Sergey Kurenoy Novosibirsk 10 November, 2019

Test problem for Perfect Art interview (version 2): Expected shortfall distribution using Monte-Carlo simulation

To generate the required three-year (N = 750 trading day) time series, we presume that one-day returns can be modeled using the stable distribution (also known as Lévy alpha-stable distribution, or Pareto-Lévy stable distribution). The time series of N independent one-day proportional returns r_i^1 is produced simply as $r_i^1 = X_L$, for random variables X_L of the stable distribution with parameters α, β, δ, γ, generated in a batch of 750 using `SciPy.stats.levy_stable.rvs`. (Note that for a symmetric distribution, having skewness parameter $\beta = 0$, the location parameters $\delta = \mu$ and scale parameters $\gamma = c$ coincide in the two competing parametrizations of the stable distribution, eliminating a common source of confusion.) Then, the time series of N–9 overlapping ten-day proportional returns r_i^{10} is calculated using the given definitions: $r_i^1=(P_{i+1}-P_i)/P_i=(P_{i+1}/P_i)-1$ and $r_i^{10}=(P_{i+10}-P_i)/P_i=(P_{i+10}/P_i)-1$ where P_i are daily prices. Ten-day returns can be expressed in terms of ten consecutive one-day returns by “un-telescoping” the ten-day price ratio as a product of ratios of next day price ratios:

$$r_i^{10}=\frac{P_{i+10}}{P_i}-1=\frac{P_{i+10}}{P_{i+9}}\frac{P_{i+9}}{P_{i+8}}\dots\frac{P_{i+2}}{P_{i+1}}\frac{P_{i+1}}{P_i}-1=(r_{i+9}^1+1)(r_{i+8}^1+1)\dots(r_{i+1}^1+1)(r_i^1+1)-1 \quad .$$

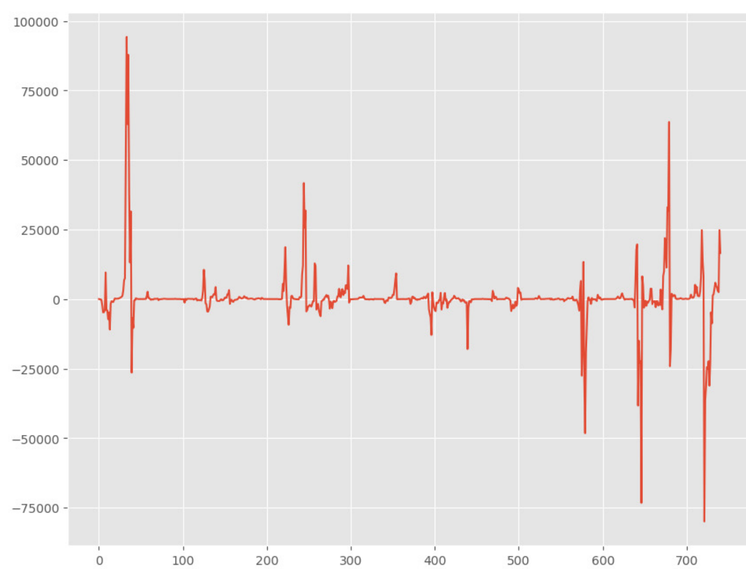
Numerically, the required product is computed efficiently by multiplying element-with-element ten overlapping `numpy` sub-arrays of N–9 = 741 consecutive one-day returns plus one. (It was verified that the results of this method coincide exactly with a direct method, and are faster and more error-resistant than a running product method.) Finally, the 1% percentile worst-case two week (ten-day) return is found by applying the `NumPy.quantile` function to the generated r_i^{10} time series and quantile $q = 0.01$.

The given stability parameter $\alpha = 1.7$ and skewness parameter $\beta = 0$ of the stable distribution given agree exactly with Mandelbrot's original analysis[1] of cotton prices, although he advocates modeling logarithmic returns according to the stable distribution rather than proportional returns. This choice of α and β is also consistent with existing fitting[2] to real-world logarithmic returns of US-listed stocks, where α generally varies between 1.5 and 1.8, and β is small ($\beta = 0 \pm 0.01$).

The left/negative tail of the stable distribution extends to $-\infty$, so the one-day and ten-day returns produced with this model are unbounded and can be very negative (while with stable-distributed logarithmic returns, the lowest possible proportional return would be -1, or 100% loss). This property of the model can produce a strange effect near -1; see the third qualitative observation on page 4.

Although they will be analyzed, the given location parameter $\delta = 1.0$ and scale parameter $\gamma = 1.0$ are unrealistically large, in the sense that the corresponding expected daily return would be $\delta \approx 100\%$ with a volatility of roughly $\gamma\sqrt{2} \approx 140\%$, which lead to unrestricted exponential growth and wild fluctuations in prices P_i. respectively. It would be reasonable to apply the model to much smaller γ and δ parameters, perhaps so that expected daily return and volatility do not exceed around 10% under "normal" conditions: that is, realistic scale and location parameters are in the following ranges:
Figure 1: a typical time series of ten-day returns for parameters $\delta = 1.0$ and $\gamma = 1.0$,
with 1% percentile value = –28317 (between the 7th and 8th lowest returns).$0 < \gamma \leq 0.10$ and $-0.10 \leq \delta \leq 0.10$. Due to limited time and computing power, these parameters are further restricted to the most interesting ranges: $0.01 \leq \gamma \leq 0.05$ and $-0.02 \leq \delta \leq 0.04$ to model more realistic situations.

Figure 1: a typical time series of ten-day returns for parameters $\delta = 1.0$ and $\gamma = 1.0$,
with 1% percentile value = –28317 (between the 7th and 8th lowest returns, a bit closer to the 7th).



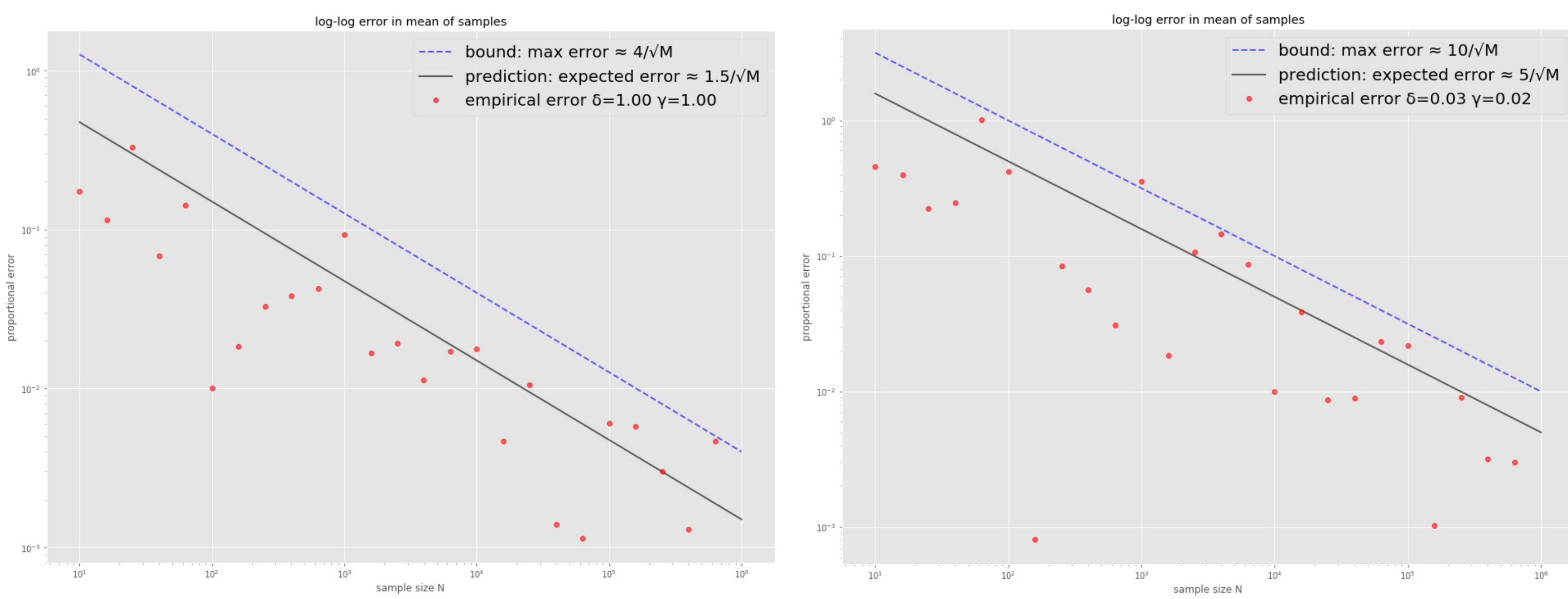
1 Benoit Mandelbrot, "The Variation of Certain Speculative Prices", *The Journal of Business*, Vol. 36, No. 4 (Oct., 1963), pp. 394-419, The University of Chicago Press. See http://www.jstor.org/stable/2350970

2 Riccardo Donati and Alice Pisani, Redexe "Risk Management and Finance" and University of Parma, presentation "Pareto-Lévy stable distributions in Action!", 23 Nov., 2011. See http://www.redexe.net/docs/redesfull.pdf

To estimate the necessary sample size of Monte Carlo simulations of the time series, we compare the mean values of the samples for different sample sizes, up to 10^6 simulations. According to the central limit theorem, provided that samples are independent and given any well-behaved statistical measurement, such a measurement should converge to some limit value for large sample size M, with an error that should scale proportionally to $1/\sqrt{M}$. Since the exact limit value of the mean is not known, we empirically calculate error relative to the mean of the largest sample with 10^6 simulations. Indeed, this kind of error scaling is observed for two cases of $\{\delta, \gamma\}$ parameters, and suggests how large the sample size must be to guarantee a given accuracy of other statistical results (or to achieve a given expected accuracy).

Given an error ε and an empirically determined scaling law $\varepsilon = k/\sqrt{M}$, the required sample size is $M = k^2/\varepsilon^2$. Thus, for $\delta = 1.0$ and $\gamma = 1.0$, to be certain that error is less than 1%, 160 000 simulations are required (while for 1% expected error, only about 22 500 simulations are needed). However, for $\delta = 0.03$ and $\gamma = 0.02$, with $M = 160\,000$ simulations, error is certainly less than 2.5% (and expected to be about 1.25%). We note that the scaling law constants k are different for other statistical measurements, such as, specifically, the fraction of observed 1% percentile ten-day returns in a given histogram bin, as used for fitting to a distribution or for a chi-square test, so these numbers are merely an estimate. We can also infer that for smaller $\{\delta, \gamma\}$ parameters, roughly 6 to 12 times more simulations are needed for the same accuracy as with $\delta = 1.0$ and $\gamma = 1.0$.

Figure 2: logarithmic plots demonstrate the scaling of error as $1/\sqrt{M}$ for sample size M.

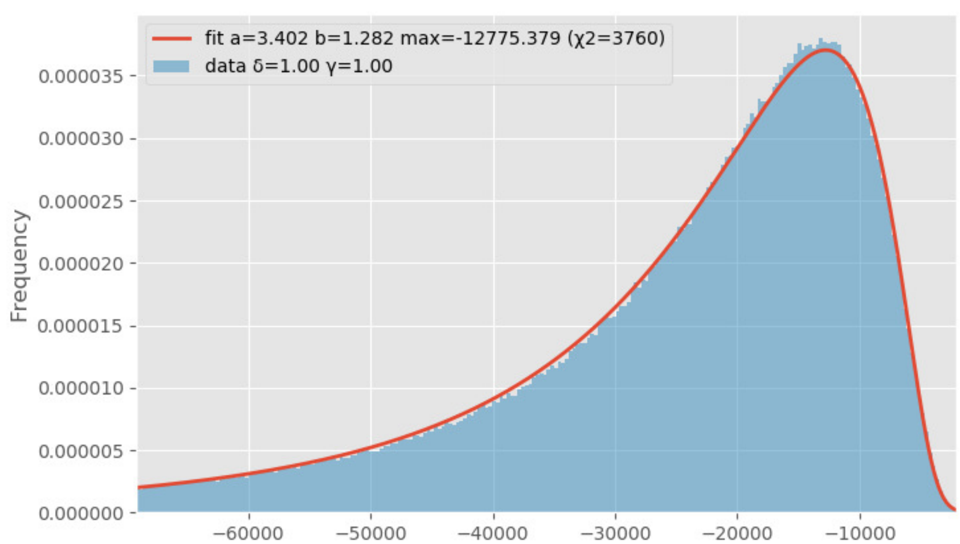


Qualitative observations about distributions that emerge when looking at samples of 1% percentile overlapping ten-day returns from independent time series instances with fixed $\{\delta, \gamma\}$ parameters, along with *proposed "financial interpretations"*, are as follows:

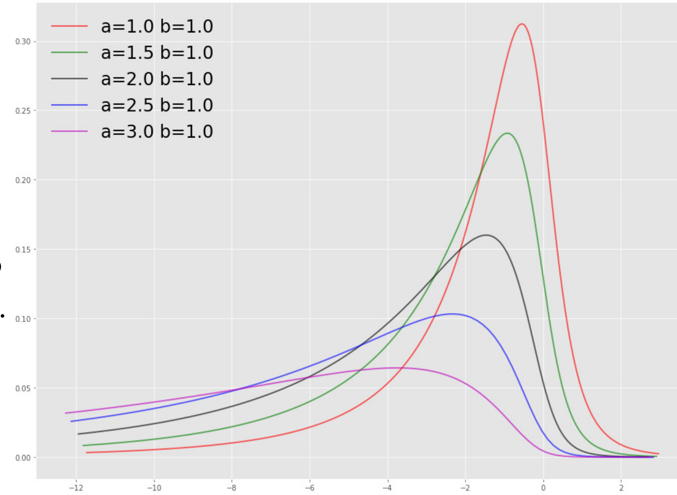
- A continuous non-symmetric or skewed distribution with a prominent peak and a long left/negative tail, but no right/positive tail. That is, *1% percentile ten-day returns are expected most frequently at a particular expectation value, although they can sometimes be much worse, but rarely any better than this fixed value.*
- The mode/maximum of the peak is generally centered at a negative value, unless $\delta \gg \gamma$. This mode is more negative when δ is small and γ is large. That is, *this aforementioned expectation value is usually a negative return, unless growth rate dominates volatility. The expected return becomes more negative when growth rate is lower or when volatility is higher.*
- A sharp cliff at return value -1 appears for $0.05 < \gamma < 0.11$ (sharpest near $\gamma = 0.08$), with an excess or pileup of frequencies slightly below -1 and a deficit slightly above -1. That is, *for a certain volatility range where the 1% percentile ten-day returns are most often near default (100% loss), the price often stagnates near zero for an extended time, which makes subsequent non-default returns less likely and incurred further debt (due to the possibility of very negative returns in this model) more likely.* (Algebraically, one-day price ratios near 0 are likely in this regime, so their product, which represents the ten-day price ratio, is consequently also near 0.)

Please see figure 5 on page 7, which more fully illustrates these observations for small $\{\delta, \gamma\}$ parameters. All distributions of the resulting samples can also be generally characterized as skewed to the left (having negative skewness) and as fat-tailed (having excess kurtosis).

Figure 3: distribution of sample data with 10^6 time series with $\delta = 1.0$ and $\gamma = 1.0$. Note that the histogram frequency is normalized to coincide with the probability density function of the fit.



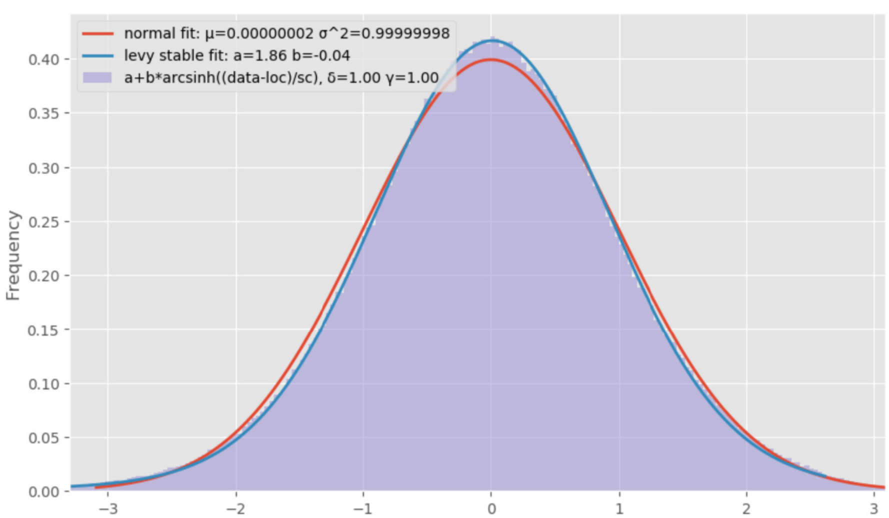
Quantitatively, the distributions of samples of 1% percentile overlapping ten-day returns most closely resemble Johnson's S_U distribution. (This can be confirmed by a brute-force method[3] that minimizes the sum of squared errors between a sample with fixed $\{\delta, \gamma\}$ and every available closest-fit distribution in `SciPy.stats`.) This distribution has four parameters: two shape parameters a and b, as well as location ξ and scale λ. The name comes from Norman Lloyd Johnson's hyperbolic sine transformation, which relates this distribution to a normal distribution:

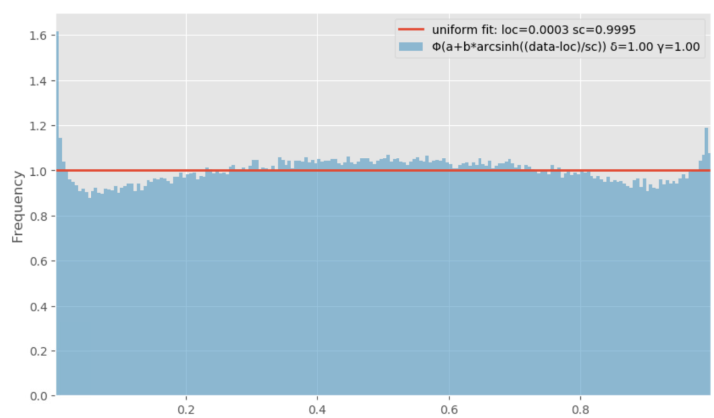


$$X_J = \xi + \lambda \sinh\left(\frac{X_N - a}{b}\right) \Leftrightarrow X_N = a + b \sinh^{-1}\left(\frac{X_J - \xi}{\lambda}\right) \quad ,$$

where X_N is distributed normally with mean 0 and variance 1, and X_J is a random variable in Johnson's S_U distribution with parameters a, b, ξ, λ. In the associated distribution found by applying the inverse hyperbolic sine transformation to the sample, one can more clearly see deviations from the theoretical distribution or verify the chi-square test statistic. Specifically, the inverse-Johnson transformed sample with 10^6 time series with $\delta = 1.0$ and $\gamma = 1.0$ actually corresponds closer to a Levy stable distribution with $\alpha = 1.86$ (not $\alpha = 1.7$ as for one-day returns, and not $\alpha = 2.0$ as for the expected perfectly normal distribution, but nearly their average). By further transforming using the normal cumulative density function Φ, deviations of the sample from the expected uniform distribution are seen even more clearly, and can be summarized as having prominent excess kurtosis or 4th moment (qualitatively, fatter tails and higher peak).

Figure 4: inverse sinh-transformed sample data with 10^6 time series with $\delta = 1.0$ and $\gamma = 1.0$ compared to a normal distribution; normal c.d.f. transformed sample data compared to a uniform distribution.





3 Code borrowed entirely from the top answer by user tmthydvnprt, saved as bestfit.py in shared repository. See https://stackoverflow.com/questions/6620471/fitting-empirical-distribution-to-theoretical-ones-with-scipy-python

To test whether the samples of 1% percentile overlapping ten-day returns plausably follow the proposed Johnson's S_U distribution, the chi-square goodness-of-fit test[4] is used. With $k = 34$ bins, the `SciPy.stats.chisquare` function is used to compute the chi-square test statistic χ^2:

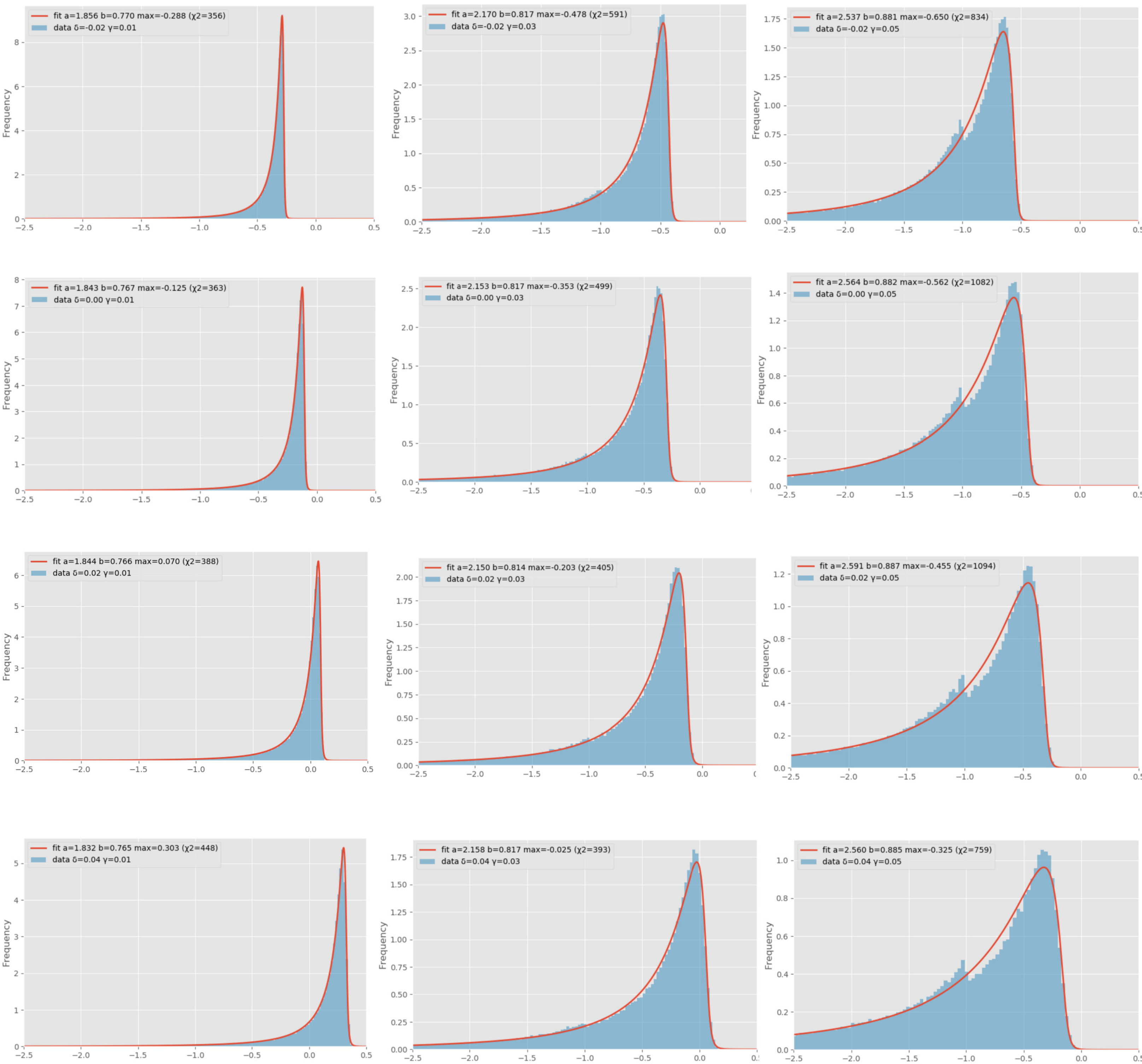
$$\chi^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i$$

Here, O_i is the observed frequency of the sample in the i-th bin, and $E_i = M(F(x_{upper}) - F(x_{lower}))$ is the expected frequency of the proposed distribution, in terms of the sample size M and the difference of its cumulative density function F evaluated at the bin edges x_{upper} and x_{lower}. It is suggested that sample size M should be large enough that every bin's expected frequency is at least 5, so in practice at least $M > 60\,000$ is needed for the bins to fill up sufficiently at the tail for this choice of bin number k and this kind of fat-tailed distribution. $M = 160\,000$ confidently satisfies this requirement.

For 95% confidence that the sample is distributed according to this theoretical distribution (with 4 parameters), with $k - c = 34 - 4 - 1 = 29$ degrees of freedom, the chi-square test statistic must not exceed the critical value of $\chi^2_{\text{crit}}(0.95, 29) = 44.557$. The resulting χ^2 of the small $\{\delta, \gamma\}$ cases range between 356 and 405 at best, but up to 1094 for prominent cliffs at –1 (described in the third observation). The χ^2 test statistic for sample data with 10^6 time series with $\delta = 1.0$ and $\gamma = 1.0$ is 3760 (while it is 3776 for the associated transformed binary-like distribution, and 2659 for the uniform-like distribution). It can be concluded that it is very unlikely that these samples are truly distributed according to the Johnson's S_U distribution: this is not surprising, because deviations from said distribution have already been mentioned.

4 National Institute of Standards and Technology (NIST) *e-Handbook of Statistical Methods* , "1.3.5.15. Chi-Square Goodness-of-Fit Test", See https://www.itl.nist.gov/div898/handbook/eda/section3/eda35f.htm

Figure 5: samples for varying small $\{\delta, \gamma\}$ with $M = 1.6{\times}10^6$, arranged with γ increasing rightward, and δ increasing downward. Again, histogram frequency is normalized to coincide with the p.d.f. of the fit.



The parameters a and b of the fitting Johnson's S_U distribution are linearly dependent on the small $\{\delta, \gamma\}$ parameters of the time series samples, and particularly strongly on the scale parameter γ. Using the method `scipy.optimize.minimize` to minimize the sum of square errors a linear function $L(\delta, \gamma) = c_1\, \delta + c_2\, \gamma + c_3$ and parameters a or b of the fitted distribution, the following linear estimation functions are found: $L_a(\delta, \gamma) \approx 0.21\, \delta + 17.45\, \gamma + 1.657$, with $\max\{|a - L_a|/a\} < 1.7\%$

$L_b(\delta, \gamma) \approx 0.085\, \delta + 2.89\, \gamma + 0.734$, with $\max\{|b - L_b|/b\} < 1.2\%$

However, the location ξ and scale λ parameters, as well as the maximum value of the probability density functions of Johnson's S_U distribution, have less obvious non-linear dependences.