Asymmetry, Fat-tail, and Autoregressive Conditional Density in Financial Return Data with Systems of Frequency Curves

Jun Yan

Department of Statistics and Actuarial Science, University of Iowa,

241 Schaeffer Hall, Iowa City, IA 52242, USA

July 7, 2005

Abstract

Asymmetry and fat-tail are both stylized facts of financial return data. Many asymmetric and fat-tailed distributions have been used to model the innovation in autoregressive conditional heteroskedasticity (ARCH) models. This article introduces two more distributions from systems of frequency curves into the ARCH context: Pearson's Type IV and Johnson's SU. Both distributions have two shape parameters and allow a wide range of skewness and kurtosis. We then impose dynamics on both shape parameters to obtain autoregressive conditional density (ARCD) models, allowing time-varying skewness and kurtosis. The quasi-maximum likelihood estimates (QMLE) of volatility parameters obtained from these distributions are found to have high efficiency in a simulation study when the true distribution is asymmetric and fat-tailed. ARCD models with these distributions are applied to the daily Standard & Poor 500 index return data. Models with time-varying shape parameters are found to give better fit than models with constant shape parameters.

KEY WORDS: ARCH; Asymmetry; Fat-tail; Johnson's transformation; Pearson system; System of Distributions

1. INTRODUCTION

Important stylized facts about financial return data have been discovered through varieties of financial applications. Three of the stylized facts relevant to this article are volatility clustering, asymmetry, and fat-tail. Volatility clustering can be explicitly modeled by a class of autoregressive conditional heteroskedasticity (ARCH) models (Engle 1982; Bollerslev 1986). There is a voluminous literature on conditional heteroskedasticity (CH) models following Engle's seminal work; see for example Bollerslev, Engle, and Nelson (1994) for a survey. Consider a time series ϵ_t, $t = 1, \ldots, T$. A simple normal-GARCH(1,1) model (Bollerslev 1986) has the form

$$\epsilon_t|\psi_{t-1} = \sqrt{h_t} z_t, \tag{1}$$

$$z_t \sim N(0,1), \tag{2}$$

$$h_t = b_0 + b_1 \epsilon_{t-1}^2 + b_2 h_{t-1}, \tag{3}$$

where ψ_{t-1} is the information set up to time $t-1$. This model is the most widely used CH model, and the consistency of the resulting quasi-maximum-likelihood estimator (QMLE) under misspecification (Bollerslev and Wooldridge 1992) makes it even popular. However, this model does not allow asymmetry and is not sufficiently fat-tailed to capture the excess kurtosis found in most financial return data. This has led to a search for more flexible conditional distribution to replace the conditional normal assumption in (2).

A desirable conditional distribution in the ARCH context should offer a wide range of shapes to approximate the truth and at the same time remain easy to handle computationally. Many distributions have been tried in the literature to to account for asymmetry or fat-tail or both. An incomplete alphabetical list of these distribution is: generalized error distribution (GED) (Nelson 1991), generalized hyperbolic distribution (GHD) (Eberlein and Keller 1995; Barndorff-Nielsen 1997), Gram-Charlier (GC) distribution (Jondeau and Rockinger 2001), noncentral t distribution (Harvey and Siddique 1999), stable distribution (McCulloch 1996), and t distribution (Bollerslev 1987; Hansen 1994). All these distributions can give thicker tails than the normal distribution. For symmetric distributions in the list, skewed versions can be constructed by perturbing the symmetric versions (Fernández and Steel 1998; Azzalini and Capitanio 2003; Jones and Faddy 2003). These distributions may be adequate for some cases but in general, there are various issues in applying them. The tail of a GED is not sufficiently thick to account for extremal events. The computation of a GHD can be challenging as its density involves a Bessel function of the third kind and there are three, instead of two, shape parameters. A GC distribution can only offer a small range of skewness and kurtosis and the nonlinear constraint on the shape parameters is hard to impose in estimation; see Figure 1 and discussions in Section 2.1. The density of a noncentral t distribution involves the sum of an infinite series and can be hard to handle. A stable distribution has the justification from the generalized central limiting theorem, but the its variance does not exist when the tail is thicker than normal. A skewed t distribution seems to be the best choice in both flexibility and computational ease. It has been successfully applied in modeling financial returns data (Hansen 1994; Jondeau and Rockinger 2003). The success of skewed t distributions motives one to reawaken the literature of systems of frequency curves to give flexible shapes for asymmetry and fat-tail. In fact, it will be seen that one of the distributions to be introduced, Pearson's Type IV, can also be viewed as a skewed t distribution.

A system of frequency curves contains distributions that have widely varying shapes (see for example Stuart and Ord 1994,Chapter 6). Most of the work was done between 1890 and early 1900s and during 1940s. Among the various systems in the literature, Pearson's system and Johnson's systems offers the maximum range of skewness and kurtosis. That is, for any given legitimate pair of skewness and kurtosis, one can find a distribution in these systems that gives the specified skewness and kurtosis. Pearson's system seeks to ascertain a family of distributions with a small number of parameters to represent observed data satisfactorily. Johnson's system seeks a transformation such that the transformed variate is at least approximately normal. Within these systems, Pearson's Type IV distribution and Johnson's SU distribution are ideal candidate to approximate the unknown true distributions of the ARCH innovations; see details in Section 2. Pearson's Type IV was first introduced into the ARCH context by Premaratne and Bera (2001) for modeling asymmetry and fat-tail. The contribution of this article differs from Premaratne and Bera (2001) in two aspects: 1) Instead of using numerical integration as in Premaratne and Bera (2001), we use an accurate implementation to compute the closed-form expression of the normalizing constant, which has probably been the main obstacle for the wide spread of Pearson's Type IV; and 2) We use a standardized version of Pearson's Type IV with mean zero and variance one to replace the normal assumption in (2). Extra emphasis will be put on the computation of normalizing constant when Pearson's Type IV distribution is introduced in Section 2.1. Johnson's SU distribution can offer a slightly larger range of skewness-kurtosis combinations, and as a result, its tail is not as thick as Pearson's Type IV; see more detail in Section 2 and Figure 1. It has not been applied in the ARCH context yet. But, as will be demonstrated, it can be used as a simpler alternative to Pearson's Type IV distribution when asymmetry and fat-tail are present. A Johnson's SU variate has the interpretation of "transform from normal". Therefore, it has the advantage that the density, distribution, quantile, and random number generator are easily obtained from transformation of a standard normal.

Both Pearson's Type IV and Johnson's SU have two shape parameters and offer a wide range of skewness and kurtosis that may be encountered in financial returns data. These parameters can be made time-varying in the same fashion in the line of Hansen (1994) to form autoregressive conditional density (ARCD) models. An ARCD model can be very useful when skewness and kurtosis are desired to be time-varying, for example, in asset pricing (Harvey and Siddique 2000). It has a clear computational advantage over autoregressive conditional moments (ARCM) model, where dynamics are imposed directly on skewness or kurtosis and the shape parameters are backed out from the skewness or kurtosis (Harvey and Siddique 1999; Brooks, Burke, and Persand 2002). Furthermore, unlike higher moments such as skewness and kurtosis, the variation in the shape parameters may be smaller and easier to manage numerically. Therefore, we adopt the ARCD approach in this article.

As volatility is more important than higher moments in the ARCH context, it is of practical interest to investigate the performance of the QMLE of volatility parameter using various assumed distributions under various true distributions. With a different parameterization of h_t than (3), Newey and Steigerwald (1997) gives general conditions under which the QMLE is consistent. However, there has not been much research on this aspect using the most widely used ARCH specification in (3) when asymmetry and fat-tail are allowed through distributions with flexible shapes. We demonstrate through a simulation study that both Pearson's Type IV and Johnson's SU can give more efficient QMLE of volatility parameters than obtained under the normal assumption. Johnson's SU can give QMLE of volatility parameters with little efficiency loss when the tail of the true distribution is thicker. On the other hand, Pearson's Type IV gives QMLE of volatility parameter with some efficiency loss when the true distribution does not have as thick tails. This efficiency comparison result, in combination with concerns of computational ease, makes Johnson's

SU distribution a very useful tool in financial return data analysis.

The rest of this article is organized as follows. In Section 2, Pearson's Type IV distribution and Johnson's SU distribution are briefly reviewed and their properties discussed in the relevance of modeling financial returns. These families are then used as the innovation distribution in a GARCH type specification with time-varying shape parameters in Section 3. A simulation study is conducted in Section 4 to examine the performance of the QMLE under correct- and misspecification of the innovation distribution. The return of the Standard & Poor (S&P) 500 daily index is used to illustrate the method in Section 5. Discussions conclude in Section 6.

2. DISTRIBUTIONS FROM SYSTEMS OF FREQUENCY CURVES

2.1 Pearson's Type IV Distribution

Back in the 1890s, Pearson (1895) introduced the now called Pearson's system of frequency curves to give a wide range of distributions to fit data that can not be well fitted by the normal distribution. The density functions in the system are defined through a differential equation which leads to various families under different conditions (see for example Johnson, Kotz, and Balakrishnan 1994, Chapter 12). One of the members in the system, Pearson's Type IV distribution, has density of the form

$$f(x;\nu,m,\xi,\lambda) = k\left[1+\left(\frac{x-\xi}{\lambda}\right)^2\right]^{-m}\exp\left[-\nu\tan^{-1}\left(\frac{x-\xi}{\lambda}\right)\right], \qquad (4)$$

where $m > 1/2$, ν, ξ, and $\lambda > 0$ are parameters, $x \in R$, and k is the normalizing constant depending on ν, m, and λ. Clearly, ξ and λ are location and scale parameters, respectively. Parameter ν can be interpreted as a skewness parameter. When $\nu = 0$, the distribution becomes symmetric. The distribution is negatively (or positively) skewed when $\nu > 0$ (or $\nu < 0$). Parameter m controls the tail thickness and can be interpreted as a kurtosis parameter. Increasing m decreases the kurtosis. The distribution reduces to a t distribution with degrees of freedom n when $\nu = 0$, $m = (n+1)/2$ and $\delta = \sqrt{n}$. The normal density is then obtained as $m \to \infty$. With nonzero ν, Pearson's Type IV distribution can be viewed as a skewed t distribution. It should be noted that ν and m are not, respectively, purely the skewness parameter and kurtosis parameter; see the moments expressions below.

The moments of Pearson's Type IV distribution can be found in classic references (e.g. Stuart and Ord 1994). The mean μ, variance σ^2, skewness s, and kurtosis κ are

$$\mu = \xi - \frac{\lambda\nu}{r} \qquad (m > 1), \tag{5}$$

$$\sigma^2 = \frac{\lambda^2}{r^2(r-1)}(r^2+\nu^2) \qquad (m > 3/2), \tag{6}$$

$$s = \frac{-4\nu}{r-2}\sqrt{\frac{r-1}{r^2+\nu^2}} \qquad (m > 2), \tag{7}$$

$$\kappa = \frac{3(r-1)[(r+6)(r^2+\nu^2)-8r^2]}{(r-2)(r-3)(r^2+\nu^2)} \qquad (m > 5/2), \tag{8}$$

where $r = 2(m-1)$. The jth moment exists only when $r+1 > j$, a resemblance to t distributions. The skewness and kurtosis are jointly determined by the two shape parameters ν and m. Let $\beta_1 = s^2$ and $\beta_2 = \kappa$. The range of β_1 and β_2 is often used in comparing the capability of modeling asymmetry and fat-tail across different distributions. Figure 1 shows the feasible range of (β_1, β_2) on the β_1-β_2 plane for four distributions, all having mean zero and variance one. Pearson's Type IV offers a wide range below the dotted line. In contrast,, the range of the Gram-Charlier density (Jondeau and Rockinger 2001) shown below the dot-dashed line, is rather limited.

[Figure 1 about here.]

The main obstacle of applying Pearson's Type IV distribution has been the evaluation of the normalizing constant k. Premaratne and Bera (2001) used numerical integration to obtain it in each evaluation of the likelihood function. Actually, the normalizing constant has a closed-form expression which involves the gamma function with a complex-valued argument (Pearson 1895; Nagahara 1999)

$$k(\nu, m, \lambda) = \frac{\Gamma(m)}{\sqrt{\pi}\Gamma(m-1/2)}\left|\frac{\Gamma(m+i\nu/2)}{\Gamma(m)}\right|^2 \tag{9}$$

where $i = \sqrt{-1}$ and $|\cdot|$ is the module of a complex number. A single precision complex gamma function is available from `www.netlib.org`. The computation in this article, however, uses the `C` code of Heinrich (2004) to compute the squared module in (9) directly with highest machine allowable precision. In particular,

$$\left|\frac{\Gamma(x+iy/2)}{\Gamma(x)}\right|^2 = \frac{1}{F(-iy, iy; x; 1)}, \tag{10}$$

where F is the hypergeometric function, sometimes written as ${}_2F_1$. Heinrich (2004) utlized two equations to compute $F(-iy, iy; x; 1)$. The first equation is a series expansion

$$F(-iy, iy; x; 1) = 1 + \frac{y^2}{x1!} + \frac{y^2(y^2+1^2)}{x(x+1)2!} + \frac{y^2(y^2+1^2)(y^2+2^2)}{x(x+1)(x+2)3!} + \cdots, \tag{11}$$

which is absolutely convergent and converges rapidly only when $x \gg 1$. The second equation is a recursion

$$\left|\frac{\Gamma(x+iy/2)}{\Gamma(x)}\right|^2 = \left[1 + \left(\frac{y}{x}\right)\right]^{-1} \left|\frac{\Gamma(x+1+iy/2)}{\Gamma(x+1)}\right|^2. \tag{12}$$

For large x, Heinrich (2004) computes $F(-iy, iy; x; 1)$ with the series (11). For small x, Heinrich (2004) calculates $F(-iy, iy; x+n; 1)$ via the series (11) for some n chosen to be sufficiently large and work down to $n = 0$ using the recursion (12). This algorithm gives very accurate result efficiently. Heinrich (2004) also gives code to generate random numbers from Pearson's Type IV distribution, implementing an exercise in Devroye (1986). This implementation is used in the simulation study.

2.2 Johnson's SU Distribution

Johnson's SU distribution is one of the three systems that Johnson (1949) introduced using transformations of the standard normal variate. Let Z be a $N(0,1)$ variable. The SU transformation is defined with

$$Z = \gamma + \delta \sinh^{-1}\left(\frac{X-\xi}{\lambda}\right), \tag{13}$$

where $\sinh^{-1}$ is the inverse hyperbolic sine function, and ξ, $\lambda > 0$, γ, $\delta > 0$ are parameters. The density function of Johnson's SU distribution can be easily found in closed-form from variable transformation:

$$f(x; \gamma, \delta, \xi, \lambda) = \frac{\delta}{\lambda\sqrt{1 + \left(\frac{x-\xi}{\lambda}\right)^2}} \phi\left[\gamma + \delta \sinh^{-1}\left(\frac{x-\xi}{\lambda}\right)\right], \tag{14}$$

where $x \in R$, ϕ is the density function of $N(0,1)$, ξ and $\lambda > 0$ are location and scale parameters, respectively, γ can be interpreted as a skewness parameter, and $\delta > 0$ can be interpreted as a kurtosis parameter. The distribution is positively or negatively skewed according as γ is negative or positive. Holding γ, increasing δ reduces the kurtosis. However, similar to the case of Pearson's Type IV distribution, γ and δ can not be viewed as purely skewness and kurtosis parameters, respectively.

We give the first four moments of Johnson's SU distribution. The mean and variance are

$$\mu = \xi + \lambda\omega^{1/2}\sinh\Omega, \tag{15}$$

$$\sigma^2 = \frac{\lambda^2}{2}(\omega - 1)(\omega\cosh 2\Omega + 1), \tag{16}$$

where $\omega = \exp(\delta^{-2})$ and $\Omega = \gamma/\delta$. Since there is not much simplification in the expressions for skewness and kurtosis, we give the third and fourth central moments μ_3 and μ_4, respectively,

$$\mu_3 = -\frac{1}{4}\omega^2(\omega^2 - 1)^2[\omega^2(\omega^2 + 2)\sinh 3\Omega + 3\sinh\Omega], \tag{17}$$

$$\mu_4 = \frac{1}{8}(\omega^2 - 1)^2[\omega^4(\omega^8 + 2\omega^6 + 3\omega^4 - 3)\cosh 4\Omega \tag{18}$$

$$+ 4\omega^4(\omega^2 + 2)\cosh 2\Omega + 3(2\omega^2 + 1)]. \tag{19}$$

The regions of β_1 and β_2 offered by Johnson's SU distribution is the area below the dashed line illustrated in Figure 1 One can conclude that Johnson's SU distribution has lighter tail than Pearson's Type IV, and as a result, it has a slightly wider region in the β_1-β_2 plane.

As Johnson's SU distribution is obtained from a transformation of a standard normal variate, its computation is a lot easier relative to Pearson's Type IV distribution. The distribution function, quantile function, and random number generator are all straightforwardly available from transforming those of the standard normal. The gradient and Hessian of the log-likelihood are analytically available, which is an advantage relative to Pearson's Type IV distribution.

2.3 Remarks

Both Pearson's Type IV and Johnson's SU families are location-scale distributions. The standard versions are obtained when $\xi = 0$ and $\lambda = 1$. Let μ_s and σ_s^2 be the mean and variance of the standard version a distribution. By constraining

$$\lambda = 1/\sqrt{\sigma_s^2}, \tag{20}$$

$$\xi = -\lambda\mu_s, \tag{21}$$

we obtain standardized versions of these families with mean zero, variance one, and two shape parameters. These standardized versions can be applied in an ARCH context; see discussions in Section 3.

[Figure 2 about here.]

In Figure 2, we present the log-densities of four standardized distributions: normal, hyperbolic, Johnson's SU, and Pearson's Type IV. The hyperbolic distribution is a sub-family in the GH distribution with one of the three shape parameters fixed at 1 (Barndorff-Nielsen and Stelzer 2004) and hence can be standardized similarly using the other two shape parameters. All four distributions in the plot have mean zero and variance one. All but normal have skewness -0.5 and kurtosis 6. It is clear from the plot that the log-density of normal is a parabola while that of hyperbolic is a hyperbola. In the range $(-5, 5)$, the three fat-tailed distributions have very close log-densities. However, when we zoom out to the range $(-20, 20)$, we observe that the tail thickness are in the order: hyperbolic, Johnson's SU, and Pearson's Type IV.

Compared to Pearson's Type IV distribution, Johnson's SU distribution has several advantages: its density is available in easy-to-evaluate closed-form; its skewness and kurtosis has a slightly wider range; and it provides more efficient estimate for volatility parameters when the true unknown density has tails not as thick; see details in Section 4.

3. AUTOREGRESSIVE CONDITIONAL DENSITY

Suppose that the observed time series is y_i, $i = 1, \ldots, T$. A general model of GARCH(p,q) is

$$y_t|\psi_{t-1} = \mu_t(a) + \epsilon_t, \tag{22}$$

$$\epsilon_t|\psi_{t-1} = \sqrt{h_t} z_t, \tag{23}$$

$$h_t = b_0 + \sum_{i=1}^{q} b_q \epsilon_{t-q}^2 + \sum_{i=1}^{p} b_{q+p} h_{t-p}, \tag{24}$$

$$E(z_t) = 0, \qquad \text{Var}(z_t) = 1, \tag{25}$$

where $\mu_t(a)$ is the conditional mean with parameter vector a, z_t are independent and identically distributed (iid) with density $f(\cdot;\eta)$, and η is a vector of shape parameters. A standardized version of Pearson's Type IV (4) or Johnson's SU (14) can be used as $f(\cdot;\eta)$ to model asymmetry and fat-tail. The mean component $\mu_t(a)$ can, for example, be specified by an ARMA formulation.

More flexibility can be introduced by allowing asymmetry or fat-tail or both to be time-varying. There are two approaches to impose dynamics on skewness or kurtosis. The first approach is the

ARCM, where dynamics are imposed directly on skewness or kurtosis and the shape parameters are backed out from the skewness or kurtosis (Harvey and Siddique 1999; Brooks et al. 2002). This approach can be very computing intensive. For example, in the case of noncentral t distribution used by Harvey and Siddique (1999), the shape parameters have to be numerically obtained from skewness or kurtosis. One exception is the Gram-Charlier distribution, which uses the skewness and kurtosis directly as parameters. The second approach is the ARCD, where dynamics are imposed on shape parameters and skewness or kurtosis are derived from the time-varying shape parameters (Hansen 1994; Jondeau and Rockinger 2003). This approach is less computationally demanding than the first approach. The skewness and kurtosis are allowed to explode, which can be very useful in modeling extremal events, even though the shape parameters are stationary (Jondeau and Rockinger 2003). The ARCD approach is adopted in the sequel.

An ARCD model replaces $f(\cdot;\eta)$ with $f(\cdot;\eta_t)$. Suppose the assumed distribution is Johnson's SU distribution where $\eta_t = (\gamma_t, \delta_t)^\top$. As these parameters determines the skewness and kurtosis, it is tempting to include the cubic or quartic of ϵ_{t-1}. However, these terms have high variations which may lead to numerical problems in estimation. Hansen (1994) used quadratic equations while Jondeau and Rockinger (2003) used piecewise linear equations of ϵ_{t-1}. We use the threshold specification similar to Jondeau and Rockinger (2003) for shape parameters $\eta_t = (\gamma_t, \delta_t)^\top$, except that we use the standardized innovation z_{t-1} instead of non-standardized ϵ_{t-1}:

$$\gamma_t = c_0 + c_1 z_{t-1}^+ + c_2 z_{t-1}^- + c_3 \gamma_{t-1}, \tag{26}$$

$$\delta_t = d_0 + d_1 z_{t-1}^+ + d_2 z_{t-1}^- + d_3 \delta_{t-1}, \tag{27}$$

where $z_{t-1}^+ = \max(z_{t-1}, 0)$, $z_{t-1}^- = \min(z_{t-1}, 0)$, and c_i's and d_i's, $i = 0, 1, 2$ and 3, are parameters to be estimated. As noted by Jondeau and Rockinger (2003,p.1709), care should be taken when the effect of z_{t-1}^+ and z_{t-1}^- are insignificant. Consider the model for γ_t to illustrate. When $c_1 = d_2 = 0$, the model reduces to $\gamma_t = c_0 + c_3\gamma_{t-1}$, which can not be distinguished from a constant model $\gamma_t = \gamma$ after γ_t quickly converges to its stationary level $c_0/(1 - c_3)$. A model with $c_3 = 0$ should be fitted first to confirm whether past observations affect γ_t. Therefore, when all c_1 and c_2 are insignificant while c_3 is significant in an estimation, the result is spurious.

When there is constraints on parameters, transformations can be used remove the constraints and the dynamics can then be imposed on the transformed parameters. For example, the shape parameter m in Pearson's Type IV distribution has to be greater than $3/2$ for the variance to exist. One can impose the dynamics on the transformed parameter $g(m_t) = \log(m_t - 3/2)$:

$$g(m_t) = d_0 + d_1 z_{t-1}^{+} + d_2 z_{t-1}^{-} + d_3 g(m_{t-1}). \tag{28}$$

The conditional log-likelihood of the full ARCD model is

$$LLK = \sum_{t=\max(p,q)+1}^{T} \left\{ \log f(z_t; \eta_t) - \frac{1}{2} \log h_t \right\}. \tag{29}$$

When autoregressive terms of shape parameters η_t are used in the ARCD specification, some reasonably arbitrary initial value η_1 can be used to get the iteration started, and its effect is negligible.

4. SIMULATION

It is of practical interest to examine the performance of quasi-maximum likelihood estimates (QMLE) when the unknown distribution is approximated by systems of frequency curves (White 1982). The QMLE under normal specification of z_t in (22) has the advantage of being consistent and asymptotically normal when the mean and variance are correctly specified. Newey and Steigerwald (1997) gave general conditions under which the QMLE is consistent using a different parameterization, . However, the parameterization of (22) is mostly widely used and it is more interesting to study the finite sample performance of the QMLE of the volatility parameters in GARCH model (22) when flexible shapes are allowed in the assumed distribution. The purpose of the simulation in this section is to compare the bias and efficiency of the QMLE from a range of assumed distributions under a range of true distributions.

The simulation is done with assuming the mean component $\mu_t(a) = 0$. Three true distributions of z_t are used: hyperbolic, Johnson's SU, and Pearson's Type IV. The three distributions has different thickness in their tails as illustrated in Figure 2. They all have two shape parameters such that the mean is 0, variance 1, skewness -0.5, and kurtosis 6. Other values of skewness and kurtosis are also tried, with similar results found and hence omitted. The higher moments of the hyperbolic distribution are available from Barndorff-Nielsen and Stelzer (2004). Numerical method is used to search the parameters that yield the specified skewness and kurtosis. These three distributions plus normal are used as assumed distributions. Dataset with 2,000 observations are generated from a GARCH(1,1) model with $b_0 = 0.1$, $b_1 = 0.1$, and $b_2 = 0.8$ under the three true distributions of z_t. Parameter are estimated with all four assumed distributions for each generated dataset. This process is repeated 1,000 times.

The simulation result of volatility parameters are summarized Table 1. As these QMLEs are not necessarily consistent, their square root of the mean squared errors (rmse) are reported instead of standard errors. At sample size 2,000, the average of the QMLEs from all assumed distributions are virtually very close to the true parameter values. The normal-QMLE is clearly at a disadvantage relative to others in terms of the efficiency measured by rmse. When the assumed distribution and the true distribution match, MLE is obtained, which have the highest efficiency as expected. When the true distribution has the lightest tail (hyperbolic), there is efficiency loss with both Johnson-SU-QMLE and Pearson-Type-IV-QMLE, the loss from Johnson's SU is less than from Pearson's Type IV. When the true distribution has medium thick tail (Johnson's SU), there is some efficiency loss with Pearson-Type-IV-QMLE but little loss with hyperbolic-QMLE. When the true distribution has the thickest tail (Pearson Type IV), both hyperbolic-QMLE and Johnson-SU-QMLE are almost as efficient as the MLE. These simulation results suggest that the hyperbolic-QMLE be used in estimating the volatility parameters. However, as discussed in the next paragraph, using hyperbolic distribution leads to poor estimation about the tail behavior. Johnson's SU distribution gives a comprise between efficiency and tail behavior. In combination with the computational ease, Johnson's SU distribution makes a very useful tool in an financial analyst's toolbox.

[Table 1 about here.]

The higher efficiency of the QMLE from lighter-tailed distributions comes at a cost which involves the shape parameters. These QMLEs converge to pseudo-values that minimize the Kullback-Leibler (KL) distance between the assumed distribution and the true distribution (White 1982). Table 2 summarizes the simulation result of the QMLE of the shape parameters. The averages of the QMLE are very close to the pseudo-values. However, the assumed distribution, with the pseudo-values in Table 2 as parameters, can give skewness and kurtosis that are very different from the true skewness and kurtosis, even though they minimizes the KL distance in their family. This is clearly seen in Table 3. The true distributions all have skewness -0.5 and kurtosis 6. When the true distribution is hyperbolic, the Pearson Type IV distribution which minimizes the KL distance has skewness -2.012 and the kurtosis does not exist! When the true distribution is Pearson Type IV, the corresponding Johnson SU distribution has skewness -0.438 and kurtosis 5.174, and the hyperbolic distribution has skewness -0.377 and kurtosis 4.504, respectively. Therefore, when the true distribution has thicker tail than the assumed distribution, although the volatility parameters can be estimated with little efficiency loss, the implied higher moments can be way off the truth.

[Table 2 about here.]

[Table 3 about here.]

5. APPLICATION

In this section, we apply Johnson's SU distribution and Pearson's Type IV distribution to model the daily returns of the Standard & Poor's 500 index (S&P 500). Daily closing prices p_t of the S&P 500 index are obtained from public domain `finance.yahoo.com` with symbol "gspc", which makes it easy to reproduce and compare against the results. The daily prices p_t spans from January 2, 1990 to June 14, 2000. The daily returns are obtained as $y_t = 100 \times \log(p_t/p_{t-1})$ with 2,641 observations. The returns series is plotted in Figure 3. The unconditional skewness and kurtosis -0.342 and 8.164, respectively. indicating that the data are left skewed and highly leptokurtic.

[Figure 3 about here.]

The models considered have three components: mean, variance, and shape. The mean component can generally be modeled by an ARMA specification, but for the S&P 500 index returns, we simply use a constant. The variance component is a GARCH(1,1) model. These two components are summarized as

$$y_t = a_0 + \epsilon_t, \tag{30}$$

$$\epsilon_t|\psi_{t-1} = \sqrt{h_t} z_t, \tag{31}$$

$$h_t = b_0 + b_1\epsilon_{t-1}^2 + b_2 h_{t-1}, \tag{32}$$

$$E(z_t) = 0, \qquad \mathrm{Var}(z_t) = 1. \tag{33}$$

The conditional distribution of the innovation z_t is assumed to be the standardized version of Johnson's SU or Pearson's Type IV. The two shape parameters of the innovation distribution can be time-invariant or time-varying. All the fitted models are nested in a full conditional autoregressive density model. In the case of Johnson's SU, the shape parameters are specified as

$$z_t \sim \text{Johnson's SU}(\cdot; \gamma_t, \delta_t), \tag{34}$$

$$\gamma_t = c_0 + c_1 z_{t-1}^+ + c_2 z_{t-1}^- + c_3 \gamma_{t-1}, \tag{35}$$

$$\delta_t = d_0 + d_1 z_{t-1}^+ + d_2 z_{t-1}^- + d_3 \gamma_{t-1}. \tag{36}$$

In the case of Pearson's Type IV, the specification of shape parameters are replaced with

$$z_t \sim \text{Pearson's Type IV}(\cdot; \nu_t, m_t), \tag{37}$$

$$\nu_t = c_0 + c_1 z_{t-1}^+ + c_2 z_{t-1}^- + c_3 \gamma_{t-1}, \tag{38}$$

$$g(m_t) = d_0 + d_1 z_{t-1}^+ + d_2 z_{t-1}^- + d_3 g(m_{t-1}), \tag{39}$$

where transformation $g(m_t) = \log(m_t - 3/2)$ is used to impose the constraint $m_t > 3/2$ for the existence of the variance. For numerical stability, we have also constrained the variance parameters such that $b_1 > 0$, $b_2 > 0$, and $b_1 + b_2 < 1$.

[Table 4 about here.]

Table 4 summarizes the estimation result for all the models under Johnson's SU distribution. In addition to parameter estimates and their standard errors, the maximized log-likelihood and model selection criteria AIC and BIC are also reported for each model. We first notice that the variance component changes little across all models. The conditional variance is highly persistent as the sum of b_1 and b_2 is very close to 1. The basic model M1 has constant shape parameters with implied skewness -0.185 and kurtosis 5.097, which are reasonably close to the empirical skewness -0.490 and kurtosis 5.565 of the standardized innovation. More flexibilities are added to model M1 with care until the full model M8 is reached. Model M2 introduces time-varying skewness parameter into M1, while the kurtosis parameter is hold constant. The estimate of c_1 is significant, implying that positive returns tend to increase the skewness. Model M3 differs from model M2 only by the inclusion of the autoregressive term γ_{t-1} in the equation of γ_t. The persistence parameter c_3 is found to be significant. From the significance of c_1 in both model M2 and M3, we conclude that the persistence is not spurious. Model M4 and M5 are similar to model M2 and M3, except that the kurtosis parameter is time-varying while the skewness is hold constant. Positive returns is found to increase the kurtosis in Model M4 and M5 from the significance of parameter d_1. The persistence in the kurtosis parameter, implied by parameter d_3, is not spurious either in Model M5 with the skewness parameter fixed. Model M6 allows both shape parameters to be time-varying, without including their lagged values in the model. The parameters c_2 and d_1 are found significant, which motivates putting the lagged shape parameters into their dynamics. Model M7 introduced γ_{t-1} into model M6 and model M8 in turn introduced δ_{t-1} into model M7. The persistence in both shape parameters are not spurious. From the change in the log-likelihood, it is well worth including these terms by AIC. The full model M8 increase the log-likelihood by a significant amount from model M7, and is selected by both AIC and BIC. From the nesting feature of these models, likelihood ratio tests can be easily constructed using the log-likelihood reported in Table 4 to test hypothesis of time-invariant parameters. One would reject the null hypothesis of constant skewness parameter or constant kurtosis parameter. The time-varying shape parameters γ_t and δ_t as well as the conditional heteroskedasticity h_t from model M8, are plotted in Figure 4. For each shape parameter, the time-invariant estimate and its 95% confidence interval from model M3 and M5, respectively, are also plotted. It is interesting to note that jumps in conditional heteroskedasticities tend to be associated with jumps in both conditional shape parameters. This observation gives empirical evidence that conditional heteroskedasticity alone may not be sufficiently flexible to capture the dynamics of the conditional density.

[Figure 4 about here.]

The estimation results with Pearson's Type IV distribution are summarized in Table 5. The variance components in all models are of little different and are close to the results in Table 4. Model M1 with constant shape parameters implies skewness -0.226 and kurtosis 5.726, comparing to the empirical skewness -0.490 and kurtosis 5.565 of the standardized innovation. The fitting process from model M2 to model M8 is the same as in the case of Johnson's SU distribution. Given the closeness of the two distributions, it is not surprising to see that inference about the dynamics of the two shape parameters ν_t and $g(m_t)$ is almost identical to that under Johnson's SU distribution. That is, both shape parameters are time-varying and the persistence in both of them are not spurious. The log-likelihood of the fitted models are also very close to those obtained with Johnson's SU distribution. There is one point, however, worth noting that these models are fitted assuming that the variance of the innovation exists but not necessarily the higher moments. This suggests that Pearson's Type IV distribution may particularly be useful when extreme events are present. Comparing the loglikelihoods of all the models in Table 4 and Table 5, we find that the two distributions give very close fit for models M1–M3, but Pearson's Type IV yields noticeable higher loglikelihoods for models M4–M8. The dynamics plot of the conditional heteroskedasticity and shape parameters are similar to those from Johnson's SU distribution in Figure 4, and are therefore omitted.

[Table 5 about here.]

6. DISCUSSION

This article introduces two distributions from systems of frequency curves into the ARCH context and illustrates their usefulness in flexible modeling asymmetry and fat-tail of financial returns data. Both Pearson's Type IV and Johnson's SU have two shape parameters and span a large range in the legitimate domain of skewness and kurtosis. Both distributions has location and scale parameters, which can be used to constructed standardized distributions with mean 0 and variance 1. The standardized distributiona are then used as the innovation distribution in an ARCH framework. Dynamics are imposed on the two shape parameters, forming autoregressive conditional density models. Pearson's Type IV distribution has even thicker tail than Johnson's SU. It can be viewed as an alternative forumlation of the increasingly popular skewed t distribution. The normalizing constant of Pearson's Type IV has been implemented to facilitate the application of this distribution. Johnson's SU distribution has an advantage that the density is available in easy-to-compute closed-form. It also has the interpretation of "transform to normal". Our simulation study shows that the QMLE of the volatility parameters from both distributions are more efficient that the normal-QMLE and their efficiency relative to the MLE depends on the shape of the true but unknown innovation distribution. Even though the QMLE from a lighter tailed distribution can be highly efficient in volatility parameters, their implied skewness and kurtosis can be off the empirical measurements. Both distributions should be added to practitioner's toolbox in approximating the unknown truth.

Equipped with a fleet of asymmetric and fat-tailed distributions, one may consider an adaptive procedure to fit a real dataset. Although financial returns data often has fat-tails, it is not necessarily the case that the thicker-tail the better. It is possible that a fat-tail distribution gives kurtosis much higher than the empirical kurtosis. When the tail is not too thick, a hyperbolic, Gram-Charlier, or many other distributions may give equally good fit. It is deemed useful to comprehensively survey all these asymmetric and fat-tailed distributions. The final choices of which distributions to use may be made from some preliminary analysis. For a given dataset, one can fit a simple model with a robust method, for example, normal-QMLE. From the resulting standardized innovations, one can compute simple statistics that measure asymmetry and fat-tail without even assuming the existence of skewness and kurtosis. These measures can then be used to guide the selection of which distributions to use (e.g. Hogg 1974).

REFERENCES

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society, Series B: Statistical Methodology 65 (2), 367–389.

Barndorff-Nielsen, O. E., 1997. Normal inverse Gaussian distributions and stochastic volatility modeling. Scandinavian Journal of Statistics 24, 1–13.

Barndorff-Nielsen, O. E., Stelzer, R., 2004. Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Levy-Processes. Working Paper 178, Centre for Analytical Finance, University of Aarthus.

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

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

Bollerslev, T., Engle, R., Nelson, D., 1994. ARCH models. In: Engle, R., McFadden, D. (Eds.), The Handbook of Econometrics. Amsterdam: North-Holland, pp. 2959–3038.

Bollerslev, T., Wooldridge, J. M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.

Brooks, C., Burke, S. P., Persand, G., 2002. Autoregressive conditional kurtosis. Discussion paper, ISMA Center, University of Reading.

Devroye, L., 1986. Non-uniform Random Variate Generation. Springer-Verlag Inc.

Eberlein, E., Keller, U., 1995. Hyperbolic distributions in finance. Bernoulli 1, 281–299.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation (STMA V24 501). Econometrica 50, 987–1007.

Fernández, C., Steel, M. F. J., 1998. On Bayesian modeling of fat tails and skewness. Journal of the American Statistical Association 93, 359–371.

Hansen, B., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

Harvey, C. R., Siddique, A., 1999. Autoregressive conditional skewness. Journal of Financial and Quantitative Analysis 34 (4), 465–487.

Harvey, C. R., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55 (3), 1263–1295.

Heinrich, J., 2004. A guide to the Pearson Type IV distribution. Note 6820, Collider Detector at Fermilab, Fermilab, Batavia, Illinois.
URL www-cdf.fnal.gov/publications/cdf6820_pearson4.pdf

Hogg, R. V., 1974. Adaptive robust procedures: A partial review and some suggestions for future applications and theory (C/R: P923-27). Journal of the American Statistical Association 69, 909–923.

Johnson, N. L., 1949. Systems of frequency curves generated by methods of translation. Biometrika 36, 149–176.

Johnson, N. L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions. Volume 1. (Second Edition). John Wiley & Sons.

Jondeau, E., Rockinger, M., 2001. Gram-charlier densities. Journal of Economic Dynamics and Control 25 (10), 1457–1483.

Jondeau, E., Rockinger, M., 2003. Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. Journal of Economic Dynamics and Control 27 (10), 1699–1737.

Jones, M. C., Faddy, M. J., 2003. A skew extension of the t-distribution, with applications. Journal of the Royal Statistical Society, Series B: Statistical Methodology 65 (1), 159–174.

McCulloch, J. H., 1996. Financial applications of stable distributions. In: Maddala, G. S., Rao, C. R. (Eds.), Handbook of Statistics Volume 14: Statistical Methods in Finance. Elsevier/North-Holland [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam], pp. 393–425.

Nagahara, Y., 1999. The PDF and CF of Pearson type IV distributions and the ML estimation of the parameters. Statistics & Probability Letters 43, 251–264.

Nelson, D. B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59, 347–370.

Newey, W. K., Steigerwald, D. G., 1997. Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroskedasticity models (STMA V39 1541). Econometrica 65, 587–599.

Pearson, K., 1895. Contributions to the mathematical theory of evolution. ii. skew variation in homogeneous material. Philosophical Transactions of the Royal Society of London. A 186, 343–414.

Premaratne, G., Bera, A. K., 2001. Modeling asymmetry and excess kurtosis in stock return data. Working paper, Department of Economics, University of Illinois.

Stuart, A., Ord, J. K., 1994. Kendall's Advanced Theory of Statistics. Volume 1. Distribution Theory (Sixth Edition). Edward Arnold Publishers Ltd.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–26.

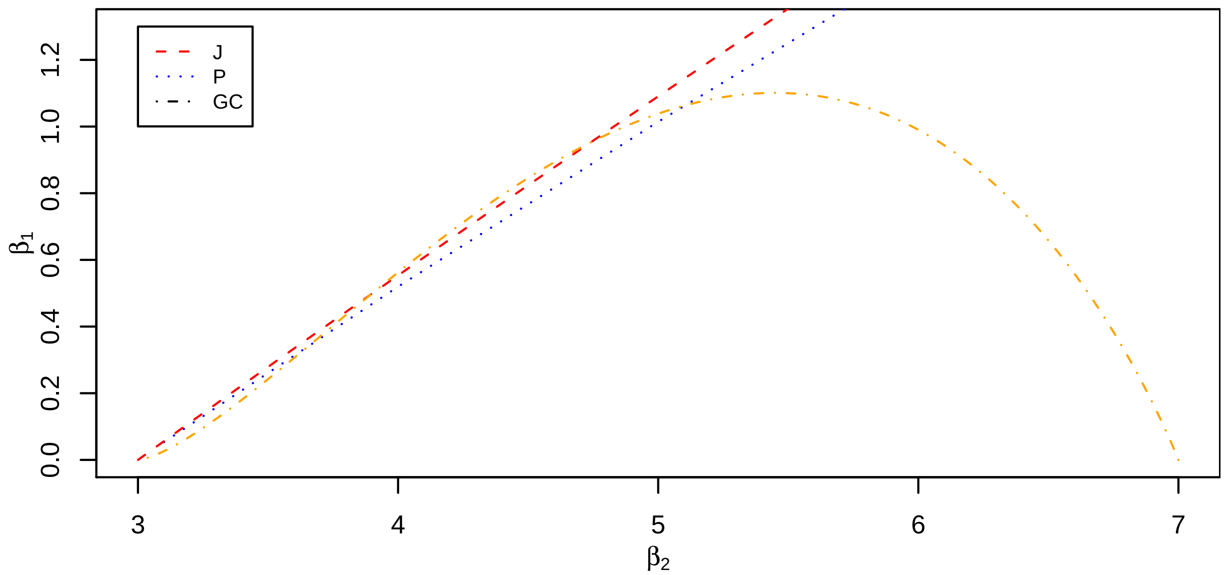

Figure 1: Comparison charts of the β_1 and β_2 regions for different distributions. The dashed line is the boundary of Johnson's SU. The dotted line is the boundary of Pearson's Type IV. The dash-dotted line is the boundary of Gram-Charlier.

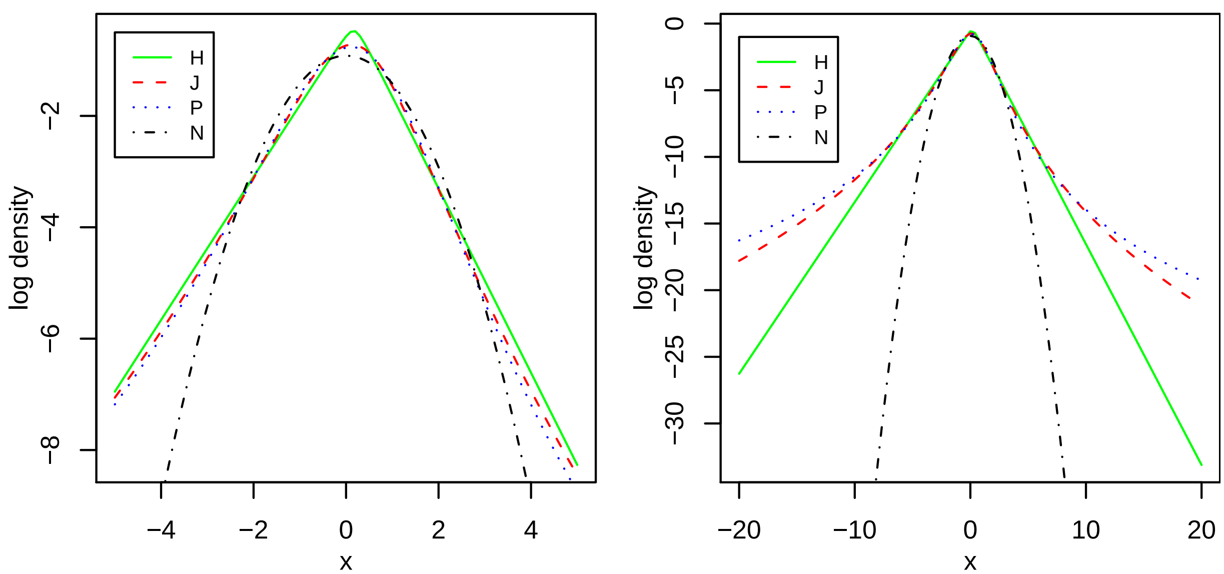

Figure 2: Comparison of log-densities, zoomed in and out. The solid line is hyperbolic. The dashed line is Johnson's SU. The dotted line is Pearson's Type IV. The dash-dotted line is normal.

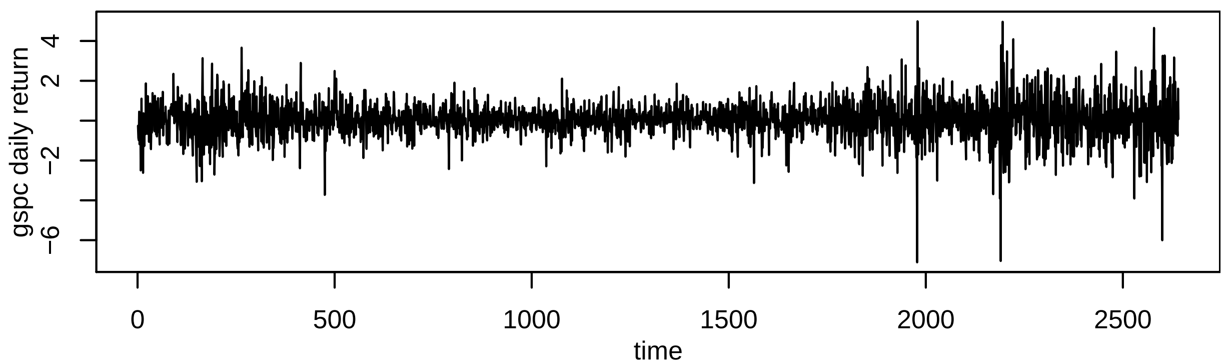

Figure 3: Daily returns of S&P 500 index from January 3, 1990 to June 14, 2000.

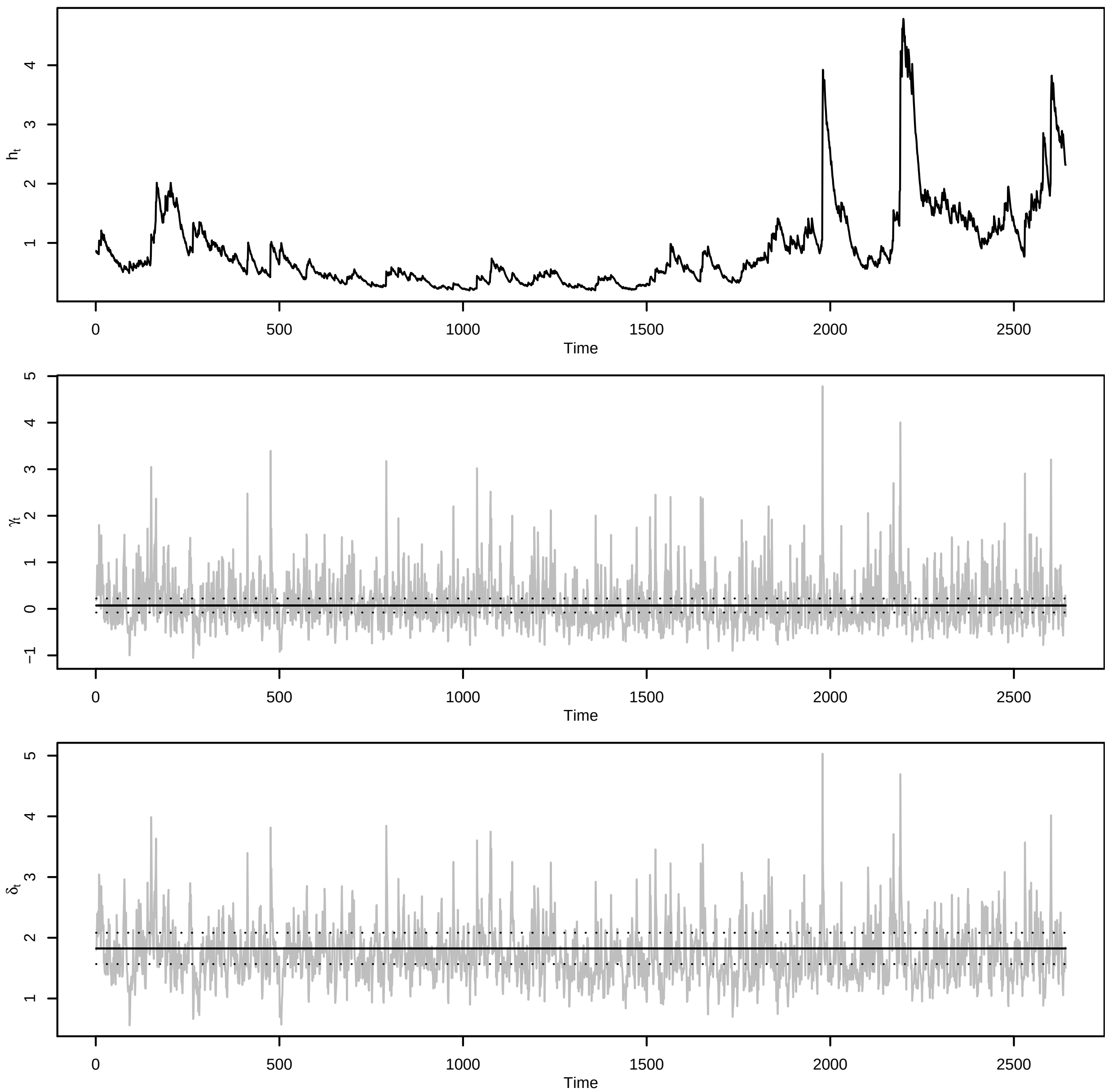

Figure 4: Time-varying conditional heteroskedasticity and shape parameters. The overlapped solid line and dotted lines in the plots for shape parameters are the time-invariant estimates and their 95% condifence interval from model M3 and M5, respectively.

Table 1: Quasi-maximum likelihood estimates of volatility parameters based on 1000 replications of sample size 2000.

True Distribution		Assumed Distribution							
		Normal		Hyperbolic		Johnson SU		Pearson IV	
Parms	True	Est	rmse	Est	rmse	Est	rmse	Est	rmse
Hyperbolic									
b_0	0.1	0.109	0.042	0.106	0.035	0.114	0.040	0.122	0.046
b_1	0.1	0.102	0.028	0.101	0.024	0.109	0.029	0.116	0.034
b_2	0.8	0.788	0.062	0.792	0.051	0.793	0.052	0.793	0.053
Johnson SU									
b_0	0.1	0.110	0.048	0.105	0.034	0.106	0.035	0.108	0.036
b_1	0.1	0.102	0.030	0.099	0.023	0.101	0.023	0.102	0.024
b_2	0.8	0.788	0.068	0.793	0.051	0.793	0.051	0.793	0.050
Pearson IV									
b_0	0.1	0.112	0.046	0.107	0.036	0.107	0.036	0.108	0.036
b_1	0.1	0.103	0.029	0.100	0.023	0.100	0.023	0.101	0.023
b_2	0.8	0.784	0.065	0.790	0.052	0.791	0.052	0.791	0.052

Table 2: Quasi-maximum likelihood estimates of shape parameters based on 1000 replications of sample size 2000.

True Distribution		Assumed Distribution								
		Hyperbolic			Johnson SU			Pearson IV		
Parms	True	Psudo	Est	Se	Psudo	Est	Se	Psudo	Est	Se
Hyperbolic										
β	−0.024		−0.025	0.017	0.243	0.242	0.046	0.561	0.558	0.121
α	0.194		0.200	0.119	1.263	1.270	0.067	2.266	2.290	0.153
Johnson SU										
γ	0.327	−0.174	−0.190	0.075		0.340	0.086	0.905	0.959	0.278
δ	1.671	1.224	1.289	0.336		1.695	0.132	3.216	3.296	0.370
Pearson IV										
ν	1.129	−0.262	−0.288	0.118	0.378	0.394	0.107		1.206	0.388
m	3.714	1.713	1.824	0.454	1.851	1.884	0.159		3.833	0.487

Table 3: Skewness and kurtosis comparison across true distributions and its minimum Kullback-Leibler Distance distributions.

True Distribution			Assumed Distribution		
	Moments	True	Hyperbolic	Johnson SU	Pearson IV
Hyperbolic	skewness	−0.500		−0.865	−2.012
	kurtosis	6.000		12.346	Inf
Johnson SU	skewness	−0.500	−0.400		−0.610
	kurtosis	6.000	4.824		8.137
Pearson IV	skewness	−0.500	−0.377	−0.438	
	kurtosis	6.000	4.504	5.174	

Table 4: Parameter estimates and standard errors of autoregressive conditional density models with Johnson's SU distribution for the S&P 500 index returns.

Parms	M1	M2	M3	M4	M5	M6	M7	M8
a_0	0.053	0.066	0.074	0.053	0.051	0.065	0.081	0.072
	0.014	0.014	0.015	0.014	0.014	0.014	0.015	0.014
b_0	0.002	0.004	0.003	0.003	0.003	0.002	0.003	0.004
	0.001	0.002	0.002	0.001	0.001	0.001	0.001	0.001
b_1	0.040	0.041	0.042	0.043	0.051	0.045	0.046	0.050
	0.008	0.009	0.009	0.009	0.010	0.008	0.009	0.009
b_2	0.958	0.953	0.955	0.957	0.949	0.954	0.954	0.950
	0.009	0.010	0.009	0.009	0.010	0.008	0.009	0.009
c_0	0.148	0.242	0.152	0.084	0.072	−0.076	0.020	−0.091
	0.095	0.158	0.086	0.085	0.081	0.115	0.111	0.077
c_1		−0.551	−0.526			−0.071	−0.336	−0.217
		0.250	0.184			0.114	0.153	0.097
c_2		−0.125	−0.081			−0.957	−0.273	−0.613
		0.081	0.063			0.415	0.244	0.232
c_3			0.660				0.584	0.527
			0.125				0.201	0.102
d_0	1.807	1.903	1.824	1.798	0.470	1.713	1.652	0.473
	0.096	0.144	0.129	0.088	0.145	0.121	0.114	0.107
d_1				−0.296	−0.282	−0.291	−0.243	−0.226
				0.065	0.049	0.055	0.071	0.035
d_2				−0.051	0.059	−0.616	−0.277	−0.407
				0.191	0.070	0.272	0.208	0.163
d_3					0.799			0.685
					0.084			0.068
LLK	−3178.0	−3172.0	−3166.2	−3173.2	−3168.1	−3164.0	−3160.6	−3150.8
AIC	6368.0	6360.0	6350.4	6362.4	6354.3	6348.0	6343.2	6325.5
BIC	6403.3	6407.0	6403.4	6409.4	6407.2	6406.8	6407.9	6396.1

Table 5: Parameter estimates and standard errors of autoregressive conditional density models with Pearson's Type IV distribution for the S&P 500 index returns.

Parms	M1	M2	M3	M4	M5	M6	M7	M8
a_0	0.053	0.066	0.081	0.058	0.060	0.056	0.066	0.066
	0.014	0.014	0.015	0.014	0.015	0.014	0.015	0.015
b_0	0.002	0.002	0.003	0.004	0.006	0.003	0.004	0.006
	0.001	0.001	0.002	0.002	0.002	0.001	0.002	0.002
b_1	0.040	0.040	0.043	0.050	0.063	0.047	0.050	0.063
	0.008	0.009	0.009	0.010	0.011	0.009	0.009	0.012
b_2	0.958	0.960	0.956	0.948	0.935	0.952	0.949	0.937
	0.009	0.009	0.010	0.010	0.012	0.009	0.009	0.012
c_0	0.478	0.742	0.491	0.050	−0.242	−0.125	−0.333	−0.197
	0.423	0.797	0.433	0.225	0.146	0.325	0.287	0.223
c_1		−1.719	−1.849			−0.145	−0.143	−0.483
		0.860	0.500			0.272	0.271	0.288
c_2		−0.219	−0.236			−3.401	−3.338	−1.499
		0.259	0.203			1.444	1.526	0.665
c_3			0.648				0.029	0.507
			0.120				0.079	0.123
d_0	0.764	0.791	0.746	0.996	0.497	0.894	0.904	0.348
	0.380	0.164	0.210	0.196	0.098	0.209	0.233	0.092
d_1				−0.919	−0.975	−0.921	−0.943	−0.804
				0.200	0.157	0.179	0.195	0.133
d_2				0.133	0.011	−0.397	−0.359	−0.169
				0.185	0.155	0.159	0.145	0.090
d_3					0.740			0.714
					0.063			0.059
LLK	−3177.9	−3171.6	−3165.8	−3170.9	−3158.7	−3158.9	−3158.3	−3144.8
AIC	6367.7	6359.3	6349.6	6357.8	6335.4	6337.8	6338.6	6313.6
BIC	6403.0	6406.3	6402.5	6404.8	6388.3	6396.6	6403.2	6384.1