Generalized Black-Scholes option pricing and investor sentiment

K. Ahn†‡, C. Ha‡, Y. Sun‡, B. Yang‡

†Moon Soul Graduate School of Future Strategy, Korea Advanced Institute of Science and Technology, 34141 Daejeon, Republic of Korea
‡HSBS Business School, Peking University, 518055 Shenzhen, P.R. China

This paper extends the standard Black-Scholes (BS) option pricing framework by utilizing the generalized solution to the heat equation proposed by Choi *et al.* (2017). We present the closed-form solution for a generalized option pricing model (GBS) and show that the modification to the standard call option price comes from two additional augments interpreted as factors associated with investor sentiment toward the underlying asset. Our model outperforms the standard BS model in both in-sample fit and out-of-sample prediction on 50ETF option data. Robustness tests also corroborate our results. Further analysis shows that the parameters for the newly incorporated terms strongly reflect investors expectation and help better explain how option market prices tend to drift from the BS model.

Keywords: Option pricing; Investor sentiment; Generalized Black-Scholes model; Expectation of return; Expectation of volatility

JEL Classification: C13, D30

1. Introduction

The seminal work of Black and Scholes (1973) laid the foundation of modern option pricing theory. While their theory, still serving as the fundamental workhorse in the derivative world, has drawn tremendous impact on both academics and practitioners alike, its empirical performance has not been fully entertaining, which is largely due to its restrictive assumptions. Consequently, researchers have proposed, for the past four decades, a myriad of option pricing models based on more realistic assumptions and theories, attempting to correct the mispricing of the standard Black-Scholes model.[1]

In this paper, we extend the standard Black-Scholes option pricing framework and derive a closed-form solution for a generalized pricing model by utilizing the general solution to the heat equation. Specifically, by transforming the partial differential equation of option price to the heat equation whose *generalized* solution has newly been obtained (Choi *et al.* (2017)), we derive a generalized BS (called GBS hereafter) pricing model. Our analytic pricing formula contains additional pricing factors for a call option. In particular, the option price of our GBS model is shown to be a linear combination of the original BS price and the option's Δ (*delta*) and Γ (*gamma*). We then further refine our GBS pricing model using a judicious parametrization of additional pricing terms to show that they are likely to be related to investor sentiment. The corresponding parameters are

[1] Volatility smile, which was initially identified by market professionals, for instance, is among the early empirical challenges of the validity of the BS assumptions, which led to a broader question of how to correct the Black-Scholes pricing bias. See Rubinstein (1994) for an early studies concerning the *smile*.

interpreted as implied sentiment factors because they tend to reflect investors' perception of future return prospect, more specifically, expectation of future return and its volatility.

The GBS model proposed in the current study is a novel modification to the standard BS model in that it provides a new, economically meaningful way to fix the mispricing extant in the existing option models by directly incorporating the impact of investor sentiment into the theoretical pricing model. So the standard BS model can be viewed as a special case of the GBS, wherein investor sentiment plays no role for pricing options. Empirical results confirm our interpretation: When investors have a bullish (bearish) expectation of future return as captured by the sentiment-related paratmeters, they tend to long (short) more call options, driving the option prices higher (lower). Thus, our empirical analysis provides direct evidence that investor sentiment may serve as a significant pricing determinant in the derivative market as well, not just in the stock market.

Using 50ETF call option prices from Shanghai Stock Exchange, we conduct both in-sample fit and out-of-sample prediction tests to see if our proposed pricing model and the corresponding interpretations are to be empirically validated. Overall, results from both tests clearly corroborate that the GBS model outperforms the BS model in a consistent manner with much smaller pricing errors. For in-sample pricing fit, we estimate the parameters using four different approaches[2] to compare the relative pricing errors of the two models considered. We document the dominant advantage of GBS model that average daily sum of squared errors (SSE) of the GBS model is less than half of that of the BS model over the entire sample period with 100% win-ratio. Following Bakshi *et al.* (1997), we further divide our option data into different categories according to moneyness and maturity for two reasons. First, we want to examine how well the GBS model captures the potential impact of underlying properties of the option contracts. Second, this subsample investigation also serves for the robustness test of in-sample fit and out-of-sample prediction performances. The correlation and regression results in Tables 6 - 9 show that the new structural parameters in the GBS model are closely related to a popular investor sentiment proxy, confirming our interpretation of the parameters. For out-of-sample pricing performance, we employ the same estimation methodologies as used for in-sample tests to find that the GBS model does not only have a better in-sample pricing fit, but also has strong predictive power. It is interesting to note that the GBS model, along with its lower pricing errors, yields lower call option prices than the BS model over the sample period. This is because bearish sentiment has prevailed in Chinese stock market since the market turbulence in 2015, and the GBS pricing model, as opposed to the BS model, is able to capture its influence on the option prices.

The major innovations in the option pricing research follwing Black and Scholes (1973) have been made in three broad areas. First, the stochastic nature of the underlying determinants of option prices is incorporated into the models. Notably, temporal variations in interest rates and the volatility of underlying assets are shown to be important aspects that need to be accounted for in any option pricing models aiming for satisfactory empirical performances. [3]

[2]Refer to Appendix D for a brief description of the four estimation methods employed in the paper.

[3]Since Merton (1973) developed a stochastic interest rate pricing model for equity options, a large number of studies have followed by further relaxing the restrictive BS assumptions to investigate different markets. To name a few among many others, Amin and Jarrow (1991), pointing out Merton's model cannot be extended to pricing American options due to the lack of a continuum of distinct bonds, propose an alternative approach based on the equivalent martingale measure and stochastic interest rate in Heath *et al.* (1992) with the application to foreign currency and currency futures options. Amin and Jarrow (1992) further extend their stochastic interest rate option pricing model by including risky asset, and derive the closed-form solutions for European type call and put options on a risky asset, forward and future contracts. They also explore the valuation of an American option whose payoff depends on the term structure of interest rate. Miltersen and Schwartz (1998) further extend this research field to commodity option market, assuming stochastic interest rate as well as convenience yield. Constant volatility, another restrictive assumption of the BS model, has also been challenged. Following Scott (1987), Hull and White (1987) and Wiggins (1987), stochastic volatility is built in option models with substantial pricing and hedging improvement (Bakshi *et al.* (1997) and Melino and Turnbull (1990, 1991) for currency options). Following largely numerical-analysis-based studies till then, Heston (1993) derives a closed-form solution to European options with stochastic volatility. Amin and Ng (1993) introduce both stochastic volatility and stochastic interest rate in a different approach, assuming the underlying stock price is correlated with consumption growth or market return, and generalizing the consumption based equilibrium of Rubinstein

Secondly, researchers began to realize that the Gaussian assumption for underlying asset return distribution might not perfectly be in line with what they observed in the actual data. In particular, higher moments such as skewness and kurtosis have been shown to affect the market participants' perception toward risk and hence the risk premiums. Alternative distributions and jump component have been proposed to capture the departure from the Gaussian distribution of underlying asset returns and abrupt movements in stock prices.[4]

Finally, the most recent improvements in option pricing research have been made by questioning the full rationality of investors. There exists ample evidence that investors may not be fully rational to the extent that behavioral aspect of investors does help better account for various empirical observations including stock returns, insider trading, mergers and acquisitions, etc.[5]

In particular, a growing body of literature shows that investor sentiment is a key determinant for asset prices. Investor sentiment refers to the overall attitude or aggregate belief of investors towards the anticipated price development of a particular security or financial market. Starting from early studies of the general mechanism through which mispricing is formed by investor sentiment (Barberis *et al.* (1998)), substantial amount of research in this regard has been devoted to its effect on the cross-section of stock returns (Baker and Wurgler (2006)). Because option prices are strictly tied down by no-arbitrage constraints such as put-call parity regardless of the model emplyed, one might be tempted to say: There is little, if any, room for behavioral factors, particularly investor sentiment, to play a role for asset return determination. However, since the volatility of the underlying asset is unknown and subject to temporal variation, option can be considered a speculative instrument of current and future volatilities, which might lead to irrational investor behavior and mispricing.

Indeed, as in the stock market, empirical evidence has been documented that sentiment-driven mispricing also exists in the option market (Stein (1989) and Poteshman (2001)).[6] Based on the finding that the risk-neutral distribution extracted from option prices could reflect investor sentiment, researchers in recent years have attempted to take a new routes to the relation between investor sentiment and the pricing behavior in the option market.[7] Since the risk-neutral skewness is proven to be determined by the slop of pricing kernel (Aït-Sahalia and Lo 1998, 2000), a significant relation between investor sentiment and the risk-neutral skewness would suggest that investor sentiment might affect option prices. By testing the relation of risk-neutral skewness with three sentiment proxies, namely bull-bear spread, net position of large speculators and the valuation error of index proposed by Sharpe (2002), Han (2008) shows when investors are bearish, the risk-neutral skewness is more negative and volatility smile is deeper; when investors turn more bullish, skewness becomes less negative and volatility smile gets flatter. Han concludes that investor sentiment is one

(1976) and Brennan (1979). Bakshi and Chen (1997a,b) study the valuation and hedging of foreign currency options with Lucas (1982) two-country model, assuming a stochastic structure for international economy.

[4]Bates (1991, 1996) derives a pricing model with jump-diffusion process to test whether there exists expectation of market crash before 1987, and later combines stochastic volatility model with jump-diffusion process by extending the methodology of Stein and Stein (1991) and Heston (1993). Scott (1997), and Bakshi and Chen (1997a) further extends the jump-diffusion model to incorporate both stochastic volatility and stochastic interest rate to reflect such empirical characteristics of underlying stock return as leptokurtosis, random change of volatility and negative correlation between stock returns and volatility. Some other studies relax the Gaussian assumption by applying Edgeworth series expansion (Jarrow and Rudd (1982)) or Gram-Charlier expansion (Corrado and Su (1996)). Borland (2002a,b) and Borland and Bouchaud (2004) develop a non-Gaussian stock pricing model by employing a Tsallis distribution of entropic index to capture the driving noise of underlying price. McCauley and Gunaratne (2003) use the exponential distribution generated from a Fokker-Planck equation instead to capture the probability of extreme outcomes.

[5]Traditional studies believe that the trading behavior of unsophisticated investors can be ignored, because their trades are random and thus cancel each other out or are exploited by competitive rational arbitrageurs who drive the asset prices back to their fundamentals. However, it is shown that noise trading and various limits to arbitrage can significantly drive the market price away from the fundamental value, and reduce the profit opportunity for arbitrageurs ((De Long *et al.* (1990), and Shleifer and Vishny (1997)).

[6]Note that the evidence in the option market is consistent with the general investor sentiment model proposed by Barberis *et al.* (1998). Also see other mispricing ....

[7]Earlier studies (e.g. Jarrow and Rudd (1982), Corrado and Su (1996)) build up non-Gaussian option pricing models with a correction part incorporating skewness and kurtosis without connecting to investor sentiment or any other behavioral factors.

of the key determinants element of option prices, but existing popular *rational* models including the ones with stochastic volatility, stochastic volatility with jumps, or asymmetric jumps, cannot account for the impact of investor sentiment.

As such, existing studies, by and large, attempt to relate empirical sentiment proxies with certain features of option prices such as risk-neural skewness and kurtosis to draw some indirect evidence on the pricing implications of investor sentiment. This is largely because theoretical option pricing models cannot accommodate the impact of investor sentiment in isolation. Our paper, on the other hand, directly admits investor sentiment to the analytic option pricing formula which we *theoretically* derive and provides direct empirical support that taking into account investor sentiment substantially reduces the pricing errors in option market.

The rest of the paper proceeds as follows. Section 2 shows in detail how we derive a closed-form solution for our GBS option pricing model, and the interpretation of the two key parameters newly incorporated in the model. Section 3 introduces the 50ETF option data set and provide the descriptive information. Section 4 presents the in-sample fit and out-of-sample prediction performances and the discussion on the economic meaning of the GBS model. Section 5 concludes.

2. Theoretical framework

2.1. *Standard BS model*

Black and Scholes (1973) show how to solve for option price with heat equation. We briefly summarize the key steps here because part of the derivation will be used to obtain the GBS pricing equation.[8]

The call option price is governed by the following partial differential equation:

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rC = 0 \tag{1}$$

with boundary condition

$$C\left(T,\ S\right) = \max\left(S - E,\ 0\right),$$

where E is the strike price. Through change of variables, we get

$$C\left(t,\ S\right) = E\mathrm{e}^{-\frac{1}{2}(k-1)x - \frac{1}{4}(k+1)^2\tau} u\left(\tau,\ x\right),$$

[8]Further details are presented in Appendix A.

where

$$x = \ln\left(\frac{S}{E}\right)$$

$$\tau = \frac{\sigma^2}{2}(T-t)$$

$$k = \frac{2r}{\sigma^2}$$

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2}$$

with boundary condition

$$\phi(x) = u(0,\ x) = \max\left[\mathrm{e}^{(1-\alpha)x} - \mathrm{e}^{-\alpha x},\ 0\right]$$

And the solution to the heat equation is

$$u(\tau,\ x) = \frac{1}{\sqrt{4\pi\tau}}\int_{-\infty}^{\infty}\phi(y)\,\mathrm{e}^{-\frac{(x-y)^2}{4\tau}}\mathrm{d}y \tag{2}$$

The final solution to the price of a call option is

$$C_{BS}(t,\ S) = S\mathrm{N}(d_1) - E\mathrm{e}^{-r(T-t)}\mathrm{N}(d_2)$$

where

$$d_1 = \frac{\log(S/E) + \left(r + \sigma^2/2\right)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = \frac{\log(S/E) + \left(r - \sigma^2/2\right)(T-t)}{\sigma\sqrt{T-t}}$$

2.2. *Generalized BS model*

Choi *et al.* (2017) show a generalized solution to the heat equation. Starting from Eq. (2), we define

$$\psi(\tau,\ x) \equiv u(\tau,\ x) = \frac{1}{\sqrt{4\pi\tau}}\int_{-\infty}^{\infty}\phi(y)\,\mathrm{e}^{-\frac{(x-y)^2}{4\tau}}\mathrm{d}y$$

$$\theta(\tau,\ x) \equiv -\frac{\partial\psi}{\partial x} = -\frac{1}{\sqrt{4\pi\tau}}\int_{-\infty}^{\infty}\phi(y)\left(-\frac{x-y}{2\tau}\right)\mathrm{e}^{-\frac{(x-y)^2}{4\tau}}\mathrm{d}y$$

$$\eta(\tau,\ x) \equiv -\frac{\partial\theta}{\partial x} = \frac{1}{\sqrt{4\pi\tau}}\int_{-\infty}^{\infty}\phi(y)\left[-\frac{1}{2\tau} + \frac{(x-y)^2}{4\tau^2}\right]\mathrm{e}^{-\frac{(x-y)^2}{4\tau}}\mathrm{d}y$$

In fact, $\theta(\tau,\ x)$ is the negative of the first-order derivative and $\eta(\tau,\ x)$ is the second-order derivative.

It is straightforward to prove that $\theta(\tau,\ x)$ and $\eta(\tau,\ x)$ satisfy the heat equations as specified below.[9]

$$\frac{\partial\psi}{\partial\tau}=\frac{\partial^2\psi}{\partial x^2}$$
$$\frac{\partial\theta}{\partial\tau}=\frac{\partial^2\theta}{\partial x^2}$$
$$\frac{\partial\eta}{\partial\tau}=\frac{\partial^2\eta}{\partial x^2}$$

Then we linearly combine the three terms to obtain the generalized solution to the heat equation. Here, λ and κ are introduced as coefficients and assumed to be constants for simplicity.

$$\varphi(\tau,\ x)=\psi(\tau,\ x)+\lambda\theta(\tau,\ x)+\kappa\eta(\tau,\ x) \tag{3}$$

We obtain the expression of $\varphi(\tau,\ x)$ below.[10]

$$\varphi(\tau,\ x)=I_0+I_1+I_2+I_3+I_4,$$

where

$$I_0=\left(1-\frac{\kappa}{2\tau}\right)\left[\mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}\mathrm{N}(d_1)-\mathrm{e}^{\frac{1}{2}(k-1)x+\frac{1}{4}(k-1)^2\tau}\mathrm{N}(d_2)\right]$$
$$I_1=-\frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}\left[\mathrm{e}^{-\frac{1}{2}d_1^2}+(k+1)\sqrt{\pi\tau}\mathrm{N}(d_1)\right]$$
$$I_2=\frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{\frac{1}{2}(k-1)x+\frac{1}{4}(k-1)^2\tau}\left[\mathrm{e}^{-\frac{1}{2}d_2^2}+(k-1)\sqrt{\pi\tau}\mathrm{N}(d_2)\right]$$
$$I_3=\frac{\kappa}{\sqrt{2\pi}2\tau}\mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}\left\{\left[1+\frac{1}{2}(k+1)^2\tau\right]\sqrt{2\pi}\mathrm{N}(d_1)+\left[(k+1)\sqrt{2\tau}-d_1\right]\mathrm{e}^{-\frac{1}{2}d_1^2}\right\}$$
$$I_4=-\frac{\kappa}{\sqrt{2\pi}2\tau}\mathrm{e}^{\frac{1}{2}(k-1)x+\frac{1}{4}(k-1)^2\tau}\left\{\left[1+\frac{1}{2}(k-1)^2\tau\right]\sqrt{2\pi}\mathrm{N}(d_2)+\left[(k-1)\sqrt{2\tau}-d_2\right]\mathrm{e}^{-\frac{1}{2}d_2^2}\right\}$$

And the boundary condition still holds

$$\lim_{\tau\to0}\varphi(\tau,\ x)=\max\left[\mathrm{e}^{\frac{1}{2}(k+1)x}-\mathrm{e}^{\frac{1}{2}(k-1)x},\ 0\right]$$

Then we obtain

$$\begin{aligned}V(\tau,\ x)=&\left[1-\frac{\lambda}{2}(k+1)+\frac{\kappa}{4}(k+1)^2\right]\mathrm{e}^{x}\mathrm{N}(d_1)-\left[1-\frac{\lambda}{2}(k-1)+\frac{\kappa}{4}(k-1)^2\right]\mathrm{e}^{-\tau k}\mathrm{N}(d_2)\\&-\frac{1}{\sqrt{4\pi\tau}}\left[\lambda-\frac{\kappa}{2}(k+1)+\frac{\kappa}{2\tau}x\right]\mathrm{e}^{x-\frac{1}{2}d_1^2}+\frac{1}{\sqrt{4\pi\tau}}\left[\lambda-\frac{\kappa}{2}(k-1)+\frac{\kappa}{2\tau}x\right]\mathrm{e}^{-\tau k-\frac{1}{2}d_2^2}\end{aligned}$$

Note that we can get the following relationship, which helps us greatly simplify the final solution.[11]

$$\mathrm{e}^{x-\frac{1}{2}d_1^2}=\mathrm{e}^{-\tau k-\frac{1}{2}d_2^2}$$

[9] Later we will prove that the boundary condition still holds.
[10] See Appendix B for further details
[11] See Appendix C for further details

Thus

$$
\begin{aligned}
V(\tau,\ x) = & \left[1 - \frac{\lambda}{2}(k+1) + \frac{\kappa}{4}(k+1)^2\right] \mathrm{e}^{x} \mathrm{N}(d_1) \\
& - \left[1 - \frac{\lambda}{2}(k-1) + \frac{\kappa}{4}(k-1)^2\right] \mathrm{e}^{-\tau k} \mathrm{N}(d_2) + \frac{\kappa}{\sqrt{4\pi\tau}} \mathrm{e}^{x - \frac{1}{2}d_1^2}
\end{aligned}
$$

So the final closed-form solution to the price of a call option is

$$
\begin{aligned}
C_{GBS}(t,\ S) = & \left[1 - \frac{\lambda}{2}(k+1) + \frac{\kappa}{4}(k+1)^2\right] S\mathrm{N}(d_1) \\
& - \left[1 - \frac{\lambda}{2}(k-1) + \frac{\kappa}{4}(k-1)^2\right] E\mathrm{e}^{-r(T-t)} \mathrm{N}(d_2) + \frac{\kappa}{\sigma\sqrt{2\pi(T-t)}} S\mathrm{e}^{-\frac{1}{2}d_1^2}
\end{aligned} \tag{4}
$$

where

$$
k = \frac{2r}{\sigma^2}
$$

We can obtain the relation between the two models because

$$
C_{GBS}(t,\ S) = E\mathrm{e}^{-\frac{1}{2}(k-1)x - \frac{1}{4}(k+1)^2\tau} \left[\psi - \lambda \frac{\partial \psi}{\partial x} + \kappa \frac{\partial^2 \psi}{\partial x^2}\right]
$$

Letting $\omega = -\frac{1}{2}(k-1)x - \frac{1}{4}(k+1)^2\tau$,

$$
C_{GBS}(t,\ S) = C_{BS} - \lambda E\mathrm{e}^{\omega} \frac{\partial \psi}{\partial x} + \kappa E\mathrm{e}^{\omega} \frac{\partial^2 \psi}{\partial x^2}
$$

First calculate ψ, $\frac{\partial \psi}{\partial x}$, $\frac{\partial^2 \psi}{\partial x^2}$

$$
\begin{aligned}
\psi(\tau,\ x) &= \frac{C_{BS}}{E\mathrm{e}^{\omega}} \\
\frac{\partial \psi}{\partial x} &= \mathrm{e}^{-\omega}\left[\frac{(k-1)}{2E} C_{BS} + \mathrm{e}^{x} \frac{\partial C_{BS}}{\partial S}\right] \\
\frac{\partial^2 \psi}{\partial x^2} &= \mathrm{e}^{-\omega}\left[\frac{(k-1)^2}{4E} C_{BS} + k\mathrm{e}^{x} \frac{\partial C_{BS}}{\partial S} + E\mathrm{e}^{2x} \frac{\partial C_{BS}^2}{\partial S^2}\right]
\end{aligned}
$$

The call option price of GBS model is then equal to:

$$
C_{GBS}(t,\ S) = C_{BS} - \lambda\left[\frac{(k-1)}{2} C_{BS} + S\Delta\right] + \kappa\left[\frac{(k-1)^2}{4} C_{BS} + kS\Delta + S^2\Gamma\right], \tag{5}
$$

where

$$
\begin{aligned}
\Delta &= \frac{\partial C_{BS}}{\partial S} = \mathrm{N}(d_1) \\
\Gamma &= \frac{\partial C_{BS}^2}{\partial S^2} = \frac{\mathrm{N}'(d_1)}{S\sigma\sqrt{T-t}} = \frac{1}{S\sigma\sqrt{T-t}} \frac{\mathrm{e}^{-\frac{1}{2}d_1^2}}{\sqrt{2\pi}}
\end{aligned}
$$

Note that Equation (4) is equivalent to Eq. (5), and can be expressed as

$$C_{GBS} = J_1 C_{BS} + J_2 \Delta + J_3 \Gamma, \tag{6}$$

where

$$\begin{aligned} J_1 &= 1 - \lambda \frac{(k-1)}{2} + \kappa \frac{(k-1)^2}{4} \\ J_2 &= (-\lambda + \kappa k)\, S \\ J_3 &= \kappa S^2 \end{aligned}$$

Equation (6) shows that C_{GBS} can be understood as a modification of C_{BS} by incorporating option's Δ and Γ, with adjustment parameters λ and κ. Note that while no restrictions are imposed on λ and κ, it can easily be shown that λ negatively affects and κ positively affects the call price:

$$\begin{aligned} -\left[\frac{(k-1)}{2} C_{BS} + S\Delta\right] &< -\left[-\frac{1}{2} C_{BS} + S\Delta\right] \\ &= -\frac{1}{2}\left[S\mathrm{N}(d_1) + \mathrm{e}^{-r(T-t)}\mathrm{N}(d_2)\right] \\ &< 0 \\ \frac{(k-1)^2}{4} C_{BS} + kS\Delta + S^2\Gamma &> 0 \end{aligned}$$

2.3. *Interpretation of the GBS model*

To better understand the pricing implications of λ and κ, we first extend the pricing formula in Equation (6) using the following result.
If we assume the n-1st order derivative (with a negative sign) satisfies the heat equation

$$\psi_{n-1}(\tau,\ x) \equiv (-1)^{n-1} \frac{\partial^{n-1}\psi}{\partial x^{n-1}},$$

then the negative n-th order derivative (with a negative sign)

$$\psi_n(\tau,\ x) \equiv -\frac{\partial \psi_{n-1}}{\partial x}$$

also satisfies the heat equation, because

$$\frac{\partial \psi_n}{\partial \tau} = -\frac{\partial^2 \psi_{n-1}}{\partial \tau \partial x} = -\frac{\partial}{\partial x}\frac{\partial^2 \psi_{n-1}}{\partial x^2} = \frac{\partial^2 \psi_n}{\partial x^2}$$

Hence we can expand Eq. (3) to an infinite series with additional derivatives

$$\varphi'(\tau,\ x) = \psi(\tau,\ x) + \lambda\theta(\tau,\ x) + \kappa\eta(\tau,\ x) + \ldots \tag{7}$$

Since the coefficients in Eq. (7) are not constrained *ex ante*, we take the liberty of setting them equal to the corresponding coefficients in a Taylor series expansion. Namely,

$$\lambda = -a,\ \ \kappa = \frac{1}{2}a^2,\ \ \ldots$$

By doing so and applying the Taylor series expansion inversely to the right-hand side of Eq. (7), we endow the parameters with economically meaningful interpretations as further derivation below demonstrates.

$$\begin{aligned}\varphi'(\tau,\ x) &= \psi(\tau,\ x) + a\frac{\partial\psi}{\partial x} + \frac{1}{2}a^2\frac{\partial^2\psi}{\partial x^2} + \ldots \\ &= \psi\left(\tau,\ x+a\right)\end{aligned}$$

The resultant call option price from the GBS model under this parametrization is

$$\begin{aligned}C'_{GBS}\left(\tau,\ x\right) &= E\mathrm{e}^{-\frac{1}{2}(k-1)x-\frac{1}{4}(k+1)^2\tau}\varphi\left(\tau,\ x\right) \\ &= E\mathrm{e}^{-\frac{1}{2}(k-1)x-\frac{1}{4}(k+1)^2\tau}\psi\left(\tau,\ x+a\right) \\ &= \mathrm{e}^{\frac{1}{2}(k-1)a}C_{BS}\left(\tau,\ x+a\right)\end{aligned}$$

By changing the notation in terms of time and spot price, we obtain

$$C'_{GBS}\left(t,\ S\right) = \left(\frac{S+\mathit{\Delta}S}{S}\right)^{\left(\frac{r}{\sigma^2}-\frac{1}{2}\right)} C_{BS}\left(t,\ S+\mathit{\Delta}S\right), \tag{8}$$

where

$$\begin{aligned}\mathit{\Delta}S = E\mathrm{e}^{x+a} - E\mathrm{e}^{x} = S\left(\mathrm{e}^{a}-1\right) \\ a = \ln\left(\frac{S+\mathit{\Delta}S}{S}\right)\end{aligned}$$

The result in Eq. (8) is quite intuitive. The option price derived from the GBS model is proportional to the price of the standard BS model when the spot price used is $S+\mathit{\Delta}S$ rather than S. In other words, when investors assess the value of an option with GBS model, the situation is equivalent to the one where they are still applying the BS formula, but with a *potentially* diverted spot value. Therefore, $\mathit{\Delta}S$ can be interpreted as a mental deviation from the actual spot price due to investor sentiment arising from the future return prospect they have in their mind.[12] Furthermore, a is shown to be exactly equal to the continuously compounded rate of return, which renders λ, and κ intuitive sentiment measures, each related to investors' belief in regard to future return and its volatility. For example, when the market keeps going up, no matter what the underlying reason might be, investors tend to become increasingly confident that stock prices will go even higher. Under this situation, they are likely to long more call options for some speculative profits, leading to higher call prices. In the BS framework, the mechanism of such phenomenon can not be explained; in GBS model, on the other hand, the option price can be adjusted by λ and κ. When investors are bullish (bearish), we expect λ to be negative (positive) and κ to be larger (smaller), both leading to positive (negative) adjustments to the BS price. From a broader perspective, it is noteworthy that our GBS model and the benchmark BS model coincide if investor sentiment is *neutral*. In other words, the GBS model nests the BS model as a special case in which there exists no room for sentiment effect, just as in a *fully rational* world. [13]

[12]We empirically validate this insight in later sections.

[13]At this juncture, it should be noted that we derive the GBS option price in equation (8) by first expanding the generalized solution for the heat equation to an infinite series. We then convert the series back to the BS framework using an economically intuitive parametrization. For the purpose of the empirical analysis of the model, however, we truncate the series at the second order derivative for two reasons. First, as a practical matter, it is simply impossible to estimate an infinite number of parameters. Second, only the first- and second-order derivatives of options are widely used in reality which is consistent with the option's Δ and Γ in the final solution in equation (6).

2.4. *Sensitivity analysis*

The main purpose of the analysis in this subsection is to examine i) how underlying pricing determinants of the BS model will interact with λ and κ of the GBS model, and ii) whether the pricing effects of those parameters in the GBS model are consistent with the interpretation we addressed in the previous section.

Figure 1 shows the sensitivity analysis of spot and strike prices for the four different cases of (λ, κ) pair. Overall, the GBS model and the BS model share a quite similar response pattern for each case, and the GBS-induced adjustment is small in magnitude partly due to parameter values assigned. More importantly, when we set λ to be positive, GBS model yields a lower price than BS model; and if we set κ to be positive, the GBS price is higher, which is consistent with our interpretation.

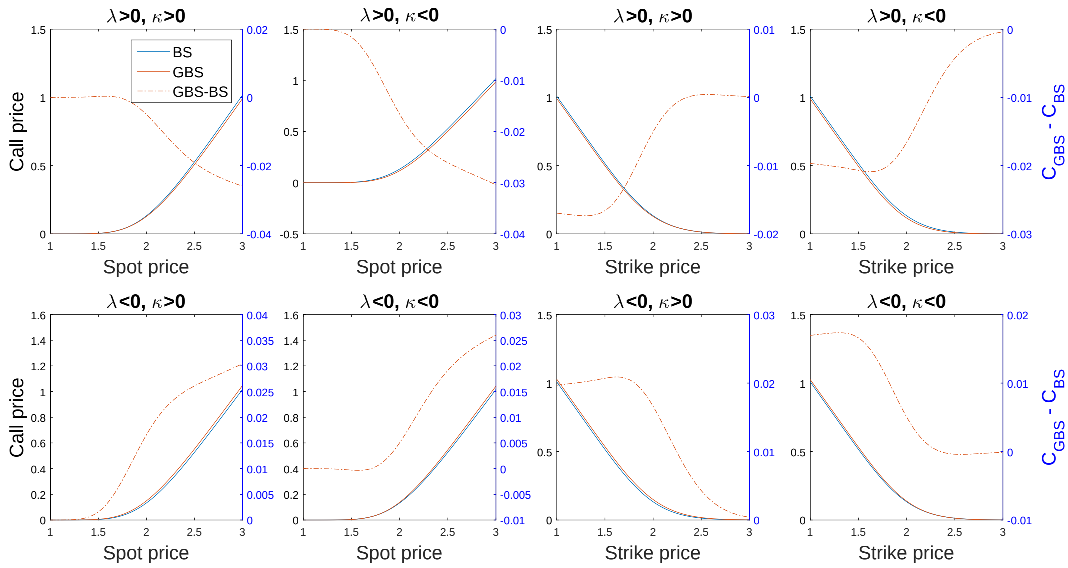

Figure 1. **Sensitivity analysis 1)**. The figure shows the sensitivity analysis of spot price and strike price. Holding the other parameters constant, in particular, $r = 0.03$, $vol = 0.3$, $delta\_t = 100/365$, we set strike price K to be 2 and let spot price S increase from 1 to 3 in the left four figures in which λ are set to be 0.01 and -0.01, κ are set to be 0.001 and -0.001, respectively. Likewise, we set spot price S to be 2 and let strike price K increase from 1 to 3 in the right four figures.

Figure 2 shows the sensitivity analysis of interest rate and volatility. Option price increases with both interest rate and volatility for the same reason in the BS model. As seen in Figure 1, positive investor sentiment consistently leads to the higher prices of the GBS model.

Figure 3 shows the sensitivity analysis of maturity, λ and κ. In particular, the right four figures illustrate the influence of λ and κ. Again, the result is consistent with our analysis that λ negatively affects option prices, whereas price effect of κ is positive.

3. Data

3.1. *Introduction of 50ETF option*

This study uses a dataset of daily close prices of 50ETF call option (delisted and on board) from February 9, 2015 to February 28, 2017. 50ETF option was listed on the Shanghai Stock Exchange (SSE) on February 9, 2015 (the same date as our data start), and is the first publicly traded option in Chinese market. Currently, there are more options being traded on SSE including options on

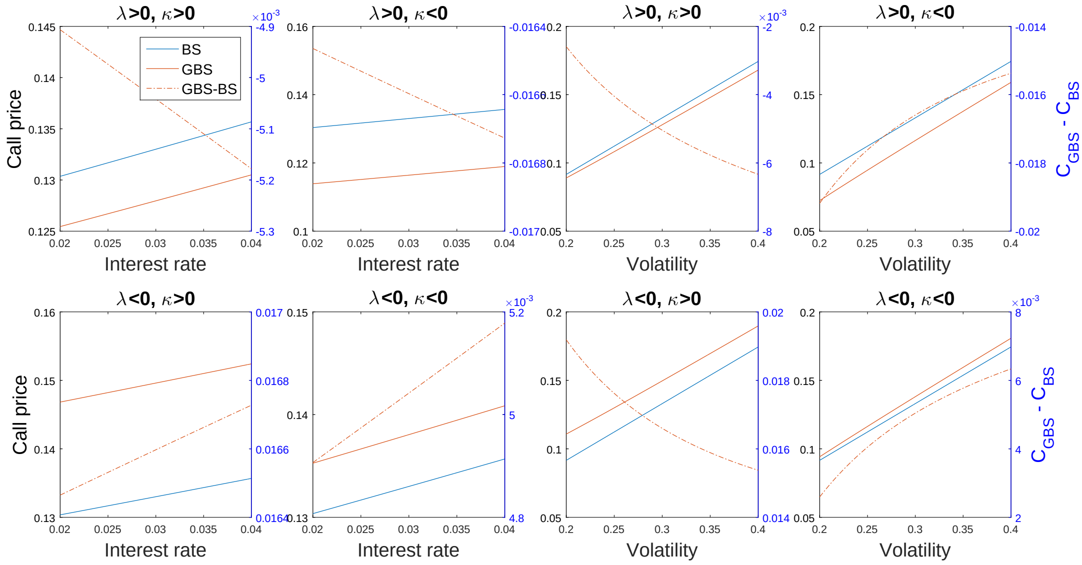

Figure 2. **Sensitivity analysis 2)**. The figure shows the sensitivity analysis of interest rate and volatility. Holding other parameters constant, in particular, $S = 2$, $K = 2$, $delta\_t = 100/365$, we set volatility σ to be 0.3 and let interest rate r increase from 0.02 to 0.04 in the left four figures in which λ are set to be 0.01 and -0.01, κ are set to be 0.001 and -0.001, respectively. Likewise, we set interest rate r to be 0.03 and let volatility σ increase from 0.2 to 0.4 in the right four figures.

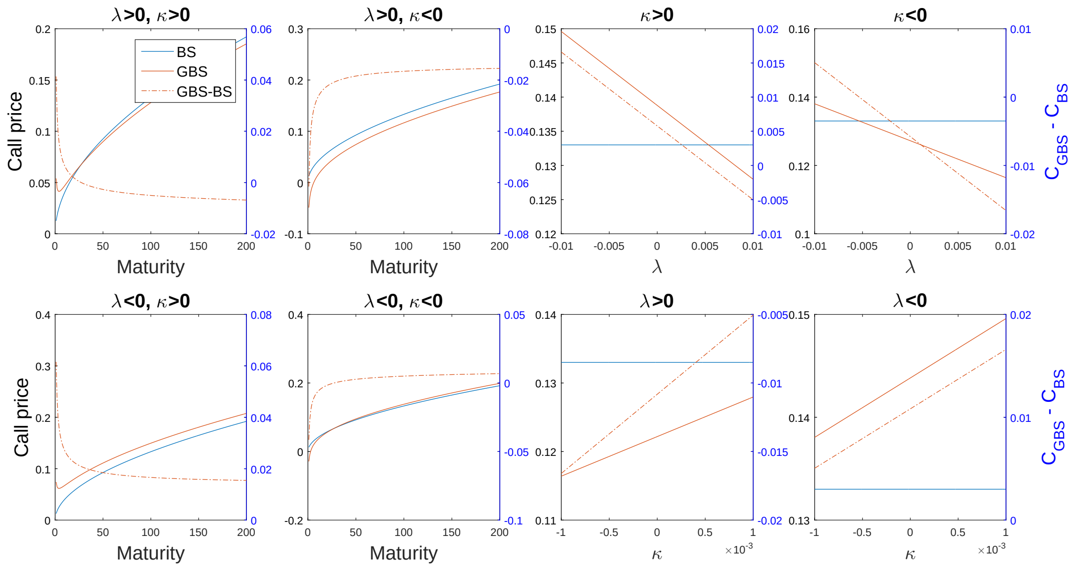

Figure 3. **Sensitivity analysis 3)**. The figure shows sensitivity analysis of maturity, λ and κ. Holding other parameters constant, in particular, $S = 2$, $K = 2$, $r = 0.03$, $vol = 0.3$, we let maturity $delta\_t$ increase from 1/365 to 200/365 in the left four figures in which λ is set to be 0.01 and -0.01, κ is set to be 0.001 and -0.001, respectively. Then we set $delta\_t = 100/365$, and let λ increase from -0.01 to 0.01 in the upper right two figures in which κ is set to be 0.001 and -0.001,respectively. Likewise, we let κ increase from -0.001 to 0.001 in the lower right two figures in which λ is set to be 0.01 and -0.01, respectively.

ETFs, stocks, indexes ,and commodity futures. Table 1 shows all the existing option contracts in China. We believe that 50ETF option data is among the best datasets for studying the effect of investor sentiment on option prices partly because the vast majority of investors in China are

Table 1. **Summary of Options in Chinese Market**

Exchange	Name	Style	Status
Shanghai Stock Exchange(SSE)	50ETF option	European	Trading
	Pingan Bank option	European	Simulating
	Shanghai Automobile option	European	Simulating
	Huaan SSE 180ETF option	European	Simulating
Shenzhen Stock Exchange (SZE)	Vanke A option	European	Simulating
	SZ 100ETF option	European	Simulating
	Originwater option	European	Simulating
	Efund Second-Board option	European	Simulating
	Nf-bearings option	European	Simulating
	HS 300ETF option	European	Simulating
	Goertek option	European	Simulating
	SME board option	European	Simulating
China Financial Futures Exchange (CFFEX)	HS 300 Index option	European	Simulating
	SH 50 Index option	European	Simulating
Dalian Commodity Exchange (DCE)	Bean pulp option	American	Trading
Shanghai Futures Exchange (SHFE)	Gold option	American	Simulating
	Copper option	American	Simulating
Zhengzhou Commodity Exchange (ZCE)	White Sugar option	American	Trading

Source: WIND database

Table 2. **Main Features of 50ETF Option Contract**

Underlying asset	Huaxia SSE 50 Index open exchange traded securities investment fund (510050.SH)
Contract type	Call options and put options
Contract unit	10000
Expiration month	Current month, next month and the following two consecutive quarters
Strike price	5 (1 at-the-money option, 2 in-the-money options, 2 out-of-money options)
Expiration date	The fourth Wednesday of each expiration month
Option Style	European style
Contract delivery	Physical delivery

Source: WIND database

individual investors, who tend to be much more prone to sentiment or subjective beliefs. Also the 50ETF option is the only publicly traded European option in China while most of other options are either in experimentation phase or American type. To help better understand the nature of our dataset, Table 2 presents the important contract details of 50ETF option.

The listing practice in the SSE requires that call options and put options be listed as a pair with the same strike price and expiration date. For example, on February 9, 2015, call option '10000001.SH' was listed on SSE with strike price 2.2000 and expiration date 2015/03/25, while on the same day, put option '10000006.SH' was also on board with strike price 2.2000 and expiration date 2015/03/25. Due to this unique listing feature, we only pay our attention to call options while expecting similar conclusions for puts.

3.2. *Descriptive statistics*

Following Bakshi *et al.* (1997), we divide all call option data into different categories. This methodology has two main benefits. First, we can examine the volatility smile phenomenon by calculating the average implied volatility of each category, thereby investigate whether there exist pricing biases in Chinese market. Second, the estimation of implied volatility (or implied parameters) of each category help us understand how pricing performances of an option pricing model can be improved. Whereas Bakshi *et al.* (1997) obtain 18 categories in total, we divide options into 24 categories for a smoother distribution in each category. We take the following steps for the classification and present the result in Table 3.

(i) First, according to moneyness measured by S/K, options are divided into out-of-the-money (OTM) if $S/K < 0.97$, at-the-money (ATM) if $0.97 \leq S/K < 1.03$ and in-the-money (ITM) if $1.03 \leq S/K$. Then, in each group, options are divided into several categories. There are 8 categories in total;

(ii) Second, according to maturity measured by days-to-expiration, options are divided into short-term (ST) if $T - t < 60$, medium-term (MT) if $60 \leq T - t < 150$ and long-term (LT) if $150 \leq T - t$.

Table 3 summarizes the mean, standard deviation (in parenthesis), and total number of option prices in each category, as well as the subtotal number of each group.

4. Discussion

4.1. *Volatility smile*

Volatility smile indicates pricing biases across both moneyness and maturity. To investigate whether there exist such pricing biases in Chinese market, we calculate the implied volatility of each option from the market price to estimate the average value for each category.

Table 4 shows that BS implied volatility exhibits an apparent smile shape. Implied volatility decreases sharply as option strike price decreases. When option goes deep in-the-money (ITM), implied volatility increases again. However, implied volatility is quite steady across different maturities. This result shows that the BS model is indeed subject to mispricing in Chinese market, especially across moneyness. Figure 4 clearly displays the implied volatility structure.

4.2. *In-sample fit performance*

To compare the in-sample fit performances of the two models, we estimate the implied parameters in the following steps:

(i) Collect all call option prices from the same day. We set the annual interest rate $r = 0.03$ for a good reason that will be explained below. Denote the parameter set by $\Phi_{it} = \{S_t,\ K_i,\ T_i,\ t,\ \sigma_t,\ \lambda_t,\ \kappa_t\}$, theoretical option price by $\hat{C}_{it}(\Phi_{it})$, and the market price by C_{it}. Then we define the pricing error of option i in day t:

$$\epsilon_{it}(\Phi_{it}) = \hat{C}_{it}(\Phi_{it}) - C_{it}$$

Table 3. **Descriptive statistics**

		Maturity (Days-to-expiration)			
Moneyness (S/K)		< 60	$60 - 150$	≥ 150	Subtotal
OTM	< 0.91	0.0095	0.0433	0.1221	5637
		(0.0209)	(0.0449)	(0.0882)	
		2799	2053	785	
	$0.91 - 0.94$	0.0263	0.0758	0.1362	1894
		(0.0385)	(0.0668)	(0.1001)	
		854	592	448	
	$0.94 - 0.97$	0.0331	0.0866	0.1410	2901
		(0.0421)	(0.0700)	(0.0970)	
		1369	808	724	
ATM	$0.97 - 1$	0.0504	0.1111	0.1656	2961
		(0.0480)	(0.0730)	(0.0989)	
		1410	808	743	
	$1 - 1.03$	0.0820	0.1425	0.1946	2745
		(0.0505)	(0.1425)	(0.1946)	
		1315	738	692	
ITM	$1.03 - 1.06$	0.1298	0.1824	0.2330	2623
		(0.0512)	(0.0816)	(0.1086)	
		1232	734	657	
	$1.06 - 1.09$	0.1836	0.2288	0.2746	2172
		(0.0547)	(0.0878)	(0.1137)	
		1033	638	501	
	≥ 1.09	0.4089	0.4644	0.4283	6159
		(0.1978)	(0.2252)	(0.1834)	
		2892	2394	873	
	Subtotal	12904	8765	5423	27092

This table calculates the average, standard deviation, and total number of option prices in each category, as well as the subtotal number of each subgroup.

(ii) Find parameters $\sigma_t,\ \lambda_t,\ \kappa_t$ that minimize

$$SSE_t = \sum_{i=1}^{N} \epsilon_{it}^2(\Phi_{it})$$

Table 4. **Implied volatility**

Implied volatility (%)		Maturity (Days-to-expiration)		
Moneyness (S/K)		< 60	$60 - 150$	≥ 150
OTM	< 0.91	36.55	35.72	36.94
	$0.91 - 0.94$	30.91	26.85	27.27
	$0.94 - 0.97$	26.16	24.22	24.12
ATM	$0.97 - 1$	23.69	23.21	22.96
	$1 - 1.03$	21.81	22.20	21.92
ITM	$1.03 - 1.06$	20.79	20.99	21.11
	$1.06 - 1.09$	20.37	19.17	20.20
	≥ 1.09	27.18	23.94	21.96

This table shows the BS implied volatility in each of the 24 categories. The implied volatility of each option is first calculated from the market prices and then averaged in each category.

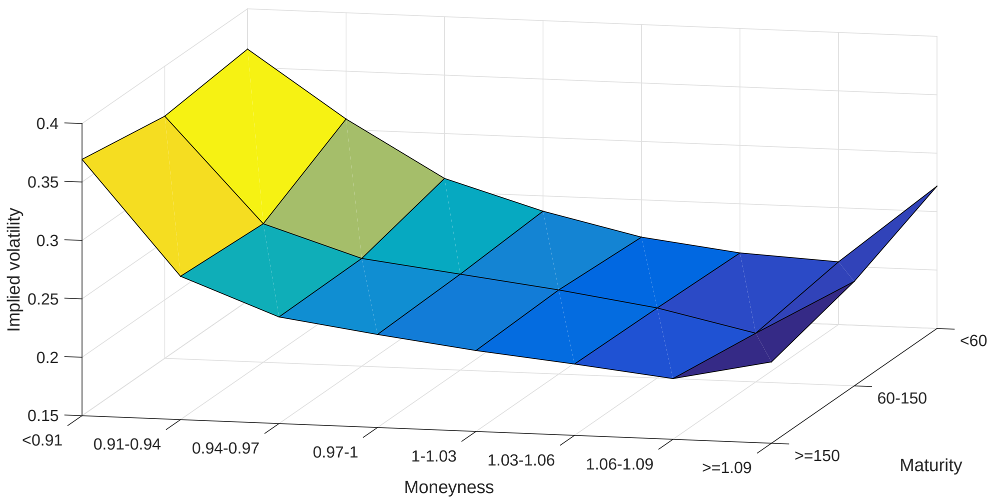

Figure 4. **Volatility smile**. This figure exhibits the implied volatility surface in Chinese option market. The x-axis represents different moneynesses, y-axis represents different maturities, and z-axis represents the average implied volatility.

(iii) Go back to step 1 and repeat the above procedure for each day.

We assume the annual interest rate to be constant at 3% because of the following three reasons. 1) The main purpose of this paper is not to study the influence of interest rate, and as such, a constant interest rate can highlight our empirical emphasis and simplify the estimation procedure.; 2) Wang *et al.* (2017) propose an option pricing model under the Students t distribution, assuming a constant interest rate at 5%; 3) AS of April 19, 20017, the overnight Shanghai Interbank Offered Rate (Shibor) is about 2.5% and 6-month rate is about 4%. It seems reasonable, therefore, to assume the risk-free rate is 3%. The objective function SSE_t represents the overall daily pricing

Table 5. **Fit parameters**

Parameters		$\sigma(\%)$	λ	κ	SSE
BS	average	23.68	-	-	0.0137
	std. err.	0.54	-	-	-
GBS	average	24.79	0.0101	0.0010	0.0066
	std. err.	0.53	0.0005	0.0001	-

This table shows means and standard errors of daily parameter estimates and the average SSE. σ denotes the annualized volatility. We estimate the parameters of the BS model and GBS model on a daily basis using all available option prices each day before calculating the statistics reported in the table.

bias of each particular group of options.[14]

The GBS model estimates slightly higher average and lower standard deviation of implied volatility, and a positive average of the two parameters each. Strikingly, the GBS model yields a much lower average SSE, i.e., less than half of the BS counterpart. Figure 5 shows that the GBS model outperforms the BS model every single day with a 100% win-ratio as measured by the incremental SSE of the GBS model over the BS model.

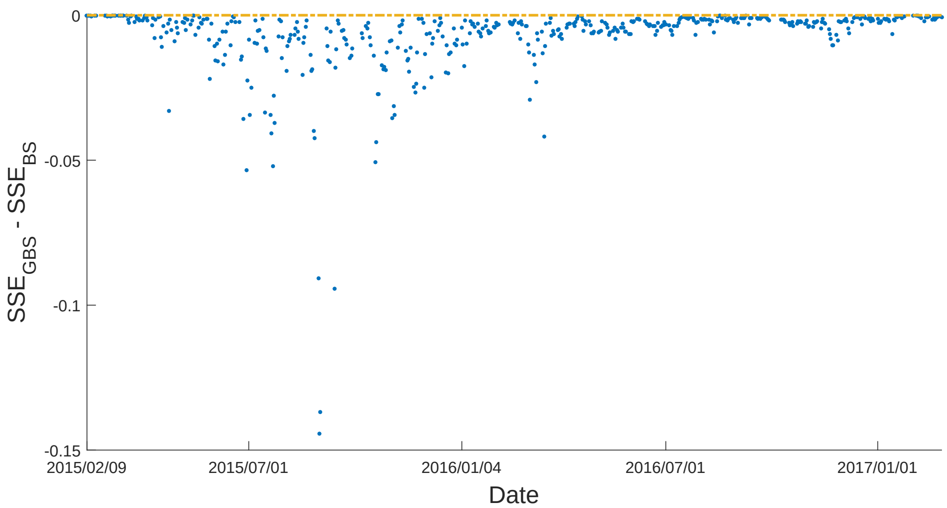

Figure 5. **Daily relative fits of GBS and BS**. This figure shows the daily difference in fit performance between the two models measured by SSE of the GBS model less SSE of the BS model.

We further compare the relative fits of the models in differen moneyness- and maturity-sorted categories. Figure 6 shows that in most cases the GBS model outperforms the BS model by a large margin with one exception in which the BS model is just slightly outperforms. Note that

[14]For this part of the analysis, we treat all option prices in each day as a single group to back out the same implied parameter for options with different moneyness levels and maturities. We take this approach to examine the overall impact of investor sentiment on option prices although it deviates in spirit from our analysis of implied volatility. In a later part we further divide options into different categories according to moneyness and maturity to investigate how well the GBS model captures the structure of the implied parameters.

this result is not in contradiction to the previous 100% win-ratio, because when we solve for the implied parameters, our objective function is the overall pricing error in each day, rather than in each category. Hence, it does not necessarily mean the GBS model will outperform in all categories.

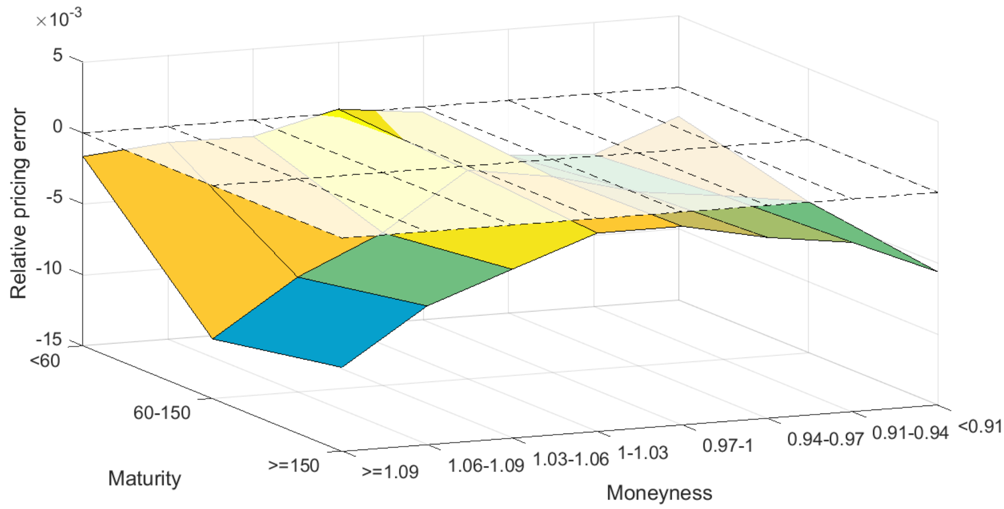

Figure 6. **Relative fitness of two models in each category**. For each category, we calculate the average pricing error of two models and calculate $|\bar{\epsilon}_{GBS}| - |\bar{\epsilon}_{BS}|$ to measure the relative fitness.

Recall that based on the theoretical analysis in section 2, we interpret that λ and κ both reflect investor sentiment toward future return prospect. Further, we expect λ is negatively related to return expectation and κ is positively related to the squared expectation. We run the regressions to investigate whether there exists a significant relationship between each of the two GBS parameters and a set of past index returns as our sentiment proxies.

For our empirical investigation, we first calculate the weekly average values of λ, κ and NAV of 50ETF. We use the weekly averages because investor sentiment is not deemed as volatile as daily trading data or stock price movements. Indeed, existing literature usually uses one week or one month as an appropriate time period. The recent returns of past 1, 2, 3 and 4 weeks are calculated based on the weekly NAV of 50ETF.

It is widely known that stock index returns, in most cases, exhibit positive autocorrelations (Lo and MacKinlay (1988)). Ample empirical evidence also suggests that investors tend to form their expectations of future returns, heavily relying on the past performances (Amin *et al.* (2004)) to the extent that recent return might serve as the most important determinant of investor sentiment (Brown and Cliff (2004)).[15] Thus we choose the historical returns as the proxies of investors' belief in regard to future return prospect. We report the correlation coefficient estimates in Table 6.[16] All negative correlations between λ and the historical returns strongly suggest that the parameter is significantly related to the investor sentiment captured by the past returns. Interestingly, the correlation becomes increasingly stronger as the return horizon extends up to a month. Although this result seems to warrant a deeper look, a plausible story, as we postulate, is that investors might rely more heavily on the monthly return history in forming their subjective belief or sentiment compared to the past returns over shorter time span.

We then proceed to regress λ on the historical returns of past 1, 2, 3 and 4 weeks, respectively. We also regress κ on the squared historical returns of past 1, 2, 3 and 4 weeks in the same manner. The first-order lag term is included in both regressions to correct for the potential autocorrelations.

[15] Amin *et al.* (2004) demonstrate that the momentum of stock market can have impact on option price, and investors' expectation of future return could be one of the channels.
[16] λ_w denotes weekly average λ.

Table 6. **Correlation coefficients 1**

	λ_w	ret_{1w}	ret_{2w}	ret_{3w}	ret_{4w}
λ_w	1	-0.3212	-0.5040	-0.5082	-0.5279
ret_{1w}		1	0.7739	0.6143	0.5524
ret_{2w}			1	0.8625	0.7519
ret_{3w}				1	0.9077
ret_{4w}					1

This table presents the correlation matrix among the estimated λ and historical returns. λ_w denotes weekly average λ, and ret_{1w}, ret_{2w}, ret_{3w} and ret_{4w} denote past recent returns of 1, 2, 3 and 4 weeks, respectively.

Table 7. **Regression results 1**

λ_w	(1)	(2)	(3)	(4)
$(Intercept)$	0.0041***	0.0035***	0.0042***	0.0046***
	(4.3915)	(3.6495)	(3.9440)	(4.3732)
λ_w_lag	0.6048***	0.6575***	0.5936***	0.5534***
	(8.8578)	(8.8958)	(7.3308)	(6.8598)
ind_1	-0.0621***	-0.0765***	-0.0385**	-0.0191
	(-3.2200)	(-5.1584)	(-2.3715)	(-1.2152)
ind_1_lag	-0.0650***	0.0263	0.0043	-0.0168
	(-3.2049)	(1.5921)	(0.2531)	(-1.0263)
Adj. R^2	0.584	0.594	0.520	0.532

This table shows the regression results of λ on historical returns. λ_w denotes weekly average λ, ind_1 denotes ret_{1w}, ret_{2w}, ret_{3w} and ret_{4w} respectively. And the first-order lag terms are introduced to remove auto-correlation. t-statistics are reported in the parenthesis. *, ** and *** denote significant at 10%, 5% and 1% level, respectively.

Table 7 presents the regression results of λ, where ind_1 denotes ret_{1w} , ret_{2w}, ret_{3w} and ret_{4w}, respectively. In all four regressions, we find that λ is negatively related to the sentiment proxy. When the market has been going up over a certain time, investors are likely to expect that the upward trend will continue and become highly bullish as a consequence. In this situation, they tend to seek to exploit this expectation by buying more calls to speculate, thereby pushing up the call prices.[17] and such an effect of positive expectation is reflected in a negative λ.

[17] In line with this reasoning, several researches have proposed demand-based option pricing models that connect buying pressure with the daily changes of implied volatility(Bollen and Whaley (2004)) or pricing kernel (Garleanu *et al.* (2009)). Garleanu *et al.* (2009) conclude that when investors anticipate positive expected returns, they take actions by longing more call, which turned to a positive demand pressure.

Table 8. **Correlation coefficients 2**

	κ_w	ret^2_{1w}	ret^2_{2w}	ret^2_{3w}	ret^2_{4w}
κ_w	1	0.5624	0.7472	0.6731	0.5066
ret^2_{1w}		1	0.7836	0.4612	0.4552
ret^2_{2w}			1	0.8140	0.6472
ret^2_{3w}				1	0.8657
ret^2_{4w}					1

This table shows the correlation matrix among estimated κ and squared historical returns. κ_w denotes weekly average κ, ret^2_{1w}, ret^2_{2w}, ret^2_{3w} and ret^2_{4w} denote squared past recent returns of 1, 2, 3 and 4 weeks, respectively.

Table 9. **Regression results 2**

κ_w	(1)	(2)	(3)	(4)
$(Intercept)$	0.0002**	0.0002	0.0003*	0.0004*
	(2.0843)	(1.4475)	(2.5415)	(2.6054)
κ_w_lag	0.2548***	0.3803***	0.4491***	0.4229***
	(3.4070)	(4.1736)	(5.0323)	(4.6189)
ind_2	0.2761***	0.2063***	0.1561***	0.0892***
	(8.4322)	(10.0847)	(8.8767)	(5.1193)
ind_2_lag	0.2006***	-0.0282	-0.0968***	-0.0559***
	(5.0707)	(-0.9498)	(-4.6290)	(-3.0657)
Adj. R^2	0.621	0.644	0.572	0.389

This table shows the regression results of κ on squared historical returns. κ_w denotes weekly average κ, ind_2 denotes ret^2_{1w}, ret^2_{2w}, ret^2_{3w} and ret^2_{4w}, respectively. And the first-order lag terms are introduced to remove auto-correlation. t-statistics are reported in the parentheses. *, ** and *** represent significance at 10%, 5% and 1% level, respectively.

Table 8 checks the correlation coefficient matrix of κ and squared historical returns. κ_w denotes weekly average κ, and ret^2_{1w}, ret^2_{2w}, ret^2_{3w} and ret^2_{4w} denote squared historical returns of past 1, 2, 3 and 4 weeks, respectively. The correlation estimates suggest that κ positively impacts investor sentiment, driving up the option price.

Table 9 reports the regression results of κ on squared historical returns. ind_2 denotes ret^2_{1w}, ret^2_{2w}, ret^2_{3w} and ret^2_{4w}, respectively.

The results in Table 9 further support our interpretation that κ is positively related to investor sentiment. All parameter estimates for the sentiment proxy, ind_2, are positive and statistically significant at 1% level. Moreover, as we see in the sensitivity analysis (Figure 3), option price becomes higher as κ gets larger. With higher κ, we would know that investor sentiment is stronger

Table 10. **In-sample fit performance**

SSE	All options	Moneyness	Maturity	MM
BS	0.0137	0.0100	0.0114	0.0074
GBS	0.0066	0.0043	0.0028	0.0018

This table shows the average SSE of in-sample fit using four estimation methods. See appendix D for details of the estimation methods.

in the market, but we cannot distinguish whether investors are more bullish or more bearish in this situation. Why then does κ have a positive impact under both scenarios? This question is important to understand how κ drives option prices. We want to go back to equation (6) to answer this question. Recall that λ is the coefficient of Δ, and κ is the coefficient of Γ. If option price has a linear relationship with spot price, the change in option price due to spot price change would be proportional to Δ. It is well known that the relation between option and underlying prices is non-linear, as is reflected by Γ , the option convexity. When spot price increases, therefore, the option price increases, and yet Γ makes the increase larger; likewise, when spot price decreases, the option price also decreases. However, Γ makes the decrease not as large. This is why κ always gets larger when investors become either more bullish or bearish.

To improve the pricing performance and validate the robustness of our model, we employ three additional classification schemes based on two important option characteristics – moneyness and maturity. Following Bakshi *et al.* (1997), i) options are divided into OTM, ATM and ITM by moneyness, called moneyness-based method; or ii) options are divided to ST, MT and LT by maturities, called maturity-based method; or iii) options are divided into 9 categories (3*3 matrix) according to both moneyness and maturity, called MM-based method. Table 10 shows that 1) all three methods improve the in-sample pricing fits of both the GBS and the BS models; 2) GBS model consistently outperforms the BS model by a large margin regardless of the classification scheme; 3) MM-based fit greatly reduces the pricing error. The results in Table 10 are very striking because the percentage reduction in pricing error from using the GBS ranges between 52% and 76% with MM-based and maturity-based classifications saving more than three quarters of the pricing errors. Note that such enormous improvements on pricing accuracy in the GBS model derives from the additional parameters (and the corresponding pricing terms) related to investor sentiment.

4.3. *out-of-sample prediction performance*

We have shown that GBS model substantially outperforms BS model with regard to in-sample pricing fit. One may argue, however, that its superior in-sample fit could be just a result of over-fitting with more parameters. To substantiate that our GBS model has a comparative advantage, we compare the out-of-sample prediction performances of the two models in the following steps.

(i) Estimate the implied parameters each day with the four methods used for in-sample tests
(ii) In each day t, calculate the predicted option prices using the implied parameters estimated in day $t-1$ (σ_{t-1}, λ_{t-1}, κ_{t-1}) and the observed parameters in day t
(iii) Calculate the prediction error: Let the parameter set denoted by $\Phi_{it} = \{S_t,\ K_i,\ T_i,\ t,\ \sigma_{t-1},\ \lambda_{t-1},\ \kappa_{t-1}\}$, the theoretical option price by $\tilde{C}_{it}(\Phi_{it})$, and market price by C_{it}. Pricing errors are calculated as

$$\epsilon_{it}(\Phi_{it}) = \tilde{C}_{it}(\Phi_{it}) - C_{it}$$

(iv) Compare the daily average SSE as is defined in the in-sample performance tests following each of the four methods

Table 11. **Out-of-sample prediction performance**

SSE	All options	Moneyness	Maturity	MM
BS	0.0173	0.0148	0.0161	0.0131
GBS	0.0125	0.0114	0.0099	0.0100
T-stat	-8.5442***	-9.6429***	-7.4049***	-5.9574***

This table shows the average SSE of out-of-sample prediction using four estimation methods. See appendix D for details of the estimation methods.

Table 12. **MM based prediction error**

MM based		BS			GBS		
Overall average error		0.0020			-0.0004		
Category average error		ST	MT	LT	ST	MT	LT
OTM	< 0.91	-0.0017	-0.0036	-0.0030	-0.0002	0.0002	0.0013
	$0.91 - 0.94$	-0.0003	0.0026	0.0037	0.0009	0.0004	0.0013
	$0.94 - 0.97$	0.0015	0.0053	0.0069	0.0002	-0.0003	0.0004
ATM	$0.97 - 1$	-0.0012	-0.0017	-0.0024	-0.0003	-0.0006	-0.0011
	$1 - 1.03$	0.0020	0.0030	0.0045	0.0001	-0.0006	0.0006
ITM	$1.03 - 1.06$	0.0003	-0.0040	-0.0068	0.0004	-0.0012	-0.0013
	$1.06 - 1.09$	0.0040	0.0029	0.0014	-0.0002	-0.0014	-0.0006
	≥ 1.09	0.0062	0.0104	0.0094	-0.0006	-0.0020	-0.0029

This table shows the average pricing errors in each category of the BS model and GBS model through the MM based estimation method. See appendix D for details of MM based estimation method.

(v) Compare the average prediction error of all categories.

Following the estimation procedure specified above, we calculate the daily average SSE to directly compare the out-of-sample pricing performances of the GBS and the BS models. We find that the SSE of the GBS model is invariably smaller compared to the BS model, which is additional strong evidence of the dominant advantage of the GBS model. Although the reductions in the pricing error are not as large as in in-sample fit tests, 23% to 39% of the BS pricing errors are saved by employing the GBS model. For each method, Table 11 reports the daily average SSE of two models. The t-tests for mean differences of the SSE show that GBS model yield statistically significant smaller prediction errors.

Table 12 examines the average predicting error estimated by the four methods we use throughout the current paper. To save space, we only present the results of MM-based prediction with the other three tables provided in Appendix D.

We report the following findings from out-of-sample prediction tests.

(i) From all four tables here and in Appendix D, the GBS model shows consistently superior out-of-sample pricing performances than the BS model, demonstrating that the improved prediction power of the GBS model is not an overfitting result.;

(ii) The three alternative methods all outperform the original method, indicating we can estimate implied parameters more precisely by classifying the option data along the significant option characteristics such as moneyness and maturity. Furthermore, the MM-based prediction approach beats the other two methods in most cases.

(iii) All four tables show that the GBS model tends to prescribe lower estimated option prices compared to the BS model. This result indicates that, more or less, bearish sentiment prevails in Chinese market over the sample period. The average daily λ of all options from in-sample fit analysis is indeed positive, which is consistent with the fact that China A share market has experienced a bearish phase since June of 2015.

We further compare the relative pricing performances in each category using MM-based method. A negative value indicates a smaller absolute average pricing error from the GBS model. Figure 7 shows clearly that the GBS model outperforms the BS model in most cases with only two exceptions at trivial disadvantages.

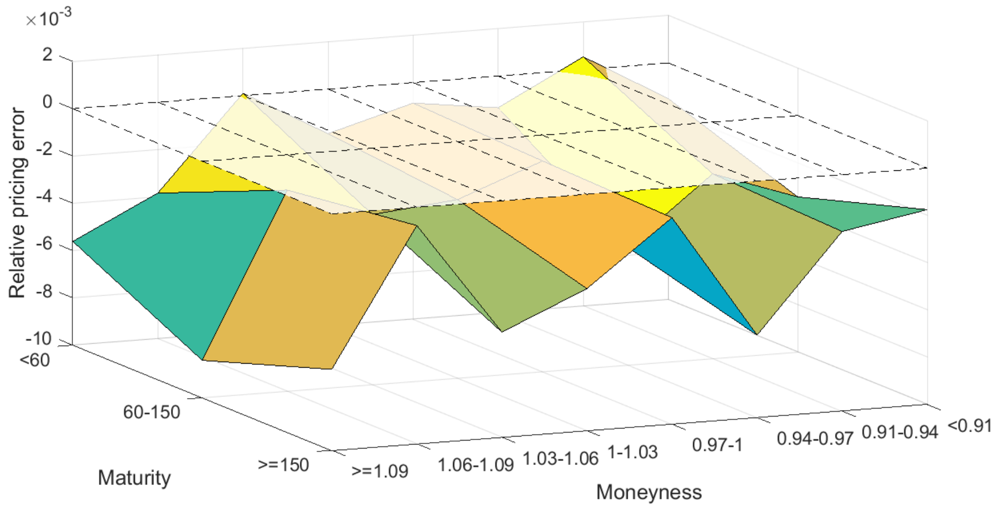

Figure 7. For each category, we calculate the average out-of-sample pricing errors of the two models ($|\bar{\epsilon}_{GBS}|-|\bar{\epsilon}_{BS}|$)to compare the relative predictive power.

5. Conclusion

We propose a generalized Black-Scholes (GBS) option pricing model by utilizing the generalized solution to the heat equation recently obtained. We provide the closed-form pricing formula for the model, wherein two newly incorporated parameters are interpreted as implied investor sentiment factors.

While prior research in recent years has documented that behavioral factors, particularly investor sentiment has significant pricing implications in option market, not just stock market, most papers have attempted to relate investor sentiment to option prices either relying on purely empirical approach or by studying the association between investor sentiment and certain aspects of the option or its underlying security (e.g., volatility smile and risk-neutral skewness). Our approach is distinguished from those studies in that our analytical option pricing formula *endogenously* incorporates additional terms which, through judicious parametrization, turn into factors related to investor sentiment.

Our empirical results show that our model invariably outperforms the standard BS model in terms of both in-sample fit and out-of sample pricing performances. Further analysis based on four subsamples that represent different option characteristics confirms that the reduction in pricing errors by the GBS is between 52% and 76% for in-sample fit performance and between 23% to 39% for out-of-sample prediction. Moreover, our results demonstrate that when we divide options based on moneyness and maturity, namely in the MM-based test, the GBS model can precisely capture the term structure of implied parameters, thereby exhibiting a significant pricing impact.

Our empirical analysis provides strong evidence that our interpretation of the two additional parameters in our GBS is consistent with the option pricing behavior salient in 50ETF option data. The first implied sentiment factor tends to capture investors belief about the future return prospect, while the second factor is related to their expectation of return volatility. Thus our results shed some light on a possible mechanism through which investor sentiment works *directly* into the option price.

Our study focuses on call options, leaving the investor sentiment reflected in put options to future research. Because the prices of call and put options are tightly linked by put-call parity, we expect that put options will give us a similar conclusion. Yet, it would still be interesting to investigate how different characteristics in sentiment might be reflected distinctively in calls and puts and the investors' heterogeneous responses. Also, this study examines the investor sentiment mainly from a time series perspective, looking into how temporal variations in investor sentiment affect the option pricing behavior. It will also be meaningful to extend our study to investigate differential impact of investor sentiment across different markets, different economic and regulatory conditions, and so on. Comparison with other option pricing models for relative pricing and hedging performances will add to the existing literature as well.

References

Aït-Sahalia, Y. and Lo, A. W., Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance,* 1998, **53(2)**, 499-547.

Aït-Sahalia, Y. and Lo, A. W., Nonparametric risk management and implied risk aversion. *Journal of Econometrics,* 2000, **94(1)**, 9-51.

Amin, K. I. and Jarrow. R. A., Pricing foreign currency options under stochastic interest rates. *Journal of International Money and Finance,* 1991, **10(3)**, 310-329.

Amin, K. I. and Jarrow. R. A., Pricing options on risky assets in a stochastic interest rate economy. *Mathematical Finance,* 1992, **2(4)**, 217-237.

Amin, K. I. and Ng. V. K., Option valuation with systematic stochastic volatility. *The Journal of Finance,* 1993, **48(3)**, 881-910.

Amin, K., Coval, J. D. and Seyhun, H. N., Index option prices and stock market momentum. *The Journal of Business,* 2004, **77(4)**, 835-874.

Bahra, B., Implied risk-neutral probability density functions from option prices: theory and application. *Working Paper,* 1997.

Baker, M. and Wurgler, J., Investor sentiment and the cross-section of stock returns. *The Journal of Finance,* 2006, **61(4)**, 1645-1680.

Bakshi, G. S. and Chen, Z., An alternative valuation model for contingent claims. *Journal of Financial Economics,* 1997a, **44(1)**, 123-165.

Bakshi, G. S. and Chen, Z., Equilibrium valuation of foreign exchange claims. *The Journal of Finance,* 1997b, **52(2)**, 799-826.

Bakshi, G., Cao, C. and Chen, Z., Empirical performance of alternative option pricing models. *The Journal of Finance,* 1997, **52(5)**, 2003-2049.

Bakshi, G., Kapadia, N. and Madan, D., Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies,* 2003, **16(1)**, 101-143.

Bali, T. G., Hu, J. and Murray, S., Option implied volatility, skewness, and kurtosis and the cross-section of expected stock returns. *Working Paper,* 2016.

Bali, T. G. and Murray, S., Does risk-neutral skewness predict the cross-section of equity option portfolio returns? *Journal of Financial and Quantitative Analysis,* 2013, **48(04)**, 1145-1171.

Barberis, N., Shleifer, A. and Vishny, R., A model of investor sentiment. *Journal of Financial Economics,* 1998, **49(3)**, 307-343.

Bates, D. S., Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies,* 1996, **9(1)**, 69-107.

Bates, D. S., The crash of'87: Was it expected? The evidence from options markets. *The Journal of Finance,* 1991, **46(3)**, 1009-1044.

Birru, J. and Figlewski, S., Anatomy of a Meltdown: The Risk Neutral Density for the S&P 500 in the Fall of 2008. *Journal of Financial Markets,* 2012, **15(2)**, 151-180.

Black, F. and Scholes, M., The pricing of options and corporate liabilities,*Journal of Political Economy,* 1973, **81(3)**, 637-654.

Blair, B. J. and Poon, S. H., Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics,* 2001, **105(1)**.

Bollen, N. P. B. and Whaley, R. E., Does net buying pressure affect the shape of implied volatility functions?,*The Journal of Finance,* 2004, **59(2)**, 711-753.

Bondarenko, O.,Why are put options so expensive?. *Working Paper,* 2003.

Borland, L. and Bouchaud, J. P., A non-Gaussian option pricing model with skew. *Quantitative Finance,* 2004, **4(5)**, 499-514.

Borland, L., A theory of non-Gaussian option pricing. *Quantitative Finance,* 2002a, **2(6)**, 415-431.

Borland, L., Option pricing formulas based on a non-Gaussian stock price model. *Physical Review Letters,* 2002b, **89(9)**, 098701.

Brennan, M. J., The pricing of contingent claims in discrete time models. *The Journal of Finance,* 1979, **34(1)**, 53-68.

Britten-Jones, M. and Neuberger, A. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance,* 2000, **55(2)**, 839-866.

Brown, G. W. and Cliff, M. T., Investor sentiment and the near-term stock market. *Journal of Empirical Finance,* 2004, **11(1)**, 1-27.

Busch, T., Christensen, B. J., Nielsen, M. Φ., The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics,* 2011, **160(1)**, 48-57.

Canina, L., Figlewski, S. The informational content of implied volatility. *Review of Financial Studies,* 1993, **6(3)**, 659-681.

Choi, B. S., Kang, H. and Choi, M. Y., Emergence of heavy-tailed skew distributions from the heat equation. *Physica A: Statistical Mechanics and Its Applications,* 2017, **470**, 88-93.

Christensen, B. J., Prabhala, N. R. The relation between implied and realized volatility. *Journal of Financial Economics,* 1998, **50(2)**, 125-150.

Coakley, J., Dotsis, G., Liu, X. and Zhai, J., Investor Sentiment and Option Prices: Evidence from Value and Growth Index Options. *Working Paper*, 2011.

Constantinides, G. M., Jackwerth, J. C. and Perrakis, S., Mispricing of S&P 500 index options. *Review of Financial Studies,* 2009, **22(3)**, 1247-1277.

Corrado, C. J. and Su, T., Skewness and kurtosis in S&P 500 index returns implied by option prices. *Journal of Financial Research,* 1996, **19(2)**, 175-192.

Darlington, R. B., Is kurtosis really peakedness?. *The American Statistician,* 1970, **24(2)**, 19-22.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J., Noise trader risk in financial markets. *Journal of Political Economy,* 1990, **98(4)**, 703-738.

Dittmar, R. F. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance,* 2002, **57(1)**, 369-403.

Dupire, B., Pricing and Hedging with Smiles. *Mathematics of derivative securities. Dempster and Pliska eds., Cambridge Uni. Press,* 1997.

Fleming, J. The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance,* 1998, **5(4)**, 317-345.

Garleanu, N., Pedersen, L. H. and Poteshman, A. M., Demand-based option pricing. *Review of Financial Studies,* 2009, **22(10)**, 4259-4299.

Guidolin, M. and Timmermann, A., International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies,* 2008, **21(2)**, 889-935.

Haas, M., Do investors dislike kurtosis? *Working Paper,* 2007.

Han, B., Investor sentiment and option prices. *Review of Financial Studies,* 2008, **21(1)**, 387-414.

Heath, D., Jarrow. R. and Morton, A., Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica: Journal of the Econometric Society,* 1992, 77-105.

Heston, S. L., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies,* 1993, **6(2)**, 327-343.

Hull, J. and White, A., The pricing of options on assets with stochastic volatilities. *The Journal of Finance,* 1987, **42(2)**, 281-300.

Jackwerth, J. C., Recovering risk aversion from option prices and realized returns. *Review of Financial Studies,* 2000, **13(2)**, 433-451.

Jarrow, R, and Rudd, A., Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics,* 1982, **10(3)** 347-369.

Jiang, G. J., Tian, Y. S. The model-free implied volatility and its information content. *Review of Financial Studies,* 2005, **18(4)**, 1305-1342.

Jorion, P. Predicting volatility in the foreign exchange market. *The Journal of Finance,* 1995, **50(2)**, 507-528.

Lemmon, M. L. and Ni, S. X., The effects of investor sentiment on speculative trading and prices of stock and index options. 2008.

Lo, A. W. and MacKinlay, A. C., Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies,* 1988, **1(1)**, 41-66.

Lucas, R. E., Interest rates and currency prices in a two-country world.*Journal of Monetary Economics,* 1982, **10(3)**, 335-359.

McCauley, J. L. and Gunaratne, G. H., An empirical model of volatility of returns and option pricing.*Physica A: Statistical Mechanics and its Applications,* 2003, **329(1)**, 178-198.

Melino, A. and Turnbull, S. M., Pricing foreign currency options with stochastic volatility.*Journal of Econometrics,* 1990, **45(1-2)**, 239-265.

Melino, A. and Turnbull, S. M., The pricing of foreign currency options.*Canadian Journal of Economics,* 1991, 251-281.

Merton, R. C., Theory of rational option pricing.*The Bell Journal of Economics and Management Science,* 1973, 141-183.

Miltersen, K. R. and Schwartz, E. S., Pricing of options on commodity futures with stochastic term structures of convenience yields and interest rates.*Journal of Financial and Quantitative Analysis,* 1998, **33(01)**, 33-59.

Poteshman, A. M., Underreaction, overreaction, and increasing misreaction to information in the options market.*The Journal of Finance,* 2001, **56(3)**, 851-876.

Rehman, Z., Vilkov, G. Risk-neutral skewness: Return predictability and its sources.

Rubinstein, M., The valuation of uncertain income streams and the pricing of options.*The Bell Journal of Economics,* 1976, 407-425.

Scott, L. O., Option pricing when the variance changes randomly: Theory, estimation, and an application.*Journal of Financial and Quantitative Analysis,* 1987, **22(04)**, 419-438.

Scott, L. O., Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates.*Mathematical Finance,* 1997, **7(4)**, 413-426.

Sharpe, S. A., Reexamining stock valuation and inflation: the implications of analysts' earnings forecasts.*Review of Economics and Statistics,* 2002, **84(4)**, 632-648.

Shleifer, A. and Vishny, R. W., The limits of arbitrage.*The Journal of Finance,* 1997, **52(1)**, 35-55.

Stein, E. M. and Stein, J. C., Stock price distributions with stochastic volatility: an analytic approach.*Review of Financial Studies,* 1991, **4(4)**, 727-752.

Stein, J., Overreactions in the options market.*The Journal of Finance,* 1989, **44(4)**, 1011-1023.

Szu, W. M., Wang, Y. C. and Yang, W. R., How Does Investor Sentiment Affect Implied Risk-Neutral Distributions of Call and Put Options?.*Review of Pacific Basin Financial Markets and Policies,* 2015, **18(02)**, 1550010.

Wang, X. T., Li, Z., Zhuang, L., European option pricing under the Student's t noise with jumps.*Physica A Statistical Mechanics & Its Applicationss,* 2017, **469**, 848-858.

Wiggins, J. B. Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics,* 1987, **19(2)**, 351-372.

Appendix

Appendix A-Solution of BS model

Assume the price of underlying asset is represented by S, and follows

$$\mathrm{d}S = \mu S\mathrm{d}t + \sigma S\mathrm{d}W$$

where μ is the drift term and σ is the volatility of underlying assets return.

C is the price of call option, according to Ito's Lemma

$$\mathrm{d}C = \left(\frac{\partial C}{\partial S}\mu S + \frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2\right)\mathrm{d}t + \frac{\partial C}{\partial S}\sigma S\mathrm{d}W$$

Now construct a portfolio Π in which we short one share of call option and long $\frac{\partial C}{\partial S}$ shares of underlying assets. Thus

$$\Pi = -C + \frac{\partial C}{\partial S}S$$

With a slight change of time $\mathrm{d}t$, the value of this portfolio is

$$\begin{aligned}\mathrm{d}\Pi &= -\mathrm{d}C + \frac{\partial C}{\partial S}\mathrm{d}S \\ &= \left(-\frac{\partial C}{\partial t} - \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2\right)\mathrm{d}t\end{aligned}$$

The return of this portfolio should be equal to risk-free rate, or there will exists opportunity to arbitrage. Thus we get the following partial differential equation

$$\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2 S^2\frac{\partial^2 C}{\partial S^2} + rS\frac{\partial C}{\partial S} - rC = 0 \tag{9}$$

with boundary condition

$$C\left(T,\ S\right) = \max\left(S - E,\ 0\right)$$

where E is the strike price. Let

$$\begin{aligned}S &= E\mathrm{e}^x \\ t &= T - \frac{2\tau}{\sigma^2} \\ C\left(t,\ S\right) &= EV\left(\tau,\ x\right)\end{aligned}$$

and change the variables in Eq. (9), we get

$$\frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial x^2} + \left(r - \frac{1}{2}\sigma^2\right)\frac{\partial V}{\partial x} - \frac{1}{2}\sigma^2\frac{\partial V}{\partial t} - rV = 0 \tag{10}$$

Let $k = \frac{2r}{\sigma^2}$, then Eq. (10) is transformed to

$$\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2} + (k-1)\frac{\partial V}{\partial x} - kV \tag{11}$$

The boundary condition is changed to

$$V(0,\ x) = \max\left(\mathrm{e}^x - 1,\ 0\right)$$

We will change the variable again and transform Eq. (11) to heat equation, whose solution is already known, then we can easily obtain the solution of call price. Let

$$V(\tau,\ x) = \mathrm{e}^{\alpha x + \beta \tau} u(\tau,\ x)$$

where α and β will be defined later. Eq. (11) will be transformed to

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + (2\alpha + k - 1)\frac{\partial u}{\partial x} - \left[(\alpha + k)(\alpha - 1) - \beta\right] u$$

In order to remove u and $\frac{\partial u}{\partial x}$, let the coefficients be zero. Thus

$$\begin{aligned} \alpha &= -\frac{1}{2}(k-1) \\ \beta &= -\frac{1}{4}(k+1)^2 \end{aligned}$$

Then the solution of Eq. (11) is

$$V(\tau,\ x) = \mathrm{e}^{-\frac{1}{2}(k-1)x - \frac{1}{4}(k+1)^2\tau} u(\tau,\ x)$$

where

$$\frac{\partial u}{\partial \tau} = \frac{\partial^2 u}{\partial x^2}, \qquad -\infty < x < \infty \text{ and } \tau > 0 \tag{12}$$

with boundary condition

$$\phi(x) = u(0,\ x) = \max\left[\mathrm{e}^{(1-\alpha)x} - \mathrm{e}^{-\alpha x},\ 0\right]$$

And the solution of heat equation is

$$u(\tau,\ x) = \frac{1}{\sqrt{4\pi\tau}} \int_{-\infty}^{\infty} \phi(y)\, \mathrm{e}^{-\frac{(x-y)^2}{4\tau}} \mathrm{d}y \tag{13}$$

Because the initial condition equals to

$$\begin{aligned} \phi(y) &= \max\left[\mathrm{e}^{\frac{1}{2}(k+1)y} - \mathrm{e}^{\frac{1}{2}(k-1)y},\ 0\right] \\ &= \mathrm{e}^{\frac{1}{2}(k+1)y} - \mathrm{e}^{\frac{1}{2}(k-1)y} \end{aligned}$$

when $y > 0$. So Eq. (13) is equivalent to

$$u(\tau,\ x) = \frac{1}{\sqrt{4\pi\tau}} \int_0^{\infty} \left[\mathrm{e}^{\frac{1}{2}(k+1)y} - \mathrm{e}^{\frac{1}{2}(k-1)y} \right] \mathrm{e}^{-\frac{(x-y)^2}{4\tau}} \mathrm{d}y$$

We solve the integral

$$u(\tau,\ x) = \mathrm{e}^{\frac{1}{2}(k+1)x + \frac{1}{4}(k+1)^2\tau} \mathrm{N}(d_1) - \mathrm{e}^{\frac{1}{2}(k-1)x + \frac{1}{4}(k-1)^2\tau} \mathrm{N}(d_2)$$

then

$$V(\tau,\ x) = \mathrm{e}^{x} \mathrm{N}(d_1) - \mathrm{e}^{-\tau k} \mathrm{N}(d_2)$$

The final solution of price of call option is

$$C_{BS}(t,\ S) = S\mathrm{N}(d_1) - E\mathrm{e}^{-r(T-t)} \mathrm{N}(d_2)$$

where

$$d_1 = \frac{\log(S/E) + \left(r + \sigma^2/2\right)(T-t)}{\sigma\sqrt{T-t}}$$
$$d_2 = \frac{\log(S/E) + \left(r - \sigma^2/2\right)(T-t)}{\sigma\sqrt{T-t}}$$

Appendix B-Solution of generalized BS model

$$\begin{aligned}
\varphi(\tau,\ x) &= \psi(\tau,\ x) + \lambda\theta(\tau,\ x) + \gamma\eta(\tau,\ x) \\
&= \frac{1}{\sqrt{4\pi\tau}} \int_{-\infty}^{\infty} \phi(y) \left[1 - \frac{\lambda(y-x)}{2\tau} - \frac{\kappa}{2\tau} + \frac{\kappa(y-x)^2}{4\tau^2} \right] \mathrm{e}^{-\frac{(y-x)^2}{4\tau}} \mathrm{d}y \\
&= \frac{1}{\sqrt{4\pi\tau}} \int_0^{\infty} \left[\mathrm{e}^{\frac{1}{2}(k+1)y} - \mathrm{e}^{\frac{1}{2}(k-1)y} \right] \left[1 - \frac{\lambda(y-x)}{2\tau} - \frac{\kappa}{2\tau} + \frac{\kappa(y-x)^2}{4\tau^2} \right] \mathrm{e}^{-\frac{(y-x)^2}{4\tau}} \mathrm{d}y
\end{aligned}$$

Let $\rho = \frac{y-x}{\sqrt{2\tau}}$, $y = \sqrt{2\tau}\rho + x$

$$u(\tau,\ x) = \frac{1}{\sqrt{2\pi}} \int_{\frac{-x}{\sqrt{2\tau}}}^{\infty} \left[\mathrm{e}^{\frac{1}{2}(k+1)\left(\sqrt{2\tau}\rho + x\right)} - \mathrm{e}^{\frac{1}{2}(k-1)\left(\sqrt{2\tau}\rho + x\right)} \right] \left[1 - \frac{\kappa}{2\tau} - \frac{\lambda\rho}{\sqrt{2\tau}} + \frac{\kappa\rho^2}{2\tau} \right] \mathrm{e}^{-\frac{1}{2}\rho^2} \mathrm{d}\rho$$

Seperate Eq.(8) to five parts

$$\begin{aligned}
I_0 &= \frac{1}{\sqrt{2\pi}} \left(1 - \frac{\kappa}{2\tau} \right) \int_{frac{-x}\sqrt{2\tau}}^{\infty} \left[\mathrm{e}^{\frac{1}{2}(k+1)\left(\sqrt{2\tau}\rho + x\right)} - \mathrm{e}^{\frac{1}{2}(k-1)\left(\sqrt{2\tau}\rho + x\right)} \right] \mathrm{e}^{-\frac{1}{2}\rho^2} \mathrm{d}\rho \\
&= \left(1 - \frac{\kappa}{2\tau} \right) \left[\mathrm{e}^{\frac{1}{2}(k+1)x + \frac{1}{4}(k+1)^2\tau} \mathrm{N}(d_1) - \mathrm{e}^{\frac{1}{2}(k-1)x + \frac{1}{4}(k-1)^2\tau} \mathrm{N}(d_2) \right]
\end{aligned}$$

where

$$d_1 = \frac{x}{\sqrt{2\tau}} + \frac{(k+1)\sqrt{2\tau}}{2}$$
$$d_2 = \frac{x}{\sqrt{2\tau}} + \frac{(k-1)\sqrt{2\tau}}{2}$$

Then the second one

$$\begin{aligned}
I_1 &= -\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\frac{1}{2}(k+1)x}\int_{frac-x\sqrt{2\tau}}^{\infty}\mathrm{e}^{\frac{1}{2}(k+1)\left(\sqrt{2\tau}\rho\right)}\frac{\lambda\rho}{\sqrt{2\tau}}\mathrm{e}^{-\frac{1}{2}\rho^2}\mathrm{d}\rho \\
&= -\frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{\frac{1}{2}(k+1)x}\int_{\frac{-x}{\sqrt{2\tau}}}^{\infty}\rho\mathrm{e}^{-\frac{1}{2}\rho^2+\frac{1}{2}(k+1)\left(\sqrt{2\tau}\rho\right)}\mathrm{d}\rho \\
&= -\frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}\int_{frac-x\sqrt{2\tau}}^{\infty}\rho\mathrm{e}^{-\frac{1}{2}\left(\rho-\frac{1}{2}(k+1)\sqrt{2\tau}\right)^2}\mathrm{d}\rho
\end{aligned}$$

Let $\xi = \rho - \frac{1}{2}(k+1)\sqrt{2\tau}$, $\rho = \xi + \frac{1}{2}(k+1)\sqrt{2\tau}$

$$I_1 = -\frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}\left[\int_{-d_1}^{\infty}\xi\mathrm{e}^{-\frac{1}{2}\xi^2}\mathrm{d}\xi + (k+1)\sqrt{2\tau}\int_{-d_1}^{\infty}\mathrm{e}^{-\frac{1}{2}\xi^2}\mathrm{d}\xi\right]$$

Due to

$$\int \xi\mathrm{e}^{c\xi^2}\mathrm{d}\xi = \frac{1}{2c}\mathrm{e}^{c\xi^2} + Const$$
$$I_1 = -\frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}\left[\mathrm{e}^{-\frac{1}{2}d_1^2} + (k+1)\sqrt{\pi\tau}\mathrm{N}(d_1)\right]$$

Similarly

$$I_2 = \frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{\frac{1}{2}(k-1)x+\frac{1}{4}(k-1)^2\tau}\left[\mathrm{e}^{-\frac{1}{2}d_2^2} + (k-1)\sqrt{\pi\tau}\mathrm{N}(d_2)\right]$$

Then come to the forth part

$$\begin{aligned}
I_3 &= \frac{\kappa}{\sqrt{2\pi}2\tau}\mathrm{e}^{\frac{1}{2}(k+1)x}\int_{\frac{-x}{\sqrt{2\tau}}}^{\infty}\rho^2\mathrm{e}^{-\frac{1}{2}\rho^2+\frac{1}{2}(k+1)\sqrt{2\tau}\rho}\mathrm{d}\rho \\
&= \frac{\kappa}{\sqrt{2\pi}2\tau}\mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau}\int_{frac-x\sqrt{2\tau}}^{\infty}\rho^2\mathrm{e}^{-\frac{1}{2}\left(\rho-\frac{1}{2}(k+1)\sqrt{2\tau}\right)^2}\mathrm{d}\rho
\end{aligned}$$

Let $\xi = \rho - \frac{1}{2}(k+1)\sqrt{2\tau}$, $\rho = \xi + \frac{1}{2}(k+1)\sqrt{2\tau}$

$$
\begin{aligned}
I_3 &= \frac{\kappa}{\sqrt{2\pi}2\tau} \mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau} \int_{-d_1}^{\infty} \left(\xi + \frac{1}{2}(k+1)\sqrt{2\tau}\right)^2 \mathrm{e}^{-\frac{1}{2}\xi^2} \mathrm{d}\xi \\
&= \frac{\kappa}{\sqrt{2\pi}2\tau} \mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau} \\
&\left[\int_{-d_1}^{\infty} \xi^2 \mathrm{e}^{-\frac{1}{2}\xi^2} \mathrm{d}\xi + (k+1)\sqrt{2\tau} \int_{-d_1}^{\infty} \xi \mathrm{e}^{-\frac{1}{2}\xi^2} \mathrm{d}\xi + \frac{1}{2}(k+1)^2 \tau \int_{-d_1}^{\infty} \mathrm{e}^{-\frac{1}{2}\xi^2} \mathrm{d}\xi\right]
\end{aligned}
$$

Due to

$$
\begin{gathered}
\mathrm{d}\left(\xi \mathrm{e}^{\mathrm{c}\xi^2}\right) = \mathrm{e}^{\mathrm{c}\xi^2} \mathrm{d}\xi + 2c\xi^2 e^{c\xi^2} \mathrm{d}\xi \\
\int \xi^2 e^{c\xi^2} \mathrm{d}\xi = -\frac{1}{2c} \int \mathrm{e}^{\mathrm{c}\xi^2} \mathrm{d}\xi + \frac{1}{2c} \xi \mathrm{e}^{\mathrm{c}\xi^2} + Const
\end{gathered}
$$

Then

$$
\begin{aligned}
I_3 &= \frac{\kappa}{\sqrt{2\pi}2\tau} \mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau} \left[\sqrt{2\pi}\mathrm{N}(d_1) - d_1 \mathrm{e}^{-\frac{1}{2}d_1^2} + (k+1)\sqrt{2\tau}\mathrm{e}^{-\frac{1}{2}d_1^2} + \frac{\sqrt{2\pi}}{2}(k+1)^2 \tau \mathrm{N}(d_1)\right] \\
&= \frac{\kappa}{\sqrt{2\pi}2\tau} \mathrm{e}^{\frac{1}{2}(k+1)x+\frac{1}{4}(k+1)^2\tau} \left\{\left[1 + \frac{1}{2}(k+1)^2 \tau\right] \sqrt{2\pi}\mathrm{N}(d_1) + \left[(k+1)\sqrt{2\tau} - d_1\right] \mathrm{e}^{-\frac{1}{2}d_1^2}\right\}
\end{aligned}
$$

Similiarly

$$
I_4 = -\frac{\kappa}{\sqrt{2\pi}2\tau} \mathrm{e}^{\frac{1}{2}(k-1)x+\frac{1}{4}(k-1)^2\tau} \left\{\left[1 + \frac{1}{2}(k-1)^2 \tau\right] \sqrt{2\pi}\mathrm{N}(d_2) + \left[(k-1)\sqrt{2\tau} - d_2\right] \mathrm{e}^{-\frac{1}{2}d_2^2}\right\}
$$

$u(\tau,\ x)$ is the summation of five parts up above

$$
\varphi(\tau,\ x) = I_0 + I_1 + I_2 + I_3 + I_4
$$

Appendix C-Relation between GBS and BS model

First we solve the expression of

$$
\begin{aligned}
V(\tau,\ x) &= \mathrm{e}^{-\frac{1}{2}(k-1)x-\frac{1}{4}(k+1)^2\tau}u(\tau,\ x) \\
&= \left(1-\frac{\kappa}{2\tau}\right)\left[\mathrm{e}^{x}\mathrm{N}(d_1)-\mathrm{e}^{-\tau k}\mathrm{N}(d_2)\right] \\
&\quad -\frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{x}\left[\mathrm{e}^{-\frac{1}{2}d_1^2}+(k+1)\sqrt{\pi\tau}\mathrm{N}(d_1)\right] \\
&\quad +\frac{\lambda}{\sqrt{4\pi\tau}}\mathrm{e}^{-\tau k}\left[\mathrm{e}^{-\frac{1}{2}d_2^2}+(k-1)\sqrt{\pi\tau}\mathrm{N}(d_2)\right] \\
&\quad +\frac{\kappa}{\sqrt{2\pi}2\tau}\mathrm{e}^{x}\left\{\left[1+\frac{1}{2}(k+1)^2\tau\right]\sqrt{2\pi}\mathrm{N}(d_1)+\left[\frac{(k+1)\sqrt{2\tau}}{2}-\frac{x}{\sqrt{2\tau}}\right]\mathrm{e}^{-\frac{1}{2}d_1^2}\right\} \\
&\quad -\frac{\kappa}{\sqrt{2\pi}2\tau}\mathrm{e}^{-\tau k}\left\{\left[1+\frac{1}{2}(k-1)^2\tau\right]\sqrt{2\pi}\mathrm{N}(d_2)+\left[\frac{(k-1)\sqrt{2\tau}}{2}-\frac{x}{\sqrt{2\tau}}\right]\mathrm{e}^{-\frac{1}{2}d_2^2}\right\} \\
&= \left[1-\frac{\lambda}{2}(k+1)+\frac{\kappa}{4}(k+1)^2\right]\mathrm{e}^{x}\mathrm{N}(d_1)-\left[1-\frac{\lambda}{2}(k-1)+\frac{\kappa}{4}(k-1)^2\right]\mathrm{e}^{-\tau k}\mathrm{N}(d_2) \\
&\quad -\frac{1}{\sqrt{4\pi\tau}}\left[\lambda-\frac{\kappa}{2}(k+1)+\frac{\kappa}{2\tau}x\right]\mathrm{e}^{x-\frac{1}{2}d_1^2}+\frac{1}{\sqrt{4\pi\tau}}\left[\lambda-\frac{\kappa}{2}(k-1)+\frac{\kappa}{2\tau}x\right]\mathrm{e}^{-\tau k-\frac{1}{2}d_2^2}
\end{aligned}
$$

It can be proven

$$
\begin{aligned}
-\tau k-\frac{1}{2}d_2^2 &= -r((T-t))-\frac{\left[\ln\frac{S}{E}+\left(r-\frac{\sigma^2}{2}\right)(T-t)\right]^2}{2\sigma^2((T-t))} \\
&= -r((T-t))-\frac{\left(\ln\frac{S}{E}\right)^2+\ln\frac{S}{E}\left(r-\frac{\sigma^2}{2}\right)(T-t)+\left(r-\frac{\sigma^2}{2}\right)^2(T-t)^2}{2\sigma^2(T-t)} \\
&= -\frac{\left(\ln\frac{S}{E}\right)^2+\ln\frac{S}{E}\left(2r-\sigma^2\right)(T-t)+\left(r-\frac{\sigma^2}{2}\right)^2(T-t)^2}{2\sigma^2(T-t)} \\
&= \ln\frac{S}{E}-\frac{\left[\ln\frac{S}{E}+\left(r+\frac{\sigma^2}{2}\right)(T-t)\right]^2}{2\sigma^2(T-t)} \\
&= x-\frac{1}{2}d_1^2
\end{aligned}
$$

Appendix D

Followings are the descriptions to the four methods we use for parameter estimation.

(i) All options fit: In each day we estimate the parameters using all the option prices in that day. Thus, we get one set of parameters.

(ii) Moneyness based fit: In each day we divide the options into three groups according to moneyness, i.e., OTM, ATM and ITM. Then we estimate the parameters using the option prices in each group, and get three set of parameters.

(iii) Maturity based fit: In each day we divide the options into three groups according to time to maturity, i.e., ST, MT and LT. Then we estimate the parameters using the options prices in each group, and get three set of parameters.

(iv) MM based fit: In each day we divide the options into nine groups according to time to maturity and moneyness, i.e., OTM-ST, OTM-MT, etc. Then we estimate the parameters using the options prices in each group, and get nine set of parameters.

Table A1 and A2 show the correlation coefficients between estimated λ, κ and subsequent returns. And table A3, A4 and A5 show the results for average pricing error of two models.

Table A1. **Correlation coefficients 3**

	λ_w	ret_{1w}	ret_{2w}	ret_{3w}	ret_{4w}
λ_w	1	-0.0960	-0.0852	-0.1555	-0.1585
ret_{1w}		1	0.7864	0.6207	0.5594
ret_{2w}			1	0.8666	0.7511
ret_{3w}				1	0.9113
ret_{4w}					1

This table shows the correlation matrix among estimated λ and subsequent returns. λ_w denotes weekly average λ, ret_{1w}, ret_{2w}, ret_{3w} and ret_{4w} denote subsequent return of 1, 2, 3 and 4 weeks respectively.

Table A2. **Correlation coefficients 4**

	κ_w	ret^2_{1w}	ret^2_{2w}	ret^2_{3w}	ret^2_{4w}
κ_w	1	0.0472	0.0312	0.0241	0.0010
ret^2_{1w}		1	0.5042	0.3748	0.2925
ret^2_{2w}			1	0.7653	0.5785
ret^2_{3w}				1	0.8051
ret^2_{4w}					1

This table shows the correlation matrix among estimated κ and squared subsequent returns. κ_w denotes weekly average κ, ret^2_{1w}, ret^2_{2w}, ret^2_{3w} and ret^2_{4w} denote squared subsequent return of 1, 2, 3 and 4 weeks respectively.

Table A3. **All options prediction error**

All options based		BS			GBS		
Overall average error		0.0013			-0.0006		
Category average error		ST	MT	LT	ST	MT	LT
OTM	< 0.91	-0.0038	-0.0112	-0.0090	-0.0015	-0.0070	-0.0036
	0.91 − 0.94	-0.0043	-0.0088	-0.0058	0.0000	-0.0054	-0.0028
	0.94 − 0.97	-0.0039	-0.0067	-0.0046	0.0011	-0.0039	-0.0024
ATM	0.97 − 1	-0.0023	-0.0036	-0.0017	0.0023	-0.0025	-0.0007
	1 − 1.03	0.0009	0.0010	0.0052	0.0011	-0.0003	0.0041
ITM	1.03 − 1.06	0.0035	0.0069	0.0100	-0.0024	0.0026	0.0068
	1.06 − 1.09	0.0056	0.0109	0.0157	-0.0045	0.0038	0.0104
	≥ 1.09	0.0060	0.0135	0.0199	-0.0054	0.0027	0.0109

This table shows the average pricing errors in each category of the BS model and GBS model through all option based estimation method.

Table A4. **Moneyness based prediction error**

Moneyness based		BS			GBS		
Overall average error		0.0014			-0.0001		
Category average error		ST	MT	LT	ST	MT	LT
OTM	< 0.91	-0.0031	-0.0060	0.0008	-0.0002	-0.0033	0.0023
	0.91 − 0.94	-0.0020	-0.0004	0.0061	0.0012	-0.0018	0.0045
	0.94 − 0.97	-0.0005	0.0021	0.0081	0.0013	-0.0011	0.0048
ATM	0.97 − 1	-0.0021	-0.0031	-0.0008	0.0000	-0.0027	-0.0005
	1 − 1.03	0.0011	0.0013	0.0057	0.0004	0.0001	0.0047
ITM	1.03 − 1.06	-0.0004	-0.0030	-0.0056	0.0015	-0.0004	-0.0038
	1.06 − 1.09	0.0033	0.0035	0.0025	-0.0010	0.0008	0.0001
	≥ 1.09	0.0053	0.0100	0.0105	-0.0030	0.0016	0.0028

This table shows the average pricing errors in each category of the BS model and GBS model through the moneyness based estimation method.

Table A5. **Maturity based prediction error**

Maturity based		BS			GBS		
Overall average error			0.0018			-0.0003	
Category average error		ST	MT	LT	ST	MT	LT
OTM	< 0.91	-0.0029	-0.0097	-0.0108	-0.0015	-0.0018	0.0003
	$0.91 - 0.94$	-0.0028	-0.0074	-0.0069	-0.0009	-0.0008	0.0015
	$0.94 - 0.97$	-0.0028	-0.0054	-0.0062	-0.0008	0.0001	0.0010
ATM	$0.97 - 1$	-0.0014	-0.0025	-0.0035	0.0002	0.0008	0.0003
	$1 - 1.03$	0.0018	0.0023	0.0033	0.0010	0.0015	0.0016
ITM	$1.03 - 1.06$	0.0045	0.0079	0.0083	0.0005	0.0020	0.0002
	$1.06 - 1.09$	0.0067	0.0119	0.0138	0.0003	0.0013	0.0002
	≥ 1.09	0.0069	0.0144	0.0173	-0.0006	-0.0015	-0.0027

This table shows the average pricing errors in each category of the BS model and GBS model through the maturity based estimation method.