The Special Elliptic Option Pricing Model and Volatility Smile

Stephen H-T. Lihn*

e-mail: stevelihn@gmail.com

Abstract: A special version of option pricing model based on elliptic distribution is developed to explain the volatility smile for short-maturity options. A skew fractional exponential distribution, called λ distribution, is formulated to facilitate the so-called λ transformation for option pricing. The distribution has a single shape parameter that covers the normal distribution, Laplace distribution, and the elliptic cusp distribution. The option prices are hypothesized to have two regimes: The local regime is the high-kurtosis regime where the prices can be calculated, but not observed directly. Mathematical procedures are developed to handle moment explosion in this regime. The global regime is based on the normal distribution and Black-Scholes model where the prices can be observed by the market. The λ transformation is a 4-step transformation projecting between the local regime and the global regime. The implied volatility curve is hypothesized to be movable by the expectation of market momentum during the regime transformation. This model captures the major features of the volatility smile for short-maturity SPX options from a few days to a few weeks.

MSC 2010 subject classifications: Primary 60K35, 14H52, 62P05; secondary 91B28.
Keywords and phrases: distribution, option pricing, volatility smile, high frequency, fat tail, leptokurtic.

Contents

*Mr. Lihn is currently managing director of equity quantitative strategy and marekt data at Novus Partners, Inc.. LinkedIn: https://www.linkedin.com/pub/stephen-horng-twu-lihn/0/71a/65

1. Introduction

This paper is a continuation of the work on the elliptic distribution[15]. The main goal of this paper is to provide an analytic framework to explain volatility smile. The importance of understanding volatility smile in financial economics is well written on Wikipedia[1], and quoted here:

> In fact, Benoît Mandelbrot had discovered already in the 1960s that changes in financial prices do not follow a Gaussian distribution, the basis for much option pricing theory, although this observation was slow to find its way into mainstream financial economics. Financial models with long-tailed distributions and volatility clustering have been introduced to overcome problems with the realism of classical financial models; jump diffusion models allow for (option) pricing incorporating "jumps" in the spot price. Risk managers, similarly, complement (or substitute) the standard value at risk models with historical simulations, volatility clustering, mixture models, principal component analysis and extreme value theory...
>
> Closely related is the volatility smile, where implied volatility is observed to differ as a function of strike price (i.e. moneyness). The term structure of volatility describes how (implied) volatility differs for related options with different maturities; an implied volatility surface is a three-dimensional surface plot of volatility smile and term structure. These empirical phenomena negate the assumption of constant volatility — and log-normality— upon which Black-Scholes is built. Approaches developed here in response include local volatility and stochastic volatility (the Heston, SABR and CEV models, amongst others). Alternatively, implied-binomial and -trinomial trees instead of directly modeling volatility, return a lattice consistent with — in an arbitrage-free sense — (all) observed prices, facilitating the pricing of other, i.e. non-quoted, strike/maturity combinations. Edgeworth binomial trees allow for a specified (i.e. non-Gaussian) skew and kurtosis in the spot price. Priced here, options with differing strikes will return differing implied volatilities, and the tree can thus be calibrated to the smile if required. Similarly purposed closed-form models include: Jarrow and Rudd (1982); Corrado and Su (1996); Backus, Foresi, and Wu (2004).
>
> Particularly following the financial crisis of 2007–2010, financial economics and mathematical finance have been subjected to criticism; notable here is Nassim Nicholas Taleb, who claims that the prices of financial assets cannot be characterized by the simple models currently in use, rendering much of current practice at best irrelevant, and, at worst, dangerously misleading...

In this paper, the foundation is based on the study of elliptic distribution, which captures the highly leptokurtic nature of the log-return distributions in major financial assets[2]. The main observation there is that, due to the high kurtosis, the fits have to be located in the small (α, γ) region (in the polar coordinate language, small R), where they have close proximity to the standard cusp distribution.

1.1. Introducing The Option Generating Function

If the log-return distribution of the underlying asset is believed to be known, the most wanted question next is - what does the option pricing model look like

[1] https://en.wikipedia.org/wiki/Financial_economics#Challenges_and_criticism

[2] My reference of high kurtosis is 12.257, that of the standard cusp distribution. See Table 2.1.

with that distribution? And one of the most pronounced issues in option pricing is the origin of the volatility smile. These days, even the popular financial data portal, Yahoo Finance, is publishing volatility smile every day[3]. This paper is an attempt to answer this question.

Assume $\widetilde{\mathcal{S}}_t$ is the price of the underlying asset at time t. In the logarithmic framework, where X is the log-return process of the underlying asset between time 0 and T, we have $\widetilde{\mathcal{S}}_T = \widetilde{\mathcal{S}}_0 e^X$. The forward prices of European option can be expressed as the expected profit in excess of the strike price, discounted by the risk-free rate,

$$\begin{aligned} \widetilde{\mathcal{C}} &= E\left[\max\left(\widetilde{\mathcal{S}}_0 e^X - \widetilde{\mathcal{K}} e^{-rT}, 0\right)\right], \quad \text{for call option;} \\ \widetilde{\mathcal{P}} &= E\left[\max\left(\widetilde{\mathcal{K}} e^{-rT} - \widetilde{\mathcal{S}}_0 e^X, 0\right)\right], \quad \text{for put option;} \end{aligned} \tag{1.1}$$

where $\widetilde{\mathcal{C}}, \widetilde{\mathcal{P}}, \widetilde{\mathcal{S}}_0, \widetilde{\mathcal{K}}, r, T$ are standard notations for call price, put price, underlying price at time 0, strike price, risk-free rate, and time to maturity. They are often abbreviated as $\widetilde{\mathcal{C}} = \left(\widetilde{\mathcal{S}} - \widetilde{\mathcal{K}}\right)_+$ and $\widetilde{\mathcal{P}} = \left(\widetilde{\mathcal{K}} - \widetilde{\mathcal{S}}\right)_+$ in textbooks (E.g. Eq. (6.1) of [4]). From Eq. (1.1), assume X has the probability distribution $P(x)$, the $\widetilde{\mathcal{S}}_0$-normalized forward prices of European option, $\widetilde{\mathcal{C}}/\widetilde{\mathcal{S}}_0$ and $\widetilde{\mathcal{P}}/\widetilde{\mathcal{S}}_0$, can be abstracted as functions of log-strike k,

$$\begin{aligned} \mathbb{L}_c(k) &= \int_k^\infty \left(e^x - e^k\right) P(x)\, dx, \quad \text{for call option;} \\ \mathbb{L}_p(k) &= -\int_{-\infty}^k \left(e^x - e^k\right) P(x)\, dx, \quad \text{for put option;} \end{aligned} \tag{1.2}$$

where $k = \log(\widetilde{\mathcal{K}}/\widetilde{\mathcal{S}}_0) - rT$. The transformational functions, $\mathbb{L}_c(k)$ and $\mathbb{L}_p(k)$, play the central role in option pricing model. For instance, they are the starting point of Dupire's local volatility model[10, 11]. These two functions are called "option generating function" (OGF) in this paper. They are viewed as a Laplace-style transform with a kernel of $\pm\left(e^x - e^k\right)$ in incomplete integrals. They can also be viewed as "operators" (like in quantum mechanics) that project the distribution from the log-return space to the log-strike space.

The OGF, $\mathbb{L}_{c,p}(k)$, can be decomposed to the subtraction of two integrals - the incomplete moment generating function (IMGF, $\int_k e^x P(x)\, dx$) and cumulative density function (CDF, $\int_k P(x)\, dx$). However, naive attempt to solve such integrals soon encounters infinity not too far along the way. The bulk of this paper is spent on how to properly solve CDF, IMGF, OGF; and insights obtained from solving them.

1.2. Introducing The λ Distribution

Although the origin of the elliptic distribution is very simple - just the elliptic curves in one of their simplest forms, $x^2 = -y^3 - \gamma y - \beta x y + \alpha$, the resulting probability distribution rarely has analytic solution. Most of the calculations

[3] E.g. see http://finance.yahoo.com/q/op?s=SPY+Options

are carried out numerically. In order to obtain analytic solutions for OGF, it is necessary to simplify the elliptic distribution. Therefore, a special version of elliptic distribution called the λ distribution is formulated. This distribution comes to existence with several purposes:

1. My numerical experiments with SPX option data have indicated that $P(x)$ is very close to the standard cusp distribution (in the polar coordinate language, R is very small). The specific shape in the valley of the volatility smile may vary somewhat, but the general shape of the smile doesn't change much. This implies the standard cusp distribution plays an important role.
2. The output of $\mathbb{L}_{c,p}(k)$ must be translated to implied volatility via the Black-Scholes equation, which is based on the normal distribution. However, in the context of elliptic distribution, it is not convenient to express normal distribution cleanly since it is the asymptotic distribution when $R \to \infty$ (and $\theta \neq \frac{7}{4}\pi$). It is also unlikely to perform precise calculation using the distribution in the infinity (I've discussed how slow kurtosis converges to 3 in [15]). Therefore, it calls for a novel way to encompass the normal distribution into the framework more elegantly, which motivates the creation of λ distribution.
3. Coincidentally, the Laplace distribution can be incorporated into the framework of λ distribution nicely. It is the asymptotic distribution of the general cusp distribution ($R \to \infty$ at $\theta = \frac{7}{4}\pi$). It is also the tail distribution of hyperbolic distribution. It turns out to be the boundary of the local regime too.

The λ distribution has the tail of $\log P(x) \sim |x|^{-\frac{2}{\lambda}}$ where $\lambda \in [1, 3]$. The risk-neutral drift, $\mu_D = E\left[e^X\right] = \int_{-\infty}^{\infty} e^x P(x)\, dx$, diverges when $\lambda \geq 2$. The boundary $\lambda = 2$ separates the local regime ($\lambda \geq 2$) from the global regime ($\lambda = 1$). In the global regime, it is well known the risk neutrality yields the put-call parity of option prices. However, in the local regime, a major hypothesis of this paper is that the risk neutrality can be violated, which produces volatility skew[4].

1.3. Introducing The λ Transformation

The second point above needs to be elaborated more since it is one of the key features of the so-called λ transformation. The option pricing model doesn't stop at calculating $\mathbb{L}_{c,p}(k)$ from a peculiar distribution $P(x)$, although this itself is already a lengthy process. There are more complications. The output of $\mathbb{L}_{c,p}(k)$ (which is a form of price) must be transformed to implied volatility, $\sigma_{imp}(k)$, which is then transported by a certain amount of log-strike, then transformed back to option prices, at which point the result matches what's observed in the market data.

[4]The volatility skew is generally observed as the stylized fact that downstrike volatilities are higher than upstrike volatilities. However our result indicates this is only true for call options. The put options can have upstrike volatilities higher than downstrike volatilities. This is explained in Section 4.8.

To put them together, the option prices, $\widetilde{\mathcal{C}}(k)$ and $\widetilde{\mathcal{P}}(k)$, observed in the market, is a composite transformation of

$$\begin{array}{ll} \frac{\widetilde{\mathcal{C}}(k)}{\widetilde{S}_0} = \mathbb{O}_c \cdot \mathbb{M}\left(r_M\right) \cdot \mathbb{V}_c \cdot \mathbb{L}_c \left|P(x)\right\rangle, & \text{for call option;} \\ \frac{\widetilde{\mathcal{P}}(k)}{\widetilde{S}_0} = \mathbb{O}_p \cdot \mathbb{M}\left(r_M\right) \cdot \mathbb{V}_p \cdot \mathbb{L}_p \left|P(x)\right\rangle, & \text{for put option.} \end{array} \tag{1.3}$$

Here we use the quantum-mechanics like operator notations, and the four operators of λ transformation are outlined below:

- $\mathbb{L}_{c,p}$: Input is a probability distribution $P(x)$; output is the normalized option prices in the local regime (in log-strike).
- $\mathbb{V}_{c,p}$: Input is normalized option prices; output is implied volatility curve.
- $\mathbb{M}\left(r_M\right)$: Input is implied volatility curve; output is also implied volatility curve with log-strike transported by market momentum r_M.
- $\mathbb{O}_{c,p}$: Input is implied volatility curve; output is the normalized option prices, observable in the global regime.

The concept of the local regime and the global regime is developed to provide philosophical support for such transformation. The transportation of implied volatility along log-strike, $\mathbb{M}\left(r_M\right)$, is a new, intriguing mechanism in this paper. The shift r_M is interpreted as a form of "market momentum"[3]. It is thought of as an expression of market expectation by bullish or bearish market participants. Such expectation shifts the volatility smile horizontally and produces the ATM skew, often denoted as $\sigma_{ATM}(T)$. The amount of shift is different for different time to maturity as well as different assets. The view on the sources of ATM skew in this paper is quite different from that of e.g. [20].

The formulae developed in this paper have been validated by `R` with hundreds of test cases. Calculations and charts are generated by the `ecd` package, which is uploaded to CRAN. Feedback and collaboration are welcomed.

2. The Development of λ Distribution

2.1. The Motivation

The motivation of the λ distribution is to construct a simple distribution that covers the normal distribution, Laplace distribution, and the elliptic cusp distribution. These are the three asymptotic cases in the elliptic distribution family[15]-

1. The normal distribution: This is the asymptotic distribution when $R \to \infty$ and $\theta \neq \frac{7}{4}\pi$. In the elliptic distribution language, it is $y \sim -x^2$. This distribution is also called "the global regime" in the λ transformation. Obviously, in option pricing model, it is connected to the Black-Scholes model and the industry-standard implied volatility calculation.
2. The Laplace distribution: This is the asymptotic distribution when $R \to \infty$ and $\theta = \frac{7}{4}\pi$. In the elliptic distribution language, it is $(-y)^2 \sim x^2$. This

distribution is the boundary of the local regime. Due to its simple mathematical form $P(x) \sim e^{-x}$, many analytic solutions can be developed. It provides good anchor for the rest of local regime, where analytic solutions are much harder to get.

3. The elliptic cusp distribution: This is $\texttt{ECD}(0,0)$ at the center of the (α,γ) plane, or simply $R=0$. In the elliptic distribution language, it is $y^3 \sim -x^2$. In addition, its asymmetric distribution, $x^2 = -y^3 - \beta xy$, also has analytic solution. This is the area that I am most interested in, since this is where the solution of SPX options resides.

If the new distribution can be constructed in the form of $(-y)^\lambda \sim x^2 + ...$ where $\lambda = 1, 2, 3$, then the integer λ can represent the three cases outlined above. Furthermore, when λ is extended to real number, it provides analytic continuity between the three distributions. In a sense, these three distributions become one. The integer λ is also more favorable in a numeric package. It can index the distributions and solutions precisely. It avoids concern of floating-point rounding issue.

The simple form of λ distribution attempts to capture the influence of the tail exponent, but foregoes the delicate structure of the elliptic distribution near the peak. The rationale here is that volatility smile is dominated by the behavior of the tails. This can be illustrated by the order-of-magnitude estimate. For SPX index, the daily log-returns have standard deviation of $\sim 1\%$[15]. Due to the low market volatility in Q2 of 2015 (when the sample data is taken for this paper), the standard deviation of the λ distribution that fits SPX options maturing in a day in June (see Figure 6.1) is only about 0.6%. So for log-strike $k = 0.05$, this is 8 standard deviations away from the peak of the distribution.

In real-world applications where fine structures are important, a general model based on the elliptic distribution should be developed. That is a separate project beyond the scope of this paper.

2.2. Basic Notations

Now that the motivation is outlined, we can start developing the distribution. The curves of the so-called "unit distribution" takes the form of

$$\hat{x}^2 = (-\hat{y})^\lambda - \beta\hat{x}\hat{y}, \text{ whose symmetric case is } \hat{y}(x) = -|\hat{x}|^{\frac{2}{\lambda}}. \tag{2.1}$$

The $\hat{}$ symbol (hat) is used to label this particularly simple parametrization across the paper. By setting $\hat{x} = \frac{x-\mu}{\sigma}$, we make the distribution qualified as a location-scale distribution family, and the curves become

$$\frac{(x-\mu)^2}{\sigma^2} = (-y)^\lambda - \beta\left(\frac{x-\mu}{\sigma}\right)y. \tag{2.2}$$

Thus, the unit distribution is a special case when $\sigma = 1, \mu = 0$. The curve $y(x) = y(x;\lambda,\sigma,\beta,\mu)$ is solved as the smallest real roots of Eq. (2.2) for $x \in [-\infty, +\infty]$

with the boundary condition: $y(x) \to -\infty$ when $x \to \pm\infty$. It is also obvious that $y(\mu) = 0$.

The distribution is primarily controlled by the shape parameter λ. It becomes a normal distribution when $\lambda = 1$; a Laplace distribution when $\lambda = 2$; a cusp distribution when $\lambda = 3$. The main source of the volatility smile is $\lambda \neq 1$. Although our primary interest are these three integer cases, λ can be any positive real number due to analytic continuity. It forms a continuous distribution family.

The σ parameter is the volatility parameter that affects the standard deviation (stdev) of the distribution. It is also directly related to the minimum implied volatility of a volatility smile curve (See Eq. (4.14)). The focus in this paper is the high frequency regime where σ is reasonably small, e.g., from the order of 0.01 down to 0.001. The integrals involving small σ typically require the MPFR library to achieve decent precision unless certain scaling rules are applied to transform the raw formula to that of a unit distribution. For this reason, in the later part of the paper, the log-strike is often normalized by σ via $\hat{k} = \frac{k-\mu}{\sigma}$, such that $\hat{k}$ is measured in the units of σ.

The μ parameter is the drift term of the stochastic process. In the classic option pricing framework, its value is determined via the risk neutral requirement, which is denoted as μ_D. In the global regime, we have $\mu_D = \frac{\sigma^2}{4}$. However, in the elliptic framework, this doesn't need to be held true in the local regime. The volatility skew is produced from $\mu \neq \mu_D$.

The β parameter is the skew parameter. It is the coefficient of the xy term. If β is assumed not to change the tail behavior, then we must have $\lambda \geq 2$ for non-zero β [5]. This is one of the indications for the local regime vs the global regime. The role of β in our model is somewhat subtle. It creates the first moment and skewness (the third moment). For instance, the stock market indices generally have a negative skewness around -0.4 to -0.7. In general, β is expected to be in the range of $[-0.8, 0.8]$, in which the analysis is performed in this paper. However, its most meaningful effect is to provide the proper scale for the risk-neutral violation, $\frac{\mu-\mu_D}{\sigma} \sim O(\beta)$. See Section 4.7 for more detail.

The probability density function (PDF) of a λ distribution is defined as

$$P(x;\lambda,\sigma,\beta,\mu) = \frac{1}{C} e^{y(x)}, \text{ where } C = \int_{-\infty}^{\infty} e^{y(x)} dx. \tag{2.3}$$

C is the normalization constant to maintain the unity of the density function, $\int_{-\infty}^{\infty} P(x)\, dx = 1$. The cumulative density function (CDF) is obviously $\Phi(x) = E\left[\mathbf{1}_{\{X \leq x\}}\right] = \int_{-\infty}^{x} P(x) dx$; while at times it may be more suitable to study the tail behavior with the complementary CDF (CCDF), $1 - \Phi(x) = \int_{x}^{\infty} P(x) dx$. Due to the canonical structure of $\frac{x-\mu}{\sigma}$ in Eq. (2.2), the CDF has the following

[5] When $\lambda < 2$ and $\beta \neq 0$, then $y \sim -x$, and βxy always dominates the tails, instead of $(-y)^{\lambda}$. This is not good.

scaling law on μ, σ:

$$\Phi(x;\lambda,\sigma,\beta,\mu) = \begin{cases} \Phi(x-\mu;\lambda,\sigma,\beta,\mu=0), & \text{shifting } \mu; \\ \Phi(\frac{x}{\sigma};\lambda,\sigma=1,\beta,\frac{\mu}{\sigma}), & \text{scaling } \sigma; \\ \Phi(\frac{x-\mu}{\sigma};\lambda,\sigma=1,\beta,\mu=0), & \text{via unit distribution.} \end{cases} \tag{2.4}$$

When $\beta = 0$, we have the symmetric distribution,

$$\frac{(x-\mu)^2}{\sigma^2} = (-y)^\lambda \text{ , whose solution is } y(x) = -\left|\frac{x-\mu}{\sigma}\right|^{\frac{2}{\lambda}}. \tag{2.5}$$

This is the so-called "stretched exponential function"[6]. It has been used in many areas of physics in which most noticeable interpretation of σ is the mean relaxation time of a decay process. Survey of literatures shows that it has been treated as a distribution family since Subbotin (1923)[19][7]. Its CDF has been described as the "generalized error function"[8]. More recently a skew parameter has been added to it, and Hansen, McDonald, and Theodossiou [5] have categorized it as "generalized error distribution" (GED) within the larger "skew generalized t-distribution" family (SGT). Its skew form (SGED) can be viewed as a step-function approximation to the skew λ distribution (See Panel (2) of Figure A.1), and therefore is treated rigorously in Appendix A, following the same framework laid out in this paper [9].

The change of variable, $x = z^\lambda$, leads to the analytic solution of C,

$$C(\lambda,\sigma) = 2\sigma\int_0^\infty \lambda z^{\lambda-1}e^{-z^2}dz = \sigma\lambda\Gamma\left(\frac{\lambda}{2}\right) = \sigma\hat{C}(\lambda), \tag{2.6}$$

where $\hat{C}(\lambda) = \lambda\Gamma\left(\frac{\lambda}{2}\right)$ is the scale-independent normalization constant for unit distribution. When $\beta \neq 0$, $y(x)$ has known analytic forms at $\lambda = 2, 3$. So λ distribution can be viewed as a distribution family based on a peculiar kind of "skew stretched exponential function". Non-zero β has limited effect to C, typically within 8% of variation for $\beta \in [-0.8, 0.8]$, and most of the change can be captured by the first β term. This is addressed in Section 2.6.

The slope of $y(x)$ will be used in the truncation of IMGF in the local regime. It is

$$\frac{dy}{dx} = \frac{\frac{2}{\sigma}(x-\mu)+\beta y}{-\lambda\sigma(-y)^{\lambda-1}-\beta(x-\mu)}. \tag{2.7}$$

When $x = \mu$, $\frac{dy}{dx}$ is singular for $\lambda \geq 2$, else $\frac{dy}{dx} = 0$. On one hand, $\frac{dy}{dx} < 0$ when $x > \mu$; on the other hand, $\frac{dy}{dx} > 0$ when $x < \mu$. When $\beta = 0$, it is pretty

[6] https://en.wikipedia.org/wiki/Stretched_exponential_function

[7] See discussion about its history at http://mathoverflow.net/questions/144202/whats-the-name-of-this-distribution

[8] http://dlmf.nist.gov/7.16

[9] See Figure 1 of [5] for the family tree of SGT distribution. Its skew form (SGED) is constructed by simply multiplying σ with $(1+\text{sgn}(x-\mu)\beta)$. This is different from our origin of elliptic curves.

λ	C	var	kurtosis
1	$\sqrt{\pi}\sigma \approx 1.772\sigma$	$\frac{1}{2}\sigma^2$	3
2	2σ	$2\sigma^2$	6
3	$\frac{3}{2}\sqrt{\pi}\sigma \approx 2.659\sigma$	$\frac{105}{8}\sigma^2 \approx 13.125\sigma^2$	$\frac{429}{35} \approx 12.257$

TABLE 2.1
The normalization constant, variance and kurtosis of $\lambda = 1, 2, 3$

straightforward that $\frac{dy}{dx} = \frac{2}{\lambda}\frac{y}{(x-\mu)} = -\operatorname{sgn}(x-\mu)\frac{2}{\lambda\sigma}\left|\frac{x-\mu}{\sigma}\right|^{\frac{2}{\lambda}-1}$. Most of the complexity is caused by the presence of β. Since β is small, its effect on $\frac{dy}{dx}$ can be thought of as a perturbation to the symmetric solution.

2.3. Statistics of The Symmetric Distribution

When $\beta = 0$, the distribution is symmetric. The CDF and moments have analytic form. With $\mu = 0$, the moments can be integrated in a similar fashion as C. The odd moments are zero, and the even moments are

$$m_{(n)} = \frac{2\lambda}{C}\sigma^{n+1}\int_0^\infty z^{n\lambda+\lambda-1}e^{-z^2}dz = \sigma^n\frac{\Gamma\left(\frac{\lambda(n+1)}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)}. \tag{2.8}$$

[10] Therefore, the variance and kurtosis are

$$\text{var} = m_{(2)} = \sigma^2\frac{\Gamma\left(\frac{3\lambda}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)};\quad \text{kurtosis} = \frac{\Gamma\left(\frac{\lambda}{2}\right)\Gamma\left(\frac{5\lambda}{2}\right)}{\Gamma\left(\frac{3\lambda}{2}\right)^2}. \tag{2.9}$$

Both the variance and kurtosis are increasing with λ, as listed in Table 2.1.

The CDF, $\Phi(x)$, and CCDF, $1-\Phi(x)$, of a symmetric λ distribution can be expressed by the incomplete gamma function $\Gamma(s,x) = \int_x^\infty t^{s-1}e^{-t}dt$,[11]

$$\begin{aligned}\Phi(x;\lambda,\sigma,\mu) &= \frac{1}{2\,\Gamma\left(\frac{\lambda}{2}\right)}\Gamma\left(\frac{\lambda}{2},\left|\frac{x-\mu}{\sigma}\right|^{\frac{2}{\lambda}}\right), \quad \text{when } (x-\mu)<0;\\ 1-\Phi(x;\lambda,\sigma,\mu) &= \frac{1}{2\,\Gamma\left(\frac{\lambda}{2}\right)}\Gamma\left(\frac{\lambda}{2},\left|\frac{x-\mu}{\sigma}\right|^{\frac{2}{\lambda}}\right), \quad \text{when } (x-\mu)\geq 0.\end{aligned} \tag{2.10}$$

Obviously, $\Phi(\mu) = \frac{1}{2}$ due to $\Gamma\left(\frac{\lambda}{2},0\right) = \Gamma\left(\frac{\lambda}{2}\right)$. For $\lambda = 1$, $\Gamma\left(\frac{1}{2},x\right) = \Gamma\left(\frac{1}{2}\right)(1-\operatorname{erf}(x))$ and the CDF becomes $\Phi(x;\lambda=1) = \frac{1}{2}\left(\operatorname{erf}\left(\frac{x-\mu}{\sigma}\right)+1\right)$ for all x. For $\lambda = 2$, $\Gamma(1,x) = e^{-x}$ and the CDF becomes that of a Laplace distribution (See Eq. (2.26) with $\beta = 0$). For $\lambda = 3$, the CDF of a cusp distribution has been solved in the elliptic distribution framework [15]:

$$\begin{aligned}\Phi(x;\lambda=3) &= \Phi_{\Lambda C}\left(\left|\frac{x-\mu}{\sigma}\right|\right), \quad \text{when } (x-\mu)<0;\\ 1-\Phi(x;\lambda=3) &= \Phi_{\Lambda C}\left(\left|\frac{x-\mu}{\sigma}\right|\right), \quad \text{when } (x-\mu)\geq 0;\end{aligned} \tag{2.11}$$

[10] This quantity will show up again in Eq. (3.46).

[11] See http://dlmf.nist.gov/8.2#i. In R, $\Gamma(s,x)/\Gamma(s)$ = `pgamma(x, s, lower = FALSE)`. One short coming of `Rmpfr` package is that it didn't re-implement `pgamma` as it did to `erf`.

$$\text{with the cusp CCDF: } \Phi_{\Lambda C}(x) = \frac{1}{\sqrt{\pi}} x^{\frac{1}{3}} e^{-x^{2/3}} + \frac{1}{2}\left(1 - \text{erf}\left(x^{\frac{1}{3}}\right)\right). \quad (2.12)$$

Since all the moments exist, the characteristic function (CF) can be derived as

$$\varphi_{sym}(t;\lambda,\sigma,\mu) = E\left[e^{itX}\right] = e^{it\mu}\left[1 + \sum_{n=2,4,\ldots}^{\infty} (-1)^{\frac{n}{2}} \frac{m_{(n)} t^n}{n!}\right]. \quad (2.13)$$

Likewise, the moment generating function (MGF) for the symmetric distribution is

$$M_{sym}(t;\lambda,\sigma,\mu) = E\left[e^{tX}\right] = e^{t\mu}\left[1 + \sum_{n=2,4,\ldots}^{\infty} \frac{m_{(n)} t^n}{n!}\right]. \quad (2.14)$$

For $\lambda = 1, 2$, there are well-known analytic formulae from the normal distribution and Laplace distribution,

$$\begin{aligned} M_{sym}(t;\lambda=1,\sigma,\mu) &= e^{t\mu+\sigma^2 t^2/4}; \\ M_{sym}(t;\lambda=2,\sigma,\mu) &= e^{t\mu}\left(1-\sigma^2 t^2\right)^{-1}, \quad \text{for } \sigma t < 1. \end{aligned} \quad (2.15)$$

The asymmetric counterpart for $\lambda = 2$ is in Eq. (3.6). The subtlety to notice here is that, at $\lambda = 2$, σ has an upper limit t^{-1} if MGF must converge. This is an important theme that I want to convey repeatedly in this paper - as λ increases in the sub-linear local regime, σ must be confined to smaller ranges if MGF "must" converge. That is, local regime only exists for high frequency log-returns, and therefore, for short maturity options.

2.4. Local Regime vs Global Regime

In this paper, the local regime is defined as $\lambda \geq 2$. The global regime is $\lambda = 1$. I don't find any use for λ between 1 and 2 in this paper, so this range is ignored. In the local regime, we encounter the phenomena called "MGF explosion" (In other literatures such as [2], it may be called "moment explosion"). This is the major obstacle in many stochastic volatility models. The volatility σ needs an upper limit, depending on λ and β. This is required in order for MGF integral to converge.

In addition, in local regime, the right tail of the distribution must be truncated in the MGF (and IMGF) integrals. This is a central subject in this paper. There are two ways to approach the subject:

1. The truncation in the MGF summation;
2. The truncation of the right tail in the MGF integral.

Exact mathematical procedures are developed to handle them. My attempt is to make them as mathematically elegant as possible. Here we study the first approach, in which case we use $n_{\overrightarrow{\infty}}$ to label the largest n in $\sum_{n=1}^{N}$. The second approach is studied in Section 3.1. When the integrals are truncated, we use

$x_{\overrightarrow{\infty}}$ to label the upper limit of such integrals, or simply $\int^{\overrightarrow{\infty}}$. The location of truncation (for either $x_{\overrightarrow{\infty}}$ and $n_{\overrightarrow{\infty}}$) is at infinity for $\lambda = 2$; and it gets smaller as λ increases. Therefore, $\lambda = 2$ is the boundary of the local regime.

The truncation justifies the violation of risk neutrality. The option prices calculated in the local regime is not directly observable in general. The prices must be transformed to the global regime to make them observable. This transformation is called λ transformation because it is thought of as projecting the data from one λ to another. The exception is the options with one-day maturity where the effect of λ transformation is small.

2.5. Truncation of MGF Summation

Each term in the MGF of a symmetric distribution in Eq. (2.14) is denoted as ($e^{t\mu}$ ignored)

$$G(n;\lambda,\sigma t) = \frac{m_{(n)}t^n}{n!} = (\sigma t)^n \frac{\Gamma\left(\frac{\lambda(n+1)}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)\Gamma\left(n+1\right)}. \tag{2.16}$$

Here we merge t with σ since most of the time they show up in pairs, and bear in mind that t ends up to be 1 in the option pricing model. For a reasonably small σ, $G(n;\lambda,\sigma t)$ is decreasing as n increases; but when $\lambda > 2$, $G(n;\lambda,\sigma t)$ always becomes increasing when n is large enough. Where $G(n;\lambda,\sigma t)$ reaches minimum is where the summation of MGF should be truncated. Expressed in terms of the digamma function $\psi(x) = \frac{d}{dx}\log\Gamma(x)$, the truncation point, denoted as $n_{\overrightarrow{\infty}}$, is the root of

$$\frac{d}{dn}\log G(n;\lambda,\sigma t) = \log(\sigma t) + \frac{\lambda}{2}\psi\left(\frac{\lambda\left(n+1\right)}{2}\right) - \psi(n+1) = 0. \tag{2.17}$$

We can solve it numerically at $t = 1$. For instance, for $\lambda = 3$, when $\sigma = 0.1$, we have $n_{\overrightarrow{\infty}} = 28.63$ and $M_{sym}(t=1) = 1.076985$. When $\sigma = 0.01$, $n_{\overrightarrow{\infty}} = 2961.96$ and $M_{sym}(t=1) = 1.000657$. These truncations are satisfactory since they occur far enough in the summation. For reasonably large n, the asymptotic expansion of the digamma function can be used to obtain closed form solution. With $\psi(x) = \log(x) + ...$[12], we have

$$n_{\overrightarrow{\infty}} + 1 \approx \sigma t\left(\frac{2}{\sigma t\lambda}\right)^{\frac{\lambda}{\lambda-2}}. \tag{2.18}$$

This formula turns out to be quite precise. At $n_{\overrightarrow{\infty}} = 28.63$, its error is less than 10^{-4}. This result is very interesting as it shows up again in Eq. (3.8) such that $n_{\overrightarrow{\infty}}+1 \approx x_{\overrightarrow{\infty}}$. That is, where the power series is truncated is where the right tail is truncated in MGF integral. This is the foundation for the incomplete moment expansion in Section 3.8.3 and small σ limit in Section 4.

[12] https://en.wikipedia.org/wiki/Digamma_function#Computation_and_approximation

On the other hand, there is an upper limit how large σ can be for each λ. Assume we have a minimum requirement $n_{\min}$ for $n_{\not\approx}$ in Eq. (2.17), we get (assume $t = 1$)

$$\log(\sigma_{\max}(\lambda)) = \psi(n_{\min}+1) - \frac{\lambda}{2}\psi\left(\frac{\lambda\,(n_{\min}+1)}{2}\right) \tag{2.19}$$

When $\lambda = 2$, the two digamma functions cancel out and $\sigma_{\max}(\lambda) = 1$. This is consistent with the limit set by Eq. (2.15). For $\lambda > 2$, $\sigma_{\max}(\lambda)$ decreases as λ increases. Assume $n_{\min} = 1$, then, at $\lambda = 3$, we have $\sigma_{\max} = 0.38$, which is $\exp\left(\frac{1}{2}\gamma - \frac{5}{4}\right)$ via digamma's recurrence relation, $\psi(n+1) = \psi(n) + \frac{1}{n}$ and $\psi(1) = -\gamma$, where γ is the Euler-Mascheroni constant. The real-world application should use σ much smaller than this since the standard deviation is nearly 100% at this level of $\sigma_{\max}$. The MGF truncation and the existence of $\sigma_{\max}(\lambda)$ solidify the concept of the local regime.

2.6. Statistics of Asymmetric Distribution

When $\beta \neq 0$, the distribution is asymmetric. Although most of the calculations have to be numerical, there are still a few analytic solutions that can enhance our insights and serve as unit tests for numerical implementation. If β is assumed not to change the tail behavior, then we must have $\lambda \geq 2$. Therefore, we will focus our effort in the range of $\lambda \in [2, 3]$. $y(x)$ has known analytic formulae only at $\lambda = 2, 3$.

2.6.1. Case I: $\lambda = 2$

For $\lambda = 2$, with $B_0 = \sqrt{1 + \frac{\beta^2}{4}}$ and $B^{\pm} = B_0 \pm \frac{\beta}{2}$, we reach a form of skew Laplace distribution,

$$\begin{array}{ll} y^-(x,\beta) = -B^+\left|\frac{x-\mu}{\sigma}\right|, & \text{when } (x-\mu) < 0; \\ y^+(x,\beta) = -B^-\left|\frac{x-\mu}{\sigma}\right|, & \text{when } (x-\mu) \geq 0. \end{array} \tag{2.20}$$

Its normalization constant is

$$C(\lambda = 2, \sigma, \beta) = \sigma\hat{C}(\lambda = 2, \beta) = 2\sigma B_0, \tag{2.21}$$

which is approximately $2\sigma\left(1 + \frac{1}{8}\beta^2\right)$. The algebraic rules for $B^{\pm}$ are

$$\begin{array}{c} B^+B^- = 1, \\ B^+ + B^- = 2B_0 = \hat{C}(\lambda = 2, \beta), \\ \triangle B = B^+ - B^- = \beta. \end{array} \tag{2.22}$$

Its moments have analytic form:

$$m_{(n)}(\beta, \lambda = 2) = \sigma^n \frac{\Gamma\,(n+1)}{2B^0}\left(\left(B^+\right)^{n+1} + (-1)^n\left(B^-\right)^{n+1}\right). \tag{2.23}$$

The first 4 moments are straightforward:

$$\begin{aligned} m_{(1)} &= \sigma\beta, \\ m_{(2)} &= 2\sigma^2\left(1+\beta^2\right), \\ m_{(3)} &= 6\sigma^3\beta\left(2+\beta^2\right), \\ m_{(4)} &= 24\sigma^4\left(1+3\beta^2+\beta^4\right). \end{aligned} \tag{2.24}$$

and the variance, skewness and kurtosis are

$$\begin{aligned} \text{var} &= \sigma^2\left(2+\beta^2\right), \\ \text{skewness} &= \tfrac{2\beta\left(3+\beta^2\right)}{\left(2+\beta^2\right)^{3/2}} \approx \tfrac{3}{\sqrt{2}}\beta\left(1-\tfrac{5}{12}\beta^2\right). \\ \text{kurtosis} &= \tfrac{3\left(8+12\beta^2+3\beta^4\right)}{\left(2+\beta^2\right)^2}. \end{aligned} \tag{2.25}$$

The CDF, $\Phi(x)$, and CCDF, $1-\Phi(x)$, are

$$\begin{aligned} \Phi(x) &= \tfrac{1}{2B_0B^+}\, e^{-B^+\left|\frac{x-\mu}{\sigma}\right|}, \qquad \text{when } (x-\mu)<0; \\ 1-\Phi(x) &= \tfrac{1}{2B_0B^-}\, e^{-B^-\left|\frac{x-\mu}{\sigma}\right|}, \qquad \text{when } (x-\mu)\geq 0. \end{aligned} \tag{2.26}$$

Notice the simple result: $\Phi(\mu) = \frac{B^-}{2B_0}$ at $x=\mu$. As $\beta\to 0$, the asymmetric solutions converge to the symmetric solutions continuously. This is very nice. It is not so "fortunate" in the case of $\lambda = 3$.

2.6.2. Case II: $\lambda = 3$

For $\lambda = 3$, it is called "asymmetric standard cusp distribution", whose $y(x)$ has been solved analytically via trigonometric solutions of elliptic distribution (See Lihn 2015[15]). With $x_0 = -\frac{4\beta^3}{27}$, $V = \left|\frac{x-\mu}{x_0\sigma}\right|^{\frac{1}{2}}$ and $W = 2\left|\frac{\beta(x-\mu)}{3\sigma}\right|^{\frac{1}{2}}$, we have

$$\begin{aligned} y^-(x,\beta) &= -J\left(\tfrac{1}{3}J_{arc}(V)\right)W, \qquad \text{when } (x-\mu)<0; \\ y^+(x,\beta) &= -I\left(\tfrac{1}{3}I_{arc}(V)\right)W, \qquad \text{when } (x-\mu)\geq 0; \end{aligned} \tag{2.27}$$

where $(I,J) = (\sinh,\cosh)$ when $\beta>0$; and $(I,J) = (\cosh,\sinh)$ when $\beta<0$. The $_{arc}$ suffix in I_{arc} and J_{arc} are mapped to the arcsinh and arccosh functions respectively. Notice that V is singular at $\beta = 0$. Therefore, analytical continuity between the asymmetric solutions and symmetric solutions has to go through infinity, which is not so good for numerical implementation.

Note that $\cosh\left(\frac{1}{3}\,\text{arccosh}(x)\right)$ is peculiarly associated with Chebychev polynomial $T_n(x)$ where $n=\frac{1}{3}$. In this context, $T_1(x) = x$ seems to be associated with $y(x)$ for $\lambda = 2$; and $T_2(x) = 2x^2-1$ with $y(x)$ for $\lambda = 1$. But for $\lambda = 3$, it becomes $\sqrt{x}\,T_{\frac{1}{3}}\left(\sqrt{x}\right)$. These are my wild speculation[13].

The CDF and moments have to be integrated numerically. This is so for all λ's greater than 2. For non-integer λ, in order to achieve good numerical

[13]Furthermore, in Eq. (15.4.3) of [1], $T_n(1-2x) = F(-n,n;\frac{1}{2};x)$ where F is the hypergeometric function ${}_2F_1$.

performance when solving $y(x)$, the R implementation assumes λ is a rational number $\frac{m}{n}$ and still resorts to polynomial root solver,

$$\begin{gathered}(-y)^m - \left(\hat{x}^2 + \beta\hat{x}\hat{y}\right)^n = 0, \\ \text{where } \hat{x} = \tfrac{x-\mu}{\sigma}.\end{gathered} \tag{2.28}$$

This method works well for $m < 30$. For instance, $\lambda = 3.125$ and 2.75 are used to fit SPX options, they correspond to $\frac{25}{8}$ and $\frac{11}{4}$.

For $\lambda \geq 2$, the integral needs to be performed separately on each side of the cusp numerically. Assume $\hat{y}^{\pm}(x)$ are the solutions for the unit distribution, then the moments (and C) have the form of:

$$\hat{C}(\lambda,\beta) = \int_0^\infty \left(e^{\hat{y}^+(x)} + e^{\hat{y}^-(x)}\right) dx. \tag{2.29}$$

$$m_{(n)}(\beta,\lambda) = \frac{\sigma^n}{\hat{C}(\lambda,\beta)} \int_0^\infty x^n \left(e^{\hat{y}^+(x)} + (-1)^n\, e^{\hat{y}^-(x)}\right) dx. \tag{2.30}$$

The CDF, $\Phi(x)$, and CCDF, $1-\Phi(x)$, are

$$\begin{aligned}\Phi(x) &= \tfrac{1}{\hat{C}(\lambda,\beta)} \textstyle\int_{-\infty}^{\hat{x}} e^{\hat{y}^-(x)} dx, \qquad \text{when } \hat{x} = \tfrac{x-\mu}{\sigma} < 0; \\ 1-\Phi(x) &= \tfrac{1}{\hat{C}(\lambda,\beta)} \textstyle\int_{\hat{x}}^{\infty} e^{\hat{y}^+(x)} dx, \qquad \text{when } \hat{x} = \tfrac{x-\mu}{\sigma} \geq 0.\end{aligned} \tag{2.31}$$

The $B^{\pm}$ notation can be extended to $\lambda > 2$ as $B^{\pm}(\lambda,\beta) = \int_0^\infty e^{\hat{y}^{\pm}(x)} dx$, which are integrals of the entire tail on each side of the cusp. By this definition, we have $B^-(\lambda,\beta) = \Phi(\mu)\,\hat{C}(\lambda,\beta)$. The algebraic rules become

$$\begin{aligned}B^+(\lambda,\beta) + B^-(\lambda,\beta) &= \hat{C}(\lambda,\beta); \\ \triangle B = B^+(\lambda,\beta) - B^-(\lambda,\beta) &= \beta.\end{aligned} \tag{2.32}$$

The first rule comes directly from the definition of the CDF itself. The second rule on $\triangle B$ is new - it is based on numerical result, yet it is very precise for every λ, β. The exactness of the difference is amazing. This indicates the skew parameter enters λ distribution in an elegant way.

At $\lambda = 3$, the regression analysis in the range of $\beta \in (0.1, 0.8)$ yields simple numerical results for the first moment $m_{(1)}$ and skewness S:

$$\begin{gathered}m_{(1)} \approx 1.2353\,\sigma\beta, \\ S \approx 1.2113\,\beta, \\ m_{(1)} \approx \operatorname{sgn}(S)\,\sigma\,(0.00178 + 0.9875\,|S| + 0.0405\,|S|^2).\end{gathered} \tag{2.33}$$

Such linear relations have profound implication. The skewness (S or β), the first moment, the risk-neutral drift (Eq. (3.10)) become almost synonym to one another at $\lambda = 3$ for short-maturity options.

2.6.3. Interpolation between $\lambda = 2$ and 3

More generally, we can use $\lambda = 2, 3$ as our base, and estimate the numeric formulae up to second order of β for C, $m_{(1)}$, $m_{(2)}$, and skewness (S) in the two dimensional space of $\lambda \in [2, 3]$ and $\beta \in (0.1, 0.8)$:

$$\begin{gathered} C \approx \lambda\Gamma\left(\tfrac{\lambda}{2}\right)\sigma\left(1+\tfrac{\beta^2}{8}\left(0.98-0.335\left(\lambda-2\right)\right)\right), \\ m_{(1)} \approx \left(1+0.235\left(\lambda-2\right)\right)\sigma\beta, \\ m_{(2)} \approx \Gamma_2\left(\lambda\right)\sigma^2\left(1+\tfrac{2}{\Gamma_2(\lambda)}\beta^2\right), \text{ where } \Gamma_2\left(\lambda\right) = \tfrac{\Gamma\left(\frac{3\lambda}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)} \\ \text{var} \approx \Gamma_2\left(\lambda\right)\sigma^2\left(1+\tfrac{\beta^2}{\Gamma_2(\lambda)}\right), \\ S \approx \tfrac{3}{\sqrt{2}}\beta\left(1-\tfrac{3}{7}\left(\lambda-2\right)\right)\left(1-\tfrac{4.5}{12}\left(3-\lambda\right)\beta^2\right). \end{gathered} \tag{2.34}$$

It is commonly known that equity indices, such as SPX, have negative skewness and it affects the dynamics of volatility surface. The 80-year daily data on DJIA shows the skewness is approximately -0.5[15][14]. At this level of skewness, if σ is small, $m_{(1)}$ is much larger than $m_{(2)}$ in magnitude. In addition, $m_{(1)}$ is negative for SPX, while $m_{(2)}$ is always positive. $m_{(1)}$ and $m_{(2)}$ have opposite signs. The risk neutral drift μ_D is dominated by skewness times σ. These facts are important in the study of the risk-neutral drift.

3. Option Pricing in Local Regime

In this section, the option pricing model is laid out for the local regime, $\lambda \geq 2$. The option prices obtained at the end of this section can explain the closing prices and implied volatilities of the options to be expired in one day. A cusp in the option prices is the signature that $\lambda \approx 3$ is indeed the correct model.

However, in order to model the option prices for longer maturities, e.g. 4 days or more, the prices will have to go through the so-called "λ transformation". The prices have to be projected from the local regime to the global regime via the implied volatility, then transported by the market momentum, and finally converted back to market prices. This complex transformation is the subject of Section 5.

3.1. Moment Generating Function

As a prerequisite of the option pricing model, the moment generating function (MGF) must be studied. The integral form of MGF is defined as

$$M(t;\lambda,\sigma,\mu,\beta) = E\left[e^{tX}\right] = \int_{-\infty}^{\overrightarrow{\infty}} e^{tx}P(x;\lambda,\sigma,\mu,\beta)\,dx. \tag{3.1}$$

The special symbol $\overrightarrow{\infty}$ is used to represent the truncation of positive infinity, since, strictly speaking, the MGF integral diverges in the right tail when $\lambda > 2$.

[14] -0.5 is the asymptotic skewness from ecd fit. The statistics of the data is -1.1, much higher due to the 1987 crash.

As a location-scale distribution family, the MGF has the following scaling law on μ, σ:

$$M(t;\lambda,\sigma,\beta,\mu) = \begin{cases} e^{t\mu}\, M(t;\lambda,\sigma,\beta,\mu=0), & \text{shifting } \mu; \\ M(\sigma t;\lambda,\sigma=1,\beta,\frac{\mu}{\sigma}), & \text{scaling } \sigma; \\ e^{t\mu}\, M(\sigma t;\lambda,\sigma=1,\beta,\mu=0), & \text{via unit distribution.} \end{cases} \tag{3.2}$$

And here the risk-neutral drift is also defined via MGF,

$$\mu_D = \begin{cases} -\log\left(M\left(t=1;\lambda,\sigma,\mu=0,\beta\right)\right), & \text{or} \\ -\log\left(M\left(\sigma;\lambda,\sigma=1,\mu=0,\beta\right)\right), & \text{via unit distribution.} \end{cases} \tag{3.3}$$

Conversely, we can write $M(t=1;\lambda,\sigma,\mu=0,\beta) = e^{-\mu_D}$. However, we will only focus on the mathematics of μ_D and defer the discussion of its meaning until Section 3.7[15]. And when x represents the log-returns, $t=1$ is all that we need to consider. In a more general setting, t is in the same role as σ.

For $\lambda = 1$, $M(t)$ is the well-known analytic solution of a normal distribution (β must be zero)

$$M(t;\lambda=1,\sigma,\mu) = e^{t\mu} e^{\sigma^2 t^2/4}. \tag{3.4}$$

The risk-neutral drift is the well-known drift term of the geometric Brownian motion (GBM),

$$\mu_D(\lambda=1) = -\tfrac{\sigma^2}{4}, \tag{3.5}$$

For $\lambda = 2$, $M(t)$ has analytic solution as long as $1-\beta\sigma t - \sigma^2 t^2 > 0$,

$$M(t;\lambda=2,\sigma,\mu,\beta) = e^{t\mu}\left(1-\beta\sigma t-\sigma^2 t^2\right)^{-1}. \tag{3.6}$$

The risk-neutral drift is

$$\mu_D(\lambda=2) = \log\left(1-\beta\sigma-\sigma^2\right). \tag{3.7}$$

In the global regime ($\lambda = 1$), σ can be any value, there is no limit. But in the local regime ($\lambda \geq 2$), σ has an upper limit $\sigma_{\max}$. At $\lambda = 2$, this limit is set by $1-\beta\sigma_{\max}-\sigma_{\max}^2 = 0$, whose solution is $\sigma_{\max} = B^-$. This upper limit gets smaller as λ increases.

When $\lambda > 2$, the right tail of MGF integral must be truncated. The location of truncation is denoted as $x_{\rightrightarrows\infty}$. It is determined by the condition: $dy/dx + t = 0$, in the right tail ($x - \mu \gg \sigma$). When σ is small, the divergence in the MGF integral occurs very far into the right tail, usually several hundred standard deviations away, which justifies the truncation. When $\beta \neq 0$, solving $\frac{dy}{dx} + t = 0$ is a numerical root-finding task. With $\mu = \beta = 0$, the truncation occurs at root of $\frac{dy}{dx} + t = -\frac{2}{\lambda\sigma}\left(\frac{x}{\sigma}\right)^{\frac{2}{\lambda}-1} + t = 0$. The analytic solution is

$$x_{\rightrightarrows\infty} = \sigma\left(\frac{2}{\sigma t\lambda}\right)^{\frac{\lambda}{\lambda-2}}. \tag{3.8}$$

[15]Since $M(t;\mu) = e^{t\mu}M(t;\mu=0)$, it is intuitive that μ must to be μ_D in order to eliminate one degree of freedom.

Eq. (3.8) is identical to the asymptotic digamma solution in Eq. (2.18). That is, $n_{\overrightarrow{\infty}}+1=x_{\overrightarrow{\infty}}$, which implies where the power series is truncated is where the integral is truncated. And the truncated MGF value from integral method is exactly the same that of summation method. For $\lambda=3$ and $\sigma=0.1$, we have truncation occurring at $x_{\overrightarrow{\infty}}=29.63$ and $M(t=1)=1.076985$. For $\sigma=0.01$, we have truncation occurring at $x_{\overrightarrow{\infty}}=2962.96$ and $M(t=1)=1.000655$. The truncation at 300σ is numerically precise enough for most financial applications[16]. At a fixed σ, due to the exponent $\frac{\lambda}{\lambda-2}$ in Eq. (3.8), $x_{\overrightarrow{\infty}}$ increases exponentially as λ decreases and reaches infinity at $\lambda=2$. So tail truncation in MGF is a special phenomena for $\lambda>2$, which justifies the concept of the local regime.

The MGF integral for μ_D can also be represented in unit distribution that allows maximum numeric precision for small σ,

$$e^{-\mu_D}=M(t=1;\mu=0)=M(\sigma;\lambda,\mu=0,\beta)=1+\\ \frac{1}{\tilde{C}(\lambda,\beta)}\left(\int_0^{\overrightarrow{\infty}}\left(e^{\sigma x}-1\right)e^{\hat{y}^+(x)}dx+\int_0^{\infty}\left(e^{-\sigma x}-1\right)e^{\hat{y}^-(x)}dx\right). \tag{3.9}$$

Notice these are integrals of the entire tail on each side of the cusp. The truncation is determined by $\frac{d\hat{y}}{dx}+\sigma=0$, which is at $x_{\overrightarrow{\infty}}=\left(\frac{2}{\sigma\lambda}\right)^{\frac{\lambda}{\lambda-2}}$ for a symmetric distribution.

Furthermore, for very small σ, only the first and second moments matter in MGF and there is a short cut for μ_D,

$$M(t=1)\approx 1+m_{(1)}+\frac{m_{(2)}}{2},\\ \mu_D\approx -m_{(1)}+\frac{1}{2}\left(m_{(1)}^2-m_{(2)}\right). \tag{3.10}$$

For example, the equity market indices have skewness of -0.4 to -0.7. At $\lambda=3$, this corresponds to $\beta=-0.3$ to -0.6. Coincidentally in this range, $m_{(1)}$ can be approximated by the skewness times σ within 2% error. In Table 3.1, μ_D is calculated by three different methods and in various scenarios of σ at $\lambda=3$ and $\beta=-0.5$. We see that, for $\sigma=0.1$, $m_{(1)}$ is in the same order as $m_{(2)}$ but with opposite signs. When σ is reduced to 0.05, m_2 term cancels out half of $m_{(1)}$ term. At $\sigma=0.01$, $m_{(1)}$ is clearly much larger than $m_{(2)}$. The effect of skewness dominates μ_D. The second-order approximation is less than 0.1% error. At $\sigma\leq 0.005$, $m_{(1)}$ estimate is within less than 6% error. The first-order approximation is reasonably precise. As you can see, using the small-σ approximation can reduce the computing time significantly.

3.2. Insight

The MGF truncation procedure and the existence of an upper limit $\sigma_{\max}$ reveals an important insight for the physics of the financial market: High kurtosis distribution is only meaningful in the small variance domain. To extend it further, since the variances of log-return distributions of financial assets, such as equities currencies, commodities, and bonds, are proportional to the sampling

[16]This limit is set by `.ecd.mpfr.N.sigma` in the `ecd` package.

method	$\sigma = 0.1$	$\sigma = 0.05$	$\sigma = 0.01$	$\sigma = 0.005$	$\sigma = 0.0015$
est. maturity			30 days	4-7 days	1 day
$-\log M(t=1;\mu=0)$	-0.006920	0.01459	0.005541	0.002935	0.0009152
$-m_{(1)} + \frac{1}{2}\left(m_{(1)}^2 - m_{(2)}\right)$	-0.004431	0.01440	0.005537	0.002934	0.0009152
$-m_{(1)}$	0.062011	0.03101	0.006201	0.003101	0.0009302

TABLE 3.1

Various scenarios of μ_D calculations at $\lambda = 3$ and $\beta = -0.5$. Note that, for weekly and monthly expiries, λ would decrease as maturity increases. This will increase the accuracy of the estimates.

period, var $\sim T$, it follows that high kurtosis distribution is only meaningful in the high frequency regime, from daily to a few weeks; that is, $T \ll 1$, when T is measured by the fraction of a year.

There is a distant analogy to the quantum physics. Quantum mechanics manifests itself in the atomic scale, while the Newtonian mechanics is observable at the large scale. This is what I have in mind for the λ transformation: There is a local regime where the high kurtosis distribution exerts its physics, and such effect can't be observed directly. There is a global regime (aka the venerable normal distribution) where the implied volatility can be projected to option prices, which can be observed by the market practitioners.

The first obstacle of quantum electromagnetism is to work around infinity. Here we have a similar challenge - taming the infinity of MGF so that we can move on to build out the option pricing model in the local regime.

Since the MGF must be truncated when $\lambda > 2$, it also indicates the "true" value of MGF can not be known. Because of this, there is an uncertainty as to what the ideal μ is (think of the uncertainty principle in quantum mechanics). This leads to the violation of risk neutrality in the local regime. As we will see, in order to explain the volatility skew in the SPX option data, μ must be different from μ_D and small constants are added to the option prices.

3.3. Incomplete MGF

Given a distribution $P(x;\lambda,\sigma,\beta,\mu)$, the incomplete MGF (IMGF) is defined as

$$\begin{aligned} M_c(k,t) &= E\left[e^{tX}\right]_{X\geq k} = \int_k^{\overrightarrow{\infty}} e^{tx} P(x)\,dx, \quad \text{for call option;} \\ M_p(k,t) &= E\left[e^{tX}\right]_{X\leq k} = \int_{-\infty}^{k} e^{tx} P(x)\,dx, \quad \text{for put option.} \end{aligned} \tag{3.11}$$

Since $t = 1$ in most cases, we can define the shorthand function,

$$M_{c,p}(k) = M_{c,p}(k, t=1). \tag{3.12}$$

And we have the parity equations: $M_c(k,t) + M_p(k,t) = M(t)$ and $M_c(-\infty) = M_p(\overrightarrow{\infty}) = M(t=1)$. Remember $M(t=1) = 1$ in risk-neutral setting by definition.

As a location-scale distribution family, the IMGF has the following scaling law on μ, σ:

$$M_{c,p}(k,t;\lambda,\sigma,\beta,\mu) = \begin{cases} e^{t\mu}\, M_{c,p}(k-\mu,t;\lambda,\sigma,\beta,\mu=0), & \text{shifting } \mu; \\ M_{c,p}(\frac{k}{\sigma},\sigma t;\lambda,\sigma=1,\beta,\frac{\mu}{\sigma}), & \text{scaling } \sigma; \\ e^{t\mu}\, M_{c,p}(\frac{k-\mu}{\sigma},\sigma t;\lambda,\sigma=1,\beta,\mu=0), & \text{via unit distribution.} \end{cases} \tag{3.13}$$

Since σ and t in IMGF always show up in pairs and μ can always be factored out to $e^{t\mu}$, we can set $t=1$ without losing any information.

For $\lambda = 1$, $M_{c,p}(k)$ is straightforward:

$$\begin{aligned} M_c(k;\lambda=1,\sigma,\mu) &= \tfrac{1}{2} e^{\mu+\sigma^2/4} \left(1 - \operatorname{erf}\left(\tfrac{k-\mu}{\sigma} - \tfrac{\sigma}{2}\right)\right), \\ M_p(k;\lambda=1,\sigma,\mu) &= \tfrac{1}{2} e^{\mu+\sigma^2/4} \left(1 + \operatorname{erf}\left(\tfrac{k-\mu}{\sigma} - \tfrac{\sigma}{2}\right)\right). \end{aligned} \tag{3.14}$$

Substituting μ with the risk-neutral drift $\mu_D(\lambda=1) = -\sigma^2/4$, we arrive at the component in the Black-Scholes model,[17]

$$\begin{aligned} M_c(k;\lambda=1,\sigma,\mu_D) &= \tfrac{1}{2} - \tfrac{1}{2} \operatorname{erf}\left(\tfrac{k}{\sigma} - \tfrac{\sigma}{4}\right). \\ M_p(k;\lambda=1,\sigma,\mu_D) &= \tfrac{1}{2} + \tfrac{1}{2} \operatorname{erf}\left(\tfrac{k}{\sigma} - \tfrac{\sigma}{4}\right). \end{aligned} \tag{3.15}$$

For $\lambda = 2$, the algebra for $B^{\pm}$ must be expanded to $B^{\pm}_{\sigma} = B^{\pm} \pm \sigma$. The algebraic rules for $B^{\pm}_{\sigma}$ are (1) $B^{+}_{\sigma} + B^{-}_{\sigma} = 2B_0$; (2) $B^{+}_{\sigma} - B^{-}_{\sigma} = \beta + 2\sigma$; (3) $B^{+}_{\sigma} B^{-}_{\sigma} = 1 - \beta\sigma - \sigma^2 = e^{\mu_D}$. The third rule is quite impressive. $M_{c,p}(k;\lambda = 2,\sigma,\beta,\mu)$ can be carried out analytically on each side of the cusp:

$$\begin{aligned} M_c(k;\lambda=2,\sigma,\beta,\mu) &= \tfrac{1}{2} \tfrac{e^{\mu}}{B^{-}_{\sigma} B_0}\, e^{-B^{-}_{\sigma} \frac{k-\mu}{\sigma}}, \quad && \text{when } k-\mu \geq 0; \\ M_p(k;\lambda=2,\sigma,\beta,\mu) &= \tfrac{1}{2} \tfrac{e^{\mu}}{B^{+}_{\sigma} B_0}\, e^{-B^{+}_{\sigma} \left|\frac{k-\mu}{\sigma}\right|}, \quad && \text{when } k-\mu < 0. \end{aligned} \tag{3.16}$$

This equation is very similar to the CDF formula in Eq. (2.26).

The more general approach for $\lambda > 2$ is to integrate the tail on each side of the cusp through a unit distribution, following the truncation procedure of $x_{\rightrightarrows}$ in Section 3.1,

$$\begin{aligned} M_c(k;\lambda,\sigma,\beta,\mu) &= \tfrac{e^{\mu}}{\hat{C}(\lambda,\beta)} \int_{\hat{k}}^{\rightrightarrows} e^{\hat{y}^{+}(x)+\sigma x} dx, \quad && \text{when } \hat{k} = \tfrac{k-\mu}{\sigma} \geq 0; \\ M_p(k;\lambda,\sigma,\beta,\mu) &= \tfrac{e^{\mu}}{\hat{C}(\lambda,\beta)} \int_{-\hat{k}}^{\infty} e^{\hat{y}^{-}(x)-\sigma x} dx, \quad && \text{when } \hat{k} = \tfrac{k-\mu}{\sigma} < 0. \end{aligned} \tag{3.17}$$

We can extend the $B^{\pm}_{\sigma}$ notation to $\lambda > 2$. Define $B^{+}_{\sigma}(\lambda) = M_c(k=\mu)\,\hat{C}(\lambda,\beta)$ and $B^{-}_{\sigma}(\lambda) = M_p(k=\mu)\,\hat{C}(\lambda,\beta)$. Then the algebraic rules for $B^{\pm}_{\sigma}$ are (1) $B^{+}_{\sigma} + B^{-}_{\sigma} = \hat{C}(\lambda,\beta)$; (2) $B^{+}_{\sigma} - B^{-}_{\sigma} \approx \beta + \Gamma(\lambda+1)\,\sigma$. The first rule is the direct result of $M_c + M_p = 1$. The second rule is derived from numerical results. The precision is okay, but not perfect for $\lambda > 2$, indicating it is the first order approximation of a full, yet unknown, conservation law.

[17] Notice that $\left(\frac{k}{\sigma} - \frac{\sigma}{4}\right)$ is simply $\left(\frac{k-\mu_D}{\sigma}\right)$.

3.4. Incomplete Moments

As in the case of MGF integral, I've shown that truncation of the integral is not the only way to handle the infinity. The other alternative is to tackle it with moment expansion. This leads to the small σ limit of OGF, which produces elegant analytic solution for volatility smile.

Writing $M_{c,p}$ as moment expansion of a unit distribution, we have

$$\begin{aligned} M_c(k;\lambda,\sigma,\beta,\mu) &= e^{\mu}\sum_{n=0}^{n_{\overrightarrow{\infty}}}\frac{\sigma^n}{n!}E\left[X^n\right]_{X\geq\hat{k}}, \\ M_p(k;\lambda,\sigma,\beta,\mu) &= e^{\mu}\sum_{n=0}^{n_{\overrightarrow{\infty}}}\frac{\sigma^n}{n!}E\left[X^n\right]_{X<\hat{k}}, \end{aligned} \tag{3.18}$$

where $n_{\overrightarrow{\infty}}$ is the limit of the summation. It must be emphasized that, once e^{μ} is factored out, only the normalized $\hat{k}$ matters in the rest of this section. We can define each term in the series as incomplete moments $M_{c,p}^{(n)}(k)$, where

$$\begin{aligned} M_c^{(n)}(k;\lambda,\sigma,\beta) &= \sigma^n E\left[X^n\right]_{X\geq\hat{k}} = \sigma^n \hat{M}_c^{(n)}(\hat{k};\lambda,\beta), \\ M_p^{(n)}(k;\lambda,\sigma,\beta) &= \sigma^n E\left[X^n\right]_{X<\hat{k}} = \sigma^n \hat{M}_p^{(n)}(\hat{k};\lambda,\beta), \end{aligned} \tag{3.19}$$

On the right hand side, $\hat{M}_{c,p}^{(n)}(\hat{k})$ are the incomplete moments for the unit distribution, the smallest building blocks in moment expansion. The parity rule, $M_c^{(n)}(k)+M_p^{(n)}(k)=m_{(n)}$, allows us to derive the other half on each side of the cusp. Notice when $n=0$, $M_{c,p}^{(0)}(k)$ are just CCDF/CDF,

$$\begin{aligned} M_c^{(0)}(k;\lambda,\sigma,\beta) &= 1-\Phi(\hat{k}),\ \text{when } \hat{k}\geq 0; \\ M_p^{(0)}(k;\lambda,\sigma,\beta) &= \Phi(\hat{k}),\ \text{when } \hat{k}<0; \\ M_c^{(0)}(k;\lambda,\sigma,\beta)+M_p^{(0)}(k;\lambda,\sigma,\beta) &= 1. \end{aligned} \tag{3.20}$$

Because of the third rule, we have $M_c^{(0)}(k)=1-\Phi(k)$ and $M_p^{(0)}(k)=\Phi(k)$ for all k.

The symmetric solutions are straightforward. When $\beta=0$, they are based on the incomplete gamma function,

$$\begin{aligned} \hat{M}_c^{(n)}(\hat{k};\lambda,\beta=0) &= \tfrac{1}{2\Gamma\left(\frac{\lambda}{2}\right)}\,\Gamma\left(\tfrac{\lambda(n+1)}{2},\left|\hat{k}\right|^{\frac{2}{\lambda}}\right), \quad \text{when } \hat{k}\geq 0; \\ \hat{M}_p^{(n)}(\hat{k};\lambda,\beta=0) &= (-1)^n\ \hat{M}_c^{(n)}(\hat{k};\lambda,\beta=0), \quad \text{when } \hat{k}<0. \end{aligned} \tag{3.21}$$

The case for $\lambda=2$ is also quite simple, based on a slightly different (simpler) form of incomplete gamma function,

$$\begin{aligned} \hat{M}_c^{(n)}(\hat{k};\lambda=2,\beta) &= \tfrac{1}{\tilde{C}(\lambda,\beta)}\int_{\hat{k}}^{\infty}e^{-B^-x}x^n dx = \frac{\left(B^-\right)^{-n-1}}{2B_0}\,\Gamma\left(n+1,\hat{k}B^-\right), && \text{when } \hat{k}\geq 0; \\ \hat{M}_p^{(n)}(\hat{k};\lambda=2,\beta) &= \tfrac{(-1)^n}{\tilde{C}(\lambda,\beta)}\int_{-\hat{k}}^{\infty}e^{-B^+x}x^n dx = \frac{(-1)^n\left(B^+\right)^{-n-1}}{2B_0}\,\Gamma\left(n+1,-\hat{k}B^+\right), && \text{when } \hat{k}<0. \end{aligned} \tag{3.22}$$

The general asymmetric solutions have to be carried out numerically,

$$\begin{aligned} \hat{M}_c^{(n)}(\hat{k};\lambda,\beta) &= \tfrac{1}{\tilde{C}(\lambda,\beta)}\int_{\hat{k}}^{\infty}e^{\hat{y}^+(x)}x^n dx, \quad \text{when } \hat{k}\geq 0; \\ \hat{M}_p^{(n)}(\hat{k};\lambda,\beta) &= \tfrac{(-1)^n}{\tilde{C}(\lambda,\beta)}\int_{-\hat{k}}^{\infty}e^{\hat{y}^-(x)}x^n dx, \quad \text{when } \hat{k}<0. \end{aligned} \tag{3.23}$$

When $\hat{k}=0$, incomplete moment integral, $\hat{M}_c^n(0;\lambda,\beta)+\hat{M}_p^n(0;\lambda,\beta)$, becomes the moment integral in Eq. (2.30).

For small σ, since only the first and second terms matter, the approximated version of $M_{c,p}$ for $\lambda>2$ is

$$M_{c,p}(k;\lambda,\sigma,\beta,\mu)=e^{\mu}\left(\hat{M}_{c,p}^{(0)}(\hat{k})+\sigma\hat{M}_{c,p}^{(1)}(\hat{k})+\frac{\sigma^2}{2}\hat{M}_{c,p}^{(2)}(\hat{k})+...\right). \tag{3.24}$$

The advantage of incomplete moments is that the integral is independent of σ nor μ. Therefore, they can be pre-computed from unit distribution based on λ,β; and stored in a database for quick lookup. For $\lambda=3,\sigma=0.01$, the sum of first two orders produces accurate results with error less than 0.1%, except for very small values. In a compute-time sensitive environment, incomplete moment approach can be valuable.

On the other hand, the analytic form for $\beta=0$ can be extended to OGF. This provides solid baseline in an area where, due to its complexity, only the integral form exists otherwise.

3.5. Option Generating Function

The option generating function (OGF) is defined as

$$\begin{aligned}\mathbb{L}_c(k;\lambda,\sigma,\beta,\mu)&=\int_k^{\overrightarrow{\infty}}\left(e^x-e^k\right)P(x;\lambda,\sigma,\beta,\mu)\,dx, && \text{for call option;}\\ \mathbb{L}_p(k;\lambda,\sigma,\beta,\mu)&=-\int_{-\infty}^{k}\left(e^x-e^k\right)P(x;\lambda,\sigma,\beta,\mu)\,dx, && \text{for put option;}\end{aligned} \tag{3.25}$$

Generally speaking, $\mathbb{L}_{c,p}(k)$ can be implemented as

$$\begin{aligned}\mathbb{L}_c(k)&=M_c(k)-e^k\ (1-\Phi)\,(k),\\ \mathbb{L}_p(k)&=-M_p(k)+e^k\,\Phi(k).\end{aligned} \tag{3.26}$$

It is obvious that this leads to the put-call parity rule,

$$\mathbb{L}_c(k;\lambda,\sigma,\beta,\mu)-\mathbb{L}_p(k;\lambda,\sigma,\beta,\mu)=1+\Delta\mu-e^k. \tag{3.27}$$

where $\Delta\mu=M(t=1;\lambda,\sigma,\beta,\mu)-1=e^{\mu-\mu_D}-1$.

Since, in the local regime, analytic formula is expressed in terms of $\mathbb{L}_c$ on one side of cusp, and $\mathbb{L}_p$ for the other side of the cusp. We have to use put-call parity quite often to derive complete solutions for both option types.

$\mathbb{L}_{c,p}(k)$ is translationally invariant, by which μ in the distribution can be factored out:

$$\mathbb{L}_{c,p}(k;\lambda,\sigma,\beta,\mu)=e^{\mu}\mathbb{L}_{c,p}(k-\mu;\lambda,\sigma,\beta,\mu=0). \tag{3.28}$$

This invariant has an important implication: μ affects the absolute values of option prices, but it has no effect on the "shape" of $\log\mathbb{L}_{c,p}(...)$. The most relevant quantity is $\hat{k}=\frac{k-\mu}{\sigma}$, that is, the log-strike relative to μ, measured in

the unit of σ. And the best representation of option prices is in its logarithm form, $\log \mathbb{L}_{c,p}(...)$.

Here we see that a constant $\Delta\mu$ is added to the put-call parity when $\mu \neq \mu_D$. There is an uncertainty whether such constant should be added on the side of $k \geq \mu$ or $k < \mu$. Furthermore, the tail truncation also introduces an uncertainty on the true value of $\mathbb{L}_{c,p}$ in the local regime. As the market data indicates, the option prices should be padded with small constants in the local regime. In addition, the mathematics of volatility smile also dictates that there must two different μ's for call option chain and put option chain (See Eq. (4.30)). So the local (normalized) option prices should be calculated as

$$\begin{aligned} \text{call option prices} = &\quad \mathbb{L}_c(k;\lambda,\sigma,\beta,\mu_c) + \epsilon_c, \\ \text{put option prices} = &\quad \mathbb{L}_p(k;\lambda,\sigma,\beta,\mu_p) + \epsilon_p. \end{aligned} \tag{3.29}$$

Both ϵ_c and ϵ_p are very small, typically in the order of the smallest quotation unit of option prices, but nevertheless are necessary to account for small premiums that the market adds to the deeply out-of-money options. And for $\lambda \geq 2$, $\mathbb{L}_c$ and $\mathbb{L}_p$ have to be calculated separately for each side of the cusp. To put these concepts together, the (normalized) option prices in the local regime are determined by

$$\begin{aligned} \text{call option prices} &= \begin{cases} \mathbb{L}_c(k;\lambda,\sigma,\beta,\mu_c) + \epsilon_c, & \text{when } k-\mu_c \geq 0; \\ \mathbb{L}_p(k;\lambda,\sigma,\beta,\mu_c) + 1 - e^k + \Delta\mu_c + \epsilon_c, & \text{when } k-\mu_c < 0. \end{cases} \\ \text{put option prices} &= \begin{cases} \mathbb{L}_c(k;\lambda,\sigma,\beta,\mu_p) - 1 + e^k - \Delta\mu_p + \epsilon_p, & \text{when } k-\mu_p \geq 0; \\ \mathbb{L}_p(k;\lambda,\sigma,\beta,\mu_p) + \epsilon_p, & \text{when } k-\mu_p < 0. \end{cases} \end{aligned} \tag{3.30}$$

3.6. From OGF to Volatility Smile

To calculate implied volatilities, it is simply equating $\mathbb{L}_{c,p}$ in the local regime to $\mathbb{L}_{c,p}$ in the global regime. The high-level procedure is described in this section. The notation $\sigma_1(k)$ is used for the volatility curve with a suffix 1 to emphasize its global nature ($\lambda = 1$). If the time to maturity T is known, the quantity $\sigma_{imp}(k) = \sigma_1(k)/\sqrt{2T}$ is typically called the "implied volatility" in the industry. In this paper, we also call $\sigma_1(k)$ the "implied volatility" for simplicity sake.

When equating $\mathbb{L}_{c,p}$, we shall use risk neutral drift $\mu = \mu_{D1} = -\frac{\sigma_1^2(k)}{4}$ from Eq. (3.5) for the global regime. So from Eq. (3.30), we reach the high-level implied volatility formula based on OGF,

$$\begin{aligned} \mathbb{L}_c(k;\lambda=1,\sigma_1(k),\beta=0,\mu_{D1}) &= \begin{cases} \mathbb{L}_c(k;\lambda,\sigma,\beta,\mu_c) + \epsilon_c, & \text{when } k-\mu_c \geq 0; \\ \mathbb{L}_p(k;\lambda,\sigma,\beta,\mu_c) + 1 - e^k + \Delta\mu_c + \epsilon_c, & \text{when } k-\mu_c < 0. \end{cases} \\ \mathbb{L}_p(k;\lambda=1,\sigma_1(k),\beta=0,\mu_{D1}) &= \begin{cases} \mathbb{L}_c(k;\lambda,\sigma,\beta,\mu_p) - 1 + e^k - \Delta\mu_p + \epsilon_p, & \text{when } k-\mu_p \geq 0; \\ \mathbb{L}_p(k;\lambda,\sigma,\beta,\mu_p) + \epsilon_p, & \text{when } k-\mu_p < 0. \end{cases} \end{aligned} \tag{3.31}$$

Given the inputs - $\lambda, \sigma, \beta, \mu_c, \epsilon_c$ for calls, the implied volatility $\sigma_1(k)$ can be solved for each k by root-finding utility. Likewise, given $\lambda, \sigma, \beta, \mu_p, \epsilon_p$ for puts, another curve of $\sigma_1(k)$ can be solved. The operation to calculate $\sigma_1(k)$ from $\mathbb{L}_{c,p}$ is called the $\mathbb{V}_{c,p}$ operator. $\sigma_1(k)$ is also called "volatility smile" because the ATM region has smaller values and the further out k is from zero, the larger $\sigma_1(k)$.

3.7. Risk Neutrality

The risk neutral drift μ_D can be derived from the boundary condition on the moneyness of options. When the option is very deep in the money, the option price should approximately reflect that of the underlying (ignoring ϵ). This is the so-called "option delta equal to 1" condition. That is, in the normalized scale, $\mathbb{L}_c(k) \to 1$ when $k \to -\infty$. Since the CCDF term diminishes in Eq. (3.26), we have $M_c(k) \to 1$ when $k \to -\infty$, which is simply

$$M(t=1; \lambda, \sigma, \beta, \mu_D) = 1 \tag{3.32}$$

This is called the risk-neutral condition of MGF. Since the term $e^{t\mu}$ can be taken out of MGF (Eq. (3.2)), we have

$$e^{\mu_D} M(t=1; \lambda, \sigma, \beta, \mu=0) = 1. \tag{3.33}$$

This is equivalent to the definition of μ_D from Eq. (3.3). Therefore, the risk-neutral put-call parity is

$$\mathbb{L}_c(k; \lambda, \sigma, \beta, \mu_D) - \mathbb{L}_p(k; \lambda, \sigma, \beta, \mu_D) = 1 - e^k. \tag{3.34}$$

3.8. Solutions For Option Generating Function

For a λ distribution $|P(x; \lambda, \sigma, \beta, \mu)\rangle$ where x is log-return of underlying asset, $\mathbb{L}_{c,p}$ can be viewed as a (more complicated) Laplace-style operator that transform the probability from the x-domain (log-return) to the k-domain (log-strike):

$$L_{c,p}(k; \lambda, \sigma, \beta, \mu) = \mathbb{L}_{c,p} |P(x; \lambda, \sigma, \beta, \mu)\rangle . \tag{3.35}$$

The $\mathbb{L}_c$ operator is the first of the 4 operators in the λ transformation. There are several scenarios where analytic forms can be known. They are explored in this section.

3.8.1. Case I: $\lambda = 1$ and Global Regime Operators

For $\lambda = 1$, the risk-neutral $L_{c,p}(k)$ is the venerable Black-Scholes equation in our notation:

$$\begin{aligned} L_c(k; \lambda=1, \sigma_1, \mu_D) &= \tfrac{1}{2}\left(1 - e^k\right) - \tfrac{1}{2} \operatorname{erf}\left(\tfrac{k}{\sigma_1} - \tfrac{\sigma_1}{4}\right) + \tfrac{e^k}{2} \operatorname{erf}\left(\tfrac{k}{\sigma_1} + \tfrac{\sigma_1}{4}\right), \\ L_p(k; \lambda=1, \sigma_1, \mu_D) &= L_c(k; \lambda=1, \sigma_1, \mu_D) - \left(1 - e^k\right). \end{aligned} \tag{3.36}$$

There are two ways Eq. (3.36) will be used in the λ transformation framework. The first operator is to calculate normalized option prices from implied volatilities: the $\mathbb{O}_c$ operator for call, and the $\mathbb{O}_p$ for put. They take a vector of implied volatilities, $\sigma_1(k)$, and calculate the normalized option prices, $L_{c,p}(k)$,

$$\begin{aligned}\mathbb{O}_c:\ L_c(k) &= \tfrac{1}{2}\left(1-e^k\right) - \tfrac{1}{2}\,\mathrm{erf}\left(\tfrac{k}{\sigma_1(k)} - \tfrac{\sigma_1(k)}{4}\right) + \tfrac{e^k}{2}\,\mathrm{erf}\left(\tfrac{k}{\sigma_1(k)} + \tfrac{\sigma_1(k)}{4}\right),\\ \mathbb{O}_p:\ L_p(k) &= -\tfrac{1}{2}\left(1-e^k\right) - \tfrac{1}{2}\,\mathrm{erf}\left(\tfrac{k}{\sigma_1(k)} - \tfrac{\sigma_1(k)}{4}\right) + \tfrac{e^k}{2}\,\mathrm{erf}\left(\tfrac{k}{\sigma_1(k)} + \tfrac{\sigma_1(k)}{4}\right).\end{aligned} \tag{3.37}$$

The second operation is the $\mathbb{V}_{c,p}$ operator: to calculate implied volatilities, $\sigma_1(k)$, from a vector of normalized option prices, $L_c(k)$ or $L_p(k)$. This is a root-finding computation based on the same formula, Eq. (3.37). There are two possible sources of input to the $\mathbb{V}_{c,p}$ operator in λ transformation: The first source is the normalized option prices from market data. They are simply option prices divided by the price of the underlying, $\frac{\widetilde{\mathcal{C}}(k)}{\widetilde{S}}$ for call and $\frac{\widetilde{\mathcal{P}}(k)}{\widetilde{S}}$ for put. Most vendors would convert the option prices to $\sigma_{imp}(k)$, following a prescribed calendar convention for T. This is a very standard procedure[18]. Since there is no regime change here, I prefer to use $\mathbb{V}_{c,p}(\lambda=1)$ to label this operation.

The second kind of data source is $L_{c,p}(k)$ generated from a λ distribution in the local regime. Since there is a regime change, I would use $\mathbb{V}_{c,p}(\lambda\to 1)$ to label the source (λ) and the target (1) regimes. So taking $\mu_{c,p}$ and $\epsilon_{c,p}$ into account, we have

$$\sigma_1(k;\lambda,\sigma,\beta,\mu) = \mathbb{V}_{c,p}(\lambda\to 1)\cdot\left(\mathbb{L}_{c,p}+\epsilon_{c,p}\right)|P(x;\lambda,\sigma,\beta,\mu_{c,p})\rangle\,, \tag{3.38}$$

Since these two operators are inverse functions to each other, we have the identity rule:

$$\mathbb{O}_{c,p}\cdot\mathbb{V}_{c,p} = \mathbb{V}_{c,p}\cdot\mathbb{O}_{c,p} = \mathbb{I} \tag{3.39}$$

3.8.2. Case II: $\lambda = 2$

For $\lambda=2$, with the normalized $\hat{k}=\frac{k-\mu}{\sigma}$, the analytic form for each side of the cusp is

$$\begin{aligned}L_c(k;\lambda=2,\sigma,\beta,\mu) &= \tfrac{\sigma e^\mu}{2B_0}\left(\tfrac{1}{B_\sigma^- B^-}\right)e^{-B_\sigma^-\hat{k}}, &&\text{when } \hat{k}\geq 0;\\ L_p(k;\lambda=2,\sigma,\beta,\mu) &= \tfrac{\sigma e^\mu}{2B_0}\left(\tfrac{1}{B_\sigma^+ B^+}\right)e^{-B_\sigma^+|\hat{k}|}, &&\text{when } \hat{k}<0;\end{aligned} \tag{3.40}$$

and the put-call parity, $L_c(k)-L_p(k)=e^{\mu-\mu_D}-e^k$, is used to derive the other half[19]. We see that, to the first order of σ, option prices are proportional to σ. When $k=\mu$, we have

[18] The dividend yield and risk-free rate can be added to Eq. (3.36). However, for the purpose of this paper, they have very little effect on the short-maturity options. Especially in today's environment, the risk-free rate is nearly zero and SPX dividend yield is less than 2% per year.

[19] Here we use $\frac{1}{B_\sigma^\pm}-\frac{1}{B^\pm}=\frac{\mp\sigma}{B_\sigma^\pm B_\sigma^\pm}$.

$$\begin{aligned} L_c(k=\mu) &= \tfrac{\sigma e^{\mu}}{2B_0}\left(\tfrac{1}{B_\sigma^- B^-}\right), \\ L_p(k=\mu) &= \tfrac{\sigma e^{\mu}}{2B_0}\left(\tfrac{1}{B_\sigma^+ B^+}\right). \end{aligned} \tag{3.41}$$

Furthermore, when $\mu = \mu_D$,[20]

$$\begin{aligned} L_c(k=\mu_D) &= \tfrac{\sigma}{2B_0} B_\sigma^+ B^+, \\ L_p(k=\mu_D) &= \tfrac{\sigma}{2B_0} B_\sigma^- B^-. \end{aligned} \tag{3.42}$$

It follows that $L_c(k=\mu_D) + L_p(k=\mu_D) = \frac{\sigma}{2B_0}\left(2+\beta^2+\beta\sigma\right)$. When $\beta = 0$, $L_c(k=\mu_D) = L_p(k=\mu_D) = \frac{\sigma}{2}$.

3.8.3. Case III: Incomplete Moment Approach and $\beta = 0$

By replacing incomplete MGF and CDF with incomplete moments, $L_{c,p}$ can be expressed as,

$$\begin{aligned} L_c(k;\lambda,\sigma,\beta,\mu) &= \tfrac{e^{\mu}}{2}\left(\textstyle\sum_{n=1}^{n\to\infty} \frac{1}{n!}\, m_c^{(n)}(\hat{k}) + \left(1-e^{\sigma\hat{k}}\right) m_c^{(0)}(\hat{k})\right), && \text{when } \hat{k}\geq 0; \\ L_p(k;\lambda,\sigma,\beta,\mu) &= -\tfrac{e^{\mu}}{2}\left(\textstyle\sum_{n=1}^{n\to\infty} \frac{(-1)^n}{n!}\, m_c^{(n)}(\hat{k}) + \left(1-e^{\sigma\hat{k}}\right) m_p^{(0)}(\hat{k})\right), && \text{when } \hat{k}<0; \end{aligned} \tag{3.43}$$

where $m_{c,p}^{(n)}(\hat{k}) = 2\sigma^n\, \hat{M}_{c,p}^{(n)}(\hat{k};\lambda,\beta)$ from Eq. (3.23); and the put-call parity, $L_c(k) - L_p(k) = e^{\mu-\mu_D} - e^{k}$, is used to derive the other half on each side of the cusp. By expanding $\left(1-e^{\sigma\hat{k}}\right)$ into Taylor series, we show that, to the first order of σ, option prices are proportional to σ:

$$\begin{aligned} L_c(k;\lambda,\sigma,\beta,\mu) = {} & \sigma e^{\mu} \textstyle\sum_{n=1}^{n\to\infty} \frac{\sigma^{n-1}}{n!} \\ & \left(\hat{M}_c^{(n)}(\hat{k};\lambda,\beta) - \hat{k}^n\, \hat{M}_c^{(0)}(\hat{k};\lambda,\beta)\right), && \text{when } \hat{k}\geq 0; \\ L_p(k;\lambda,\sigma,\beta,\mu) = {} & \sigma e^{\mu} \textstyle\sum_{n=1}^{n\to\infty} \frac{(-\sigma)^{n-1}}{n!} \\ & \left(\hat{M}_p^{(n)}(\hat{k};\lambda,\beta) - \left|\hat{k}\right|^n\, \hat{M}_p^{(0)}(\hat{k};\lambda,\beta)\right), && \text{when } \hat{k}<0; \end{aligned} \tag{3.44}$$

This Taylor expansion result is the very reason that $L_{c,p}$ is called "generating function".

When $\beta = 0$, using Eq. (3.21), the option prices become sums of incomplete gamma functions,

[20] Here we use $B_\sigma^- B_\sigma^+ = e^{\mu_D}$; $B_\sigma^- B^+ = 1-\sigma B^+$ and $B_\sigma^+ B^- = 1+\sigma B^-$.

$$
\begin{aligned}
L_c(k;\lambda,\sigma,\beta=0,\mu) = & \\
& \frac{\sigma e^{\mu}}{2\Gamma\left(\frac{\lambda}{2}\right)} \sum_{n=1}^{n_{\overrightarrow{\infty}}} \frac{\sigma^{n-1}}{n!} \\
& \left(\Gamma\left(\frac{\lambda(n+1)}{2}, \left|\hat{k}\right|^{\frac{2}{\lambda}}\right) - \hat{k}^n \, \Gamma\left(\frac{\lambda}{2}, \left|\hat{k}\right|^{\frac{2}{\lambda}}\right)\right), \qquad \text{when } \hat{k} \geq 0; \\
L_p(k;\lambda,\sigma,\beta=0,\mu_D) = & \\
& \frac{\sigma e^{\mu}}{2\Gamma\left(\frac{\lambda}{2}\right)} \sum_{n=1}^{n_{\overrightarrow{\infty}}} \frac{(-\sigma)^{n-1}}{n!} \\
& \left(\Gamma\left(\frac{\lambda(n+1)}{2}, \left|\hat{k}\right|^{\frac{2}{\lambda}}\right) - \left|\hat{k}\right|^{n} \, \Gamma\left(\frac{\lambda}{2}, \left|\hat{k}\right|^{\frac{2}{\lambda}}\right)\right), \qquad \text{when } \hat{k} < 0;
\end{aligned}
\tag{3.45}
$$

The symmetric sum-of-incomplete-gamma expression of option prices provides good baseline for option prices across all λ. It also sheds some light into the formation of option prices: Call option is the sum of all "moment" terms regardless of odd n or even n; while put option is the sum of the differences between even "moment" terms and odd "moment" terms.

To illustrate this, we can calculate $L_{c,p}$ at $k=\mu$. Since $\Gamma(s,0)=\Gamma(s)$, we arrive at particularly simple expressions,

$$
\begin{aligned}
L_c(k=\mu;\lambda,\sigma,\beta=0,\mu) &= \frac{\sigma e^{\mu}}{2}\left(\sum_{n=1}^{n_{\overrightarrow{\infty}}} \frac{\sigma^{n-1}}{n!}\, \Gamma\left(\frac{\lambda(n+1)}{2}\right) / \Gamma\left(\frac{\lambda}{2}\right)\right), \\
L_p(k=\mu;\lambda,\sigma,\beta=0,\mu) &= \frac{\sigma e^{\mu}}{2}\left(\sum_{n=1}^{n_{\overrightarrow{\infty}}} \frac{(-\sigma)^{n-1}}{n!} \Gamma\left(\frac{\lambda(n+1)}{2}\right) / \Gamma\left(\frac{\lambda}{2}\right)\right).
\end{aligned}
\tag{3.46}
$$

Notice the similarity to Eq. (2.8). $L_c(k=\mu)$ and $L_p(k=\mu)$ are just taking different segments of symmetric moments in the series summation of MGF. And the difference, $L_c - L_p$, comes back to $e^{\mu-\mu_D} - e^{\mu}$ by the very definition of $M(t=1;\mu=0) = e^{-\mu_D}$.

3.8.4. Case III: $\lambda > 2$

There is no known analytic solution for $\lambda > 2$ except when $\beta = 0$. We resort to the integral form of OGF for each side of the cusp, based on unit distribution,

$$
\begin{aligned}
L_c(k;\lambda,\sigma,\beta,\mu) &= \frac{e^{\mu}}{\hat{C}(\lambda,\beta)} \int_{\hat{k}}^{\overrightarrow{\infty}} e^{\hat{y}^+(x)} \left(e^{\sigma x} - e^{\sigma\hat{k}}\right) dx, \qquad \text{when } \hat{k} \geq 0; \\
L_p(k;\lambda,\sigma,\beta,\mu) &= \frac{-e^{\mu}}{\hat{C}(\lambda,\beta)} \int_{-\hat{k}}^{\infty} e^{\hat{y}^-(x)} \left(e^{-\sigma x} - e^{\sigma\hat{k}}\right) dx, \qquad \text{when } \hat{k} < 0;
\end{aligned}
\tag{3.47}
$$

and use put-call parity, $L_c(k) - L_p(k) = e^{\mu-\mu_D} - e^{k}$, to derive the other half. This formula provides the most general approach to calculate option prices. However, one must understand how to estimate numerical error in such direct integral. In particular, for short-maturity options, σ can be as small as 0.001. In R, the `integrate` function has a default tolerance of `.Machine$double.eps^0.25`, which is about 0.0001. As you can see, there is not much room for error. This is one major reason that I use the unit distribution extensively and attempt to factor out μ and σ as much as possible in this paper.

4. Small σ Limit of Option Prices and Volatility Smile

I've developed several ways to calculate $L_{c,p}(k)$. As it has been emphasized repeatedly in this paper, the short-maturity options have particularly simple behavior. This area corresponds to $\sigma \ll 1$, $|\mu_D| \ll 1$. This section is dedicated to solutions under such condition. One major achievement here is that the analytic formula for volatility smile is derived.

4.1. The $\mathbb{V}$ Operator for Small σ and Small k

Small k corresponds to the neighborhood of ATM options. Eq. (3.36) can be simplified when $L_{c,p}(k)$ is small and $\left(\frac{k}{\sigma_1(k)} \pm \frac{\sigma_1(k)}{4}\right)$ is small, that is, $\sigma, \sigma_1(k) \ll 1$ and $k \ll \sigma_1(k)$. By expanding the erf function at zero[21] and keep the linear terms, we have (ϵ is ignored here)

$$L_{c,p}(k) = \frac{\sigma_1(k)}{2\sqrt{\pi}} \mp \frac{k}{2}. \tag{4.1}$$

Rearranging the terms, we get the first order approximation of the $\mathbb{V}$ operator for small k,

$$\mathbb{V}_{c,p}(\lambda \to 1): \ \sigma_1(k) = 2\sqrt{\pi}\left(L_{c,p}(k) \pm \frac{k}{2}\right). \tag{4.2}$$

This approximation works only for a very small range of k near μ, nevertheless, it is a (somewhat exaggerated) illustration of the volatility smile. Take call option as example, as k increases on the positive side, L_c decreases exponentially and $k/2$ term takes over to move the volatility $\sigma_1(k)$ up. As k decreases on the negative side, L_c is dominated by $e^{\mu-\mu_D} - e^k$, which also moves $\sigma_1(k)$ up. Therefore, a smile is formed with the local minimum near $k = \mu \approx 0$: $\min(\sigma_1) = 2\sqrt{\pi}L_{c,p}(k = \mu)$.

4.2. OGF for Small σ

From both the incomplete moment expansion, Eq. (3.43), and the integral form, Eq. (3.47), $L_{c,p}$'s first order approximation in σ is

$$\begin{aligned} L_c(k;\lambda,\sigma,\beta,\mu) &= \sigma e^{\mu}\left(\hat{M}_c^{(1)}(\hat{k}) - \hat{k}\,\hat{M}_c^{(0)}(\hat{k})\right), \quad \text{when } \hat{k} \geq 0; \\ L_p(k;\lambda,\sigma,\beta,\mu) &= -\sigma e^{\mu}\left(\hat{M}_p^{(1)}(\hat{k}) - \hat{k}\,\hat{M}_p^{(0)}(\hat{k})\right), \quad \text{when } \hat{k} < 0. \end{aligned} \tag{4.3}$$

At $\lambda = 3$ and $\sigma = 0.001$, the error of this approximation is less than 1% even for $\left|\hat{k}\right|$ as large as 5. The error becomes smaller for smaller $\left|\hat{k}\right|$. At $k = \mu$, $L_{c,p}$ is

[21] $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\left(z - \frac{1}{3}z^3 + ...\right)$. See https://en.wikipedia.org/wiki/Error_function

just the first-order incomplete moments, integrating the entire tail on each side of the cusp,

$$\begin{aligned} L_c(k=\mu) &= \sigma e^{\mu}\hat{M}_c^{(1)}(0) = \tfrac{\sigma\, e^{\mu}}{\tilde{C}(\lambda,\beta)} \int_0^{\infty} e^{\hat{y}^+(x)} x\, dx \\ L_p(k=\mu) &= -\sigma e^{\mu}\hat{M}_p^{(1)}(0) = \tfrac{\sigma\, e^{\mu}}{\tilde{C}(\lambda,\beta)} \int_0^{\infty} e^{\hat{y}^-(x)} x\, dx \end{aligned} \tag{4.4}$$

At $\lambda = 3$ and $\sigma = 0.001$, the error of this approximation is less than 0.4%.

For small σ and $\beta = 0$, we define the star OGF $L^*_\lambda(\hat{k})$, which is independent of σ and μ,

$$L^*_\lambda(\hat{k}) = \frac{1}{2\Gamma\left(\frac{\lambda}{2}\right)} \left(\Gamma\left(\lambda, \left|\hat{k}\right|^{\frac{2}{\lambda}}\right) - \left|\hat{k}\right| \Gamma\left(\frac{\lambda}{2}, \left|\hat{k}\right|^{\frac{2}{\lambda}}\right)\right), \tag{4.5}$$

and the OGF reaches its simplest analytic form (non-zero $\Delta\mu_{c,p}$ and $\epsilon_{c,p}$ only for the local regime),

$$\begin{aligned} L_c\left(k;\lambda,\sigma,\beta=0,\mu_c\right) + \epsilon_c = & \begin{cases} \sigma e^{\mu_c} L^*_\lambda(\hat{k}) + \epsilon_c, & \text{when } \hat{k} \geq 0; \\ \sigma e^{\mu_c} L^*_\lambda(\hat{k}) + \left(1 - e^{k}\right) + \Delta\mu_c + \epsilon_c, & \text{when } \hat{k} < 0; \end{cases} \\ L_p\left(k;\lambda,\sigma,\beta=0,\mu_p\right) + \epsilon_p = & \begin{cases} \sigma e^{\mu_p} L^*_\lambda(\hat{k}) - \left(1 - e^{k}\right) - \Delta\mu_p + \epsilon_p, & \text{when } \hat{k} \geq 0; \\ \sigma e^{\mu_p} L^*_\lambda(\hat{k}) + \epsilon_p, & \text{when } \hat{k} < 0; \end{cases} \end{aligned} \tag{4.6}$$

where $\Delta\mu_{c,p} = e^{\mu_{c,p} - \mu_D} - 1$. Note that $L^*_\lambda(\hat{k})$ is symmetric on $\hat{k}$.

Eq. (4.5) allows us to simplify the $\mathbb{V}_{c,p}\,(\lambda \to 1)\cdot\mathbb{L}_{c,p}$ operation into a one-line formula which can be applied with the root-finding utility to solve $\sigma_1(k)$. For $\lambda = 1$, we must use $\hat{k}_1 = \frac{k-\mu_{D1}}{\sigma_1(k)} = \frac{k}{\sigma_1(k)} + \frac{\sigma_1(k)}{4}$. So from $L_{c,p}$ in Eq. (4.6), we reach the volatility smile formula based on incomplete gamma function,

$$\sigma_1(k)\, L^*_1(\hat{k}_1) = \sigma L^*_\lambda(\hat{k}) + \delta_{c,p}(\hat{k})\, \Delta\mu_{c,p} + \epsilon_{c,p}, \tag{4.7}$$

where

$$\begin{aligned} \hat{k}_1 &= \tfrac{\sigma\,\hat{k}}{\sigma_1(k)} + \left(\tfrac{\mu_{c,p}}{\sigma_1(k)} + \tfrac{\sigma_1(k)}{4}\right), \\ \delta_c(\hat{k}) &= \delta(\hat{k} < 0), \\ \delta_p(\hat{k}) &= -\delta(\hat{k} \geq 0), \\ \hat{k} &= \tfrac{k-\mu_{c,p}}{\sigma}, \end{aligned} \tag{4.8}$$

and $\delta(x)$ is the indicator function: 1 when x is true, else 0. Therefore, $\delta_{c,p}(\hat{k})$ is the indicator function for in-the-money option (with a negative sign for put). Notice that the e^{μ} term in $\sigma e^{\mu} L^*$ is dropped on both sides since it is basically 1 for small σ. For reasonably large $\left|\hat{k}\right|$, the $\frac{\sigma_1(k)}{4}$ term in $\hat{k}_1$ can also be dropped. The term is retained here for the sake of analytic tractability.

Now we choose to express the volatility smile in terms of the dimensionless ratio, $Q(\hat{k}) = \frac{\sigma_1(k)}{\sigma}$, by which we can replace $\hat{k}_1$ with $\frac{1}{Q(\hat{k})}\left(\hat{k} + \frac{\mu_{c,p}}{\sigma}\right)$. Eq. (4.7) becomes

$$Q(\hat{k})\, L^*_1\left(\tfrac{1}{Q(\hat{k})}\left(\hat{k} + \tfrac{\mu_{c,p}}{\sigma}\right)\right) = L^*_\lambda(\hat{k}) + \left(\delta_{c,p}(\hat{k})\Delta\mu_{c,p} + \epsilon_{c,p}\right)\tfrac{1}{\sigma} \tag{4.9}$$

in which $Q(\hat{k})$ can be solved by the root-finding utility, for each $\hat{k}$ at a given λ. Since $L^*_\lambda(\hat{k})$ is symmetric over $\hat{k}$, the volatility smile is symmetric if and only if $\mu_{c,p} = \mu_D$. On the other hand, when $\mu_{c,p} \neq \mu_D$, the risk neutrality is violated; and the volatility smile is skewed, which is observed very commonly in the market data and will be studied more extensively in Section 4.7.

4.3. Exact Solutions of L^* for Integer λ

For integer $\lambda = 1, 2, 3$, $L^*_\lambda(x)$ have exact solutions. They shall be mentioned here before we dive into more complex expansions and asymptotic series. For the global regime, it is

$$L^*_1(x) = \frac{1}{2\sqrt{\pi}}\, e^{-x^2} - \frac{x}{2}\, \mathrm{erfc}\,(x)\,,\ x \geq 0. \tag{4.10}$$

For $\lambda = 2$, as it has always been simple, it is

$$L^*_2(x) = \frac{1}{2} e^{-x},\ x \geq 0. \tag{4.11}$$

For $\lambda = 3$, the elliptic solution is

$$L^*_3(x) = \frac{2}{\sqrt{\pi}}\, e^{-x^{\frac{2}{3}}} \left(1 + x^{\frac{2}{3}}\right) - \frac{x}{2}\, \mathrm{erfc}\left(x^{\frac{1}{3}}\right),\ x \geq 0. \tag{4.12}$$

They can be derived via the recurrence relations of incomplete gamma function (See Eq. (6.5.22) of [1]) and terminated at $\Gamma\,(1, x) = e^{-x}$ and $\Gamma\left(\frac{1}{2}, x\right) = \sqrt{\pi}\,\mathrm{erfc}\,(\sqrt{x})$. The logarithm of $L^*_2(x)$ is precisely linear; and, for $\lambda > 2$, the logarithm of $L^*_\lambda(x)$ is approximately linear.

The essense of calculating volatility smile in Eq. (4.7) is an exercise of estimating $\hat{k}_1$ in the logarithm of $\sigma_1(k) L^*_1(\hat{k}_1)$ that matches the quantity of $\sigma L^*_\lambda(\hat{k})$ from the local regime.

4.4. Minimum Implied Volatility in the Smile

The minimum implied volatility in the valley of the volatility smile, near $k = 0$, provides an important anchor for the curve. Since it is related to the ATM options, it is also in the most heavily traded range of k. When $k = \mu = \mu_D$, the option prices are simply ($\epsilon_{c,p}$ is negligible here)

$$L_{c,p}(k = \mu; \lambda, \sigma, \beta = 0, \mu) \to \frac{\sigma e^{\mu}}{2}\, \frac{\Gamma\,(\lambda)}{\Gamma\left(\frac{\lambda}{2}\right)}. \tag{4.13}$$

For very small σ and $\mu, \mu_D \ll \sigma$, from Eq. (4.13), we have

$$\min\,(\sigma_1(k)) = \sigma_1(\mu_D) = \sqrt{\pi}\,\sigma\, \frac{\Gamma\,(\lambda)}{\Gamma\left(\frac{\lambda}{2}\right)}. \tag{4.14}$$

When $\lambda = 2$, it is $\min(\sigma_1(k)) = \sqrt{\pi}\,\sigma$; when $\lambda = 3$, it becomes $\min(\sigma_1(k)) = 4\sigma$. That is, $\min Q(\hat{k}) = \sqrt{\pi}$ and 4 respectively. These are very elegant results.

However, this still ties too much into our model. We can push further for a more "model-free" result. Replacing $\sigma_1(k)$ with $\sigma_{imp}(k)\sqrt{2T}$ and σ with the variance from Eq. (2.34) , we get an extended relation between minimum implied volatility and the standard deviation of the underlying distribution applicable to all $\lambda \geq 2$ and $|\beta| \leq 0.8$,

$$\frac{\min(\sigma_{imp}(k))\sqrt{T}}{\sqrt{\text{var}}} \approx \sqrt{\frac{\pi}{2}}\,\frac{\Gamma(\lambda)}{\sqrt{\Gamma\left(\frac{\lambda}{2}\right)\Gamma\left(\frac{3\lambda}{2}\right)}}\,\frac{1}{\sqrt{1+\frac{1}{\Gamma_2(\lambda)}\beta^2}}, \tag{4.15}$$

where $\Gamma_2(\lambda) = \frac{\Gamma\left(\frac{3\lambda}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)}$. This formula has less than 1% numerical error when $\sigma = 0.001$. Since we are most concerned about $\lambda \approx 3$, we can use the following nearly "model-free" formula

$$\frac{\min(\sigma_{imp}(k))\sqrt{T}}{\sqrt{\text{var}}} \approx \frac{0.781}{\sqrt{1+0.0762\,\beta^2}},\ \text{near } \lambda \approx 3. \tag{4.16}$$

We found that, intuitively speaking, $\min(\sigma_{imp}(k))\sqrt{T}$ is almost equivalent to the measurement of 1-stdev move by the market for the duration of time T. This result gives important meaning to what the minimum implied volatility is.

4.5. Taylor Expansion of L^* at Small $\hat{k}$

The L^* function is the difference of two incomplete gamma functions. From Eq. (6.5.29) of [1], the limiting gamma function can be expanded as

$$\gamma^*(s,x) = e^{-x}\sum_{n=0}^{\infty}\frac{x^n}{\Gamma(s+n+1)}, \tag{4.17}$$

by which we have

$$\Gamma(s,x) = \Gamma(s)\left(1 - x^s\gamma^*(s,x)\right). \tag{4.18}$$

Since $L^*_\lambda(\hat{k})$ is symmetric on $\hat{k}$, we will only consider the case of $\hat{k} \geq 0$. Then L^* can be represented by γ^*,

$$L^*_\lambda(\hat{k}) = \frac{\Gamma(\lambda)}{2\Gamma\left(\frac{\lambda}{2}\right)} - \frac{1}{2}\hat{k} + \frac{1}{2}\hat{k}^2\left(\gamma^*\left(\frac{\lambda}{2}, \hat{k}^{\frac{2}{\lambda}}\right) - \frac{\Gamma(\lambda)}{\Gamma\left(\frac{\lambda}{2}\right)}\gamma^*\left(\lambda, \hat{k}^{\frac{2}{\lambda}}\right)\right), \tag{4.19}$$

Using the power expansion of γ^*, we get

$$L^*_\lambda(\hat{k}) = \frac{\Gamma(\lambda)}{2\Gamma\left(\frac{\lambda}{2}\right)} - \frac{\hat{k}}{2} + \frac{\hat{k}^2}{2}\,e^{-\hat{k}^{\frac{2}{\lambda}}}\sum_{n=0}^{\infty}\hat{k}^{\frac{2n}{\lambda}}\left(\frac{1}{\Gamma\left(\frac{\lambda}{2}+n+1\right)} - \frac{\Gamma(\lambda)}{\Gamma\left(\frac{\lambda}{2}\right)\Gamma(\lambda+n+1)}\right), \tag{4.20}$$

It is obvious that $L^*_\lambda(\hat{k})$ starts with the constant $\frac{\Gamma(\lambda)}{2\Gamma\left(\frac{\lambda}{2}\right)}$ at $\hat{k}=0$, which is associated with $\min(\sigma_1(k))$ in Eq. (4.14). Then $L^*_\lambda(\hat{k})$ decreases as $\hat{k}$ increases because of the negative $\frac{\hat{k}}{2}$ term. But $L^*_\lambda(\hat{k})$ is always positive - it must be because it represents a form of price. To understand $L^*_\lambda(\hat{k})$ at large $\hat{k}$ requires hyergeometric expansion in the following section.

4.6. Hypergeometric Asymptotics of L^* at Large $\hat{k}$

The asymptotics of incomplete gamma function allows us to study the inner workings of volatility smile in Eq. (4.7). $\Gamma(s,x)$ can be expanded asymptotically via Tricomi confluent hypergeometric function $U(a,b,x)$ and generalized hypergeometric series ${}_2F_0$,

$$\begin{aligned}\Gamma(s,x) \approx & \quad e^{-x}x^{s-1}\,{}_2F_0\left(1,1-s;;-\tfrac{1}{x}\right)\\ &= e^{-x}x^{s-1}\left(1+\tfrac{s-1}{x}+\tfrac{(s-1)(s-2)}{x^2}+\tfrac{(s-1)(s-2)(s-3)}{x^3}+...\right)\\ &= e^{-x}x^{s-1}\textstyle\sum_{n=0}^{N}\frac{(s-1)_{(n)}}{x^n},\end{aligned} \tag{4.21}$$

where $(s-1)_{(n)}$ is the Pochhammer symbol for falling factorial[22]. This result is stated in Eq. (6.5.31) of [1] [23].

When s is a positive integer, the series terminates on $n=s$. So we have

$$\begin{aligned}\Gamma(s=1,x) \approx & \quad e^{-x};\\ \Gamma(s=2,x) \approx & \quad e^{-x}x\left(1+\tfrac{1}{x}\right);\\ \Gamma(s=3,x) \approx & \quad e^{-x}x^2\left(1+\tfrac{2}{x}+\tfrac{2}{x^2}\right).\end{aligned} \tag{4.22}$$

However, in our case, s can be half integer $\left(\frac{1}{2},\frac{3}{2}\right)$ and the series won't terminate [24]. These infinite ${}_2F_0$ series have zero radius of convergence. So the terms in ${}_2F_0$ must be cut off, which is what N stands for. When x is large enough, only a few terms are needed to obtain a good approximation. Since s is at most 3 for our purpose, we shouldn't go too far beyond x^{-3}.

The expansion of $L^*_\lambda(\hat{k})$ is performed in terms of the inverse of $\hat{k}^{\frac{2}{\lambda}}$,

$$L^*_\lambda(\hat{k}) = \frac{e^{-\hat{k}^{\frac{2}{\lambda}}}\,\hat{k}^{2-\frac{2}{\lambda}}}{2\Gamma\left(\frac{\lambda}{2}\right)}\left({}_2F_0\left(1,1-\lambda;;-\hat{k}^{-\frac{2}{\lambda}}\right)-{}_2F_0\left(1,1-\frac{\lambda}{2};;-\hat{k}^{-\frac{2}{\lambda}}\right)\right), \tag{4.24}$$

[22] See https://en.wikipedia.org/wiki/Pochhammer_symbol

[23] It comes from combining (a) $\Gamma(s,x)=e^{-x}x^sU(1,1+s,x)$, which can be derived from Eq. (13.6.28) $\Gamma(s,x)=e^{-x}U(1-s,1-s,x)$ and Eq. (13.1.29) $U(a,b,x)=x^{1-b}U(1+a-b,2-b,z)$ of [1]; and (b) $U(a,b,x)\approx x^{-a}\,{}_2F_0\left(a,a-b+1;;-\frac{1}{x}\right)$ from Eq. (13.5.2) of [1].

See also https://en.wikipedia.org/wiki/Incomplete_gamma_function and https://en.wikipedia.org/wiki/Generalized_hypergeometric_function#Terminology

[24] The hypergeometric expansion at $s=\frac{1}{2}$ is identical to the asymptotic expansion of the error function whereby

$$\Gamma\left(\frac{1}{2},x\right)=\sqrt{\pi}\,\mathrm{erfc}\left(\sqrt{x}\right)=e^{-x}x^{-\frac{1}{2}}\sum_{n=0}^{N}(-1)^n\frac{(2n-1)!!}{(2x)^n}. \tag{4.23}$$

See https://en.wikipedia.org/wiki/Error_function#Asymptotic_expansion

by which we get

$$L^*_\lambda(\hat{k}) = \frac{\lambda}{4\Gamma\left(\frac{\lambda}{2}\right)} e^{-\hat{k}^{\frac{2}{\lambda}}}\, \hat{k}^{2-\frac{4}{\lambda}} \left(1 + \frac{3}{2}\left(\lambda - 2\right)\hat{k}^{-\frac{2}{\lambda}} + \frac{1}{4}\left(7\lambda^2 - 36\lambda + 44\right)\hat{k}^{-\frac{4}{\lambda}} + ...\right), \hat{k} \gg 1. \tag{4.25}$$

This result is most useful for the local regime. At $\lambda = 3$, $\left|\hat{k}\right|$ has to be greater than 2 to avoid the singularity at $\hat{k} = 0$. We see clearly that $L^*_\lambda(\hat{k})$ decays exponentially as $\hat{k}$ increases. And notice that the leading term of $L^*_\lambda(\hat{k})$ becomes $\frac{1}{4\sqrt{\pi}\,\hat{k}^2}\, e^{-\hat{k}^2}$ at $\lambda = 1$; and $\frac{1}{2}\, e^{-\hat{k}}$ at $\lambda = 2$ (which is exactly that); and $\frac{3}{2\sqrt{\pi}}\, e^{-\hat{k}^{\frac{2}{3}}}\, \hat{k}^{\frac{2}{3}}$ at $\lambda = 3$.

4.7. Volatility Smile

With the help of hypergeometric expansion, the volatility smile formula, Eq. (4.7), becomes

$$\begin{aligned} Q\, L^*_1(\tfrac{\hat{k}}{Q}) &= \tfrac{Q}{2\sqrt{\pi}} \left(e^{-\left|\frac{\hat{k}}{Q}\right|^2} - \sqrt{\pi}\left|\tfrac{\hat{k}}{Q}\right| \operatorname{erfc}\left(\left|\tfrac{\hat{k}}{Q}\right|\right)\right) \\ &= \left(\delta_{c,p}(\hat{k})\Delta\mu_{c,p} + \epsilon_{c,p}\right)\tfrac{1}{\sigma} + \\ \tfrac{\lambda}{4\Gamma\left(\frac{\lambda}{2}\right)}\, e^{-\hat{k}^{\frac{2}{\lambda}}}\, \hat{k}^{2-\frac{4}{\lambda}} &\left(1 + \tfrac{3}{2}\left(\lambda - 2\right)\hat{k}^{-\frac{2}{\lambda}} + \tfrac{1}{4}\left(7\lambda^2 - 36\lambda + 44\right)\hat{k}^{-\frac{4}{\lambda}} + ...\right), \left|\hat{k}\right| \geq 2, \end{aligned} \tag{4.26}$$

in which $Q = Q(\hat{k})$ can be solved by the root-finding utility, for each $\hat{k}$ at a given λ. The precision of this formula is quite high for $\left|\hat{k}\right| \geq 2$. For the local regime (on the right-hand side), due to the convergence issue, only the leading term and the second order term ($\hat{k}^{-\frac{4}{\lambda}}$) estimates produce meaningful results.[25]

For large positive $\hat{k}$ where $\hat{k}^{\frac{2}{\lambda}} - \left(2 - \frac{4}{\lambda}\right)\log\hat{k} > 8$, we can use asymptotic hypergeometric expansion for the global regime as well, which yields

$$\begin{aligned} &\tfrac{Q^3}{\hat{k}^2}\, e^{-\left(\frac{\hat{k}}{R}\right)^2} \left(1 - \tfrac{3}{2}\left(\tfrac{Q}{\hat{k}}\right)^2 + \tfrac{15}{4}\left(\tfrac{Q}{\hat{k}}\right)^4 + ...\right) = \left(\delta_{c,p}(\hat{k})\Delta\mu_{c,p} + \epsilon_{c,p}\right)\tfrac{1}{\sigma} + \\ &\tfrac{\sqrt{\pi}\lambda}{\Gamma\left(\frac{\lambda}{2}\right)}\, e^{-\hat{k}^{\frac{2}{\lambda}}}\, \hat{k}^{2-\frac{4}{\lambda}} \left(1 + \tfrac{3}{2}\left(\lambda - 2\right)\hat{k}^{-\frac{2}{\lambda}} + \tfrac{1}{4}\left(7\lambda^2 - 36\lambda + 44\right)\hat{k}^{-\frac{4}{\lambda}} + ...\right). \end{aligned} \tag{4.27}$$

When $\delta_{c,p}(\hat{k})$ is zero, it corresponds to the trading ranges for out-of-money options. If the entire constant term $\left(\delta_{c,p}(\hat{k})\Delta\mu_{c,p} + \epsilon_{c,p}\right)$ is zero, we only have the $e^{-\hat{k}^{\frac{2}{\lambda}}}\, \hat{k}^{4-\frac{4}{\lambda}}$ term on the right hand side of Eq. (4.26), which determines the asymptotic behavior of $Q(\hat{k})$ at large $\left|\hat{k}\right|$. The volatility smile is curved. By taking logarithm, retaining the exponential terms and the constant $D =$

[25] For this reason, the next term, $\frac{1}{8}\left(15\lambda^3 - 140\lambda^2 + 420\lambda - 400\right)\hat{k}^{-\frac{6}{\lambda}}$, is not useful for our purpose.

$\log\left(\frac{\sqrt{\pi}\lambda}{\Gamma\left(\frac{\lambda}{2}\right)}\right)$, and dropping the rest of the terms, we reach the elegant result of volatility smile for very large out-of-money $\hat{k}$,

$$\frac{\sigma_1(k)}{\sigma} = Q(\hat{k}) = \left|\hat{k}\right| \left(\hat{k}^{\frac{2}{\lambda}} - D\right)^{-\frac{1}{2}}. \tag{4.28}$$

Use the 1-day fit of SPX as an example, at $\hat{k} = 40$, $Q \approx 12.71$ is within 0.3% error. Notice that $Q(\hat{k}) \to \hat{k}^{1-\frac{1}{\lambda}}$ as $\hat{k} \to \infty$. This indicates the slope $\frac{d\log Q(\hat{k})}{d\log \hat{k}} \to 1 - \frac{1}{\lambda}$ as $\hat{k} \to \infty$. However, when the constant term $\left(\delta_{c,p}(\hat{k})\Delta\mu_{c,p} + \epsilon_{c,p}\right)$ is non-zero, it will eventually dominate the $e^{-\hat{k}^{\frac{2}{\lambda}}}\hat{k}^{4-\frac{4}{\lambda}}$ term. This leads to the discussion of volatility skew.

4.8. Volatility Skew

When $\mu \neq \mu_D$, volatility smile is morphed into volatility skew. This condition is called "violation of risk neutrality". In addition, $\epsilon_{c,p}$ can also be non-zero. When either $\delta_{c,p}(\hat{k})$ or $\epsilon_{c,p}$ is not zero, the dynamics of a volatility skew is very different from what has been described above. $\delta_{c,p}(\hat{k}) \neq 0$ corresponds to the in-the-money options; while $\epsilon_{c,p} \neq 0$ changes the smile behavior of the deeply out-of-money options. In both cases, at large $\left|\hat{k}\right|$, the constant term $\left(\delta_{c,p}(\hat{k})\Delta\mu_{c,p} + \epsilon_{c,p}\right)$ eventually dominates the right hand side of Eq. (4.26) since the $e^{-\hat{k}^{\frac{2}{\lambda}}}\hat{k}^{4-\frac{4}{\lambda}}$ term diminishes exponentially,

$$\frac{Q}{2\sqrt{\pi}}\left(e^{-\left|\frac{\hat{k}}{Q}\right|^2} - \sqrt{\pi}\left|\frac{\hat{k}}{Q}\right| \operatorname{erfc}\left(\left|\frac{\hat{k}}{Q}\right|\right)\right) = \left(\delta_{c,p}(\hat{k})\Delta\mu_{c,p} + \epsilon_{c,p}\right)\frac{1}{\sigma},\ \left|\hat{k}\right| \gg 1. \tag{4.29}$$

This produces nearly linear volatility curves asymptotically at large $\left|\hat{k}\right|$. But $\epsilon_{c,p}$ itself doesn't produce volatility skew. The skew is generated by $\delta_{c,p}(\hat{k})\Delta\mu_{c,p}$. Furthermore, notice that there is no explicit dependency on λ. This is quite amazing. The in-the-money skew is all about how much risk-neutral violation there is: $\frac{\Delta\mu_{c,p}+\epsilon_{c,p}}{\sigma} \approx \frac{\mu_{c,p}-\mu_D+\epsilon_{c,p}}{\sigma}$. The out-of-money smile comes from the minimum risk premium $\frac{\epsilon_{c,p}}{\sigma}$. Not only so, it is also required by mathematics that $\delta_{c,p}(\hat{k})\frac{\Delta\mu_{c,p}}{\sigma} \geq 0$ (assume $\epsilon_{c,p}$ is much smaller than $\Delta\mu_{c,p}$). This indicates $(\mu_{c,p} - \mu_D)$ must have the same sign as $\delta_{c,p}(\hat{k})$. Therefore, we have

$$\begin{aligned} &\mu_c \geq \mu_D, \text{ for call option,} \\ &\mu_p \leq \mu_D, \text{ for put option.} \end{aligned} \tag{4.30}$$

The fact that μ is directional and there are two different μ's for calls and puts is intriguing. The mathematics dictates that the steeper slope of a volatility skew is always on the side of in-the-money options. It appears that the call-option market expects more upward drift; while the put-option market expects more

downward drift. More premium is required to insure the in-the-money options. In the case of SPX options, this is more so for in-the-money call options than put options. Is this psychology? Or there is a market dynamics driving it? I don't have a deeper explanation as to what is the cause behind it.

The quantity, $\frac{\Delta\mu}{\sigma} \approx \frac{\mu-\mu_D}{\sigma}$, determines the magnitude of the volatility skew. Now if the distribution is symmetric, $\mu_D \sim m_{(2)} \sim \sigma^2$, this order of magnitude is too small to produce observable volatility skew. Only when $\mu-\mu_D \sim m_{(1)} \sim \sigma\beta$, it is large enough to be observable. This insight is very important for the role of the skew parameter β. The effect of β on the magnitude of $L_{c,p}$ is far smaller than exponential. Therefore, it has relatively small impact on the outcome of implied volatility curve. This is shown in Panel (5) of Figure A.1. The difference of implied volatility generated from a skew λ distribution (blue line) vs a symmetric distribution (black line) is very small. However, the main effect of β is to produce the right magnitude for $\Delta\mu$. In Panel (6) of Figure A.1, the large enough skew is produced by $\Delta\mu = \frac{1}{2}\mu_D(\lambda,\sigma,\beta)$ which is in the order of $\sigma\beta$. Without reference to β, a symmetric distribution will not be able to determine what $\Delta\mu$ should be for the skew. We also see a very small ϵ changes the upstrike curve quite a bit at large $\hat{k}$. This concludes the asymptotic analysis of volatility smile for small σ.

5. The Lambda Transformation

In the previous sections, I've developed the $\mathbb{V}_{c,p}(\lambda\to 1)$ and $\mathbb{L}_{c,p}$ operators in great length. The implied volatility ($\sigma_1(k)$) can be calculated by the combination of these two operators. However, it takes two more steps to connect to the market data observed in the real world. This 4-step transformation is called "the λ transformation". The transformation is best described in the operator form, similar to the quantum mechanics.

5.1. The Option Pricing Framework

The option prices, $\widetilde{\mathcal{C}}(k)$ and $\widetilde{\mathcal{P}}(k)$, observed in the market at a given time ($\widetilde{\mathcal{S}}$ is fixed) and time to maturity (T is fixed) is a composite transformation of

$$\begin{aligned}\frac{\widetilde{\mathcal{C}}(k)}{\widetilde{\mathcal{S}}} &= \mathbb{O}_c(\lambda=1)\cdot\mathbb{M}(r_M)\cdot\mathbb{V}_c(\lambda\to 1)\cdot(\mathbb{L}_c+\epsilon_c)\,|P(x;\lambda,\sigma,\beta,\mu_c)\rangle\,;\\ \frac{\widetilde{\mathcal{P}}(k)}{\widetilde{\mathcal{S}}} &= \mathbb{O}_p(\lambda=1)\cdot\mathbb{M}(r_M)\cdot\mathbb{V}_p(\lambda\to 1)\cdot(\mathbb{L}_p+\epsilon_p)\,|P(x;\lambda,\sigma,\beta,\mu_p)\rangle\,.\end{aligned} \tag{5.1}$$

This equation is called the λ transformation since it is projecting quantities from the $\lambda\geq 2$ regime on the right side towards the $\lambda=1$ regime on the left side.

First, the $\mathbb{L}_c$ operator takes a distribution $|P(x;\lambda,\sigma,\beta,\mu_{c,cp})\rangle$, and generate the normalized option prices (for each log-strike) in the local regime. This has been implemented in Section 3.5 in great detail.

Secondly, the $\mathbb{V}_{c,p}$ operator takes the normalized prices and generate the implied volatilities (for each log-strike) in the global regime. Conversely, the

$\mathbb{O}_{c,p}$ operator takes the implied volatilities and generates the normalized prices in the global regime. They have been implemented in Section 3.8.1.

Last, the $\mathbb{M}(r_M)$ operator is called the momentum operator. It is a new concept here in our option pricing model. It moves the implied volatility horizontally by the quantity r_M,

$$\mathbb{M}(r_M):\ \sigma_1(k) \mapsto \sigma_1(k + r_M) \tag{5.2}$$

The quantity r_M is called "market momentum". It is typically in the order of the standard deviation of log-returns of the underlying at the sampling frequency T, which is proportional to σ. The difference between r_M and σ is that r_M is directional. If the market participants expect the market to go up, r_M is positive, such is the case for SPX most of the time. On the other hand, if the general expectation on the market is bearish, r_M is negative (or non-positive), such is the case for preciosu metals.

Since the market prices are often quoted in terms of implied volatilities, $\sigma_{imp}(k)$, The $\mathbb{O}_{c,p}(\lambda = 1)$ operator can be moved to the left hand side via Equation 3.39, and we have the three-step λ transformation,

$$\begin{aligned}
\sigma_{imp}(k) = \mathbb{V}_c(\lambda = 1) \cdot \frac{\widetilde{\mathcal{C}}(k)}{\widetilde{\mathcal{S}}} = \quad & \mathbb{M}(r_M) \cdot \mathbb{V}_c(\lambda \to 1) \cdot (\mathbb{L}_c + \epsilon_c) \left| P(x; \lambda, \sigma, \beta, \mu_c) \right\rangle ; \\
\sigma_{imp}(k) = \mathbb{V}_p(\lambda = 1) \cdot \frac{\widetilde{\mathcal{P}}(k)}{\widetilde{\mathcal{S}}} = \quad & \mathbb{M}(r_M) \cdot \mathbb{V}_p(\lambda \to 1) \cdot (\mathbb{L}_p + \epsilon_p) \left| P(x; \lambda, \sigma, \beta, \mu_p) \right\rangle .
\end{aligned} \tag{5.3}$$

5.2. Market Momentum

The quantity $r_M(T)$ is an input from the option market. It is a new degree of freedom. The magnitude of r_M can be as large as annualized 30%. The effect is far more significant than the risk-free rate ($\ll 1\%$) and dividend yield ($\sim 2\%$). In our framework, this is a major contributor to the ATM skew.

I'd like to interpret $r_M(T)$ as an expression on market momentum. The collective expectation of the market participants moves the volatility smile not only up and down (via σ), but also left and right (via r_M). Market momentum has been observed and put to practice for two decades that generates fantastic investment results[3]. Whether they are really related will require further research.

Since the valley of a volatility smile can be fit with a quadratic curve, its location can be determined with reasonable accuracy. This distance is approximately r_M. Therefore, r_M can be determined from the market data in a nonparametric model-free fashion. It is interesting to investigate if such knowledge has any predictive power and provides any arbitrage opportunity.

Nevertheless, I do want to mention that, in the stochastic volatility models, the ATM skew is generated by the correlation (typically the ρ notation in $dW_1 dW_2 = \rho dt$) between the price process and the volatility process. In this frame of thought, r_M could be just an expression of such correlation. This would require more delicate comparison between models and large amount of market data.

5.3. The Cusp in Option Prices

One major question the reader may ask is - Does the "cusp" really exist? Or, can the cusp be observed "directly" from the market data? Although there are some complication due to the effect of λ transformation, I'd like to say, the answer is yes - especially in the option chain to expire in one day, in which the "market momentum" is smallest.

One of the most direct way to observe the cusp of the distribution, if it exists, is through the log-slope of $L_c(k)$, that is, the derivative of $\log L_c(k)$, normalized by the stdev of the distribution. We define this log-slope as

$$U_c(k) = \sqrt{\text{var}}\,\frac{d}{dk}\log L_c(k) = -\sqrt{\text{var}}\,\frac{e^k\ (1-\Phi(k))}{L_c(k)}, \tag{5.4}$$

$$U_p(k) = \sqrt{\text{var}}\,\frac{d}{dk}\log L_p(k) = \sqrt{\text{var}}\,\frac{e^k\,\Phi(k)}{L_p(k)}, \tag{5.5}$$

with the relation $\frac{dL_c(k)}{dk} = -e^k\ (1-\Phi(k))$ and $\frac{dL_p(k)}{dk} = e^k\,\Phi(k)$. The interpretation of $U_c(k)$ is how much option price changes in log scale per log-strike, measured in unit of standard deviation of the underlying distribution. Notice that this quantity is independent of the actual option prices nor the underlying. It is also almost independent of the level of volatility σ in a theoretical sense. It is a pure measure on the "shape" of the distribution itself. However, $\sqrt{\text{var}}$ doesn't not come from market data directly. But indirectly it can be estimated by Eq. (4.16) with minimum assumption. Comparing this curve between theory and data should give a fairly strong evidence that the cusp exists.

Figure 5.1 shows 6 panels for combinations of $\lambda = 2, 3$; and $\beta = 0$ and ± 0.7 (blue solid lines). The shape of $U_c(k)$ clearly shows the difference between $\lambda = 2$ and $\lambda = 3$. When $\lambda = 2$, the log-slope is flat in the right tail; while it is an obvious cusp for $\lambda = 3$. Once normalized by the stdev, the tip of the cusp is nearly a constant.

For $\lambda = 2$, the flat side of $U_{c,p}(k)$ can be fully analyzed. From Eqs. (2.26) and (3.40), the terms related to exponential cancel out nicely. With the variance of $\sigma^2\left(2+\beta^2\right)$, we have

$$\begin{aligned} U_c(k;\lambda=2,\sigma,\beta,\mu) &= \tfrac{-\sqrt{\text{var}}}{B^-}\left(\tfrac{1}{B^-_\sigma}-\tfrac{1}{B^-}\right)^{-1} = -B^-_\sigma\sqrt{2+\beta^2}, \quad \text{when } \hat{k}\geq 0;\\ U_p(k;\lambda=2,\sigma,\beta,\mu) &= \tfrac{-\sqrt{\text{var}}}{B^+}\left(\tfrac{1}{B^+_\sigma}-\tfrac{1}{B^+}\right)^{-1} = B^+_\sigma\sqrt{2+\beta^2}, \quad \text{when } \hat{k}<0. \end{aligned} \tag{5.6}$$

It is obvious that the right side of $U_c(k)$ and the left side of $U_p(k)$ are flat for $\lambda = 2$. As $\sigma \to 0$ and $\beta \to 0$, this level settles at $U_c = -U_p = -\sqrt{2}$.

This analysis can be extended to locate the tip of the cusp for the symmetric case. With the variance of $\sigma^2\Gamma\left(\frac{3\lambda}{2}\right)/\Gamma\left(\frac{\lambda}{2}\right)$, we have

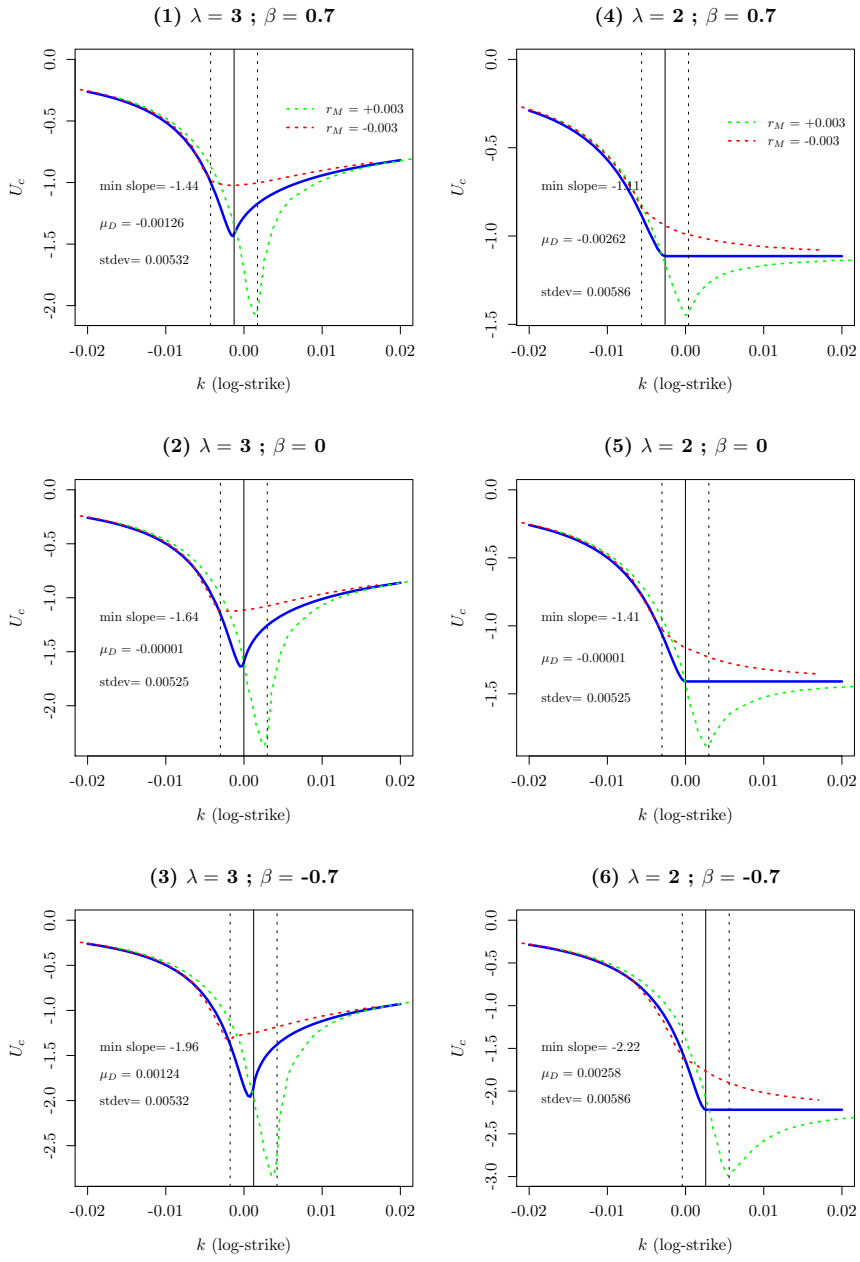

FIG 5.1. *The log-slope of call option prices, $U_c(k)$, for six combinations of $\lambda = 2, 3$; and $\beta = 0$ and ± 0.7.*

$$U_c(k=\mu;\lambda=2,\sigma,\beta,\mu) = -\sqrt{\Gamma\left(\frac{3\lambda}{2}\right)\Gamma\left(\frac{\lambda}{2}\right)}\left(\sum_{n=1}^{n_{\infty}}\frac{\sigma^{n-1}}{n!}\Gamma\left(\frac{\lambda(n+1)}{2}\right)\right)^{-1},$$
$$U_p(k;\lambda=2,\sigma,\beta,\mu) = \sqrt{\Gamma\left(\frac{3\lambda}{2}\right)\Gamma\left(\frac{\lambda}{2}\right)}\left(\sum_{n=1}^{n_{\infty}}\frac{(-\sigma)^{n-1}}{n!}\Gamma\left(\frac{\lambda(n+1)}{2}\right)\right)^{-1} \tag{5.7}$$

As $\sigma \to 0$, this level settles at $U_c = -U_p = -\frac{1}{\Gamma(\lambda)}\sqrt{\Gamma\left(\frac{3\lambda}{2}\right)\Gamma\left(\frac{\lambda}{2}\right)}$. For $\lambda = 3$, this is -1.605.

In Panel (2) of Figures 6.1 and 6.2, the log-slopes of SPX options are illustrated. However, the tip of the cusp is much higher than the theoretical value. This is explained next.

5.4. Mixing Market Momentum with The Cusp

In last section, we see that $U_{c,p}(k)$ have relatively stable shapes, primarily determined by λ. Its shapes can be observed fairly clearly in SPX option data. However, when we apply λ transformation, especially the market momentum operator, its shapes can change dramatically. When the momentum moves implied volatilities towards out-of-money, the cusp becomes more pronounced. On the other hand, when the momentum moves implied volatilities towards in-the-money, it produces a flattening effect that is quite unusual. This flattening shape is confirmed in the 4-day market data (Panel (5) of Figure 6.2). I consider this quite impressive.

If we consider $U_{c,p}(k)$ as the output of the corresponding operators, $\mathbb{U} = \sqrt{\text{var}}\,\frac{d}{dk}\log(...)$, then we have

$$U_c(k;r_M) = \mathbb{U}\cdot\mathbb{O}_c(\lambda=1)\cdot\mathbb{M}(r_M)\cdot\mathbb{V}_c(\lambda\to 1)\cdot(\mathbb{L}_c+\epsilon_c)\,|P(x;\lambda,\sigma,\beta,\mu_c)\rangle\,;$$
$$U_p(k;r_M) = \mathbb{U}\cdot\mathbb{O}_p(\lambda=1)\cdot\mathbb{M}(r_M)\cdot\mathbb{V}_p(\lambda\to 1)\cdot(\mathbb{L}_p+\epsilon_p)\,|P(x;\lambda,\sigma,\beta,\mu_p)\rangle\,. \tag{5.8}$$

In Figure 5.1, the positive momentum and negative momentum are added as dotted green and red lines. We see that the positive momentum enhances the cusp. It even morphs the flat slope of $\lambda = 2$ into a cusp. We also see that the negative momentum flattens the cusp into a round top. The λ transformation framework can produce the shapes of log-slope that are observed in the market data.

6. Fitting Volatility Smiles of SPX Options

I've finished the theoretical aspect of the option pricing model. In this section, the fits to the SPX options are presented to validate the framework developed so far. The option data is from CBOE for the month of June, 2015. Two fits are performed - The first fit is on an option chain to expire in a day. The second fit is on an option chain to expire in 4 days.

The bids, asks and mid-points (average of bid-ask) from market data are drawn on the charts. The attempt is to fit to the mid-points as best as possible.

On the theoretical side, the blue line represents the full λ transformation; the red line takes r_M out; and lastly, the dotted red line takes both r_M and ϵ out. This helps the reader to visualize the effect in each step of λ transformation.

In the middle panels, the log-slopes of market data are fit with polynomials of 6 orders on each side of the cusp at k_{cusp}. This fit helps us understand the existence of the cusp in the market data.

6.1. Fitting SPX Options to Expire in One Day

The first fit in Figure 6.1 is from the option chain closed as of June 18, to expire in a day on the quad-witching day, June 19. $\lambda = 3.125$ is chosen to fit the very sharp cusp. The market momentum r_M is in the same order as σ, its effect is relatively small but still noticeable by comparing the tips of log-slopes. The cusps are very clear in both calls and puts. $\Delta\mu$ is fairly large for calls but not much so for puts. And $\epsilon_{c,p}$ are fairly small.

6.2. Fitting SPX Options to Expire in 4 Days

The second fit in Figure 6.2 is from the option chain closed as of June 16, to expire in 4 days on the quad-witching day, June 19. $\lambda = 2.8$ is smaller and it is in line with decreasing kurtosis for larger T. However, the term structure of the kurtosis is still an unsolved research area. The enhanced cusp in log-slope for calls and the flattened cusp for puts are very obvious. The fits from the theory are reasonable well done. $r_M = 0.017$ in calls is very noticeable and is much larger than μ and μ_D. However, in puts, $r_M = 0.012$ is slightly smaller but is still significant.

7. Summary

In this paper, I have developed the option pricing model based on λ distribution. When the distribution is combined with the 4-step λ transformation, its accuracy is very good when fitting the short-maturity SPX options. More research should be conducted to extend the term structure on maturity. The model should be also tested on different asset classes, especially on those that have drastically different smile structure.

Finally, I'd like to thank my family members that this work can't be accomplished without their tolerance. I am very grateful for them.

Appendix A: Skew Generalized Error Distribution (SGED)

The SGED can be viewed as a variation of symmetric λ distribution by multiplying σ with $(1 + \text{sgn}(x - \mu)\beta)$ on each side of the cusp. The skew is different from our origin of elliptic curve at $\lambda = 3$. However, since its mathematics is fully compatible with that of incomplete gamma function, lots of mathematical

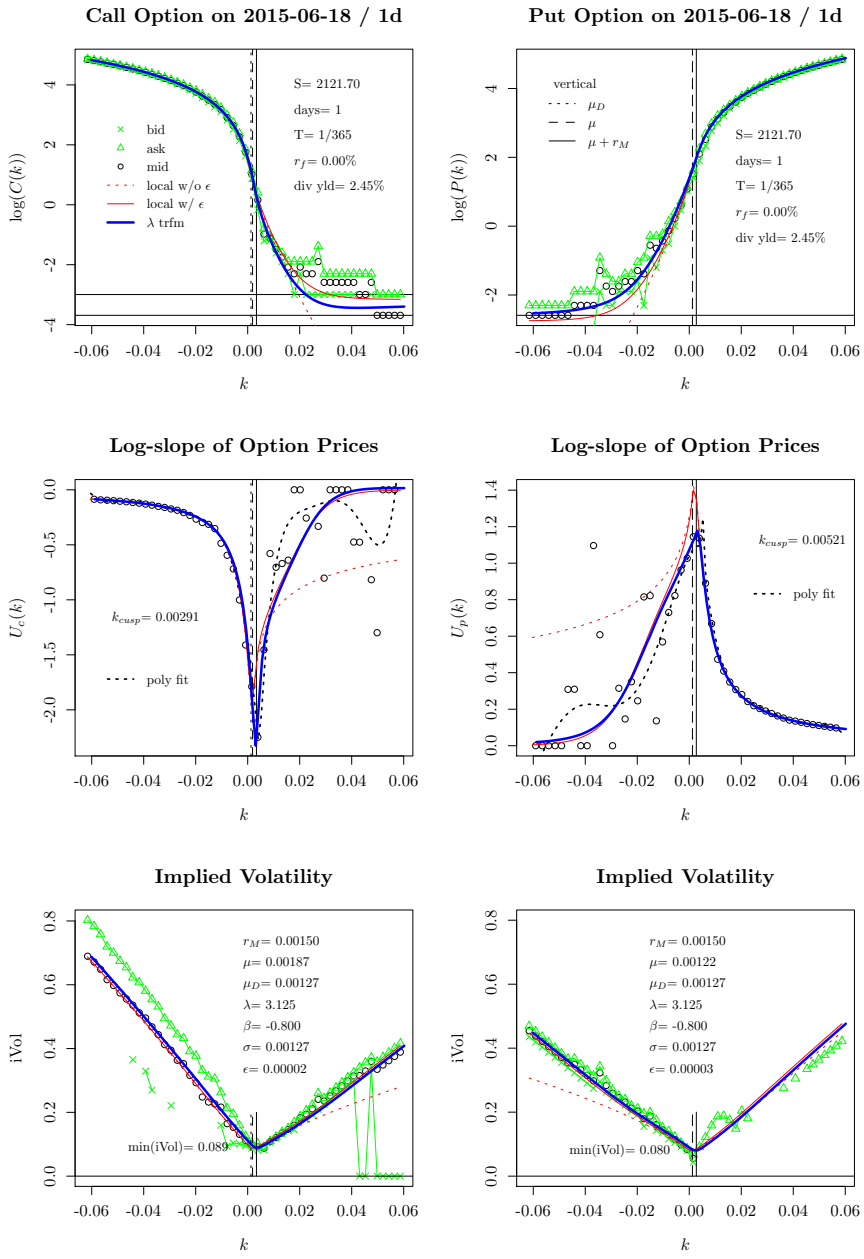

FIG 6.1. *SPX options to expire in one day. Fit with λ distribution.*

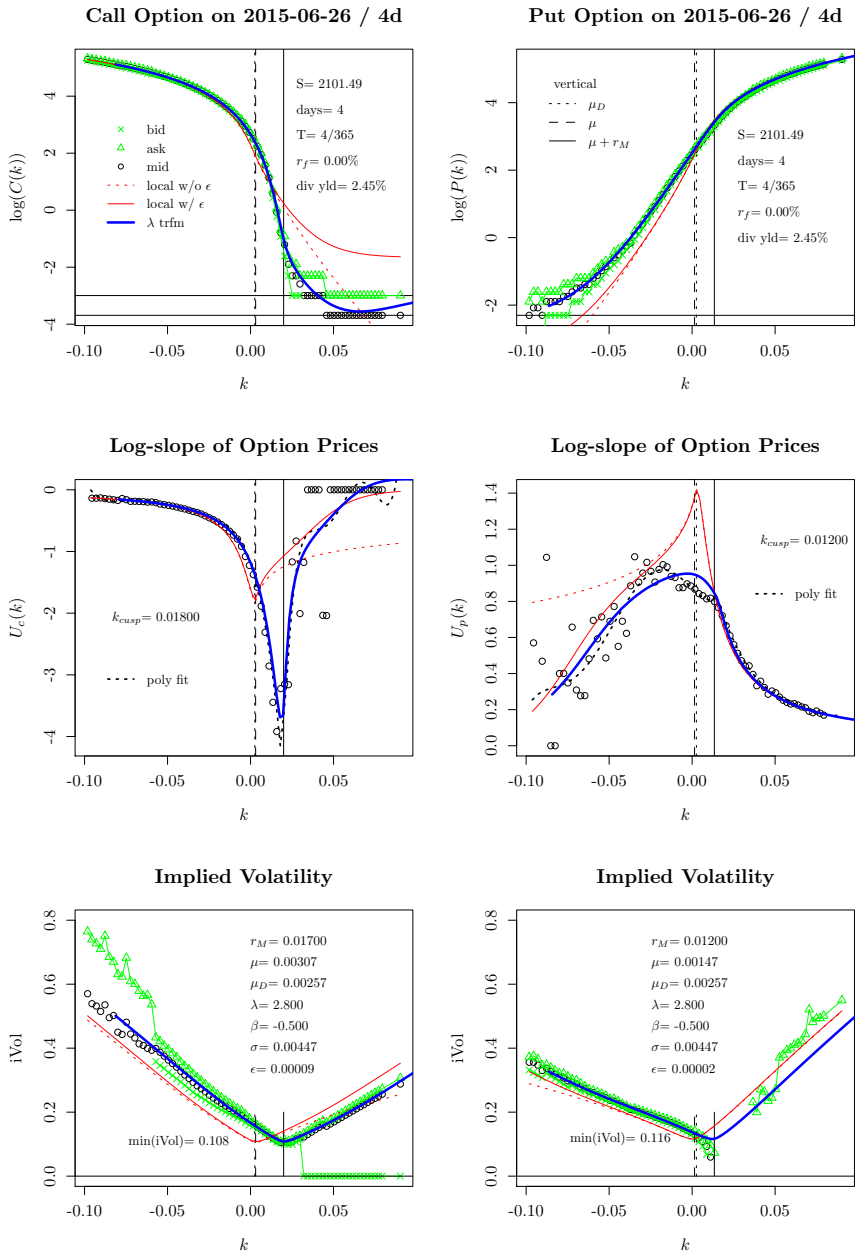

FIG 6.2. *SPX options to expire in 4 days. Fit with λ distribution.*

frameworks developed here can be easily modified to cover SGED in the option pricing setting. In this appendix, I attempt to summary the results in a similar order that λ distribution is laid out in this paper.

The simplest view of SGED is that it has two different σ's for each side of cusp. It fits well with our framework of thought. Using $\pm$ notation, we have

$$\begin{aligned} \sigma_- &= \sigma\,(1-\beta)\,, \quad \text{when } (x-\mu)<0, \\ \sigma_+ &= \sigma\,(1+\beta)\,, \quad \text{when } (x-\mu)\geq 0; \end{aligned} \tag{A.1}$$

where σ_+ is the volatility parameter assigned to the right tail of a (symmetric) λ distribution, and σ_- for the left tail. And the following algebraic rules hold,

$$\begin{aligned} \sigma_+ + \sigma_- &= \sigma, \\ \sigma_+ - \sigma_- &= 2\beta\sigma, \end{aligned} \tag{A.2}$$

Note that β in SGED context has a very different scale, compared to λ distribution. It modifies the standard deviation in a much significant way.

A.1. The Distribution

The curves are defined for each side of the cusp,

$$\begin{aligned} y^-(x) &= -\left|\tfrac{x-\mu}{\sigma_-}\right|^{\frac{2}{\lambda}}, \quad \text{when } (x-\mu)<0; \\ y^+(x) &= -\left|\tfrac{x-\mu}{\sigma_+}\right|^{\frac{2}{\lambda}}, \quad \text{when } (x-\mu)\geq 0. \end{aligned} \tag{A.3}$$

The slope is $\frac{dy}{dx} = \frac{2}{\lambda}\frac{y}{(x-\mu)} = -\operatorname{sgn}(x-\mu)\,\frac{2}{\lambda\sigma_\pm}\left|\frac{x-\mu}{\sigma_\pm}\right|^{\frac{2}{\lambda}-1}$. The normalization constant is independent of β,

$$C = \sigma\hat{C} = \sigma\lambda\Gamma\left(\frac{\lambda}{2}\right) \tag{A.4}$$

With $\hat{y}(x) = x^{-2/\lambda}$ and $\hat{x}_\pm = \frac{x-\mu}{\sigma_\pm}$, the CDF, $\Phi(x)$, and CCDF, $1-\Phi(x)$, are

$$\begin{aligned} \Phi(x) &= \tfrac{\sigma_-}{\sigma\hat{C}(\lambda,\beta)}\int_{-\infty}^{\hat{x}_-} e^{\hat{y}(x)}dx = \tfrac{\sigma_-}{2\sigma\,\Gamma\left(\frac{\lambda}{2}\right)}\Gamma\left(\tfrac{\lambda}{2}, \left|\tfrac{x-\mu}{\sigma_-}\right|^{\frac{2}{\lambda}}\right), \quad \text{when } (x-\mu)<0; \\ 1-\Phi(x) &= \tfrac{\sigma_+}{\sigma\hat{C}(\lambda,\beta)}\int_{\hat{x}_+}^{\infty} e^{\hat{y}(x)}dx = \tfrac{\sigma_+}{2\sigma\,\Gamma\left(\frac{\lambda}{2}\right)}\Gamma\left(\tfrac{\lambda}{2}, \left|\tfrac{x-\mu}{\sigma_+}\right|^{\frac{2}{\lambda}}\right), \quad \text{when } (x-\mu)\geq 0. \end{aligned} \tag{A.5}$$

The $B^\pm$ notation is defined as the tail integrals, $B^\pm = \int_0^\infty e^{\hat{y}^\pm(x)}dx$, whose algebraic rules are

$$\begin{aligned} B^+ + B^- &= \hat{C}, \\ \Delta B = B^+ - B^- &= \beta\lambda\Gamma\left(\tfrac{\lambda}{2}\right). \end{aligned} \tag{A.6}$$

We see that ΔB, the difference of the tail integrals, is not the same as that of λ distribution in Eq. (2.32).

A.2. Basic Statistics

The n-th moment is

$$m_{(n)} = \frac{\Gamma\left(\frac{\lambda(n+1)}{2}\right)}{2\sigma\,\Gamma\left(\frac{\lambda}{2}\right)}\left(\sigma_+^{n+1} + (-1)^n\,\sigma_-^{n+1}\right). \tag{A.7}$$

The first 4 moments are straightforward:

$$\begin{array}{c} m_{(1)} = 2\sigma\beta\,\Gamma(\lambda)/\Gamma\left(\frac{\lambda}{2}\right), \\ m_{(2)} = \sigma^2\left(1+3\beta^2\right)\,\Gamma\left(\frac{3}{2}\lambda\right)/\Gamma\left(\frac{\lambda}{2}\right), \\ m_{(3)} = 4\sigma^3\beta\left(1+\beta^2\right)\,\Gamma(2\lambda)/\Gamma\left(\frac{\lambda}{2}\right), \\ m_{(4)} = \sigma^4\left(1+10\beta^2+5\beta^4\right)\,\Gamma\left(\frac{5}{2}\lambda\right)/\Gamma\left(\frac{\lambda}{2}\right). \end{array} \tag{A.8}$$

and the variance is (with $\mathit{\Gamma}_2(\lambda) = \frac{\Gamma\left(\frac{3\lambda}{2}\right)}{\Gamma\left(\frac{\lambda}{2}\right)}$)

$$\text{var} = \mathit{\Gamma}_2(\lambda)\,\sigma^2\left(1+3\beta^2-4\beta^2\,\frac{\Gamma(\lambda)^2}{\Gamma\left(\frac{3}{2}\lambda\right)\Gamma\left(\frac{\lambda}{2}\right)}\right). \tag{A.9}$$

At $\lambda = 3$, SGED has $m_{(1)} = 4.5\,\sigma\beta$; while λ distribution has $m_{(1)} = 1.235\,\sigma\beta$. β's influence on $m_{(1)}$ in SGED is about 3.6 times that of λ distribution. For variance, SGED has var $= \mathit{\Gamma}_2(\lambda)\,\sigma^2\left(1+1.45\beta^2\right)$; while λ distribution has var $= \mathit{\Gamma}_2(\lambda)\,\sigma^2\left(1+0.076\beta^2\right)$. The coefficient on β^2 in SGED is about 19 times that of λ distribution. Intuitively speaking, we can estimate that β's influence in SGED is about 3.6-4.3 times that of β in λ distribution.

A.3. MGF and Summation Truncation

When $\lambda = 2$, the MGF is

$$M(t) = \left(1 - 2\beta\sigma t - \left(1-\beta^2\right)\sigma^2 t^2\right)^{-1}. \tag{A.10}$$

When $\lambda > 2$, the MGF is

$$M(t) = e^{t\mu}\left[1 + \sum_{n=1,2,\ldots}^{n_{\rightrightarrows}} \frac{m_{(n)}t^n}{n!}\right], \tag{A.11}$$

in which the summation is truncated at $n_{\rightrightarrows}$ where $\frac{m_n t^n}{n!}$ reaches its minimum, using the even moment formula. So $n_{\rightrightarrows}$ is the root of slightly different digamma equation,

$$\log(\sigma t) + \frac{\lambda}{2}\psi\left(\frac{\lambda(n+1)}{2}\right) - \psi(n+1) + \frac{H(n+1,\beta)}{F(n+1,\beta)} = 0, \tag{A.12}$$

where $\frac{H(n+1,\beta)}{F(n+1,\beta)}$ is $\frac{d}{dn}\log F(n+1,\beta)$ with $F(n,\beta) = \frac{1}{2}\left((1+\beta)^n + (1-\beta)^n\right)$ and $H(n,\beta) = \frac{1}{2}\left(\log(1+\beta)\,(1+\beta)^n + \log(1-\beta)\,(1-\beta)^n\right)$. When $n\,\log(1+|\beta|) \gg$

1, $\frac{H(n+1,\beta)}{F(n+1,\beta)} \to \log(1+|\beta|)$ which can be combined with $\log(\sigma t)$. So we see that the skew parameter has an effect of being a multiplier to σ: $\sigma \mapsto \sigma(1+|\beta|)$ in the summation truncation.

The rate of risk-neutral drift, μ_D is

$$\mu_D = -\log\left(M\left(t=1;\mu=0\right)\right), \tag{A.13}$$

whose small σ approximation is

$$\mu_D \approx -m_{(1)} - \frac{\text{var}}{2} = -2\sigma\beta\frac{\Gamma(\lambda)}{\Gamma\left(\frac{\lambda}{2}\right)} - \frac{\Gamma_2(\lambda)}{2}\sigma^2\left(1+3\beta^2-4\beta^2\frac{\Gamma(\lambda)^2}{\Gamma\left(\frac{3}{2}\lambda\right)\Gamma\left(\frac{\lambda}{2}\right)}\right). \tag{A.14}$$

A.4. MGF Integral Truncation

The MGF integral for μ_D can be represented in unit distribution as,

$$\begin{gathered} e^{-\mu_D} = M(t=1;\mu=0) = 1+ \\ \frac{1}{\hat{C}(\lambda,\beta)}\left(\frac{\sigma_+}{\sigma}\int_0^{\overrightarrow{\infty}}\left(e^{\sigma_+ x}-1\right)e^{\hat{y}(x)}dx + \frac{\sigma_-}{\sigma}\int_0^{\infty}\left(e^{-\sigma_- x}-1\right)e^{\hat{y}(x)}dx\right) \\ = 1+\frac{1}{\hat{C}(\lambda,\beta)}\int_0^{\overrightarrow{\infty}}\left(\frac{\sigma_+}{\sigma}e^{\sigma_+ x}+\frac{\sigma_-}{\sigma}e^{-\sigma_- x}-1\right)e^{-x^{2/\lambda}}dx. \end{gathered} \tag{A.15}$$

When $\lambda > 2$, the truncation occurs in the right tail at root of $\frac{dy}{dx} + t = -\frac{2}{\lambda\sigma_+}\left(\frac{x}{\sigma_+}\right)^{\frac{2}{\lambda}-1} + t = 0$. The analytic solution is

$$x_{\overrightarrow{\infty}} = \sigma_+\left(\frac{2}{\sigma_+ t\lambda}\right)^{\frac{\lambda}{\lambda-2}}. \tag{A.16}$$

However, the relation $n_{\overrightarrow{\infty}} + 1 = x_{\overrightarrow{\infty}}$ doesn't hold exactly for $\beta \neq 0$. Especially when β is negative, $x_{\overrightarrow{\infty}}$ is much larger than $n_{\overrightarrow{\infty}} + 1$. For instance, at $\lambda = 3, \sigma = 0.1, \beta = -0.2$, $x_{\overrightarrow{\infty}} = 46.3$ and $n_{\overrightarrow{\infty}} = 19.58$. But this doesn't affect the consistency between the summation method and the integral method. The result for μ_D has 0.3% error between the two methods. The error is reduced to 10^{-5} if σ is lowered to 0.01.

A.5. Incomplete MGF and Incomplete Moments

The incomplete MGF can be obtained by integrating the tail on each side of the cusp through a unit distribution, following the truncation procedure of $x_{\overrightarrow{\infty}}$ outlined above,

$$\begin{aligned} M_c(k) &= \frac{\sigma_+ e^{\mu}}{\sigma\hat{C}}\int_{\hat{k}_+}^{\overrightarrow{\infty}} e^{-x^{2/\lambda}+\sigma_+ x}dx, \quad \text{when } \hat{k}_+ = \tfrac{k-\mu}{\sigma_+} \geq 0; \\ M_p(k) &= \frac{\sigma_- e^{\mu}}{\sigma\hat{C}}\int_{-\hat{k}_-}^{\infty} e^{-x^{2/\lambda}-\sigma_- x}dx, \quad \text{when } \hat{k}_- = \tfrac{k-\mu}{\sigma_-} < 0; \end{aligned} \tag{A.17}$$

The parity equation is: $M_c(k) + M_p(k) = M(t=1) = e^{\mu-\mu_D}$. And $M(t=1) = M_c(k=\mu) + M_p(k=\mu)$.

The incomplete MGF can also expanded as sum of incomplete moments of a unit distribution,

$$\begin{array}{ll} M_c(k) = e^{\mu}\sum_{n=0}^{n_{\overrightarrow{\infty}}} \frac{\sigma_+^n}{n!}\hat{M}_c^{(n)}(\hat{k}_+), & \text{when } \hat{k}_+ \geq 0; \\ M_p(k) = e^{\mu}\sum_{n=0}^{n_{\overrightarrow{\infty}}} \frac{\sigma_-^n}{n!}\hat{M}_p^{(n)}(\hat{k}_-), & \text{when } \hat{k}_- < 0. \end{array} \tag{A.18}$$

where $n_{\overrightarrow{\infty}}$ is the limit of the summation, and $\hat{M}_{c,p}^{(n)}(\hat{k}_\pm)$ are based on the incomplete gamma function,

$$\begin{array}{ll} \hat{M}_c^{(n)}(\hat{k}_+) = \frac{\sigma_+}{2\sigma\Gamma\left(\frac{\lambda}{2}\right)}\,\Gamma\left(\frac{\lambda(n+1)}{2}, \left|\hat{k}_+\right|^{\frac{2}{\lambda}}\right), & \text{when } \hat{k}_+ \geq 0; \\ \hat{M}_p^{(n)}(\hat{k}_-) = \frac{(-1)^n\sigma_-}{2\sigma\Gamma\left(\frac{\lambda}{2}\right)}\,\Gamma\left(\frac{\lambda(n+1)}{2}, \left|\hat{k}_-\right|^{\frac{2}{\lambda}}\right), & \text{when } \hat{k}_- < 0. \end{array} \tag{A.19}$$

A.6. OGF

The incomplete MGF can be obtained by integrating the tail on each side of the cusp through a unit distribution, following the truncation procedure of $x_{\overrightarrow{\infty}}$ outlined above,

$$\begin{array}{ll} L_c(k) = \frac{\sigma_+ e^{\mu}}{\sigma\hat{C}}\int_{\hat{k}_+}^{\overrightarrow{\infty}} e^{-x^{2/\lambda}}\left(e^{\sigma_+ x} - e^{\sigma_+\hat{k}_+}\right)dx, & \text{when } \hat{k}_+ = \frac{k-\mu}{\sigma_+} \geq 0; \\ L_p(k) = -\frac{\sigma_- e^{\mu}}{\sigma\hat{C}}\int_{-\hat{k}_-}^{\infty} e^{-x^{2/\lambda}}\left(e^{-\sigma_- x} - e^{\sigma_-\hat{k}_-}\right)dx, & \text{when } \hat{k}_- = \frac{k-\mu}{\sigma_-} < 0; \end{array} \tag{A.20}$$

By replacing incomplete MGF and CDF with incomplete moments, the OGF, $L_{c,p}(k)$, can be expressed as,

$$\begin{array}{rcll} L_c(k) & = & \sigma_+ e^{\mu}\sum_{n=1}^{n_{\overrightarrow{\infty}}} \frac{\sigma_+^{n-1}}{n!}\left(\hat{M}_c^{(n)}(\hat{k}_+) - \hat{k}_+^n\,\hat{M}_c^{(0)}(\hat{k}_+)\right), & \text{when } \hat{k}_+ \geq 0; \\ & = & \frac{\sigma_+^2 e^{\mu}}{2\sigma\Gamma\left(\frac{\lambda}{2}\right)}\sum_{n=1}^{n_{\overrightarrow{\infty}}} \frac{\sigma_+^{n-1}}{n!}\left(\Gamma\left(\frac{\lambda(n+1)}{2}, \left|\hat{k}_+\right|^{\frac{2}{\lambda}}\right) - \hat{k}_+^n\,\Gamma\left(\frac{\lambda}{2}, \left|\hat{k}_+\right|^{\frac{2}{\lambda}}\right)\right); & \\ L_p(k) & = & \sigma_- e^{\mu}\sum_{n=1}^{n_{\overrightarrow{\infty}}} \frac{(-1)^{n-1}\sigma_-^{n-1}}{n!}\left(\hat{M}_p^{(n)}(\hat{k}_-) - \left|\hat{k}_-^n\right|\,\hat{M}_p^{(0)}(\hat{k}_-)\right), & \text{when } \hat{k}_- < 0; \\ & = & \frac{\sigma_-^2 e^{\mu}}{2\sigma\Gamma\left(\frac{\lambda}{2}\right)}\sum_{n=1}^{n_{\overrightarrow{\infty}}} \frac{(-1)^{n-1}\sigma_-^{n-1}}{n!}\left(\Gamma\left(\frac{\lambda(n+1)}{2}, \left|\hat{k}_-\right|^{\frac{2}{\lambda}}\right) - \left|\hat{k}_-^n\right|\,\Gamma\left(\frac{\lambda}{2}, \left|\hat{k}_-\right|^{\frac{2}{\lambda}}\right)\right). & \end{array} \tag{A.21}$$

The put-call parity, $L_c(k) - L_p(k) = e^{\mu-\mu_D} - e^k$, is used to derive the other half on each side of the cusp. We can see that, to the first order of σ, option prices are proportional to σ.

When $k = \mu$, we have $\hat{M}_{c,p}^{(n)}(0) = \frac{\sigma_\pm}{2\sigma\Gamma\left(\frac{\lambda}{2}\right)}\,\Gamma\left(\frac{\lambda(n+1)}{2}\right)$. Therefore, the option prices that contributes to minimum implied volatility are

$$\begin{aligned} L_c(\mu) &= \frac{\sigma_+^2 e^{\mu}}{2\sigma\Gamma\left(\frac{\lambda}{2}\right)} \sum_{n=1}^{n\to\infty} \frac{\sigma_+^{n-1}}{n!} \Gamma\left(\frac{\lambda(n+1)}{2}\right), \\ L_p(\mu) &= \frac{\sigma_-^2 e^{\mu}}{2\sigma\Gamma\left(\frac{\lambda}{2}\right)} \sum_{n=1}^{n\to\infty} \frac{(-1)^{n-1}\sigma_-^{n-1}}{n!} \Gamma\left(\frac{\lambda(n+1)}{2}\right). \end{aligned} \tag{A.22}$$

The λ transformation is then used to translate option prices from the local regime to the global regime.

A.7. Small σ Limit of OGF

For very small σ, the OGF reaches its simplest analytic form using $L_\lambda^*(\hat{k})$, but $\hat{k}$ must be β-adjusted on each side of the cusp,

$$\begin{aligned} L_c(k;\lambda,\sigma,\beta=0,\mu) &= \begin{cases} \frac{\sigma_+^2}{\sigma} e^{\mu} L_\lambda^*(\hat{k}_+), & \text{when } \hat{k}_+ \geq 0; \\ \frac{\sigma_-^2}{\sigma} e^{\mu} L_\lambda^*(\hat{k}_-) + \left(1+\Delta\mu - e^k\right), & \text{when } \hat{k}_- < 0; \end{cases} \\ L_p(k;\lambda,\sigma,\beta=0,\mu) &= \begin{cases} \frac{\sigma_+^2}{\sigma} e^{\mu} L_\lambda^*(\hat{k}_+) - \left(1+\Delta\mu - e^k\right), & \text{when } \hat{k}_+ \geq 0; \\ \frac{\sigma_-^2}{\sigma} e^{\mu} L_\lambda^*(\hat{k}_-), & \text{when } \hat{k}_- < 0; \end{cases} \end{aligned} \tag{A.23}$$

where $\Delta\mu = e^{\mu-\mu_D} - 1$. The $\left(1+\Delta\mu - e^k\right)$ term can be further reduced to $-e^{\mu}\sigma_\pm\hat{k}_\pm$ for small $\left|\hat{k}_\pm\right|$ and when $\mu_D \sim o(\sigma)$.

A.8. SGED as A Step-Function Approximation of λ Distribution

SGED can be viewed as a step-function approximation of λ distribution. This should become obvious visually from Panel (2) of Figure A.1. For a classroom modelling exercise, this simpler distribution may be quite useful. In Figure A.1, we take the parameters from SPX 1-day option fits, and compare various behaviors of the skew λ distribution and SGED to the symmetric distribution. The black dotted line is the symmetric distribution, with $\lambda = 3.125, \sigma = 0.00127$. The blue line adds $\beta = -0.5$ to the λ distribution; and the red line is SGED with $\beta = -0.18$. Panel (1) is the comparison of $y(x)$ among the three distributions. The next three plots walk through the relative changes of $y(x)$, PDF, and CDF.

However, it gets interesting in Panel (5) where the risk neutral implied volatility from three distributions are very similar. The differences among them don't produce large volatility skew. In Panel (6), $\Delta\mu = 0.0004$ and $\epsilon = 0.0002$ are added to the distributions. The size of $\Delta\mu$ is chosen to be about half of μ_D of the skew λ distribution. The volatility skew is generated and all three curves look very much alike. What this tells us is that the main contributors of the volatility skew are non-zero $\Delta\mu$ and ϵ.

A.9. Fitting SPX Options

SGED is fully compatible with the λ transformation. It can produce equally well fits to SPX options. Figure A.2 reproduces the fits to SPX options to expire in

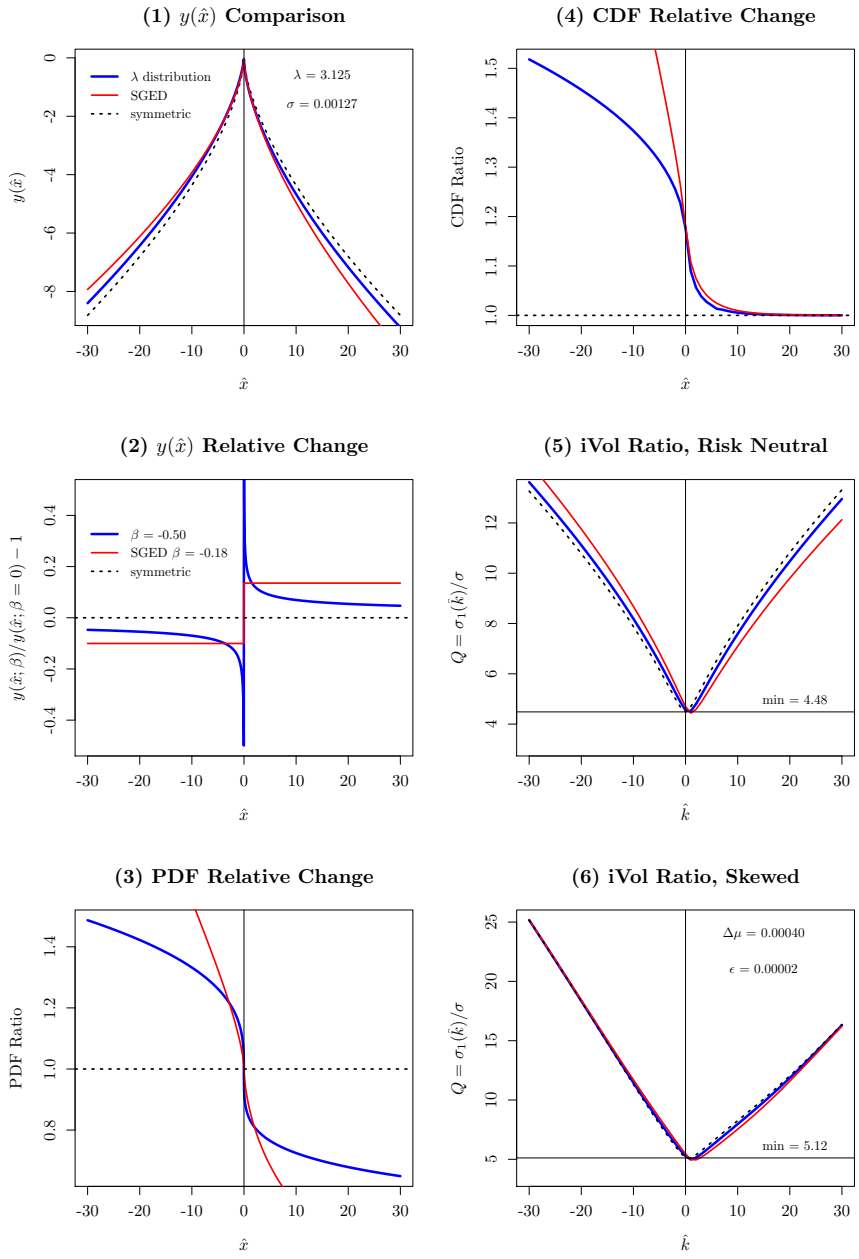

FIG A.1. *SGED as a step-function approximation of λ distribution. Panel (5) and Panel (6) shows that the main contributors of the volatility skew are non-zero $\Delta\mu$ and ϵ.*

one day with SGED parameters; and Figure A.3 shows the fits to SPX options to expire in 4 days.

The fits uses slightly different values for the parameters. One major difference is much smaller scale for β; while λ stays the same.

A.9.1. Discussion

SGED presents a good alternative to λ distribution since it is a step-function approximation. It fits very well within the λ transformation framework. It is simpler to manuver and captures most of the features in the volatility smile. The difference is probably only within the small region near ATM. If this matters, more research will have to be conducted.

References

[1] Abramowitz, M. and Stegun, I. A. (1964). "Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables", Dover, New York.

[2] Andersen, Leif B. G., Piterbarg, Vladimir V. (2007). "Moment explosions in stochastic volatility models". Finance and Stochastics, January 2007, 11(1), 29-50. Also available on SSRN: 559481.

[3] Asness, Clifford S., Moskowitz, Tobias J., Pedersen, Lasse H. (2012). "Value and Momentum Everywhere". Chicago Booth Research Paper No. 12-53. SSRN: 2174501.

[4] Austing, Peter (2014). "Smile Pricing Explained", Palgrave and Macmillan. ISBN: 978-1137-323571-5.

[5] Hansen, C., McDonald, J. and Theodossiou, P. (2008). Some Flexible Parametric Models for Partially Adaptive Estimators of Econometric Models. Economics: The Open-Access, Open-Assessment E-Journal, Vol. 1, 2007-7. Also available on SSRN: 1726799.

[6] Barndorff-Nielsen, Halgreen (1977). "Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions". Wahrscheinlichkeitstheorie und verwandte Gebiete 38, p. 309-311.

[7] Black, Fischer; Myron Scholes (1973). "The Pricing of Options and Corporate Liabilities". Journal of Political Economy 81 (3), 637–654.

[8] Breymann, Wolfgang, and Luthi, David (2013), "ghyp: A package on generalized hyperbolic distributions". Vignette of `ghyp` package on CRAN.

[9] Das, S. R. and Sundaram, R. K. (1999). "Of Smiles and Smirks: A Term Structure Perspective". Journal of Financial and Quantitative Analysis, 34(2), 211-240.

[10] Dupire, B., 1992, "Arbitrage Pricing with Stochastic Volatility", Proceedings of AFFI Conference, Paris, June.

[11] Dupire, B., 1993b, "Pricing and Hedging with Smiles", Proceedings of AFFI Conference, La Baule, June (also presented at IAFE meeting, New York, December 1993).

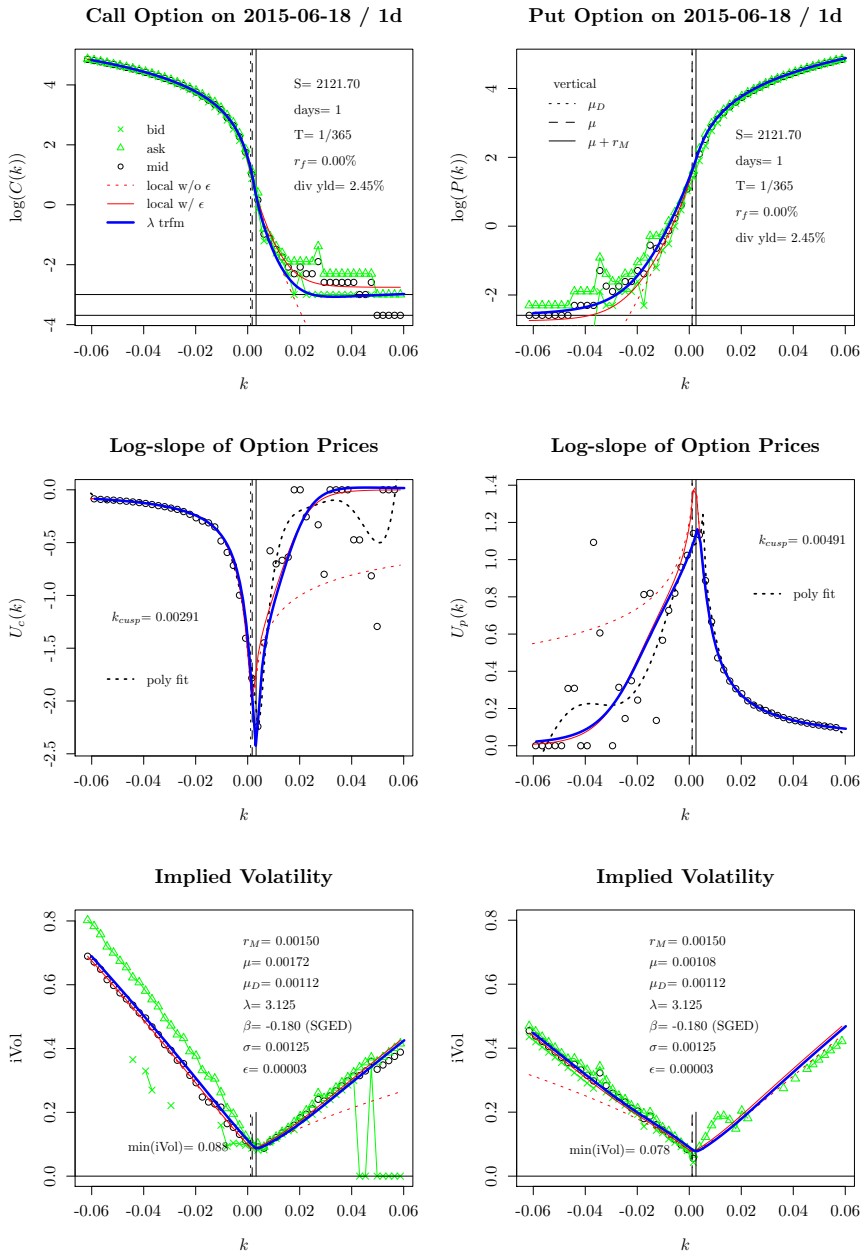

FIG A.2. *SPX options to expire in one day. Fit with SGED.*

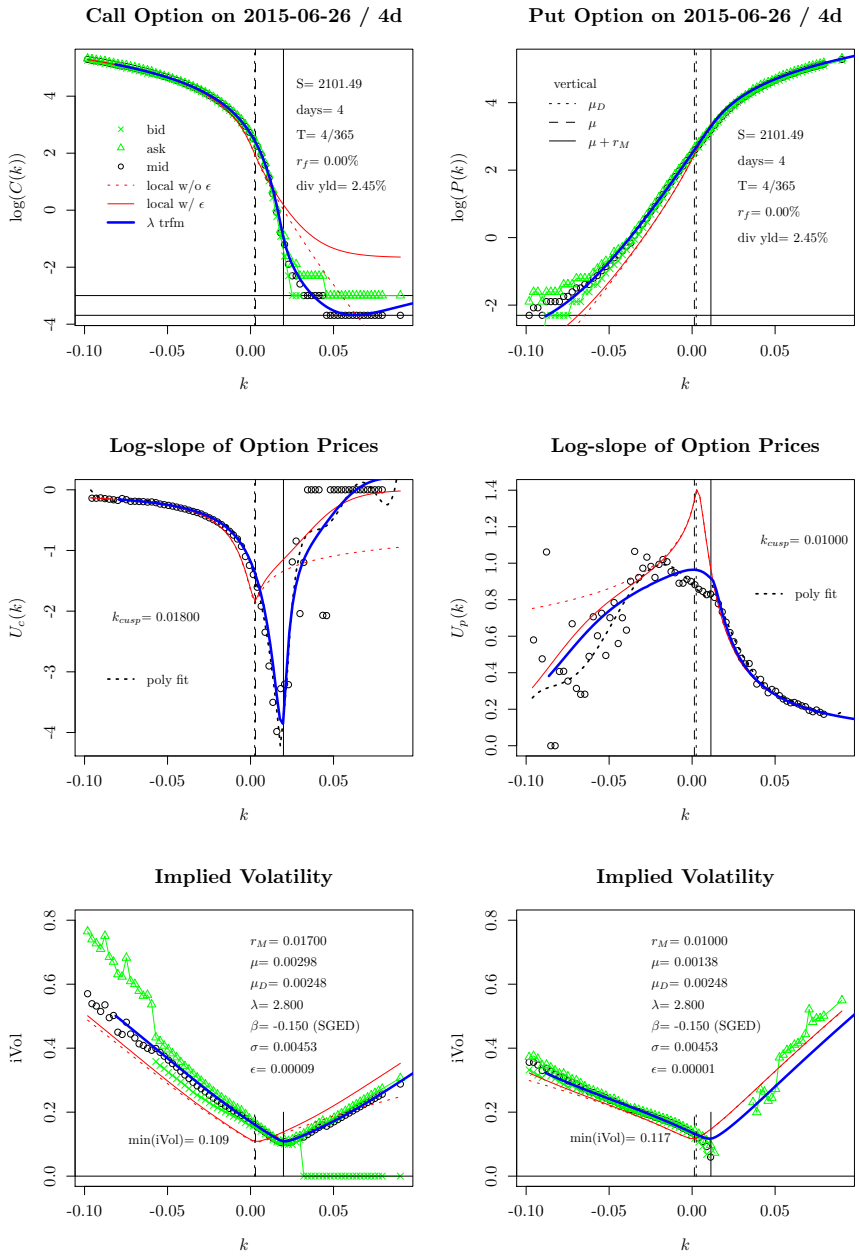

FIG A.3. *SPX options to expire in 4 days. Fit with SGED.*

[12] Gatheral, J (2006). "The Volatility Surface: A Practitioner's Guide". Wiley, ISBN: 0471792519.
[13] Gabaix, X. (2009). "Power Laws in Economics and Finance," Annual Review of Economics, 1, p. 255-93.
[14] Mantegna, R.N. and Stanley, H.E. (1994). "Stochastic processes with ultraslow convergence to a Gaussian: The truncated Lévy flight," Physical Review Letters 73: 2946-2949.
[15] Lihn, S. H.-T. (2015). "On the Elliptic Distribution and Its Application to Leptokurtic Financial Data". SSRN: 2697911.
[16] Rachev, S. T., Fabozzi, F. J., Menn, C. (2005). "Fat-Tailed and Skewed Asset Return Distributions : Implications for Risk Management, Portfolio Selection, and Option Pricing". Wiley, ISBN: 0471718866.
[17] Tate, John T. (1974). "The Arithmetic of Elliptic Curves". Inventiones Mathematicae, Volume 23, Issue 3-4, p. 179-206, Springer-Verlag.
[18] Silverman, J. H. (2008). "The Arithmetic of Elliptic Curves". Second Edition, ISBN 978-0-387-09493-9, Springer.
[19] Subbotin, M. T. (1923). "On the law of frequency of error", Matematicheskii Sbornik, 31, 296-301.
[20] Vargas, V., Dao, T.-L., and Bouchaud, J.-P. (2013). "Skew and implied leverage effect: smile dynamics revisited". arXiv:1311.4078v1[q-fin.ST].