The Skew-*t* Option Pricing Model

Claudia Yeap, S. T. Boris Choy$^{(\boxtimes)}$, and S. Simon Kwok

University of Sydney, Sydney, Australia
claudia.yeap@gmail.com, {boris.choy,simon.kwok}@sydney.edu.au

Abstract. This paper presents a skew-*t* option pricing model. It is constructed analogously to the variance gamma option pricing model proposed by [14]. This proposed skew-*t* model inherits the variance gamma model's three parameters and their respective interpretations. In addition, it also has a fat-tailed, skewed distribution and infinite-activity (pure jump) stock dynamics, which is achieved through modelling the length of time intervals as stochastic. This paper has three main insights. From a theoretical perspective, a result is obtained for the correlation between the variance gamma model's logarithm returns and its gamma stochastic variance. This result holds for the skew-*t* model as well, which has reciprocal gamma variance, and it provides a new way to quantify the leverage effect under each model. The focus then shifts to the numerical procedures required for estimating the skew-*t* model's parameters. Finally, an empirical comparison between the skew-*t*, variance gamma and Black-Scholes models is conducted. The discussion links four pieces of analysis - pricing errors, pricing biases, the higher moments of the distributions and the market's implied volatility.

Keywords: Black-Scholes · Mean-variance mixture
Variance gamma · Reciprocal gamma

1 Introduction

For almost twenty years, the variance gamma (VG) option pricing model [14] has risen to prominence as the pure-jump, skewed and fat-tailed alternative to the Black-Scholes option pricing model [3]. While the Black-Scholes model assumes normality on the underlying stock's logarithm returns, the VG model assumes them to be skew-VG distributed. In this paper, the VG option pricing model is referred to as the skew-VG model to remain cognisant of this definition. The merits of the skew-VG model lie in its enhanced ability to model extreme returns in the market and to capture shocks around announcement periods. In 2007, it achieved the milestone of becoming a built-in pricing function on Bloomberg terminals [4] for traders' everyday use.

And yet, for most practitioners, students and academics, the most well-known fat-tailed alternative to the normal distribution is the Student-*t* distribution. It might reasonably be expected then that an option pricing model based on a

skewed t-distribution would also be an available alternative to the skew-VG option pricing model in the extant literature. This was not found to be the case.

This paper fills the gap in the literature for a skew-t option pricing model that is comparable to the skew-VG option pricing model. Along side the contribution of the new, stand alone option pricing model, comes fresh empirical insights into whether the skew-VG's documented pricing accuracy improvements over Black-Scholes is attributable to the option market's particular preference for the skew-VG distribution or rather a more generic preference for a skewed, fat-tailed distribution.

The theoretical conduit for developing the skew-t model has been the skew-t and skew-VG distributions' location and scale mixture of normals representation, as it unveils a striking parallel between the two distributions. Where the skew-VG uses a gamma mixing density, the skew-t uses a reciprocal (or more commonly known as inverse) gamma mixing density. It is due to this ability to reduce the two distributions' differences to a singular theoretical point, that after switching the mixing densities, we can use all other techniques implemented by Madan et al. [14] and the result is an original skew-t model.

The skew-t model inherits all the same properties as the skew-VG model, including being a pure-jump model, having a time-deformed Geometric Brownian motion interpretation, as well as the skew-VG's three parameters with exactly the same interpretation. The skew-VG's three parameters are the Brownian motion diffusion parameters, σ (which is comparable to Black-Scholes' sole parameter of volatility), a skewness parameter, θ, that governs how much the latent mixing variable impacts the drift, and ν, the variance of the mixing density, which drives the kurtosis of the mixture distributions.

After the model is specified, this paper has three further sections. The first is the comparison of the skew-t model to other option pricing models. This paper offers a quantifiable relationship via moments matching to the skew-VG model and parameter matching to Heston's stochastic volatility model [10]. The second section concerns numerical strategies for estimating the parameters of the skew-t model using options data, that is under the risk-neutral measure. Gradient-free algorithms were employed and the surface of the objective function and convergence are portrayed in this section in preparation for the empirical test that follows. The third part is an empirical test that examines three years of data commencing with the Global Financial Crisis. Both a cross-sectional and time-series treatment of the pricing results are provided, and we investigate the predictability of the pricing errors to measure model misspecification.

In pursuit of a closed-form solution, the proposed option pricing model is applicable to European call and put options. The resulting model is semi-analytic one. For pricing American options using this model we would need to appeal to simulation methods. Using the S&P500 index as the underlying asset of the data set, there was no shortage of European options data available. Over three years, 39,667 data points are accounted for, each for puts and calls.

Section 2 specifies the properties of the skew-*t* option pricing model. Section 3 covers parameter estimation. Section 4 encompasses the empirical results. Section 5 concludes this paper.

2   The Skew-*t* Option Pricing Model

The skew-*t* option pricing model is the central contribution of this paper. It has been derived using the same principles as Madan, Carr and Chang's skew-VG option pricing model. This section shows the steps used to derive the skew-*t* model. Once the model has been defined, we investigate how the proposed model compares to other option pricing models. In particular we compare it with other *t*-distributions, the skew-VG model and Heston's stochastic volatility model.

2.1   Skew-*t* Model Specification

In parallel with the skew-VG option pricing model, we disaggregate the derivation of the skew-*t* model into three tiers: the distribution of the stationary increment, the dynamic process for log stock price and the option pricing model.

2.1.1   Step 1: The Skew-*t* Distribution

The skew-*t* distribution that we use in this paper is constructed via the normal mean-variance mixture representation. Introducing the latent scale mixture variable λ, the probability density function (pdf) of the skew-*t* distribution with location μ, scale σ, skewness θ and degrees of freedom ν is given by

$$f(x) = \int_0^\infty N\left(x \mid \mu + \theta\lambda, \sigma^2\lambda\right) R\Gamma\left(\lambda \,\middle|\, \alpha, \beta\right) d\lambda \tag{1}$$

where $R\Gamma$ is the reciprocal gamma distribution with pdf given by

$$R\Gamma\left(\lambda \mid \alpha, \beta\right) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-(\alpha+1)} \exp\left(-\frac{\beta}{\lambda}\right), \quad \lambda, \alpha, \beta > 0. \tag{2}$$

The objective of our skew-*t* option pricing model is to be able to compare it against the skew-VG option pricing model. As such we select parameters of the mixing density so that our model's parameters have the exact same interpretation as skew-VG's parameters. We set the mean of the reciprocal gamma density to be 1 (so that the σ still remains comparable with the Brownian motion diffusion parameter as the standard deviation of log returns over a time unit), and then we further choose the parameters so that the variance of the time change density is ν, for $\nu > 0$. Based on the mean and variance for the reciprocal gamma density, the resulting parameters to use are

$$\text{For } R\Gamma(\alpha, \beta): \quad E[\lambda] = \frac{\beta}{\alpha - 1} = 1 \quad \text{and} \quad V[\lambda] = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)} = \nu.$$

The solution is

$$\alpha = \frac{1}{\nu} + 2 \quad \text{and} \quad \beta = \frac{1}{\nu} + 1.$$

Finlay and Seneta [8] also parameterise their skew-t distribution as a normal mean-variance mixture form with the reciprocal gamma distribution having the unit mean. In addition, it has the degrees of freedom equal to 2α or $\frac{2}{\nu} + 4$. In this paper, the skew-t distribution is reparameterised to have variance equal to ν.

Substituting the parameters into the reciprocal gamma density provided above, and taking the normal density function, the resulting conditional density for the proposed skew-t distribution is:

$$
\begin{aligned}
f(x) &= \int_0^\infty N\left(x \mid \mu + \theta\lambda, \sigma^2\lambda \right) R\Gamma\left(\lambda \,\Big|\, \frac{1}{\nu} + 2, \frac{1}{\nu} + 1 \right) d\lambda \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2\lambda}} \exp\left(-\frac{(x - (\mu + \theta\lambda))^2}{2\sigma^2\lambda} \right) \\
&\quad \frac{(1+\nu)^{(\frac{1}{\nu}+2)}\lambda^{-\frac{1}{\nu}-3} \exp\left(\frac{-(1+\nu)}{\nu\lambda} \right)}{\nu^{(\frac{1}{\nu}+2)}\Gamma(\frac{1}{\nu}+2)} d\lambda.
\end{aligned}
\tag{3}
$$

Finlay and Seneta [8] provide the marginal density for the skew-t distribution. After reparameterising with $R\Gamma(\frac{1}{\nu}+2, \frac{1}{\nu}+1)$, we obtain the marginal pdf of the skew-t distribution as

$$
f(x) = \begin{cases}
\frac{\Gamma(\frac{1}{\nu}+\frac{5}{2})\sqrt{\nu}}{\sqrt{2\sigma^2(1+\nu)}\sqrt{\pi}\Gamma(\frac{1}{\nu}+2)} \left(1 + \left(\frac{(x-\mu)\sqrt{\nu}}{\sqrt{2\sigma^2(1+\nu)}} \right)^2 \right)^{-(\frac{1}{\nu}+\frac{5}{2})}, & \text{for } \theta = 0 \\[2ex]
\sqrt{\frac{2}{\pi}} \frac{(1+\nu)^{(\frac{1}{\nu}+2)} \exp\left(\frac{(x-\mu)\theta}{\sigma^2} \right)}{\nu^{(\frac{1}{\nu}+2)}\sigma\Gamma(\frac{1}{\nu}+2)} \left(\frac{\theta^2\nu}{2\sigma^2(1+\nu)+(x-\mu)^2\nu} \right)^{\frac{1}{2}(\frac{1}{\nu}+\frac{5}{2})} & \text{for } \theta \neq 0 \\[2ex]
K_{\frac{1}{\nu}+\frac{5}{2}}\left(\frac{|\theta|\sqrt{2\sigma^2(1+\nu)+(x-\mu)^2\nu}}{\sigma^2\sqrt{\nu}} \right)
\end{cases}
\tag{4}
$$

where $K(\cdot)$ is a modified Bessel function of the second kind.

2.1.2   Step 2: The Skew-t Process for Log Stock Prices

The second step is to verify whether we can apply the Lévy process theory to make a continuous, dynamic process out of the skew-t distribution. We saw above that Lévy processes require the distribution to be infinitely divisible. Since the t-distribution is a special case of the GH distribution and the GH distribution and all its special cases are infinitely divisible (Barndorff-Nielsen and Shephard [2]), we can invoke Lévy process theory here, the sum of independent and identically distributed increments. Using X_τ to denote the log returns over period, τ, the skew-t process is defined as follows.

$$
\begin{aligned}
X_1 &= C_j \\
C_j &\sim \text{skew-}t \\
X_\tau &= \sum_{j=1}^{\tau} C_j.
\end{aligned}
\tag{5}
$$

2.1.3 Step 3: Risk-Neutral Pricing and the Skew-t Option Pricing Model

The third step is to extend the log stock process into an option pricing context by first fixing the drift μ to enforce the martingale condition. By deriving the characteristic functions for a normal and a reciprocal gamma distributions, it can be shown that μ is given by

$$\mu = (r - q) - \ln\left(\frac{2\left(-\frac{1+\nu}{\nu}(\theta + \frac{\sigma^2}{2})\right)^{\frac{(\frac{1}{\nu}+2)}{2}}}{\Gamma(\frac{1}{\nu}+2)} K_{\frac{1}{\nu}+2}\left(\sqrt{-4\frac{1+\nu}{\nu}(\theta + \frac{\sigma^2}{2})}\right)\right). \quad (6)$$

Since $\nu > 0$, the real drift constraint for the skew-t model is

$$\theta + \frac{\sigma^2}{2} < 0. \quad (7)$$

Moreover, the constraint that $\sigma > 0$ forces θ to be negative. However under the risk-neutral pricing framework, we expect any skewness parameter to be negative as it corresponds to investors' risk-averse behaviour anyway. This is also supported empirically, by the work of Konikov and Madan [13] on the skew-VG option pricing model for S&P500 index options (which is the data used in this paper), all estimates for θ were negative.

The final step is to provide the option pricing model itself. The t-distribution is not closed in convolution density and so we use the semi-analytic option pricing model from Carr and Madan [5]. Now the skew-t option pricing model is presented as follows.

$$C_0 = e^{q\tau} S_0 \Pi_1 - e^{-r\tau} X \Pi_2$$
$$P_0 = e^{-r\tau} X (1 - \Pi_2) - e^{q\tau} S_0 (1 - \Pi_1)$$
$$\Pi_1 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re\left(\frac{e^{-iux}\phi_{s_\tau}(u-i)}{iu\phi_{s_\tau}(-i)}\right) du \quad (8)$$
$$\Pi_2 = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re\left(\frac{e^{-iux}\phi_{s_\tau}(u)}{iu}\right) du$$

$$\phi_{s_\tau}(u) = \text{Characteristic function of the log stock price at time, } \tau$$

For skew-$t_{\frac{2}{\nu}+4}$:

$$\phi_{s_\tau}(u)^{\{T\}} = e^{iu(s_0+\mu\tau)}$$

$$\left[\frac{2\left(-\frac{1+\nu}{\nu}\left(iu\theta - \frac{\sigma^2 u^2}{2}\right)\right)^{\frac{\frac{1}{\nu}+2}{2}}}{\Gamma\left(\frac{1}{\nu}+2\right)} K_{\frac{1}{\nu}+2}\left(\sqrt{-4\frac{1+\nu}{\nu}\left(iu\theta - \frac{\sigma^2 u^2}{2}\right)}\right)\right]^\tau \quad (9)$$

where X is the strike price not the log returns, C_0 is the current call price, P_0 is the current put price, the relationship between the call and put price is

defined by Put-Call parity. τ is the time to maturity in years, r is the annual risk-free rate of return, q is the annual dividend yield, Π_1 and Π_2 are risk-neutral probabilities, $s_\tau = \ln S_\tau$, the log of the spot price, $x = \ln X$, the log of the strike price and $i = \sqrt{-1}$. μ is the martingale corrected mean given in equation (6). The degrees of freedom is $\frac{2}{\nu} + 4$. $\sigma > 0$, $\nu > 0$, and $\theta + \frac{\sigma^2}{2} < 0$.

2.1.4 The Skew-t, Skew-VG and Comparison with Heston's Stochastic Volatility Model

From the literature review, we saw that an unresolved rivalry exists between the skew-VG and Heston's models (Kim and Kim [12]). It is interesting therefore to directly compare these two models and the new skew-t model theoretically. This section provides new insights into how we can match the parameters of Heston's model with skew-VG and skew-t.

Under Heston's model [10], we model the log returns to be normally distributed with a fixed drift but a stochastic variance. The pair of formulas is provided below. That variance process, V_t, is mean-reverting by rate, κ, to the long run mean variance, μ_V, and with a Gaussian innovation term and volatility of stochastic variance parameter, σ.

As a result of the time varying variance, the initial V_0 becomes another parameter to estimate for Heston's model. It then has a fifth parameter for the correlation between its two processes, the log returns process and variance process to be estimated through ρ. In the expression below, log returns are denoted with X_t (τ notation is done away with in this section as it was used previously to be consistent with the option time to maturity, τ but the insights offered here are on a stock dynamics level).

$$\text{Heston's model:} \quad X_t = \mu + \sqrt{V_t} z_t^{(1)}$$
$$V_t = V_{t-1} + \kappa(\mu_V - V_{t-1}) + \sigma\sqrt{V_{t-1}} z_t^{(2)} \qquad (10)$$
$$z_t^{(1)} \sim N(0,1),\ z_t^{(2)} \sim N(0,1)$$

For the skew-VG and skew-t models, through the random time change interpretation to the mixing variable, λ, the interpretation described so far has been that log returns are distributed normally, with fixed mean and variance, *conditional* upon a *random* length of the time increment. Another way to interpret this is that, modelling the process over fixed time intervals, the mean and the variance of the log returns are changing each period, as controlled by the λ. The interpretation of changing drift and variance for skew-VG and skew-t is more conducive to comparison with Heston's model.

We can subscript λ as λ_t without loss of generality, where it can be interpreted as the latent variable's disturbance to the average drift $(\mu + \theta)$ and variance (σ^2) of the skew-VG process. The result is that, where Heston's model has two processes, the log returns and stochastic variance, the skew-VG and skew-t both have three processes, with the addition of stochastic behaviour in the drift $(r_t = \mu + \theta\lambda_t)$.

The skew-VG and skew-t model can be written in terms of the three processes as such:

$$X_t = \mu + \theta\lambda_t + \sigma\sqrt{\lambda_t}z_t$$
$$r_t = \mu + \theta\lambda_t$$
$$V_t = \sigma^2\lambda_t$$
$$z_t \sim N(0,1) \tag{11}$$
$$\lambda_t \sim \text{Reciprocal gamma for skew-}t,$$
$$\text{Gamma for skew-VG, mean} = 1 \text{ and variance} = \nu$$

Table 1 provides a side-by-side comparison between Heston and the skew-t (and VG) models. It also expresses the processes as distributions to aid further comparison of the means and variances. This analysis reveals to us a close-knit relationship between the skew-t (and VG) and Heston models. Specifically, we are able to retrieve the skew-t (VG) model from Heston's model by:

1. Setting Heston's mean reversion parameter, $\kappa = 0$, such that the expected variance is constant over time. Under Heston's model, as a result, $\mu_V = V_0$, the long run mean variance is the same as the initial variance,
2. Instead of the stochastic variance following a normal distribution, we ascribe it a gamma distribution,
3. We then introduce a stochastic component to the drift of the returns process which is driven by the same gamma innovation in the stochastic variance process from (2) and scaled by a parameter, θ,
4. We do not directly estimate the correlation ρ between the log returns and variance process.

In net, the VG process sets a restriction on two of Heston's give parameters, $\kappa = 0$ and $V_0 = \mu_V$, does not estimate another of its parameters, the correlation, ρ and introduces its own drift skewness parameter, θ, to come to its total of three parameters. $\kappa = 0$ for skew-t and skew-VG means that the models will likely perform poorly during times of volatility clustering such as crisis periods.

Two further remarks on the comparison—Heston's stochastic variance is normally distributed and so requires a further condition to make it positive, but the skew-t (VG)'s stochastic variance will always be positive as it is driven by non-negative distributions, the gamma or reciprocal gamma. Indeed this would be the case for any time subordinator model (time cannot be negative). Another note, is that Heston's σ is comparable to skew-t (VG)'s ν in that they both drive the volatility of the stochastic variance, and in turn kurtosis.

From this section, the similarities that have been formalised between the skew-t (VG) and Heston model shed light on why there is the rivalry between the skew-VG and Heston model in the literature to date—*both* models have stochastic variance processes and can model the leverage effect. The deficiency however of the skew-VG and skew-t model is that restricts Heston's $\kappa = 0$.

Table 1. Comparison between Heston's model and the skew-t and skew-VG models. Conditional drifts and variances are shown.

	Heston's model	Skew-t and Skew-VG model *Replace reciprocal gamma with gamma for skew-VG results*
Log returns process	$X_t = \mu + \sqrt{V_t}z_t^{(1)}$	$X_t = \mu + \theta\lambda_t + \sigma\sqrt{\lambda_t}z_t$
Drift of log returns process	$r_t = \mu$ (constant)	$r_t = \mu + \theta\lambda_t$
Variance of log returns process	$V_t = V_{t-1} + \kappa(\mu_V - V_{t-1}) + \sigma\sqrt{V_{t-1}}z_t^{(2)}$	$V_t = (\sigma\sqrt{\lambda_t})^2 = \sigma^2\lambda_t$
Innovation terms	$z_t^{(1)} \sim N(0,1),$ $z_t^{(2)} \sim N(0,1)$	$z_t \sim N(0,1)$ $\lambda_t \sim$ Reciprocal gamma mean = 1 var = ν, $\nu > 0$
Distribution of drift	r is constant over time	$r_t \sim$ Reciprocal gamma mean = $\mu + \theta$ var = $\theta^2\nu$
Distribution of variance	$V_t \sim$ Normal mean = $V_{t-1} + \kappa(\mu_V - V_{t-1})$ var = $\sigma^2 V_{t-1}$	$V_t \sim$ Reciprocal gamma mean = σ^2 var = $\sigma^4\nu$
Correlation between log returns, X_t, and variance, V_t	$-1 \le \rho \le 1$	$\rho = \dfrac{\theta\nu}{\sqrt{\nu(\theta^2\nu + \sigma^2)}}$ $1 < \rho < 1$, since $\sigma > 0$
Correlation between drift, r_t, and variance, V_t	0, drift is deterministic	$\mathrm{corr}(r_t, V_t) =$ $\begin{cases} -1 & \text{for } \theta < 0 \\ 1 & \text{for } \theta > 0 \\ 0 & \text{for } \theta = 0 \end{cases}$

3   Parameter Estimation Procedures

This section discusses how the parameters of the option pricing models were calibrated. It has two focuses. First, I formulate the optimisation problem and then test the calibration methods on a simulated data set without noise and then with noise introduced. The purpose of this section is to verify that we are able to recover estimates close to the true parameter values in preparation for the empirical section that follows.

3.1 Formulating the Optimisation Problem

3.1.1 The Objective Function

The objective function is the root mean square percentage pricing errors (RMSPE), where the percentage errors are found through taking the log difference. This was used by Madan et al. [14], the percentage errors used due to the range of call prices across different strikes and maturities. The authors also give the result that the optimisation problem is asymptotically equivalent to the maximum likelihood method.

The minimisation of square percentage errors has since been adopted in other empirical tests on the skew-VG model (see Kim and Kim [12] where instead of using log difference for percentage errors, a discrete percentage error measure was used, $([C - C^*]/C)$ where C^* is the call price estimated by the model and C is the observed call price. In this paper, with the chosen optimisation algorithm, outlined below, it was found that the log-difference aided the calibration compared to using the discrete measure from Kim and Kim [12]. As such the objective function used in this paper, from Madan, Carr and Chang (their Equation (30)) is:

$$\text{Objective function} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\ln(C^*) \ln(C))^2}. \tag{12}$$

where n is the number of call prices, and ln is the natural logarithm.

3.1.2 Constraints

We impose the positivity constraints on the Black-Scholes σ and the skew-VG and skew-t's, ν and σ through log transformations. For the constraints required to ensure that the drift, μ is real in the skew-VG and skew-t model, we use a penalty function that adds an arbitrarily large quantity to the minimisation objective value in the case of violation.

Recalling from above these constraints, for skew-VG: $\nu \left(\theta + \frac{\sigma^2}{2} \right) < 1$ and for skew-t: $\theta + \frac{\sigma^2}{2} < 0$, skew-$t$'s θ is restricted to negative values but skew-VG's θ is not. For our optimisation, this restricted domain on skew-t's θ was an advantage for obtaining more accurate estimates compared to the skew-VG. It was therefore decided that a negativity constraint would be imposed on skew-VG's θ as well. This firstly, makes for a fairer competition between the models but also is theoretically sound based on the arguments raised above, namely investors' risk-aversion and previous empirical studies on the skew-VG model estimating all negative θs.

To ensure the negativity of skew-VG and skew-t's θs, where we directly estimate the $-\log(\theta)$ and retrieve θ through taking the exponential and then negative of that value. Together, the log transformations and penalty functions allow us to then use an unconstrained optimisation procedure.

3.1.3   The Optimisation Algorithm

The optimisation algorithm used is the Nelder-mead simplex algorithm, instructed by Matlab's $fminsearch$ function. $fminsearch$ is Matlab's unconstrained, non-linear optimiser where a simplex (a 2-dimensional line for the Black-Scholes with only one parameter and tetrahedron (a simplex with $3 + 1$ faces) for the skew-VG and skew-t models each with three parameters) descends the surface of the objective function. At each step, the simplex can be reflected, expanded, contracted or shrunk. Its changing step size makes it a preferred algorithm for noisy objectives, such as that for the option pricing models with numerical integration (Gilli and Schumann [9]).

In other papers on the calibration of such option pricing models using Matlab, the $lsqnonlin$ function (non-linear least squares) has also been implemented (Moodley [15]). However $lsqnonlin$, $fminunc$ and $fmincon$ performed poorly on calibrating our models. It was observed that the global optimiser, $patternsearch$, could recover the true parameter values as well as $fminsearch$, but it took an unworkable about of time. $fminsearch$ on the other hand took only eight hours to estimate the parameters for all 161 weeks (close to 40,000 prices) using parallel computing and four workers.

3.1.4   Initial Values, Stopping Criterion and Convergence

The initial values were set based on the average parameter values estimated in Madan, Carr and Chang's empirical testing of the skew-VG model [14]. The initial values for the skew-VG and skew-t model were selected to be the same. Respectively, for θ, ν and σ, they are -0.2, 0.2 and 0.15. For Black-Scholes the initial value for σ is 0.2, set higher than the initial σ in the other two models because it does not have the availability of the extra parameters to further increase the variance of the distribution.

There are two operative stopping criteria in our algorithm. The first is a minimum tolerance on gains made on the objective value, $1e{-}04$, and the second is a minimum tolerance on movement in the variables, $1e{-}04$. These criteria are adequate to reach objective values that are small enough to allow us to observe convergence in the parameters estimated in the empirical study.

3.1.5   Obtaining Standard Errors

We encountered the issue that, unlike the gradient-based optimisation functions in Matlab, the $fminsearch$ function does not output the Jacobian or Hessian from which we can calculate the standard errors. In their calibration study, Gilli and Schumann [9] did not provide standard errors.

Calculating the standard errors required first obtaining the Jacobian matrix ($n \times p$, where p is the number of parameters), which was found through evaluating the slope of the objective function for each individual call price and for each parameter. The interval over which the slope was evaluated was set to a small value, $1e{-}03$. Then it was also required that we convert the standard errors obtained on the log-transformed parameters to the standard errors for the

parameters in their original form, the exponential of the log-transformed parameters. To so this, we invoked the Delta method and so multiplied the standard errors by the absolute value of the first derivative of the exponential (or negative exponential for θ) function of the log-parameters.

4   Empirical Study

We cast the following empirical hypotheses. Given how different the skew-VG and skew-*t* distributions are with the same parameter values, it is likely that in their attempt to fit a mutual distribution they will return rather differing parameter estimates. Another empirical observation we are likely to see is that during times of volatility clustering or financial crisis, the skew-VG and skew-*t* may perform worst.

It should be noted that one of the key advantages of focusing on the comparison between two models with the same number of parameters is that no advantage is given to one of them in in-sample fitting. Black-Scholes however, with only one parameter will be at a disadvantage for the in-sample fitting, which along with the practicality of being interested forecasting prices, is the motivation for conducting out-of-sample tests as well.

This empirical section is based on the methodology employed by Bakshi et al. [1] for comparing option pricing models based on two main criteria: minimising pricing error and minimising the predictability of the pricing errors by moneyness (strike:spot ratio) and maturity. Subsequent empirical tests such as Madan et al. [14], Kim and Kim [12] and Eberlein et al. [6] use similar test designs. Unlike the aforementioned tests, this paper has a distributional focus and provides additional analysis of the higher moments, tail-fitting and quantile plots.

4.1   The Data

The data used are the daily prices of European call and put options on the S&P500 index. In order to cull highly illiquid options data we restrict the data to options with moneyness between 0.97 and 1.03 and maturities between 1 month away and 1 year away (Kim and Kim [12]). All option data was obtained from Option Metrics. The risk-free rate is the rate of return on a one month US treasury bill and the dividend is that on the S&P500 index.

In order to account for frequently changing parameter values over time, the training period was only one week long. The size of the data available caters for the use of short training samples. There are 245 data points per week on average. For out-of-sample we tested the model on the data one day out of the sample.

Table 2 summarises the data used. There are 39,667 data points each for calls and puts, spanning 162 weeks (1 to 161 for training and 2 to 162 for testing) starting on the 23 February 2009. It was desired that the three years encompassed bust, boom and quiet periods but otherwise the particular start date for the sample is arbitrary. This period captures part of the GFC, the Euro-debt crisis,

Table 2. Sample properties of the S&P500 Index European call options and put option in parenthesis. Categorisation is in terms of moneyness (strike to spot price ratio) and time to maturity. Dates span from 23 February 2009 29 March 2012, 162 weeks (161 training samples).

Call (Put)	Time to maturity			{Sample size}
Moneyness	<3 months	3–6 months	>6 months	Subtotal
In (Out of) 0.97–0.99	$50.68 ($30.59) {7226}	$68.60 ($52.15) {3197}	$96.65 ($88.12) {2684}	$64.47 ($47.63) {13107}
At 0.99–1.01	$36.00 ($39.48) {7340}	$54.59 ($61.52) {3300}	$82.92 ($97.70) {2760}	$50.24 ($56.90) {13400}
Out of (In) 1.01–1.03	$24.04 ($51.08) {7344}	$41.98 ($72.42) {3314}	$71.02 ($109.40) {2502}	$37.49 ($67.54) {13160}
Subtotal	$36.83 ($40.44) {21910}	$54.90 ($62.15) {9811}	$83.81 ($98.15) {7946}	$50.71 ($57.37) {39667}
Number of weeks				162
Average number of data points in a training period (a week)				245

the US debt-ceiling crisis and some recovery periods in between. As expected, we can see that for calls, the price of the options increases with the time left until maturity and the deeper the option is in-the-money.

Table 2 is also useful for highlighting how differing the option prices are within our data set. The range on the averages for each category is from $36.83 to $96.65. This motivates the use of percentage pricing errors rather than pricing errors in dollar terms. The main body of this paper therefore will only include the percentage pricing errors (mean, mean absolute and root mean square percentage errors).

4.2   Empirical Parameter Estimation and In-Sample Fitting

There are a few notes to make regarding the summary of parameters given in Table 3. The skew-VG and skew-t use different parameter values to fit a common distribution. For θ, on average, the skew-VG will fit a more negative skewness parameter than skew-t and in turn, skew-t appears to model a higher σ over skew-VG. For ν, looking just at the average value estimated, there is not much between the two models, skew-VG at 0.234 and skew-t at 0.238. In the context of the skew-t distribution, this is on average 12.4 degrees of freedom ($d.f. = \frac{2}{\nu} + 4$ from above).

Although the average ν's are similar, it is noticeable that skew-t's maximum value for ν is extremely high. Specifically we observed four kurtoses for skew-t that exceeded 10 (11.06, 40.30, 324.67 and 733.52). This means that the distribution has very fat tails and is very peaked at these points. Theoretically, the kurtosis is able to be infinite. While the skew-t can have very large values for ν, we notice that that the skew-VG can take on very large (negative) values of θ, -2.781.

Table 3. The parameters of the risk-neutral densities of log annual S&P500 Index returns based on the Black-Scholes option pricing model, skew-VG and skew-t option pricing models. There are 161 training samples. Standard errors have been omitted for neatness and because, just as was seen in the simulation, they were generally close to zero. The objective function is the root mean square percentage pricing error (RMSPE). Results are for calls. Puts results were similar and so are not provided to save space.

Model	BSM	Skew VG			Skew t		
Parameter	σ	θ	ν	σ	θ	ν	σ
Mean	0.220	−0.397	0.234	0.206	−0.174	0.238	0.239
Standard deviation	0.059	0.463	0.143	0.072	0.067	0.587	0.058
Minimum	0.134	−2.781	0.000	0.025	−0.697	0.025	0.152
Maximum	0.412	0.000	0.558	0.412	−0.064	5.520	0.434
Optimised objective value	0.114	0.0547			0.0695		

This is further evidence of differing behaviour across skew-VG and skew-t when it comes to data fitting. (By way of reference for the values on skew-VG's θ, given the length of our sample (161 weeks) our estimates seem reasonable compared to previous empirical tests. Konikov and Madan's ([13] empirical fitting of the skew-VG model to S&P500 data using only five training periods had a minimum $\theta = -1.789$).

The objective value (root mean square percentage pricing error) is used as an yardstick for in-sample fitting (Bakshi et al. [1]). On average, we can see that the skew-VG achieved the best fit of 5.47%, skew-t of 6.95% and both are vast improvements on Black-Scholes at 11.39%.

4.2.1   Moments of the Fitted Distributions

Following on from the parameter values, the moments for underlying log stock returns distribution are next provided (Table 4). For Black-Scholes, the variance is just the square of the σ parameter. For skew-VG and skew-t, all three parameters contribute to all three of the moments, that is, it is not the case that the skewness is solely determined by, but increases in magnitude with, the skewness parameter, θ for example. The moments were obtained through simulating the distributions based on the parameter values (1,000,000 simulations) instead of relying on the closed-form expressions for the moments as they cannot be used around $\nu = 0.5$ or 1 for the skew-t model.

As we would expect from the larger θ for skew-VG, on average the distribution fit under the skew-VG model has greater skewness (−0.596) compared to the skew-t (−0.494). Due to the outlying kurtoses (such as the two > 300), we also provide the median kurtosis (in addition to the mean kurtosis). It is interesting to note that while the skew-t's maximum value for ν and largest kurtosis estimated is a lot higher than for skew-VG, the median kurtosis fit by the skew-t is lower than the median kurtosis for the skew-VG. This suggests that the skew-VG's

Table 4. The average moments for the underlying log returns distributions for calls. Puts were similar and so not displayed in the interest of space.

Model	BSM	Skew-VG	Skew-t
Mean	−0.047	−0.054	−0.053
Variance	0.052	0.068	0.066
Skewness	0	−0.596	−0.494
Kurtosis	3	3.989	10.778
Median kurtosis	3	3.930	3.898

parameter values together, very likely driven by its higher θ, are overall fitting just as fat-tailed distributions as the skew-t.

4.3   Out-of-Sample Pricing Performance

This section presents three different perspectives on pricing performance. The first is cross-sectionally, that is how to the different models perform for options of varying moneyness ratios and maturities, second time-series analysis is provided and a comparison made between above average market volatility levels and below average. Third, we return to a more in depth understanding of the moments under high and low volatility to understand how the two distributions respond to the different financial climates.

4.3.1   Cross-Sectional Analysis and Tail-Fitting

In relation to Table 5, as a measure of overall fit, we can compare the root mean squared percentage errors (RMSPE) for all option types (bottom right corner of table), to those in the in-sample fitting. It appears that the out-of-sample fitting (7.39%, 4.10% and 4.86% for Black-Scholes, skew-VG and skew-t) is better than the in-sample fitting (11.39%, 5.47% and 6.95% respectively). However, this is a misleading comparison since the out-of-sample fit is on data only for one day (49 data points on average), where as the in-sampling fitting tried to fit one set of parameters to a whole week of data (245 data points on average). The results would therefore, instead, show signs of the variation that exists within a week.

Turning to the cross-sectional analysis and MPEs, (log predicted - log observed). Across moneyness, all models overprice OTM calls and underprice ITM calls. Black-Scholes has the greatest error differential between the moneyness categories and skew-VG has the least. Across maturities, the MPEs indicate a lower average error for short-term contracts compared to long-term, but checking the MAPE and RSMPE, we can see that the short term contracts (less than 3 months) actually have the highest absolute and root squared errors, especially OTM calls. We can also note that while for Black-Scholes the most accurately priced maturities are mid-term (3–6 months), for skew-VG and skew-t, the fit is best in the long-term contracts (6–12 months).

Table 5. 1 day out-of-sample percentage pricing errors using a week long training period and parameters estimated on option data for all maturities and moneyness (strike to spot price) levels. The percentage error is found by taking the log difference. A corresponding table of results for non-percentage errors is provided in the appendix. 161 training samples.

Calls			Months to maturity			
			<3	3–6	>6	All
		Strike/Spot	***A. Mean percentage errors (%)***			
Black-Scholes	ITM	0.97–0.99	−1.51	−4.62	−7.71	−3.53
	ATM	0.99–1.01	1.99	−3.00	−7.10	−1.12
	OTM	1.01–1.03	8.45	−0.18	−5.87	3.61
		All	2.99	−2.59	−7.00	−0.44
Skew-VG	ITM	0.97–0.99	0.97	−0.25	−1.73	0.12
	ATM	0.99–1.01	0.82	0.54	−0.88	0.38
	OTM	1.01–1.03	1.01	1.72	0.27	1.11
		All	0.92	0.68	−0.82	0.50
Skew-*t*	ITM	0.97–0.99	−1.33	−1.03	−2.42	−1.46
	ATM	0.99–1.01	−1.40	0.64	−1.22	−0.89
	OTM	1.01–1.03	2.92	3.62	0.51	2.68
		All	0.07	1.08	−1.10	0.07
		Strike/Spot	***B. Mean absolute percentage errors (%)***			
Black-Scholes	ITM	0.97–0.99	3.16	4.95	8.03	4.56
	ATM	0.99–1.01	4.83	4.20	7.72	5.23
	OTM	1.01–1.03	10.19	4.20	7.05	8.03
		All	6.07	4.42	7.69	5.92
Skew-VG	ITM	0.97–0.99	3.19	2.22	2.51	2.78
	ATM	0.99–1.01	3.97	2.80	2.50	3.36
	OTM	1.01–1.03	5.54	3.75	2.73	4.55
		All	4.22	2.92	2.59	3.53
Skew-*t*	ITM	0.97–0.99	2.89	2.43	3.26	2.81
	ATM	0.99–1.01	4.26	2.99	2.86	3.63
	OTM	1.01–1.03	6.64	4.99	2.92	5.51
		All	4.59	3.46	3.06	3.95
		Strike/Spot	***C. Root mean squared percentage errors (%)***			
Black-Scholes	ITM	0.97–0.99	3.42	5.06	8.08	5.33
	ATM	0.99–1.01	5.39	4.40	7.78	6.14
	OTM	1.01–1.03	11.15	4.53	7.16	9.47
		All	7.73	4.87	7.85	7.39
Skew-VG	ITM	0.97–0.99	3.27	2.30	2.57	3.08
	ATM	0.99–1.01	4.15	2.86	2.57	3.68
	OTM	1.01–1.03	6.03	3.83	2.81	5.15
		All	4.72	3.16	2.80	4.10
Skew-*t*	ITM	0.97–0.99	3.07	2.50	3.33	3.11
	ATM	0.99–1.01	4.68	3.09	2.98	4.16
	OTM	1.01–1.03	7.25	5.18	3.07	6.26
		All	5.55	4.02	3.34	4.86

There is one category where skew-t offers the best model by MAPE and RSMPE: in-the-money, short-term. However, this is not a compelling result, rather it seems an anomaly cross-sectionally. Indeed for puts - skew-VG is superior in all categories. Further, it is the OTM rather than the ITM options fitting that yields the most direct interpretation. Due to the limited downside for options, their pricing performance only reveals information about the fit of the stock price distribution to the right (left) of the strike price for calls (puts). Therefore, in-the-money options show more of a general fitting of the distribution (all but one tail), where as OTM options carry *only* the tail information. Table 6 gives the results for OTM puts and OTM calls to compare the left and right tail fitting of the skew-VG and skew-t distributions.

Table 6. Relative fit of the tails by the skew-VG and skew-t distributions

	Left tail fitting of the stock price distribution *(Reflects the probability that the spot price will fall below a strike price that is less than 99% of the current spot price)*	Right tail fitting of the stock price distribution *(Reflects the probability that the spot price will rise above a strike price that is more than 101% of the current spot price)*
	Out-of-the-money puts	Out-of-the-money calls
Mean absolute percentage error		
Skew-VG	3.62%	4.55%
Skew-t	3.77%	5.51%
Root mean square percentage error		
Skew-VG	4.06%	5.15%
Skew-t	4.27%	6.26%

From the out-of-the-money options fitting, it is evident that both models fit the left tail of the stock price distribution better than the right tail and that the skew-t's performance is more competitive with skew-VG's performance in the left tail fitting as well (0.15% greater MAPE for OTM puts, where as 0.96% greater MAPE for OTM calls).

We hypothesised in Sect. 2.1.4 that during crisis periods, the skew-t and skew-VG models would perform the worst. Based on the instances when neither the skew-VG or skew-t are able to improve upon the Black-Scholes model, these indeed occur when the volatility is at its highest. Particularly, these periods are at the start of our sample, 23 February 2009 during the Global Financial Crisis, and from the 25th July 2011, coinciding with the US debt-ceiling crisis. It is found that 4% of the time (161 weeks), neither the skew-VG or the skew-t could improve upon Black-Scholes. Outside this 4%, we are interested in the specific comparison between skew-t and skew-VG. Overall, the skew-VG outperforms skew-t more often than not and this is consistent with the cross-sectional results. The break-down between the percentage of times the skew-VG versus the skew-t outperforms is shown in Table 7.

Table 7. The percentage of weeks each model is the best performer across low versus high volatility for call options. The data sample begins on the 23 February 2009 (during the Global Financial Crisis) to 29 March 2012

Superior model	Skew-VG	Skew-*t*	Black-Scholes	Total
All levels of volatility	65%	30%	4%	161
Above average volatility	57%	34%	9%	70
Below average volatility	71%	27%	1%	91

To conclude on the pricing analysis for calls, we see that the skew-VG outperforms skew-*t* in almost all cross-sectional categories when the results are averaged across time. But to look at the results from a time-series perspective, it is apparent that while in below average volatility, the skew-VG is the superior model, in above average volatility we observe skew-*t*'s performance becomes more competitive with the skew-VGs.

5   Conclusion

The resounding conclusion is that without any information about the market volatility, the skew-VG option pricing model is more likely to outperform the skew-*t* than otherwise. However, if we use the S&P500 volatility index (VIX) or Black-Scholes implied volatility, then under more volatile conditions, the skew-VG's performance starts to deteriorate and the skew-*t* model becomes more competitive with the skew-VG. A core contribution of this paper has therefore been that, for the skew-VG, we now have evidence that its superior performance in past empirical studies showed a more powerful result than merely the option market's preference for a fat-tailed, skewed distribution and/or infinite activity, pure jump model. Now that we can compare it against another model that bears all these traits as well, (and with the same few number of parameters,) we can confirm that empirically the skew-VG distribution performs well due to some its idiosyncrasies that distinguish it from the skew-*t*. Considering the variability in the relatively performance of the skew-VG and skew-*t* models, a generalised model which nests both as special cases, enables pricing fitting and predicting that is at least as good as the superior model of the skew-VG or skew-*t*. Such a generalised model is proposed in Yeap et al. [7], and assumes the underlying log return distribution is generalised hyperbolic.

The final model that we gained deeper insight into, only in the theory section, was Heston's Stochastic Volatility model. The five parameters of Heston's model were matched to skew-VG and skew-*t* who each only have three. The result was that we can obtain the symmetric VG and symmetric t from Heston's model through setting the rate of mean reversion, κ, to 0, and by having a gamma stochastic variance process rather than normal. This paper also provided the skew-*t* and skew-VG models' equivalent values for Heston's correlation coefficient, ρ, such that the leverage effect can now be quantified within their framework as well.

Further research would be aimed at conducting empirical tests between the skew-t and skew-VG with models that can model time-dependency. This paper nominated three such models: Heston's model, Heyde and Gay's dependent t-option pricing model [11] and a two-state Markov model such as Konikov and Madan's [13]. The motivations for these extensions were discussed as each arose.

References

1. Bakshi, G., Cao, C., Chen, Z.: Empirical performance of alternative option pricing models. J. Finance **52**, 2003–2049 (1997)
2. Barndorff-Nielsen, O.E., Shephard, N.: Basics of Lévy processes. Draft chapter of Lévy Driven Volatility Models (2012)
3. Black, F., Scholes, M.: The pricing of options and corporate liabilities. J. Polit. Econ. **81**, 637–654 (1973)
4. Carr, P., Hogan, A., Stein, H.: The for a change: the variance gamma model and option pricing. Working paper (2007)
5. Carr, P., Madan, D.B.: Option valuation using the fast fourier transform. J. Comput. Finance **2**, 61–73 (1998)
6. Eberlein, E., Keller, U., Prause, K.: New insights into smile, mispricing and value at risk: the hyperbolic model. J. Bus. **71**(3), 371–405 (1998)
7. Yeap, C., Choy, S.T., Kwok, S.: A flexible generalised hyperbolic option pricing model and its special cases. J. Financ. Econometrics (2017, to appear)
8. Finlay, R., Seneta, E.: Stationary-increment variance gamma and t models: simulation and parameter estimation. Int. Stat. Rev. **76**(2), 167–186 (2008)
9. Gilli, M., Schumann, E.: Calibrating optionpricing models with heuristics. Natural Comput. Comput. Finance **380**, 9–37 (2012)
10. Heston, S.L.: A closed-form solution fo roptions with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**(2), 327–343 (1993)
11. Heyde, C.C., Gay, R.: Fractals and contingent claims. Australian National University, Preprint (2002)
12. Kim, I.J., Kim, S.: Empirical comparison of alternative stochastic volatility option pricing models: evidence from Korean KOSPI 200 index options market. Pacific-Basin Financ. J. **12**, 117–142 (2004)
13. Konikov, M., Madan, D.B.: Option pricing using variance gamma Markov chains. Rev. Deriv. Res. **5**, 81–115 (2002)
14. Madan, D.B., Carr, P.P., Chang, E.C.: The variance gamma process and option pricing. Eur. Finance Rev. **2**, 79–105 (1998)
15. Moodley, N.: The Heston model: A practical approach. Faculty of Science, University of Witwatersrand, Johannsesburg, Honours Project (2005)