

## Bayesian Inference for Markov-switching Skewed Autoregressive Models

Stéphane Lhuissier<sup>1</sup>

August 2019, WP #726

### ABSTRACT

We examine Markov-switching autoregressive models where the commonly used Gaussian assumption for disturbances is replaced with a skew-normal distribution. This allows us to detect regime changes not only in the mean and the variance of a specified time series, but also in its skewness. A Bayesian framework is developed based on Markov chain Monte Carlo sampling. Our informative prior distributions lead to closed-form full conditional posterior distributions, whose sampling can be efficiently conducted within a Gibbs sampling scheme. The usefulness of the methodology is illustrated with a real-data example from U.S. stock markets.

**Keywords:** Regime switching, Skewness, Gibbs-sampler, time series analysis, upside and downside risks.

**JEL classification:** C01; C11; C2; G11

---

<sup>1</sup> Banque de France, 31, Rue Croix des Petits Champs, DGSEI-DEMFI-POMONE 41-1422, 75049 Paris Cedex 01, FRANCE (Email : [stephane\\_lhuissier@club.fr](mailto:stephane_lhuissier@club.fr) ; URL: <http://www.stephanelhuissier.eu> ).

I thank Tobias Adrian, Isaac Baley, Fabrice Collard (Banque de France discussant), Marco Del Negro, Kyle Jurado, Pierre-Alain Pionnier (PSE Discussant), Jean-Marc Robin, Moritz Schularick, Mathias Trabandt and participants at the 10th French Econometrics Conference (PSE), CFE 2018 (University of Pisa), the 2019 Banque de France seminar, ICMAIF 2019 (University of Crete), and the 2019 Padova Workshop (Padova University) for their helpful comments. This paper previously circulated as "The Switching Skewness over the Business Cycle".

## NON-TECHNICAL SUMMARY

Markov-switching models have been a popular tool used in econometric time series analysis. So far almost all extensions and applications of these models have been devoted to detecting abrupt changes in the behavior of the first two moments of a given series --- namely, the mean and the variance. Notable examples include Hamilton (1989)'s model of long-term mean rate of economic growth; stock return volatility models of Turner, Startz, and Nelson (1989) and Pagan and Schwert (1990), the exchange rate dynamics model of Engel and Hamilton (1990) and the real interest rate model of Garcia and Perron (1996).

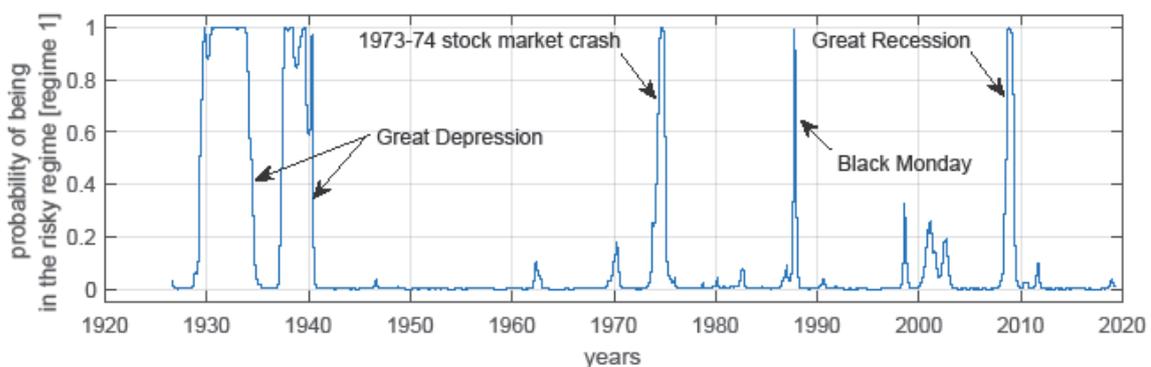
By contrast, skewness, the third moment, has received little attention in the Markov-switching literature, though it is found in many economic time series, such as stock returns and exchange rate returns, and it appears to vary over time (see, for example, Alles and Kling (1994) and Harvey and Siddique (1999) for analysis of U.S. monthly stock indices, and Johnson (2002), Carr and Wu (2007), and Bakshi, Carr, and Wu (2008) for analysis of exchange rate returns). This important omission is due to the commonly used Gaussian assumption for disturbances, which does not allow for possible departure from symmetry.

In this paper, we work with the autoregressive time series (AR) model with Markov-switching introduced by Hamilton (1989), but relax the normality assumption. Instead, we consider a skew-normal distribution proposed by Azzalini (1985, 1986). The key innovation in his work is to account for several degree of asymmetry. With respect to the normal distribution, the skew-normal family is a class of density functions that depends on an additional shape parameter that affects the tails of the density. Such a distribution has already been intensively studied in statistics, biologists, engineers, and medical researchers, but remains largely unexplored in economics.

Our approach here is to propose a simple and easy-to-implement Bayesian framework for such models. More specifically, we develop a Gibbs sampler for Bayesian inferences of AR time series subject to Markov mean, variance and skewness shifts. Our Gibbs sampling procedure can be seen as an extension of Albert and Chib (1993) to account for time variations in the skewness. Specifically, we take advantage of the stochastic representation of skew-normal variables, which is based on a convolution of normal and truncated-normal variables, in order to obtain a straightforward Markov Chain Monte Carlo (MCMC) sampling sequence that involves a 7-block Gibbs sampler for Markov-switching models, in which one can generate in a flexible and straightforward manner alternatively draws from full conditional posterior distributions. In order to make computationally feasible estimation and inference, we provide a companion software package for anyone interested in such models.

As an empirical illustration of our approach, we analyze the time-varying distribution of NYSE/AMEX/NASDAQ stock index returns. We establish two regimes. The first, prevailing during the periods of financial distress, is marked by negative expected returns, large conditional volatility, and positive conditional skewness. The second, frequently observed during tranquil periods, exhibits positive expected returns, low conditional volatility, and negative conditional skewness. Therefore, our result shows that stocks are particularly risky to hold in bad times, like the Great Recession. However, during such times, the positive degree of skewness, indicating that extreme values on the right side of the mean are more likely than the extreme values of the same magnitude on the left side of the mean, allow sometimes to perform large positive returns. Say it differently, stocks produce negative average returns in bad times, but sometimes take large positive hits.

**U.S. stock returns: probability of being in a risky regime**




---

## Inférence bayésienne pour les modèles autorégressifs asymétriques à changements de régimes markoviens

Nous examinons les modèles autorégressifs à changements de régimes markoviens dans lesquels l'hypothèse gaussienne communément utilisée est remplacée par une loi normale asymétrique. Ceci permet de détecter des changements de régimes non seulement dans la moyenne et la variance d'une série temporelle donnée, mais également dans son asymétrie. Un cadre bayésien est proposé à l'aide d'un algorithme de Monte Carlo par chaîne de Markov. Nos distributions informatives a priori permettent d'obtenir une expression analytique des distributions postérieures conditionnelles, dont l'échantillonnage peut être conduit efficacement à l'aide d'un échantillonneur de Gibbs. Un exemple de résultats obtenus à partir de données réelles issues des marchés boursiers américains illustre la pertinence de la méthodologie.

**Mots-clés :** Changement de régime ; Asymétrie ; Échantillonneur de Gibbs ; Analyse de séries temporelles ; Risques à la hausse et à la baisse.

Les Documents de travail reflètent les idées personnelles de leurs auteurs et n'expriment pas nécessairement la position de la Banque de France ou de l'Eurosystème. Ce document est disponible sur [publications.banque-france.fr](http://publications.banque-france.fr)

## I. INTRODUCTION

Markov-switching models have been a popular tool used in econometric time series analysis. So far almost all extensions and applications of these models have been devoted to detecting abrupt changes in the behavior of the first two moments of a given series — namely, the mean and the variance. Notable examples include Hamilton (1989)’s model of long-term mean rate of economic growth; stock return volatility models of Turner, Startz, and Nelson (1989) and Pagan and Schwert (1990); the exchange rate dynamics model of Engel and Hamilton (1990); and the real interest rate model of Garcia and Perron (1996).

By contrast, skewness, the third moment, has received little attention in the Markov-switching literature, though it is found in many economic time series, such as stock returns and exchange rate returns, and it appears to vary over time (see, for example, Alles and Kling (1994) and Harvey and Siddique (1999) for analysis of U.S. monthly stock indices, and Johnson (2002), Carr and Wu (2007), and Bakshi, Carr, and Wu (2008) for analysis of exchange rate returns). This important omission is due to the commonly used Gaussian assumption for disturbances, which does not allow for possible departure from symmetry. In this paper, we work with the autoregressive time series (AR) model with Markov-switching introduced by Hamilton (1989), but relax the normality assumption. Instead, we consider a skew-normal distribution proposed by Azzalini (1985, 1986). The key innovation in his work is to account for several degree of asymmetry. With respect to the normal distribution, the skew-normal family is a class of density functions that depends on an additional shape parameter that affects the tails of the density. Such a distribution has already been intensively studied in statistics, biologists, engineers, and medical researchers, but remains largely unexplored in economics.

Our approach here is to propose a simple and easy-to-implement Bayesian framework for such models. More specifically, we develop a Gibbs sampler for Bayesian inferences of AR time series subject to Markov mean, variance and skewness shifts. Our Gibbs sampling procedure can be seen as an extension of Albert and Chib (1993) to account for time variations in the skewness. Specifically, we take advantage of the stochastic representation of skew-normal variables, which is based on a convolution of normal and truncated-normal variables, in order to obtain a straightforward Markov Chain Monte Carlo (MCMC) sampling sequence that involves a 7-block Gibbs sampler for Markov-switching models, in which one can generate

in a flexible and straightforward manner alternatively draws from full conditional posterior distributions. In order to make computationally feasible estimation and inference, we provide a companion software package for anyone interested in such models.

As an empirical illustration of our approach, we analyze the NYSE/AMEX/NASDAQ stock index. We establish two regimes. The first, prevailing during the periods of financial distress, is marked by negative expected returns, large conditional volatility, and positive conditional skewness. The second, frequently observed during tranquil periods, exhibits positive expected returns, low conditional volatility, and negative conditional skewness. Therefore, our result shows that stocks are particularly risky to hold in bad times, like the Great Recession. However, during such times, the positive degree of skewness, indicating that extreme values on the right side of the mean are more likely than the extreme values of the same magnitude on the left side of the mean, allow sometimes to perform large positive returns. Say it differently, stocks produce negative average returns in bad times, but sometimes take large positive hits. Finally, we show that our Markov-switching skewed AR model is very strongly preferred to a Markov-switching symmetric AR model (i.e., errors are governed by normal shocks) by standard model selection criteria. Overall, our results corroborate with the literature that time-varying skewness is a real feature of U.S. stock index.

In the literature, time variation in skewness has been, in the first place, modelled through the generalized autoregressive conditional heteroskedasticity (GARCH) models. Notable examples include Harvey and Siddique (1999), Jondeau and Rockinger (2003), and Christoffersen, Heston, and Jacobs (2006). The deterministic behavior of such systems lead, however, to limited implications. Feunou and Tédongap (2012) and Iseringhausen (2018) go a step further by modelling time-varying skewness as stochastic by extending the standard stochastic volatility (SV) model; the parameter that governs the asymmetry of the distribution evolves according to an autoregressive process. By contrast, our Markov-switching framework offers another way of modelling stochastically time-varying skewness. This choice is justified by the fact that many economic and financial data sets exhibit rapid shifts in their behaviors due, for example, to financial or currency crises, and thus models with smooth and drifting coefficients seem to be less suited for capturing such changes. Crises are well-known for hitting the economy instantaneously, which favor models with abrupt changes like Markov-switching models.

In this context, the Nakajima (2013) specification is closest to our approach: namely, allowing the parameter that governs the asymmetry of the (skew-normal) distribution<sup>1</sup> to vary over time according to a first-order Markov-switching process. There are, however, some major differences. First, our Gibbs sampler leads to closed-form full conditional posterior distributions for any parameter of the Markov-switching skewed AR model, whereas a Random-Walk Metropolis-Hastings (RWMH) algorithm is needed to sample the shape parameters in Nakajima (2013). Second, our algorithm assumes that mean, variance, and skewness switch at the same time, whereas only skewness can switch in Nakajima (2013). This considerably restricts the distributional flexibility and complicates the interpretation of the parameters of the model since in the skew-normal distribution the shape parameter also affects the mean and the variance. The potential problem is that if a time series specified is in fact subject to only mean/variance shifts, then the shape parameter may incorrectly switch to compensate for the absence of mean or variance shifts within framework. Third, as an empirical illustration of our approach, we examine the U.S. excess stock returns, not the exchange rate returns.

The paper is organised as follows. Section II presents a brief overview of the skew-normal family of distributions. Section III outlines the Markov-switching model with skew-normal distributions, and explains how to estimate it. Section IV presents a MCMC method to carry out posterior inference. Section V deals with real data set from U.S. stock markets. Section VI concludes.

## II. THE SKEW-NORMAL DISTRIBUTION: A PRELIMINARY

In this section, we first review the necessary properties of the skew-normal distribution, and next we describe a constant-parameters AR model with skew-normal errors.

**II.1. Basic notions.** The skew-normal family was introduced by Azzalini (1985, 1986) as the extension of the normal family from a symmetric form to an asymmetric form. It is a distribution that has an additional parameter: a shape parameter  $\alpha \in \mathbb{R}$ , which allow for possible deviation from symmetry. The following paragraphs provide the general framework of such distribution.

---

<sup>1</sup>Note also that Nakajima (2013) uses a generalized hyperbolic skew Student's  $t$  distribution instead of a skew-normal one.

Let  $Y$  a random variable with the following density

$$p(Y|\xi, \omega^2, \alpha) = \frac{2}{\omega} \phi\left(\frac{Y - \xi}{\omega}\right) \Phi\left(\alpha \frac{Y - \xi}{\omega}\right), \quad (1)$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the standard normal density function and cumulative distribution function, respectively. We say that the random variable  $Y$  follows a univariate skew-normal distribution with location parameter  $\xi$ , scale parameter  $\omega^2$ , and a skewness parameter  $\alpha$ :

$$\text{skew-normal}(Y|\xi, \omega^2, \alpha). \quad (2)$$

If the skewness parameter is equal to zero, then the density of  $Y$  is a normal distribution with mean  $\xi$ , and standard deviation  $\omega$ .

The moments of the skew-normal distribution can be summarized as follows

$$E[Y] = \xi + \omega\delta\sqrt{\frac{2}{\pi}}, \quad \text{var}[Y] = \omega^2\left(1 - \frac{2}{\pi}\delta^2\right), \quad (3)$$

where  $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$  and  $\delta \in (-1, 1)$ .

As an illustration, Figure 1 displays skew-normal density functions when  $\alpha = 0, -1, -4, -10$  in the left-hand panel, and  $\alpha = 0, 1, 4, 10$  in the right-hand panel. For the remaining parameters, we set  $\xi = 0$  and  $\omega^2 = 1$ . As can be seen, the skewness parameter strongly alters the tails of the distribution. When  $\alpha$  is negative, the distribution tends to be skewed to the left, while when it is positive, the distribution is skewed to the right.

Following Henze (1986), an interesting characteristic of the skew-normal distribution is that it can be represented stochastically. In particular, the skew-normal distribution in (2) is equivalent to

$$Y = \xi + \omega\delta Z + \omega\sqrt{(1 - \delta^2)}U, \quad (4)$$

where  $Z$  and  $U$  are independent random variables defined, respectively, as follows:

$$Z = \text{truncated-normal}(Z|0, 1)_{Z>0} \quad \text{and} \quad U = \text{normal}(U|0, 1), \quad (5)$$

with  $\text{truncated-normal}(x|\mu, \Sigma)$  denotes the truncated-normal distribution of  $x$  with mean  $\mu$ , variance  $\Sigma$ , and truncation below zero, and  $\text{normal}(x|\mu, \Sigma)$  denotes the normal distribution of  $x$  with mean  $\mu$  and variance  $\Sigma$ . Say it differently, the skew-normal distribution may be seen as the combination of a normal random variable and a truncated standard normal variable.

In next sections, we will show that this elegant and stochastic representation is crucial in order to obtain our Gibbs-sampling procedure.

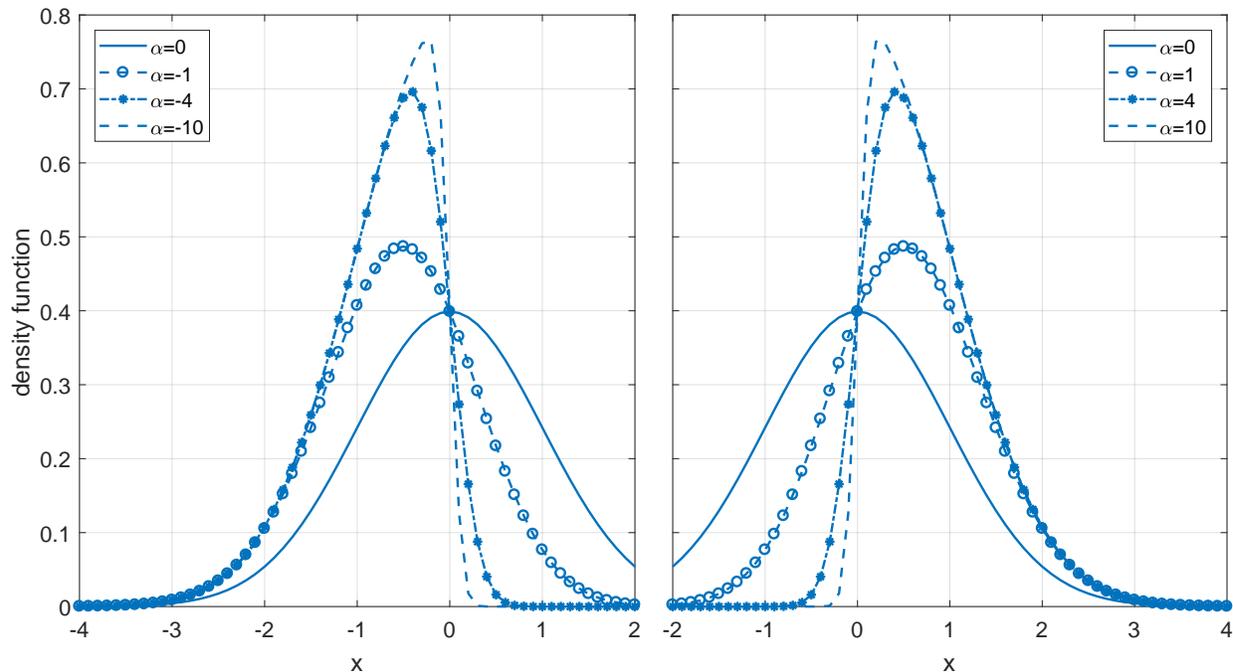


FIGURE 1. Skew-normal density functions when  $\alpha = 0, -1, -4, -10$  in the left-hand panel, and  $\alpha = 0, 1, 4, 10$  in the right-hand panel. The location parameter,  $\xi$ , is set to zero, and the scale parameter,  $\omega^2$  to one.

**II.2. Skewed AR models.** Consider now the following AR model in which the observation at time  $t$ ,  $y_t$ , leads to the following representation:

$$y_t = c + \phi_1(y_{t-1} - c) + \dots + \phi_\tau(y_{t-\tau} - c) + \epsilon_t, \quad t = 1, \dots, T, \quad (6)$$

where the vector  $\phi = (\phi_1, \dots, \phi_\tau)$  contains the coefficients at the lag  $\tau$ ;  $T$  is the sample size;  $c$  is a constant, and  $\epsilon_t$  follows a skew-normal distribution as follows:

$$\text{skew-normal}(\epsilon_t | b\Delta, \omega^2, \alpha), \quad (7)$$

where  $\Delta = \omega\delta$ ,  $\omega^2$ , and  $\alpha$  are unknown parameters. We assume that  $b = -\sqrt{2/\pi}$ . By doing so, it guarantees that  $E(\epsilon_t) = 0$ . By considering equations (6) and (7), it can be shown that this model is a stochastic process constructed by skew-normal process. Therefore, following the work of Minozzo and Ferracuti (2012), we conclude the stationarity of the model.

A compact form of the AR model in equation (6) is given by:

$$y_t = c + \phi x_t + \epsilon_t, \quad (8)$$

with  $x_t = [y_{t-1}, \dots, y_{t-\tau}]'$ . Equations (6) and (7) are equivalent to

$$p(y_t|Y_{t-1}, c, \Delta, \phi, \omega^2, \alpha) = \text{skew-normal}(y_t|c + \phi x_t + b\Delta, \omega^2, \alpha), \quad (9)$$

with  $Y_t = [y_1, \dots, y_t]$ . The stochastic representation of equation (9) can be conveniently reformulated as

$$y_t = \mu + \phi x_t + \omega \delta z_t + \omega \sqrt{1 - \delta^2} \epsilon_t, \quad (10)$$

$$p(z_t) = \text{truncated-normal}(z_t|0, 1)_{z_t > 0}, \quad (11)$$

$$p(\epsilon_t) = \text{normal}(\epsilon_t|0, 1), \quad (12)$$

where  $\mu = c + b\Delta$ .

### III. SKEWED AUTOREGRESSIVE MODELS WITH MARKOV SHIFTS

We now extend constant-parameters skewed AR models to a setting in which parameters follow a Markov-switching process. For  $1 \leq i, j \leq h$ , the discrete and unobservable variable  $s_t$  is an exogenous first order Markov process with the following transition matrix  $Q$ :

$$Q = \begin{bmatrix} q_{1,1} & \cdots & q_{1,h} \\ \vdots & \ddots & \vdots \\ q_{h,1} & \cdots & q_{h,h} \end{bmatrix}, \quad (13)$$

where  $h$  is the total number of regimes; and  $q_{i,j} = \Pr(s_t = i | s_{t-1} = j)$  denote the transition probabilities that  $s_t$  is equal to  $i$  given that  $s_{t-1}$  is equal to  $j$ , with  $q_{i,j} \geq 0$  and  $\sum_{i=1}^h q_{i,j} = 1$ . We assume that the conditional densities of  $y_t$ , given  $s_t$ , arise from a skew-normal distribution. By integrating out  $s_t$ , the marginal density of  $y_t$  leads to a weighted average of conditional densities as given by

$$p(y_t|Y_{t-1}, \theta) = \sum_{s_t \in H} p(y_t|Y_{t-1}, s_t, \theta) \Pr(s_t, \theta), \quad (14)$$

where  $H$  is a finite set of  $h$  elements and is taken to be the set  $\{1, \dots, h\}$ , and  $\theta = (\theta_i)_{i \in H}$  with  $\theta_i = (\mu_i, \phi_i, \omega_i^2, \alpha_i)$ . The conditional likelihood at time  $t$ ,  $p(y_t|Y_{t-1}, s_t, \theta)$ , is generated by

$$p(y_t|Y_{t-1}, s_t, \theta) = \frac{2}{\omega_{s_t}} \phi \left( \frac{y_t - \phi x_t - \mu_{s_t}}{\omega_{s_t}} \right) \Phi \left( \alpha_{s_t} \frac{y_t - \phi x_t - \mu_{s_t}}{\omega_{s_t}} \right). \quad (15)$$

Equation (14) can be evaluated recursively by updating  $\Pr(s_t, \theta)$  according to the Hamilton (1989)'s filter (See Appendix A). Interestingly, the inclusion of the additional shape parameter does not require to modify the original filter.

For mixtures defined in (14), it follows that each conditional density leads to the following stochastic representation:

$$y_t = \mu_{s_t} + \phi x_t + \omega_{s_t} \delta_{s_t} z_t + \omega_{s_t} \sqrt{1 - \delta_{s_t}^2} \epsilon_t, \quad (16)$$

where  $z_t$  and  $\epsilon_t$  are defined in equations (11) and (12). Note that Markov switching affects the intercept, the scale, and the shape parameters, not the coefficient parameter vector  $\phi$ .

Given (15), it follows that the overall likelihood of  $Y_T$  is

$$p(Y_T|\theta) = \prod_{t=1}^T \left[ \sum_{s_t \in H} p(y_t|Y_{t-1}, s_t, \theta) \Pr(s_t, \theta) \right]. \quad (17)$$

To form the posterior density,  $p(\theta|Y_T)$ , we combine the overall likelihood function  $p(Y_T|\theta)$  with the prior  $p(\theta)$ :

$$p(\theta|Y_T) \propto p(Y_T|\theta)p(\theta), \quad (18)$$

The posterior density  $p(\theta|Y_T)$  is not of standard form, but we will show in the next section that it is possible to use the idea of Gibbs-sampling by sampling alternatively from conditional posterior distributions.

For computational reasons, we employ a logarithm transformation in equation (18) to obtain the log-posterior function as follows

$$\log \{p(\theta|Y_T)\} \propto \log \{p(Y_T|\theta)\} + \log \{p(\theta)\}, \quad (19)$$

where the conditional log-likelihood at time  $t$ , given  $s_t$ , is as follows

$$\log \{p(y_t|Y_{t-1}, s_t, \theta)\} = \text{constant} - \log \{\omega_{s_t}\} - \frac{(y_t - \phi x_t - \mu_{s_t})^2}{2\omega_{s_t}^2} + \log \left\{ \Phi \left( \alpha_{s_t} \frac{y_t - \phi x_t - \mu_{s_t}}{\omega_{s_t}} \right) \right\}.$$

The strategy to find the posterior mode of (19) is to generate a sufficient number of draws from the prior distribution of each parameter. Each set of points is then used as starting points to the CSMINWEL program, the optimization routine developed by Christopher A. Sims. Starting the optimization process at different values allows us to correctly cover the parameter space and avoid getting stuck in a ‘‘local’’ peak. Note, however, that we do not need to use a more complicated method for finding the mode like the blockwise optimization method developed by Sims, Waggoner, and Zha (2008), in which the authors break the

parameters into several subblocks of parameters and apply a standard hill-climbing quasi-Newton optimization routine to each block, while keeping the other subblocks constant, in order to maximize the posterior density<sup>2</sup>. The size of the Markov-switching univariate model in (16) remains relatively small and allows us employ a standard technique.

#### IV. A GIBBS SAMPLER

In the existing statistical literature, efficient posterior simulation algorithms have been applied to finite mixtures of skew-normal distributions. See, for example, Lin, Lee, Yen, and Chung (2007) and Frühwirth-Schnatter and Pyne (2010). Our work differs from this literature along several dimensions. First, we assume that regime shifts evolve according to a Markov chain. Finite mixture models seems to be less suited for time series analysis as they consider unrealistically rapid switching regimes. By contrast, Markov-switching models can be seen as an extension of mixture models with a general solution to the problem of state persistence. Second, we introduce an autoregressive process of finite order, as naturally modelled in the macroeconomics literature. Third, our MCMC algorithm is able to directly generate draws of the shape parameters from a closed-form full conditional posterior distribution, and thus avoiding to employ a RWMH algorithm. Overall, Our MCMC approach can be seen as an extension of Albert and Chib (1993) to Markov mean, variance and skewness shifts.<sup>3</sup>

A MCMC simulation method is employed to approximate the joint posterior density  $p(\theta, Z_T, S_T | Y_T)$ , where  $S_t = [s_1, \dots, s_t]$ , and  $Z_t = [z_1, \dots, z_t]$  for  $t \geq 1$ . Here, a key to Bayesian estimation of a Markov-switching skewed AR model is to exploit the stochastic representation as defined in equation (16).

Because we consider a Bayesian approach to inference of the complete model, as defined by equations (11), (12), (13), and (16), we now explicit our priors. For  $k = 1, \dots, h$ , the prior on the set of parameters  $\theta$  is given by:

$$\phi = \text{normal}(\phi | \bar{b}, \bar{B}), \quad (20)$$

$$\mu_k = \text{normal}(\mu_k | \bar{a}, \bar{A}), \quad (21)$$

---

<sup>2</sup>See, for example, Lhuissier (2017) and Lhuissier and Tripier (2019), for applications of such a method in the context of multivariate-equation Markov-switching models.

<sup>3</sup>Albert and Chib (1993) develop a Gibbs sampling for AR time series subject to Markov mean and variance shifts.

$$\omega_k = \text{inv-gamma}(\omega_k | \bar{\alpha}, \bar{\beta}), \quad (22)$$

$$q_k = \text{dirichlet}(q_k | \bar{\alpha}_{1k}, \dots, \bar{\alpha}_{hk}), \quad (23)$$

$$\alpha_k = \text{normal}(\alpha_k | \alpha_0, \psi_0), \quad (24)$$

where  $\bar{b}, \bar{B}, \bar{a}, \bar{A}, \bar{\alpha}, \bar{\beta}, \bar{\alpha}_{1k}, \dots, \bar{\alpha}_{hk}$  are the hyperparameters;  $q_k$  denotes the  $k$ th column of  $Q$ ; and  $\text{dirichlet}(q_k | \alpha_1, \dots, \alpha_h)$  is the Dirichlet distribution of  $q_k$  as follows:

$$\frac{1}{B(\alpha)} \prod_{i=1}^h q_i^{\alpha_i - 1} \quad (25)$$

with  $B(\alpha) = \frac{\prod_{i=1}^h \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^h \alpha_i)}$ , where  $\Gamma$  denotes the standard gamma function. As can be seen, we directly specify informative priors for the shape parameter  $\alpha_k$  rather than for  $\delta_k$ , the transformed shape parameters.<sup>4</sup>

The stochastic representation leads us to exploit the idea of Gibbs-sampling. Let  $\theta_{\neq x}$  contain the model's parameters, except for  $x$ . The MCMC sampling scheme at the  $(i)$ st iteration, for  $i = 1, \dots, N_1 + N_2$ , consists of sampling from the following conditional posterior distributions

- (1)  $p\left(S_T^{(i)} | Y_T, \theta^{(i-1)}\right)$ ;
- (2)  $p\left(Q^{(i)} | S_T^{(i)}\right)$ ;
- (3)  $p\left(Z_T^{(i)} | Y_T, S_T^{(i)}, \theta^{(i-1)}\right)$ ;
- (4)  $p\left(\phi^{(i)} | Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq \phi}^{(i-1)}\right)$ ;
- (5)  $p\left(\mu_k^{(i)} | Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq \mu_k}^{(i-1)}\right)$ ;
- (6)  $p\left(\omega_k^{(i)} | Y_T, S_T^{(i)}, Z_T^{(i)}, \phi^{(i)}, \delta_k^{(i-1)}\right)$ ;
- (7)  $p\left(\alpha_k^{(i)} | Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq \alpha}^{(i)}\right)$ .

A few items deserve discussion. First, simulation from the conditional posterior density  $p\left(S_T^{(i)} | Y_T, \theta^{(i-1)}\right)$ , given  $Z_T$  and  $\theta$ , is standard and in closed form. Second, simulation from the conditional posterior density  $p\left(Q^{(i)} | S_T^{(i)}\right)$  is independent of the time series  $Y_T$ ,

---

<sup>4</sup>When specifying priors for  $\delta_k$ , instead of  $\alpha_k$ , there is no closed form for the posterior distribution of  $\delta_k$ , and one must impose a non-informative prior (i.e., uniform distribution on a bounded interval between  $-1.00$  and  $1.00$ ), and use a RWMH algorithm. This explains our difference with the Nakajima (2013)'s MCMC algorithm.

the random variable  $Z_T$  and the model's other parameters. Third, simulation from the conditional posterior density  $p\left(Z_T^{(i)}|Y_T, S_T^{(i)}, \theta^{(i-1)}\right)$ , given  $Y_t, Z_t$  and  $\theta$ , is available in closed form due to the stochastic representation of the Markov-switching model through normal and truncated-normal variables. Fourth, simulations from the conditional posterior densities  $p\left(\phi^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq\phi}^{(i-1)}\right)$  and  $p\left(\sigma_k^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \phi^{(i)}, \delta_k^{(i-1)}\right)$  reduces to Bayesian inference for Markov-switching models with known allocations,  $S_T$ . Finally, simulation from the conditional posterior density  $p\left(\alpha_k^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq\alpha}^{(i)}\right)$  is in closed form, and follows an unified skew-normal distribution introduced by Arellano-Valle and Azzalini (2006).

This sampler begins with setting parameters at the peak of the posterior density function. We collect  $N_1 + N_2$  draws of the MCMC sequence and keep only the last  $N_2$  values. The only computational complication involves the simulation from the posterior distribution of  $\alpha$ , which requires to sample from a truncated multivariate normal distribution. With respect to Albert and Chib (1993), our Gibbs-sampling procedure involves two more blocks, namely the conditional posterior distribution of  $Z_T$ , given the parameters and the states, and the conditional posterior distribution of  $\alpha_k$ , given  $Z_T, S_T$  and the remaining parameters.

The researcher can use our companion computer program, written in C++, to estimate and simulate an AR skewed model with Markov shifts. The user just needs to provide an input file in which he/she must mention each specification (such as the number of lags, prior settings, the number of draws, the number of burn-in, etc...) of the AR model.<sup>5</sup> Due to its simplicity and efficiency, we believe that our companion computer code is relevant for anyone interested in inference of such models.

The subsections that follow provide the computational details for each conditional posterior distribution.

**IV.1. Conditional posterior densities,  $p\left(S_T^{(i)}|Y_T, \theta^{(i-1)}\right)$ .** For  $t = 1, 2, \dots, T$ , we can generate  $S_T^{(i)}$  using the Carter and Kohn (1994)'s multi-move Gibbs-sampling as following

$$p(S_T^{(i)}|Y_T, \theta^{(i-1)}) = p(s_T^{(i)}|Y_T, \theta^{(i-1)}) \prod_{t=1}^{T-1} p(s_t^{(i-1)}|s_{t+1}^{(i)}, Y_T, \theta^{(i-1)}). \quad (26)$$

---

<sup>5</sup>The software is available at <http://stephanelhuissier.eu/assets/skewcodes.zip>. In Appendix B, we provide a concrete example of how to use the interface with our C++ computer code. The software program was written in modern C++11 and mainly uses the GNU Scientific Library (GSL-2.5) and the Eigen library (3.3.5). Both libraries are open source.

Drawing  $S_T^{(i)}$  from the full conditional distribution based on this equation is standard. We begin with a draw from  $p(s_T|Y_T, \theta)$  obtained with the Hamilton (1989) basic filter, and then iterate recursively backward to draw  $s_{T-1}, s_{T-2}, \dots, 1$  according to

$$p(s_t|Y_T, \theta) = \sum_{s_{t+1}} p(s_t|Y_T, \theta, s_{t+1})p(s_{t+1}|Y_T, \theta), \quad (27)$$

where

$$p(s_t|Y_T, \theta, s_{t+1}) = \frac{\Pr[s_{t+1}|s_t] p(s_t|Y_T, \theta)}{p(s_{t+1}|Y_t, \theta)}. \quad (28)$$

Appendix A provides the details for derivation of the Hamilton (1989) filter.

**IV.2. Conditional posterior densities,  $p(Q^{(i)}|S_T^{(i)})$ .** The conditional posterior distribution of  $Q^{(i)}$  is as follows:

$$p(q_k^{(i)}|S_T) = \text{dirichlet}(q_k^{(i)}|\bar{\alpha}_{1k} + n_{1k}, \dots, \alpha_{Hk}^- n_{Hk}) \quad (29)$$

where  $q_k^{(i)}$  is the  $k$ th column of  $Q^{(i)}$ ,  $n_{ij}$  is the total number of transitions from state  $j$  to state  $i$  over the entire sample.

Drawing  $Q^{(i)}$  from the above full conditional distribution is also standard.

**IV.3. Conditional posterior densities,  $p(Z_T^{(i)}|Y_T, S_T^{(i)}, \theta^{(i-1)})$ .** Here, the nice property of such a model is that the full conditional distribution of  $Z_t$  given  $Y_t, S_T^{(i)}$ , and  $\theta^{(i)}$  is available in closed form.

For  $t = 1, 2, \dots, T$ , we generate  $Z_T^{(i)}$  according to

$$p(Z_T^{(i)}|Y_T, S_T^{(i)}, \theta^{(i-1)}) = \prod_{t=1}^T p(z_t^{(i)}|Y_t, S_t^{(i)}, \theta^{(i-1)}), \quad (30)$$

where

$$p(z_t^{(i)}|Y_t, S_t^{(i)}, \theta^{(i-1)}) = \text{truncated-normal}\left(z_t^{(i)}|\delta_{s_t}^{(i-1)}(y_t - \phi^{(i-1)}x_t - \mu_{s_t}^{(i-1)}), \omega_{s_t}^{(i-1)2}\left(1 - \delta_{s_t}^{(i-1)2}\right)\right)_{z_t^{(i)} > 0}.$$

**IV.4. Conditional posterior densities,  $p(\phi^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq \phi}^{(i-1)})$ .** If we let  $y_t^* = \frac{y_t - \mu_{s_t} - \delta_{s_t} z_t}{\omega_{s_t} \sqrt{1 - \delta_{s_t}^2}}$ , and  $x_t^* = \frac{x_t}{\omega_{s_t} \sqrt{1 - \delta_{s_t}^2}}$ , we obtain an homoskedastic model as follows

$$y_t^* = \phi x_t^* + \nu_t, \quad (31)$$

where  $\nu_t$  follows a standard normal distribution. Then, simulation from the full conditional distribution of  $\phi^{(i)}$ , given  $Y_T, S_T^{(i)}, Z_T^{(i)}$  and  $\theta_{\neq \phi}^{(i-1)}$ , becomes straightforward, given a conjugate prior distribution. The posterior is defined as

$$p\left(\phi^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq\phi}^{(i-1)}\right) = \text{normal}\left(u_\phi^{(i)}, U_\phi^{(i)}\right), \quad (32)$$

where

$$u_\phi^{(i)} = (\bar{B}^{-1} + X'X)^{-1} (\bar{B}^{-1}\bar{b} + X'y_t^*), \quad (33)$$

$$U_\phi^{(i)} = (\bar{B}^{-1} + X'X)^{-1}, \quad (34)$$

and  $\bar{b}$  and  $\bar{B}$  are known hyperparameters of the prior distribution — the mean and the variance, respectively — and  $X = [x_1^*, \dots, x_T^*]'$ .

**IV.5. Conditional posterior densities,  $p\left(\mu_k^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq\mu_k}^{(i-1)}\right)$ .** If we let  $y_t^{**} = \frac{y_t - \phi x_t - \delta_{s_t} z_t}{\omega_{s_t} \sqrt{1 - \delta_{s_t}^2}}$ , and  $x_t^{**} = \frac{1}{\omega_{s_t} \sqrt{1 - \delta_{s_t}^2}}$ , we obtain an homoskedastic model as follows

$$y_t^{**} = \mu_{s_t} x_t^{**} + \varsigma_t, \quad (35)$$

where  $\varsigma_t$  follows a standard normal distribution.

Therefore, the posterior can be defined as

$$p\left(\mu_k^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \theta_{\neq\mu}^{(i-1)}\right) = \text{normal}(v_{\mu,k}, V_{\mu,k}), \quad (36)$$

where

$$v_{\mu,k} = \left(\bar{A}^{-1} + \sum_{t \in \{t:s_t=k\}} x_t^{**2}\right)^{-1} \left(\bar{A}^{-1}\bar{a} + \sum_{t \in \{t:s_t=k\}} x_t^{**} y_t^{**}\right), \quad (37)$$

$$V_{\mu,k} = \left(\bar{A}^{-1} + \sum_{t \in \{t:s_t=k\}} x_t^{**2}\right)^{-1}, \quad (38)$$

with  $\bar{a}$  and  $\bar{A}$  are known hyperparameters of the prior distribution — the mean and the variance, respectively.

**IV.6. Conditional posterior densities,  $p\left(\omega_k^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \phi^{(i)}, \mu_k^{(i)}, \delta_k^{(i-1)}\right)$ .** Given  $Y_t, S_T, Z_T, \theta$ , and  $S_T$ , the scale parameter  $\omega$  can be drawn using the following inverse-gamma distribution

$$p\left(\omega_k^{(i)}|Y_T, S_T^{(i)}, Z_T^{(i)}, \phi^{(i)}, \mu_k^{(i)}, \delta_k^{(i-1)}\right) = \text{inv-gamma}(\bar{\alpha} + s s r_k, \bar{\beta} + T_k), \quad (39)$$

where  $T_k$  is the number of elements in  $\{t : s_t = k\}$  for  $k = 1, \dots, h$ , and  $ssr_k$  is the sum of squared residual defined as

$$ssr_k = \sum_{t \in \{t: s_t=k\}} \left( \frac{y_t - \phi^{(i)} x_t - \mu_{s_t}^{(i)} - \delta_{s_t}^{(i-1)} z_t^{(i)}}{\sqrt{1 - \delta_{s_t}^{(i-1)2}}} \right)^2, \quad (40)$$

where  $\bar{\alpha}$  and  $\bar{\beta}$  are the shape hyperparameters implied by the choice for the prior mean and variance.

**IV.7. Conditional posterior densities,  $p(\alpha_k^{(i)} | Y_T, S_T^{(i)}, \theta_{\neq \alpha}^{(i)})$ .** Let  $\bar{y}_t = \frac{y_t - \phi x_t - \mu_{s_t}}{\omega_{s_t}}$  and  $\bar{Y}_T = [\bar{y}_1, \dots, \bar{y}_T]'$ . Consider the following derivation for the full conditional distribution of  $\alpha_k$ , given  $Y_T$ ,  $S_T^{(i)}$ , and  $\theta_{\neq \alpha}^{(i)}$ :

$$\begin{aligned} p(\alpha_k^{(i)} | Y_T, S_T^{(i)}, \theta_{\neq \alpha}^{(i)}) &\propto \phi\left(\frac{\alpha_k - \alpha_0}{\psi_0}\right) \prod_{t=1}^T \Phi(\alpha_k \bar{y}_t) \\ &\propto \phi\left(\frac{\alpha_k - \alpha_0}{\psi_0}\right) \Phi_T(\alpha_k \bar{Y}_T; I_T) \\ &\propto \phi\left(\frac{\alpha_k - \alpha_0}{\psi_0}\right) \Phi_T(\bar{Y}_T \alpha_0 + \bar{Y}_T(\alpha_k - \alpha_0); I_T) \\ &\propto \text{SUN}_{1,T}(\alpha_k^{(i)} | \alpha_0, \Delta_{1,k} \alpha_0 / \psi_0, \psi_0, 1, \Delta_{1,k}, \Gamma_{1,k}) \end{aligned}$$

where  $\text{SUN}_{d,m}(x | \xi, \tau, \omega, \Omega, \Delta, \Gamma)$  refers to the unified skew-normal (SUN) distribution introduced by Arellano-Valle and Azzalini (2006) as follows

$$\phi_d(z - \xi; \omega \Omega \omega) \frac{\Phi_m(\gamma + \Delta \Omega^{-1} \omega^{-1} (z - \xi); \Gamma - \Delta \Omega^{-1} \Delta')}{\Phi_m(\gamma; \Gamma)^{-1}}, \quad (41)$$

with  $\Phi_d$  is the cumulative density function of  $d$ -variate Gaussian distribution with variance-covariance matrix  $\Sigma$ ,  $\Omega$ ,  $\Gamma$ , and  $\Omega^* = ((\Gamma, \Delta)', (\Delta', \Omega)')$  are correlations matrices, and  $\omega$  is a  $d \times d$  diagonal matrix;  $\Delta_{1,k} = [\zeta_t]_{t \in \{t: s_t=k\}}$  with  $\zeta_t = \psi_0 \bar{y}_t^2 (\psi_0^2 \bar{y}_t^2 + 1)^{-1/2}$ ;  $\Gamma_{1,k} = I - \text{diag}(\Delta_{1,k})^2 + \Delta_{1,k} \Delta_{1,k}^2$ ; and where  $\text{diag}(V)$  is a diagonal matrix, the elements of which coincide with those of vector  $V$ .<sup>6</sup>

<sup>6</sup>Canale, Pagui, and Scarpa (2016) demonstrate that informative priors (i.e., normal or skew-normal distribution) for the shape parameter of a constant skew-normal model lead to closed-form full conditional distributions.

To simulate draws from the SUN distribution, one can use its stochastic representation. Let  $U_0$  and  $U_{1,-\gamma}$  have the following distribution

$$U_0 = \text{normal}(U_0|0, \bar{\Psi}_\Delta), \quad \text{and} \quad U_{1,-\gamma} = \text{truncated-normal}(U_1|0, \Gamma)_{-\gamma}. \quad (42)$$

Then, it can be show the SUN distribution can be generated as follows

$$\xi + \omega (U_0 + \Delta\Gamma^{-1}U_{1,-\gamma}). \quad (43)$$

Once we obtain  $\alpha_k$ , we can directly transform it to recover  $\delta_k = \frac{\alpha_k}{\sqrt{1+\alpha_k^2}}$ .

**IV.8. Label-switching.** Due to the label-switching problem, we normalize the labels of regimes to obtain accurate posterior distributions, such as, for example,  $\alpha_1 < \dots < \alpha_h$ . To achieve this constraint, we adopt rejection sampling.

## V. APPLICATION: U.S. EXCESS STOCK RETURNS

In this section, we apply the proposed algorithm to the U.S. excess stock return for value-weighted portfolio of all CRSP firms listed on the NYSE, AMEX, or NASDAQ. The time series, shown in Figure 2, is organized monthly from July 1926 to April 2019.<sup>7</sup> We consider a Markov-switching skewed AR(1) model with  $h=2$  regimes.

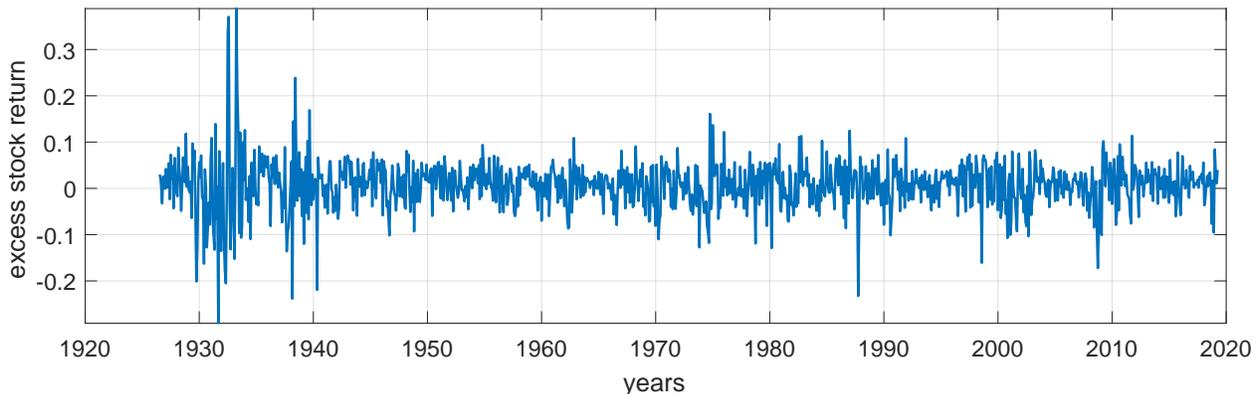


FIGURE 2. Sample period: 1926.M07 — 2019.M04. Historical path of U.S. excess stock returns.

The priors are defined in Table 1, which reports the specific distribution, the mean and the standard deviation for each parameter. A few of them deserve further discussion. First,

<sup>7</sup>The excess return dataset is freely available at the Kenneth R. French's homepage: [https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html)

for  $\mu_k$  and  $\phi_1$  we choose a normal prior with the mean 0.00 and the standard deviation 2.00. These priors are rather dispersed and cover a large parameter space. Second, The prior for the scale parameter,  $\omega_k$ , follows an inverse-gamma distribution, with the mean 0.05 and the standard deviation 0.10. Third, the prior for the shape parameter,  $\alpha_k$ , has a normal density with the mean 0.00 and the standard deviation 2.00. Fourth, it may be worth noting that we impose the exact same prior across regimes, so that differences between regimes result from data rather than priors. Finally, the prior duration of each regime is about twenty months, meaning that the average probability of staying in the same regime is equal to 0.95 and a standard deviation equal to 0.05.

The results shown in this paper are based on 11,000 draws with our Gibbs-sampling procedure developed in Section IV. We discard the first 1,000 draws as burn-in, and keep every 10-th draw in order to achieve an approximately independent sample. On the right-hand side of Table 1, we report the posterior mode, mean, and median with the 90 percent probability interval for each parameter of the estimated model.

TABLE 1. AR(1) Markov-switching model for U.S. excess stock returns.

Coefficient	Description	Prior			Posterior				
		Density	para(1)	para(2)	Mode	Mean	Median	[5;	95]
$\mu_1$	location	N	0.00	2.00	-0.1094	-0.0961	-0.1004	-0.1509	-0.0236
$\mu_2$	location	N	0.00	2.00	0.0431	0.0378	0.0396	0.0162	0.0481
$\omega_1$	scale	I-G	0.05	0.10	0.1465	0.1414	0.1360	0.1057	0.1957
$\omega_2$	scale	I-G	0.05	0.10	0.0520	0.0486	0.0488	0.0384	0.0573
$\alpha_1$	shape	N	0.00	2.00	1.5655	1.4306	1.3976	0.1455	2.7200
$\alpha_2$	shape	N	0.00	2.00	-1.4640	-1.2065	-1.2571	-1.8731	-0.2209
$q_{11}$	prob.	D	0.95	0.05	0.9449	0.9222	0.9256	0.8668	0.9667
$q_{22}$	prob.	D	0.95	0.05	0.9921	0.9873	0.9881	0.9765	0.9955
$\phi$	persistence	N	0.00	5.00	0.0095	0.0135	0.0126	-0.0413	0.0680

*Note:* Sample period: 1926.M07—2019.M04. N stands for Normal, D for Dirichlet, and Inv-G for Inverted-Gamma distributions. The 5 percent and 95 percent demarcate the bounds of the 90 percent probability interval. Para(1) and Para(2) correspond to the means and standard deviations.

The first finding that is evident is the remarkable difference in the estimated parameters across the two states. Regarding the shape parameter, the first state gives a positive value

(1.5655 at the mode), while its value in the second state is negative ( $-1.4640$  at the mode). The probability intervals for  $\alpha_1$  and  $\alpha_2$  lie exclusively within positive and negative regions, respectively. This reinforces our estimates, and reveals strong evidence of time variation in the skewness of U.S. excess stock returns. Regarding other regime-switching parameters, both location and scale parameters are subject to important shifts across regimes. The location parameter turns out to be lower in the first regime than in the second one, where  $\mu_1$  at the mode is robustly negative ( $-0.1094$  at the mode), and  $\mu_2$  is positive ( $0.0431$  at the mode). Regarding the scale parameters, the estimates for  $\omega_1$  and  $\omega_2$  are  $0.1465$  and  $0.0488$ , respectively, with their corresponding error bands appearing relatively tight. Thus, volatility turns out to be three times higher in the first regime. To sum up, the first (second) regime is characterized by low (high) first moment, high (low) second moment, and positive (negative) third moment.

Regarding the persistence parameter,  $\phi$ , its estimates gives a value close to zero, and its probability intervals lies within both the negative and positive regions, suggesting that  $\phi$  could be dropped from the model.

Regarding the posterior probabilities ( $q_{11}$  and  $q_{22}$ ) of the Markov-switching process, it is apparent that the persistence of staying in each state is relatively high. The 90% probability intervals for  $q_{11}$  are  $0.8668$  and  $0.9667$ , and those for  $q_{22}$  are  $0.9765$  and  $0.9955$ , indicating that the first regime is much less persistent than the second regime. Once again, posterior modes, means and medians are concentrated in tight ranges, reinforcing estimated transition parameters.

Figure 3 displays marginal posterior density estimates for parameters of the model using normal kernel density estimates. Several comments can be made from viewing these plots. First, the distributions for  $\mu_2$  and  $\alpha_2$  parameters are bimodals. Their MCMC draws tend to occasionally visit values close to 0. Such a behavior results directly from the integrated effect of the non-conjugate joint posterior distribution of all parameters that have multiple peaks. Second, the distributions for  $\alpha_1$  and  $\alpha_2$  are almost entirely displayed in positive and negative regions, respectively, meaning that the differences between skewness parameters are apparent.

Figure 4 reports the probabilities — evaluated at the mode — of being in Regime 1 over time. The probabilities are smoothed in the sense of Kim (1994); i.e., full sample information

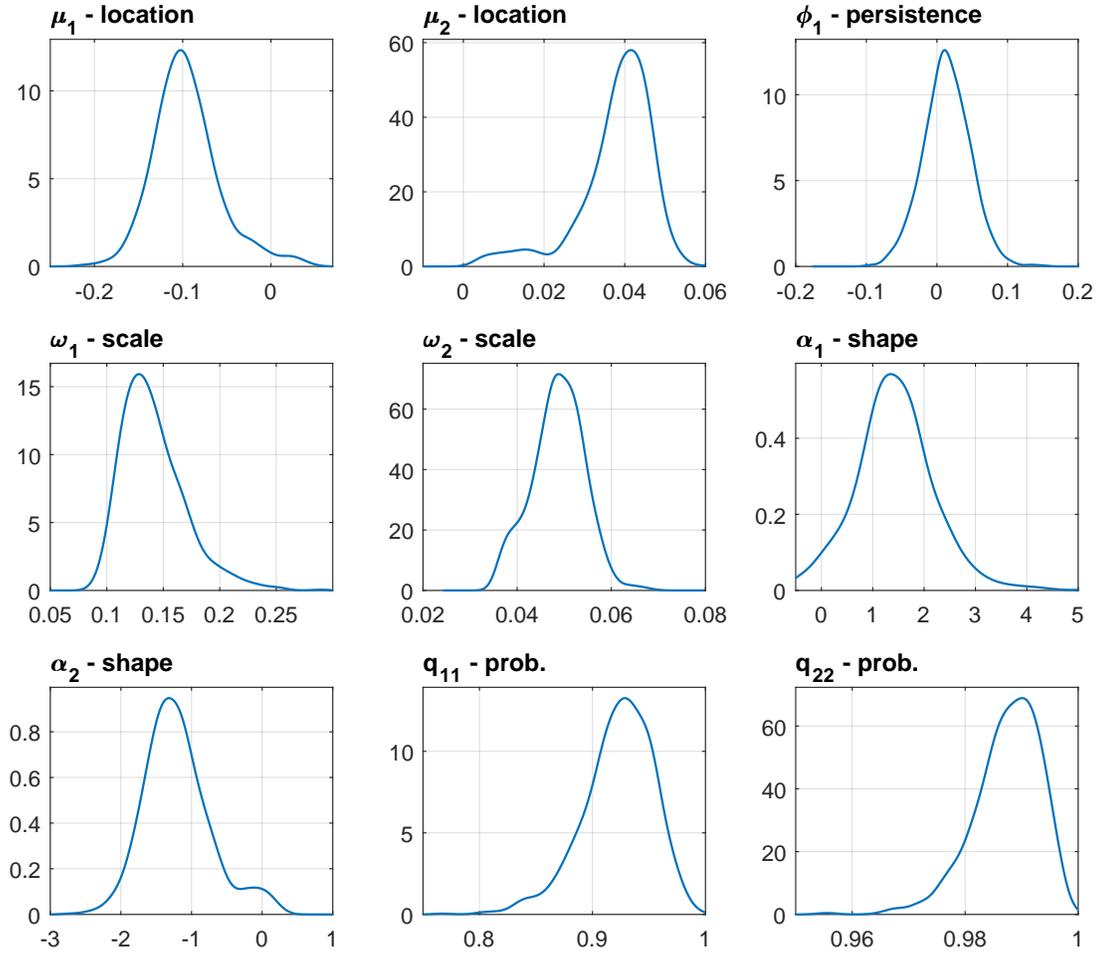


FIGURE 3. Marginal posterior densities using normal kernel density estimates from skewed AR(1) model with Markov shifts.

is used in getting the regime probabilities at each date. One can see from the figure that U.S. economy has been characterized by switches between the two regimes over time. The first regime coincides remarkably well with periods of financial distress such as the Great Depression of the late 1920s and 1930s, the 1973-74 stock market crash, 1987’s Black Monday market crash, and the 2007-2009 Great Recession. Interestingly, our results seem to reveal that the U.S. stock return is skewed to the right in periods of financial distress, and skewed to the right during the remaining periods. This is somewhat similar to Alles and Kling (1994), we report that the skewness of stock indices is more negative during economic upturns and less negative, even positive, during downturns. In other words, extreme values on the right side of the mean are more likely than the extreme values of the same magnitude on the left

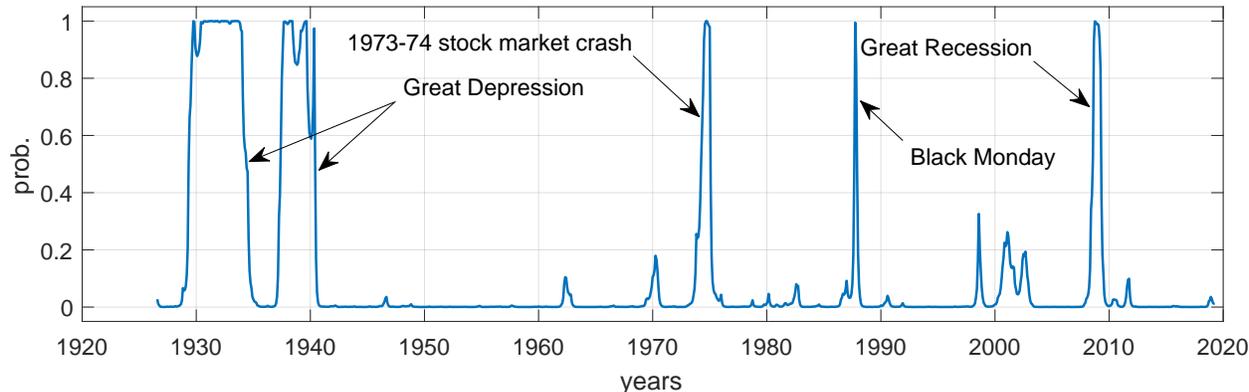


FIGURE 4. Sample period: 1926.M08 — 2019.M04. Smoothed probabilities of Regime 1.

side of the mean, due to the positive degree of skewness. Those results might be, at first sight, quite surprising, but can be largely understood through the existing information-based theories. See Alles (2004) for a comprehensive explanation of such patterns.

TABLE 2. Information criteria

Markov-switching				
model	$m$	log-likelihood	AIC	BIC
Skew-normal				
shocks	9	1863.0	-3708.1	-3662.9
Normal shocks	7	1854.2	-3694.3	-3659.4

*Note:* Akaike information criteria (AIC):  $AIC = -2 * (\log\text{-likelihood} - m)$ . Bayesian information criteria (BIC):  $BIC = -2 * \{\log\text{-likelihood} - 0.5m\log(T)\}$ . The number of parameters and the size of sample are denoted by  $m$  and  $T$ , respectively.

For comparison purposes, we also fit a Markov-switching AR(1) model with normal shocks ( $\alpha_1 = \alpha_2 = 0$ ). The log-likelihood at the peak and two information-based criteria, AIC (Akaike (1973)) and BIC (Schwarz (1978)), are shown in Table 2. Clearly, our Markov-switching model with skew-normal shocks outperforms the one with normal shocks, since it has the largest log-likelihood, and the smallest AIC and BIC.

## VI. CONCLUSION

Our main goal in this paper was to develop a MCMC procedure for skewed autoregressive models subject to Markov shifts. We use the stochastic representation of the skew-normal family to obtain closed-form full conditional posterior distributions, whose sampling can be efficiently conducted within a Gibbs sampling scheme. An application of this procedure to U.S. excess stock returns demonstrates evidence of time-varying skewness.

Extending univariate AR models with Markov skewness shifts to a multivariate framework, like vector autoregression, would seem to be a natural next step. Another area of future work would be to relax the assumption of exogeneity of regime switching in order to better understand the sources of changes in the skewness of a time series. As such, the works by Kim, Piger, and Startz (2008) and Chang, Choi, and Park (2017) on endogenous Markov-switching AR models could then be used in this direction. All in all, we believe those extensions certainly represent an interesting avenue for future research and would be suited to a variety of economic problems.

## REFERENCES

- AKAIKE, H. (1973): *Information Theory and an Extension of the Maximum Likelihood Principle*pp. 199–213. Springer New York, New York, NY.
- ALBERT, J. H., AND S. CHIB (1993): “Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts,” *Journal of Business & Economic Statistics*, 11(1), 1–15.
- ALLES, L. (2004): “Time-Varying Skewness in Stock Returns: An Information-Based Explanation,” *Quarterly Journal of Business and Economics*, 43(1/2), 45–55.
- ALLES, L. A., AND J. L. KLING (1994): “Regularities In The Variation Of Skewness In Asset Returns,” *Journal of Financial Research*, 17(3), 427–438.
- ARELLANO-VALLE, R. B., AND A. AZZALINI (2006): “On the Unification of Families of Skew-Normal Distributions,” *Scandinavian Journal of Statistics*, 33(3), 561–574.
- AZZALINI, A. (1985): “A Class of Distributions Which Includes the Normal Ones,” *Scandinavian Journal of Statistics*, 12(2), 171–178.
- (1986): “Further Results on a Class of Distributions Which Includes the Normal Ones,” *Statistica*, 46, 199–208.
- BAKSHI, G., P. CARR, AND L. WU (2008): “Stochastic risk premiums, stochastic skewness in currency options, and stochastic discount factors in international economies,” *Journal of Financial Economics*, 87(1), 132 – 156.
- CANALE, A., E. PAGUI, AND B. SCARPA (2016): “Bayesian Modeling of University First-year Students’ Grades after Placement Test,” *Journal of Applied Statistics*, 43(16), 3015–3029.
- CARR, P., AND L. WU (2007): “Stochastic skew in currency options,” *Journal of Financial Economics*, 86(1), 213 – 247.
- CARTER, C. K., AND R. KOHN (1994): “On Gibbs Sampling for State Space Models,” *Biometrika*, 81(3), 541–553.
- CHANG, Y., Y. CHOI, AND J. Y. PARK (2017): “A new approach to model regime switching,” *Journal of Econometrics*, 196(1), 127 – 143.
- CHRISTOFFERSEN, P., S. HESTON, AND K. JACOBS (2006): “Option valuation with conditional skewness,” *Journal of Econometrics*, 131(1-2), 253–284.

- ENGEL, C., AND J. D. HAMILTON (1990): “Long Swings in the Dollar: Are They in the Data and Do Markets Know It?,” *American Economic Review*, 80(4), 689–713.
- FEUNOU, B., AND R. TÉDONGAP (2012): “A Stochastic Volatility Model With Conditional Skewness,” *Journal of Business & Economic Statistics*, 30(4), 576–591.
- FRÜHWIRTH-SCHNATTER, S., AND S. PYNE (2010): “Bayesian Inference for Finite Mixtures of Univariate and Multivariate Skew-Normal and Skew-t Distributions,” *Biostatistics*, 11, 317–336.
- GARCIA, R., AND P. PERRON (1996): “An Analysis of the Real Interest Rate under Regime Shifts,” *The Review of Economics and Statistics*, 78(1), 111–125.
- HAMILTON, J. D. (1989): “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica*, 57, 357–384.
- HARVEY, C. R., AND A. SIDDIQUE (1999): “Autoregressive Conditional Skewness,” *The Journal of Financial and Quantitative Analysis*, 34(4), 465–487.
- HENZE, N. (1986): “A Probabilistic Representation of the ‘Skew-Normal’ Distribution,” *Scandinavian Journal of Statistics*, 13(4), 271–275.
- ISERINGHAUSEN, M. (2018): “The Time-Varying Asymmetry Of Exchange Rate Returns: A Stochastic Volatility Stochastic Skewness Model,” Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 18/944, Ghent University, Faculty of Economics and Business Administration.
- JOHNSON, T. C. (2002): “Volatility, Momentum, and Time-Varying Skewness in Foreign Exchange Returns,” *Journal of Business & Economic Statistics*, 20(3), 390–411.
- JONDEAU, E., AND M. ROCKINGER (2003): “Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements,” *Journal of Economic Dynamics and Control*, 27(10), 1699–1737.
- KIM, C.-J. (1994): “Dynamic Linear Models with Markov-switching,” *Journal of Econometrics*, 60, 1–22.
- KIM, C.-J., AND C. R. NELSON (1999): *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, vol. 1 of *MIT Press Books*. 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(2), 263 – 273.

- LHUISSIER, S. (2017): “Financial Intermediaries’ Instability and Euro Area Macroeconomic Dynamics,” *European Economic Review*, 98, 49 – 72.
- LHUISSIER, S., AND F. TRIPIER (2019): “Regime-Dependent Effects of Uncertainty Shocks: A Structural Interpretation,” Working papers 714, Banque de France.
- LIN, T. I., J. C. LEE, S. Y. YEN, AND N. CHUNG (2007): “Finite mixture modelling using the skew normal distribution,” *Statistica Sinica*.
- MINOZZO, M., AND L. FERRACUTI (2012): “On the Existence of some Skew-normal Stationary Processes,” *Chilean Journal of Statistics*, 3(2), 157–170.
- NAKAJIMA, J. (2013): “Stochastic volatility model with regime-switching skewness in heavy-tailed errors for exchange rate returns,” *Studies in Nonlinear Dynamics & Econometrics*, 17(5), 499–520.
- PAGAN, A. R., AND G. W. SCHWERT (1990): “Alternative models for conditional stock volatility,” *Journal of Econometrics*, 45(1-2), 267–290.
- SCHWARZ, G. (1978): “Estimating the dimension of a model,” *The Annals of Statistics*, 6, 461–464.
- SIMS, C. A., D. F. WAGGONER, AND T. ZHA (2008): “Methods for Inference in Large Multiple-equation Markov-switching Models,” *Journal of Econometrics*, 146, 255–274.
- TURNER, C. M., R. STARTZ, AND C. R. NELSON (1989): “A Markov model of heteroskedasticity, risk, and learning in the stock market,” *Journal of Financial Economics*, 25(1), 3–22.

APPENDIX A. THE LIKELIHOOD,  $p(Y_T|\theta)$

The evaluation of the overall likelihood function is obtained using the standard Hamilton (1989) filter. The likelihood of  $Y_T$  is

$$p(Y_T|\theta) = \prod_{t=1}^T p(y_t|Y_{t-1}, \theta), \quad (44)$$

where the conditional likelihood function  $p(y_t|Y_{t-1}, \theta)$ , given  $\theta$ , at date  $t$  is obtained by integrating the density  $p(y_t, s_t|Y_{t-1}, \theta)$  over  $s_t$  as follows

$$p(y_t|Y_{t-1}, \theta) = \sum_{s_t \in H} p(y_t, s_t|Y_{t-1}, \theta), \quad (45)$$

$$= \sum_{s_t \in H} p(y_t|s_t, Y_{t-1}, \theta) \Pr[s_t|Y_{t-1}, \theta], \quad (46)$$

Using the Hamilton (1989) filter, we can recursively compute  $\Pr[s_t|Y_t, \theta]$  forward. Specifically,

$$\Pr[s_t|Y_{t-1}, \theta] = \sum_{s_{t-1} \in H} q_{s_t, s_{t-1}} \Pr(s_{t-1}|Y_{t-1}, \theta), \quad \text{for } t > 0, \quad (47)$$

where  $q_{s_t, s_{t-1}} = \Pr[s_t|s_{t-1}]$  is the transition probability described in (13).

We then update the joint probability term in the following way:

$$\Pr[s_t|Y_t, \theta] = \frac{p(y_t, s_t|Y_{t-1}, \theta)}{p(y_t|Y_{t-1}, \theta)} \quad (48)$$

$$= \frac{p(y_t|s_t, Y_{t-1}, \theta) \cdot \Pr(s_t|Y_{t-1}, \theta)}{p(y_t|Y_{t-1}, \theta)}, \quad \text{for } t > 0, \quad (49)$$

Once the parameters of the model are estimated, we follow Kim (1994) and Kim and Nelson (1999) by making inference on  $s_T$ , the smoothed probabilities, in the following way:

$$\Pr[s_t|Y_T, \theta] = \sum_{s_{t+1} \in H} \Pr[s_t, s_{t+1}|Y_T, \theta], \quad (50)$$

where

$$\Pr[s_t, s_{t+1}|Y_T, \theta] = \frac{\Pr[s_{t+1}|Y_T, \theta] \cdot \Pr[s_t|Y_T, \theta] \cdot \Pr[s_{t+1}|s_t]}{\Pr[s_{t+1}|Y_T, \theta]}. \quad (51)$$

The advantage of such a method is that it allows us to infer the unobservable variable  $s_t$  using all the information in the sample.

## APPENDIX B. COMPUTER SOFTWARE

Once compiled, our companion C++ computer code for this paper, available at the author's website, is easy to use. One must provide an input file that indicates prior specifications, the structure of AR process, MCMC options, and time series data. An example of this interface is provided below.

```
1 //== Number Lagged Variables ==//
2 1
3
4 //== Size Sample ==//
5 267
6
7 //== Number Regimes ==//
8 2
9
10 //== Prior for Phi ==//
11 0.0000 5.0000
12
13 //== Prior for Scale ==//
14 1.1782 2.5891
15
16 //== Prior for Shape ==//
17 0.0000 3.0000
18
19 //== Prior for Transition Matrix ==//
20 12.00 3.000
21 3.000 12.00
22
23 //== Number Optimization Runs ==//
24 100
25
26 //== Number Draws ==//
27 11000
28
```

```

29 //== Number Burn-in ==//
30 1000
31
32 //== Thinning Factor ==//
33 10
34
35 //== Y-data ==//
36 2.1999843197342273e-01
37 1.0628799925268773e+00
38 ...

```

This input file concerns an AR model of order 1 where the shape parameter follows a two-states Markov process. Regarding the MCMC procedure, the input file asks for 11,000 draws, whose 1,000 as burn-in, and 10 as thinning factor. In this file, the header bracketed by

```

1 //== ... ==//

```

communicates with the software what kind of data is expected. The number below the header “Number Lagged Variables” indicates how many lags are defined for the model. The number below the header “Size Sample” indicates the size of the sample. The number below the header “Number Regimes” indicates the number of regimes for the Markov process. The numbers below the headers “Prior for Phi”, “Prior for Scale”, “Prior for Shape”, and “Prior for Transition Matrix” indicate the hyperparameters for the parameters  $\phi$ ,  $\omega$ ,  $\alpha$ , and  $Q$ , respectively. The number below the header “Number Optimization Runs” indicates the number of times the optimization process is repeated. At each time, a new set of points, generated from the prior, is used. The number below the header “Number Draws” indicates the number of draws in the MCMC algorithm. The number below the header “Number Burn-in” indicates the number of burn-in in the MCMC algorithm. The number below the header “Thinning Factor” indicates the thinning factor in the MCMC algorithm. Finally, the values below the header “Y-data” contain time series data.