Some Useful Densities for Risk Management and their Properties

Eric Ghysels* Fangfang Wang†

This version: July 14, 2011

Abstract

Suppose a risk manager cares about the (conditional) variance, skewness and kurtosis of returns and uses a parametric density function based on the first four (conditional) moments. There are quite a few densities to choose from and depending on which is selected, the risk manager implicitly assumes very different tail behavior and very different feasible skewness/kurtosis combinations. Surprisingly, there is no systematic analysis of the trade-off he or she faces. It is the purpose of the paper to address this. We focus on the tail behavior and the range of skewness and kurtosis as these are key for risk management.

Keywords: Normal Inverse Gaussian, Variance Gamma, Generalized Skewed t distribution, tail behavior, feasible domain, risk neutral measure, affine jump-diffusion model

*Department of Economics, University of North Carolina at Chapel Hill and Department of Finance, Kenan-Flagler Business School at UNC, Chapel Hill, NC 27599. Email: eghysels@unc.edu

†Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL 60607. Email: ffwang@uic.edu

1 Introduction

Suppose a risk manager cares about the (conditional) moments of returns, in particular variance, skewness and kurtosis. In addition, assume that he or she wants to use a parametric density function with the first four (conditional) moments given. The idea of keeping the number of moments small and characterizing densities by those moments has been suggested in various papers, see Madan, Carr, and Chang (1998), Theodossiou (1998), Aas and Haff (2006), Eriksson, Ghysels, and Wang (2009), among others.[1] It turns out that there are quite a few densities to choose from. Depending on which density is selected, the risk manager implicitly assumes very different tail behavior and very different feasible skewness/kurtosis combinations. Surprisingly, there is no systematic analysis of the trade-off he or she faces. It is the purpose of the paper to address this.

Many appealing distributions commonly used in financial modeling belong to a larger class of densities called the generalized hyperbolic (GH) class of distributions which, as a rich family, has wide applications in risk management and financial modeling.[2] The GH class is characterized by five parameters which, when further narrowed down to subclasses of four-, three-, or two-parameter distributions, yields widely used distributions such as the normal inverse Gaussian distribution, the hyperbolic distribution, the variance gamma distribution, the generalized skewed t distribution, the student t distribution, the gamma distribution, the Cauchy distribution, the normal distribution, etc.

In this paper we focus on distributions fully characterized by their first four moments, and extensively used in the risk management literature, namely the normal inverse Gaussian (NIG) distribution, the variance gamma (VG) distribution, and the generalized skewed t (GST) distribution.[3] We use the distributions to model the risk neutral density of asset returns, with moments

[1]Figlewski (2007) provides a comprehensive literature review of various approaches to derive risk neutral densities - we focus on moment-based methods.

[2]See for example Eberlein, Keller, and Prause (1998), Rydberg (1999), Eberlein (2001a), Eberlein and Prause (2002), Eberlein and Hammerstein (2004), Bibby and Sørensen (2003), Chen, Hardle, and Jeong (2008), among many others.

[3]The application of the NIG distribution in the context of risk management appears in Eberlein and Keller (1995a), Barndorff-Nielsen (1997a), Barndorff-Nielsen (1997b), Rydberg (1997), Eberlein (2001b), Venter and de Jongh (2002), Aas, Haff, Dimakos, and Center (2006), Kalemanova (2007), Eriksson, Ghysels, and Wang (2009), among others. The VG distribution was introduced in Madan and Seneta (1990). It was further studied by Madan, Carr, and Chang (1998), Carr, Geman, Madan, and Yor (2002), Konikov and Madan (2002), Ribeiro and Webber (2004), Hirsa and Madan (2004), Seneta (2004), Avramidis and L'Ecuyer (2006), Moosbrucker (2006), among others. Finally, the GST distribution also has had many applications, including Frecka and Hopwood (1983), Theodossiou (1998), Prause

extracted from derivative contracts. In particular, the risk neutral moments are formulated by a portfolio of the out-of-money European Call/Put options indexed by their strikes (Bakshi, Kapadia, and Madan (2003)). The risk neutral density is of course important to price derivative contracts. Focusing on risk neutral densities allows us to use relevant parameter settings - given the widely availability of options data and its applications. Note that Aït-Sahalia and Lo (2000) argue in more general terms for the use of risk neutral distributions for the purpose of risk management. Yet, many of the issues we address pertain to distributions in general - not just risk neutral ones. In particular, we focus on the tail behavior and so-called feasible domain - that is the skewness/kurtosis combinations that are feasible for each of the densities. We derive closed-form expressions for the moments (which can be used for the purpose of estimation) for the three aforementioned classes of distributions.

To appraise how well the various density approximations perform, we consider affine jump-diffusion models, which yield closed form expressions for the risk neutral density. This allows us to appraise how well the various classes of distributions approximate the density implied by realistic jump diffusions and its resluting derivative contracts. In addition we also compare the goodness of fit to some existing approaches such as: the Gram-Charlier series expansion (GCSE) and the Edgeworth expansion, previously suggested as approximating densities.[4]

The rest of this paper is outlined as follows: we start with a review on the GH family of distributions in section 2, and then study their tail behavior and moment-based parameter estimation. In section 3 we characterize the feasible domains of various distributions using S&P 500 index options data and report findings of a simulation study based on jump diffusion processes. Section 4 covers empirical applications in option pricing and value-at-risk forecasting. Concluding remarks appear in section 5. The technical details are in an Appendix.

(1999), Wang (2000), Bams, Lehnert, and Wolff (2005), Bauwens and Laurent (2005), Aas and Haff (2006), Kuester, Mittnik, and Paolella (2006), Rosenberg and Schuermann (2006), and Bali and Theodossiou (2007).

[4]There are two types of Gram-Charlier series expansion discussed in the literature – A-type GCSE, applied in the context of derivatives pricing by Madan and Milne (1994) and C-type GCSE applied to option pricing by Rompolis and Tzavalis (2007), and Rompolis and Tzavalis (2008). We refer to Rubinstein (1998) and Eriksson, Ghysels, and Wang (2009) for a discretion on the Edgeworth expansion.

2 Moment Conditions of the Generalized Hyperbolic Distribution

The Generalized Hyperbolic distribution was introduced by Barndorff-Nielsen (1977) to study aeolian sand deposits, and it was first applied in a financial context by Eberlein and Keller (1995b). In this section we will give a brief review of the GH family of distributions and then discuss their tail behavior and moments.

2.1 The Generalized Hyperbolic Distribution

The GH distribution is a normal variance-mean mixture where the mixture is a Generalized Inverse Gaussian (GIG) distribution. Suppose that Y is a GIG distributed variable with density:

$$f(y;a,b,p) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} y^{p-1} \exp[-\frac{1}{2}(ay + b/y)], \quad y > 0$$

where $K_p(z) = 1/2 \int_0^\infty y^{p-1} \exp[-1/2z(y+1/y)]dy$ (for $z > 0$) is a modified Bessel function of the third kind with index p. We then write $Y \overset{\mathcal{L}}{=} GIG(a,b,p)$. The parameter space of $GIG(a,b,p)$ is $\{a > 0, b > 0, p = 0\} \cup \{a > 0, b \geq 0, p > 0\} \cup \{a \geq 0, b > 0, p < 0\}$. A GH random variable is constructed by allowing for the mean and variance of a normal random variable to be GIG distributed. Namely, a random variable X is said to be GH distributed (or $X \overset{\mathcal{L}}{=} GH(\alpha,\beta,\mu,b,p)$) if $X \overset{\mathcal{L}}{=} \mu + \beta Y + \sqrt{Y}Z$ where $Y \overset{\mathcal{L}}{=} GIG(\alpha^2-\beta^2, b^2, p)$, $Z \overset{\mathcal{L}}{=} N(0,1)$, and Y is independent of Z. The density function of X is therefore

$$f_{GH}(x;\alpha,\beta,\mu,b,p) = \frac{\alpha^{1/2-p}(\alpha^2-\beta^2)^{p/2}e^{(x-\mu)\beta}}{\sqrt{2\pi}bK_p(b\sqrt{\alpha^2-\beta^2})} K_{p-1/2}\left(\alpha b\sqrt{1+\frac{(x-\mu)^2}{b^2}}\right)\left(1+\frac{(x-\mu)^2}{b^2}\right)^{p/2-1/4} \tag{1}$$

with parameters satisfying $\alpha > |\beta|, b > 0$, $p \in \mathbb{R}$, $\mu \in \mathbb{R}$.

The GH distribution is closed under linear transformations, which is a desirable property notably in portfolio management. Suppose X is $GH(\alpha,\beta,\mu,b,p)$ distributed, then $tX + l$ is GH$(\alpha/|t|, \beta/t, t\mu + l, |t|b, p)$ for $t \neq 0$ due to the scaling property of the GIG distribution (i.e., if $Y \overset{\mathcal{L}}{=} GIG(a,b,p)$, then $tY \overset{\mathcal{L}}{=} GIG(a/t, tb, p)$ for $t > 0$). Therefore, the set $\{\mu + \beta Y + \sigma\sqrt{Y}Z : Y \overset{\mathcal{L}}{=} GIG(\alpha^2-\beta^2, b^2, p), Z \overset{\mathcal{L}}{=} N(0,1), Y \perp Z, \alpha > |\beta|, b > 0, p \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0\}$ is equivalent to the set $\{\mu + \beta Y + \sqrt{Y}Z : Y \overset{\mathcal{L}}{=} GIG(\alpha^2-\beta^2, b^2, p), Z \overset{\mathcal{L}}{=} N(0,1), Y \perp Z, \alpha > |\beta|, b > 0, p \in \mathbb{R}, \mu \in \mathbb{R}\}$

under an affine transform, which implies that it is sufficient to characterize the GH distribution with five parameters. It also follows that the GH distribution is infinitely divisible, a property that yields GH Lévy processes by subordinating Brownian motions. However, the GH distribution is not closed under convolution in general except when $p = -1/2$.

Various subclasses of the GH distribution can be derived by confining the parameters to a subset of the parameter space. The widely used distributions which form subclasses of the GH distribution are the Symmetric GH distribution $GH(\alpha, 0, \mu, b, p)$, the Hyperbolic distribution $GH(\alpha, \beta, \mu, b, 1)$, and the Normal Inverse Gaussian (NIG) distribution $GH(\alpha, \beta, \mu, b, -1/2)$. Note that the parameter space of the GH distribution excludes $\{\alpha > |\beta|, b = 0, p > 0\}$ and $\{\alpha = |\beta|, b > 0, p < 0\}$ which are permitted by the GIG distribution. If we allow parameters to take values on the boundary of parameter space, we can obtain various limiting distributions. These include the (1) Variance Gamma distribution, (2) Generalized Skewed T distribution, (3) Skewed T distribution, (3) Noncentral Student T distribution, (4) Cauchy distribution, (5) Normal distribution, among others (see, for instance, Eberlein and Hammerstein (2004), Bibby and Sørensen (2003), Haas and Pigorsch (2007)).

2.2 Tail Behavior

The GH family covers a wide range of distributions and therefore exhibits various tail patterns. We discuss its tail behavior in general.[5] We write $A(x) \sim B(x)$ as $x \to \infty$ for functions A and B if $\lim_{x\to\infty} A(x)/B(x) = c$ for some constant c.

Note that $f_{GH}(x; \alpha, \beta, \mu, b, p) \sim |x - \mu|^{p-1} \exp(-\alpha|x - \mu| + \beta(x - \mu))$ (see Haas and Pigorsch (2007)). An application of L'hôpital's rule yields the following:

Proposition 2.1. *Suppose that X is $GH(\alpha, \beta, \mu, b, p)$ distributed with $\alpha > |\beta|$, $b > 0$, $p \in \mathbb{R}, \mu \in \mathbb{R}$. Its right tail and left tail satisfy $P(X - \mu > x) \sim x^{p-1}e^{-(\alpha-\beta)x}$ and $P(X - \mu < -x) \sim x^{p-1}e^{-(\alpha+\beta)x}$ respectively where $x > 0$ is sufficiently large.*

Therefore α, β and p control the tail behavior. A small α and a large p yield heavy tails. The right tail is heavier when $\beta > 0$ whereas the left tail is heavier when $\beta < 0$. Hence β relates to

[5] Bibby and Sørensen (2003) discussed the tail behavior of the GH family under a different parameterization.

skewness and $\beta = 0$ yields a symmetric distribution. The tails of subclasses of the GH distribution can be derived from Proposition 2.1 directly.

Next we look at the tails of 'limiting' distributions derived from the GH distribution. The Variance Gamma (VG) distribution is derived from the GH distribution by letting b go to 0 and keeping p positive, and hence it has the same tail behavior as the GH distribution. If we set $\alpha = |\beta|$ in (1), we obtain the Generalized Skewed T (GST) distribution. Its tail behavior has the following properties:

Corollary 2.1. *Suppose that X is GST(β, μ, b, p) distributed with $|\beta| > 0$, $\mu \in \mathbb{R}$, $b > 0$, and $p < 0$. Then for $x > 0$ sufficiently large, $P(X - \mu > x) \sim x^p$ and $P(X - \mu < -x) \sim x^{p-1}e^{-2\beta x}$ for $\beta > 0$, while $P(X - \mu > x) \sim x^{p-1}e^{2\beta x}$ and $P(X - \mu < -x) \sim x^p$ for $\beta < 0$.*

The Skewed T distribution is derived from the GST distribution by setting $p = -b^2/2$. Letting β go to 0 in the GST distribution yields the Noncentral Student T distribution with $-2p$ degrees of freedom. Its density behaves like $f_{GH}(x; 0, 0, \mu, b, p) \sim |x - \mu|^{2p-1}$ for sufficiently large $|x - \mu|$, and hence $P(|X - \mu| > x) \sim x^{2p}$, when $x > 0$ is sufficiently large. The tail property of the Cauchy distribution, as a special case of the Noncentral Student T distribution, is obtained by $p = -\frac{1}{2}$.

The Normal distribution can be viewed as a limiting case of the GH law (the Hyperbolic distribution in particular, i.e., $p = 1$) as well with mean $\mu + \beta\sigma^2$ and variance σ^2, where $\sigma^2 = \lim_{\alpha\to\infty}(b/\alpha)$.

The above discussion indicates that the tails of the GH and VG distributions are exponentially decaying, which is slower than the normal distribution but faster than the GST and the Skewed T distribution. The GH and VG distributions are therefore called semi-heavy tailed (see Barndorff-Nielsen and Shephard (2001)), and they posses moments of arbitrary order. The GST distribution does not have moments of arbitrary order. The r^{th} moment exists if and only if $r < -p$. However, its tails are a mixture of polynomial and exponential decays – one heavy tail and one semi-heavy tail, which distinguishes the GST law from the others (see Aas and Haff (2006)).

2.3 Skewness and Kurtosis

The question how skewness and excess kurtosis affect the distribution of financial returns is the primary interest of this paper. We will first present some general results regarding the GH distribution in order to characterize the mapping between moments and parameters. For a centered GH distribution (i.e., $\mu = 0$), the moments of arbitrary order can be expanded as an infinite series of Bessel functions of the third kind with gamma weights (see for example Barndorff-Nielsen and Stelzer (2005)). Using this representation, we have the following proposition:

Proposition 2.2. *Suppose that X is $GH(\alpha, \beta, 0, b, p)$ distributed. The first four moments, $m_n = EX^n$, can be explicitly expressed as*

$$\begin{aligned}
m_1 &= \frac{b\beta K_{p+1}(b\gamma)}{\gamma K_p(b\gamma)}, \\
m_2 &= \frac{bK_{p+1}(b\gamma)}{\gamma K_p(b\gamma)} + \frac{\beta^2 b^2 K_{p+2}(b\gamma)}{\gamma^2 K_p(b\gamma)}, \\
m_3 &= \frac{3\beta b^2 K_{p+2}(b\gamma)}{\gamma^2 K_p(b\gamma)} + \frac{\beta^3 b^3 K_{p+3}(b\gamma)}{\gamma^3 K_p(b\gamma)}, \\
m_4 &= \frac{\beta^4 b^4 K_{p+4}(b\gamma)}{\gamma^4 K_p(b\gamma)} + \frac{6\beta^2 b^3 K_{p+3}(b\gamma)}{\gamma^3 K_p(b\gamma)} + \frac{3b^2 K_{p+2}(b\gamma)}{\gamma^2 K_p(b\gamma)}
\end{aligned} \tag{2}$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$.

Proof. See Appendix A.

Therefore the mean M, variance V, skewness S and excess kurtosis K of a $GH(\alpha, \beta, \mu, b, p)$ random variable can be expressed explicitly as functions of the five parameters, which yields a mapping (call it $\mathcal{T}$) from the parameter space to the space spanned by (M, V, S, K). Note, however, that $\mathcal{T}$ may not be a bijection and therefore its inverse may not exist. Since this paper is aimed at modeling financial returns which are skewed and leptokurtic and aimed at building densities based on skewness and (excess) kurtosis, we restrict our attention to subclasses of the GH family which have a four-parameter characterization and yield a bijection between moments and parameters. It is impossible to express in general parameters explicitly via the first four moments due to the presence of Bessel functions. We will focus in the remainder of the paper on the cases where we have an explicit mapping between moments and parameters, namely we will focus on the NIG, VG and GST distributions.

The Normal Inverse Gaussian Distribution

The NIG distribution is obtained from the GH distribution by letting $p = -1/2$. Therefore, as an application of Proposition 2.2, we have the following,

Proposition 2.3. *Denote by M, V, S, K the mean, variance, skewness and excess kurtosis of a NIG(α, β, μ, b) random variable with $\alpha > |\beta|$, $\mu \in \mathbb{R}$, and $b > 0$. Then we have the following: (1)*

$$M = \mu + \frac{\beta b}{\gamma}, \quad V = \frac{b\alpha^2}{\gamma^3}, \quad S = \frac{3\beta}{\alpha\sqrt{b\gamma}}, \quad K = \frac{3(4\beta^2 + \alpha^2)}{b\alpha^2\gamma},$$

(2) if $D \equiv 3K - 5S^2 > 0$,

$$\alpha = 3\frac{\sqrt{D + S^2}}{D}V^{-1/2}, \quad \beta = \frac{3S}{D}V^{-1/2}, \quad \mu = M - \frac{3S}{D + S^2}V^{1/2}, \quad b = \frac{3\sqrt{D}}{D + S^2}V^{1/2}.$$

The proof of (1) uses the fact that $K_{n+1/2}(z) = K_{1/2}(z)\left(1 + \sum_{i=1}^{n}[(n+i)!2^{-i}z^{-i}]/[i!(n-i)!]\right)$. The formal derivation of (2) appears in Eriksson, Forsberg, and Ghysels (2004).

The Variance Gamma Distribution

Recall that the VG distribution is obtained by keeping $\alpha > |\beta|$, $\mu \in \mathbb{R}$, $p > 0$ fixed and letting b go to 0. Proposition 2.2 is stated for the GH distribution. For the limiting cases, similar results are derived by applying the dominant convergence theorem. Particularly, when b approaches 0, we have the following result:

Proposition 2.4. *Denote by M, V, S, K the mean, variance, skewness and excess kurtosis of a $VG(\alpha, \beta, \mu, p)$ random variable with $\alpha > |\beta|$, $\mu \in \mathbb{R}$, and $p > 0$. Then, (1)*

$$M = \mu + \frac{\beta p}{\eta}, \quad V = \frac{p}{\eta^2}(\eta + \beta^2), \quad S = \frac{\beta(3\eta + 2\beta^2)}{(\eta + \beta^2)^{3/2}p^{1/2}}, \quad K = \frac{3(\eta^2 + 4\eta\beta^2 + 2\beta^4)}{p(\eta + \beta^2)^2} \tag{3}$$

where $\eta = \frac{\alpha^2 - \beta^2}{2} > 0$. (2) If $2K > 3S^2$, letting $C = 3S^2/2K$, the equation $(C-1)R^3 + (7C-6)R^2 + (7C-9)R + C = 0$ has a unique solution in $(0, 1)$, denoted by R, and

$$\alpha = \frac{2\sqrt{R}(3 + R)}{\sqrt{V}|S|(1 - R^2)}, \quad \beta = \frac{2R(3 + R)}{\sqrt{V}S(1 - R^2)}, \quad p = \frac{2R(3 + R)^2}{S^2(1 + R)^3}, \quad \mu = M - \frac{2\sqrt{V}R(3 + R)}{S(1 + R)^2}.$$

Proof. See Appendix A.

The GST distribution is obtained from the GH distribution by $\alpha \to |\beta|$, with parameters satisfying $\beta \in \mathbb{R}$, $\mu \in \mathbb{R}$, $b > 0$, and $p < -4$ so that the 4^{th} moment exists. An application of the dominant convergence theorem to Proposition 2.2 yields the following:

Proposition 2.5. *Denote by M, V, S, K the mean, variance, skewness and excess kurtosis of a GST(β, μ, b, p) random variable with $\beta \in \mathbb{R}, \mu \in \mathbb{R}, b > 0,$ and $p < -4$. Then, (1)*

$$\begin{aligned}
M &= \mu + \tfrac{b^2\beta}{v-2} \\
V &= \tfrac{b^2}{v-2} + \tfrac{2b^4\beta^2}{(v-2)^2(v-4)} \\
S &= \left[6(v-2) + \tfrac{16b^2\beta^2}{v-6}\right] \tfrac{b\beta(v-4)^{1/2}}{[(v-2)(v-4)+2b^2\beta^2]^{3/2}} \\
K &= \left[\tfrac{8b^4\beta^4(5v-22)}{(v-6)(v-8)} + \tfrac{16b^2\beta^2(v-2)(v-4)}{v-6} + (v-2)^2(v-4)\right] \tfrac{6}{[(v-2)(v-4)+2b^2\beta^2]^2}
\end{aligned} \tag{4}$$

where $v = -2p > 8$. (2) If (K, S^2) satisfy

$$\frac{32(y_U+4)}{(y_U+2)^2} > S^2 \quad where \quad y_U \equiv \frac{\sqrt{K^2+156K+900} - K + 30}{K} > 0, \tag{5}$$

the following system of equations:

$$\begin{aligned}
0 &= 2\rho[3(v-6) + 4(v-4)\rho]^2 - S^2(v-4)(v-6)^2(1+\rho)^3 \\
0 &= 12(v-4)(5v-22)\rho^2 + 48(v-4)(v-8)\rho + 6(v-6)(v-8) \\
&\quad -K(v-4)(v-6)(v-8)(1+\rho)^2
\end{aligned} \tag{6}$$

has a unique solution satisfying $\rho > 0$ and $v > 8$, denoted by (ρ, v). We then have

$$\beta = sig(S)\sqrt{\frac{\rho(1+\rho)(v-4)}{2V}}, \quad \mu = M - sig(S)\sqrt{\frac{\rho(v-4)V}{2(1+\rho)}}, \quad b = \sqrt{\frac{V(v-2)}{1+\rho}}, \quad p = -v/2.$$

Proof. See Appendix A.

Feasible Domain

Proposition 2.3 indicates that the range of excess kurtosis and skewness admitted by the NIG distribution is $D_{nig} \equiv \{(K, S^2) : 3K > 5S^2\}$, which is referred to as the *feasible domain* of the

NIG distribution. Therefore, the *feasible domain* of the VG distribution read from Proposition 2.4 is $D_{vg} \equiv \{(K, S^2) : 2K > 3S^2\}$ and it includes D_{nig}. The GST distribution has a *feasible domain* $D_{gst} \equiv \{(K, S^2) : S^2 < \frac{32(y_U+4)}{(y_U+2)^2}\}$ where $y_U = \sqrt{1 + 156/K + 900/K^2} - 1 + 30/K$. Note that D_{gst} is included in the set $\{(K, S^2) : S^2 < \min(32, 8K/15)\}$, which is a subset of D_{nig} and D_{vg}. Therefore the GST distribution has bounded skewness, and the VG distribution has the largest feasible domain among the three. It should also be noted that $\{(K, S^2) : 2K = 3S^2\}$ is the skewness-kurtosis combination of the Pearson Type III distribution. Further details regarding feasible domains appear in Section 3.2.

3 Risk Neutral Moments and Risk Management

We are interested in modeling asset returns with NIG, VG and GST distributions for the purpose of risk management. It is therefore natural to think in terms of the risk neutral distribution since it plays an important role in derivative pricing. In this section, we present risk neutral moment-based estimation methods using European put and call contracts. Affine jump-diffusion models will be used to evaluate the performance of the NIG, VG and GST approximations.

3.1 Estimating Moments of Risk Neutral Distributions

Given an asset price process $\{S_t\}$, Bakshi, Kapadia, and Madan (2003) show that the risk neutral conditional moments of $\tau-$period log return $R_t(\tau) = \ln(S_{t+\tau}) - \ln(S_t)$ given time t information can be written as an integral of out-of-the-money (OTM) call and put option prices. In particular, the arbitrage-free prices of the volatility contract $V(t,\tau) = E_t^Q(e^{-r\tau}R_t(\tau)^2)$, cubic contract $W(t,\tau) = E_t^Q(e^{-r\tau}R_t(\tau)^3)$ and quartic contract $X(t,\tau) = E_t^Q(e^{-r\tau}R_t(\tau)^4)$ at time t can be expressed as

$$V(t,\tau) = \int_{S_t}^{\infty} \frac{2(1-\ln(K/S_t))}{K^2} C(t,\tau;K)dK + \int_0^{S_t} \frac{2(1-\ln(K/S_t))}{K^2} P(t,\tau;K)dK \tag{7}$$

$$W(t,\tau) = \int_{S_t}^{\infty} \frac{6\ln(K/S_t) - 3(\ln(K/S_t))^2)}{K^2} C(t,\tau;K)dK + \int_0^{S_t} \frac{6\ln(K/S_t) - 3(\ln(K/S_t))^2}{K^2} P(t,\tau;K)dK \tag{8}$$

$$X(t,\tau) = \int_{S_t}^{\infty} \frac{12(\ln(K/S_t))^2 - 4(\ln(K/S_t))^3)}{K^2} C(t,\tau;K)dK + \int_0^{S_t} \frac{12(\ln(K/S_t))^2 - 4(\ln(K/S_t))^3}{K^2} P(t,\tau;K)dK \tag{9}$$

where Q represents the risk neutral measure, r is risk-free rate, while $C(t,\tau;K)$ and $P(t,\tau;K)$ are the prices of European calls and puts written on the underlying asset with strike price K and expiration τ periods from time t. Therefore, the time t conditional risk neutral moments (mean, variance, skewness, and excess kurtosis) of $\ln(S_{t+\tau})$ are:

$$Mean(t,\tau) = \mu(t,\tau) + \ln(S_t) \tag{10}$$

$$Var(t,\tau) = e^{r\tau}V(t,\tau) - \mu^2(t,\tau) \tag{11}$$

$$Skew(t,\tau) = \frac{e^{r\tau}W(t,\tau) - 3\mu(t,\tau)e^{r\tau}V(t,\tau) + 2\mu(t,\tau)^3}{[e^{r\tau}V(t,\tau) - \mu(t,\tau)^2]^{3/2}} \tag{12}$$

$$EKurt(t,\tau) = \frac{e^{r\tau}X(t,\tau) - 4\mu(t,\tau)e^{r\tau}W(t,\tau) + 6e^{r\tau}\mu(t,\tau)^2V(t,\tau) - 3\mu(t,\tau)^4}{[e^{r\tau}V(t,\tau) - \mu(t,\tau)^2]^2} - 3 \tag{13}$$

where $\mu(t,\tau) = e^{r\tau} - 1 - e^{r\tau}V(t,\tau)/2 - e^{r\tau}W(t,\tau)/6 - e^{r\tau}X(t,\tau)/24$.

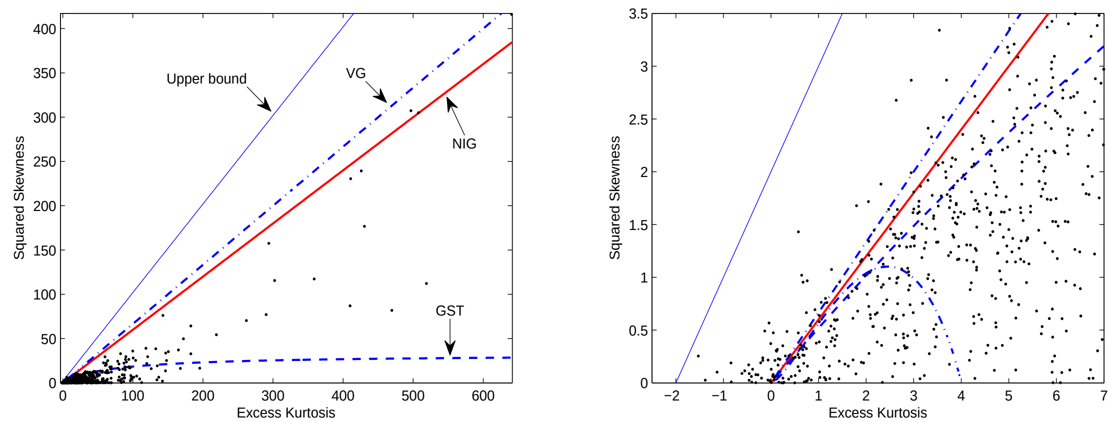
Typically we cannot implement directly equations (7), (8) and (9) since we do not have a continuum of strike prices available. Hence, the integrals are replaced by approximations involving weighted sums of OTM puts and calls across (a subset of) available strike prices. While the approximation entails a discretization bias, Dennis and Mayhew (2002) report that such biases are typically small even with a small set of puts and calls.[6] In particular, the integrals in equations (7), (8) and (9) are evaluated by a trapezoid approximation method described in Conrad, Dittmar, and Ghysels (2010). Therefore the above formulas - computed using discrete weighted sums - yield estimates of the mean, variance, skewness and excess kurtosis of the risk neutral density. In the empirical work, we follow the practical implementation discussed by Conrad, Dittmar, and Ghysels (2010).

Note that the approach pursued here is different from statistical analysis based on return-based estimation via sample counterparts of population moments. The use of derivative contracts for the purpose of pricing and risk management is widespread in the financial industry. We follow exactly this strategy, by computing moments of risk neutral densities obtained from extracting information from existing derivatives contracts. Then we will use parametric densities based on the extracted moments to compute various objects of interest to a risk manager, ranging from pricing other derivatives contracts to value-at-risk computations, etc.

[6] As noted by Dennis and Mayhew (2002) and Conrad, Dittmar, and Ghysels (2010), it is critical to select a set of puts and calls that symmetric in strike prices. In contrast, discretely weighted sums of asymmetrically positioned puts/calls result in biases.

3.2 Risk Neutral Moments and Feasible Domains

The very first question we address is whether the range of moments that are extracted from market data fall within the feasible domain of the respective densities. Figure 1 plots daily skewness and



(a) (b)

Figure 1: S&P 500 index options from 1996 to 2009: 5∼14 days to maturity

kurtosis extracted from S&P 500 index options with 5∼14 days to maturity for a sample covering January 4, 1996 to October 30, 2009.[7] There are 1258 dots in Figure 1 (a). Each dot represents a daily combination of squared skewness and excess kurtosis calculated via expressions (12) and (13). The pairs of OTM call/put options involved in the calculation of risk neutral moments per day ranges from 3 to 85, with an average of 21.26 (see Table 1).

Superimposed on the dots are the feasible domains of the NIG, VG and GST distributions, i.e., the area under the curves with labels 'VG', 'NIG', and 'GST' respectively. The line labeled 'Upper bound' represents the largest possible skewness-kurtosis combination of an arbitrary random variable, so the region above the bound is the so-called impossible region which is obtained by the formula $S^2 = K + 2$. As reported in Table 1, the VG feasible region covers 94.54% of the data

[7]The data is obtained from Optionmetrics through Wharton Research Data Services. A number of data filters are applied to screen for recording errors - these are standard in the literature and described elsewhere, see e.g. in Battalio and Schultz (2006) and Conrad, Dittmar, and Ghysels (2010), among others. Notably we use filters to try to ensure that our results are not driven by stale or misleading prices. In addition to eliminating option prices below 50 cents and performing robustness checks with additional constraints on option liquidity, we also remove options with less than one week to maturity, and eliminate days in which closing quotes on put-call pairs violate no-arbitrage restrictions. We encountered some days with two pairs of OTM call/put options. The literature often puts a lower bound of three pairs to avoid noisy moment estimates.

points, and the NIG feasible region covers 92.59%. Finally, the coverage of the GST feasible region is 81.44% of the data points.[8]

Table 1 summarizes coverage of VG, NIG, and GST distributions, and number of contracts used to compute the risk neutral moments. Though the VG, NIG, and GST distributions can not accommodate any arbitrary skewness-kurtosis combinations, most of the S&P 500 options with short maturities feature skewness-kurtosis combinations within the feasible region of the VG and NIG distributions.[9]

When we zoom in Figure 1(a) we obtain the next plot (b). The area under the curve with a range of excess kurtosis from 0 and 4 is the feasible domain of A-type Gram-Charlier series expansion (GCSE). The A-type GCSE and the Edgeworth expansion have been studied by Madan and Milne (1994), Rubinstein (1998), Eriksson, Ghysels, and Wang (2009), among others as a way to approximate the unknown risk neutral density. Since the Edgeworth expansion admits a smaller feasible region than the A-type GCSE (see Barton and Dennis (1952) for more detail), we only draw the feasible domain of the A-type GCSE in Figure 1. Clearly, most of dots are outside the feasible region of the A-type GCSE and hence outside the feasible region of the Edgeworth expansion.

	Total	Pairs of call/put			VG	NIG	GST	Impossible Region
		min	max	average				
$5 \sim 14$	1258	3	85	21.26	94.54%	92.59%	81.44%	0
$17 \sim 31$	1855	3	90	22.22	74.45%	69.06%	50.08%	0
$81 \sim 94$	1220	3	61	13.96	23.89%	15.66%	3.63%	34/1220
$171 \sim 199$	1294	3	39	11.54	8.06%	5.43%	3.47%	25/1294

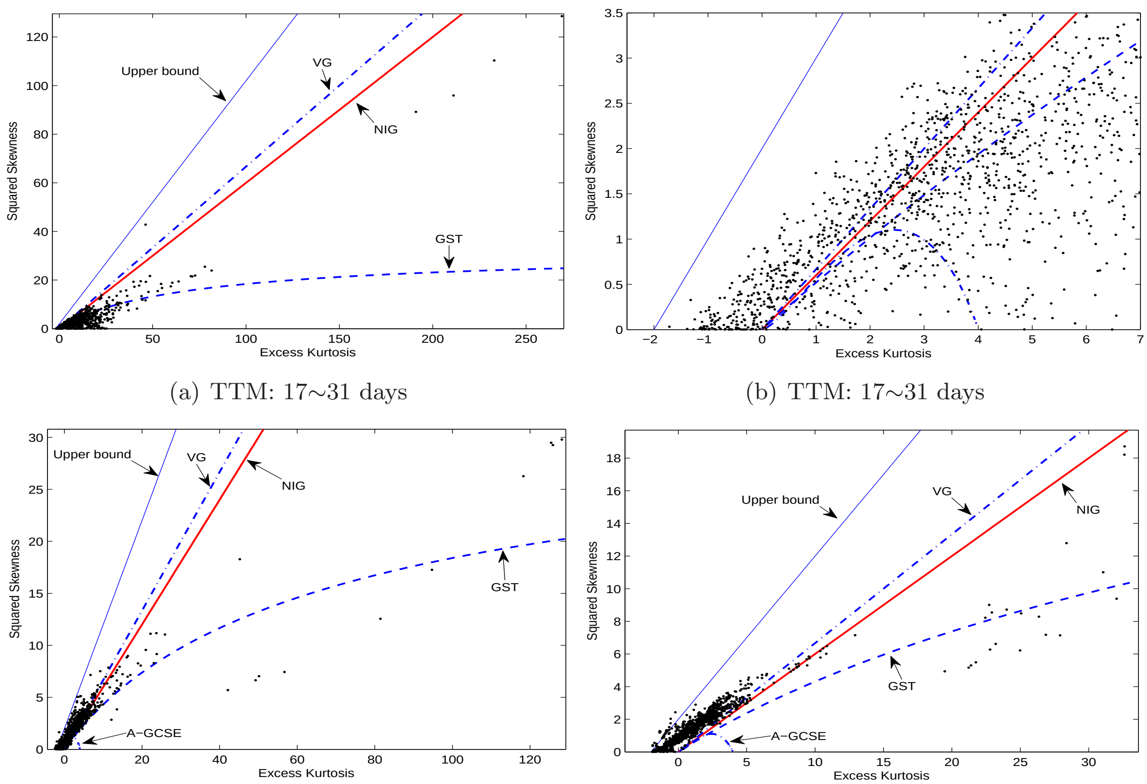
Table 1: Coverage of VG, NIG, and GST distributions for SPX from Jan 1996 to Oct. 2009

Figure 2 plots squared skewness and excess kurtosis using options with longer time to expiration: 17∼31 days (around one month), 81∼94 days (around three months), and 171∼199 days (around six

[8] Following the suggestion of a Referee we also considered the Johnson family of distributions which consists of three distributions: S_L (or Log-Normal), S_B, and S_U. For contracts shown in Figure 1(a), only 11.69% of the dots are in the S_B feasible region and 88.31% are in the S_U region. Given that the NIG, VG and GST distributions seem to be more suitable for our analysis we refrained from including the Johnson class.

[9] Contrary to the VG, NIG and GST distributions, it should be noted that the Johnson family of distributions does cover the entire domain of admissible skewness-kurtosis combinations. However, the S_B distribution in the Johnson family has bounded support (with estimated boundary parameters) - an unappealing feature for our risk management applications.

months). Table 1 contains again the numerical values of the feasible domain coverage for the various distributions. The VG feasible region covers 74.45% of the data points (out of 1855 contracts) for 17∼31 days to maturity, and 23.89% (out of 1220) for 81∼94 days to maturity. Moreover, when the time-to-maturity increases to 171∼199 days, the coverage drops to 8.06% (out of 1294). These observations are consistent with the fact that the returns are more leptokurtic when sampled more frequently. It should also be noted that a few data points in Figure 2 (c,d) are in the impossible region – 2.78% (81∼94) and 1.93% (171∼199) respectively according to the figures in the last column of Table 1. This could be due to estimation error in the moments, or perhaps arbitrage opportunity – because risk neutral moments are calculated under arbitrage-free assumptions.[10]



(a) TTM: 17∼31 days

(b) TTM: 17∼31 days

(c) TTM: 81∼94 days

(d) TTM: 171∼199 days

Figure 2: S&P 500 index options from 1996 to 2009, continued

The overall picture that emerging from our analysis so far is that the classes of distributions we examine have appealing properties in terms of skewness and kurtosis coverage - in particular,

[10]We thank one of the referees for pointing out this.

when compared with approximating densities (e.g. A-type GCSE and the Edgeworth expansion) proposed in the prior literature. The feasible domain does become more restrictive for longer term maturities beyond three months.

3.3 Simulation Evidence

We want to assess the accuracy of the various distributions via a simulation experiment. To do so we characterize risk neutral densities with a commonly used framework in financial asset pricing and risk management, namely continuous time jump diffusion processes. In particular, we use affine jump-diffusions which yield explicit expressions for the risk neutral density. We select parameters that are empirically relevant, allowing us to assess numerically how accurate the various approximating distributions are in realistic settings. The discussion will focus again on the VG, NIG, and GST distributions.

Suppose that the log price process $Y_t = \log(S_t)$ is generated from the following affine jump-diffusion model under risk neutral measure:

$$\begin{aligned} dY_t &= (r - \lambda_J \bar{\mu} - \tfrac{1}{2} V_t)dt + \sqrt{V_t} dW_t^1 + dN_t \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\rho\sqrt{V_t} dW_t^1 + \sigma\sqrt{1-\rho^2}\sqrt{V_t} dW_t^2 \end{aligned} \tag{14}$$

where N_t is a compound Poisson process with Lévy measure $\nu(dy) = \lambda_J f(dy)$, and $\bar{\mu} = \int_{\mathbb{R}} e^y f(dy) - 1$ is the mean jump size. Moreover, W^1, W^2 are two independent standard Brownian motions, and are independent of N_t as well.

For any $u \in \mathbb{C}$, the conditional characteristic function of Y_T at time t is

$$\Psi(u; t, T, x_t) \doteq E(e^{uY_T} | \mathcal{F}_t) = \exp(\psi_1(u, T-t) + \psi_2(u, T-t)V_t + uY_t), \tag{15}$$

where $x_t \doteq (Y_t, V_t)$. $\psi_1(u,\tau) = ru\tau - \kappa\theta(\frac{\gamma+b}{\sigma^2}\tau + \frac{2}{\sigma^2}\ln\left[1 - \frac{\gamma+b}{2\gamma}(1-e^{-\gamma\tau})\right]) - \lambda_J\tau(1+\bar{\mu}u) + \lambda_J\tau\int_{\mathbb{R}} e^{uy} f(dy)$, $\psi_2(u,\tau) = -\frac{a(1-e^{-\gamma\tau})}{2\gamma-(\gamma+b)(1-e^{-\gamma\tau})}$, and $b = \sigma\rho u - \kappa$, $a = u(1-u)$, and $\gamma = \sqrt{b^2 + a\sigma^2}$ (see Duffie, Pan, and Singleton (2000)). The density function of Y_T conditional on information up to time t is therefore $f(y; t, T, x_t) = \frac{1}{\pi}\int_0^\infty e^{-iuy}\Psi(iu; t, T, x_t)du$ which follows from inverse Fourier transform. The price of European call option written on Y with maturity T and strike price K

is defined as $C_t(K;T,x_t) = E((e^{Y_T} - K)^+|\mathcal{F}_t)$, and it has an explicit expression: $C_t(K;T,x_t) = P_1(K,t,T,x_t) - KP_2(K,t,T,x_t)$ where $P_1(K,t,T,x_t) = \frac{1}{2}s_t - \frac{e^{-r(T-t)}}{\pi}\int_0^\infty Im\left[\frac{e^{iv\ln(K)}\Psi(1-iv;t,T,x_t)}{v}\right]dv$, $P_2(K,t,T,x_t) = \frac{1}{2}e^{-r(T-t)} - \frac{e^{-r(T-t)}}{\pi}\int_0^\infty Im\left[\frac{e^{iv\ln(K)}\Psi(-iv;t,T,x_t)}{v}\right]dv$ and Im denotes the imaginary part of a complex number (see Duffie, Pan, and Singleton (2000)). In the numeric calibration, we consider the fast Fourier transform of Carr and Madan (1999) for both $f(y;t,T,x_t)$ and $C_t(K;T,x_t)$ (see also Lee (2004)). The details are in Appendix B.

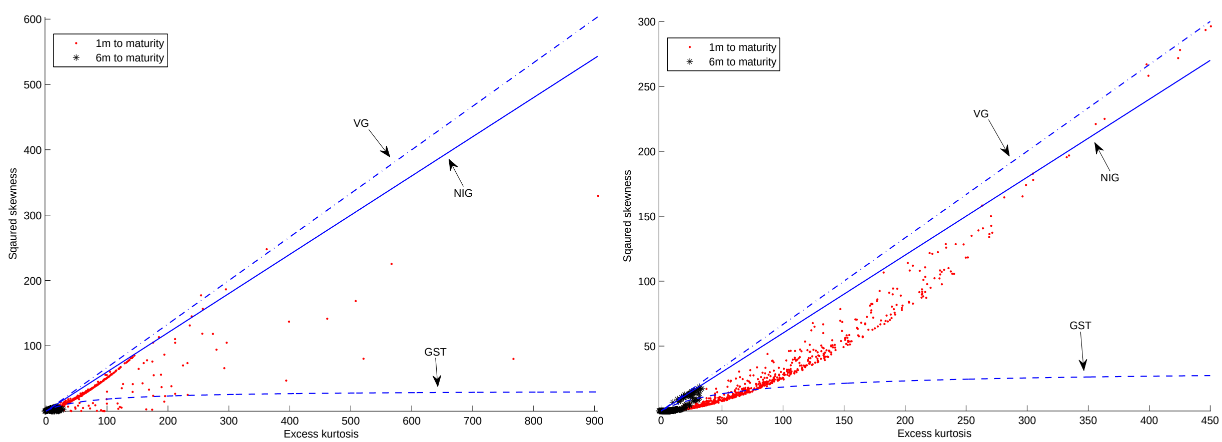
When jumps are excluded from model (14), the marginal distribution of Y_t for large t is close to the NIG distribution. Inclusion of jumps will make the asymptotic distribution of the log-price process deviate from the NIG distribution by a factor controlled by the compensated cumulant generating function of the jumps (see Duffie, Filipovic, and Schachermayer (2003) and Keller-Ressel (2011)). We therefore consider Gaussian jumps, i.e., $\int_{\mathbb{R}} e^{uy}f(dy) = \exp(\mu_J u + \sigma_J^2 u^2/2)$, and exponential jumps, i.e., $\int_{\mathbb{R}} e^{uy}f(dy) = (1-u\mu_J)^{-1}$ in model (14). The model parameters are taken from Duffie, Pan, and Singleton (2000): $r = 3.19\%$, $\rho = -0.79$, $\theta = 0.014$, $\kappa = 3.99$, $\sigma = 0.27$, $\lambda_J = 0.11$. And $\mu_J = -0.14$, and $\sigma_J = 0.15$ for Gaussian jumps, while $\mu_J = 0.14$ for exponential jumps.

We also focus exclusively on the two cases involving jumps - as they represent the most realistic processes. We use the approximating densities to price European call options and compare them with the correctly priced contract $C_0(K;T,(y_0,v_0))$ obtained from the diffusion, where (y_0,v_0) represent values of the state variables at time 0. Two measures of pricing errors are considered. The first is based on absolute price difference, denoted by L_a. The second is defined in terms of relative price difference, denoted by L_r. The two measures are defined as follows:

$$L_a = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(C_i^{model} - C_i^{true})^2} \quad \text{and} \quad L_r = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{C_i^{model} - C_i^{true}}{C_i^{true}}\right)^2} \tag{16}$$

where C_i^{true} and C_i^{model} represent the true price and the price estimated from the approximating distribution. And the sum is taken over a range of strikes (or moneyness, which is the ratio of asset price S_t and strike price K_i) for contracts written on the same security. These measures represent respectively absolute and relative pricing errors. To determine the value of (y_0,v_0), we simulate the log price process starting from value 0 and draw from the invariant distribution of the volatility process which is a Gamma distribution with characteristic function $\phi(u) = (1 - iu\sigma^2/(2\kappa))^{-2\theta\kappa/\sigma^2}$

(see, for instance, Keller-Ressel (2011)). We simulate 1000 sample paths. For each, we drop the first 1000 observations and use the 1001^{th} observation from simulation as the value of (y_0, v_0). Figure 3 displays the skewness and kurtosis calculated from the simulated AJD models with Gaussian jumps and Exponential jumps respectively, for $T = 1$ month and 6 months.[11] Table 2 reports the mean pricing errors $\bar{L}_a$ and $\bar{L}_r$, an average of L_a and L_r over 1000 iterations, for three different ranges of moneyness. Figure 3 shows that the simulated processes are within the feasible domain of the NIG and VG distributions, but the GST is inadequate in terms of feasible domain coverage - for both the Gaussian and Exponential jump cases. This is why we cover in Table 2 only the NIG and VG cases. The pricing errors of the two approximating densities are quite similar - particularly in absolute terms. In relative terms (using the L_r loss function) it seems that the NIG has a slight edge for short maturities (1 month) while the reverse is true for the longer maturity (6 months). Yet, both in terms of feasible domain and pricing errors it is fair to say that the two classes of distributions are comparable.



(a) Gaussian Jumps

(b) Exponential Jumps

Figure 3: Skewness and kurtosis of simulated affine jump-diffusion models

[11] It is worth noting that we applied Kolmogorov-Smirnov tests to see whether there were significant differences between the model-based densities and the approximating ones. In almost all cases we rejected the null of identical distributions. These statistics reflect overall fit, however, while we will focus on the parts of the distributions that matter for risk management.

Moneyness	Gaussian Jump (0.7, 0.8) $\bar{L}_a$	$\bar{L}_r$	(0.8, 0.9) $\bar{L}_a$	$\bar{L}_r$	(0.97, 1.03) $\bar{L}_a$	$\bar{L}_r$	Exponential Jump (0.7, 0.8) $\bar{L}_a$	$\bar{L}_r$	(0.8, 0.9) $\bar{L}_a$	$\bar{L}_r$	(0.97, 1.03) $\bar{L}_a$	$\bar{L}_r$
						1 month						
NIG	0.0000	1.1505	0.0000	1.5409	0.0014	0.2153	0.0001	0.2961	0.0002	0.2903	0.0019	0.2226
VG	0.0000	1.2690	0.0001	2.8927	0.0022	0.3272	0.0001	0.3082	0.0002	0.4270	0.0032	0.2979
						6 months						
NIG	0.0001	1.6900	0.0002	0.8589	0.0015	0.0434	0.0003	0.2480	0.0010	0.2590	0.0023	0.0645
VG	0.0000	1.5052	0.0003	0.8700	0.0019	0.0605	0.0003	0.2299	0.0014	0.3611	0.0037	0.1121

Table 2: Mean pricing errors for call options based on simulated Affine Jump-Diffusion model.

4 Empirical Applications

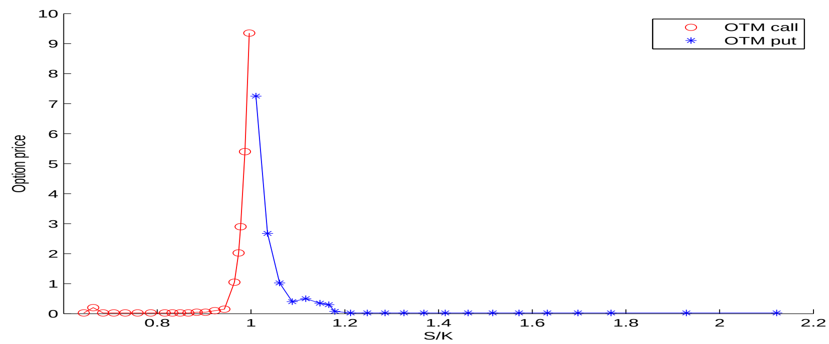
In this section, we will cover empirical applications in option pricing and value-at-risk forecasting. A subsection is devoted to each topic.

4.1 Option Pricing

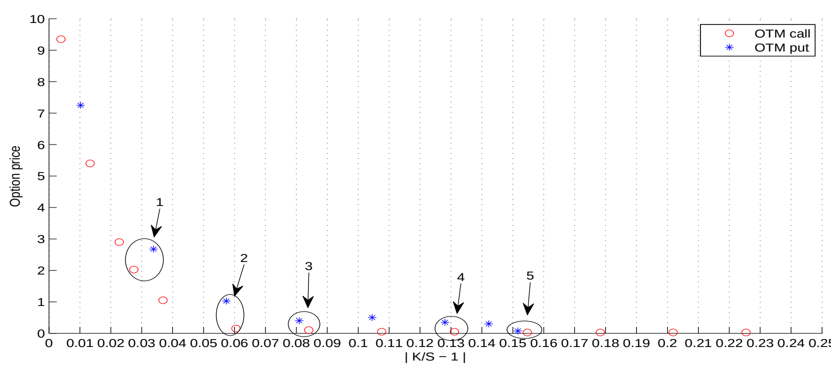
We start from the observation that a good method of option pricing should be able to price contracts in situations where only a few contracts are traded. We therefore design experiments where we compute risk neutral moments using a smaller set of option contracts than is available. We then price contracts which are not used to compute the moments via the approximate densities. This means we use a subset of existing market prices to extract risk neutral moments and another set of existing market prices to appraise the accuracy of the approximate option prices. Hence, we will examine, through an experimental design, how data sparseness affects option pricing via VG, NIG, and GST density approximations. We do this in such a way that we can appraise how well our methods perform to price options that are deep out-of-the-money versus options that are not. The former is the most challenging task to achieve for any method - and we show that we do very well. It will be helpful to explain the empirical investigation first with an illustrative example - which is covered first - followed by a full sample implementation.

4.1.1 An Illustrative Empirical Case

We start with an illustrative example and then proceed to a full sample formal analysis. To illustrate the design we use SPX options (European options on the S&P 500 index) priced on September 23, 2009 (which is a Wednesday) with 7 days to maturity (i.e., September 30, 2009) to illustrate the procedure. There are 42 call options and 42 put options with 7 days to maturity on 2009/9/23. The data consists of 20 OTM calls and 22 OTM puts. Figure 4(a) plots all the available OTM options. We then focus on the options with moneyness (i.e., the ratio of index level S and strike price K) between 0.8 and 1.2, which are displayed in Figure 4(b). Since the valuation of formulae (7), (8) and (9) require the calls and puts to have the same (or similar) distance from strike price to the index level in order to mitigate estimation error (recall the discussion in footnote 6), we use $|K/S-1|$ as x-axis in Figure 4(b). To evaluate the option pricing via the VG, NIG, and GST



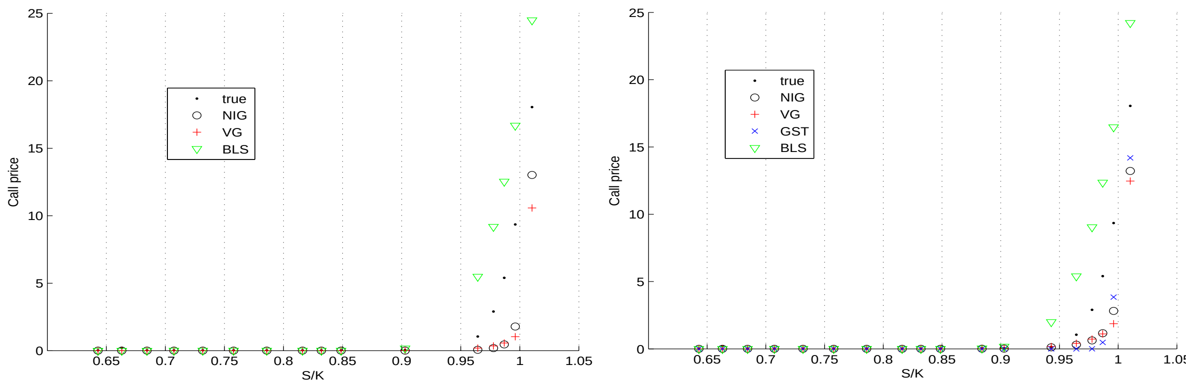
(a) OTM calls/puts



(b) Data used

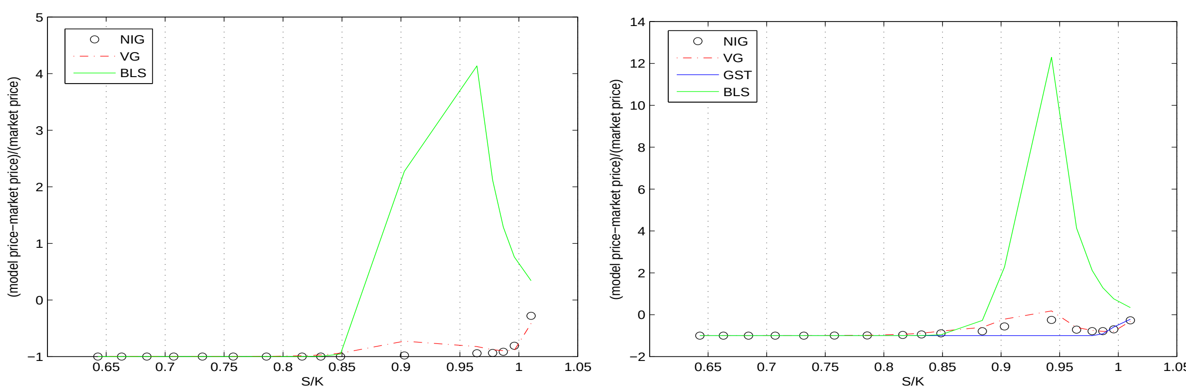
Figure 4: OTM S&P 500 index calls and puts, 2009/9/23, expired on 2009/9/30

approximations and assess how data sparseness affects pricing accuracy, we consider two strategies: (1) select the call/put pairs with $|K/S-1|$ closest to 0.03, 0.06, 0.09, 0.12, and 0.15, and (2) select the call/put pairs with $|K/S-1|$ closest to 0.03, 0.09, and 0.15. Therefore we pick call/put pairs in Circles 1-5 in Figure 4(b) for Strategy 1, and call/put pairs in Circles 1, 3, and 5 for Strategy 2. For each strategy, we derive the approximating densities – the VG, NIG, and GST distributions, and then price all the 'unused' call options.[12]



(a) Strategy 1 (b) Strategy 2

Figure 5: Price Call options, 2009/9/23, expired on 2009/9/30



(a) Strategy 1 (b) Strategy 2

Figure 6: (Model - Market Price)/(Market Price) Call options, 2009/9/23, expired on 2009/9/30

To appraise the performance of the VG, NIG, and GST approximations, we use the price computed from the Black-Scholes model as a benchmark - which will be denoted BLS in the tables. The volatility that enters the Black-Scholes formula is the Black-Scholes implied volatility of the previous-day (2009/9/22 in this case) at-the-money (henceforth denoted ATM - defined as S/K

[12]We do not consider call options with moneyness greater than 1.03, because they are not liquid, i.e. infrequently traded.

between 0.97 and 1.03) call option which has the shortest time-to-maturity and strike price closest to the index level.[13]

Figure 5 plots the observed market prices of the 'unused' call options and the prices derived from the VG, NIG, and GST approximations using Strategies 1 and 2. Table 3 reports the pricing error measured by L_a and L_r (defined in (16) where we replace 'true' by 'market', i.e. market options data). Suppose the range of moneyness of call options that are used for pricing is $[a, b]$. We divide all the 'unused' calls into three groups: Group 1 contains 'unused' options with moneyness within $[a, b]$, Group 2 contains 'unused' options with moneyness less than a, and Group 3 contains 'unused' calls with moneyness bigger than b. In particular:

- Strategy 1: $a = 0.8660$ and $b = 0.9733$. There are 2 points in $[a, b]$, 10 points in '$< a$', 4 points in '$> b$', a total of 16 unused points.

- Strategy 2: $a = 0.8660$ and $b = 0.9733$. There are 4 points in $[a, b]$, 10 points in '$< a$', 4 points in '$> b$', a total of 18 unused points.

The three groups are labeled as $[a, b]$, $< a$, and $> b$ respectively in Table 3. The following observations emerge from examining Table 3:

- The GST approximation is not feasible using Strategy 1, while it is for Strategy 2.

- The pricing errors are relatively small in strategy 2. In other words, the accuracy is *improved* when we use *fewer* options to derive the NIG, VG, and GST approximations.

- Overall, the VG approximation is best, though Strategy 1 picks NIG and strategy 2 picks GST for group 3 – perhaps in light of a small sample size in that group.

4.1.2 Full Sample Analysis

We examine now all the SPX options from January 4, 1996 to October 30, 2009 which are priced on *Wednesdays*.[14] For each Wednesday and time-to-maturity T_0, we pick the ATM call option with

[13] This is considered the most heavily traded and accurately priced contract, see Bates (1995) and Chernov and Ghysels (2000).

[14] We pick Wednesday - as this is common in the empirical option pricing literature - since it avoids some of the day-of-the-week effects that occur particularly on Fridays and Mondays.

	Overall		$[a,b]$		$< a$		$> b$	
	L_a	L_r	L_a	L_r	L_a	L_r	L_a	L_r
				Strategy 1				
BLS	3.4794	1.5616	3.0719	3.3376	0.0675	0.9949	6.6103	1.3079
NIG	2.6837	0.9449	0.6992	0.9606	0.0675	0.9998	5.3434	0.7816
VG	3.1201	0.9194	0.6105	0.7757	0.0674	0.9893	6.2244	0.7962
GST	NA	NA	NA	NA	NA	NA	NA	NA
				Strategy 2				
BLS	3.3091	3.2507	2.3598	6.5854	0.0675	0.9949	6.6103	1.3079
NIG	2.2326	0.8451	0.3739	0.6122	0.0673	0.9782	4.7200	0.6685
VG	2.4797	0.8147	0.3193	0.4496	0.0672	0.9597	5.2494	0.6942
GST	2.0950	0.9493	0.5315	1.0000	0.0675	1.0000	4.4110	0.7454

Table 3: Pricing S&P 500 index call options on 2009/9/23

strike price closest to the index level, and then calculate its Black-Scholes implied volatility (IV). We therefore obtain a sample of IV's for time-to-maturity T_0. Based on this sample we obtain the empirical first and third quartiles, i.e., Q_1 and Q_3. Denote by $D(1, T_0)$ all the Wednesday calls whose IV is less than Q_1, $D(2, T_0)$ the Wednesday calls whose IV is between Q_1 and Q_3, and $D(3, T_0)$ the set of Wednesday calls whose IV is larger than Q_3. Based on observations in Section 3.2, we consider T_0 as one week (i.e., 5$\sim$14 days), one month (i.e., 17$\sim$31 days), and three months (i.e., 81$\sim$94 days). Therefore we end up with 9 different combinations. Table 4 lists, for each combination, the number of days which yield skewness-kurtosis combination within the feasible domain of the VG, NIG, GST distributions using Strategy 1 and 2 respectively.[15] Namely, the triplets in Table 4 are the numbers of days feasible for the VG, NIG, and GST distributions. For instance, the triplet $(27, 21, 0)$ means that the VG distribution can model 27 dates out of 46 (i.e., first quartile yields 46 data points), NIG can model 21 data points while none of the 46 days yields a skewness-kurtosis combination in the feasible domain of the GST distribution. We note that the GST is the most restrictive - as expected. We therefore consider two sample configurations - one where all three distributions are applicable. This is a relatively small sample that arguably gives unfair advantage to the more restrictive GST distribution. Hence, we also examine a sample where only the VG and NIG distributions are feasible. This is a larger and hence more reliable sample. We analyze the pricing errors for days identified in Table 4 starting with the most restrictive sample configurations involving all days where GST, VG and NIG distributions apply. We focus on the

[15]We removed days which have fewer than 5 calls or 5 puts.

T_0	Total	Strategy	$D(1, T_0)$	$D(2, T_0)$	$D(3, T_0)$
$5 \sim 14$	184	1	(27, 21, 0)	(66, 55, 3)	(41, 39, 19)
		2	(34, 28, 2)	(78, 76, 32)	(43, 40, 32)
$17 \sim 31$	352	1	(34, 22, 1)	(114, 88, 15)	(60, 48, 25)
		2	(65, 55, 2)	(142, 126, 40)	(62, 53, 31)
$81 \sim 94$	255	1	(16, 7, 0)	(52, 36, 17)	(5, 4, 4)
		2	(22, 10, 0)	(43, 30, 14)	(4, 4, 3)

Table 4: Number of days which yield skewness-kurtosis combination within the feasible domain of the VG, NIG, GST distributions, SPX option (1996∼2009)

set $D(2, T_0)$ with $T_0 = 17 \sim 31$ and $81 \sim 94$, $D(3, T_0)$ with $T_0 = 5 \sim 14$ and $17 \sim 31$ because they have relatively more observations than the other cells. For days identified in each of the cells, we repeat the analysis described for the single day 2009/9/23, and derive L_a and L_r. We then average L_a and L_r within each cell. These averages, denoted by $\bar{L}_a$ and $\bar{L}_r$ respectively, are reported in Table 6.

What do we learn from Table 6? For median volatility regimes ($D(2, .)$) we note that option pricing based on Black Scholes implieds perform best in absolute terms (L_a) but not in relative terms. This observation applies to both strategies and across moneyness. In terms of relative pricing errors the picture is quite different. Namely, in terms of relative pricing errors, we find that either the VG approximation or the GST one is the best. It should also be noted, however, that the NIG and VG approximations typically yield similar relative pricing errors. The GST appears to perform better for ATM options $> b$. It is also worth noting that Strategies 1 and 2 yield comparable pricing errors. This is rather impressive as it means that we can use less data (i.e. contracts) and still find similar pricing errors.

In turbulent times ($D(3, .)$) we note that BLS starts to perform poorly both in absolute and relative pricing error terms. Overall - all maturities pooled - we find GST to perform best in absolute terms, but the VG approximation is again clearly dominant - with NIG quite similar in terms of performance. It is clear that the distributional approximations make a big difference compared to Black Scholes implied option pricing as the pricing errors are substantially larger for BLS. In terms of relative pricing errors, the differences between BLS and distributional approximations are particularly important.

What happens when we drop the GST distribution - which is most restrictive in terms of feasibility? The results - where we remove the GST distribution - appear in Table 7. Given the different sample configuration, we look at a larger set of combinations, namely we look at the sets $D(1, T_0)$ with $T_0 = 5 \sim 14$ and $17 \sim 31$, $D(2, T_0)$ with $T_0 = 5 \sim 14$, $17 \sim 31$ and $81 \sim 94$, and finally $D(3, T_0)$ with $T_0 = 5 \sim 14$ and $17 \sim 31$. Hence, we not only looking at a larger data set within each cell (cfr. Table 4) but also a broader set of maturity/moneyness/volatility-state combinations. We start with the cases that are common between Tables 6 and 7. We note that whenever GST is best when feasible the NIG distribution fills the void. However, since NIG and VG are again often close - it is also the case that VG is equally appealing. Overall, we find quantitatively the same results as in the more constraint sample where GST is feasible. Namely, Strategy 1 and 2 appear similar, the absolute pricing errors are dominated by BLS in the low volatility regime - as before - but in terms of relative pricing errors there are clear gains across all moneynesses and maturities. Moreover, once we move to high volatility states BLS no longer performs well.

4.2 Value at Risk

Last but certainly not least, we consider value-at-risk (VaR) forecasting. We use the notation introduced in Section 3.1. Hence, $-R_t(\tau)$ can be viewed as loss from time t to $t+\tau$. The $\tau-$period-ahead VaR forecast at level $100(1-\alpha)\%$ is then defined as $VaR_t(\tau, \alpha) \equiv \inf_y \{y : P^Q(-R_t(\tau) \geq y|\mathcal{F}_t) \leq \alpha\}$, and P^Q implies that the VaR is evaluated under risk neutral measure.

Denote by $\widehat{VaR}_t(\tau, \alpha)$ the estimate of $VaR_t(\tau, \alpha)$, where the estimated is obtained using the VG, NIG, and GST distributions. We consider again the SPX options from January 4, 1996 to October 30, 2009, with 7 days and 14 days to maturity. Hence, we do as if we hold the market portfolio, or something similar to that and compute its VaR forecast. Figures 7 and 8 plot the losses $\{-R_t(\tau)\}$ over the life time of options, i.e., $\tau = 7$ and 14 with 229 and 172 observations respectively. The vertical line divides the series into two subperiods: (1) Prior to August 2008 and (2) the Financial crisis period.[16] Superimposed, we also plot the out-of-sample VaR forecasts at 95% level.

[16]We also remove days which have only 2 pairs of contracts.

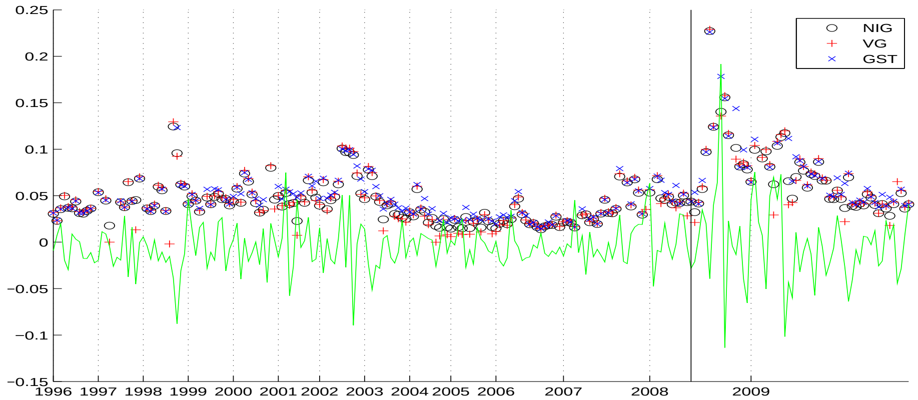


Figure 7: 95% VaR forecast for S&P 500 index over 7 days

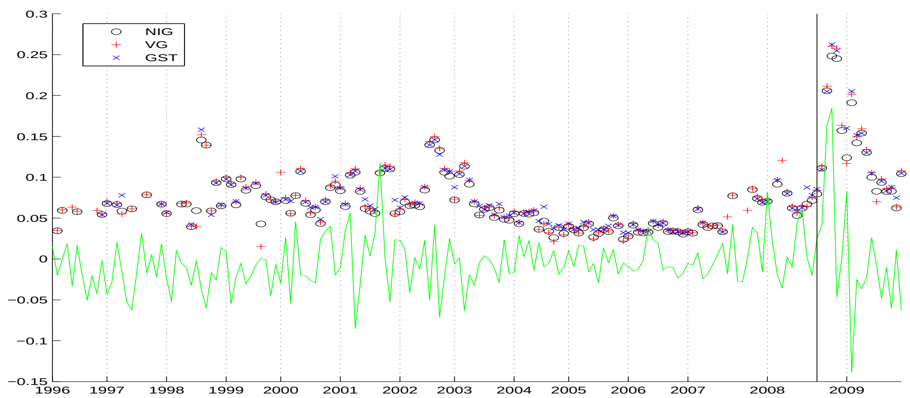


Figure 8: 95% VaR forecast for S&P 500 index over 14 days

Define the indicator of VaR failure as $I_t(\tau, \alpha) = I_{\{R_t(\tau) < -\widehat{VaR}_t(\tau,\alpha)\}}$. Then $\sum_{t=1}^{T} I_t(\tau, \alpha)$ is the total number of exceedances of VaR forecasts, where T is the number of forecasts. We report in Table 5 the empirical risk level (see the columns with label 'Emp. Risk') - which is the total exceedances divided by T - for the full sample and the two subsamples: prior to the financial crisis and during the crisis. For nominal risk level 0.05 and $\tau = 7$, we note that NIG overestimates the risk for the high volatility period but underestimates the risk for the relative low volatility times. In contrast, the VG and GST distributions underestimate risk. Overall, the NIG provides a more accurate estimate of the risk. For the combinations ($\alpha = 0.025$, $\tau = 7$) and ($\alpha = 0.05$, $\tau = 14$), all the three distributions underestimate the risk. VG yields the smallest departure from the nominal level for $\alpha = 0.025$ and $\tau = 7$ while NIG does for $\alpha = 0.05$ and $\tau = 14$.

	Uncond.	Tail	Emp. Risk	Uncond.	Tail	Emp. Risk	Uncond.	Tail	Emp. Risk
	Full Sample			Prior to August 2008			Financial Crisis Period		
				$\alpha = 0.025$, $\tau = 7$					
NIG	0.0951	<0.01	0.0094	0.2729	0.0029	0.0127	NaN	0.2485	0
VG	0.5228	<0.01	**0.0185**	1.0000	<0.01	**0.0250**	NaN	0.2422	0
GST	0.0347	0.0172	0.0053	0.1085	0.0509	0.0071	NaN	0.2892	0
				$\alpha = 0.05$, $\tau = 7$					
NIG	0.8365	<0.01	**0.0469**	0.7378	<0.01	**0.0443**	0.8788	<0.01	**0.0545**
VG	0.7126	<0.01	0.0556	0.4842	<0.01	0.0625	0.6060	<0.01	0.0357
GST	0.6082	<0.01	0.0421	0.6775	<0.01	0.0426	0.7609	<0.01	0.0408
				$\alpha = 0.05$, $\tau = 14$					
NIG	0.0584	<0.01	**0.0201**	0.1128	<0.01	**0.0229**	NaN	0.3972	0
VG	0.0467	<0.01	0.0194	0.0909	<0.01	0.0219	NaN	0.3972	0
GST	0.0413	<0.01	0.0159	0.0798	<0.01	0.0182	NaN	0.4401	0

Table 5: VaR Backtesting

To backtest the accuracy of VaR forecasts, we consider two tests: the unconditional coverage test of Christoffersen (1998), and the tail density test of Berkowitz (2001). In the unconditional coverage test, the null hypothesis is $E(I_t(\tau,\alpha)) = \alpha$ and the alternative is $E(I_t(\tau,\alpha)) \neq \alpha$ under an independence assumption of $\{I_t(\tau,\alpha)\}_{t=1}^{T}$. The tail density test of Berkowitz (2001) works with the left tail of the return distribution, or expected shortfall $E(R_t(\tau)|R_t(\tau) < -\widehat{VaR}_t(\tau,\alpha))$. Define $z_t = \Phi^{-1}(\hat{F}(R_t(\tau)))$, where $\hat{F}$ is either NIG, or VG, or GST and Φ is cdf of standard normal distribution. Hence, examining $E(R_t(\tau)|R_t(\tau) < -\widehat{VaR}_t(\tau,\alpha))$ translates into examining $E(z_t|z_t < \Phi^{-1}(\alpha))$. Further define $z_t^* = \min(z_t, \Phi^{-1}(\alpha))$. The tail density test is a likelihood ratio test based on the fact that z_t^* is a censored normal variable if $\hat{F}$ provides a good fit. The columns labeled 'Uncond.' and 'Tail' in Table 5 are the p-values of unconditional coverage test and tail density test respectively.

Note that the unconditional coverage test is not applicable for ($\alpha = 0.025$, $\tau = 7$) and ($\alpha = 0.05$, $\tau = 14$) during the Crisis, because the exceedance of VaR forecast is 0. The p-values of the unconditional coverage test, when applicable, indicate that there is no significant departure from adequacy except for the GST at the short maturity and $\alpha = 0.025$. The tail density test overall provides significant results for the full sample and the period prior to Crisis, which is probably caused by the strong serial correlation in the data. Particularly the tail density test seems to tell us that the distributional approximations do not do so well as far as VaR goes.

5 Discussion

The primary focus of the paper is to study the Generalized Hyperbolic family of distributions, which covers a wide range of distributions commonly used in the literature to model the financial risk. We specifically study densities determined by the first four moments, that is the NIG distribution, the VG distribution and the GST distribution. We study the properties of the NIG, VG and GST distributions in terms of tail behavior and feasible domain. Among them, the VG distribution admits the largest possible combinations of skewness and (excess) kurtosis while the GST is the most restrictive.

In the context of risk management we analyze option pricing - assuming an unknown risk neutral density which is approximated by the class of distributions we study as well as Gram-Charlier and Edgeworth expansions. We show, through numerical and empirical evidence, that the VG and NIG distributions are roughly similar as candidate approximating risk neutral densities for option pricing applications. The GST - with a more restrictive feasible domain - also performs well when it can be applied.

We also find that the NIG and VG approximations work extremely well in terms of option pricing in particular during high volatility periods, compared to the industry standard of using previous period ATM implied Black Scholes volatilities. We also find that with only a small set of quoted contracts, we can extrapolate and price very well options not used in the computations of risk neutral moments and hence options not used to compute the approximations.

Nevertheless, the NIG and VG still have their limitations, as their feasible regions are still not large enough to cover all option pricing applications we encountered. This fact should prompt us to think about a new family of distributions which can accommodate a wider range of skewness and kurtosis, and we leave this as a topic for future research. In addition, we also find from the tail density test that the distributional approximations do not do so well as far as VaR goes.

Hamburger's theory (see Widder (1946), Chihara (1989)) proves the existence of a distribution for *any* given skewness-kurtosis combination, but unfortunately it does not show how to construct the density. If we think of the ARCH class of models - mixtures driven by time-varying volatility - one realizes that there are many ways to formulate mixture models. While this is an attractive idea, it would take us far away from the current paper.

	Strategy 1								Strategy 2							
	Overall		$[a,b]$		$< a$		$> b$		Overall		$[a,b]$		$< a$		$> b$	
	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$
								$D(2,T_0)$, T_0=17 $\sim$ 31								
BLS	4.96	1.92	5.06	3.19	0.38	2.34	6.46	0.23	4.26	2.50	4.09	3.25	0.50	3.60	5.57	0.24
NIG	7.05	0.64	3.52	0.71	0.19	0.92	9.45	0.40	5.06	0.63	2.67	0.70	0.18	0.94	7.28	0.35
VG	7.59	0.61	3.37	0.60	0.17	0.84	10.28	0.44	5.49	0.58	2.54	0.56	0.17	0.86	8.05	0.40
GST	6.27	0.70	4.10	0.94	0.20	1.00	8.18	0.35	4.61	0.73	3.11	0.92	0.19	1.00	6.34	0.32
								$D(2,T_0)$, T_0=81 $\sim$ 94								
BLS	5.14	1.68	5.79	0.47	3.04	4.42	5.64	0.10	5.84	1.86	6.58	0.58	3.57	5.11	6.19	0.11
NIG	9.62	0.55	9.45	0.58	0.78	0.89	11.99	0.25	9.29	0.58	9.09	0.59	0.83	0.90	11.88	0.24
VG	10.01	0.51	10.04	0.58	0.64	0.74	12.41	0.26	9.63	0.54	9.59	0.59	0.70	0.78	12.21	0.25
GST	9.25	0.58	9.30	0.60	0.89	0.98	11.39	0.23	9.13	0.60	9.02	0.62	0.93	0.97	11.60	0.23
								$D(3,T_0)$, T_0=5 $\sim$ 14								
BLS	4.27	6.03	5.93	5.63	1.10	5.57	8.44	0.40	4.07	4.97	5.03	5.19	0.73	4.02	7.37	0.43
NIG	3.43	0.81	2.29	0.67	0.19	0.94	7.92	0.43	3.17	0.78	1.83	0.66	0.19	0.94	7.03	0.39
VG	3.73	0.77	2.23	0.60	0.18	0.89	8.61	0.46	3.42	0.75	1.77	0.57	0.18	0.89	7.63	0.42
GST	3.13	0.91	2.59	0.92	0.20	1.00	7.05	0.40	2.87	0.89	2.10	0.90	0.20	1.00	6.21	0.35
								$D(3,T_0)$, T_0=17 $\sim$ 31								
BLS	6.92	7.97	9.77	2.51	2.81	11.77	10.34	0.32	6.04	6.45	8.05	2.92	2.03	9.28	8.99	0.30
NIG	4.70	0.79	4.88	0.68	0.29	0.96	8.72	0.28	4.23	0.77	4.05	0.65	0.25	0.97	7.63	0.26
VG	4.94	0.74	4.97	0.63	0.26	0.89	9.26	0.30	4.46	0.72	4.16	0.58	0.23	0.91	8.10	0.28
GST	4.50	0.83	5.03	0.75	0.31	1.00	8.08	0.26	4.03	0.81	4.16	0.74	0.27	1.00	7.08	0.24

Table 6: Pricing S&P 500 index call options (1996$\sim$2009) - GST, VG and NIG

	Strategy 1								Strategy 2							
	Overall		$[a, b]$		$< a$		$> b$		Overall		$[a, b]$		$< a$		$> b$	
	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$	$\bar{L}_a$	$\bar{L}_r$
$D(1, T_0)$, T_0=5 $\sim$ 14																
BLS	2.38	3.43	0.71	5.36	0.19	1.00	2.92	1.67	2.05	2.94	0.56	4.31	0.16	1.00	2.63	1.46
NIG	2.49	0.70	0.15	0.77	0.19	0.99	3.18	0.62	2.27	0.69	0.15	0.78	0.16	0.99	2.98	0.59
VG	8.39	0.75	0.13	0.67	0.18	0.98	10.70	0.75	6.41	0.73	0.13	0.75	0.16	0.98	8.38	0.68
$D(1, T_0)$, T_0=17 $\sim$ 31																
BLS	4.09	4.09	2.74	6.40	0.16	0.95	4.78	0.78	3.53	3.41	2.30	5.22	0.17	0.97	4.25	0.67
NIG	4.29	0.75	1.01	0.93	0.16	0.99	5.36	0.61	3.38	0.70	0.74	0.86	0.17	0.99	4.45	0.54
VG	5.54	0.70	0.88	0.77	0.16	0.98	6.97	0.63	4.15	0.62	0.60	0.66	0.17	0.98	5.47	0.55
$D(2, T_0)$, T_0=5 $\sim$ 14																
BLS	2.15	2.33	1.33	3.42	0.15	1.21	3.28	0.43	2.16	3.13	1.24	4.39	0.15	1.17	3.26	0.46
NIG	4.01	0.79	0.64	0.85	0.14	0.98	6.16	0.60	3.30	0.70	0.48	0.72	0.14	0.96	5.08	0.52
VG	5.25	0.74	0.55	0.69	0.14	0.96	7.94	0.64	3.82	0.66	0.39	0.59	0.14	0.94	5.86	0.54
$D(2, T_0)$, T_0=17 $\sim$ 31																
BLS	4.65	4.01	4.87	5.26	0.53	4.95	6.24	0.29	5.12	5.88	5.03	7.87	0.49	4.54	6.68	0.34
NIG	5.86	0.76	3.55	0.90	0.18	0.98	8.47	0.44	4.59	0.72	2.61	0.83	0.18	0.98	6.83	0.36
VG	6.27	0.72	3.28	0.78	0.18	0.94	9.23	0.48	4.96	0.65	2.41	0.67	0.18	0.94	7.54	0.41
$D(2, T_0)$, T_0=81 $\sim$ 94																
BLS	6.71	3.69	8.05	0.80	4.36	8.29	7.77	0.15	6.44	3.18	7.26	0.73	4.20	8.33	7.16	0.13
NIG	8.10	0.60	8.77	0.60	0.71	0.95	10.19	0.21	7.85	0.59	8.17	0.63	0.65	0.95	9.67	0.21
VG	8.38	0.56	9.20	0.59	0.62	0.85	10.50	0.22	8.10	0.56	8.53	0.62	0.56	0.85	9.93	0.22
$D(3, T_0)$, T_0=5 $\sim$ 14																
BLS	3.86	4.73	4.74	5.19	0.69	4.13	7.57	0.44	3.85	4.75	4.73	5.11	0.68	4.04	7.31	0.43
NIG	3.68	0.85	2.09	0.79	0.19	0.97	7.95	0.46	3.06	0.80	1.69	0.71	0.19	0.95	6.79	0.39
VG	3.95	0.81	1.98	0.69	0.18	0.94	8.57	0.49	3.30	0.76	1.61	0.61	0.18	0.91	7.38	0.43
$D(3, T_0)$, T_0=17 $\sim$ 31																
BLS	7.00	8.96	9.12	4.40	2.29	11.67	10.82	0.36	7.00	8.54	9.04	5.74	2.01	11.27	10.22	0.36
NIG	4.92	0.81	4.82	0.76	0.26	0.98	8.74	0.30	4.18	0.80	3.91	0.73	0.25	0.98	7.50	0.26
VG	5.18	0.77	4.83	0.69	0.24	0.93	9.32	0.32	4.38	0.75	3.94	0.65	0.23	0.93	7.95	0.28

Table 7: Pricing S&P 500 index call options (1996∼2009) - VG and NIG

References

Aas, K., and I. Haff (2006): "The Generalized Hyperbolic Skew Student's t-Distribution," *Journal of Financial Econometrics*, 4(2), 275–309.

Aas, K., I. Haff, X. Dimakos, and N. Center (2006): "Risk estimation using the multivariate Normal Inverse Gaussian distribution," *Journal of Risk*, 8(2).

Aït-Sahalia, Y., and A. Lo (2000): "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics*, 94, 9–51.

Avramidis, A., and P. L'Ecuyer (2006): "Efficient Monte Carlo and quasi-Monte Carlo option pricing under the variance-gamma model," *Management Science*, 52(12), 1930–1994.

Bakshi, G., N. Kapadia, and D. Madan (2003): "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *The Review of Financial Studies*, 16(1), 101–143.

Bali, T., and P. Theodossiou (2007): "A conditional-SGT-VaR approach with alternative GARCH models," *Annals of Operations Research*, 151(1), 241–267.

Bams, D., T. Lehnert, and C. Wolff (2005): "An evaluation framework for alternative VaR-models," *Journal of International Money and Finance*, 24(6), 944–958.

Barndorff-Nielsen, O. (1977): "Exponentially decreasing distributions for the logarithm of particle size," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pp. 401–419.

——— (1997a): "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling," *Scandinavian Journal of Statistics*, 24(1), 1–13.

——— (1997b): "Processes of normal inverse Gaussian type," *Finance and Stochastics*, 2(1), 41–68.

Barndorff-Nielsen, O., and N. Shephard (2001): "Normal modified stable processes.," *The- ory of Probability and Mathematical Statistics*, 65, 1–20.

Barndorff-Nielsen, O., and R. Stelzer (2005): "Absolute Moments of Generalized Hyperbolic Distributions and Approximate Scaling of Normal Inverse Gaussian Levy Processes," *Scandinavian Journal of Statistics*, 32(4), 617–637.

Barton, D., and K. Dennis (1952): "The Conditions Under Which Gram-Charlier and Edgeworth Curves are Positive Definite and Unimodal," *Biometrika*, 39(3/4), 425–427.

Bates, D. (1995): "Testing option pricing models," .

Battalio, R., and P. Schultz (2006): "Options and the bubble," *The Journal of Finance*, 61, 2071–2102.

Bauwens, L., and S. Laurent (2005): "A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models," *Journal of Business and Economic Statistics*, 23(3), 346–354.

Berkowitz, J. (2001): "Testing density forecasts, with applications to risk management," *Journal of Business and Economic Statistics*, 19(4), 465–474.

Bibby, B., and M. Sørensen (2003): “Hyperbolic processes in finance,” *Handbook of Heavy Tailed Distributions in Finance*, pp. 211–248.

Carr, P., H. Geman, D. Madan, and M. Yor (2002): “The Fine Structure of Asset Returns: An Empirical Investigation*,” *The Journal of Business*, 75(2), 305–332.

Carr, P., and D. Madan (1999): “Option valuation using the fast Fourier transform,” *Journal of Computational Finance*, 2(4), 61–73.

Chen, Y., W. Hardle, and S. Jeong (2008): “Nonparametric risk management with generalized hyperbolic distributions,” *Journal of the American Statistical Association*, 103(483), 910–923.

Chernov, M., and E. Ghysels (2000): “A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation,” *Journal of Financial Economics*, 56(3), 407–458.

Chihara, T. (1989): “Hamburger moment problems and orthogonal polynomials,” *American Mathematical Society*, 315(1).

Christoffersen, P. (1998): “Evaluating interval forecasts,” *International economic review*, 39(4), 841–862.

Conrad, J., R. Dittmar, and E. Ghysels (2010): “Ex Ante Skewness and Expected Stock Returns,” Discussion paper, Working Paper, University of North Carolina at Chapell Hill.

Dennis, P., and S. Mayhew (2002): “Risk-neutral skewness: Evidence from stock options,” *Journal of Financial and Quantitative Analysis*, 37(03), 471–493.

Duffie, D., D. Filipovic, and W. Schachermayer (2003): “Affine processes and applications in finance,” *The Annals of Applied Probability*, 13(3), 984–1053.

Duffie, D., J. Pan, and K. Singleton (2000): “Transform Analysis and Asset Pricing for Affine Jump-diffusions,” *Econometrica*, 68(6), 1343–1376.

Eberlein, E. (2001a): “Application of generalized hyperbolic Lévy motions to finance,” in *Lévy processes: theory and applications*, ed. by O. Barndorff-Nielsen, T. Mikosch, and S. Resnick, pp. 319–337. Birkhäuser Verlag.

——— (2001b): “Application of generalized hyperbolic Levy motions to finance,” *Levy processes: theory and applications*, pp. 319–336.

Eberlein, E., and E. Hammerstein (2004): “Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes,” in *Seminar on Stochastic Analysis, Random Fields, and Applications IV: Centro Stefano Franscini, Ascona, May 2002*, p. 221. Birkhauser.

Eberlein, E., and U. Keller (1995a): “Hyperbolic distributions in finance,” *Bernoulli*, 1(3), 281–299.

——— (1995b): “Hyperbolic distributions in finance,” *Bernoulli*, pp. 281–299.

Eberlein, E., U. Keller, and K. Prause (1998): “New insights into smile, mispricing, and value at risk: The hyperbolic model,” *Journal of Business*, 71(3), 371–405.

Eberlein, E., and K. Prause (2002): "The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures," in *Mathematical Finance - Bachelier Finance Congress 2000*, ed. by H. Geman, D. Madan, S. Pliska, and T. Vorst, p. 245267. Springer Verlag.

Eriksson, A., L. Forsberg, and E. Ghysels (2004): "Approximating the Probability Distribution of Functions of Random Variables: A New Approach," Discussion paper CIRANO.

Eriksson, A., E. Ghysels, and F. Wang (2009): "The Normal Inverse Gaussian Distribution and the Pricing of Derivatives," *The Journal of Derivatives*, 16(3), 23–37.

Figlewski, S. (2007): "The Implied Risk Neutral Density for the US Market Portfolio," Discussion Paper NYU.

Frecka, T., and W. Hopwood (1983): "The Effects of Outliers on the Cross-Sectional Distributional Properties of Financial Ratios," *The Accounting Review*, 58(1), 115–128.

Haas, M., and C. Pigorsch (2007): "Financial Economics, Fat-tailed Distributions," *Encyclopedia of Complexity and Systems Science*, 4.

Hirsa, A., and D. Madan (2004): "Pricing American options under variance gamma," *Journal of Computational Finance*, 7(2), 63–80.

Kalemanova, A. (2007): "The normal inverse Gaussian distribution for synthetic CDO pricing," *The Journal of Derivatives*, 14(3), 80–93.

Keller-Ressel, M. (2011): "Moment Explosions and Long-Term Behavior of Affine Stochastic Volatility Models," *Mathematical Finance*, 21(1), 73–98.

Konikov, M., and D. Madan (2002): "Option pricing using variance gamma markov chains," *Review of Derivatives Research*, 5(1), 81–115.

Kuester, K., S. Mittnik, and M. Paolella (2006): "Value-at-risk prediction: A comparison of alternative strategies," *Journal of Financial Econometrics*, 4(1), 53–89.

Lee, R. (2004): "Option pricing by transform methods: extensions, unification and error control," *Journal of Computational Finance*, 7(3), 51–86.

Madan, D., P. Carr, and E. Chang (1998): "The variance gamma process and option pricing," *European Finance Review*, 2(1), 79–105.

Madan, D., and F. Milne (1994): "Contingent Claims Valued and Hedged by Pricing and Investing in a basis," *Mathematical Finance*, 4, 223–245.

Madan, D., and E. Seneta (1990): "The Variance Gamma (VG) Model for Share Market Returns," *Journal of Business*, 63(4), 511.

Moosbrucker, T. (2006): "Pricing CDOs with correlated variance gamma distributions," *Research Report, Department of Banking, University of Cologne.*

Prause, K. (1999): *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures.* Freiburg (Breisgau).

Ribeiro, C., and N. Webber (2004): "Valuing path-dependent options in the variance-gamma model by Monte Carlo with a gamma bridge," *Journal of Computational Finance*, 7(2), 81–100.

Rompolis, L., and E. Tzavalis (2007): "Retrieving risk neutral densities based on risk neutral moments through a Gram–Charlier series expansion," *Mathematical and Computer Modelling*, 46(1-2), 225–234.

——— (2008): "Recovering Risk Neutral Densities from Option Prices: A New Approach," *Journal of Financial and Quantitative Analysis*, 43(4), 1037–1054.

Rosenberg, J., and T. Schuermann (2006): "A general approach to integrated risk management with skewed, fat-tailed risks," *Journal of Financial Economics*, 79(3), 569–614.

Rubinstein, M. (1998): "Edgeworth Binomial Trees," *Journal of Derivatives*, 5, 20–27.

Rydberg, T. (1997): "The normal inverse Gaussian Levy process: simulation and approximation," *Communications in statistics. Stochastic models*, 13(4), 887–910.

Rydberg, T. (1999): "Generalized hyperbolic diffusion processes with applications in finance," *Mathematical Finance*, 9(2), 183–201.

Seneta, E. (2004): "Fitting the variance-gamma model to financial data," *Journal of Applied Probability*, pp. 177–187.

Theodossiou, P. (1998): "Financial data and the skewed generalized T distribution," *Management Science*, 44(12 Pt 1), 1650–1661.

Venter, J., and P. de Jongh (2002): "Risk estimation using the normal inverse Gaussian distribution," *Journal of Risk*, 4, 1–24.

Wang, S. (2000): "A class of distortion operators for pricing financial and insurance risks," *Journal of risk and insurance*, pp. 15–36.

Widder, D. (1946): *The Laplace transform.* Princeton Mathematical series, 6, Princeton University Press Princeton, NJ.

Technical Appendices

A Proofs

Proof of Proposition 2.2: Note that

$$m_n = \frac{2^{\lceil \frac{n}{2} \rceil} \bar{\gamma}^p b^{2\lceil \frac{n}{2} \rceil} \beta^{\tilde{n}}}{\sqrt{\pi} K_p(\bar{\gamma}) \bar{\alpha}^{p+\lceil \frac{n}{2} \rceil}} \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma(k + \lceil \frac{n}{2} \rceil + \frac{1}{2})}{\bar{\alpha}^k (2k+\tilde{n})!} K_{p+k+\lceil \frac{n}{2} \rceil}(\bar{\alpha}) \tag{A.1}$$

where $\tilde{n} \equiv n(\text{mod } 2)$ $\bar{\alpha} = b\alpha$, $\bar{\beta} = b\beta$, $\bar{\gamma} = \sqrt{\bar{\alpha}^2 - \bar{\beta}^2}$(see Barndorff-Nielsen and Stelzer (2005)). Expression (2) is derived from (A.1) together with the fact that

$$K_v(z) = \frac{z^v}{x^v} \sum_{k=0}^{\infty} \frac{1}{2^k k!} \frac{y^{2k}}{x^k} K_{v+k}(x), \quad z = \sqrt{x^2 - y^2}, x > 0, y > 0, v \in \mathbb{R}.$$

$$\sum_{k=0}^{\infty} \frac{\bar{\beta}^{2k}}{\bar{\alpha}^k} \frac{(2k+3)}{2^k k!} K_{p+k+2}(\bar{\alpha}) = K_{p+3}(\bar{\gamma}) \frac{\bar{\alpha}^{p+2} \bar{\beta}^2}{\bar{\gamma}^{p+3}} + K_{p+2}(\bar{\gamma}) \frac{3\bar{\alpha}^{p+2}}{\bar{\gamma}^{p+2}},$$

$$\sum_{k=0}^{\infty} \frac{\bar{\beta}^{2k}}{\bar{\alpha}^k} \frac{(2k+3)(2k+1)}{2^k k!} K_{p+k+2}(\bar{\alpha}) = K_{p+4}(\bar{\gamma}) \frac{\bar{\beta}^4 \bar{\alpha}^{p+2}}{\bar{\gamma}^{p+4}} + K_{p+3}(\bar{\gamma}) \frac{6\bar{\beta}^2 \bar{\alpha}^{p+2}}{\bar{\gamma}^{p+3}} + K_{p+2}(\bar{\gamma}) \frac{3\bar{\alpha}^{p+2}}{\bar{\gamma}^{p+2}}.$$

■

Proof of Proposition 2.4: (1) Use the fact that $\lim_{b \to 0} \frac{b^k K_{p+k}(\bar{\gamma})}{K_p(\bar{\gamma})} = \frac{2^k \Gamma(p+k)}{\gamma^k \Gamma(p)}$ for $k \in \mathbb{Z}^+$ and $p > 0$, and the Dominant Convergence Theorem.

(2) Note that (3) implies

$$S = \frac{2\rho(3+R)}{\sqrt{V}\alpha(1-R^2)}, \quad K = \frac{6(1+6R+R^2)}{V\alpha^2(1-R)^2(1+R)}$$

where $\rho = \frac{\beta}{\alpha}$ ($|\rho| < 1$) and $R = \rho^2 < 1$. Therefore,

$$\frac{3S^2}{2K} = \frac{R(3+R)^2}{(1+6R+R^2)(1+R)} \tag{A.2}$$

Define $C = \frac{3S^2}{2K}$ and $C \in (0,1)$. Note that $f(R) = \frac{R(3+R)^2}{(1+6R+R^2)(1+R)}$ is continuous and strictly increasing on $(0,1)$ with range $(0,1)$. Given S, K and $3S^2 < 2K$, there exists a unique $R \in (0,1)$ satisfying (A.2), or $(C-1)R^3 + (7C-6)R^2 + (7C-9)R + C = 0$. It follows immediately that

$$\alpha = \frac{2\rho(3+R)}{\sqrt{V}S(1-R^2)} = \frac{2\sqrt{R}(3+R)}{\sqrt{V}|S|(1-R^2)}, \quad \beta = \rho\alpha = \frac{2R(3+R)}{\sqrt{V}S(1-R^2)}$$
$$p = \frac{V(1-R)^2\alpha^2}{2(1+R)} = \frac{2R(3+R)^2}{S^2(1+R)^3}, \quad \mu = M - \frac{2\beta p}{\alpha^2 - \beta^2} = M - \frac{2\sqrt{V}R(3+R)}{S(1+R)^2}$$

■

Proof of Proposition 2.5 : (1) It follows from the fact that $\lim_{\gamma \to 0} \frac{K_{p+k}(\bar{\gamma})}{\gamma^k K_p(\bar{\gamma})} = \frac{b^k \Gamma(-p-k)}{2^k \Gamma(-p)}$ for $k \in \mathbb{Z}^+$ and $p < 0$, and an application of the Dominated Convergence theorem.

(2) Define $\rho = \frac{2b^2\beta^2}{(v-2)(v-4)}$. $0 < \rho < b^2\beta^2/12$. Expression (4) implies that

$$V = \frac{(1+\rho)b^2}{v-2}$$
$$S^2 = \left(3 + \frac{4(v-4)\rho}{v-6}\right)^2 \frac{2\rho}{(v-4)(1+\rho)^3}$$

$$K = \left[\frac{2(5v-22)(v-4)\rho^2}{(v-6)(v-8)} + \frac{8(v-4)\rho}{v-6} + 1\right]\frac{6}{(1+\rho)^2(v-4)}$$

and thus $b^2 = V(v-2)/(1+\rho)$, and $\beta^2 = \rho(1+\rho)(v-4)/(2V)$, where $\rho > 0$ and $v > 8$ are solutions of the following system of equations

$$\begin{array}{lcl} S^2 & = & \left(3 + \frac{4(v-4)\rho}{v-6}\right)^2 \frac{2\rho}{(v-4)(1+\rho)^3} \\ K & = & \left[\frac{2(5v-22)(v-4)\rho^2}{(v-6)(v-8)} + \frac{8(v-4)\rho}{v-6} + 1\right] \frac{6}{(1+\rho)^2(v-4)} \end{array} \tag{A.3}$$

or

$$\begin{array}{lcl} 0 & = & 2\rho[3(v-6)+4(v-4)\rho]^2 - S^2(v-4)(v-6)^2(1+\rho)^3 \\ 0 & = & 12(v-4)(5v-22)\rho^2 + 48(v-4)(v-8)\rho + 6(v-6)(v-8) \\ & & -K(v-4)(v-6)(v-8)(1+\rho)^2 \end{array}$$

Note that (A.3) may not have solutions with arbitrary combination of K and S^2. We next need to justify that the necessary and sufficient (N&S) condition under which (A.3) has one and only one solution is (5). Let $x = \rho$ and $y = v - 8$. Both are positive. Fix y, S, K, and define

$$f_1(x; y, K) \equiv x^2\left[\frac{12(5y+18)}{(y+2)y} - K\right] + 2x\left[\frac{24}{y+2} - K\right] + \frac{6}{y+4} - K \equiv Ax^2 + 2Bx + C \tag{A.4}$$

$$f_2(x; y, S) \equiv x^3\left[\frac{32(y+4)}{(y+2)^2} - S^2\right] + 3x^2\left[\frac{16}{y+2} - S^2\right] + 3x\left[\frac{6}{y+4} - S^2\right] - S^2 \equiv Dx^3 + 3Ex^2 + 3Fx - S^2 \tag{A.5}$$

Since

$$\frac{12(5y+18)}{(y+2)y} > \frac{32(y+4)}{(y+2)^2} > \frac{24}{y+2} > \frac{16}{y+2} > \frac{6}{y+4} > 0,$$

the N&S condition that (A.4) has one and only one positive root is $\frac{12(5y+18)}{(y+2)y} > K > \frac{6}{y+4}$. The positive root is $x_1(y; K) = \frac{\sqrt{B^2-AC}-B}{A}$ and $y_L < y < y_U$ where $y_L = \max(0, 6/K - 4)$ and $y_U = \sqrt{1 + 156/K + 900/K^2} - 1 + 30/K$. Define

$$g(y; S, K) = f_2(x_1(y; K); y, S), \qquad \text{for } y_L < y < y_U. \tag{A.6}$$

Lemma A.3 implies that g has root(s) in (y_L, y_U) if and only if

$$\lim_{y \to y_U^-} D > 0 \quad \text{or} \quad \frac{32(y_U+4)}{(y_U+2)^2} > S^2 \tag{A.7}$$

And the root is unique as well. Therefore, the N&S condition that one can estimate the GST parameters via the first four moments is (5). ■

To complete the proof of Proposition 2.5 , we need to justify the following lemmas:

Lemma A.1. *$x_1(y; K)$ is increasing in y, and $\lim_{y \to y_L^+} x_1(y; K) = 0$, $\lim_{y \to y_U^-} x_1(y; K) = +\infty$.*

Proof: (1) Note that $x_1' = -\frac{A'x_1^2 + 2B'x_1 + C'}{2\sqrt{B^2-AC}} > 0$. x_1 is increasing in y.

(2) If $K < 3/2$, then $y_L = 6/K - 4 > 0$. Note that $\lim_{y \to y_L^+} A > 0$, $\lim_{y \to y_L^+} B > 0$, $\lim_{y \to y_L^+} C \nearrow 0$. We have $\lim_{y \to y_L^+} x_1 = 0$. If $K \geq 3/2$, then $y_L = 0$. Similarly we have $\lim_{y \to y_L^+} x_1 \searrow 0$.

(3) Note that $\lim_{y \to y_U^-} A \searrow 0$, $\lim_{y \to y_U^-} B < 0$, and $\lim_{y \to y_U^-} C < 0$. $\lim_{y \to y_U^-} x_1 = \lim_{y \to y_U^-} \frac{-C}{\sqrt{B^2-AC}+B} = +\infty$. ■

Lemma A.2. *For $g(y; S, K)$ defined in (A.6), if there exists $y_0 > 0$ such that $g(y_0) = 0$, then $g'(y_0) > 0$.*

Proof: Define $a = A + K$, $b = B + K$, $c = C + K (= F + S^2)$, $d = D + S^2$, and $e = E + S^2$. Then

$$K = \frac{ax_1^2 + 2bx_1 + c}{(x_1+1)^2}, \quad S^2 = \frac{dx_1^3 + 3ex_1^2 + 3cx_1}{(x_1+1)^3}\Big|_{y=y_0}.$$

Therefore,

$$\frac{dx_1}{dy} = -\frac{A'x_1^2 + 2B'x_1 + C'}{2(Ax_1 + B)} = -\frac{(a'x_1^2 + 2b'x_1 + c')(1 + x_1)}{2((a-b)x_1 + (b-c))},$$

$$\begin{aligned}
\frac{dg}{dy}\Big|_{y=y_0} &= D'x_1^3 + 3E'x_1^2 + 3F'x_1 + 3(Dx_1^2 + 2Ex_1 + F)\frac{dx_1}{dy}\Big|_{y=y_0} \\
&= (d'x_1^2 + 3e'x_1 + 3c')x_1 - 3\left((d-e)x_1^2 + 2(e-c)x_1 + c\right)\frac{(a'x_1^2 + 2b'x_1 + c')x_1}{2((a-b)x_1^2 + (b-c)x_1)}\Big|_{y=y_0} \\
&= -(a'x_1^2 + 2b'x_1 + c')x_1\left(\frac{3}{2}\frac{(d-e)x_1^2 + 2(e-c)x_1 + c}{(a-b)x_1^2 + (b-c)x_1} - \frac{d'x_1^2 + 3e'x_1 + 3c'}{a'x_1^2 + 2b'x_1 + c'}\right)\Big|_{y=y_0}
\end{aligned}$$

With some algebra, one can show that

$$\frac{3}{2}\frac{(d-e)x_1^2 + 2(e-c)x_1 + c}{(a-b)x_1^2 + (b-c)x_1} - \frac{d'x_1^2 + 3e'x_1 + 3c'}{a'x_1^2 + 2b'x_1 + c'} > 0$$

for all $y > 0$, and hence $g'(y_0) > 0$. ∎

Lemma A.3. *Consider $g(y; S, K)$ defined in (A.6). If $\lim_{y \to y_U^-} D > 0$, g has one and only one root in (y_L, y_U). If $\lim_{y \to y_U^-} D \leq 0$, $g(y) < 0$ for all $y \in (y_L, y_U)$.*

Proof: If $\lim_{y \to y_U^-} D > 0$, then $D(y) > 0$ for all $y \in (y_L, y_U)$. It follows from Lemma A.1 that $\lim_{y \to y_L^+} g = -S^2 < 0$ and $\lim_{y \to y_U^-} g = +\infty$. Therefore $g(y)$ has one and only one root in (y_L, y_U) due to intermediate value theorem and Lemma A.2.

Next we will show that $g(y) < 0$ for all $y \in (y_L, y_U)$ if $\lim_{y \to y_U^-} D \leq 0$. Define $y_D = \inf\{y \in (y_L, y_U) : D(y) \leq 0\}$. If $y_D = y_L$, then $F < E < D \leq 0$ on (y_L, y_U) and hence $g(y) < 0$ for $y \in (y_L, y_U)$. Suppose $y_D \in (y_L, y_U)$. We have $D(y) > 0$ for $y \in (y_L, y_D)$. Note that $\lim_{y \to y_D} D = 0$, $\lim_{y \to y_D} E < 0$, $\lim_{y \to y_D} F < 0$, and $\lim_{y \to y_D} x_1$ is finite and positive. Thus $\lim_{y \to y_D} g < 0$. Note also that $\lim_{y \to y_L^+} g = -S^2 < 0$. It follows from Lemma A.2 that $g(y) < 0$ for $y \in (y_L, y_D]$ and hence $g(y) < 0$ for $y \in (y_L, y_U)$. ∎

B Fast Fourier transform for density function and call price

We use Fast Fourier transform of Carr and Madan (1999) to numerically evaluate conditional density function and call option prices derived from Model (14). The conditional density $f(y; t, T, x_t) = \frac{1}{\pi}\int_0^\infty e^{-iuy}\Psi(iu; t, T, x_t)du$ which is inverse Fourier transform of characteristic function, is approximated by

$$f_t(y_l; T) \approx \sum_{j=1}^{N} e^{-(l-1)(j-1)\frac{2\pi}{N}i}\tilde{x}(j), \quad \tilde{x}(j) = \frac{\eta}{\pi}\delta_j e^{ibu_j}\Psi(iu_j; t, T, x_t) \tag{B.8}$$

where $y_l = -b + \lambda(l-1)$ with $b = N\lambda/2$ and $\lambda = 2\pi/(\eta N)$ for $l = 1, 2, \ldots, N$. $u_j = \eta(j-1)$, and $\delta_j = 1/2$ when $j = 1$ and 1 otherwise. For call price, consider the fact that

$$C_t(K; T, x_t) \equiv E((e^{Y_T} - K)^+ | \mathcal{F}_t) = \frac{e^{-\xi \log(K)}}{\pi}\int_0^\infty e^{-iu\log(K)}\phi(u)du, \quad \phi(u) = \frac{e^{-r(T-t)}\Psi(iu + (1+\xi); t, T, x_t)}{\xi^2 + \xi - u^2 + iu(2\xi + 1)}$$

where ξ is a dampening coefficient. Therefore $C_t(K; T, x_t)$ is approximated by

$$C_t(K_l; T, x_t) \approx e^{-\xi \log(K_l)}\sum_{j=1}^{N} e^{-(l-1)(j-1)\frac{2\pi}{N}i}\tilde{x}(j), \quad \tilde{x}(j) = \delta_j\frac{\eta}{\pi}e^{iu_j b}\phi(u_j) \tag{B.9}$$

where $\log(K_l)$=$-b + \lambda(l-1)$ with $b = N\lambda/2$ and $\lambda = 2\pi/(\eta N)$ for $l = 1, 2, \ldots, N$.

Considering the criterion stated in Carr and Madan (1999), we take in numerical calibration N=2^{13}, $\eta = 0.25$ for both (B.8) and (B.9), and $\xi = 4$ in (B.9).