The Applications of Mixtures of Normal Distributions in Empirical Finance: A Selected Survey

Tony S. Wirjanto* and Dinghai Xu †

Abstract

This paper provides a selected review of the recent developments and applications of mixtures of normal (MN) distribution models in empirical finance. One attractive property of the MN model is that it is flexible enough to accommodate various shapes of continuous distributions, and able to capture leptokurtic, skewed and multimodal characteristics of financial time series data. In addition, the MN-based analysis fits well with the related regime-switching literature. The survey is conducted under two broad themes: (1) minimum-distance estimation methods, and (2) financial modeling and its applications.

Keywords: Mixtures of Normal, Maximum Likelihood, Moment Generating Function, Characteristic Function, Switching Regression Model, (G)ARCH Model, Stochastic Volatility Model, Autoregressive Conditional Duration Model, Stochastic Duration Model, Value at Risk.

JEL classification: C01, C13.

We would like to thank the participants in the Chinese Economists Society (CES) 2009 China conference for helpful comments. Especially, we wish to thank Li Gan, Yongmiao Hong, John Knight and Hongyi Li for valuable discussions and suggestions.

*School of Accounting & Finance and Department of Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario, N2L 3G1; Tel: 001-519-888-4567 ext. 35210; Email: twirjant@uwaterloo.ca

†Department of Economics, University of Waterloo, Waterloo, Ontario, N2L 3G1; Tel: 001-519-888-4567 ext. 32047; Email: dhxu@uwaterloo.ca

1 Introduction

A distributional assumption of returns on financial assets is known to play an important role in both financial modeling and its applications. The most convenient assumption, until recently, has been that asset returns follow a stationary Gaussian/normal process. This is partly motivated by the view that in the long run, asset returns are approximately normally distributed. However, the distribution of returns on financial asset has been found to exhibit substantial leptokurtosis (fat tails) and, in many cases, also skewness (asymmetry around the mean) relative to those of a Gaussian distribution. One way to accommodate this stylized fact is to introduce a more flexible distribution model. In recent years, increasing attention has been focused on the Gaussian mixture family since any continuous distribution can be approximated arbitrarily well by an appropriate finite Gaussian mixtures. Related applications of this mixtures of normal (MN) family can be found across various disciplines that include astronomy, biology, economics, engineering, and finance.

The earliest recorded application of the MN family was undertaken by Simon Newcomb in his study in Astronomy in 1886. This is followed by Karl Pearson in his classic work on Method of Moments in 1894. The most intuitive underlying assumption of the MN family is that the data is sampled from different sub-groups. In finance applications, for example, the stock returns can be profitably viewed as one that arises from a multiple source of information, such as a firm-specific information component, a market-wide information component, and a non-information component, see Kon (1984). Applications of the MN models in other fields are documented in Everitt and Hand (1981), Titterington, Smith and Makov (1985) and Mclachlan and Peel (2000).

In this paper, we focus on the recent developments and applications of the MN models used in empirical finance. The survey is conducted under two themes: (1) minimum-distance estimation methodologies for the mixture parameters and (2) financial modeling with the MN models.

An MN model allows for great flexibility in capturing various density shapes; however, this same flexibility also turns out to lead to some estimation problems in practice. Attempts have been made to solve this problem. In particular, various methods have been proposed to estimate the parameters in the mixture models. Maximum Likelihood (ML) estimation, by far, is the most popular methods because of its optimal statistical properties. However, one of the prerequisite for implementing the ML approach is that the likelihood function must be bounded over its parameter space. In the discrete MN set-up, this condition is not always satisfied; as a result, the ML method simply breaks down. Observing this difficulty, Quandt and Ramsey

(1978) proposed an estimation method based on a Moment Generating Function (MGF). This approach minimizes the sum of squares of the distance between the theoretical MGF and its empirical counterpart. Schmidt (1982) extended the MGF to a generalized least squares (GLS) version, which is referred as the Modified MGF (MMGF). Rather than minimizing the ordinary sum of squares, the MMGF is implemented through a minimization over a generalized sum of squares involving its variance-covariance matrix. Schmidt (1982) demonstrated that the MMGF is asymptotically more efficient in general. However, the unboundedness of the MGF, as pointed out by Quandt and Ramsey (1978), can lead to a numerical instability (i.e., a failure in convergence). Alternatively, Tran (1994) introduced a Discrete Empirical Characteristic Function (DECF) method.[1] One advantage of this approach over the MGF-based methods is that the Characteristic Function (CF) is always uniformly bounded due to the Fourier transformation. The estimators based on the CF, in general, are numerically stable. In addition, the ECF contains all the information in the data because of its one-to-one correspondence with the distribution function. However, two major problems remain in the DECF method. One is the choice of the size of the grid points, and the other is the "optimal" distance among those grid points. Feuerverger and McDunnough (1981) suggested that given a fixed number of grid points, an equal space could be taken to reduce the parameter dimension in the estimation. But the first problem has been shown to be difficult to address. Schmidt (1982) and Tran (1994) suggested that for practical purposes, a 10-grid-point set should be sufficient. However, Schmidt (1982) conjectured that as the number of grid points approaches infinity, the asymptotic variance of the estimates would be close the Cramer-Rao lower bound. Hence, more grid points are always preferred to less. On the other hand, increasing the number of the grid points may yield singularity in the variance-covariance matrix, rendering the numerical estimation difficult to carry out. Xu and Knight (2009) extend the DECF method by using an iterated procedure based on the continuous ECF (CECF) to efficiently estimate the parameters in the MN models. The proposed estimation method does not suffer from the two aforementioned problems associated with the grid points since in this new method, the theoretical CF is continuously matched with its empirical component. Another important class of the estimation approaches for the mixture models is tooted in the Bayesian methodology. However, for reason of space,this review only focuses on classical estimation methods. For a survey on the Bayesian approaches to estimate the MN models, see e.g. McLanchlan and Peel (2000).

The second part of the paper discusses recent development in financial modeling which uses the MN models. As mentioned earlier, the discrete MN model is an appealing candidate for financial modeling due to its flexibility in accommodating any shape of continuous distribu-

[1]The ECF estimating procedure was formally proposed by Feuerverger and Mureika (1977) and Heathcote (1977).

tion. Notably, Kon (1984) discusses its applications to 30 stocks in the Dow-Jones Industrial Average and shows that the MN model has a substantially more descriptive validity than a student-t model. Venkataraman (1997) applies the MN model with two mixture components to construct the Value at Risk (VaR) measures. His empirical results show that the MN model provides a reasonable fit for the data. Subsequently, Chin, Weigend and Zimmermann (1999) combine Gaussian mixtures with Independent Component Analysis (ICA) to construct large portfolios of risk measures. More recently, empirical research has indicated that financial asset returns display not only excess skewness and leptokurtosis, but also time-varying volatility and volatility clustering over time. The MN model is not designed to capture thus type of dynamic characteristics. Consequently, in recent years, time-dependent models have gained much attention in the empirical finance literature. A benchmark model of this was developed by Engle (1982) and known as the Autoregressive Conditional Heteroscedasticity (ARCH) model. In a standard ARCH model, the conditional variance is a linear deterministic function of past squared errors. Bollerslev (1986) proposed a Generalized ARCH (GARCH) specification by allowing the conditional variance to be a linear deterministic function of both the past squared errors and past conditional variances. This is done to avoid estimation of the ARCH models with long lags. Stochastic Volatility (SV) models popularized by Taylor (1986) provide an alternative specification of the dynamics of the financial asset returns. In essence, the SV structure allows for additional randomness in the volatility process. The ARCH/GARCH and SV models are currently the most popular non-linear financial models. One important lesson that one can draw from this literature is that while the Gaussian GARCH and SV can usually generate a heavy tail feature, they cannot produce a sufficient amount of leptokurtosis relative to that observed in the data - see Bai, Russell and Tiao (2003) and the reference therein. Since there is no reason a priori to assume that the innovations' distribution of financial returns must be normal, the normality assumption for the innovations of these models can in principle be relaxed by replacing it with a flexible MN family. To this end, Wong and Li (2001), Haas, Mittnik and Paolella (2004), Alexander and Lazar (2006), and Xu and Wirjanto (2008) combine the MN model and the ARCH/GARCH model, giving rise to a so-called GARCH-MN model. Similarly, Mahieu and Schotman(1998), Kim, Shephard and Chib (1998) Xu and Knight (2009) and Omori, Chib, Shephard and Nakajima (2007) accommodate different numbers of the MN components in the SV specifications. In addition, De Luca nd Gallo (2004, 2009) and Hujer and Vuletic (2004), respectively, combine a mixture of exponential and Burr distributions with an autoregressive conditional duration (ACD) model proposed by Engle and Russell (1998) and Xu, Knight and Wirjanto (2008) combine MN with a stochastic conditional duration (SCD) model originally proposed by Bauwens and Veredas (2004). To be sure, there are other interesting applications of the MN models in finance as well, which, for reason of space, we do not cover in this review, including the valuation of option pricing - see e.g. Ritchey (1990) and

Melick and Thomas (1997), the local volatility models - see e.g. Brigo and Mercurio (2001), Brigo, Mercurio and Sartorelli (2002), Brigo and Mercurio (2002), and Alexander (2004), and the portfolio theory - see e.g. Buckley, Comeza-Na, Djerroud and Seco (2002), etc.

The rest of the paper is organized as follows. Section 2 reviews a number of methods used to estimate the parameters of the MN models. In section 3, we review some recent developments and applications of the MN family, and also mention briefly other mixture families, in the financial modeling. Concluding remarks are provided in Section 4.

2 Minimum-Distance Estimation Methods

In this section, we review a number of minimum-distance methods for estimating the parameters of the MN model. In a general set up of a discrete MN model, we define an independently and identically distributed (iid) random variable X drawn from K different normal distributions with probability p_k by specifying the component's probability density function as:

$$pdf(x_j) = \sum_{k=1}^{K} p_k N(\mu_k, \sigma_k^2) \tag{1}$$

where $j = 1, 2, ..., n$. Let $\theta = (p', \mu', \sigma')'$, where p, μ, σ are each K $\times$ 1 vectors. Thus we have (3K-1) unknown parameters to estimate from (1).

For simplicity and for ease of comparison, we restrict our attention to the case where the mixture has two components only; that is, we consider the case where K = 2 in (1). In other words, the random variable X is assumed to be generated from a mixture of two normal distributions. Next we recast the model as:

$$\begin{aligned} x &\sim N(\mu_1, \sigma_1^2) \quad \text{with probability } p; \\ x &\sim N(\mu_2, \sigma_2^2) \quad \text{with probability } 1-p; \end{aligned}$$

Then the mixture density can be defined as:

$$f(x; p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x-\mu_1)^2}{2\sigma_1^2}] + \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{(x-\mu_2)^2}{2\sigma_2^2}] \tag{2}$$

2.1 Maximum Likelihood Estimation

In a classic paper, Quandt (1972) used the ML method for estimating the parameters in the MN model. Theoretically, the ML approach is the most efficient method because of its well-

established statistical properties under the appropriate regularity conditions. However, in practice, the ML estimation process can fail to converge in cases that include the regime-switching model. The reason for this is that the likelihood function is not always bounded in this type of models. For illustrative purpose we provide an example of this below.

The log-transformation of the mixture likelihood of (2) can be expressed as:

$$l(\theta; x) = \sum_{i=1}^{n} ln \left(\frac{p}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x_i - \mu_1)^2}{2\sigma_1^2}] + \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{(x_i - \mu_2)^2}{2\sigma_2^2}] \right) \tag{3}$$

The idea of the ML approach is to maximize the transformed likelihood function specified in (3) with respect to the five unknown parameters, $\theta = (p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, and set the resulting score vector to a zero vector:

$$\frac{\partial lnL(\theta; r)}{\partial \theta'} = 0 \tag{4}$$

Equation (4) contains a system of five equations:

$$\begin{aligned}
\frac{\partial lnL}{\partial p} &= \sum_{i=1}^{n} \left(\frac{1}{\Delta} (\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x_i - \mu_1)^2}{2\sigma_1^2}] - \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{(x_i - \mu_2)^2}{2\sigma_2^2}]) \right) = 0 \\
\frac{\partial lnL}{\partial \mu_1} &= \sum_{i=1}^{n} \left(\frac{1}{\Delta} \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x_i - \mu_1)^2}{2\sigma_1^2}] \times \frac{(x_i - \mu_1)}{\sigma_1^2} \right) = 0 \\
\frac{\partial lnL}{\partial \mu_2} &= \sum_{i=1}^{n} \left(\frac{1}{\Delta} \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{(x_i - \mu_2)^2}{2\sigma_2^2}] \times \frac{(x_i - \mu_2)}{\sigma_2^2} \right) = 0 \\
\frac{\partial lnL}{\partial \sigma_1^2} &= \sum_{i=1}^{n} \left(\frac{1}{\Delta} [-\frac{p}{2\sqrt{2\pi}} (\sigma_1^2)^{-\frac{3}{2}} \exp[-\frac{(x_i - \mu_1)^2}{2\sigma_1^2}] + \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x_i - \mu_1)^2}{2\sigma_1^2}] \frac{(x_i - \mu_1)^2}{2(\sigma_1^2)^2}] \right) = 0 \\
\frac{\partial lnL}{\partial \sigma_2^2} &= \sum_{i=1}^{n} \left(\frac{1}{\Delta} [-\frac{1-p}{2\sqrt{2\pi}} (\sigma_2^2)^{-\frac{3}{2}} \exp[-\frac{(r_i - \mu_2)^2}{2\sigma_2^2}] + \frac{1-\alpha}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{(x_i - \mu_2)^2}{2\sigma_2^2}] \frac{(x_i - \mu_2)^2}{2(\sigma_2^2)^2}] \right) = 0
\end{aligned} \tag{5}$$

where $\Delta = \left(\frac{p}{\sqrt{2\pi\sigma_1^2}} \exp[-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}] + \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp[-\frac{(x_i-\mu_2)^2}{2\sigma_2^2}] \right)$.

Note that the system of equations in (5) cannot be solved explicitly and, in general, has non-unique roots. A more serious problem encountered by the ML estimation of (5) is that the mixture likelihood function is not always well behaved (i.e., it is unbounded from above). As a consequence, the standard numerical optimization may converge to a local maximum instead of the global one. To illustrate this problem, we arbitrarily set the values of μ_2 and σ_2^2 and choose μ_1 equal to the pth element in the random variable, x, i.e. $\mu_1 = x_p$. In other words, the pth residual vanishes from the first regime. Then for any $p \in (0, 1)$ with chosen μ_1 , μ_2 and σ_2^2, we examine the behavior of the log-likelihood over a sequence of points as σ_1^2 approaches zero.

In (4), for i = 1, 2, ... ,n, $\frac{1-p}{\sqrt{2\pi\sigma_2^2}}\exp[-\frac{(x_i-\mu_2)^2}{2\sigma_2^2}] > 0$. As $\sigma_1^2 \to 0$, for $i = 1, 2, ..., p-1, p+1, ..., n$, $\frac{p}{\sqrt{2\pi\sigma_1^2}}\exp[-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}] \to 0$. When i = p, $\frac{\alpha}{\sqrt{2\pi\sigma_1^2}}\exp[-\frac{(x_p-\mu_1)^2}{2\sigma_1^2}] = \frac{p}{\sqrt{2\pi\sigma_1^2}}e^0$. As σ_1^2 approaches to zero, the first part of the pth term in (5) will become arbitrarily large. A similar analysis applies to μ_2.[2] Thus, the ML estimation, working through (5), can incorrectly maximize an unbounded likelihood and this can lead to a numerical instability problem.

2.2 Moment Generating Function

As mentioned earlier, the ML approach can sometimes fail when it is used in estimating the MN model due to the singularity of the matrix of the second partial derivatives for the log-likelihood function. For this reason, Quandt and Ramsey (1978) introduced a method based on the MGF since the MGF has a one-to-one mapping to the distribution function and, therefore can solve the estimability problem involved in the ML procedure. Based on (2), the corresponding MGF can be expressed as:

$$g(t,\theta) = E(e^{tx}) = p\exp(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2) + (1-p)\exp(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2) \tag{6}$$

Next we define its empirical counterpart as:

$$g_n(t,x) = \frac{1}{n}\sum_{i=1}^{n}\exp(tx_i) \tag{7}$$

By the Law of Large Numbers (LLN), we have the following result: $g_n(t,x) \xrightarrow{P} g(t,\theta)$.

Hence, for a given set of grid points, Quandt and Ramsey (1978) proposed to match the theoretical MGF with its sampling counterpart. This is equivalent to minimizing the following distance measure:

$$h_1(t,\theta) = \sum_{j=1}^{m}\left(g_n(t_j,x) - g(t_j,\theta)\right)^2 \tag{8}$$

To obtain unique solutions, Quandt and Ramsey (1978) suggested that the number of grid points be set equal to the number of the parameters to estimate. Thus, in the mixtures of two normal, m is suggested to be 5, i.e. $t = (t_1, t_2, t_3, t_4, t_5)$. Quandt and Ramsey (1978) provided the asymptotic distribution for the MGF estimator; that is,

$$\sqrt{n}(\hat{\theta}_{MGF} - \theta) \xrightarrow{d} N(0, \Psi_1) \tag{9}$$

[2]For a detailed discussion of this issue, see Chapter 2 in Quant (1988).

where $\hat{\theta}_{MGF} \in \text{argmin}[h_1(t,\theta)]$, $\Psi_1 = (A'A)^{-1}A'\Omega A(A'A)^{-1}$. A is a 5×5 matrix with ijth element: $A_{ij} = \frac{\partial g(t_i,\theta)}{\partial \theta_j}$, $i,j = 1,2,...,5$ and Ω is also a 5×5 matrix with ijth element: $\Omega_{ij} = g(t_i + t_j, \theta) - g(t_i,\theta)g(t_j,\theta)$.

Noting that the minimization through (8) can be viewed equivalently as a non-linear least square (NLS) procedure, where the residual is defined as the difference between (6) and (7), Schmidt (1982) proposed a more efficient MGF procedure, namely a Modified MGF (MMGF). In essence, the MMGF approach extends the NLS procedure to a GLS procedure that involves the variance-covariance matrix of the disturbance terms. It was shown that the MMGF estimators are asymptotically more efficient than those from the ordinary MGF process. The efficiency gains can be substantial in cases where the number of grid points is greater than five ($m \geq 5$). The distance measure in the MMGF procedure can be usefully defined as:

$$h_2(t,\theta) = \sum_{i=1}^{m}\sum_{j=1}^{m}\left[((g_n(t_i,r) - g(t_i,\theta))) \times \Omega_{ij}^{-1} \times ((g_n(t_j,r) - g(t_j,\theta)))\right] \tag{10}$$

where $i,j = 1,2,...,m$. Ω is a $m \times m$ matrix with ijth element: $\Omega_{ij} = g(t_i+t_j,\theta) - g(t_i,\theta)g(t_j,\theta)$. Then,

$$\sqrt{n}(\hat{\theta}_{MMGF} - \theta) \xrightarrow{d} N(0, \Psi_2) \tag{11}$$

where $\hat{\theta}_{MMGF} \in \text{argmin}[h_2(t,\theta)]$, $\Psi_2 = [A'\Omega^{-1}A]^{-1}$. Here, A is a $m \times m$ matrix with ijth element: $A_{ij} = \frac{\partial g(t_i,\theta)}{\partial \theta_j}$, $i,j = 1,2,...,m$.

Both Quandt and Ramsey (1978) and Schmidt (1982) mentioned that there are two main problems associated with the MGF-based estimation; one is the optimal values and the number of the grid points, and the other relates to the numerical convergence problem. The latter problem is a result of the unboundedness of the MGF itself.

2.3 Discrete Empirical Characteristic Function

Due to the difficulties encountered in the standard ML and the MGF-type procedures mentioned previously, Tran (1994) introduced an alternative estimation process, that is based on the CF, to estimate the parameters of the MN models. This procedure is similar to the MGF-type estimation except that in this procedure we replace the MGF with the CF. One advantage of this replacement is that by the Fourier transformation, the CF is always uniformly bounded in the parameter space while the MGF is not.

The CF associated with (2) is defined as:

$$C(t,\theta) = E(e^{itx}) = p\exp(i\mu_1 t - \frac{1}{2}\sigma_1^2 t^2) + (1-p)\exp(i\mu_2 t - \frac{1}{2}\sigma_2^2 t^2) \tag{12}$$

where $i = \sqrt{-1}$.

Noting that $\exp(itx) = \cos(tx) + i\sin(tx)$, (12) can be rewritten as:

$$\begin{aligned} C(t,\theta) &= p\cos(\mu_1 t)\exp(-\frac{1}{2}\sigma_1^2 t^2) + (1-p)\cos(\mu_2 t)\exp(-\frac{1}{2}\sigma_2^2 t^2) \\ &+ i[p\sin(\mu_1 t)\exp(-\frac{1}{2}\sigma_1^2 t^2) + (1-p)\sin(\mu_2 t)\exp(-\frac{1}{2}\sigma_2^2 t^2)] \end{aligned} \tag{13}$$

Defining its empirical counterpart (ECF) by the following expression:

$$C_n(t,x) = \frac{1}{n}\sum_{j=1}^{n}\exp(itx_j) \tag{14}$$

We also can decompose the ECF into two components,

$$C_n(t,x) = \frac{1}{n}\sum_{j=1}^{n}\cos(tx_j) + i[\frac{1}{n}\sum_{j=1}^{n}\sin(tr_j)] \tag{15}$$

$C_n(t,x) \xrightarrow{P} C(t,\theta)$ by the LLN.

Then, for a given set of discrete grid points, Tran (1994) proposed a method based on the DECF, which matches (13) with (15) and estimates θ by minimizing:

$$\omega_1(t,\theta) = \sum_{j=1}^{m}\left(C_n(t_j,r) - C(t_j,\theta)\right)^2 \tag{16}$$

In terms of Schmidt (1982)'s improved NL-GLS version, we minimize:

$$\omega_2(t,\theta) = \sum_{i=1}^{m}\sum_{j=1}^{m}\left[((C_n(t_i,x) - C(t_i,\theta))) \times \Omega_{ij}^{-1} \times ((C_n(t_j,r) - C(t_j,\theta)))\right] \tag{17}$$

where the expression for Ω is readily found in Tran (1994) and Yu (1998).

The asymptotic distribution of the DECF estimator is shown by Feuerverger and McDun-

nough (1981) to be given by:

$$\sqrt{n}(\hat{\theta}_{DECF} - \theta) \xrightarrow{d} N(0, \Psi_3) \tag{18}$$

where $\hat{\theta}_{DECF} \in \text{argmin}[\omega_2(t, \theta)]$, $\Psi_3 = [A'\Omega^{-1}A]^{-1}$. Here, A is an $m \times m$ matrix with ijth element: $A_{ij} = \frac{\partial C(t_i,\theta)}{\partial \theta_j}$, $i, j = 1, 2, ..., m$.

Unfortunately the DECF procedure still suffers from the problems associated with using a discrete set of grid points. In addition, it is also difficult to obtain a closed form solution in (17). Thus, as we increase the size of t in order to improve the estimation efficiency, the numerical estimation could become computationally quite intensive.

2.4 Continuous Empirical Characteristic Function

Due to the difficulties associated with the DECF procedure, in this subsection, we present an estimator based on the CECF - for earlier development of this estimator, see for example Heathcote (1977), Besbeas and Morgan (2002) and Xu and Knight (2008). The CECF approach is essentially a generalized version for the DECF approach. In this approach, rather than evaluating the distance between the theoretical CF and the ECF over a discrete fixed-grid-point set, it matches all of the moments continuously with a continuous weighting function.

Specifically, define the CECF estimator as the minimizer of the following objective distance measure,

$$D(\theta; x) = \int_{-\infty}^{+\infty} |C_n(t) - C(t, \theta)|^2 w(t) dt \tag{19}$$

where $w(t)$ is some weighting function designed to ensure a convergence of the integral in (19). In this paper, we use the exponential weighting function, $w(t) = \exp(-bt^2)$, where b is a non-negative real number. This kernel form has been used extensively in the literature, see Paulson et al. (1975), Heathcote (1977), Knight and Yu (2002), Besbeas and Morgan (2002) and Xu and Knight (2008). There are several advantages to using this particular weighting kernel. In general, the exponential function tends to assign more weight around the origin, which is consistent with the underlying CF theory that the CF contains most information around the origin. In the MN settings, we find that under this Gaussian kernel, both the distance function in (19) and the asymptotic covariance structure of the estimator can be derived in closed form. This means that we could reduce the computational burden in the practical implementation substantially. In addition, this weighting function also continuously evaluates the distance between the theoretical CF and the ECF. By doing so, it avoids the two major problems associated with the discrete type of methods mentioned earlier, namely the choice of the size of the

evaluating grids and the choice of the distance among the grids, see Schmidt (1982). However, as Paulson et al. (1975) and Yu (2004) point out, with a special weighting form $\exp(-t^2)$, the estimation may lead to low efficiency. One way of improving the efficiency is to use the cross-validation method for the selection of the bandwidth, see Besbeas and Morgan (2002). Xu and Knight (2008) develop an efficient iterated procedure to continuously update the bandwidth via minimizing a certain precision measure of the asymptotic variance-covariance matrix of the parameter estimates.

In general, if a random sample is generated from the process (1) and the distance measure between the theoretical and empirical CF is defined as in (19), then the integral can be solved analytically and is given by:

$$\begin{aligned} D(\theta;x) &= \frac{1}{n^2}\sqrt{\frac{\pi}{b}}\sum_{i=1}^{n}\sum_{j=1}^{n}\exp(-\frac{1}{4b}(x_i-x_j)^2)+\sum_{k=1}^{K}p_k^2\sqrt{\frac{\pi}{b+\sigma_k^2}} \\ &+ 2\sum_{k=1}^{K}\sum_{h\neq k}^{K}p_kp_h\sqrt{\frac{\pi}{b+\frac{1}{2}(\sigma_k^2+\sigma_h^2)}}\exp(-\frac{(\mu_k-\mu_h)^2}{4b+2(\sigma_k^2+\sigma_h^2)}) \\ &- \frac{2}{n}\sum_{k=1}^{K}[p_k\sqrt{\frac{\pi}{\frac{1}{2}\sigma_k^2+b}}\sum_{j=1}^{n}\exp(-\frac{(x_j-\mu_k)^2}{4b+2\sigma_k^2})] \end{aligned} \tag{20}$$

The asymptotic properties of the CF based estimator are established in Heathcote (1977) and Knight and Yu (2002). Let $\hat{\theta}=\text{argmin}[D(\theta;x)]$, where $D(\theta;x)$ is defined in (20), then,

$$\sqrt{n}(\hat{\theta}-\theta)\xrightarrow{d}N(0,\Lambda^{-1}\Omega\Lambda^{-1}) \tag{21}$$

where Λ and Ω are $(3K-1)\times(3K-1)$ matrices with the $ijth$ element:

$$\Lambda_{ij}=E\left(\frac{\partial D^2(\theta;x)}{\partial\theta_i\partial\theta_j}\right) \quad \text{and} \quad \Omega_{ij}=E\left(\frac{\partial D(\theta;x)}{\partial\theta_i}\frac{\partial D(\theta;x)}{\partial\theta_j}\right) \tag{22}$$

The expressions for Λ and Ω can be readily found in Xu (2007) and Xu and Knight (2009).

Based on the closed form distance measure in (20) and the asymptotic variance-covariance matrix of the parameter estimates in (22), the iterative estimation procedure can be implemented in the following steps:

Step 1. Set the initial bandwidth value of b, say b_0 ;

Step 2. Plug in $b=b_0$ and the data of x and minimize the closed form distance function in (20) and get θ_0 , i.e. $\hat{\theta}_0=\text{argmin}[D(y;\theta,x)]$;

Step 3. Plug $\hat{\theta}_0$ into the asymptotic covariance matrix M and get M_0. Construct a precision measure, such as the trace or determinant of M, which is a function of b. Update the bandwidth b via $b_1 = \text{argmin}[\text{trace}(M_0)]$ or $b_1 = \text{argmin}[\det(M_0)]$;

Step 4. Iterate the step 2 to 3 until a stopping criterion is met,[3] for example, $|b_t - b_{t-1}| < \epsilon$ and $\epsilon = 10^{-3}$.

Xu and Knight (2009) show that the above iterated procedure can improve the efficiency of the CECF estimator based on the Asymptotic Relative Efficiency (ARE) measure between the CECF and MLE estimators.

3 Incorporating the MN Family into Financial Models

3.1 Gaussian Mixtures with Time Varying Means–Switching Regression and Mixture of AR Process

Quandt and Ramsey (1978) proposed a generalized set up based on the Gaussian mixture distribution in the form of a switching regression (SWR) model. In essence, the SWR generalizes (1) by allowing the means to be changing across observations and resulting in the following set up, for $i = 1, 2, ..., n$,

$$f(y_i) = \sum_{k=1}^{K} p_k \phi_k(y_i; \mu_{ki}, \sigma_k^2) \tag{23}$$

where $\mu_{ki} = x_i' \beta_k$. Equivalently it can be expressed as:

$$\begin{aligned} y_i &= x_i'\beta_1 + u_{1i} \qquad \text{with probability } p_1 \\ y_i &= x_i'\beta_2 + u_{2i} \qquad \text{with probability } p_2 \\ &... \\ y_i &= x_i'\beta_K + u_{Ki} \qquad \text{with probability } p_K \end{aligned} \tag{24}$$

with $u_{ki} \sim N(0, \sigma_k^2)$. In general, x_i' is $n \times m$ and the corresponding coefficient β_k is $m \times 1$. Then, we have $(K \times m + 2K - 1)$ unknown parameters, which is specified as $\theta = (p_1, ..., p_{K-1}; \beta_1, ..., \beta_K; \sigma_1^2, ..., \sigma_K^2)'$.[4]

[3] Provided that the influence function of $\hat{\theta}(b)$ is bounded, the optimal b is theoretically guaranteed to exist. This is shown in Besbeas (1999).

[4] Xu (2009) develops an efficient estimation for the SWR parameters based on the iterative procedure. See the discussion in subsection 2.4.

As an interesting observation, the SWR assumes very similar characteristics as the threshold regression(THR). In the SWR, the regime switching is characterized by the mixing proportion parameter (or a probability measure), p, while in the THR system, the regime is switching through a set of threshold variable. Both models have been successfully applied to various economic models, such as the disequilibrium market model (Fair and Jaffee, 1972, and Quandt and Ramsey, 1978), the labor-supply model (Heckman, 1974), the housing-demand model (Lee and Trost, 1978), and the union-nonunion wage model (Lee, 1978).

In a financial application, Wong and Li (2000) extend (24) to a time series model to give rise to a mixture autoregressive (MAR) model. Essentially, if x_i' in (24) is a set of different order of lagged values of y_i, the individual regression in each regime becomes an AR process with certain probability. There are several characteristics of their model that we note below. First, by mixing several linear AR processes, the MAR model produces a nonlinear time series structure. In addition, both stationary and non-stationary AR components can be accommodated in the MAR model. Furthermore, the conditional distribution at each time spot is modeled as a flexible MN, which means that the MAR models are capable of modeling the multimodal conditional distributions and with heteroscedastic characteristics. Their empirical results based on the Canadian Lynx data show some successes in capturing the features of the data. However, as Wong and Li (2001) mentioned, one main shortcoming for the MAR models stems from its simple autocorrelation structure, which is similar to that of the AR process. Furthermore, the MAR model assumes a constant conditional variance in each regime, which can not explain the time varying volatility dynamics of most financial time series data. Therefore, in the next sub-section, we discuss time varying volatility models with a discrete MN family.

3.2 Gaussian Mixtures with Time Varying Volatility

3.2.1 GARCH under the Gaussian Mixtures

So far we have seen that to the extent that financial asset returns have been found to be characterized by excess kurtosis, and, some time, also skewness (as in the case of equity returns), the MN distribution lends itself to be an attractive modeling candidate. However, financial asset returns have also been found to be characterized by time varying volatility and volatility clustering over time. The MN distribution, like any other fat tailed distributions, is not specifically designed to deal with this particular characteristic of asset returns. Instead, conditional volatility models have been proposed to capture this feature of asset returns. The first type of this is so-called deterministic volatility models which include Engle's (1982) ARCH model and Bollerslev's (1986) GARCH model. Briefly, an ARCH model specifies the conditional variance (or volatility) of asset returns to be a linear function of past squared innovations of the

conditional mean process of asset returns, while a GARCH specification allows the conditional variance to be a linear function of both past squared mean return innovations and past conditional variances.

Importantly, evidence obtained to this date seems to suggest that the GARCH models, such as those with normally distributed innovations (GARCH-N) and even those with fat tailed distributions, including the GARCH model with Student's t distributed innovations (GARCH-t), have not been able to capture the extent of skewness and, in particular, leptokurtosis typically found in the return data. This represents an important shortcoming of the GARCH models as volatility models, in particular when these volatility models are used to construct risk measures, such Value at Risk (VaR) measures. To amend this, we suggest that the traditional GARCH model be combined with the MN model to give rise to a so-called GARCH-MN model and propose a novel estimation approach for this model.

There are several advantages to using the GARCH-MN model as a volatility model: (i) unlike the ad-hoc GARCH-t model and any other GARCH models with heavy tailed distributions, the GARCH-MN model is founded upon a normally distributed assumption and, thus, allows for a component-wise application of the Central Limit Theorem; (ii) The GARCH-MN structure allows for conditional variance in each of the components as well as dynamic feedback between the components; (iii) It has tractable stationarity conditions which reduce to the usual condition for the GARCH-N model; (iv) the GARCH-MN model gives rise naturally to time-varying skewness and kurtosis, which, as some have argued, are the stylized facts of financial asset returns; (v) the GARCH-MN model captures the correlation structure of the data far better than the standard GARCH model (with or without a fat-tailed density, such as the GARCH-t model); and (vi) the GARCH-MN model can be shown to perform more superior in out-of-sample VaR forecasts than most other competing GARCH models. See, e.g. Xu and Wirjanto (2008) for evidence on this.

Focusing on the volatility modeling, we specify a return process as an m^{th}- order autoregressive, or AR(m), process

$$r_t = a_0 + \sum_{j=1}^{m} a_i r_{t-j} + e_t \tag{25}$$

where $t = 1, 2, .., T$, $r_t = 100(\log P_t - \log P_{t-1})$ with P_t being the closing price of a financial asset (such as an individual stock, a stock index, or a foreign currency) on the t^{th} trading day.

For convenience, we define X_t , in general, as the adjusted returns obtained as the residuals from the AR(m) process in (25), i.e. $X_t \equiv \hat{e}_t$, Next, we assume that X_t is generated by a K

component GARCH-MN process as follows. First, the distribution of X_t conditional on the information set up to time t-1, I_{t-1}, is specified as a K component MN

$$X_t|I_{t-1} \sim N(p_k, \mu_k, \sigma^2_{k,t}) \tag{26}$$

for $k = 1, 2, ..., K$, or more compactly

$$X_t|I_{t-1} \sim N(p, \mu, \sigma^2_t) \tag{27}$$

where $p = (p_1, p_2, ..., p_K)$, $\mu = (\mu_1, \mu_2, ..., \mu_K)$ and $\sigma^2_t = (\sigma^2_{1t}, \sigma^2_{2t}, ..., \sigma^2_{Kt})$. A zero mean process can be ensured in (26) or (27) by requiring that $\mu_j = \sum_{k=1}^{K-1}(p_k/p_j)\mu_k$. The conditional variance of each mixture component is specified as

$$\sigma^2_{k,t} = \lambda_k + \sum_{i=1}^{q} \alpha_{ki} X^2_{t-i} + \sum_{l=1}^{K}\sum_{j=1}^{s} \beta_{klj}\sigma^2_{k,t-j} \tag{28}$$

It can be expressed more compactly as

$$\sigma^2_t = \lambda + \sum_{i=1}^{q} \alpha_i X^2_{t-i} + \sum_{j=1}^{s} \beta_j \sigma^2_{k,t-j} \tag{29}$$

where $\alpha_i = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{iK})'$ is a $(K \times 1)$ vector, and β_j is a $(K \times K)$ matrix.

A considerable simplification of the GARCH-MN model can be achieved if we can assume that the past values of the l^{th} variance component have a trivially small effect on the current values of variance component and work with

$$\sigma^2_{k,t} = \lambda_k + \sum_{i=1}^{q} \alpha_{ki} X^2_{t-i} + \sum_{j=1}^{s} \beta_{kj}\sigma^2_{k,t-j} \tag{30}$$

where the $(K \times K)$ matrix β_j in (29) is a diagonal matrix. Unless otherwise stated, we will work with this version of the GARCH-MN model from now on.

The GARCH-MN model nests several previous models used in Empirical Finance: [i] the unconditional MN model with no volatility dynamics (MN); [ii] Bollerslev's (1986) GARCH-N model is obtained by setting k=1; [iii] Vlaar and Palm (1993) and Palm and Vlaar (1997) consider a GARCH model with an MN(2) error distribution and the restrictions that $\sigma^2_{2t} = \sigma^2_{1t} + \delta^2$; and [iv] Bauwens, Bos and van Dijk (1999) also propose an MN(2)-GARCH(1,1) model, but the component variances are assumed to be proportional to each other; that is, for all t, $\sigma^2_{2t} = \theta\sigma^2_{1t}$.

More recently, Haas, Mittnik and Paolella (2004) and Alexander and Lazar (2006) propose more general specifications of the GARCH-MN models. In particular, Haas, Mittnik and Paolella (2004) allow for interdependence between the variance components in each regime, while Alexander and Lazar (2006) extend the model to include asymmetric GARCH processes.

Below we briefly discuss the statistical properties of the GARCH-MN model. As before, we set $q = s = 1$ in (30) and focus on the conditional variance of each mixture component as given by a GARCH(1,1) process

$$\sigma_{k,t}^2 = \lambda_k + \alpha_k X_{t-1}^2 + \beta_k \sigma_{k,t-1}^2 \tag{31}$$

where, as before, we obtain the GARCH-N model for $k = 1$. First, from (31), it is evident that for a nonnegative conditional variance of each mixture component, we need the following restrictions: $\lambda_k > 0$, $\alpha_k \geq 0$, and $\beta_k \geq 0$.

In the GARCH-MN model in (30), a necessary and sufficient condition for the unconditional variance of the process $\{X_t\}$ to exist is given by $Det(I - \beta(1)) = \alpha(1)p') > 0$. This is equivalent to say that the the process $\{X_t\}$ is weakly stationary, if the characteristic equation $Det(I - \alpha(z)p' - \beta(z)) = 0$ lies outside the unit circle. This result allows us to obtain the following condition:

$$\left[\sum_{k=1}^{K} \frac{p_k}{1-\beta_k}(1 - \alpha_k - \beta_k)\right] \prod_{k=1}^{K}(1 - \beta_k) = 0 \tag{32}$$

It implies that, unlike the GARCH-N model, the restriction of $\alpha_k + \beta_k < 1$ needs not hold for each k. Instead, the necessary and sufficient conditions for the existence of the unconditional variance in the GARCH-MN model is given by

$$\sum_{k=1}^{K} \frac{p_k}{1-\beta_k}\alpha_k < 1 \tag{33}$$

Equation (32) also implies that the model possesses finite variance even when some of the components may not be covariance stationary, as long as the corresponding components' weights are sufficiently small. In particular, the overall unconditional variance of the model is (See Appendix A in Xu and Wirjanto, 2008)

$$E(X_t^2) = \frac{\sum_{k=1}^{K} p_k \mu_k^2 + \sum_{k=1}^{K} \frac{p_k}{1-\beta_k}\lambda_k}{\sum_{k=1}^{K} \frac{p_k}{1-\beta_k}(1 - \alpha_k - \beta_k)} \tag{34}$$

where $E(X_t^2) \equiv E(\sigma_t^2)$, and the unconditional variance of each individual mixture component of the model is

$$E(\sigma_{k,t}^2) = \frac{\lambda_k + \alpha_k E(X_t^2)}{1 - \beta_k} \tag{35}$$

Given these last two results, the conditional and unconditional third and fourth moments of the adjusted returns, $E(X_t^3|I_{t-1})$, $E(X_t^3)$, $E(X_t^4|I_{t-1})$, and $E(X_t^4)$, can be derived. See Appendix A in Xu and Wirjanto (2008). These results, in turn, can be used to calculate the conditional and unconditional skewness coefficients as $\frac{E(X_t^3|I_{t-1})}{(\sigma_t^2)^{3/2}}$ and $\frac{E(X_t^3|I_{t-1})}{[E(X_t^2)]^{3/2}}$, as well as the conditional and unconditional kurtosis coefficients as $\frac{E(X_t^4|I_{t-1})}{(\sigma_t^2)^2}$ and $\frac{E(X_t^3|I_{t-1})}{[E(X_t^2)]^2}$

3.2.2 SV under the Gaussian Mixtures

As an alternative to the GARCH model with a deterministic volatility function, the SV model proposed by Taylor (1986) allows volatility to evolve according to a stochastic process. The estimation for the SV parameters is proved to be more challenging. To illustrate the problem, we present a SV model in its standard from as,

$$x_t = \exp(h_t/2)e_t \tag{36}$$

$$h_t = \lambda + \alpha h_{t-1} + v_t \tag{37}$$

where e_t and v_t are assumed to be i.i.d[5] random disturbances as normal, i.e, $e_t \sim i.i.d\ N(0,1)$ and $v_t \sim i.i.d\ N(0, \sigma_v^2)$. The latent variable, h_t, is the log volatility at time t, and is assumed to follow a stationary AR(1) process ($|\alpha| < 1$). We call the model in (36)-(37) a SV-N model in which $(\lambda, \alpha, \sigma_v^2)$ are the unknown parameters.

From the statistical view point, the nonlinearity characterized by two random error product processes presents difficulties in the estimation. In addition, since the volatility series is latent, h_t needs to be integrated out from the likelihood function. However, the likelihood function involves a sequence of integrals with a dimension equal to the sample size. As a result, it is difficult to evaluate the integrals analytically. For this reason, alternative estimation methods have been devised and used for estimating the SV parameters.[6]. Taking the advantages of the flexibility property from the MN distribution, Kim and Shephard (1994), Kim, Shephard and Chib (1998) and Omori, Chib, Shephard and Nakajima (2007) propose an approach based on the following linearized transformation,

$$y_t = log(x_t^2) = h_t + \epsilon_t \tag{38}$$

[5]The assumption on zero correlation between the two innovations is a feature of the standard SV model; however we can relax this assumption by allowing for so-called "leverage effects".

[6]For a survey of various estimation methods for the SV-N model, see Broto and Ruiz (2003)

where $\epsilon_t = log(e_t^2)$. Under the standard normality assumption, the transformed error, ϵ_t, follows a logarithmic chi-squared distribution with 1 degrees of freedom. In (38), the dynamic characteristics of the transformed data y_t are still captured by the latent AR(1) process h_t, but through a linear specification. The new specification, defined in (37) and (38), contains all of the parameters of interest. One straightforward method to estimate the parameters of the model is the Quasi maximum likelihood (QML) method. In the QML approach, a normal density is used to approximate the $log\chi_1^2$ distribution. As is well known, the approximated normal distribution is characterized by a mean of -1.2704, and a variance of $\pi^2/2$. Under a Gaussian state space, Kalman filter techniques can be applied to the quasi-likelihood function based on (37) and (38). In essence, the quasi-likelihood function in the QML procedure is the first-order Edgeworth expansion of the exact likelihood function. But an approximation by truncating the series expansion may yield inefficient estimates of the parameters of the model.

Figure 1: $\log(\chi_1^2)$ Density and Its Corresponding Approximation

(a) (b)

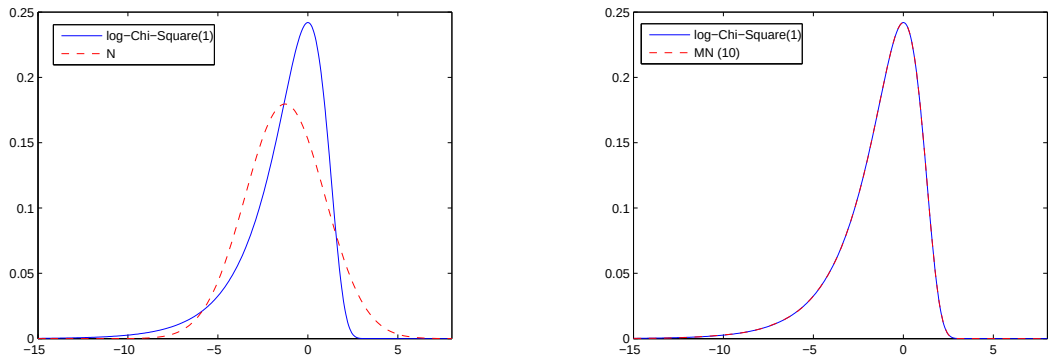

Figure 1(a) illustrates that the approximation to the density of ϵ by a normal density, instead of using the $log\chi_1^2$ density, could be inappropriate (left-skewed and tail differenced). Kim and Shephard (1994), Kim, Shephard and Chib (1998) and Omori, Chib, Shephard and Nakajima (2007) propose using the MN to approximate the $log\chi_1^2$ density. There are several advantages to this approximation. First, the flexibility MN structure provides a closer fit distribution-wise and, second, the conditional state space model with (37) and (38) is Gaussian, based on which an efficient multimove Gibbs sampling technique can be directly applied for the estimation. For comparison, a mixture of ten normal density is plotted against the $log\chi_1^2$ density in Figure 1(b).

The following ten mixture components' parameters are taken from Omori, Chib, Shephard and Nakajima (2007), which is presented in Table 1. As shown in Figure 1(b), the MN provides a close fit to the $log\chi_1^2$ distribution. However, the density approximation is based on the normality assumption for the original innovation term. A relaxation from the Normality will

Table 1. Mixtures of Ten Normal Parameters

l	p_l	μ_l	σ_l^2
1	0.00609	1.92677	0.11265
2	0.04775	1.34744	0.17788
3	0.13057	0.73504	0.26768
4	0.20674	0.02266	0.40611
5	0.22715	0.85173	0.62699
6	0.18842	1.97278	0.98583
7	0.12047	3.46788	1.57469
8	0.05591	5.55246	2.54498
9	0.01575	8.68384	4.16591
10	0.00115	14.65000	7.33342

bias the approximation further. In addition, from the empirical perspective, as Bai, Russell and Tiao (2003) pointed out, the amount of kurtosis generated from the SV-N model is far less than that from the empirical asset returns data. Subsequently, Mahieu and Schotman (1998) and Xu and Knight (2009) propose using a flexible mixture to accommodate a wider range of the distributions for the disturbance, giving rise to a linearized SV-MN (LSV-MN) model. Essentially, the SV and MN parameters are estimated simultaneously. In their methods, no prior distributional assumption for e_t is required; instead a flexible MN is used for capturing the distribution for ϵ_t. In addition, Xu and Knight (2009) derive the general closed-form cross moment conditions generated from the linearized LSV-MN model. With the model representation of (37) and (38), if $\epsilon_t \sim p_l N(\mu_l, \sigma_l^2)$ and $v_t \sim N(0, \sigma_v^2)$, then,

$$\begin{aligned} E(|x_t|^m|x_{t+k}|^n) &= \exp(\frac{n\lambda}{2}\sum_{j=1}^{k}\alpha^{j-1}) \times \exp(\frac{n^2\sigma_v^2}{8}\sum_{j=1}^{k}\alpha^{2k-2j}) \\ &\times \exp(\frac{\lambda(m+n\alpha^k)}{2(1-\alpha)} + \frac{\sigma_v^2(m+n\alpha^k)^2}{8(1-\alpha^2)}) \\ &\times \sum_{l=1}^{L} p_l \exp(\frac{m\mu_l}{2} + \frac{m^2\sigma_l^2}{8}) \\ &\times \sum_{l=1}^{L} p_l \exp(\frac{n\mu_l}{2} + \frac{n^2\sigma_l^2}{8}) \end{aligned} \tag{39}$$

With the closed-form formula in (39), it is easy to show that the LSV-MN model exhibits more flexible tail behavior than the SV-N model, see Xu and Knight (2009). From Harvey (1998), the fourth moment of x_t is:

$$E(x_t^4) = 3\exp(\frac{2\lambda}{(1-\alpha)} + \frac{2\sigma_v^2}{(1-\alpha^2)}) \tag{40}$$

Plugging $m = 4$ and $n = 0$ into (39), it yields,

$$E(x_t^4) = [\sum_{l=1}^{L} p_l \ \exp(2\mu_l + 2\sigma_l^2)] \times \exp(\frac{2\lambda}{(1-\alpha)} + \frac{2\sigma_v^2}{(1-\alpha^2)}) \tag{41}$$

It can be seen from the comparison between (40) and (41) that one part of the tail function comes from the MN distribution, while under the SV-N model, the corresponding part is only a fixed constant of 3.

The aforementioned papers utilize the flexibility of the MN family to approximate the transformed error density in the linearized SV model structure. The purpose of this is to simplify the parameter estimation involved in the SV specification. However, in the log-squared transformation, the sign information is lost in the original return time series. In addition, the conditional Normal distributional assumption of return is considered to be too restrictive in most applications. Consequently, Asai (2009), Abanto-Valle, Bandyopadhyay, Lachos and Enriquez (2009) and among others, propose imposing the MN family on e_t to capture the heavy tail characteristics of the conditional return distribution (SV-MN). And Xu (2007) derive closed-form moment conditions under a certain dependence structure. When

$$\begin{pmatrix} e_t \\ v_t \end{pmatrix} \sim \text{i.i.d.} \quad p_l \ \ N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_l^2 & \rho_l\sigma_l\sigma_v \\ \rho_l\sigma_l\sigma_v & \sigma_v^2 \end{pmatrix}\right) \tag{42}$$

then, the general cross moment conditions for x_t and x_{t+k} is given as:

$$\begin{aligned} E(x_t^m x_{t+k}^n) &= \exp(\frac{\alpha^2\sigma_v^2(m+n\alpha^k)^2}{8(1-\alpha^2)}) \\ &\times \exp(\frac{n^2\sigma_v^2}{8}\sum_{j=1}^{k-1}\alpha^{2k-2j}) \\ &\times \frac{\partial M^{(m)}(r_1,r_2)}{\partial r_1^{(m)}}\Big|_{\ r_1=0, r_2=\frac{m+n\alpha^k}{2}} \\ &\times \frac{\partial M^{(n)}(r_1,r_2)}{\partial r_1^{(n)}}\Big|_{\ r_1=0, r_2=\frac{n}{2}} \end{aligned} \tag{43}$$

where $M(r_1, r_2)$ is defined as the joint MGF of e_t and v_t.

Selected moments are provided in the Appendix in Xu (2007), which demonstrates the flexibility of the SV-MN compared to the SV-N model. Empirical evidence from Asai (2009) and Abanto-Valle, Bandyopadhyay, Lachos and Enriquez (2009) also show the better performance of the SV with the MN than other competing models including SV-t, SV-GED, GARCH-N,

GARCH-t and GARCH-MN specifications.

3.3 Other Applications and Extensions

3.3.1 Mixture-ACD and Mixture-SCD

Given an increased accessibility to ultra high frequency (UHF) data by researchers working in Finance, the modeling of financial data at the transaction level has become an active research area, beginning with the work by Hasbrouck (1991) and Engle and Russell (1998). Resent research has focused on features of financial transaction data, in particular on on the irregular spacing of data in time with the aim of gaining the full amount of information involved in financial transaction data. This irregular occurrence of transaction data is typically modeled as a financial point (or duration) process. The most common types are trade durations and quote durations as defined by the time between two consecutive trade or quote arrivals, respectively. Price durations correspond to the time between absolute cumulative price changes of given size and can be used as an alternative volatility measure.[7] A salient feature of transaction data is that market events tend to be clustered over time, rendering financial durations to be positively serially correlated over time with a strong persistence. In fact the dynamic properties of financial durations are quite similar to those of daily asset return volatilities. Accounting these features of the data leads to different types of dynamic models on the basis of a duration representation, an intensity representation or a counting representation of a point process.

Engle and Russell (1998) are the first to characterize a point process in discrete time by means of a dynamic duration model. They introduce an autoregressive conditional duration (ACD) model, in which the conditional mean of the durations is modeled as a conditionally deterministic function of past information. In contrast to the ACD models, the stochastic conditional duration (SCD) model proposed by Bauwens and Veredas (2004), specifies the conditional mean of durations as a stochastic latent process, with the conditional distribution of durations defined on a positive support. A useful analogy can be readily drawn between the differences of the two specifications with the differences of the GARCH and SV frameworks for capturing the conditional volatility of financial asset returns. In particular, the SCD model relates to the logarithmic ACD model in the same way as the stochastic volatility model relates to the exponential GARCH model of Nelson (1991).

As with the GARCH models which involve multiplicative disturbances, an interesting question relates to the choice of the process for the innovation term. This question is particularly

[7]Similarly, a volume duration is defined as the time until a cumulative order volume of given size is traded and captures an important dimension of market liquidity.

relevant since resent studies have shown that although the ACD-type models perform well in capturing the persistence in the process, the fit of the distributions tends to be poor. This suggests that a further study should be conducted on the distributional assumptions of the innovation process to the durations. Now it has been argued that informed and uninformed traders tend to interact with each other in the market through information revealing price formation processes; that is, the informed traders will act to buy an asset if the market price of the asset is lower than the intrinsic value (based on their information) and, conversely sell the asset if its price is higher than the intrinsic value. To the extent that information is costly, the actions of these two groups of traders are governed by two different innovation processes. This difference in behavior motivates the introduction of a mixture of two distributions, since the instantaneous rate of transaction can be viewed as being different across categories, and this gives rise to an observed flow of transaction in which the two types of traders are virtually indistinguishable. From a statistical point of view, the mixture of (exponential) distribution is able to deliver a variance of the innovation process which is larger than the mean. This is in line with the stylized facts derivable from the estimated residuals of the ACD model (similar to the fat tails of innovations in the GARCH model).

In particular, De Luca and Gallo (2004) assume that the innovations to the duration follow a mixture of two exponential probability density functions in which the two categories of traders are combined, while Hujer and Vuletic (2004) propose a mixture of two Burr probability density functions. Within this framework, the weights of the mixture, p and 1 - p, are interpreted as the probabilities of observing a transaction carried out by the informed and uninformed traders respectively. Although this interpretation is intuitively appealing, it seems rather restrictive to think of a constant proportion of informed and uninformed traders in a given time interval. A more interesting formulation would involve time-varying weights where the weight, p, can be assumed to follow a logistic function, similar to the formulation of the Markov Switching models with time-varying probabilities. This is presented recently in De Luca and Gallo (2009).

While several modifications of the original ACD specification have been put forward in the literature,[8] the study that focusses on the SCD model is much less forthcoming even to this day. Bauwens and Veradas (2004) were the first to propose an SCD model. In their study, they compare the empirical performance of an SCD model with an ACD model, and conclude that the former is preferable on statistical ground. The leverage term in the latent equation of the duration process is added by Feng, Jiang and Song (2004) to allow for an intertemporal correlation between the observable duration and the conditional duration, and the correlation is found to be positive.

[8]See Pacurar (2008) for an excellent survey on the use of the ACD model in Empirical Finance.

One econometric challenge with the SCD model lies in the construction of the dependence structure between the innovations driving the observation and latent equations of the duration process. Bauwens and Veredas (2004) and others deal with this issue by imposing a Weibull or Gamma distributional assumption on the observation equation innovation and a Gaussian distributional assumption on the latent equation innovation. Xu, Knight and Wirjanto (2008) introduce flexible discrete mixtures of bivariate normal distribution family into the SCD model giving rise to a mixtures of normal SCD (SCD-MN) model. The SCD-MN model imposes mixtures of bivariate normal distribution family on the innovations of the observation and latent equations of the duration process. This extension allows the model not only to capture the asymmetric behavior of the expected duration but also to easily accommodate a rich set of dependence structures between the two innovations driving the observation and latent equations of the duration process.

We begin the discussion of the SCD-MN model by first introducing the SCD model as presented in Bauwens and Veredas (2004). Let $0 =< \tau_0 < \tau_1 < ... < \tau_T$ denote the arrival times, and $d_1, d_2, ..., d_T$ denote the corresponding durations, i.e., $d_t = \tau_t - \tau_{t-1}$. Then the SCD model can be written as

$$d_t = exp(h_t)e_t \tag{44}$$

$$h_t = \lambda + \alpha h_{t-1} + \sigma_v v_t \tag{45}$$

where v_t is $i.i.dN(0,1)$, e_t denotes a distribution on the positive real line, possibly a function of some parameter γ. In Bauwen and Veredas (2004), the distribution of e_t is chosen to be either Weibull or Gamma with a shape parameter given by γ. Assuming that the distribution of e_t is parameterized, so that $E(e_t) = 1$, then h_t is the logarithm of the unobserved mean of d_t and is assumed to be generated by a Gaussian autoregressive process of order one, with $|\alpha| < 1$ is to ensure the stationarity of the process. It is also assume that $\{e_t\}$ and v_t are mutually independent sequences. The parameters to be estimated are $\theta = (\lambda, \alpha, \sigma_V, \gamma)'$. The parameter space is $Re \times (-1, 1) \times Re_+ \times Re_+$.

Notice the similarity between the SCD and SV models in their cannonical form is striking, except that the distribution of e_t in the SCD model is assumed to be non-normal since this is by definition is a positive random variable. However this assumption makes it possible to identify the parameter γ. This similarity also suggests that the estimation of the SCD model faces the same impediment as that faced by the estimation of the SV model. In particular, given a sequence d of T realizations of the duration process, the density of d given θ can be

written as

$$f(d|\theta) = \int f(d, h|\theta)dh \tag{46}$$

where $f(d, h|\theta)$ is the joint density of d and h, indexed by θ, or as

$$f(d|\theta) = \int f(d|h, \theta)f(h|\theta)dh \tag{47}$$

where $f(d|h, \theta)$ is the density of d indexed by θ, conditional on a vector h of the same dimension as d, and $f(h|\theta)$ is the density of h indexed by θ. Equation (47) makes it clear, as in the case of the SV model, that given the functional form assumed for the distribution of (e_t) (such as Weibull or Gamma), its multiple integral, which has a dimension equal to the sample size (T), cannot be solved analytically and must be computed numerically by simulation.

In (44), the duration time series, d_t, follow a nonlinear product process. To reduce the complexity involved in a product of two random error processes, Bauwens and Veradas (2004) propose to transform it into the following linear state space form,

$$y_t \;=\; log(d_t) = h_t + \epsilon_t \tag{48}$$

where the transformed disturbance is given by $\epsilon_t = log(e_t)$, and the latent variable h_t follows an AR(1) process given by (45).[9]

Bauwens and Veredas (2004) proposed a Weibull(v,1) or Gamma(v,1) distribution on e_t and the Gaussian distribution for the innovation v_t. In addition, the two innovations, e_t and v_t, are assumed to be uncorrelated. Through a logarithmic transformation, ϵ_t will have a Log-Weibull(v, 1) distribution or Log-Gamma(v, 1)distribution. The two resulting density functions are given by:

Log-Weibull(v, 1)

$$f(x) \;=\; v\exp(vx - e^{vx}) \tag{49}$$

Log-Gamma(v, 1)

$$f(x) \;=\; \frac{1}{\Gamma(v)}\exp(vx - e^{x}) \tag{50}$$

[9]The distribution of ϵ_t can in principle be approximated by Gaussian and then Kalman filter can be applied to calculate the approximate likelihood as in Bauwens and Veredas (2004).

Like the SV model, there is no closed form expression available for the likelihood function of the SCD model. However, as shown by Knight and Ning (2008), there is a closed form expression for the characteristic function (CF) of y_t. Since the CF carries the same amount of information as the distribution function itself, the SCD model can be uniquely and fully parameterized by the CF. This suggests that it is possible to estimate the model by matching the theoretical CF of the model to the ECF from the sampling observations, by minimizing the distance between the joint CF and ECF. This idea is implemented in Knight and Ning (2008) where they derive the moment conditions and joint CF expressions based on the i.i.d error distributional assumptions. However, to examine the appropriateness of the"leverage effect" captured by the SCD model, we need to specify certain dependence structure between the two innovations. As alluded to earlier, it is not straightforward to accommodate correlations between the Weibull or Gamma distribution and the Gaussian distribution. An obvious approach to model the dependence would be to use copulas with the specified marginals. Unfortunately, the estimation of such models would not be straightforward either; instead it requires simulation based estimators. Xu, Knight and Wirjanto (2008) impose distributional assumptions directly on the transformed errors, ϵ_t, and v_t. In the current literature, there are two popular specifications to model the correlations of the innovations:[10]

(a) Contemporaneous dependence structure

$$\begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix} \sim p_l \;\; N\left(\begin{pmatrix} \mu_l \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_l^2 & \rho_l\sigma_l\sigma_v \\ \rho_l\sigma_l\sigma_v & \sigma_v^2 \end{pmatrix}\right) \tag{51}$$

(b) Lagged inter-temporal dependence structure

$$\begin{pmatrix} \epsilon_{t-1} \\ v_t \end{pmatrix} \sim p_l \;\; N\left(\begin{pmatrix} \mu_l \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_l^2 & \rho_l\sigma_l\sigma_v \\ \rho_l\sigma_l\sigma_v & \sigma_v^2 \end{pmatrix}\right) \tag{52}$$

where $l = (1,\ 2,\ ...\ ,\ \mathrm{L})$, L is number of mixture components, p_l is the mixing proportion parameter, and $\sum_{l=1}^{L} p_l = 1$.

In the above specifications, the parameter ρ captures the correlation between the transformed errors ϵ_t and v_t. However we are interested in examining the relationships between e_t and v_t, which are the innovations from the original specification. So by way of transformation (i.e., $e_t = \exp(\epsilon_t)$), we need to back out the implied correlation expression from the above assumptions. This is given in the Proposition 1 in Xu, Knight and Wirjanto (2008); that is,

[10]See Jiang, Knight and Wang (2005), Yu (2005) and Xu (2007) for details of SV modeling under these two dependence structures.

under assumption (a) or (b), we have:

$$(i) \quad \text{cov}(e_t, v_t) \;=\; \sum_{l=1}^{L} p_l \rho_l \sigma_v \sigma_l \exp(\mu_l + \frac{1}{2}\sigma_l^2) \tag{53}$$

or

$$(ii) \quad \text{cov}(e_{t-1}, v_t) \;=\; \sum_{l=1}^{L} p_l \rho_l \sigma_v \sigma_l \exp(\mu_l + \frac{1}{2}\sigma_l^2) \tag{54}$$

Xu, Knight and Wirjanto (2008) also provide a general closed form moment expressions to examine the statistical properties of the model under these two dependence structures. In the case of contemporaneous dependence, then if ϵ_t and v_t satisfy assumption (a), for m, n, k ≥ 0, the closed form expression for cross-moments between d_t and d_{t+k} is given by:

$$\begin{aligned} E(d_t^m d_{t+k}^n) \;=\;& \exp(n\lambda \sum_{j=1}^{k} \alpha^{j-1}) \\ \times\;& \exp(\frac{\lambda(m+n\alpha^k)}{(1-\alpha)} + \frac{\alpha^2\sigma_v^2(m+n\alpha^k)^2}{2(1-\alpha^2)}) \\ \times\;& \exp(\frac{n^2\sigma_v^2}{2} \sum_{j=1}^{k-1} \alpha^{2(k-1-j)}) \\ \times\;& \sum_{l=1}^{L} p_l \;\exp(m\mu_l + \frac{m^2\sigma_l^2}{2} + \frac{(m+n\alpha^k)^2\sigma_v^2}{2} + m(m+n\alpha^k)\rho_l\sigma_l\sigma_v) \\ \times\;& \sum_{l=1}^{L} p_l \;\exp(n\mu_l + \frac{n^2\sigma_l^2}{2} + \frac{n^2\sigma_v^2}{2} + n^2\rho_l\sigma_l\sigma_v) \end{aligned} \tag{55}$$

In the case of lagged intertemporal dependence, if ϵ_t and v_t satisfy assumption (b), for m, n, k ≥ 0, the closed form expression for cross moments between d_t and d_{t+k} is given by:

$$\begin{aligned} E(d_t^m d_{t+k}^n) \;=\;& \exp(n\lambda \sum_{j=1}^{k} \alpha^{j-1}) \\ \times\;& \exp(\frac{\lambda(m+n\alpha^k)}{(1-\alpha)} + \frac{\sigma_v^2(m+n\alpha^k)^2}{2(1-\alpha^2)}) \\ \times\;& \exp(\frac{n^2\sigma_v^2}{2} \sum_{j=2}^{k} \alpha^{2(k-j)}) \\ \times\;& \sum_{l=1}^{L} p_l \;\exp(m\mu_l + \frac{m^2\sigma_l^2}{2} + \frac{n^2\alpha^{2k-2}\sigma_v^2}{2} + mn\alpha^{k-1}\rho_l\sigma_l\sigma_v) \\ \times\;& \sum_{l=1}^{L} p_l \;\exp(n\mu_l + \frac{n^2\sigma_l^2}{2}) \end{aligned} \tag{56}$$

Since the estimation of the SCD-MN model is closely parallel to that of the MN-SV model,

we refer the readers to that sub-section for a discussion on issues related to the estimation of the model. It suffices to say that the closed form solution for general moment conditions and joint CF derived in Xu, Knight and Wirjanto (2008) for the SCD-MN model not only renders the resulting statistical inference simpler and but also reduces the required computational costs. Another important advantage of the approach proposed in Xu, Knight and Wirjanto (2008) is that the structure of the SCD-MN model could accommodate different correlation structures between the innovations from the duration and latent autoregressive processes. This opens up an avenue for conducting an analysis on the asymmetric behavior of the expected durations and the local dynamic behavior of the observed durations. Jiang, Knight, Wang (2005) examine the properties of the SV model under different dependence specifications, i.e. contemporaneous and lagged inter-temporal correlations between the two innovations. Recognizing that a SCD model possesses a similar framework as the SV model, it would be interesting to investigate these dependence structures in the context of the SCD model.

3.3.2 MN Applications in Risk Analysis

In this subsection, we briefly mention the MN applications in the risk analysis. In recent years, risk management analysis has become increasingly important to financial institutions due to the rapid globalization and increased trading volumes with the associated potential risks. Regulators are beginning to design new regulations around it, such as bank capital standards for market risk and the reporting requirements for the risks associated with derivatives used by corporations. One of the fundamental issues in the financial risk management is to fully characterize the distribution of the returns. In other words, a good approximation for the unconditional distribution of the returns is very important for a further risk construction.

Finding suitable models for asset returns is the first step in the financial risk management. As mentioned, once the evolution of the returns are successfully modeled, the associated risk measures can be constructed accordingly. One benchmark risk-measure is the so-called Value at Risk (VaR), which is defined as the minimum expected loss at a specified probability level over a certain period. The VaR measurement is very attractive to many practitioners since VaR quantifies the potential risk exposure into a single number. In particular, statistically speaking, the VaR value corresponds to the lower quantile of the return distribution. However, in applications the VaR calculation is often based on the normality assumption. Consequently, the VaR implied from a Normal distribution can be expressed as follow,

$$\text{VaR} = \Phi^{-1}(\varsigma) \tag{57}$$

where ς is a given probability (e.g., 0.05 or 0.01), and $\Phi^{-1}(.)$ stands for the inverse of the normal

cumulative probability function.

As argued earlier, the stock returns are not well approximated by the normal distribution, particularly in the short run; as a result the normality based VaR tends to underestimate the risk. Venkataraman (1997) incorporated the MN into the construction of the VaR measures for the stock and portfolio returns. The MN-based VaR was shown to perform significantly better than that from the conventional Normal approach. Similar findings have been established in Zangari (1996), Hull and White (1998), Zhang and Cheng (2005) and etc. To further capture the time-varying volatility dynamics, Ausin and Galeano (2007) and Xu and Wirjanto (2008) used an GARCH-MN structure for the construction of the VaR measurement. By using backtesting methods, Xu and Wirjanto (2008) showed that the VaR measures obtained from the GARCH-MN model outperform those obtained from other competing models, including normal, MN and GARCH-N and GARCH-t models.

As an extension, the VaR can be constructed in the multivariate environment. It is well known that one of the main issues in the multivariate model is the dimension of its parameters. This problem becomes more acute when the MN is introduced into the multivariate structure. For this reason, several dimension reduction techniques have been introduced to solve this issue. One of the toold used for such a dimension reduction is the so-called Independent Component Analysis (ICA). Its origin can be traced back to he signal processing analysis - see Hyvarinen, Karhunen and Oja (2001) for more details. The main advantage of this technique lies in the transformation from a multivariate setting into several individual independent components, for which the univariate analysis can be easily applied. Chin, Weigend and Zimmermann (1999) combined the MN and ICA technique in computing "large" portfolio VaRs. The results showed that the MN with the ICA provides a good VaR performance in both the in-sample and out-of-sample tests. Recently, Xu and Wirjanto (2009) combine the GARCH-MN model with the ICA technique for calculating several large portfolio VaR measures. Based on the empirical results from the stock and FX market, the proposed method was shown to produce more reliable and superior VaR estimates to other competing methods, while maintaining the computational feasibility.

4 Conclusion Remarks

This paper provides a selected survey of the recent applications of the mixture models in the empirical finance. The review was carried out under two broad themes: statistical estimation methodologies via minimum-distance measures and financial applications. We showed that the incorporation of the MN distributional family allowed for great flexibility in capturing many of

the stylized facts or empirical properties of the financial asset returns. From the financial application perspective, we noted improved results from adopting the MN in the models. However, there are still many unresolved issues. Here We list a few of them as avenues for future research. First, throughout the paper, the number of the mixture components (or regimes) was taken as given. In practice, the determination of the number of the clusters remains a difficult task - see McLanchan (1987), Thode, Finch and Mendell (1988), Bozdogan (1992), Feng and McCulloch (1996), Polymedis and Titterington (1998) etc and reference therein. Second, as mentioned earlier, the unknown parameters in the MN-type models tend to increase rapidly as the number of the mixture components increases. This can lead to numerical convergence problems in practice and sometimes result in a prohibitively computational cost, especially when we work with a large sample size. This has been a major impediment to the attempts to extend the model a multivariate setting. Third, there is the identification issue; specifically, as two clusters among the mixtures become more and more similar, the identification process becomes more and more difficult to make, and so does the estimation of the model. Finally, we think the most challenging task with the MN-type models used in finance is the interpretation of each mixture component or regime. Unlike in exact or natural science, in finance, it is difficult to know exactly the compositions and the sources of the observed data. In engineering, for example, the mixed signals can be traced back to the original sources. In biology, the mixed data can be identified via the physical species or others. Although we can interpret the observed financial data as a mixture of different information components, this, to a certain level, remains to this date at best an educated guess only.

References

[1] Abanto-Valle, C.A., D. Bandyopadhyay, V.H. Lachos, I. Enriquez (2008), Robust Bayesian Analysis of Heavy-tailed Stochastic Volatility Models Using Scale Mixtures of Normal Distributions, working paper.

[2] Alexander, C. (2004), Normal mixture diffusion with uncertain volatility: Modelling short- and long-term smile effects, Journal of Banking & Finance, 28, 2957-2980

[3] Alexander, C. and E. Lazar (2006), Symmetric Normal Mixture GARCH, Journal of Applied Econometrics, 21, 307-336.

[4] Andersen, T. S. and B. Sorensen (1996), GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study Journal of Business and Econmic Statistics, 14, 329-352.

[5] Asai, M. (2009), Bayesian Analysis of Stochastic Volatility Models with Mixture-of- Normal Distributions, forthcoming in Mathematics and Computers in Simulation.

[6] Ausin, M.C. and P. Galeano (2007), Bayesian estimation of the Gaussian mixture GARCH model, Computational Statistics and Data Analysis, 51, 2636-2652.

[7] Bauwens, L., C. Bos, and H. van Dijk (1999), Adaptive Polar Sampling with an Application to a Bayes Measure of Value-at-Risk, Tinbergen Institute, Discussion Paper, TI 99-082/4.

[8] Bai, X., Russell, J. R., and G. Tiao (2003), Kurtosis of GARCH and Stochastic Volatility Models with Non-Normal Innovations, Journal of Econometrics, 114, 349-360.

[9] Bauwens,L and D. Verdas, D. (2004), The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations, Journal of Econometrics, 119 (2), 381-482.

[10] Besbeas, P. (1999), Parameter Estimation Based on Empirical Transforms. Ph.D. Thesis, University of Kent, UK.

[11] Besbeas, P. and Morgan, B.J.T. (2002), Integrated Squared Error Estimation of Normal Mixtures. Computational Statistics and Data Analysis, 44 (3), 517-526.

[12] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, Journal of Econometrics, 31, 307-327.

[13] Bozdogan. H. (1992), Choosing the Number of Component Clusters in the Mixture-Model Using a New Information Complexity Criterion of the Inverse-Fisher Information Matrix, Information and Classification: Concepts, Methods and Applications, 44 - 54.

[14] Brigo, D. and F. Mercurio (2000), A mixed-up smile, Risk 13, 123126.

[15] Brigo, D., and F. Mercurio (2001), Displaced and mixture diffusions for analytically-tractable smile models, In: Geman, H., Madan, D.B., S. R. Pliska, A. C. F. Vorst (Eds.), Mathematical Finance, Bachelier Congress 2000. Springer, Berlin.

[16] Brigo, D., F. Mercurio and G. Sartorelli (2002), Alternative asset price dynamics and volatility smiles. BancaIMI report available from www.damianobrigo.it.

[17] Broto, C. and E. Ruiz (2004), Estimation Methods for Stochastic Volatility Models: A Survey, Journal of Economic Surveys, 18, 613-649.

[18] Buckley, I., G. Comeza-Na, D. Djerroud and L. Seco (2002), Portfolio Optimization For alternative Investments, unpublished mansuscript.

[19] Chin,E., A. Wrigend and H. Zimmermann, 1999, Computing Portfolio Risk Using Gaussian Mixtures and Independent Component Analysis, Proceedings of the 1999 IEEE/IAFE/INFORMS, CIFEr'99, New York, March 1999, pp. 74-117.

[20] Christoffersen, P. (1998). Evaluating Interval Forecasts. International Economic Review, 39:841862.

[21] Cohen, A.C. (1967), Estimation In Mixtures of Two Normal Distributions, Biometrika, 56, 15-28.

[22] Day, N.E. (1969), " Estimating the Components of a Mixture of Normal Distributions", Biometrika, 56,463-474.

[23] De Luca, G. and G. M. Gallo (2004), Mixture Processes for Financial Intradaily Durations, Studies in Nonlinear Dynamics & Econometrics, 8, Article 8.

[24] De Luca, G. and G. M. Gallo (2009), Time-Varying Mixing Weights in Mixture Autoregressive Conditional Duration Models, Econometric Reviews, 28, 102-120.

[25] Engle, R. F. (1982), Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom inflation, Econometrica, 50, 987-1007.

[26] Engle, R.F. and J. R. Russell, (1998), Autoregressive Conditional Duration: a New Approach for Irregularly Spaced Transaction Data, Econometrica, 19, 69-90.

[27] Everitt, B. S. and D. J. Hand, (1981), Finite Mixture Distributions, Chapman and Hall.

[28] Fair, R.C. and D.M. Jaffee (1972) , Methods of Estimation for Markets in Disequilibrium, Econometrica, 40, 497-514.

[29] Fama, E. (1965), The Behavior of Stock Prices, Journal of Business, 47, 244-280.

[30] Feng, Z. D and McCulloch, C.E. (1995). Using Bootstrap Likelihood Ratios in Finite Mixture Models, *Journal of the Royal Statistics Society*, B (58): 609 - 617.

[31] Feng, D., G. J. Jiang, and P. X. K. Song (2003), Stochastic Conditional Duration Models with 'Leverage Effect' for Financial Transaction Data, Journal of Financial Econometrics. 2, 390-421.

[32] Feuerverger, A. and P. McDunnough (1981), On the Efficiency of Empirical Characteristic Function Procedures, Journal of The Royal Statistical Society, 43, 147-156.

[33] Feuerverger, A. and R. Mureika (1977), The Empirical Characteristic Function and Its Applications, The Annals of Statistics, 5, 88-97.

[34] Haas, M., S. Mittnik, and M. Paolella (2004), Mixed Normal Conditional Heteroskedasticity, Journal of Financial Econometrics, 2, 211-250.

[35] Hasbrouck, J. (1991), Measuring the Information Content of Stock Trades', The Journal of Finance , 46, 179-207.

[36] Heathcote, C.R. (1977), Integrated Mean Square Error Estimation of Parameters, Biometrika, 64, 255-264.

[37] Heckman, J.J. (1974), Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle, American Economic Review, American Economic Association, vol. 64(1), 188-94.

[38] Hujer, R. and Vuletic, S. (2004), Econometric analysis of financial trade processes by mixture duration models, Working Paper.

[39] Hull, J. and A. White, Value at Risk when Daily Changes in Market Variables are not Normally Distributed, Journal of Derivatives, Spring 1998, 9-19.

[40] Hyv'arinen, A. (1999), Fast and robust fixed-point algorithms for independent component analysis,'IEEE Transactions on Neural Networks, 10, 626-634.

[41] Hyv'arinen, A. and E. Oja (1997), A fast fixed-point algorithm for independent component analysis, Neural Computation, 9, 1483-1492.

[42] Jiang, G. J. and J. Knight, (2002), Estimation of Continuous - Time Processes via the Empirical Characteristic Function, Journal of Business & Economic Statistics, 20, 198 - 212.

[43] Jiang, G.J., J. L. Knight, and G. G. Wang (2005), Alternative Specifications of Stochastic Volatility Models-Theoretical and Empirical Comparison, Working Paper. U. W. O.

[44] Kim, S, N. Shephard, and S. Chib (1998), Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models, Review of Economic Studies, 45, 361-393

[45] Kon, S. J. (1984), Models of Stock Returns: Finance, 39, 147-165.

[46] Knight, J. L. and Ning, Q. (2008), Estimation of the stochastic conditional duration model via alternative methods, The Econometrics Journal 11(3).

[47] Knight, J. L. and J. Yu (2002), Empirical Characteristic Function in Time Series Estimation, Econometric Theory, 18, 691 - 721.

[48] Kon, S. (1984), Models of Stock Returns - A Comparison, The Journal of Finance, 39, 147 - 165.

[49] Lee, L.F., (1978), Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables, International Economic Review, 19(2), 415-33.

[50] Lee, L.F. and Trost, R.P. (1978), Estimation of Some Limited Dependent Variable Models with Application to Housing Demand, Journal of Econometrics, 8, 357-382.

[51] Lopez, J. (1998), Methods for Evaluating Value-at-Risk Estimates, FRBNY.

[52] Mahieu, R and Schotman, P. (1998), "An Empirical Application of Stochastic Volatility Models", Journal of Applied Econometrics, 13 (4) 333-359.

[53] McLachlan, G.(1987), On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture, Applied Statistics, 38: 318 - 324.

[54] McLachlan, G. and D. Peel (2000), Finite Mixture Models, Wiley.

[55] Melick, W.R.,and C. P. Thomas (1997), Recovering an assets implied pdf from option prices: An application to crude oil during the gulf crisis. Journal of Financial and Quantitative Analysis, 32, 91115.

[56] Nelson, D. (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, Journal of Econometrics, 43, 227-251.

[57] Omori, Y., Chib, S., Shephard, N. and Nakajima, J (2007), Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference, Journal of Econometrics, 140 (2), 425-449.

[58] Palm, F. C. and P. J. G. Vlaar (1997), Simple Diagnostic Procedures for Modeling Financial Time Series. Allgemeines Statistisches Archiv, 81, 85-101.

[59] Paulson, A. S. , Holcomb, E. W. and Leitch, R.A. (1975), The Estimation of the Parameters of the Stable Laws, Biometrika, 62, 163 - 170.

[60] Polymedis, A. and Titterington, D. M. (1998), On the Determinants of the Number of Components in a Mixture, Statistics & Probability Letters, 38: 295 - 298.

[61] Quandt, R. E. (1972), A New Approach to Estimating Switching Regressions, Journal of the American Statistical Association, 67, 306-310.

[62] Quandt, R. E. and J. B. Ramsey (1978), Estimating Mixtures of Normal Distributions and Switching Regressions, Journal of the American Statistical Association, 73, 730 - 738.

[63] Quandt, R. E. (1988), The Econometrics of Disequilibrium, Basil Blackwell.

[64] Ritchey, R.J. (1990), Call Option Valuation for Discrete Normal Mixtures. Journal of Financial Research, 13, 285-296.

[65] Schmidt, P. (1982), An Improved Version of the Quandt-Ramsey MGF Estimator for Mixtures of Normal Distributions and Switching Regressions, Econometrica, 50, 501 - 516.

[66] Taylor, S. J. (1986), Modelling Financial Time Series, Wiley: Chichester, UK.

[67] Thode. H, S. Finch and N.R. Mendell, (1988), Simulated Percentage Points for the Null Distribution of the Likelihood Ratio Test for a Mixture of Two Normals.

[68] Titterington, D. M. , A. F. M. Smith and U. E. Makov (1985), Statistical Analysis of Finite of Mixture Distributions, John Willey & Son Ltd.

[69] Tran, K (1994), Mixture, Moment and Information - Three Essays in Econometrics", Ph.D Thesis. The University of Western Ontario.

[70] Tran, K. (1998), Estimating Mixtures of Normal Distributions via Empirical Characteristic Function, Econometric Reviews, 17(2): 167-183.

[71] Venkataraman, S. (1997), Value at Risk for a Mixture of Normal Distributions: The Use of Quasi-Bayesian Estimation Techniques, Economic Perspectives, Federal Reserve Bank of Chicago, 21, 2-13.

[72] Vlaar, P., and F. Palm (1993), The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity and Jumps, Journal of Business and Economic Statistics, 11, 351-360.

[73] Wong, C. and Li, W. (2000), On a Mixture Autoregressive Model, Journal of the Royal Statistical Society: Series B, 62(1), 95-115.

[74] Wong, C. and Li, W. (2001), On a Mixture Autoregressive Conditional Heteroscedastic Model, Journal of the American Statistical Association, 96, 982-995.

[75] Xu, D. (2007), Asset Returns, Volatility and Value-at-Risk, Ph.D Thesis, The University of Western Ontario.

[76] Xu, D. (2009), An Efficient Estimation for Switching Regression Models, Working Paper, University of Waterloo.

[77] Xu, D. and Knight, J. L. (2009), Continuous Empirical Characteristic Function Estimation of Mixtures of Normal Parameters, Forthcoming in Econometric Reviews.

[78] Xu, D. and Knight, J. L. (2009), Stochastic Volatility Model under a Discrete Mixture of Normals Specification, Working Paper, University of Waterloo.

[79] Xu, D. and T. S. Wirjanto (2008), An Empirical Characteristic Function Approach to VaR under a Mixture of Normal Distribution with Time-Varying Volatility, Working paper, University of Waterloo.

[80] Xu, D. and T. S. Wirjanto (2008), Value at Risk for Stochastic Volatility Model under Bivariate Mixtures of Normal Distribution - Part I: Univariate Modeling TD-Computational Finance Research Partnership Project.

[81] Xu, D., J. Knight and T. S. Wirjanto (2008), Asymmetric Stochastic Conditional Duration Model - A Mixture of Normals Approach, Working paper, University of Waterloo.

[82] Xu, D. and T. S. Wirjanto (2009), Computation of Portfolio VaRs with GARCH Models Using Independent Component Analysis, Working paper, University of Waterloo.

[83] Yu, J. (1998), Empirical Characteristic Function In Time Series Estimation and A Test Statistic in Financial Modeling, Ph. D Thesis, The University of Western Ontario.

[84] Yu, J. (2004), Empirical Characteristic Function Estimation and Its Applications, Econometric Reviews, 2004, 23, 93-123.

[85] Yu, J. (2005), On Leverage in a Stochastic Volatility Model, Journal of Econometrics, 127, 165-178.

[86] Zhang, M.H. and Cheng, Q. S. (2005), An Approach to VaR for Capital Markets with Gaussian Mixture, Applied Mathematics and Computation, 168, 1079-1085.

[87] Zangari, P. (1996), An Improved Methodology for Measuring VaR, RiskMetrics Monitor, Reuters/JP Morgan.