(\tau)\right) \pi\left(\boldsymbol{\theta}_i^c(\tau), \boldsymbol{\Theta}_{-i}^{[g]}(\tau)\right)}{L\left(\boldsymbol{y}_t | \boldsymbol{\Theta}^{[g]}(\tau)\right) \pi\left(\boldsymbol{\Theta}^{[g]}(\tau)\right)}$$

Step3. If $g < G_1$, go to Step 1.

Upon completion, these first G_1 iterations yield the burn-in sample. Following Chen et al. (2012), I set $\rho = 0.95$, $\omega = 100$, and tune the positive number b_i so that the empirical acceptance rate lies in the range $(0.2, 0.45)$ for the ith block. Tuning is done every 100 iterations by increasing b_i when the acceptance rate in the last 100 iterations is higher than 0.45, or decreasing b_i when that rate is lower than 0.2.

At the end of the first G_1 iterations, the burn-in sample mean $\boldsymbol{\mu}_i(\tau)$ and covariance matrix $\boldsymbol{\Sigma}_i(\tau)$ of $\boldsymbol{\theta}_i(\tau)$ with corresponding lower triangular Cholesky factor $\boldsymbol{\Sigma}_i^{1/2}(\tau)$ are computed for $i = 1, ..., K+1$. The MCMC sampling scheme then continues for G_2 additional iterations according to the following independent kernel MH steps:

Step4. Increment g by 1 and set $\boldsymbol{\Theta}^{[g]}(\tau)$ equal to $\boldsymbol{\Theta}^{[g-1]}(\tau)$.

Step5. For $i = 1, ..., k+1$ in turn, generate $\boldsymbol{\theta}_i^c(\tau)$ as

$$\boldsymbol{\theta}_i^c(\tau) = \boldsymbol{\mu}_i(\tau) + \boldsymbol{\Sigma}_i^{1/2}(\tau)\boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \rho N\left(\mathbf{0}, \boldsymbol{I}\right) + (1-\rho)\, N\left(\mathbf{0}, \omega \boldsymbol{I}\right)$$

and replace $\boldsymbol{\theta}_i^{[g]}(\tau)$ in $\boldsymbol{\Theta}^{[g]}(\tau)$ by $\boldsymbol{\theta}_i^{c}(\tau)$ with the probability $min\left(\zeta_i, 1\right)$, where

$$\zeta_i = \frac{L\left(\boldsymbol{y}_t|\boldsymbol{\theta}_i^c(\tau), \boldsymbol{\Theta}_{-i}^{[g]}(\tau)\right)\pi\left(\boldsymbol{\theta}_i^c(\tau), \boldsymbol{\Theta}_{-i}^{[g]}(\tau)\right)q\left(\boldsymbol{\theta}_i^{[g]}(\tau)\right)}{L\left(\boldsymbol{y}_t|\boldsymbol{\Theta}^{[g]}(\tau)\right)\pi\left(\boldsymbol{\Theta}^{[g]}(\tau)\right)q\left(\boldsymbol{\theta}_i^c(\tau)\right)}$$

$$\begin{aligned} q\left(\boldsymbol{\theta}_i(\tau)\right) \quad \propto \quad & \rho exp\left\{-\frac{1}{2}\left(\boldsymbol{\theta}_i(\tau) - \boldsymbol{\mu}_i(\tau)\right)' \boldsymbol{\Sigma}_{i,\tau}^{-1}\left(\boldsymbol{\theta}_i(\tau) - \boldsymbol{\mu}_i(\tau)\right)\right\} \\ & + \frac{1-\rho}{\omega^{dim(\boldsymbol{\theta}_i(\tau))/2}} exp\left\{-\frac{1}{2}\left(\boldsymbol{\theta}_i(\tau) - \boldsymbol{\mu}_i(\tau)\right)' \boldsymbol{\Sigma}_i^{-1}(\tau)\left(\boldsymbol{\theta}_i(\tau) - \boldsymbol{\mu}_i(\tau)\right)\right\} \end{aligned}$$

Step6. If $g < G_1 + G_2$, go to Step 4.

Observe that the use of $\boldsymbol{\Sigma}_i(\tau)$ in Step 5 accounts for the posterior correlation among the elements of $\boldsymbol{\theta}_i(\tau)$, thereby improving the efficiency of the Markov chain. The parameter updates are sequentially repeated until the convergence of the Markov chain is achieved. The burn-in draws are discarded, and the steps are iterated a large number of times to generate draws from which the desired features (means, variances, quantiles, etc.) of the posterior distribution can be estimated consistently.

In this paper, $G_1 = 50,000$ for the random walk MH sampler and $G_2 = 50,000$ with a thinning of 5 for the independent kernel MH sampler, resulting in posterior samples comprising 10,000 draws. The convergence of the IK-MH Markov chains is assessed using the Geweke (1992) test. For each parameter, I also assess the accuracy of its posterior mean by computing the numerical standard error (NSE) according to the batch-means method (Ripley, 1987). In all simulated and real data examples of this paper, it is observed that MCMC chains are well converged inside 50,000 iterations.

5 Specification Issues

In this section I address two issues with the model specification that arise when the presence and the number of regimes are considered. I resolve both these issues with the marginal likelihood $\pi(\boldsymbol{y}|\tau)$, which is the key input to the computation of a model's posterior probability (Kass and Raftery, 1995). The marginal likelihood can easily be computed for MSQAR models estimated for different numbers of regimes, so it then becomes natural to base inference about the presence and the number of regimes by Bayesian model selection.

Denote the τth MSQAR model with k number of regimes as M_k with the model-specific parameter vector Θ_k for $k = 1, ..., K$. In this context, Bayesian model selection proceeds by pairwise comparison of the models in the collection of $\{M_k\}$ through their posterior odds ratio, which for any two models M_k and M_g is written as $Pr\left(M_k|\boldsymbol{y},\tau\right)/Pr\left(M_g|\boldsymbol{y},\tau\right) = \left[Pr\left(M_k|\tau\right)/Pr\left(M_g|\tau\right)\right]\left[\pi\left(\boldsymbol{y}|M_k,\tau\right)/\pi\left(\boldsymbol{y}|M_g,\tau\right)\right]$. The first fraction on the right-hand side of the odds ratio is known as the prior odds and the second as the Bayes factor. If the interest is to test the presence of regimes, the hypothesis should be $H_0 : \ k = 1 \ vs. \ H_1 : \ g > 1$. Note that for $k = 1$, the corresponding MSQAR model collapses to the QAR model of Koenker and Xiao (2006) which can be estimated by the same Bayesian approach described in previous section. If the interest is empirical to the choice of the number of regimes, then the hypothesis should be $H_0 : \ k = a \ vs. \ H_1 : \ g > a$ for $a > 1$.

The marginal likelihood is related to the prior, posterior, and sample density functions via the equality

$$\pi(\boldsymbol{y}|M_k,\tau) = \frac{f\left(\boldsymbol{y}|M_k,\Theta_k(\tau)\right)\pi(\Theta_k(\tau)|M_k,)}{\pi(\Theta_k(\tau)|\boldsymbol{y},M_k)}$$

which holds at all admissible points of the parameter space. So for given values $\Theta_k(\tau)^*$ of the model parameters, an estimate of $log\,\pi(\boldsymbol{y}|M_k,\tau)$ can be obtained by using

$$log\,\hat{\pi}(\boldsymbol{y}|M_k,\tau) = log\,f\left(\boldsymbol{y}|M_k,\Theta_k(\tau)^*\right) + log\,\pi\left(\Theta_k(\tau)^*|M_k\right) - log\,\hat{\pi}\left(\Theta_k(\tau)^*|\boldsymbol{y},M_k\right) \tag{5.1}$$

where $\hat{\pi}\left(\Theta_k(\tau)^*|\boldsymbol{y},M_k\right)$ is an estimate of the posterior ordinate at the chosen parameter values. In principle, $\Theta_k(\tau)^*$ could be any point in the space of admissible values. As Chib and Jeliazkov (2001) explains, however, efficiency considerations dictate that $log\,\pi\left(\boldsymbol{y}|M_k,\tau\right)$ is likely to be more accurately evaluated at a high density point, such as the posterior mean or mode, rather than at a point in the tails of the posterior. Here, I use the posterior mean for ease of computation.

The first term on the right-hand side of (5.1) is the log of the MSQAR likelihood function in (A.3) evaluated at $\Theta_k\left(\tau\right)^*$. The second term on the right-hand side of (5.1) is the log of the prior density evaluated at $\Theta_k(\tau)^*$. The log of the posterior ordinate estimate appearing as the third term on the right-hand side of (5.1) requires further computations as

$$\pi\left(\Theta_k(\tau)^*|\boldsymbol{y},M_k\right) = \frac{E_1\left\{\alpha\left(\Theta_k(\tau),\Theta_k(\tau)^*|\boldsymbol{y},M_k\right)q\left(\Theta_k(\tau),\Theta_k(\tau)^*|\boldsymbol{y},M_k\right)\right\}}{E_2\left\{\alpha\left(\Theta_k(\tau)^*,\Theta_k(\tau)|\boldsymbol{y},M_k\right)\right\}}$$

where $\alpha = min(\zeta, 1)$ denote the probability of move (probability of accepting the proposed value). The numerator expectation E_1 is with respect to the posterior distribution $\pi(\Theta_k(\tau)|\boldsymbol{y}, M_k)$ and the denominator expectation E_2 is with respect to the proposal density $q(\Theta_k(\tau)^*, \Theta_k(\tau)|\boldsymbol{y}, M_k)$. This implies that a simulation-consistent estimate of the posterior ordinate is available as

$$\hat{\pi}\left(\Theta_k(\tau)^* |\boldsymbol{y}, M_k\right) = \frac{R^{-1}\sum_{r=1}^{R} \alpha\left(\Theta_k^{[r]}(\tau), \Theta_k(\tau)^*|\boldsymbol{y}, M_k\right) q\left(\Theta_k^{[r]}(\tau), \Theta_k(\tau)^*|\boldsymbol{y}, M_k\right)}{J^{-1}\sum_{j=1}^{J} \alpha\left(\Theta_k(\tau)^*, \Theta_k^{[j]}(\tau)|\boldsymbol{y}, M_k\right)}$$

where $\left\{\Theta_k^{[r]}(\tau)\right\}$ are the sampled draws from the posterior distribution (for example, from the independent kernel MH step in this paper) and $\left\{\Theta_k^{[j]}(\tau)\right\}$ are draws from the proposal density $q(\Theta_k(\tau)^*, \Theta_k(\tau)|\boldsymbol{y}, M_k)$, given the fixed value $\Theta_k(\tau)^*$.

Chib and Jeliazkov (2001) has also provided the posterior ordinate for multiple parameter blocks as

$$\pi\left(\Theta_j(\tau)^*|\boldsymbol{y}, \Theta_1(\tau)^*, ..., \Theta_{j-1}(\tau)^*\right) = \frac{E_1\left\{\alpha\left(\Theta_j(\tau), \Theta_j(\tau)^*|\boldsymbol{y}, \Psi_{j-1}^*, \Psi^{j+1}\right) q\left(\Theta_j(\tau), \Theta_j(\tau)^*|\boldsymbol{y}, \Psi_{j-1}^*, \Psi^{j+1}\right)\right\}}{E_2\left\{\alpha\left(\Theta_j(\tau)^*, \Theta_j(\tau)|\boldsymbol{y}, \Psi_{j-1}^*, \Psi^{j+1}\right)\right\}}$$

where $\Theta_j(\tau)$ is the jth parameter block for $j = 1, ..., K+1$, $\Psi_{j-1} = (\Theta_1(\tau), ..., \Theta_{j-1}(\tau))$ and $\Psi^{j+1} = (\Theta_{j+1}(\tau), ..., \Theta_{K+1}(\tau))$. E_1 is the expectation with respect to conditional posterior $\pi\left(\Theta_j(\tau), \Psi^{j+1}|\boldsymbol{y}, \Psi_{j-1}^*\right)$ and E_2 is with respect to the conditional product measure $\pi\left(\Psi^{j+1}|\boldsymbol{y}, \Psi_j^*\right) q\left(\Theta_j(\tau)^*, \Theta_j(\tau)|\boldsymbol{y}, \Psi_{j-1}^*, \Psi^{j+1}\right)$. These two integrals can be estimated as before from the Metropolis-Hasting MCMC output. See Chib and Jeliazkov (2001) for detail steps. Upon substitution of all the estimated posterior ordinates into (5.1), the estimate of (the log of) the marginal likelihood is

$$log\,\hat{\pi}(\boldsymbol{y}|M_k, \tau) = log\,f\left(\boldsymbol{y}|M_k, \Theta_k(\tau)^*\right) + log\,\pi\left(\Theta_k(\tau)^*|M_k\right) - \sum_{j=1}^{K+1} log\,\hat{\pi}\left(\Theta_{j,k}(\tau)^*|\boldsymbol{y}, M_k\right)$$

The simulation-sample sizes in the numerator and the denominator are allowed to be different, although in practice I set them to be equal. The marginal likelihood of the MSQAR model is available almost as soon as the full MCMC run is finished.

6 Simulation

This section carries on a Monte Carlo study to examine the performance of the adaptive "block-at-a-time" MH sampler in MSQAR model estimation. There are in fact two main purposes for this study: (i) to verify the adequacy of using the asymmetric Laplace distribution as the likelihood when the true model errors follow different distributions, and (ii) to investigate the potential effect on regime inference and quantile prediction of ignoring that autoregressive coefficients might be $\tau-$dependent. Specifically, two simulation experiments are designed to achieve the second purpose: (1) an experiment using data simulated from a MSAR model of (3.7) with autoregressive coefficients independent of τ, and (2) using the simulation method in Koenker and Xiao (2006) to simulate data by allowing autoregressive coefficients depending on both s_t and τ as in (3.1).

6.1 Monte Carlo Simulation: Design 1

This experiment simulates data from a MSAR model with 2 regimes as

$$y_t = \begin{cases} 2.0 + 0.2y_{t-1} + 0.5\varepsilon_t, & s_t = 1 \\ -2.0 + 0.4y_{t-1} + \varepsilon_t, & s_t = 2 \end{cases}$$

with the transition probability matrix $P = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$. The true parameter values are referenced based on empirical data estimations in this paper. Two underlying distributions are considered for error terms, including a standard normal distribution ($N(0,1)$) and a standardized student-t distribution with 3 degrees of freedom (t_3). The τth theoretical conditional quantile of y_t can be expressed in a MSQAR form as

$$Q_{y_t}(\tau|\mathbf{y}_{t-1};\Theta(\tau)) = \begin{cases} \theta_{10}(\tau) + \theta_{11}y_{t-1}, & s_t = 1 \\ \theta_{20}(\tau) + \theta_{21}y_{t-1}, & s_t = 2 \end{cases}$$

with the corresponding quantile parameters as $\theta_{10}(\tau) = 2.0 + 0.5Q_{\varepsilon_t}(\tau)$, $\theta_{11} = 0.2$, $\theta_{20}(\tau) = -2.0 + Q_{\varepsilon_t}(\tau)$, and $\theta_{21} = 0.4$. $Q_{\varepsilon_t}(\tau)$ is the τth theoretical quantile of a underlying distribution. As seen, the autoregressive coefficients, θ_{11} and θ_{21}, are $\tau-$independent.

200 data replications are simulated for each underlying distribution. 5,000 observations are generated for each data replication but only the last 500 observations are kept for estimation in order to reduce initial effects. MSQAR models are examined in different sample sizes, $T = \{200, 500\}$ and quantile levels, $\tau = \{0.05, 0.25, 0.5, 0.75, 0.95\}$.

Table 1 reports the estimation results. This table includes the true quantile parameters (True), posterior means (PM), standard errors (Std), the root of mean squared errors (RMSE), and the mean absolute deviation (MAD). RMSE and MAD errors in Table 1 are small over different quantile levels and distributions. The small difference between the true and estimated parameters indicates the reasonable accuracy in model estimation. The small standard errors also show a favorable precision in model estimation. Furthermore, the accuracy and the precision of model estimations are improved with the increase in sample sizes considered due to the reduction in RMSEs, MADs and standard errors. As expected, the model estimation for the less extreme quantiles present smaller RMSE and MAD errors than extreme quantiles.

[Table 1 about here]

Figure 1 plotting the posterior kernel densities of parameter estimates along with true parameters indicated by the vertical lines. Figure 1 shows that the posteriors well contain the true quantile parameters with a slightly better performance for $\tau = 0.5$. In many cases, the posteriors appear skewed but still with most of the density concentrated near the true parameter values. To save space, Figure 1 plots results for $\tau = 0.05, 0.5, 0.95$ and $N = 200$ from the normal distribution. Other results are similar and available upon request.

[Figure 1 about here]

The accuracy of estimation is further assessed by examining the deviations between the true conditional quantile $Q_{y_t}(\tau|\mathbf{y}_{t-1};\Theta)$ implied by the true MSAR models and $Q_{y_t}(\tau|\mathbf{y}_{t-1};\hat{\Theta}(\tau))$, the directly estimated conditional quantile by the MSQAR model. More precisely, the mean absolute deviation error (MADE) is computed across observations as

$$MADE = \frac{1}{T}\sum_{t=1}^{T}\left|Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1};\Theta\right) - Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1};\hat{\Theta}(\tau)\right)\right|$$

where $\hat{\Theta}(\tau)$ are the posterior means of the parameters and the MADE is computed for each DGP replication. The results are shown in Table 2, where for each DGP configuration the entries are the median of the 200 MADEs reported with the corresponding lower 5% and upper 95% quantiles in square brackets. Furthermore, denote MSAR_N_N as the MSAR model estimated with the standard normal distribution assumption for data simulated from standard normal and MSAR_t_t as the MSAR model estimated with the student-t distribution assumption for data simulated from t_3. As expected, MSAR_N_N and MSAR_t_t as the true models obtain the best estimation precision.

Nonetheless, using the asymmetric Laplace distribution as the likelihood also achieves reasonable quantile predictability comparable to the true models. Compared to MSAR_N_t and MSAR_t_N, the MSQAR model estimation appears to have smaller MADE values in general with a narrowing of the range between the lower and upper MADE quantiles. Of course, increasing the sample size improves the estimation precision at all quantile levels along with a narrowing of the range between the lower and upper MADE quantiles. Basically, this simulation experiment shows that using asymmetric Laplace distribution as the likelihood achieves reasonable accuracy and precision in model estimation, regime predictability and quantile prediction when the true model errors follow different distributions.

[Table 2 about here]

6.2 Monte Carlo Simulation: Design 2

To investigate the potential effect of τ-dependent autoregressive coefficients on regime inference and quantile prediction, this experiment simulates data by modifying the approach of Koenker and Xiao (2006) as follows. Randomly draw a value z from $U(0,1)$. If $0 < z < 0.25$, then generates y_t as

$$y_t = \begin{cases} 2 + 0.1y_{t-1} + 0.5\varepsilon_t(z), & s_t = 1 \\ -2 + 0.3y_{t-1} + \varepsilon_t(z), & s_t = 2 \end{cases}$$

If $0.25 \leq z \leq 0.75$, then generates y_t as

$$y_t = \begin{cases} 2 + 0.4y_{t-1} + 0.5\varepsilon_t(z), & s_t = 1 \\ -2 + 0.6y_{t-1} + \varepsilon_t(z), & s_t = 2 \end{cases}$$

If $0.75 < z < 1$, then generates y_t as

$$y_t = \begin{cases} 2 + 0.7y_{t-1} + 0.5\varepsilon_t(z), & s_t = 1 \\ -2 + 0.9y_{t-1} + \varepsilon_t(z), & s_t = 2 \end{cases}$$

where $\varepsilon_t(z)$ is the theoretical quantile value given the probability z and the transition probability matrix $P = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$. Similarly, a standard normal distribution ($N(0,1)$) and a standardized student-t distribution with 3 degrees of freedom (t_3) are considered for $\varepsilon_t(z)$. Therefore, the corresponding theoretical conditional quantile of y_t is given by

$$Q_{y_t}(\tau|\mathbf{y}_{t-1};\Theta(\tau)) = \begin{cases} \theta_{10}(\tau) + \theta_{11}(\tau)y_{t-1}, & s_t = 1 \\ \theta_{20}(\tau) + \theta_{21}(\tau)y_{t-1}, & s_t = 2 \end{cases}$$

where $\theta_{10}(\tau) = 2+0.5Q_{\varepsilon_t}(\tau)$, $\theta_{20}(\tau) = -2+Q_{\varepsilon_t}(\tau)$ and $\{\theta_{11}(\tau), \theta_{21}(\tau)\} = \{(0.1, 0.3), (0.4, 0.6), (0.7, 0.9)\}$ for $0 < \tau < 0.25$, $0.25 \leq \tau \leq 0.75$, and $0.75 \leq \tau < 1$, respectively. Compared to Simulation Design 1, this experiment simulates data from a location-scale quantile autoregressive model with the autoregressive coefficients varying with τ from lower to upper tails. In this experiment the models are estimated only for $\tau = \{0.05, 0.5, 0.95\}$, one quantile level for each simulation segment.

Table 3 reports the simulation results for Simulation Design 2. Similar to the results in Simulation Design 1, the MSQAR model estimation using asymmetric Laplace distribution as the likelihood achieves reasonable accuracy and precision with the improvement as the increase in sample sizes.

[Table 3 about here]

However, the results from Tables 4 tell very different stories from those in Tables 2. Table 4 reporting the accuracy of quantile predictions shows that the MSQAR model with asymmetric Laplace distribution as the likelihood achieves the greatest performance of quantile predictions due to the smallest MADE values among all competing models estimated. For example, for data simulated from normal distribution with $\tau = 0.05$ and $N = 200$, the MSQAR model has MADE=1.181 smaller than 1.465 and 1.500 from the MSAR_N_N and MSAR_t_N models, respectively. In addition, the intervals between lower and upper MADE quantiles of the MSQAR model are not only narrower but also lower than those from the MSAR model. This simulation result indicates both higher estimation precision and accuracy of the MSQAR model estimation than the benchmark model. Basically, the results in Table 4 further consolidate the non-negligible effects of τ-dependent autoregressive coefficients on quantile prediction. Overall, the simulation in this subsection demonstrates that conditional quantiles implied by a MSAR model are too restrictive with the assumption of autoregressive coefficients independent of τ, and thus lead to substantial bias in quantile estimation.

[Table 4 about here]

7 Empirical Illustration

Many empirical studies have applied quantile autoregressive models to estimate financial risks. In this section I illustrate the potential of the proposed MSQAR model in estimating S&P 500 index returns with regime changes in tails for market risk assessment across quantiles. Monthly (from January 1926 to February 2013 with 1047 months) and weekly (from January 1950 to February 2013 with 3294 weeks) S&P 500 index returns are taken from the Center for Research in Security Prices (CRSP). Figure 2 plots the time series of the S&P index returns.

[Figure 2 about here]

Particularly, the interest in this empirical illustration is given to the presence of regimes, regime identification and the statistical significance of the conditional quantile estimation. Besides the MSAR model, an additional benchmark model, the CAViaR model (conditional autoregressive Value-at-Risk by regression quantiles) of Engle and Manganelli (2004), is also considered for comparison.[5] The marginal likelihood described in Section 5 is computed to test the presence of regimes by constructing the Bayesian factors.[6] This paper defines that regime 2 represents more extreme outcomes than regime 1. For instance, at lower tails, quantiles of regime 2 should be more negative or farther into the left tail areas than those of regime 1, which is mostly associated with the periods of economic recessions and crises. In contrary, at upper tails, quantiles of regime 2 should be more positive or farther into the right tail areas than those of regime 1.

Table 5 reports the MSQAR model estimation results for monthly and weekly S&P 500 returns.[7] The results show that the numerical standard errors are small and the Markov chain appears to be converged well as indicated by the generally insignificant values of the Geweke (1992) test statistic. Table 5 also shows that the quantile intercepts (θ_{10}) monotonically increase with the increase of

[5] In this paper, I estimate CAViaR model specified with the symmetric absolute value, see Engle and Manganelli (2004), page 369. Xiao and Koenker (2009) show that this CAViaR specification is conveniently the corresponding quantile function of the modified linear $GARCH(p, q)$ model of Taylor (1986).

[6] Despite that this paper focuses on the presence of regimes, the choice of the number of regimes with more than 2 regimes, such as Nalewaik (2011) and Baele et al. (2014), etc., can be tested in a similar way.

[7] The tests based on ACF and PACF (not reported here, but available upon request) reject autocorrelation in the residuals of the estimated models up to 26 lags, which indicates that the MSQAR models with the lag of order one in Table 5 are statistically sufficient to capture the dynamics of the return data. Furthermore, in nonlinear settings where the number of parameters increases with the number of regimes it is very convenient to choose parsimonious models that require a low number of parameters increases with the number of regimes.

quantile levels, and the autoregressive coefficients of regime 2 ($\theta_{21\tau}$) are larger in magnitude than those of regime 1 ($\theta_{11\tau}$). Furthermore, Figure 3 plots the estimated parameters over quantiles with the 5% and 95% credible intervals of posterior distributions. The intervals of quantile autoregressive coefficients exclude zero for lower and upper tails, while the intervals around $\tau = 0.5$ include zero. This result is interesting in that return predictability might be discovered from the nonzero autoregressive coefficients of lower and upper tails, while the zero-included coefficients around $\tau = 0.5$ do not have such ability. According to the efficient market hypothesis (EMH), at any given time, stock prices fully reflect all available information. Hence, in an efficient market, stock prices become not predictable but random. In practice, the information from lower and upper tails temporarily deviating from EMH, thus, can be useful for investors to discern investment patterns in a market. The similar empirical evidence has also been found by Bali et al. (2009).

[Table 5 about here]

[Figure 3 about here]

Looking at the plots of transition probabilities in Figure 3, their variation across quantiles in regime 1 is much smaller than in regime 2. The transition probabilities of regime 1 are ranging from 0.85 to 0.985, compared to the range of regime 2 from 0.381 to 0.945. It seems that the more extreme the quantile level is, the lower the persistence of regime 2 ($p_{22}(\tau)$) is. Despite the variation in the regime persistence, the unconditional probabilities are very similar across quantiles, i.e., $\pi_1(\tau)$ and $\pi_2(\tau)$ are around 0.84 and 0.16 for each quantile level, respectively. This result is reasonable in that the unconditional probability reflects possible changes in common and fundamental macroeconomic conditions. In contrast to the unconditional probability, regime persistence is possibly varying across quantiles. This observation is further consolidated by Figure 4 plotting the smoothed probability $\xi_{s_t=2,t|T}$ for $\tau = 0.05, 0.5$ (see *Appendix A* for the computation method.). The shaded areas are NBER-dated business cycles. In this figure, regime 2 for the 5% lower tail appears to be much less persistence, while both smoothed probabilities trace closely to the NBER-dated business cycles.

[Figure 4 about here]

Applying the method in Section 5, Table 6 reports the logarithm of the marginal likelihoods for testing the presence of regimes. Using the scale of Kass and Raftery (1995), very strong evidence of the MSQAR model over the QAR model has been found as the Bayesian factors $\pi(\boldsymbol{y}|MSQAR, \tau)/\pi(\boldsymbol{y}|QAR, \tau)$ for all τs considered exceed 150. This result supports the presence of regimes across quantiles for both monthly and weekly S&P 500 index returns.

[Table 6 about here]

Figure 5 plots conditional quantiles for $\tau = 0.05$ as $Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1}, s_t; \hat{\Theta}(\tau)\right) = \sum_{i\in S} Q_{y_t}\left(\tau|\boldsymbol{y}_{t-1}, s_t = i; \hat{\theta}_i(\tau)\right) Pr\left(s_t = i|\boldsymbol{y}_t; \hat{\Theta}(\tau)\right)$. The dark lines in Figure 5 are the conditional quantile dynamics ($Q_{y_t}(\tau|\boldsymbol{y}_{t-1}, s_t; \hat{\Theta}(\tau))$) and the top and bottom light lines are the conditional quantile dynamics of regime 1 ($Q_{y_t}(\tau|\boldsymbol{y}_{t-1}, s_t = 1; \hat{\theta}_1(\tau))$) and regime 2 ($Q_{y_t}(\tau|\boldsymbol{y}_{t-1}, s_t = 2; \hat{\theta}_2(\tau))$), respectively. The usefulness of the proposed MSQAR model can be immediately recognized from Figure 5. Specifically, it is recognized that existing methods without regime shifts cannot separate the market risk level associated with economic recessions from that associated with economic normal times. Thus, quantiles estimated from those approaches are at best the results of averaging on different economic states. However, Figure 5 demonstrates that the MSQAR model can identify market risks which are associated with different economic regimes. Quantile regimes identified by the MSQAR model are particularly beneficial for risk management, such as stress-testing financial institutions under Basel Accord regulation, since a risk manager or a central government concerns about extreme scenarios or worst possible outcomes.

[Figure 5 about here]

A commonly used criterion to compare between quantile models is the violation ratio $\hat{vr} = \hat{\tau}/\tau$ where $\hat{\tau} = \frac{1}{T}\sum_{t=1}^{T} I(y_t < Q_{y_t}(\tau|\boldsymbol{y}_{t-1}; \hat{\Theta}(\tau))$ is the number of quantile exceedances (violations) divided by the evaluation sample size. Ideally, the violation ratios should be close to one. Otherwise if $\hat{vr} < 1$, then estimates are too conservative (less violations than nominal), while a ratio $\hat{vr} > 1$ means too aggressive (more violations than nominal). Table 7 reporting violation ratios shows that the CAViaR model has the violation ratio closest to one, followed by the MSQAR model. The violation ratio results also show that the MSAR model gives aggressive estimations for lower tails but conservative for upper tails. I further assess the quantile estimation with the unconditional coverage (UC) test of Kupiec (1995), the conditional coverage (CC) test of Christoffersen (1998), the dynamic quantile (DQ) test of Engle and Manganelli (2004), and the Value-at-Risk model based on quantile regressions of Gaglianone et al. (2011).[8] These tests are quite standard in the quantile evaluation literature; see Kuester et al. (2006) for details. The test results reported in Table 8 show that all these tests considered do not reject the null hypotheses for the MSQAR and CAViaR models which have a satisfactory quantile estimation. In contrast, the MSAR models are rejected by UC, CC and VQR tests.

[Table 7 about here]

[Table 8 about here]

In addition, it is also important to evaluate the model by the monotonicity requirement on the conditional quantile functions. If the monotonicity is satisfied, there should be no crossings over quantiles. Severe crossing problems violate the theorem of equivariance to monotone transformation from (1.1) to (1.2). Figure 6 plots the estimated quantiles of each single regime. The straight lines are $Q_{y_t}(\tau|\boldsymbol{y}_{t-1}, s_t = 1; \hat{\theta}_1(\tau))$ and $Q_{y_t}(\tau|\boldsymbol{y}_{t-1}, s_t = 2; \hat{\theta}_2(\tau))$ for regime 1 and 2, respectively. The dots are the scatter plots with y_t as y-axis and y_{t-1} as x-axis. Despite that the MSQAR model is nonlinear, it takes a linear form within a single regime. Quantiles within a regime are not parallel due to quantile autoregressive coefficients of τ-dependent. Figure 6 shows that the estimated quantiles have no crossing issues for monthly S&P 500 returns, while the minor crossing problem occurs for weekly frequency data. Overall, Figure 6 does not raise any severe crossing issue for the data considered in this paper.

[Figure 6 about here]

8 Conclusion

This paper proposes a new location-scale quantile autoregression, so-called Markov-switching quantile autoregression, to characterize behaviors of different parts of a nonstationary time series distribution. The new method directly inferences and estimates dynamic quantiles by allowing the location

[8]The Theorem 4 of Engle and Manganelli (2004) is used for the DQ test using four lags

and scale parameters subject to regime shifts. Unobservable economic regimes are inferred by standard Hamilton filter approach in which quantile error terms follow a three parameter asymmetric Laplace distribution. Bayesian estimation is adopted to deal with some serious computational challenges in this nonlinear model which has non-differentiable likelihood functions. The Bayesian output also provides a convenient way to test the presence and the number of regimes by computing marginal likelihoods.

The simulation study in this paper has verified the adequacy of using the asymmetric Laplace distribution as the likelihood in regime inference and quantile prediction when the true model errors follow different distributions. In addition, the potentially substantial bias in quantile estimation has been found via the simulation study when the benchmark model (Markov-Switching autoregressive model) assumes that autoregressive coefficients are independent of τ. The empirical illustration of the proposed model in the U.S. financial market clearly supports the presence of regimes in quantile autoregression and achieves statistical significance in quantile estimation.

References

[1] Ausin, M.C. and H.F. Lopes (2010) Time-varying joint distribution through copulas. Computational Statistics and Data Analysis 54: 2383-2399

[2] Baele, L., Bekaert, G., Inghelbrecht, K., and Wei, M. (2014) Flights to Safety. Finance and Economics. Discussion Series 2014-46, Board of Governors of the Federal Reserve System.

[3] Bali, T., Demirtas, K. and Levy, H. (2009) Is There an Intertemporal Relation between Downside Risk and Expected Returns? Journal of Financial and Quantitative Analysis 44(4): 883-909

[4] Bauwens, L., A. Preminger and J.V.K. Rombouts (2010) Theory and inference for a Markov-Switching GARCH model. The Econometrics Journal 13: 218-244

[5] Cai, Y. and J. Stander (2008) Quantile Self-Exciting Threshold Autoregressive Time Series Models. Journal of Time Series Analysis 29(1): 186-202

[6] Cai, Y. (2010) Forecasting for quantile self-exciting threshold autoregressive time series models. Biometrika 97(1): 199-208

[7] Chen, Q., R. Gerlach, and Z. Lu (2012) Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution. Computational Statistics and Data Analysis 56(11): 3498-3516

[8] Chernozhukov, V. and H. Hong (2003) An MCMC approach to classical estimation. Journal of Econometrics 115: 293-346

[9] Cheung, Y. and U.G. Erlandsson (2005) Exchange rates and Markov-Switching dynamics. Journal of Business & Economic Statistics 23(3): 314-320

[10] Chib, S. and E. Greenberg (1995) Understanding the Metropolis-Hastings algorithm. American Statistician 49: 327-335

[11] Chib, S. and I. Jeliazkov (2001) Marginal Likelihood From the Metropolis-Hastings Output. Journal of the American Statistical Association 96(453): 270-281

[12] Christoffersen, P.F., F.X. Diebold, R.S. Mariano, A.S. Tay and Y.K. Tse (2007) Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. Journal of Financial Forecasting 1(2): 1-22

[13] Christoffersen, P. (1998) Evaluating interval forecasts. International Economic Review 39: 841-862

[14] Diebold, F.X., Lee, J.-H. and Weinbach, G. (1994), Regime Switching with Time-Varying Transition Probabilities, in C. Hargreaves (ed.), Nonstationary Time Series Analysis and Cointegration. (Advanced Texts in Econometrics, C.W.J. Granger and G. Mizon, eds.), 283-302. Oxford: Oxford University Press.

[15] Engel, C. (1994) Can the Markov-Switching model forecast exchange rates? Journal of International Economics 36: 151-165

[16] Engel, C. and Kim, C.J. (1999), The Long-Run U.S./U.K. Real Exchange Rate. Journal of Money, Credit, and Banking, 31, 335-356

[17] Engle, R. and S. Manganelli (2004) CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, Journal of Business & Economic Statistics, 22: 367-381

[18] Filardo, A. J. (1994) Business-Cycle Phases and Their Transitional Dynamics. Journal of Business & Economic Statistics 12: 299–308

[19] Gaglianone, W., L. Lima, O. Linton, and D. Smith (2011), Evaluating Value-at-Risk Models via Quantile Regression, Journal of Business & Economic Statistics, 29: 150-160

[20] Garcia, R. and P. Perron (1996) An Analysis of the Real Interest Rate under Regime Shifts. The Review of Economics & Statistics 78: 111-125

[21] Gerlach, R., C.W.S. Chen and N.Y.C. Chan (2011) Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets. Journal of Business & Economic Statistics 29(4): 481-492

[22] Geweke, J. and H. Tanizaki (2001) Bayesian estimation of state-space models using the Metropolis-Hastings algorithm within Gibbs sampling. Computational Statistics & Data Analysis 37: 151-170

[23] Geweke, J. (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo, J., Berger, J., David, A., Smith, A. (Eds.), Bayesian Statistics. Vol. 4. Oxford University Press, Oxford, pp. 169-193.

[24] Gray, S.F. (1996) Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42: 27-62

[25] Guerin, P. and M. Marcellino (2013) Markov-Switching MIDAS Models. Journal of Business & Economic Statistics 31(1): 45-56

[26] Guidolin, M. (2012) Markov Switching Models in Empirical Finance. Advances in Econometrics, ISBN: 978-1-78052-526-6

[27] Hamilton, J.D. (1994) Time Series Analysis. Princeton University Press

[28] Hamilton, J., D. Waggoner and T. Zha (2007) Normalization in econometrics. Econometric Reviews 26: 221-252

[29] Hamilton, J.D. and R. Susmel (1994) Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64: 307-333

[30] Kass R and Raftery A. (1995) Bayes factors. Journal of the American Statistical Association 90: 773-795

[31] Kim, C.J. (1994) Dynamic linear models with Markov-switching. Journal of Econometrics 60: 1-22

[32] Kim, C.J. and C.R. Nelson (1999) State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. The MIT Press

[33] 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

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

[35] Koenker, R. and G. Bassett (1978) Regression quantile. Econometrica 46: 33-50

[36] Koenker, R. and Z. Xiao (2006) Quantile Autoregression. Journal of the American Statistical Association 101(475): 980-990

[37] Kuester, K., Mittnik, S. and Paolella, M. (2006) Value-at-risk prediction: a comparison of alternative strategies. Journal of Financial Econometrics 4: 53-89

[38] Kupiec, P. (1995) Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives 3: 73-84

[39] Liu, X. and R. Luger (2015) Markov-Switching Quantile Autoregression: A Gibbs sampling approach. Working Paper

[40] Nalewaik, J. J. (2011) Forecasting recessions using stall speeds. Board of Governors of the Federal Reserve System - Finance and Economics Discussion Series.

[41] Ripley, B. (1987) Stochastic Simulation. John Wiley, New York

[42] Sims, C.A. and T. Zha (2006) Were There Regime Switching in US Monetary Policy. American Economic Review 96(1): 54-81

[43] Taylor, S. (1986) Modelling Financial Time Series. New York: Wiley

[44] Tierney, L. (1994) Markov Chains for Exploring Posterior Distributions. Ann. Statist. 22: 1701-1728

[45] Vrontos, I., P. Dellaportas, D. Politis (2002) Full Bayesian inference for GARCH and EGARCH models. Journal of Business and Economic Statistics 18: 187-198

[46] Xiao, Z. and R. Koenker (2009) Conditional Quantile Estimation for Generalized Autoregressive Conditional Heteroscedasticity Models. Journal of the American Statistical Association 104(488): 1696-1712

[47] Yu, K. and J. Zhang (2005) A Three-Parameter Asymmetric Laplace Distribution and Its Extension. Communications in Statistics- Theory and Methods 34: 1867-1879

[48] Yu, K. and R.A. Moyeed (2001) Bayesian quantile regression. Statistics & Probability Letters 54: 437-447

Table 1: The Simulation Design 1: Estimation Results

	(1) Simulation with Normal Distribution								
	True	$N=200$				$N=500$			
		PM	Std	RMSE	MAD	PM	Std	RMSE	MAD
$\tau=0.05$									
p_{11}	0.900	0.883	0.034	0.042	0.031	0.895	0.019	0.022	0.017
p_{22}	0.900	0.890	0.034	0.039	0.030	0.901	0.022	0.025	0.019
$\theta_{10}(\tau)$	1.178	1.262	0.113	0.120	0.099	1.249	0.081	0.092	0.075
$\theta_{11}(\tau)$	0.200	0.181	0.038	0.210	0.168	0.184	0.032	0.178	0.145
$\theta_{20}(\tau)$	-3.645	-3.584	0.295	0.082	0.065	-3.606	0.193	0.054	0.043
$\theta_{21}(\tau)$	0.400	0.401	0.097	0.172	0.200	0.400	0.059	0.147	0.116
$\tau=0.25$									
p_{11}	0.900	0.887	0.034	0.040	0.029	0.898	0.019	0.021	0.016
p_{22}	0.900	0.889	0.032	0.038	0.029	0.899	0.021	0.024	0.018
$\theta_{10}(\tau)$	1.663	1.671	0.088	0.053	0.041	1.665	0.060	0.036	0.029
$\theta_{11}(\tau)$	0.200	0.194	0.030	0.151	0.122	0.195	0.024	0.120	0.099
$\theta_{20}(\tau)$	-2.674	-2.677	0.221	0.083	0.067	-2.660	0.123	0.046	0.037
$\theta_{21}(\tau)$	0.400	0.394	0.067	0.168	0.133	0.402	0.039	0.098	0.081
$\tau=0.5$									
p_{11}	0.900	0.889	0.034	0.040	0.029	0.900	0.019	0.021	0.015
p_{22}	0.900	0.888	0.032	0.038	0.029	0.897	0.021	0.023	0.018
$\theta_{10}(\tau)$	2.000	1.997	0.085	0.043	0.035	1.988	0.055	0.028	0.022
$\theta_{11}(\tau)$	0.200	0.196	0.028	0.142	0.112	0.198	0.021	0.106	0.085
$\theta_{20}(\tau)$	-2.000	-2.057	0.209	0.108	0.089	-2.041	0.119	0.063	0.052
$\theta_{21}(\tau)$	0.400	0.382	0.064	0.165	0.139	0.388	0.037	0.097	0.080
$\tau=0.75$									
p_{11}	0.900	0.891	0.035	0.040	0.030	0.903	0.019	0.022	0.018
p_{22}	0.900	0.884	0.032	0.040	0.031	0.894	0.021	0.024	0.019
$\theta_{10}(\tau)$	2.337	2.323	0.087	0.038	0.031	2.311	0.057	0.027	0.021
$\theta_{11}(\tau)$	0.200	0.203	0.029	0.144	0.116	0.206	0.021	0.107	0.054
$\theta_{20}(\tau)$	-1.326	-1.499	0.231	0.118	0.172	-1.469	0.148	0.105	0.125
$\theta_{21}(\tau)$	0.400	0.358	0.075	0.177	0.173	0.365	0.047	0.140	0.115
$\tau=0.95$									
p_{11}	0.900	0.884	0.040	0.048	0.035	0.896	0.024	0.027	0.021
p_{22}	0.900	0.863	0.036	0.057	0.046	0.874	0.025	0.040	0.033
$\theta_{10}(\tau)$	2.822	2.779	0.123	0.046	0.037	2.772	0.080	0.034	0.027
$\theta_{11}(\tau)$	0.200	0.217	0.042	0.225	0.188	0.215	0.033	0.181	0.148
$\theta_{20}(\tau)$	-0.355	-0.453	0.124	0.144	0.380	-0.451	0.103	0.123	0.322
$\theta_{21}(\tau)$	0.400	0.337	0.086	0.187	0.215	0.328	0.067	0.145	0.204

PM, Std, RMSE and MAD are posterior mean, standard deviation, the root of mean squared errors and the mean absolute deviation errors, respectively.

Table 1 continued	(2) Simulation with t_3 Distribution								
	True	$N = 200$				$N = 500$			
		PM	Std	RMSE	MAD	PM	Std	RMSE	MAD
$\tau = 0.05$									
p_{11}	0.900	0.879	0.031	0.042	0.033	0.885	0.021	0.029	0.022
p_{22}	0.900	0.880	0.041	0.050	0.037	0.892	0.024	0.028	0.022
$\theta_{10}(\tau)$	1.321	1.428	0.101	0.111	0.094	1.441	0.077	0.108	0.094
$\theta_{11}(\tau)$	0.200	0.191	0.035	0.180	0.148	0.193	0.028	0.142	0.117
$\theta_{20}(\tau)$	-3.359	-3.319	0.398	0.119	0.093	-3.319	0.246	0.074	0.059
$\theta_{21}(\tau)$	0.400	0.416	0.118	0.197	0.240	0.407	0.080	0.121	0.161
$\tau = 0.25$									
p_{11}	0.900	0.885	0.031	0.038	0.029	0.891	0.021	0.026	0.020
p_{22}	0.900	0.879	0.035	0.045	0.033	0.889	0.021	0.027	0.021
$\theta_{10}(\tau)$	1.779	1.770	0.062	0.035	0.028	1.777	0.043	0.024	0.018
$\theta_{11}(\tau)$	0.200	0.202	0.024	0.118	0.095	0.202	0.016	0.078	0.060
$\theta_{20}(\tau)$	-2.442	-2.465	0.156	0.064	0.049	-2.443	0.091	0.037	0.031
$\theta_{21}(\tau)$	0.400	0.393	0.048	0.120	0.093	0.400	0.029	0.073	0.058
$\tau = 0.5$									
p_{11}	0.900	0.886	0.031	0.037	0.028	0.892	0.021	0.025	0.020
p_{22}	0.900	0.876	0.034	0.046	0.034	0.886	0.021	0.028	0.022
$\theta_{10}(\tau)$	2.000	1.994	0.058	0.029	0.022	1.993	0.034	0.017	0.014
$\theta_{11}(\tau)$	0.200	0.199	0.023	0.115	0.091	0.201	0.013	0.067	0.051
$\theta_{20}(\tau)$	-2.000	-2.025	0.117	0.060	0.048	-2.025	0.074	0.039	0.028
$\theta_{21}(\tau)$	0.400	0.394	0.037	0.095	0.077	0.396	0.024	0.062	0.049
$\tau = 0.75$									
p_{11}	0.900	0.886	0.032	0.039	0.029	0.892	0.022	0.026	0.020
p_{22}	0.900	0.870	0.034	0.050	0.039	0.881	0.022	0.033	0.026
$\theta_{10}(\tau)$	2.221	2.218	0.075	0.034	00.025	2.212	0.043	0.020	0.016
$\theta_{11}(\tau)$	0.200	0.199	0.027	0.133	0.101	0.202	0.016	0.080	0.063
$\theta_{20}(\tau)$	-1.558	-1.643	0.137	0.103	0.083	-1.647	0.085	0.079	0.065
$\theta_{21}(\tau)$	0.400	0.383	0.042	0.113	0.092	0.385	0.027	0.076	0.062
$\tau = 0.95$									
p_{11}	0.900	0.881	0.039	0.048	0.036	0.888	0.026	0.032	0.024
p_{22}	0.900	0.853	0.037	0.067	0.055	0.861	0.025	0.051	0.044
$\theta_{10}(\tau)$	2.679	2.690	0.186	0.069	0.053	2.675	0.128	0.048	0.038
$\theta_{11}(\tau)$	0.200	0.197	0.066	0.153	0.265	0.202	0.047	0.137	0.186
$\theta_{20}(\tau)$	-0.641	-0.596	0.110	0.185	0.141	-0.596	0.073	0.134	0.106
$\theta_{21}(\tau)$	0.400	0.354	0.066	0.201	0.167	0.358	0.048	0.158	0.130

PM, Std, RMSE and MAD are posterior mean, standard deviation, the root of mean squared errors and the mean absolute deviation errors, respectively.

Table 2: The Simulation Design 1: Summary Statistics for the Quantile Predictability

	Simulated from Normal Distribution			Simulated from t_3 Distribution		
	MSQAR	MSAR_N	MSAR_t	MSQAR	MSAR_N	MSAR_t
$N = 200$						
$\tau = 0.05$	0.197	0.125	0.249	0.174	0.243	0.144
	[0.124,0.323]	[0.050,0.223]	[0.061,0.345]	[0.070,0.538]	[0.076,0.629]	[0.059,0.623]
$\tau = 0.25$	0.099	0.100	0.122	0.112	0.188	0.108
	[0.068,0.176]	[0.044,0.172]	[0.048,0.182]	[0.033,0.179]	[0.063,0.410]	[0.040,0.240]
$\tau = 0.5$	0.093	0.093	0.091	0.098	0.118	0.099
	[0.064,0.203]	[0.044,0.150]	[0.042,0.157]	[0.044,0.159]	[0.048,0.253]	[0.039,0.150]
$\tau = 0.75$	0.119	0.097	0.113	0.131	0.189	0.105
	[0.096,0.161]	[0.046,0.147]	[0.043,0.166]	[0.074,0.228]	[0.073,0.491]	[0.044,0.219]
$\tau = 0.95$	0.168	0.118	0.232	0.175	0.235	0.135
	[0.060,0.387]	[0.053,0.215]	[0.053,0.345]	[0.059,0.456]	[0.081,0.662]	[0.055,0.612]
$N = 500$						
$\tau = 0.05$	0.134	0.084	0.156	0.155	0.200	0.103
	[0.068,0.275]	[0.041,0.148]	[0.048,0.271]	[0.053,0.323]	[0.061,0.483]	[0.051,0.317]
$\tau = 0.25$	0.090	0.068	0.084	0.087	0.176	0.076
	[0.056,0.114]	[0.034,0.107]	[0.041,0.1137]	[0.049,0.137]	[0.071,0.310]	[0.041,0.126]
$\tau = 0.5$	0.085	0.063	0.070	0.077	0.091	0.070
	[0.033,0.117]	[0.033,0.102]	[0.035,0.144]	[0.041,0.115]	[0.047,0.169]	[0.038,0.108]
$\tau = 0.75$	0.117	0.067	0.077	0.133	0.176	0.078
	[0.099,0.208]	[0.035,0.107]	[0.041,0.218]	[0.092,0.188]	[0.088,0.368]	[0.039,0.134]
$\tau = 0.95$	0.132	0.082	0.137	0.171	0.194	0.094
	[0.063,0.341]	[0.041,0.140]	[0.040,0.358]	[0.061,0.402]	[0.059,0.592]	[0.067,0.297]

Entry values in this table are mean absolute deviation errors (MADE) across 200 replications. Values in square brackets are the 5% and 95% quantiles of MADEs across 200 replications. MSAR_N and MSAR_t represent the MSAR model estimated by normal and student-t distributions, respectively.

Table 3: The Simulation Design 2: Estimation Results

	True	(1) Simulation with Normal Distribution							
		$N = 200$				$N = 500$			
		PM	Std	RMSE	MAD	PM	Std	RMSE	MAD
$\tau = 0.05$									
p_{11}	0.900	0.908	0.022	0.022	0.135	0.904	0.017	0.019	0.125
p_{22}	0.900	0.915	0.022	0.024	0.142	0.910	0.018	0.023	0.138
$\theta_{10}(\tau)$	1.178	1.164	0.048	0.050	0.204	1.168	0.043	0.044	0.192
$\theta_{11}(\tau)$	0.100	0.103	0.002	0.004	0.059	0.102	0.002	0.002	0.044
$\theta_{20}(\tau)$	-3.645	-3.975	0.059	0.332	0.575	-3.939	0.028	0.300	0.543
$\theta_{21}(\tau)$	0.300	0.326	0.005	0.027	0.162	0.322	0.003	0.022	0.148
$\tau = 0.5$									
p_{11}	0.900	0.898	0.025	0.025	0.144	0.901	0.018	0.018	0.117
p_{22}	0.900	0.892	0.022	0.023	0.140	0.895	0.018	0.019	0.123
$\theta_{10}(\tau)$	2.000	1.995	0.046	0.046	0.189	2.001	0.038	0.038	0.176
$\theta_{11}(\tau)$	0.400	0.401	0.011	0.011	0.096	0.400	0.011	0.011	0.096
$\theta_{20}(\tau)$	-2.000	-2.015	0.064	0.066	0.237	-2.006	0.047	0.048	0.198
$\theta_{21}(\tau)$	0.600	0.598	0.017	0.017	0.122	0.599	0.015	0.015	0.112
$\tau = 0.95$									
p_{11}	0.900	0.897	0.025	0.025	0.145	0.899	0.021	0.021	0.127
p_{22}	0.900	0.884	0.021	0.026	0.149	0.887	0.017	0.022	0.131
$\theta_{10}(\tau)$	2.822	3.055	0.040	0.260	0.483	3.082	0.016	0.236	0.510
$\theta_{11}(\tau)$	0.700	0.644	0.009	0.061	0.237	0.639	0.001	0.057	0.247
$\theta_{20}(\tau)$	-0.355	-0.348	0.006	0.017	0.127	-0.339	0.006	0.010	0.088
$\theta_{21}(\tau)$	0.900	0.826	0.005	0.079	0.281	0.821	0.002	0.074	0.272

PM, Std, RMSE and MAD are posterior mean, standard deviation, the root of mean squared errors and the mean absolute deviation errors, respectively.

Table 4 continued	(2) Simulation with t_3 Distribution								
	True	$N = 200$				$N = 500$			
		PM	Std	RMSE	MAD	PM	Std	RMSE	MAD
$\tau = 0.05$									
p_{11}	0.900	0.896	0.019	0.019	0.126	0.899	0.015	0.015	0.108
p_{22}	0.900	0.920	0.026	0.028	0.154	0.911	0.018	0.027	0.151
$\theta_{10}(\tau)$	1.321	1.358	0.050	0.062	0.232	1.346	0.044	0.051	0.207
$\theta_{11}(\tau)$	0.100	0.102	0.002	0.003	0.049	0.101	0.001	0.002	0.038
$\theta_{20}(\tau)$	-3.359	-3.688	0.013	0.329	0.573	-3.675	0.003	0.317	0.562
$\theta_{21}(\tau)$	0.300	0.329	0.002	0.029	0.170	0.327	0.001	0.027	0.164
$\tau = 0.5$									
p_{11}	0.900	0.896	0.023	0.023	0.139	0.896	0.017	0.017	0.117
p_{22}	0.900	0.892	0.024	0.025	0.147	0.896	0.018	0.018	0.121
$\theta_{10}(\tau)$	2.000	1.999	0.037	0.037	0.173	2.000	0.028	0.028	0.149
$\theta_{11}(\tau)$	0.400	0.400	0.008	0.007	0.079	0.400	0.008	0.007	0.076
$\theta_{20}(\tau)$	-2.000	-2.011	0.058	0.059	0.217	-2.005	0.051	0.051	0.205
$\theta_{21}(\tau)$	0.600	0.597	0.014	0.014	0.106	0.601	0.013	0.013	0.106
$\tau = 0.95$									
p_{11}	0.900	0.896	0.030	0.030	0.159	0.888	0.029	0.031	0.159
p_{22}	0.900	0.857	0.013	0.038	0.207	0.865	0.011	0.044	0.188
$\theta_{10}(\tau)$	2.679	2.927	0.054	0.249	0.498	2.895	0.022	0.222	0.468
$\theta_{11}(\tau)$	0.700	0.645	0.013	0.061	0.246	0.639	0.004	0.056	0.236
$\theta_{20}(\tau)$	-0.641	-0.607	0.012	0.048	0.219	-0.593	0.008	0.036	0.185
$\theta_{21}(\tau)$	0.900	0.822	0.010	0.079	0.280	0.824	0.012	0.076	0.275

PM, Std, RMSE and MAD are posterior mean, standard deviation, the root of mean squared errors and the mean absolute deviation errors, respectively.

Table 4: The Simulation Design 2: Summary Statistics for the Quantile Predictability

	Simulated from Normal			Simulated from t_3		
	MSQAR	MSAR_N	MSAR_t	MSQAR	MSAR_N	MSAR_t
$N = 200$						
$\tau = 0.05$	1.181	1.465	1.500	1.305	1.480	1.480
	[1.028,1.317]	[1.288,1.624]	[1.320,1.668]	[1.124,1.503]	[1.249,1.690]	[1.277,1.655]
$\tau = 0.5$	0.531	1.232	1.247	0.536	1.257	1.237
	[0.435,0.637]	[1.072,1.378]	[1.102,1.416]	[0.435,0.637]	[1.054,1.406]	[1.077,1.417]
$\tau = 0.95$	0.994	1.198	1.223	1.012	1.281	1.274
	[0.882,1.112]	[1.060,1.315]	[1.085,1.342]	[0.896,1.112]	[1.068,1.487]	[1.114,1.478]
$N = 500$						
$\tau = 0.05$	1.166	1.440	1.478	1.297	1.448	1.445
	[1.089,1.249]	[1.312,1.545]	[1.342,1.599]	[1.134,1.462]	[1.314,1.646]	[1.330,1.559]
$\tau = 0.5$	0.531	1.218	1.248	0.535	1.248	1.224
	[0.435, 0.629]	[1.122,1.313]	[1.153,1.334]	[0.480,0.595]	[1.129,1.348]	[1.151,1.322]
$\tau = 0.95$	0.984	1.178	1.202	1.007	1.259	1.243
	[0.896,1.087]	[1.081,1.269]	[1.077,1.417]	[0.942,1.066]	[1.115,1.536]	[1.140,1.358]

Entry values in this table is mean absolute deviation error (MADE) across 200 replications. Values in parentheses are the 5% and 95% quantiles of MADEs across 200 replications. MSAR_N and MSAR_t represent the MSAR model estimated by normal and student-t distributions, respectively.

Table 5: The MSQAR Model Estimation Results

	Monthly S&P 500					Weekly S&P 500				
	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
p_{11}	0.902	0.925	0.985	0.987	0.842	0.850	0.868	0.985	0.947	0.787
	(0.018)	(0.055)	(0.010)	(0.009)	(0.016)	(0.017)	(0.034)	(0.011)	(0.102)	(0.019)
	[1.440]	[-0.447]	[-0.405]	[0.062]	[-0.274]	[0.631]	[0.065]	[1.273]	[0.685]	[-1.832]
p_{22}	0.572	0.625	0.914	0.917	0.381	0.389	0.641	0.945	0.880	0.415
	(0.045)	(0.063)	(0.043)	(0.039)	(0.066)	(0.046)	(0.037)	(0.027)	(0.146)	(0.044)
	[-0.685]	[-0.570]	[-2.405]	[-0.483]	[-0.244]	[-0.276]	[0.446]	[-0.803]	[0.568]	[-0.509]
$\theta_{10}(\tau)$	-4.121	-0.114	1.045	3.522	4.971	-1.619	-0.039	0.346	1.066	1.647
	(0.089)	(0.051)	(0.064)	(0.062)	(0.063)	(0.039)	(0.041)	(0.040)	(0.132)	(0.035)
	[-2.444]	[-0.464]	[0.648]	[-0.108]	[0.836]	[-0.322]	[-0.747]	[-1.678]	[0.723]	[-1.007]
$\theta_{11}(\tau)$	0.084	0.035	0.026	-0.087	-0.082	0.085	0.015	0.035	-0.025	-0.032
	(0.042)	(0.026)	(0.027)	(0.031)	(0.042)	(0.036)	(0.026)	(0.022)	(0.081)	(0.020)
	[1.835]	[-0.307]	[-0.518]	[0.098]	[0.064]	[-0.328]	[1.246]	[-0.788]	[-0.616]	[-0.126]
$\theta_{20}(\tau)$	-15.77	-3.978	-2.124	5.784	12.05	-5.865	-1.952	-0.094	2.422	4.851
	(0.147)	(0.047)	(0.318)	(0.174)	(0.200)	(0.094)	(0.040)	(0.147)	(0.147)	(0.080)
	[-0.478]	[0.101]	[-1.615]	[-0.212]	[0.169]	[-0.296]	[1.404]	[0.965]	[-0.835]	[-0.665]
$\theta_{21}(\tau)$	0.378	0.115	0.085	0.042	0.032	0.207	0.010	0.016	-0.196	-0.251
	(0.039)	(0.045)	(0.048)	(0.054)	(0.046)	(0.038)	(0.017)	(0.056)	(0.069)	(0.018)
	[-0.224]	[-0.342]	[0.889]	[0.395]	[0.621]	[0.317]	[0.348]	[0.124]	[0.617]	[0.538]

Values in parentheses are numerical standard errors and the Geweke (1992) test statistic in square brackets. The test distribution for Geweke (1992) statistic is standard normal distribution.

Table 6: The Marginal Likelihoods and Bayesian Factors (BF)

	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
Monthly S&P 500					
QAR	-3829.97	-3324.14	-3165.05	-3200.57	-3600.99
MSQAR	-3418.9	-3205.14	-3117.42	-3142.66	-3328.27
BF: MSQAR/QAR	>150	>150	>150	>150	>150
Weekly S&P 500					
QAR	-8773.94	-7359.17	-6953.43	-7095.96	-8246.85
MSQAR	-7983.84	-7165.84	-6842.68	-6923.00	-7371.30
BF: MSQAR/QAR	>150	>150	>150	>150	>150

Table 7: Violation Ratios

	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
Monthly S&P 500					
MSQAR	**0.994**	0.990	0.998	1.007	1.009
MSAR_N	2.145	1.421	0.968	0.828	0.833
MSAR_t	2.126	1.452	0.965	0.828	0.836
CAViaR	0.993	**0.997**	**0.999**	**1.001**	**1.000**
Weekly S&P 500					
MSQAR	1.014	1.029	1.000	1.009	1.040
MSAR_N	2.186	1.211	0.964	0.921	0.947
MSAR_t	2.186	1.240	0.963	0.909	0.946
CAViaR	**0.990**	**0.996**	**1.001**	**1.001**	**1.000**

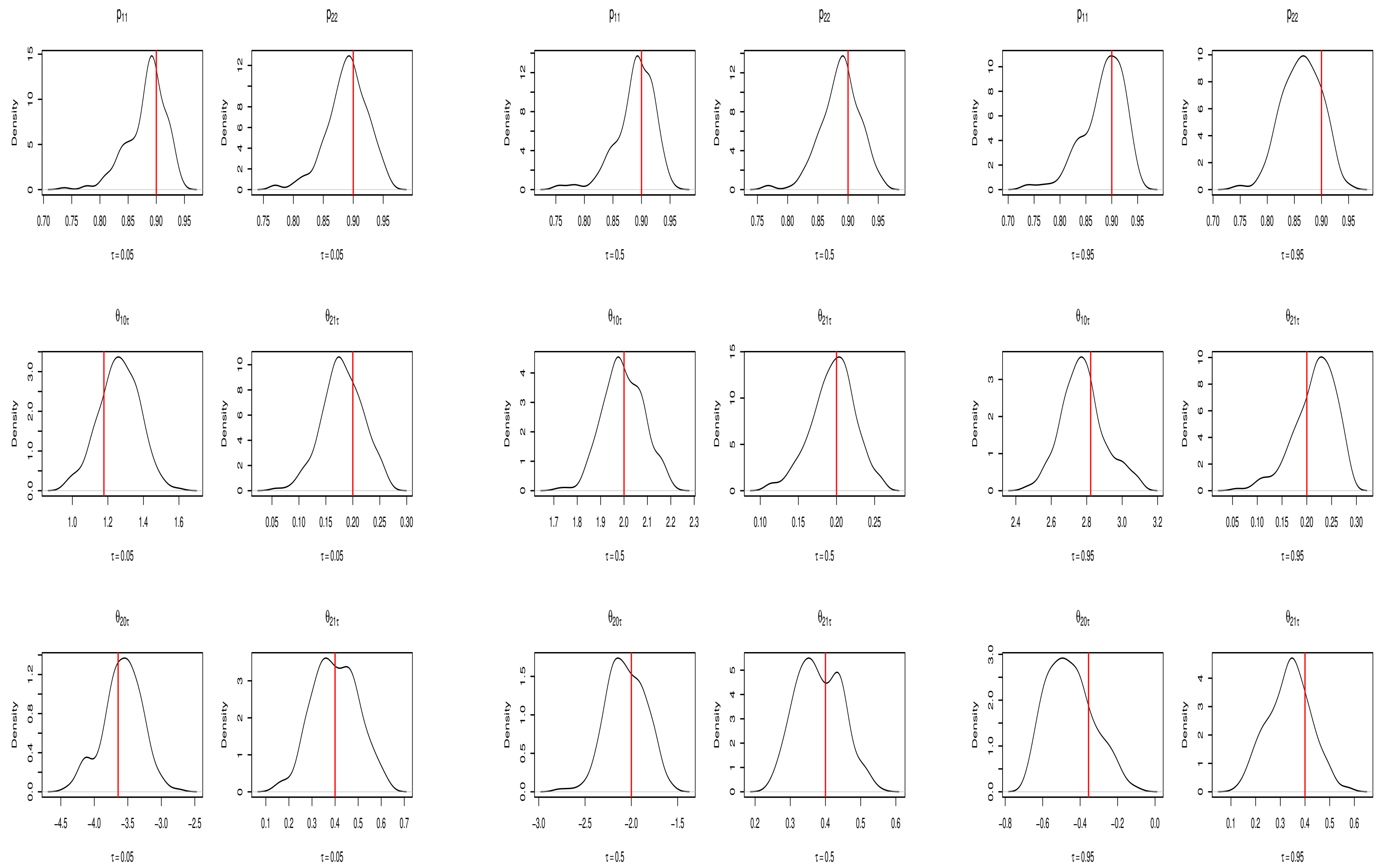
The violation ratio is defined as $\hat{\tau}/\tau$ where $\hat{\tau} = \frac{1}{T}\sum_{t=1}^{T} I(y_t < Q_{y_t}(\tau|\boldsymbol{y}_{t-1}; \hat{\Theta}(\tau))$ is the number of quantile exceedances (violations) divided by the evaluation sample size.

Table 8: Quantile Test Statistics

	Monthly S&P 500					Weekly S&P 500				
	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$	$\tau = 0.05$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.95$
UC										
MSQAR	0.966	0.858	0.951	0.694	0.174	0.978	0.666	0.986	0.369	0.785
MSAR_N	**<0.001**	**<0.001**	0.322	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.041**	**<0.001**	**<0.001**
MSAR_t	**<0.001**	**<0.001**	0.266	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.035**	**<0.001**	**<0.001**
CAViaR	0.960	0.957	0.975	0.957	0.960	0.892	0.888	0.972	0.888	0.955
CC										
MSQAR	0.051	0.205	0.867	0.840	0.271	0.340	0.323	0.592	0.281	0.214
MSAR_N	**<0.001**	**<0.001**	0.509	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.217	**<0.001**	**<0.001**
MSAR_t	**<0.001**	**<0.001**	0.471	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.201	**<0.001**	**<0.001**
CAViaR	0.152	0.989	0.953	0.225	0.503	0.780	0.801	0.864	0.710	0.240
DQ										
MSQAR	0.929	0.922	0.987	0.989	0.977	0.920	0.962	0.978	0.982	0.906
MSAR_N	0.770	0.955	0.987	0.973	0.817	0.898	0.943	0.995	0.993	0.977
MSAR_t	0.754	0.945	0.990	0.973	0.822	0.891	0.942	0.995	0.988	0.971
CAViaR	0.945	0.996	0.992	0.934	0.974	0.990	0.948	0.992	0.944	0.979
VQR										
MSQAR	0.783	0.240	0.933	0.697	0.819	0.449	0.558	0.717	0.454	0.265
MSAR_N	**0.047**	0.312	0.939	0.146	**<0.001**	**<0.001**	**<0.001**	0.167	**<0.001**	**<0.001**
MSAR_t	**0.034**	0.186	0.940	0.144	0.105	**<0.001**	**<0.001**	0.157	**<0.001**	**<0.001**
CAViaR	1.000	0.995	1.000	0.999	0.997	0.997	0.968	1.000	0.111	0.948

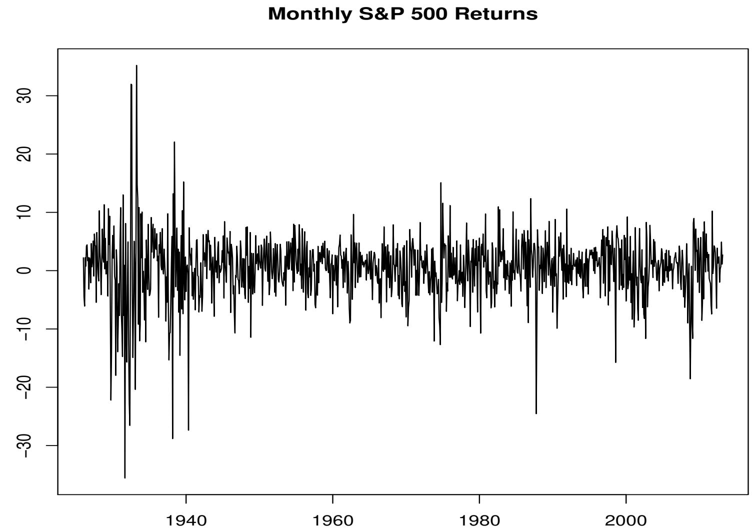
UC, CC, DQ and VQR represent the unconditional coverage test of Kupiec (1995), the conditional coverage test of Christoffersen (1998), the dynamic quantile (DQ) test of Engle and Manganelli (2004) using four lags, and the Value-at-Risk model based on quantile regressions of Gaglianone et al. (2011), respectively. Entries in this table are p-values.

Figure 1: Posteriors of the Parameter Estimates with the True Parameters Indicated By the Vertical Lines for $\tau = 0.05, 0.5, 0.95$.



(1) $\tau = 0.05$ (2) $\tau = 0.5$ (3) $\tau = 0.95$

Figure 2: Time Series Data Plots



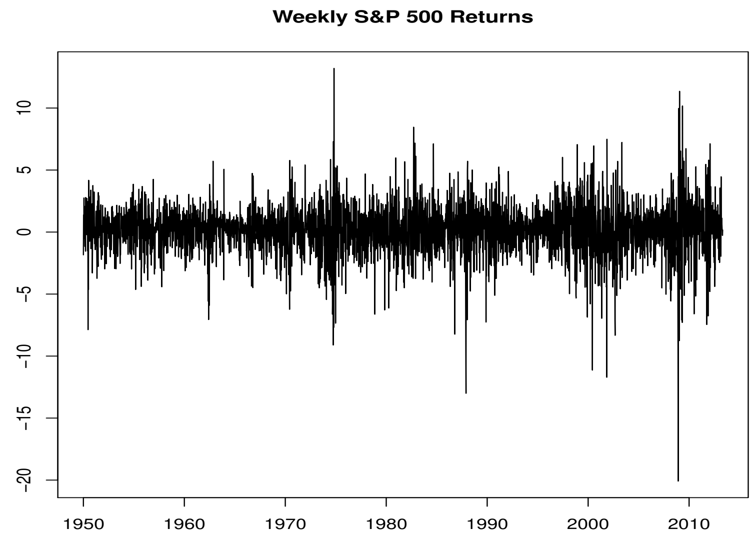
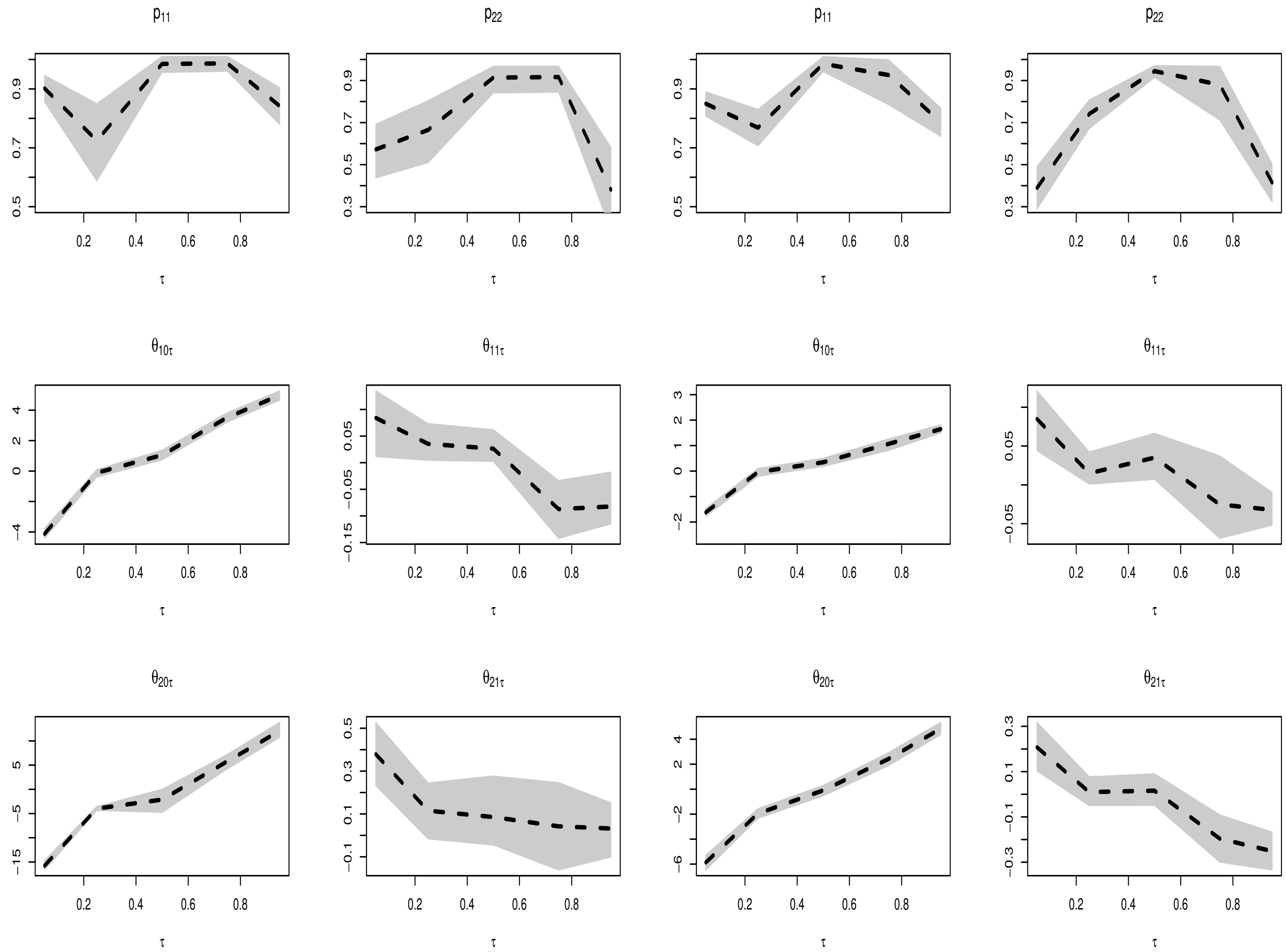


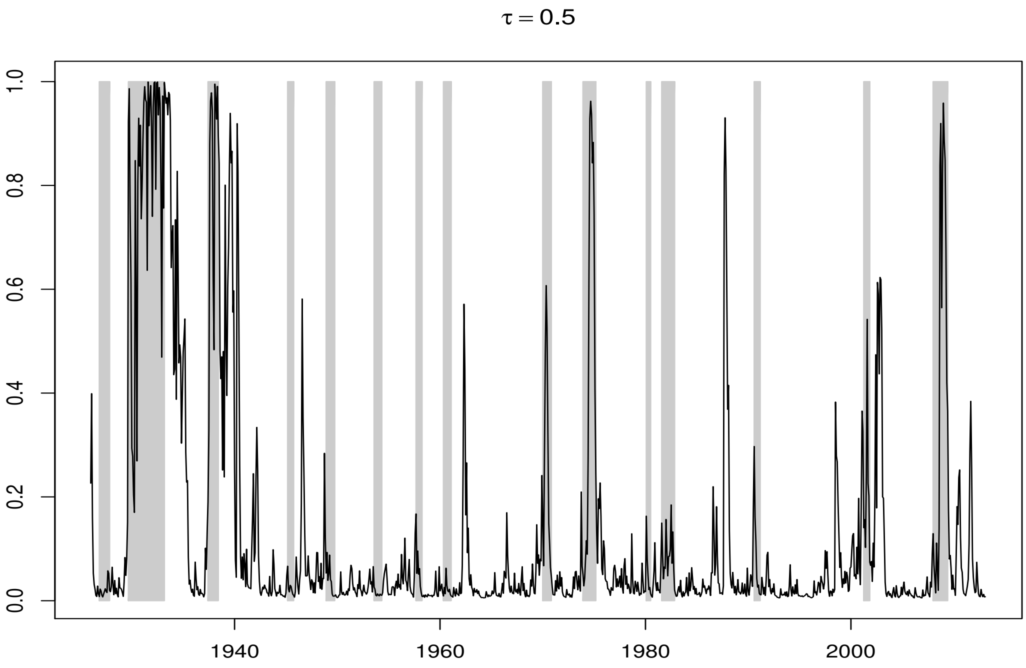
Figure 3: Plots for Quantile Parameter Estimates Across τs



(1) Monthly S&P 500 Returns

(2) Weekly S&P 500 Returns

Figure 4: Plots of Smoothed Probabilities with NBER-dated Business Cycles (shaded areas)



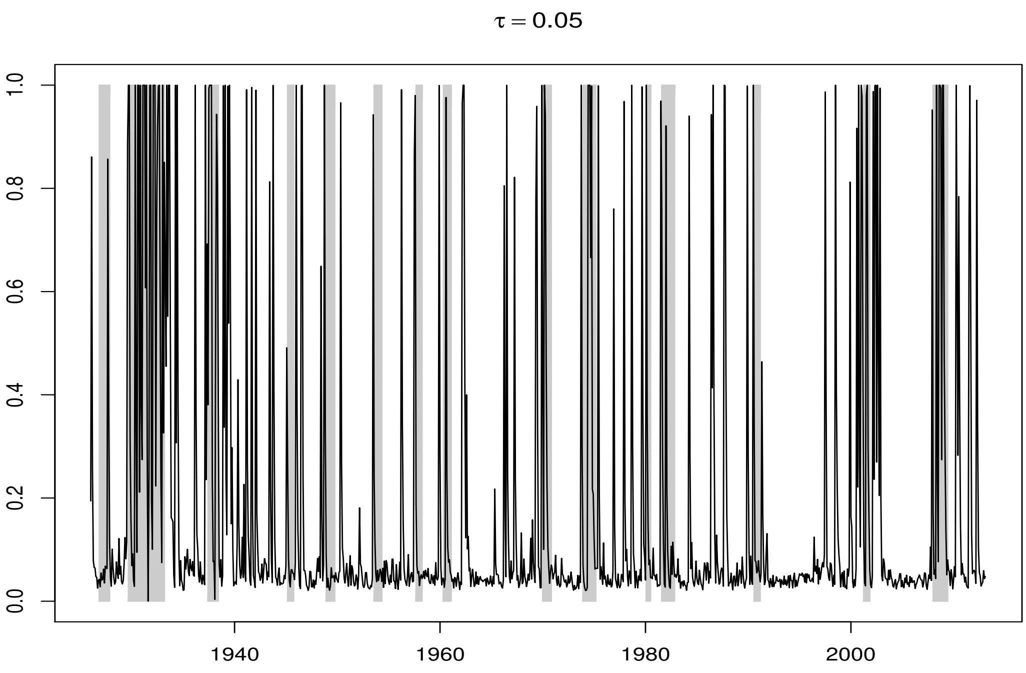


Figure 5: The Estimated Quantiles for $Q_{y_t}(\tau = 0.05|s_t = 1)$ (top light lines), $Q_{y_t}(\tau = 0.05|s_t)$ (dark lines), $Q_{y_t}(\tau = 0.05|s_t = 2)$(bottom light lines)

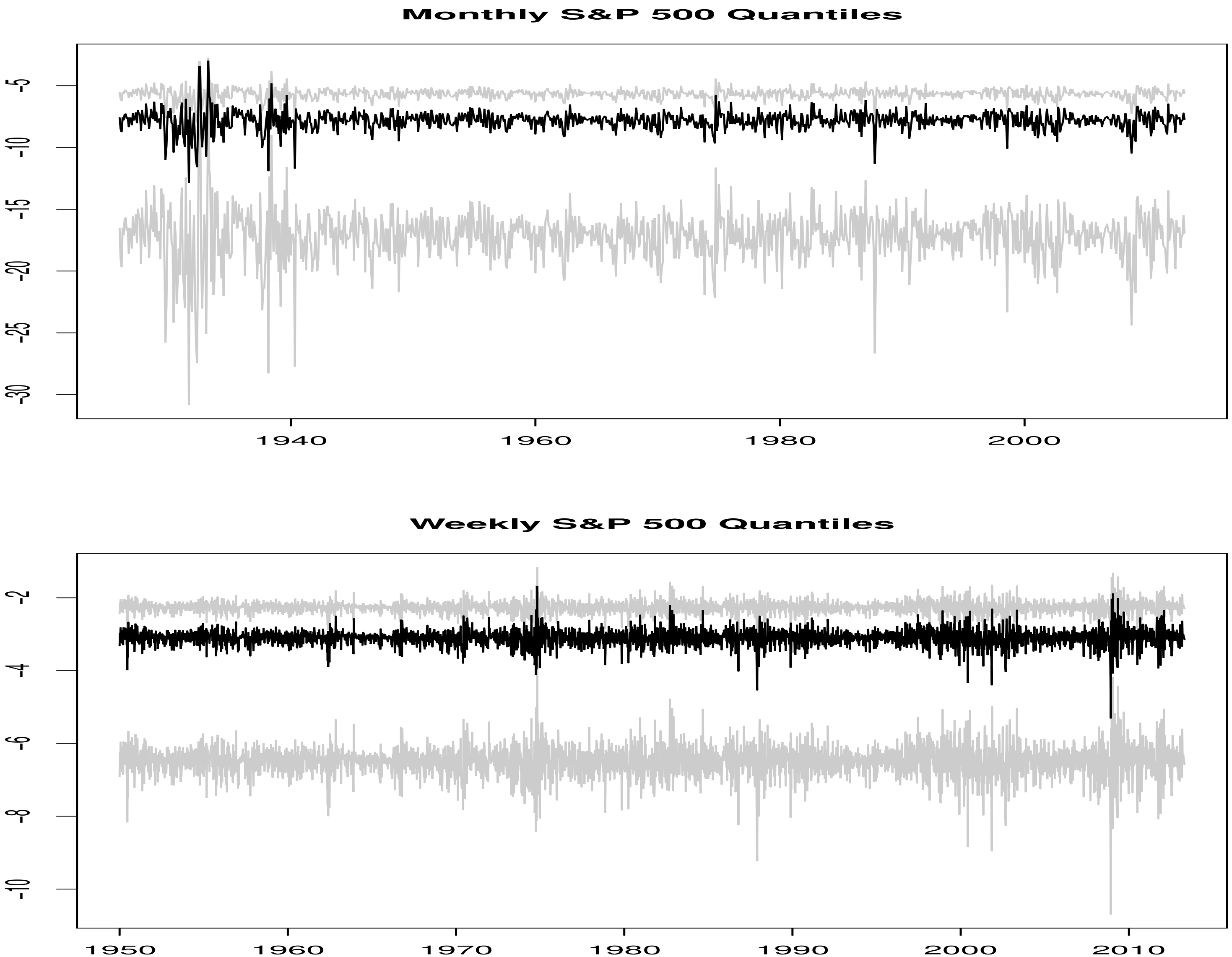
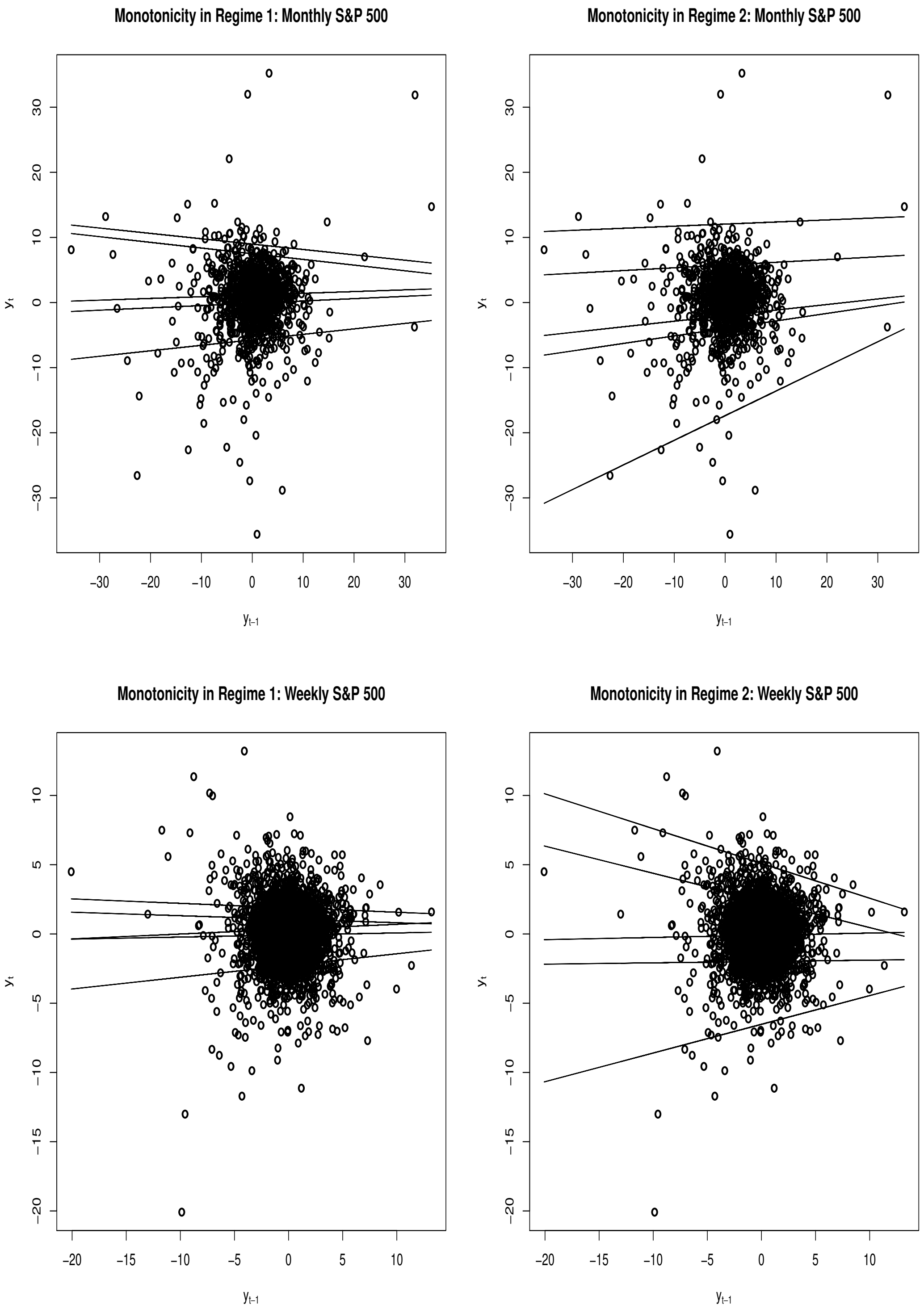


Figure 6: Quantile Monotonicity for Each Single Regime



A Filtering and MSQAR Likelihood

The following inference gives the filtering probability as

$$\xi_{j,t|t}(\tau) = Pr\left(s_t = j|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right) = \sum_{i \in S} Pr\left(s_t = j, s_{t-1} = i|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right)$$

where $\sum_{j \in S} \xi_{j,t|t}(\tau) = 1$ and $\boldsymbol{\Theta}(\tau) = (P(\tau), \boldsymbol{\theta}_{s_t}(\tau))$ is a vector of the parameters with $s_t \in S$. The formulation of filtering probabilities is obtained by Bayes theorem as

$$\xi_{j,t|t}(\tau) = \frac{\sum_{i \in S} p_{ij}(\tau)\xi_{i,t-1|t-1}(\tau)\eta_{j,t}(\tau)}{f\left(y_t|\mathbf{y}_{t-1}, \tau; \boldsymbol{\Theta}(\tau)\right)} \tag{A.1}$$

where $\eta_{j,t}(\tau)$ is the conditional likelihood as

$$\begin{aligned}\eta_{j,t}(\tau) &= f\left(y_t|s_t = j, \mathbf{y}_{t-1}, \tau; \boldsymbol{\theta}(\tau)\right) \\ &= \frac{\tau(1-\tau)}{\varsigma} exp\left\{-\frac{\left(y_t - Q_{y_t}\left(\tau|\mathbf{y}_{t-1}; \boldsymbol{\theta}_j(\tau)\right)\right)}{\varsigma}\left[\tau - I\left(y_t < Q_{y_t}\left(\tau|\mathbf{y}_{t-1}; \boldsymbol{\theta}_j(\tau)\right)\right)\right]\right\}\end{aligned}$$

and

$$f(y_t|\mathbf{y}_{t-1}, \tau; \boldsymbol{\Theta}(\tau)) = \sum_{j \in S}\sum_{i \in S} p_{ij}(\tau)\xi_{i,t-1|t-1}(\tau)\eta_{j,t}(\tau)$$

Thus, the relationship between the filtering and prediction probabilities is given by

$$\xi_{j,t+1|t}(\tau) = Pr\left(s_{t+1} = j|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right) = \sum_{i \in S} p_{ij}(\tau)\xi_{i,t|t}(\tau) \tag{A.2}$$

The inference, similar to Hamilton's filter (Hamilton, 1994), is performed iteratively for $t = 1, ..., T$ with the initial values, $\xi_{j,0|0}(\tau)$ for $j \in S$. The sample likelihood for the τth conditional quantile of y_t is then given by

$$L(\boldsymbol{\Theta}; \tau) = \prod_{t=1}^{T} f(y_t|\mathbf{y}_{t-1}, \tau; \boldsymbol{\Theta}(\tau)) \tag{A.3}$$

In addition, following the approach of Kim (1994), this paper estimates smoothing probabilities, $\xi_{j,t|T}(\tau) = Pr\left(s_t = i|\mathbf{y}_T, \tau; \theta(\tau)\right)$. Apply the Bayes theorem and the Markov property to yield

$$Pr\left(s_t = i|s_{t+1} = j,\ \mathbf{y}_T, \tau; \boldsymbol{\Theta}(\tau)\right) = \frac{p_{ji} Pr\left(s_t = j|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right)}{Pr\left(s_{t+1} = i|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right)}$$

It is therefore the case that

$$Pr\left(s_t = j, s_{t+1} = i|\mathbf{y}_T, \tau; \boldsymbol{\Theta}(\tau)\right) = Pr\left(s_{t+1} = i|\mathbf{y}_T, \tau; \boldsymbol{\Theta}(\tau)\right) \frac{p_{ji} Pr\left(s_t = j|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right)}{Pr\left(s_{t+1} = i|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right)} \tag{A.4}$$

The smoothed inference for date t is the sum of (A.4) over $i \in S$ as

$$\xi_{j,t|T}(\tau) = \sum_{i \in S} Pr\left(s_{t+1} = i|\mathbf{y}_T, \tau; \boldsymbol{\Theta}(\tau)\right) \frac{p_{ji} Pr\left(s_t = j|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right)}{Pr\left(s_{t+1} = i|\mathbf{y}_t, \tau; \boldsymbol{\Theta}(\tau)\right)} \tag{A.5}$$

The smoothed probabilities are thus obtained by iterating on (A.5) backward for $t = T-1, T-2, ..., 1$. This iteration starts with $\xi_{j,T|T}(\tau)$ for $j \in S$ which is estimated from (A.1) for $t = T$. This algorithm is valid only when s_t follows a first-order Markov chain.