On a Poisson Subordinated Distribution for Precise Statistical Measurement of Leptokurtic Financial Data

Stephen H-T. Lihn *†

March 25, 2012

Abstract

A new Poisson subordinated distribution is proposed to capture major leptokurtic features in log-return time series of financial data. This distribution is intuitive, easy to calculate, and converge quickly. It fits well to the historical daily log-return distributions of currencies, commodities, Treasury yields, VIX, and, most difficult of all, DJIA. It serves as a viable alternative to the more sophisticated truncated stable distribution.

Contents

*Email: stevelihn@gmail.com

†This is a working paper.

1 Introduction

The method of mixing Gaussian distributions of varying variances to produce a leptokurtic distribution has been known since 1970's (Praetz, P.D., 1972). However, up to today, there is no well-defined mixture distribution that is easy to work with when studying the highly leptokurtic distribution of daily log-return data from the financial market. For instance, many stock indices and commodity time series have kurtosis of more than 10. One can easily collect 90-year history of the daily log-returns of DJIA[1], which has kurtosis of more than 20. There is no good distribution other than the stable distribution

[1]For instance, download from http://finance.yahoo.com/

to describe such fat-tail data. But due to the infinite moments, the tails of the stable distribution have to be truncated.

In this paper, a Poisson subordinated distribution is proposed to fill the gap. It is easy to calculate; the summation converges quickly; analytic formula exists for all moments. Hopefully with these nice features, it is a good tool for the financial professionals to describe the kind of data they need to manipulate everyday.

I will use the acronym "PSD" for the Poisson subordinated distribution in this paper. The formula developed in this paper have been validated by GNU Maxima and R. Calculations and charts are generated by a library I wrote in R in both the double precision and MPFR (Multiple Precision Float Reliable).

2 The Development of PSD

2.1 Basic Notations

First of all, the basic notations are defined in this section. Additional investigations about the distribution will be elaborated in Section 2.2.

The normal distribution is

$$N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{1}$$

$$\Phi_N(x;\mu,\sigma^2) = \frac{1}{2}[1+\mathrm{erf}(\frac{x-\mu}{\sqrt{2}\sigma})] \tag{2}$$

where μ is the mean and σ is the volatility. In order to introduce skewness into PSD, we will use the skew normal distribution [2] with no shift, which is defined as

$$SN(x;\sigma^2,a) = 2\,N(x;0,\sigma^2)\,\Phi_N(a\,x;0,\sigma^2). \tag{3}$$

In Appendix B, I will present an alternative method of introducing skewness using shifted normal distribution. Mathematically it is simpler than skew normal distribution. The form of moments is very similar. But it has slightly less capability of generating skewness when the tails are close to normal. The reader can consult the Appendix and the companion R library.

[2] http://en.wikipedia.org/wiki/Skew_normal_distribution

The Poisson distribution is

$$Q(k,\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}, \tag{4}$$

where λ is the expected number of occurrences during the given interval and k is the number of occurrences of an event. Think of k as the magnitude of "earthquakes" in the financial market. It is obvious that

$$\sum_{k=0}^{\infty} Q(k,\lambda) = 1. \tag{5}$$

The subordination works as following. For a given financial time series, e.g., the daily log-returns of DJIA, we assume there is a unit volatility σ associated with it. Everyday we throw the dice [3] according to the Poisson distribution and get k with a certain probability $Q(k,\lambda)$. On that day, the actual volatility is determined by the scaling formula:

$$\sigma_k = \sigma\,(k+1)^{\alpha}(1+\gamma)^k \tag{6}$$

where α and γ are two scaling factors that amplify the unit volatility to the actual volatility. Both α and γ tend to be between 0 and 1. The actual volatility is fed into the skew normal distribution subordinatee (Equation 3) to produce the random move for that day. The skewness factor is denoted as β, which tends to be a small fraction between $-\sqrt{2/\pi}$ and $\sqrt{2/\pi}$ (0.798).[4] Therefore, the probability distribution function (PDF) of the combined process is

$$P(x;\sigma,\alpha,\gamma,\beta,\lambda) = \sum_{k=0}^{\infty} Q(k,\lambda)SN(x;\sigma_k^2,\frac{\beta}{\sqrt{2/\pi-\beta^2}}) \tag{7}$$

where we can attribute the PDF as the sum of all k-th terms:

$$P(x;\sigma,\alpha,\gamma,\beta,\lambda) = \sum_{k=0}^{\infty} P^{(k)}(x;\sigma,\alpha,\gamma,\beta,\lambda) \tag{8}$$

$$P^{(k)}(x;\sigma,\alpha,\gamma,\beta,\lambda) = Q(k,\lambda)SN(x;\sigma_k^2,\frac{\beta}{\sqrt{2/\pi-\beta^2}}). \tag{9}$$

[3] This naive picture can't explain the conditional heteroscedasticity, which will require more sophisticated modeling within the Poisson distribution framework, or integrating PSD with the GARCH model.

[4] The form of $\frac{\beta}{\sqrt{2/\pi-\beta^2}}$ is so chosen that the first moment is proportional to $\beta\,\sigma$.

In some applications such as tail index and numeric computation, the closed form of $d\,P(x)/dx$ could also be of importance:

$$\frac{d\,P}{dx}(x;\sigma,\alpha,\gamma,\beta,\lambda) = -\sum_{k=0}^{\infty} P^{(k)}(x;\sigma,\alpha,\gamma,\beta,\lambda)\left(\frac{x}{\sigma_k^2}\right) + Q(k,\lambda)\sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{2/\pi-\beta^2}}\left(\frac{\beta}{\sigma_k}\right) SN(\sqrt{\frac{2}{\pi}}\frac{x}{\sqrt{2/\pi-\beta^2}};\sigma_k^2,0). \quad (10)$$

In a similar fashion, the cumulative distribution function (CDF), $\Phi(x) = \int_{-\infty}^{x} P(x)dx$, is

$$\Phi(x;\sigma,\alpha,\gamma,\beta,\lambda) = \sum_{k=0}^{\infty} Q(k,\lambda)\,[\frac{1}{2} + \frac{1}{2}\mathrm{erf}(\frac{x}{\sigma_k\sqrt{2}}) - 2\,T(\frac{x}{\sigma_k}, \frac{\beta}{\sqrt{2/\pi-\beta^2}})] \quad (11)$$

where $T(h,a)$ is Owen's T function.

As a starter, we can easily check the value of the PDF at $x = 0$:

$$P(0;\sigma,\alpha,\gamma,\beta,\lambda) = \frac{1}{\sqrt{2\,\pi}\,\sigma}e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{(k+1)^{\alpha}\,k!\,(\gamma+1)^k} \quad (12)$$

This summation can be calculated numerically; therefore, is valuable in validating numerical implementation. Using the $\mathcal{L}$ notation of Equation 55, we have:

$$P(0;\sigma,\alpha,\gamma,\beta,\lambda) = \frac{1}{\sqrt{2\,\pi}\,\sigma}e^{-\frac{\lambda\,\gamma}{\gamma+1}}\,\mathcal{L}_{-\alpha}\left(\frac{\lambda}{\gamma+1}\right) \quad (13)$$

Furthermore, when $\beta = 0$, $P(0;\sigma,\alpha,\gamma,0,\lambda)$ describes the peak value of PDF of a symmetric PSD. This is our first encounter of the Lihn function $\mathcal{L}_N(x)$. More details can be found in Section 6.

2.2 Some Properties and Investigations

In this subsection, we will discuss several obvious properties of PSD, mainly about the PDF in Equation 7. The reader can obtain some intuitions about this distribution through these exercises.

2.2.1 Reduced to a Normal Distribution

The distribution is reduced to a normal distribution when $\alpha = \gamma = \beta = 0$ since σ_k is simplified to σ:

$$P(x; \sigma, 0, 0, 0, \lambda) = N(x; 0, \sigma^2). \tag{14}$$

This provides us a good starting point with respect to the Central Limit Theorem.

2.2.2 $\lambda = 1$

When $\lambda = 1$, $Q(k, \lambda)$ is simplified to $1/k!$. This allows us to remove some of the noises in the subsequent analytic solutions. This is what will be used in all numerical calculations throughout this paper unless mentioned else.

2.2.3 Symmetric When $\beta = 0$

When $\beta = 0$, the distribution is symmetric. When we study the properties of PSD other than skewness, it is easier to gain intuition under such condition since we have one less variable to worry about. Figure 1 demonstrates the various shapes of symmetric distribution and how the shapes of PDF change with increasing α and γ in the range of typical financial applications (with $\lambda = 1, \beta = 0$). When $\gamma = 0$, α alone generates moderately leptokurtic tails. When a positive γ is present, the kurtosis starts to rise rapidly. More numeric results will be presented later when kurtosis is studied in greater details.

When α is in the range of 0.66 to 0.86 and γ is zero, PSD produces Pareto-like tails, by which we mean the log PDF is linear when x is large. However, numeric simulation shows that only a particular α (˜0.62) produces a precise Pareto-tail condition. More numeric results will be presented later.

2.2.4 Asymmetric When $\beta \neq 0$

When β is not zero, the distribution is skewed. All the odd moments are not zero when β is not zero. The mean of PSD is shifted in proportion to β, $\mu_1 \propto \beta$. The skewness is positive when $\beta > 0$ and negative when $\beta < 0$. The skewness can be quite large for some β's as demonstrated in Figure 2.

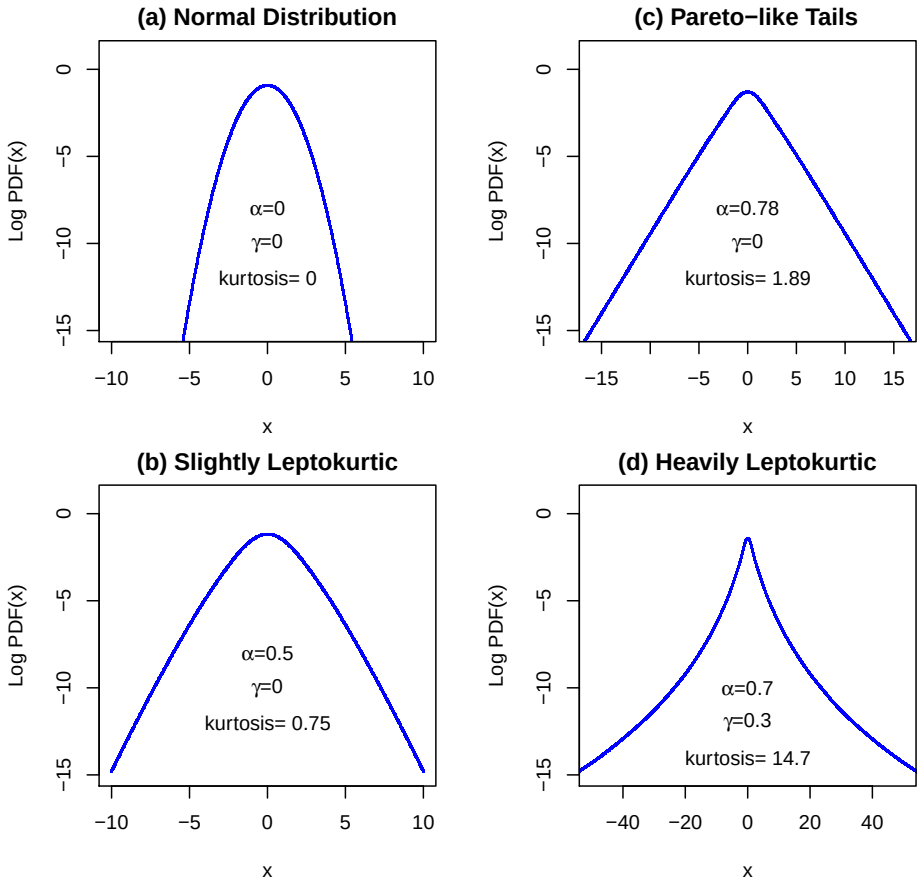

Figure 1: Demo of log PDF(x) of various symmetric shapes. Notice that kurtosis can become very large.

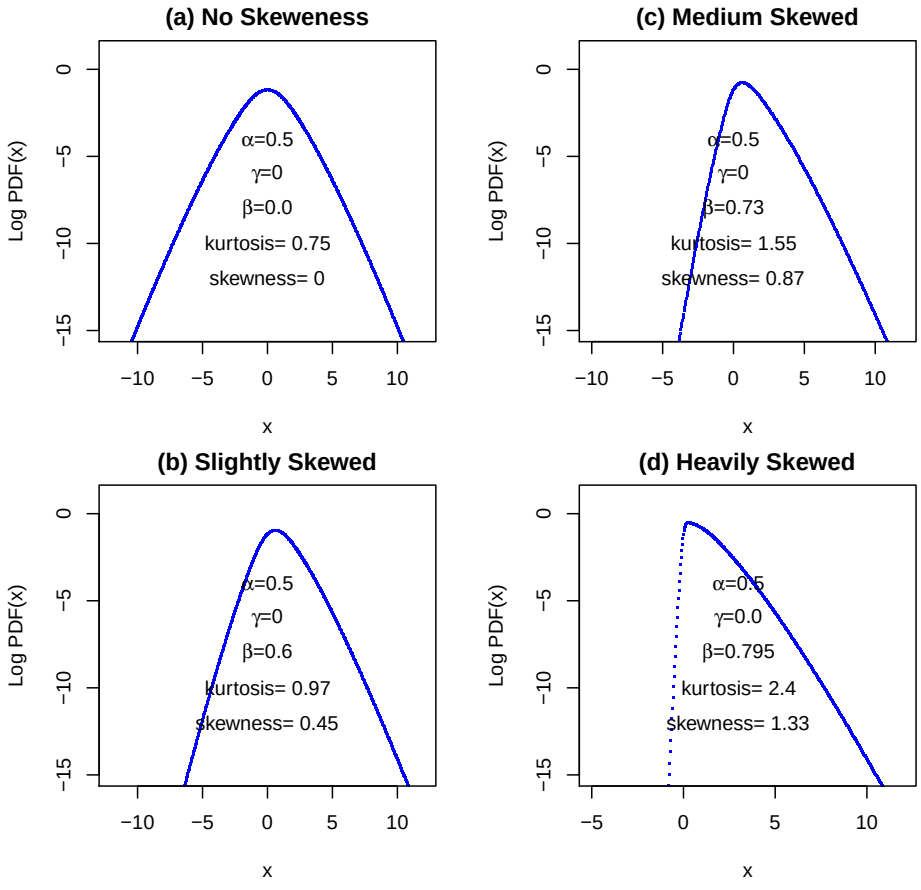

Figure 2: Demo of log PDF(x) of various skewed shapes. Notice that the shape can become very lopsided.

2.2.5 The Structure of σ_k

We now take a look at the structure of σ_k. This is the core of the PSD that determines the shape of the distribution. The relation between σ_k and k is determined by the choices of α and γ in $(k+1)^\alpha(1+\gamma)^k$. The former (α) allows $\log \sigma_k$ to scale at the rate of $\log(k+1)$ while the later (γ) linearly with k. Therefore, α generates moderate tails while γ generates very prominent tails. Figure 3 is an illustration of the log contribution of the k-th item of PDF(x) when $\alpha = 0.7$, $\gamma = 0.3$. The contour shows the level of $\log(P^{(k)}(x;\sigma,\alpha,\gamma,\beta,\lambda))$ in Equation 7. The Poisson sum converges quickly within 20 terms even for a very heavy tail. This allows very efficient numeric computation of PSD, which has been implemented in R.

3 The Statistics

The moment generating function $G(t) = \int_{-\infty}^{\infty} e^{tx}\, P(x)\, dx$ is

$$G(t;\sigma,\alpha,\gamma,\beta,\lambda) = 2\sum_{k=0}^{\infty} Q(k,\lambda) e^{\sigma_k^2 t^2/2}\, \Phi_N(\sqrt{\frac{\pi}{2}}\beta\, t\, \sigma_k; 0, 1) \tag{15}$$

by which the n-th moment is

$$\mu_n = \frac{d^n}{dt^n} G(t)|_{t=0}. \tag{16}$$

Similarly, the characteristic function $\mathcal{C}(t) = G(it)$ is

$$\mathcal{C}(t;\sigma,\alpha,\gamma,\beta,\lambda) = 2\sum_{k=0}^{\infty} Q(k,\lambda) e^{-\sigma_k^2 t^2/2}\, \Phi_N(i\sqrt{\frac{\pi}{2}}\beta\, t\, \sigma_k; 0, 1). \tag{17}$$

The separation between real and imaginary parts in $\mathcal{C}(t)$ is quite clear since

$$\Phi_N(i\sqrt{\frac{\pi}{2}}\beta\, t\, \sigma_k; 0, 1) = \frac{1}{2} + \frac{i}{2}\,\mathrm{erfi}(\frac{\sqrt{\pi}\,\beta\, t\, \sigma_k}{2}), \tag{18}$$

where erfi(x) is the imaginary error function.

It is interesting to note that if $P(x)$ represents the log-capital distribution of market constituents, then $G(t)$ is the mean market capitalization of the market, which should always exist in PSD.

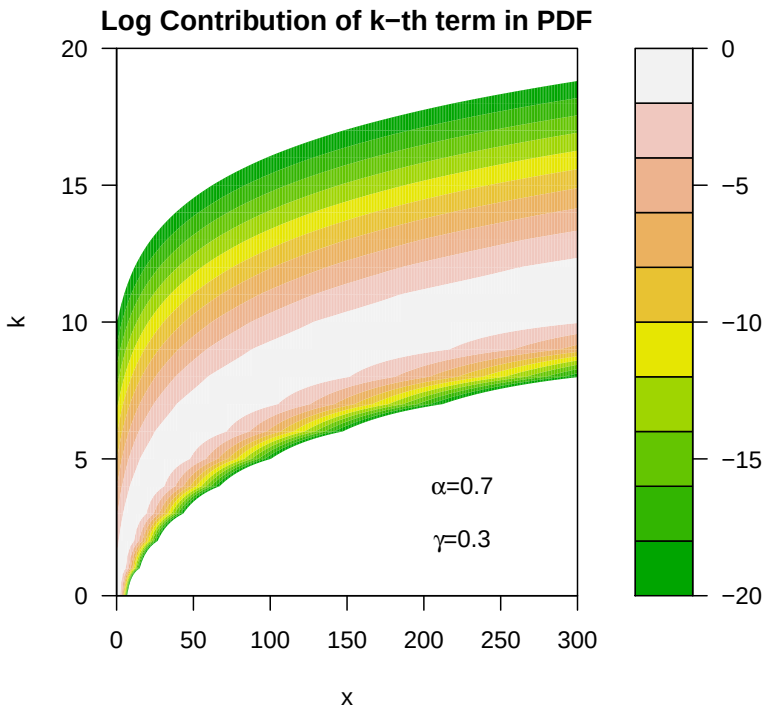

Figure 3: Log contribution of the k-th item of PDF(x) when $\alpha = 0.7$, $\gamma = 0.3$. The contour shows the level of $\log(P^{(k)}(x; \sigma, \alpha, \gamma, \beta, \lambda))$ in Equation 7. The Poisson sum converges quickly within 20 terms even for a very heavy tail. This allows very efficient numeric computation of PSD, which has been implemented in R.

We shall attempt to carry out analytic solutions for the first 4 moments. We are especially interested in the closed form of skewness and kurtosis, if possible. It should be pointed out before we proceed that (a) the odd moments are non-zero only when $\beta \neq 0$; (b) the n-th moment is proportional to $(\sigma_k)^n$.

3.1 The Mean

The first moment, the mean, is

$$\mu_1 = \beta\,\sigma\,e^{-\lambda}\sum_{k=0}^{\infty}\frac{(k+1)^{\alpha}\lambda^k(1+\gamma)^k}{k!} \tag{19}$$

which is predominantly generated by the presence of β. Equation 19 has known analytic solutions when α is a positive integer or zero. Using the notation of Equation 55, we have:

$$\mu_1 = \beta\,\sigma\,e^{\lambda\gamma}\,\mathcal{L}_{\alpha}(\lambda(1+\gamma)). \tag{20}$$

The reader may want to cross-reference Section 6 for more details on the $\mathcal{L}_N(x)$ notation.

3.2 The Variance

The second moment is

$$\mu_2 = \sigma^2\,e^{-\lambda}\sum_{k=0}^{\infty}\frac{(k+1)^{2\alpha}\lambda^k(\gamma+1)^{2k}}{k!}. \tag{21}$$

Equation 21 has known analytic solutions when 2α is a positive integer or zero. For instance, in the typical range of α between 0 and 1, the closed form is available when $\alpha = 0.5$. Using the $\mathcal{L}_N(x)$ notation, we have:

$$\mu_2 = \sigma^2\,e^{\lambda\gamma(\gamma+2)}\,\mathcal{L}_{2\alpha}(\lambda(\gamma+1)^2). \tag{22}$$

This leads to the formula of variance

$$var = \mu_2 - \mu_1^2 = \sigma^2\,e^{\lambda\gamma(\gamma+2)}\,\mathcal{L}_{2\alpha}(\lambda(\gamma+1)^2) - \beta^2\,\sigma^2\,e^{2\lambda\gamma}\,\mathcal{L}_{\alpha}(\lambda(1+\gamma))^2. \tag{23}$$

To illustrate the relation between the unity volatility (σ) and the actual volatility ($\sigma_{act} = var^{1/2}$), we studied the simplified case where $\beta = 0$ and $\lambda = 1$. Now

$$\frac{\sigma_{act}}{\sigma} = e^{\gamma(\frac{\gamma}{2}+1)}\,\mathcal{L}_{2\alpha}((\gamma+1)^2)^{\frac{1}{2}}. \tag{24}$$

Figure 4 shows the contour plot of the actual volatility ratio, Equation 24. We can examine the ratio of σ_{act}/σ with several standard parameter sets. When $\alpha = 1, \gamma = 0.25$, the ratio is 3.8. Majority of the contribution comes from the $\mathcal{L}_{2\alpha}$ term since the exponential term is not much more than one with a small γ.

It is also interesting to estimate how quick the finite integral of moments converges to theoretical value. Since the following integral is well known:

$$\int_{-x}^{x} x^2\, e^{-x^2/2}\, dx = \sqrt{2\pi}\,\mathrm{erf}\left(\frac{x}{\sqrt{2}}\right) - 2\,x\,e^{-x^2/2}, \tag{25}$$

we can study the convergence of variance in a reasonably elegant form with $\sigma = 1, \beta = 0$, that is, when the distribution is symmetric and the unity volatility is one. Let's define

$$\delta\mu_2(x;\alpha,\gamma,\lambda) = \mu_2(\sigma = 1, \beta = 0) - \int_{-x}^{x} x^2\, P(x;1,\alpha,\gamma,0,\lambda)\, dx \tag{26}$$

which is (the notation $\hat{\mathcal{L}}_{\alpha}^{(k)}$ is defined in Equation 56)

$$\begin{aligned}\delta\mu_2(x;\alpha,\gamma,\lambda) = e^{\lambda\gamma(\gamma+2)} &\sum_{k=0}^{\infty} \hat{\mathcal{L}}_{2\,\alpha}^{(k)}(\lambda\ (\gamma+1)^2)\left[1 - \mathrm{erf}\left(f_{(k)}(x;\alpha,\gamma)\right)\right] \\ &+ \sqrt{\frac{2}{\pi}}\,x\,e^{\lambda\gamma} \sum_{k=0}^{\infty} \hat{\mathcal{L}}_{\alpha}^{(k)}(\lambda\ (\gamma+1))\, e^{-f_{(k)}(x;\alpha,\gamma)^2} \\ &\text{where } f_{(k)}(x;\alpha,\gamma) = \frac{x}{\sqrt{2}}\,(k+1)^{-\,\alpha}\,(\gamma+1)^{-k}.\end{aligned} \tag{27}$$

The ratio $\delta\mu_2(x)/\mu_2$ describes the error of computing the variance by finite integral from $-x$ to x. As α and γ get bigger, the tails are more prominent, therefore, the error gets larger. The integral needs to extend to large multiple, e.g. 1000, of the unity volatility to yield high precision. This demonstrates how difficult it is to estimate tail risk with finite range of data.

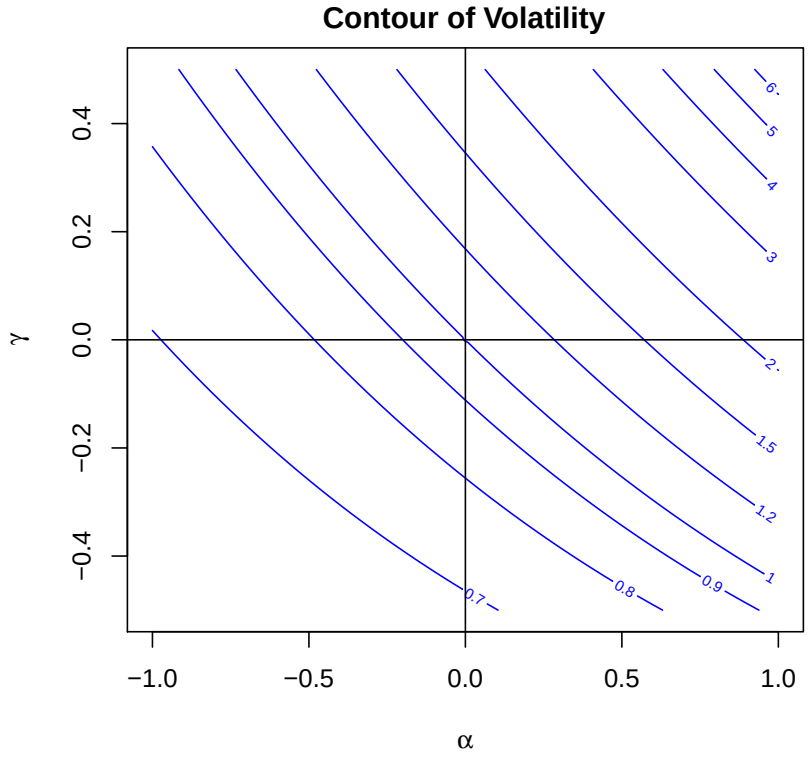

Figure 4: Volatility Contour of σ_{act}/σ as a function of α and γ while $\beta = 0$, $\lambda = 1$.

3.3 General Form of the N-th Moment

As we work through the first few moments, we begin to see a pattern emerging. The N-th moment has the general form of:

$$\mu_N = g_N(\beta)\,\sigma^N\,e^{-\lambda}\sum_{k=0}^{\infty}\frac{(k+1)^{N\,\alpha}\,\lambda^k\,(\gamma+1)^{N\,k}}{k!} \tag{28}$$

$$= g_N(\beta)\,\sigma^N\,e^{\lambda((\gamma+1)^N-1)}\,\mathcal{L}_{N\alpha}(\lambda(\gamma+1)^N). \tag{29}$$

where $g_N(\beta)$ is an N-th order polynomial of β. $g_N(\beta)$ can be symbolically [5] calculated from $\frac{d^n}{dt^n}G(t)|_{t=0}$ by setting $\alpha=\gamma=0,\sigma=\lambda=1$. Below listed $g_N(\beta)$ for the first six moments:

$$g_1(\beta) = \beta\,, \tag{30}$$

$$g_2(\beta) = 1\,, \tag{31}$$

$$g_3(\beta) = 3\beta - \frac{\pi}{2}\beta^3\,, \tag{32}$$

$$g_4(\beta) = 3\,, \tag{33}$$

$$g_5(\beta) = 15\beta - 5\pi\,\beta^3 + \frac{3\pi^2}{4}\beta^5\,, \tag{34}$$

$$g_6(\beta) = 15\,. \tag{35}$$

3.4 The Skewness

The third moment is studied here in the context of skewness. Skewness in terms of raw moments is quite complicated:

$$skewness = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2-\mu_1^2)^{3/2}}. \tag{36}$$

However, we can study a special case, $\{\alpha=1,\gamma=0,\lambda=1\}$ to reveal the general structure of the skewness. The skewness is reduced to:

$$skewness(\beta) = \frac{15\,\beta - 7.562\,\beta^3}{(5-4\,\beta^2)^{3/2}} \tag{37}$$

[5] For instance, my symbolic program is written in GNU Maxima.

which is a monotonically increasing function of β when $|\beta| < 1$. This range of β covers most use cases in real world. [6]

Figure 5 shows the contour plot of skewness when $\alpha \neq 0, \gamma = 0$. The

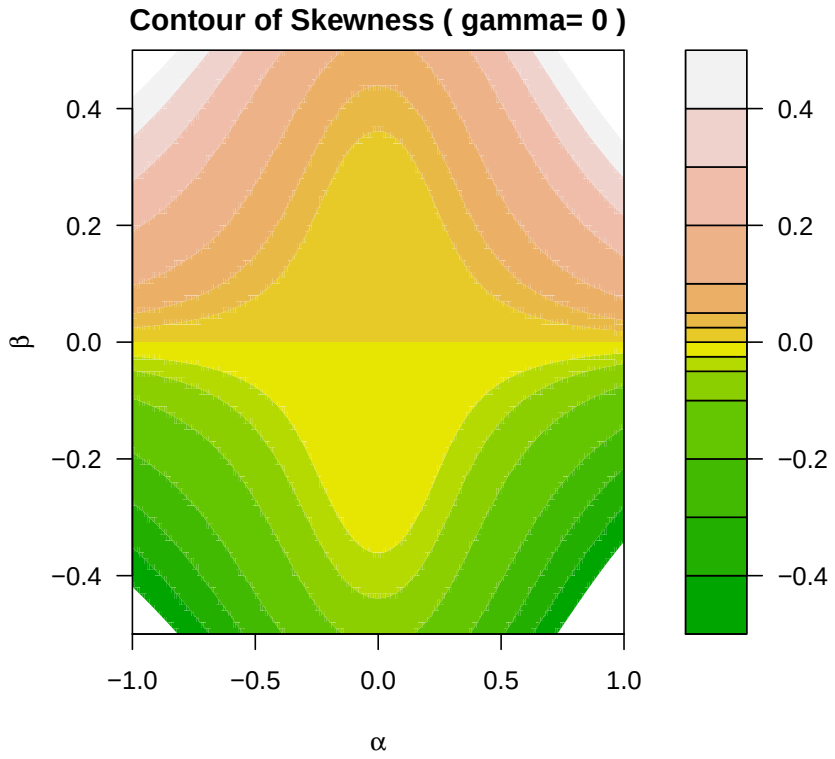

Figure 5: Skewness Contour when $\gamma = 0$. Notice that the presence of α enhances the skewness.

center of zero skewness is at where the normal distribution is. The presence of α enhances the skewness. When γ is not zero, the center of zero skewness will be shifted.

[6] The skewness of SN is $\frac{0.429\,\beta^3}{(1-\beta^2)^{\frac{3}{2}}}$. Note that the Skew Normal distribution approaches maximum skewness as β approaches $\sqrt{2/\pi}$ (0.798).

3.5 The Kurtosis

The fourth moment is studied here in the context of excess kurtosis when $\beta = 0$. This simplified version will give us better intuition.

$$kurt = \left(\frac{\mu_4}{\mu_2^2}\right)_{\beta=0} - 3 = \frac{3\, e^{\lambda} \sum_{k=0}^{\infty} \frac{(k+1)^{4\,\alpha}\, \lambda^k\, (\gamma+1)^{4\,k}}{k!}}{\left(\sum_{k=0}^{\infty} \frac{(k+1)^{2\,\alpha}\, \lambda^k\, (\gamma+1)^{2\,k}}{k!}\right)^2} - 3 \tag{38}$$

Using the $\mathcal{L}_N(x)$ notation, we have:

$$kurt = \frac{3\, e^{\lambda\, \gamma^2\, (\gamma+2)^2}\, \mathcal{L}_{4\alpha}(\lambda\, (\gamma+1)^4)}{\mathcal{L}_{2\alpha}(\lambda(\gamma+1)^2)^2} - 3. \tag{39}$$

When studying financial data, the kurtosis plays an important role since most of the log-return time series are highly leptokurtic if the data is available for long history (say, greater than 30 years). One of the basic criteria concerning any probability distribution candidate suitable for financial applications is whether the distribution is capable of handling very high kurtosis. For instance, in the analysis of daily log-returns of DJIA and commodities, the excess kurtosis can go as high as 10-20. Distributions with Pareto tails usually have excess kurtosis of less than 3.0. In the skew lognormal cascade distribution that I have studied before (Lihn 2008), it is extremely complicated to generate high kurtosis while being able to fit variance well. PSD is far superior to the skew lognormal cascade distribution in this regard. Let's take a look at the kurtosis generating capability of PSD.

Figure 6 shows the contour plot of kurtosis for various α and γ, assuming $\beta = 0, \lambda = 1$ for simplicity. With the illustrated range of $\alpha = 0 \sim 1$, $\gamma = 0 \sim 0.5$, PSD covers the kurtosis from zero up to 100. This flexibility should make PSD useful for heavily leptokurtic data.

3.6 Tail Index

We now study the tail index in terms of the reciprocal of hazard function, according to Gabaix 2009, Section 4.2. The tail index ξ is defined as

$$\xi = \lim_{x\to\infty} \xi(x), \text{where}\, \xi(x) = \frac{d}{dx} \frac{1-\Phi(x)}{P(x)}. \tag{40}$$

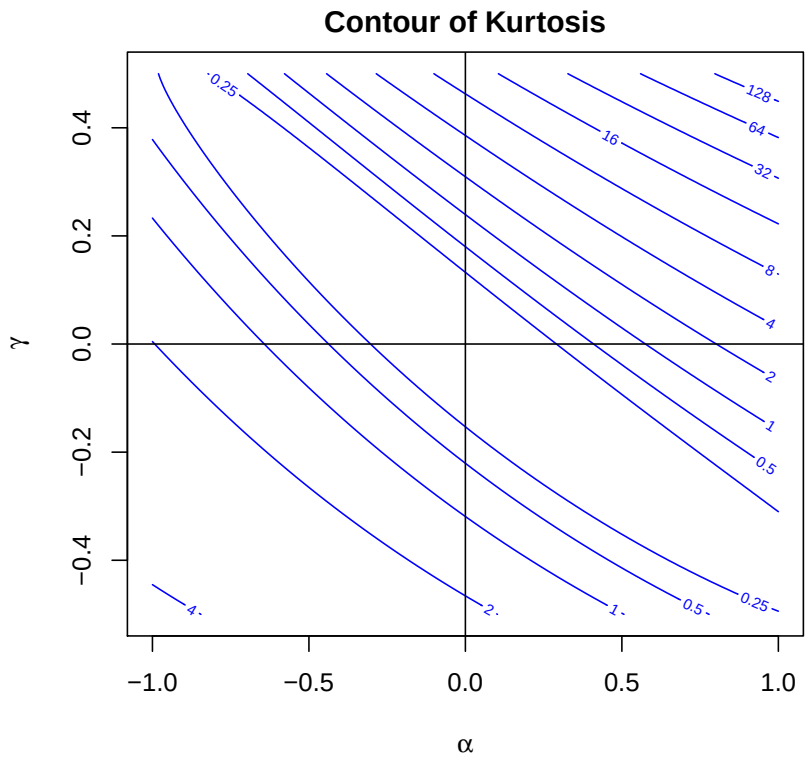

Figure 6: Kurtosis Contour.

$\xi(x)$ can be simplified to a more computationally friendly form:

$$\xi(x) = -1 - \frac{1 - \Phi(x)}{P(x)^2} \frac{d\,P(x)}{d\,x}. \tag{41}$$

Since $P(x)$, $\Phi(x)$, and $\frac{d\,P(x)}{dx}$ have concise summation formula to work with (See Section 2.1), we can compute the tail index numerically. The challenge here is that very high numeric precision can not be obtained without using MPFR (multiple precision floating number) which is very time-consuming, and the numeric error in the tails ($|x| \to \infty$) is not easy to estimate.

Numeric simulation presented in Figure 7 indicates that $\xi \sim 0$ when $\alpha \sim 0.7, \gamma = 0$. When α and γ are both small, ξ is negative. When α gets larger or when γ is positive, ξ is positive and finite. This puts this distribution in the category of "regular" in the extreme value theory (EVT).

3.7 Pareto Tail

Pareto distribution occupies a very special position in economics, physics, and mathematics since it presents an environment of scale invariance. This is illustrated by the elegant concept of 80/20 principle (See Gabaix 2009 for examples).

When we look at both the distribution and the random variable in the logarithm scale, Pareto distribution is no more than a straight line on the chart [7] , that is,

$$\lim_{x\to\infty} \frac{d^2}{dx^2} \log P(x) = 0 \tag{42}$$

which is equivalent to

$$\lim_{x\to\infty} \Delta(x) = \frac{1}{P(x)} \frac{d^2\,P(x)}{dx^2} - \frac{1}{P(x)^2} \left(\frac{d\,P(x)}{dx}\right)^2 = 0 \tag{43}$$

So now the study of Pareto tail is a matter of determining which α yields $\Delta(x) \to 0$ for large x. Under the simplified condition of $\sigma = 1, \gamma = 0, \beta = 0, \lambda = 1$, we have:

$$\Delta(x,\alpha) = \frac{x^2\,\widetilde{\mathcal{L}}_{-5\,\alpha}(x,\alpha) - \widetilde{\mathcal{L}}_{-3\,\alpha}(x,\alpha)}{\widetilde{\mathcal{L}}_{-\alpha}(x,\alpha)} - x^2\,\frac{\widetilde{\mathcal{L}}_{-3\,\alpha}(x,\alpha)^2}{\widetilde{\mathcal{L}}_{-\alpha}(x,\alpha)^2} \tag{44}$$

[7] Assume y is price return and $x = \log(y)$ is the log return. The Pareto-like PDF expressed in y, $P(y) = y^{-\gamma}$, is equivalent to the PDF expressed in x, $P(x) = e^{(1-\gamma)\,x}$. It is obvious that, $\log P(x) = (1-\gamma)\,x$, is a straight line in x; and $\frac{d^2}{dx^2} \log P(x) = 0$.

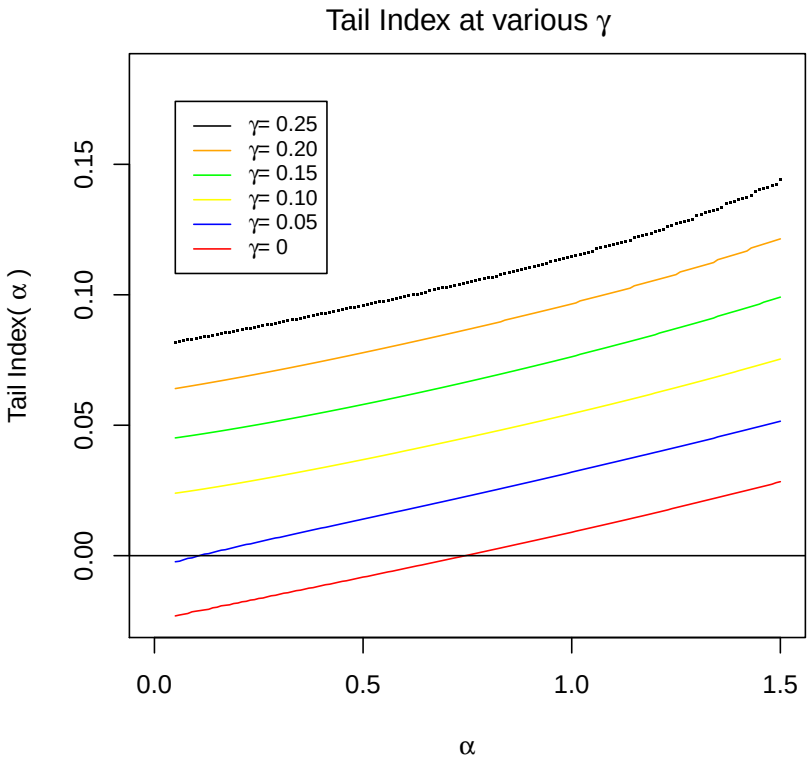

Figure 7: Tail Index computed at $P(x; \alpha, \gamma) = e^{-20}$. The larger the index is, the fatter the tails are. The result is consistent with the fact that α produces moderate tails while γ produces prominent tails.

where the $\widetilde{\mathcal{L}}$ notation is used:

$$\widetilde{\mathcal{L}}_p(x,\alpha) = e^{-1}\sum_{k=0}^{\infty}\phi\left(\frac{x}{(k+1)^{\alpha}}\right)\frac{(k+1)^p}{k!}, \tag{45}$$

and $\sqrt{2\pi}\,\phi(x) = e^{-\frac{1}{2}x^2}$. We can inspect the contour plot of $\Delta(x,\alpha)$, as shown in Figure 8, to see where the Pareto tail may be. Numeric simulation indicates that the Pareto tail occurs at $\alpha \sim 0.624$. However, the high-precision computation is quite challenging as the PDF becomes very small ($P(200) \cong 10^{-132}$). Such small PDF won't be observed in any real-world data and is calculated simply for mathematical curiosity.

3.8 Value At Risk (VAR) Related Topics

In this section, we will investigate the topics related to Value at Risk (VAR). When $\Phi(x)$ represents the CDF of the daily log-returns of a portfolio, the 1% daily VAR is

$$VAR = \exp(\Phi^{-1}(0.01)). \tag{46}$$

Therefore, the issue of calculating VAR in PSD is equivalent to finding $\Phi^{-1}(C)$. However, the fat tail tends to complicate the perceived risk. If the tail probability is not estimated correctly, most likely VAR will be underestimated. Figure 9 shows the contour of x where the CDF $\Phi(\sigma_{act}\,x;\alpha,\gamma) = 0.01$. This demonstrates the change of 1% VAR relative to standard deviation σ_{act} as α and/or γ increases. Figure 10 shows the contour of x where $\Phi(\sigma_{act}\,x;\alpha,\gamma) = 0.001$. This demonstrates the changes of 0.1% VAR relative to standard deviation as α and γ increases.

4 The Application To Financial Data

We will now apply this distribution to the log-return time series in exchange rates, commodities, volatility index, stock market indices, over long history. Long historical time series typically went through major events causing large disruptions in prices. Such disruptions were reflected in fat tails and large excess kurtosis (3-50). Tail events can devastate ignorant financial institutions and individual investors who use wrong kind of distributions to analyze risks. We will explore how to use this distribution to fit various sets of data, get intuition on the range of parameters, and have a feel of how good the distribution behaves in real world.

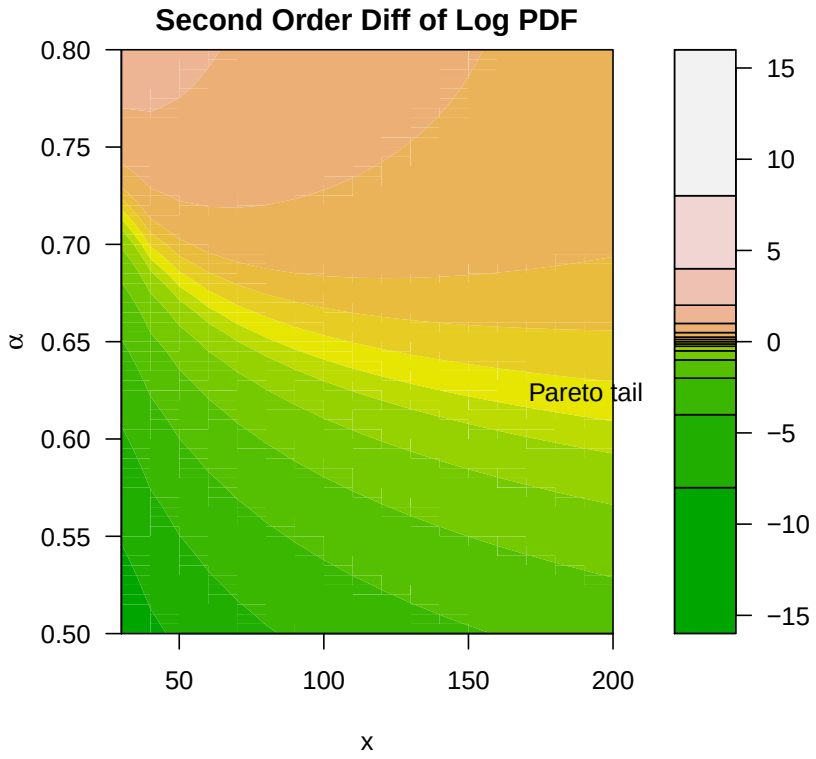

Figure 8: The contour of $\Delta(x, \alpha)$ multiplied by factor of 1000. The yellow strip illustrates the region closest to zero which is the condition of a Pareto tail. One can see that the Pareto tail occurs at $\alpha \sim 0.624$.

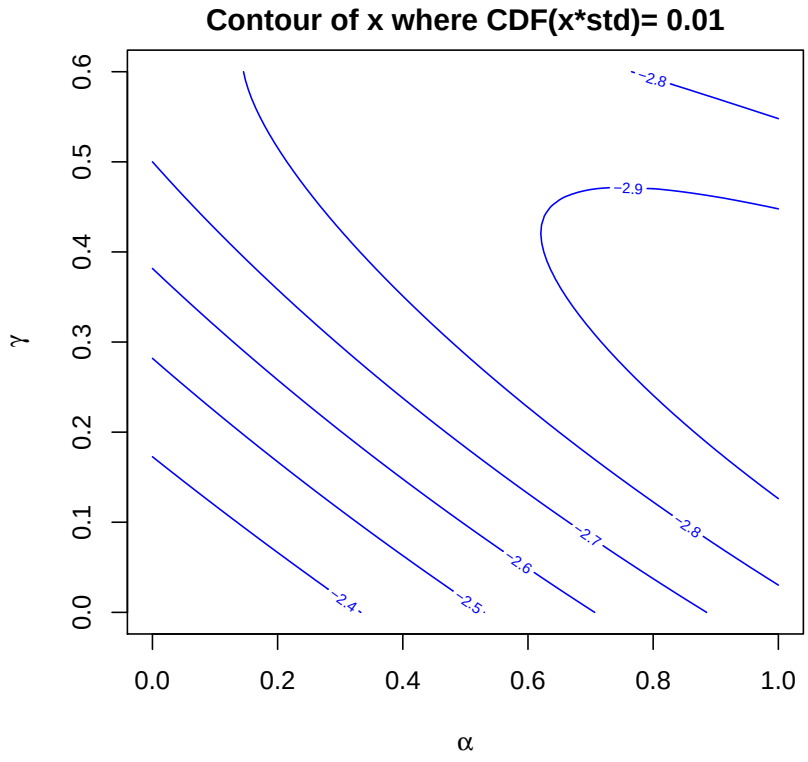

Figure 9: The contour of x where the CDF $\Phi(\sigma_{act}\, x; \alpha, \gamma) = 0.01$. This demonstrates the change of 1% VAR relative to standard deviation as α and/or γ increases. Notice that the increase is about 30% (from 2.3 times standard deviation to 2.9) which isn't particularly drastic (compared to 0.1% VAR). Also notice there are some fine structures when x gets above 2.8.

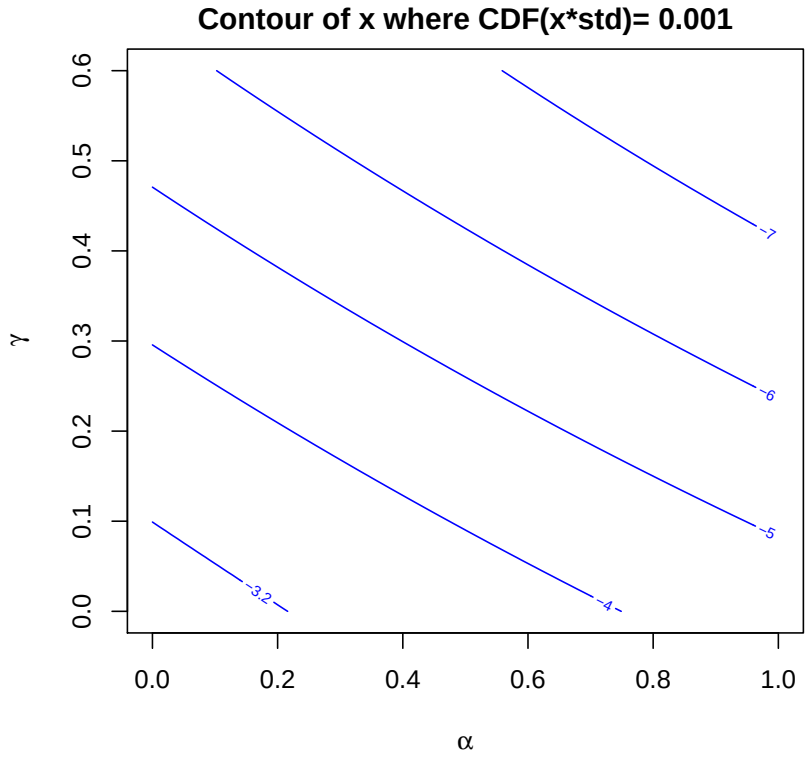

Figure 10: The contour of x where $\Phi(\sigma_{act}\, x; \alpha, \gamma) = 0.001$. This demonstrates the changes of 0.1% VAR relative to standard deviation as α and γ increases. One should notice that the increase is more than 100% (from 3 to 7.7) which is more drastic than the scenario of 1% VAR. The levels of the contour are very predictable.

4.1 Adding Location Parameter

When the PSD is applied to the real-world data, it is necessary to add the location parameter μ to offset the center of the distribution:

$$P'(x;\mu,\sigma,\alpha,\gamma,\beta,\lambda) = P(x-\mu_1-\mu;\sigma,\alpha,\gamma,\beta,\lambda) \tag{47}$$

Notice that μ_1 is subtracted out of $P'(x)$ so that $\langle x\rangle = \int x\,P'(x)dx = \mu$; and when $\mu = 0$, $\langle x\rangle = 0$.

4.2 Regression Methodology

The optimx and spg packages in R are used to perform nonlinear programming, which minimizes the diff function. The diff function is a least-mean-square combination of the following deviations between PSD fit and data.

1. Deviation of variance
2. Deviation of skewness
3. Deviation of kurtosis
4. Deviation of peak PDF
5. Deviation of PDF, $P(x)$ for x within 3 standard deviations
6. Deviation between QQ-plot and the 45° line

Each item can be given different weights to accommodate varying behaviors of the underlying data. Sometimes fitting moments are as good as fitting QQ-plot. But in other cases, one has to choose between better fits to kurtosis or QQ-plot. In these circumstances, weights can be used to influence optmix. Guards are provisioned in the program to confine α and γ between 0 and 1.

We will present the fits in a standard format of four charts: 1) PDF fit; 2) CDF fit; 3) log PDF fit; 4) QQ plot fit. The PDF and CFD fits show how good the theoretical distribution describes the peak of the population. The log PDF fit shows how good the distribution describes the tails. The QQ plot fit describes the quantile-to-quantile comparison between the theoretical distribution and observed data, which provides a stringent test on the theoretical distribution.

4.3 Swiss Franc (SZD/USD)

The first financial data set we will examine is the daily log-returns of Swiss Franc (SZD) to US dollar (USD) exchange rate from 1975 to 2008. The data is slightly leptokurtic with excess kurtosis slightly less than 3.0. Figure 11 shows the PSD fit for SZD/USD exchange rate. The tails of log PDF are very close to linear, therefore, the fit can be accomplished with zero γ. It is very impressive that the QQ-plot follows the 45° line precisely.

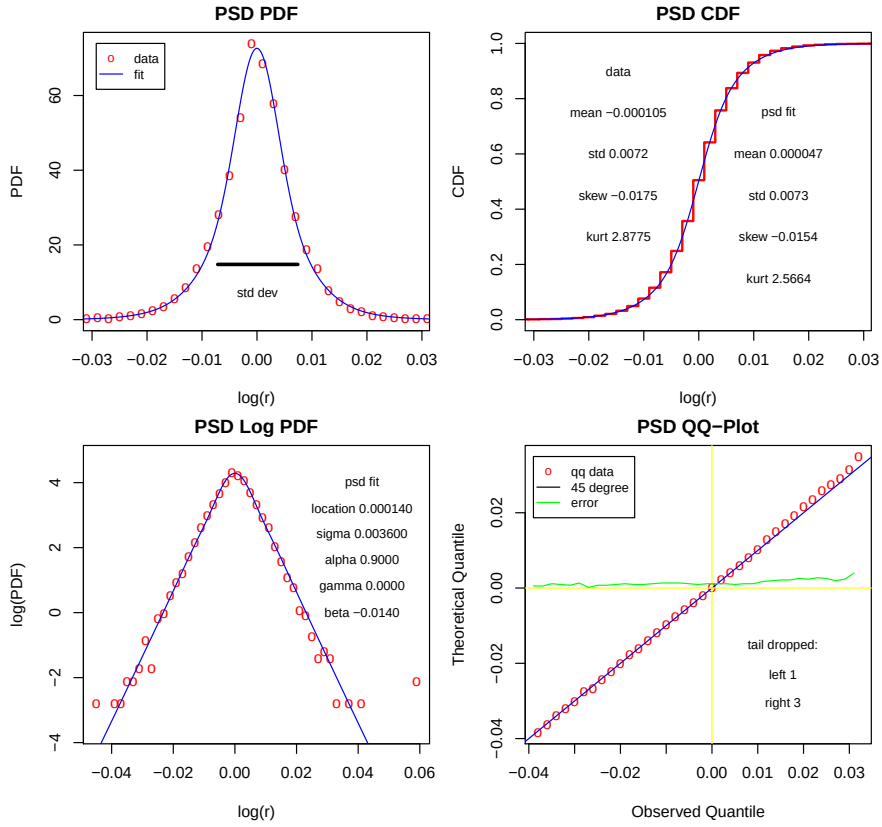

Figure 11: The daily log-returns of SZD/USD exchange rate from 1975 to 2008. The fit is overweight on fitting QQ-plot. The tails are fit very well except the few farthest points.

4.4 VIX: Volatility Index

Figure 12 is the fit for VIX from 1990 to 2011. The data is slightly leptokurtic with excess kurtosis slightly less than 3.0. The tails of log PDF bend slightly outward from linearity therefore requires a small γ to describe the shapes. it is well known that the log-returns of VIX is positively skewed, that is, volatility begets more volatility. This feature is described very well in the fit.

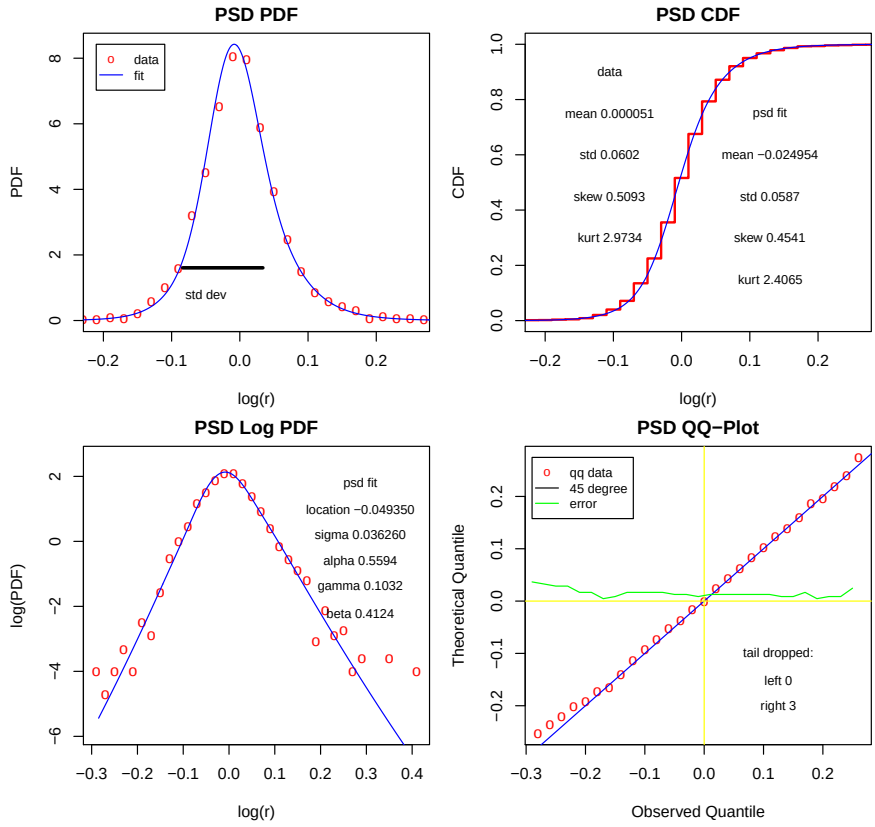

Figure 12: Daily log-return of VIX (1990-2011).

4.5 Gold

We will study gold prices from 1972 to 2009 as the representive of commodities and precious metals. Figure 13 is the fit of the log returns for gold. The data is highly leptokurtic with excess kurtosis of more than 10. In this fit, I've overweight the QQ-plot and underweight the kurtosis. The reason is that, for highly leptokurtic data set, the observed kurtosis is just an indication of the range the kurtosis is in. Its exact value should not be taken too seriously.

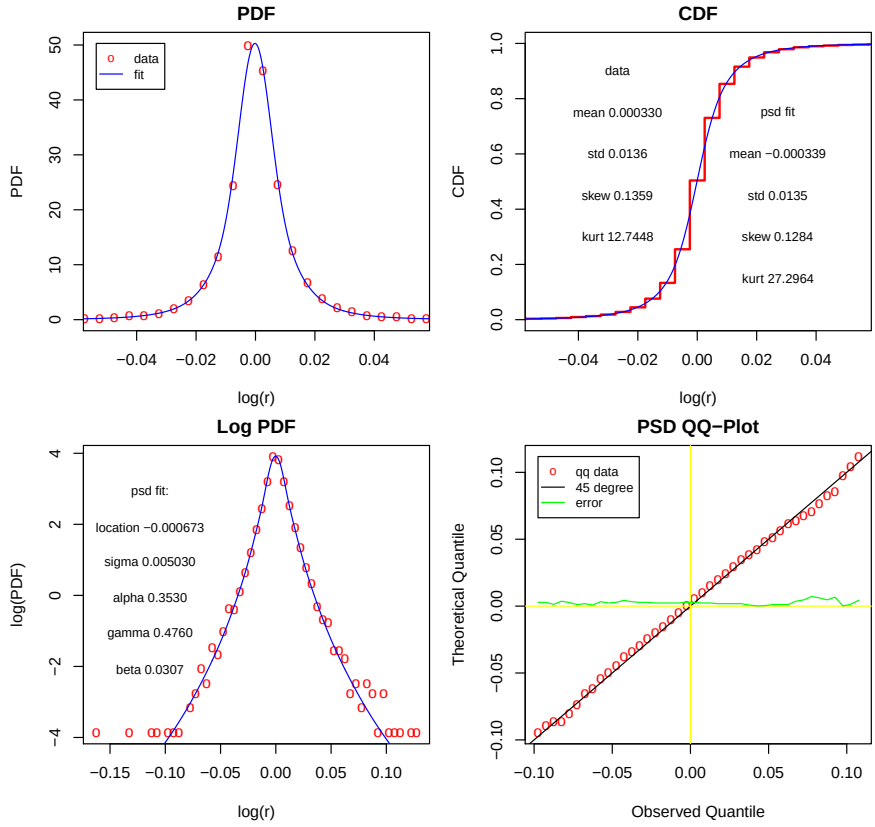

Figure 13: Gold data fit (1972-2009). Overweight on fitting QQ-plot; underweight on kurtosis. The QQ-plot is fit very well, but the implied kurtosis from the fit is twice that of the data.

4.6 R10Y: 10-Year Treasury Yield

Figure 14 is the fit for R10Y from 1962 to 2011. The data is highly leptokurtic with excess kurtosis of more than 10. Although its kurtosis is similar to that of gold, its shape of distribution is subtly different. This causes the fit to drift toward the high-end of α (~ 1.0) and a smaller γ (~ 0.2). This is different from gold's parameters of $\alpha \sim 0.35$ and $\gamma \sim 0.5$.

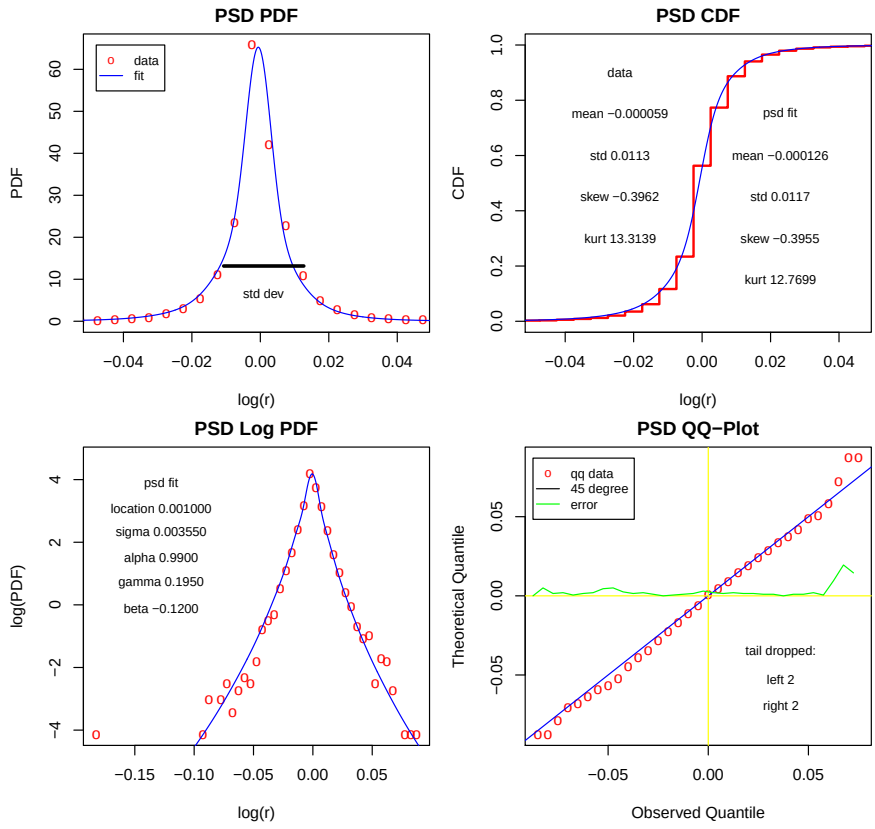

Figure 14: Daily log-return of 10-year Treasury yield (1962-2011).

4.7 DJIA: Dow Jones Industrial Average

Figure 15 is the fit for DJIA daily log-returns from 1928 to 2011. This is arguably one of the most difficult time series due to its very high kurtosis of 24. However, one can see that PSD fit handles it very well except the very far end of the tails in QQ-plot. The standard deviation, skewness, and kurtosis are spot-on in the fit. This is a very good testimony of the PSD.

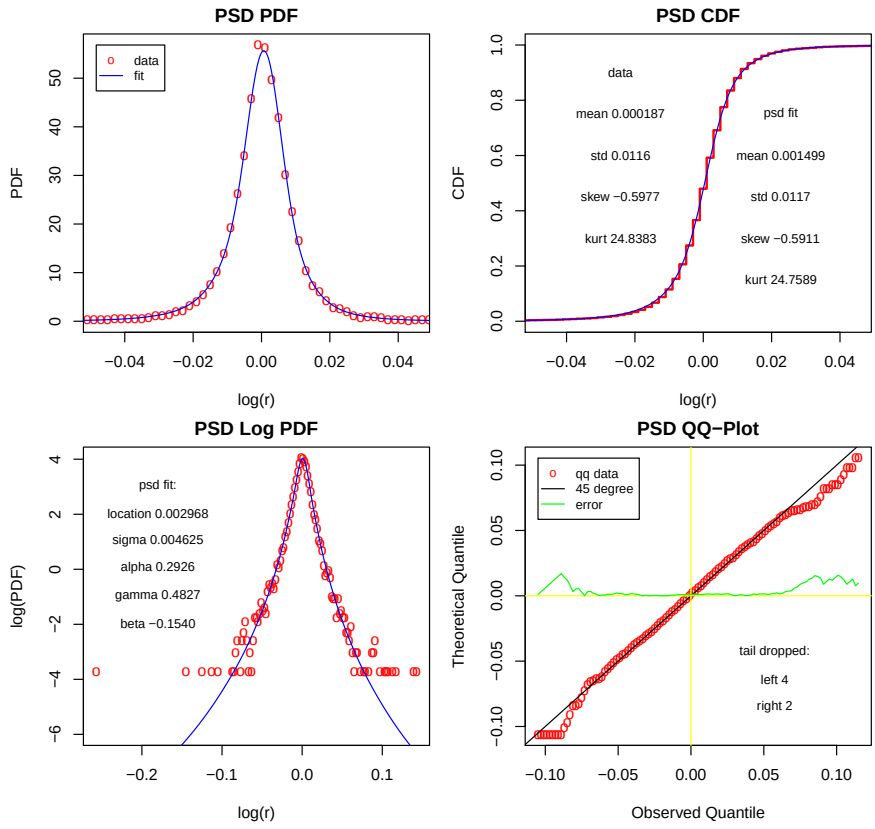

Figure 15: Daily log-returns of DJIA (1928-2011). PSD handles the fit very well except the very far end of the tails in QQ-plot. The standard deviation, skewness, and kurtosis are spot-on in the fit.

5 Summary

In this paper, I have shown a new kind of Poisson subordinated distribution. It is capable of fitting a wide range of leptokurtic financial data better than most known distributions. PDF, CDF, statistical moments and tail index can be computed through summation of Poisson series, which converges rapidly. It is very promising to be a useful statistical tool for the financial professionals.

6 Appendix A: Analytic Form of Certain Sums

The summations in the moments tend to follow a predictable format:

$$S(x,N)=\sum_{k=0}^{\infty}\frac{(k+1)^N\,x^k}{k!},x\neq 0. \tag{48}$$

Since we have $\frac{d}{dx}x^{(k+1)}=(k+1)\,x^k$ for $k=0..\infty$, it follows that

$$(k+1)^N\,x^k=[\frac{d}{dx}\,x]^N(x^k) \tag{49}$$

where the functional operator $[\frac{d}{dx}\,x]$ is to first multiply its argument by x, then differentiate the product by x. Therefore, we can eliminate the summation in Equation 48 and arrive at:

$$S(x,N)=[\frac{d}{dx}\,x]^N(e^x)=\mathcal{L}_N(x)\,e^x \tag{50}$$

where the new function $\mathcal{L}_N(x)$ is defined - which I call the ***Lihn function***. This allows us to derive the closed form of the summation through differentiations – When N is a positive integer, $\mathcal{L}_N(x)$ is an N-th order polynomial of x. For instance, the first few solutions of $\mathcal{L}_N(x)$ are

$$(x+1)\,\text{when}\,N=1, \tag{51}$$
$$(x^2+3x+1)\,\text{when}\,N=2, \tag{52}$$
$$(x^3+6x^2+7x+1)\,\text{when}\,N=3, \tag{53}$$
$$(x^4+10x^3+25x^2+15x+1)\,\text{when}\,N=4. \tag{54}$$

These analytic forms are valuable in verifying the correctness of numerical implementations. Since we often set $\lambda=1$ in this paper, the numeric value

of $\mathcal{L}_N(1)$ is of particular interest in our context, which is an integer series, {2, 5, 15, 52, 203, 877, 4140, ...}.

On the other hand, by combining Equations 48 and 50, we come up with the following equation:

$$\mathcal{L}_N(x) = e^{-x} \sum_{k=0}^{\infty} \frac{(k+1)^N x^k}{k!} \tag{55}$$

which extends $\mathcal{L}_N(x)$ to non-integer N ($N \in R$). In some places, it is also convenient to define

$$\hat{\mathcal{L}}_N^{(k)}(x) = e^{-x} \frac{(k+1)^N x^k}{k!}. \tag{56}$$

When N is a positive integer, we can reduce $\mathcal{L}_N(x)$ to the N-th order polynomial as shown above. When N is a negative integer, we can also explore the analytic solutions (but only with limited success). Since we have $x^{-1} \int_0^x x^k = (k+1)^{-1} x^k$ for $k = 0..\infty$, it follows that

$$(k+1)^N x^k = [x^{-1} \int_0^x]^{-N}(x^k) \text{ when } N < 0 \tag{57}$$

where the functional operator $[x^{-1} \int_0^x]$ is to first integrate its argument by x, then divide the integration by x. Therefore, we can eliminate the summation in Equation 48 and arrive at:

$$S(x, N) = [x^{-1} \int_0^x]^{-N}(e^x) = \mathcal{L}_N(x)\, e^x \text{ when } N < 0 \tag{58}$$

which we can attempt to derive closed form of the summation through integrations. $N = -1$ is easy to solve, which is $\mathcal{L}_{-1}(x) = x^{-1}(1 - e^{-x})$. When $N = -2$, it becomes somewhat complicated - $\mathcal{L}_{-2}(x) = (-\log(x) - \gamma(0, -x))\, x^{-1}\, e^{-x}$, where $\gamma(s, x)$ is the lower incomplete Gamma function. It is hard to move to $N = -3$. However, it should be pointed out that $\gamma(1, x) = 1 - e^{-x}$ indicating $\mathcal{L}_{N<0}(x)$ has a close relation to the mathematics of incomplete Gamma function. But it is beyond the interest of this paper.

7 Appendix B: Alternative Method to Introduce Skewness

In this appendix, an alternative way to incorporate skewness is explained. Here we don't use Skew Normal (SN) distribution. Instead, the skewness is

introduced by way of shifting the center of each normal distribution mixture, $N(x;\beta\sigma_k,\sigma_k^2)$. In doing so, the probability distribution function (PDF) of the combined process is

$$P(x;\sigma,\alpha,\gamma,\beta,\lambda)=\sum_{k=0}^{\infty}Q(k,\lambda)N(x;\beta\sigma_k,\sigma_k^2) \tag{59}$$

The moment generating function $G(t)=\int_{-\infty}^{\infty}e^{tx}\,P(x)\,dx$ is

$$G(t;\sigma,\alpha,\gamma,\beta,\lambda)=\sum_{k=0}^{\infty}Q(k,\lambda)e^{\beta\sigma_k t+\sigma_k^2 t^2/2} \tag{60}$$

The major difference is the form of the beta polynomials in the N-th moments, $g_N(\beta)$. Below listed are the beta polynomial $g_N(\beta)$ for the first six moments:

$$g_1(\beta)=\beta\,, \tag{61}$$

$$g_2(\beta)=1+\beta^2\,, \tag{62}$$

$$g_3(\beta)=3\beta+\beta^3\,, \tag{63}$$

$$g_4(\beta)=3+6\beta^2+\beta^4\,, \tag{64}$$

$$g_5(\beta)=15\beta++10\,\beta^3+\beta^5\,, \tag{65}$$

$$g_6(\beta)=15+45\,\beta^2+15\,\beta^4+\beta^6\,. \tag{66}$$

The first order of β is similar to that of SN. This means, for small beta, the two methods of generating skewness are very similar.

References

[1] Azzalini, A. (1985), "A class of distributions which includes the normal ones," Scand. J. Statist., 12, 171178.

[2] Gabaix, X. (2009), "Power Laws in Economics and Finance," Annual Review of Economics, 1, p. 255-93.

[3] Lihn, S. (2008), "Analytic Study of Skew Lognormal Cascade Distribution," SSRN.

[4] Praetz, P. D. (1972), "The Distribution of Share Price Changes,". Journal of Business 45, 49-55.