Empirical Bounds of Log-Returns Characteristics

Yoshihiro Shirai

Applied Mathematics Dept.
University of Washington

June 27, 2023

Introduction

Folly and Fantasy in Finance

- Valuations should be based on log-returns dynamics explaining both
  - i. **historical prices**, reflecting **risks** deemed acceptable by operators/regulators;
  - ii. **option prices**, reflecting **market expectations**.
- It is possible to consistently model empirical/option-implied **finite dimensional distribution** of asset prices (Madan, [2022]):
  - Bid and ask defined by a set of equivalent laws distorting a measure C, chosen by the market, reflecting options' mid prices;
  - Historical (mid) price dynamics are specified by a measure $\mathbb{P}$;
- Inconsistencies however arise over path sets of **probability zero**, i.e. $\mathbb{P}$ and C are typically not equivalent;

Possible Solution: introducing Statistical Model Uncertainty

$\Rightarrow$ Dynamics specified by **nondominated** set $\mathfrak{P}$ of laws;

$\Rightarrow$ For each such law market chooses an EMM $\mathbb{Q}$;

$\Rightarrow$ Bid and ask are inf and sup over prices generated by each $\mathbb{Q}$;

$\Rightarrow$ If $\mathfrak{P}$ is singleton, we go back to a classical set up;

Volatility Uncertainty

- $\Omega = \mathbb{D}(\mathbb{R}_+, \mathbb{R})$ denotes the Skorohod space of real valued cadlags paths $\omega = \{\omega_t\}_{t\in\mathbb{R}_+}$ on $\mathbb{R}_+$ satisfying $\omega_0 = 0$;
- $\mathcal{F}$ is the Borel σ-algebra generated by the Skorohod topology on Ω;
- X is the canonical process on $(\Omega, \mathcal{F})$ defined by $X_t(\omega) = \omega_t$;
- $\mathfrak{P}_{[\underline{\sigma},\overline{\sigma}]}$ is the set of laws on $(\Omega, \mathcal{F})$ under which X is a semimartingale with differential characteristics $(\mu, \sigma, 0)$ where the process σ evolves in $[\underline{\sigma}, \overline{\sigma}]$.

"Escalator up and elevator down"

Hard to capture local asymmetries in log-returns due e.g. to panic/immediate selloff.

Speed Uncertainty

- $(\Omega, \mathcal{F})$ and X as before;
- $\kappa_{\mathbf{k}}(x)$ is defined, for $\mathbf{k} = (b_p, c_p, b_n, c_n)$, by

$$\kappa_{\mathbf{k}}(x) = c_p \frac{e^{-x/b_p}}{x} \mathbb{1}_{\{x>0\}} + c_n \frac{e^{-|x|/b_n}}{|x|} \mathbb{1}_{\{x<0\}}$$

- Given $K \subset \mathbb{R}^4_+$, $\Theta = \{(\mu, 0, \kappa_{\mathbf{k}} dx)\}_{\mathbf{k} \in K}$ is the set of Levy triplets corresponding to the Bilateral Gamma processes with parameters $(b_p, c_p, b_n, c_n) \in K$;
- $\mathfrak{P}_\Theta$ is the corresponding set of BG laws on $(\Omega, \mathcal{F})$.

Note:

i. BG law capable to fit options mid prices;

ii. Two BG laws are equivalent iff their speed is the same;

$\Rightarrow$ no need to include local BG laws in $\mathfrak{P}_\Theta$;

$\Rightarrow$ uncertainty in statistical parameters (c_p, c_n).

Goals

To **construct** $\mathfrak{P}_\Theta$, we

- estimate $\tilde{K}$ from historical prices;
- estimate $\hat{K}$ from risk neutral prices;
- combine them to form K

Question:

How well can we match **bid-ask spreads**?

Potential Applications:

Good and fast quotes for reversals and combos.

BG Process for Log Returns

From Black-Scholes to Bilateral Gamma

- Observation: prices exhibit exponential growth
- **Black-Scholes**: log-returns are Gaussian (maximal entropy law on $\mathbb{R}$)
- Issues:
  - i **Risk Aversion**
    - $\Rightarrow$ Days with intense selloff alternate with lower activity ones
    - $\Rightarrow$ Daily **realized variance/quadratic variation** is not constant
    - $\Rightarrow$ OTM puts priced higher than OTM calls (**volatility smile**)
  - ii Prices exhibit leptokurtic features and often **jump**
    - $\Rightarrow$ need to look at **higher moments** than just variance

Possible Solution

Randomize time to track periods with higher/slower activity

From Black-Scholes to Bilateral Gamma

- VG: Quadratic variation is **gamma process** (maximum entropy law on $\mathbb{R}_+$)

$$\Rightarrow S(t) = S(0)e^{\mu g(t) + B(g(t))}$$

- **Characteristic exponent**: by conditioning on the random time $g(t)$,

$$\mathbb{E}[e^{i\theta B(g(t))}] = \mathbb{E}\left[e^{-\frac{\theta^2\sigma^2}{2}g(t)}\right] = \left(1 + \frac{\sigma^2 v\theta^2}{2}\right)^{-\frac{t}{v}}$$
$$= (1 - ia\theta)^{-\frac{t}{v}} (1 + ia\theta)^{-\frac{t}{v}}$$

where $v = \mathbb{V}[g(1)]$, $a^2 = \frac{\sigma^2 v}{2}$ and we assumed wlog $\mathbb{E}[g(1)] = 1$.

Excess Return

- $\Rightarrow$ is difference of two gamma processes
- $\Rightarrow$ has **finite variation**

From Black-Scholes to Bilateral Gamma

- VG is sum of processes of gains and losses with same mean and variance
- Issue: downward jumps > upward jumps (**escalator up and elevator down**)
- BG process defined as difference of independent gamma processes, i.e.

$$\mathbb{E}[e^{i\theta X(t)}] = (1 - i\theta b_p)^{-tc_p} (1 + i\theta b_n)^{-tc_n}$$

with Levy density $\kappa(x) = \left(\frac{c_p}{x} e^{-b_p x} \mathbb{1}_{\{(0,\infty)\}}(x) + \frac{c_n}{|x|} e^{-b_n|x|} \mathbb{1}_{\{(-\infty,0)\}}(x) \right)$

$\Rightarrow$ BG is a finite variation Levy process

$\Rightarrow$ BG is self-decomposable $\Rightarrow$ sum of "independent news"

- Moments
  - Gains: $\mu_p = c_p b_p$, $\sigma_p = \sqrt{c_p} b_p$
  - Losses: $\mu_n = c_n b_n$, $\sigma_n = \sqrt{c_n} b_n$

Estimating $\tilde{K}$

Some Intuition

- Estimating $\tilde{K}$ is equivalent to identifying relationships between the parameters (b_p, c_p, b_n, c_n) of a bilateral gamma density
- Equivalently, we can estimate relationships between $(\mu_p, \sigma_p, \mu_n, \sigma_n)$;
- In a symmetric, Black-Scholes, setting, it is reasonable to expect that a positive relationship exists between reward, μ, and risk, σ, and several studies confirm such relationships;

Question:

$\Rightarrow$ Can we identify $(\sigma_p, \mu_n, \sigma_n)$ as risks and μ_p as their compensation?

$\Rightarrow$ If so, one would expect to see bounds for μ_p given $(\sigma_p, \mu_n, \sigma_n)$, to be increasing in each of the risks.

Risks and Compensation

Theorem 1.

Let X^+, X^-, Y^+, Y^- have gamma distribution. If

$$\mathbb{E}[X^+] \geq \mathbb{E}[Y^+],\ \mathbb{E}[X^-] \leq \mathbb{E}[Y^-],\ V[X^+] \leq V[Y^+],\ V[X^-] \leq V[Y^-],$$

then $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for every concave function u.

Theorem 2.

A strictly increasing and concave function $v \in C^2((0,\infty))$ has local CRRA coefficient ϵ bounded below by 1 if and only if there is a strictly increasing and concave function $u \in C^2(\mathbb{R})$ such that $v(x) = u(\log(x))$ for every $x \in (0,\infty)$.

Kelly's Criterion

LT investors maximize $\log$ utility. ST ones are more risk averse $\Rightarrow \epsilon \geq 1$

$\Rightarrow$ For BG log-returns, 3D risks vector $(\sigma_p, \mu_n, \sigma_n)$ compensated by μ_p

Estimation of Boundaries of the set $\tilde{K}$

- **Dataset**: $(\mu_p, \sigma_p, \mu_n, \sigma_n)$ daily estimated for 184 stocks for the period between 1/1/2008 to 31/12/2020.
- Assume that, for given risks $(\sigma_p, \mu_n, \sigma_n)$, acceptable compensation ranges between $f_m(\sigma_p, \mu_n, \sigma_n)$ and $f_M(\sigma_p, \mu_n, \sigma_n)$.
- f_m and f_M estimated via **quantile regression**, i.e. we solve

$$\min_{f \in \mathcal{F}} (1-\tau) \sum_i \left[\mu_p(i) - f_M(\sigma_p(i), \mu_n(i), \sigma_n(i))\right]^+ - \tau \sum_i \left[\mu_p(i) - f_M(\sigma_p(i), \mu_n(i), \sigma_n(i))\right]^- ,$$

- We set $\tau = 0.05$ for f_m and $\tau = 0.95$ for f_M.
- For $\mathcal{F}$ we considered the class of linear and Gaussian process regressors.

Estimation of Boundaries of the set $\tilde{K}$

Results

- Linear Regression:

$$f_m(\sigma_p, \mu_n, \sigma_n) = 0.0017 + 0.2029\sigma_p + 0.9951\mu_n - 0.3711\sigma_n,$$
$$f_M(\sigma_p, \mu_n, \sigma_n) = 0.0017 + 0.2710\sigma_p + 1.0102\mu_n - 0.2311\sigma_n.$$

- Gaussian process regression:
  - $\partial f_m / \partial \sigma_n$ always negative;
  - $\partial f_M / \partial \sigma_n$ negative at all but two of 16 representative points

Observation

$\Rightarrow$ Risk seeking behavior in the pure loss prospect

Estimation of Boundaries of the set $\tilde{K}$

$\frac{\partial f_M}{\partial \sigma_p}$	$\frac{\partial f_M}{\partial \mu_n}$	$\frac{\partial f_M}{\partial \sigma_n}$	$\frac{\partial f_m}{\partial \sigma_p}$	$\frac{\partial f_m}{\partial \mu_n}$	$\frac{\partial f_m}{\partial \sigma_n}$
0.2667	2.4704	0.7577	-0.0130	2.0042	-0.2421
0.8691	1.9402	-1.3539	1.1140	1.8974	-0.8361
1.5243	1.9553	-1.1134	1.4108	1.9274	-1.2346
1.0459	2.0254	-0.4887	0.5666	1.9927	-1.2635
1.0867	1.9956	-1.0836	0.8823	2.0199	-1.2220
0.4639	2.0065	-1.4194	0.6053	2.0648	-1.1568
1.3013	2.0509	-1.4681	1.2715	2.0128	-1.4149
0.9669	2.0019	-0.2462	0.4477	1.9760	-1.0806
1.4434	2.2522	0.3978	0.5052	2.0026	-0.5761
0.9710	1.9465	-0.9840	0.9472	1.8900	-0.8995
1.0653	1.9423	-1.4702	1.3075	1.9230	-0.9990
0.9307	1.9594	-0.5941	0.6044	1.9087	-1.0416
1.3390	2.0444	-1.8664	1.4529	2.0287	-1.5394
0.8652	1.9872	-1.3001	0.9281	2.0499	-1.0898
1.1957	2.0398	-0.9586	0.9027	1.9967	-1.3931
0.9283	1.9956	-0.0906	0.3913	1.9830	-0.9539

Estimation of Boundaries of the set $\tilde{K}$

Implied Boundaries of Measure Performance

Upper Boundary	Observation	Lower Boundary	Upper Boundary	Observation	Lower Boundary
0.0856	0.0694	0.0697	0.0536	0.0467	0.0469
0.0224	0.0208	0.0190	0.0184	0.0165	0.0143
0.0348	0.0343	0.0339	0.0269	0.0260	0.0252
0.0180	0.0167	0.0153	0.0147	0.0130	0.0112
0.0706	0.0685	0.0661	0.0473	0.0453	0.0428
0.1440	0.1428	0.1421	0.1024	0.1002	0.0986
0.0329	0.0308	0.0284	0.0243	0.0225	0.0204
0.0127	0.0119	0.0107	0.0092	0.0088	0.0081

Table: μ_p boundaries (estimated via Quantile GPR), at 16 representative points.

Estimation of Boundaries of the set $\tilde{K}$

Upper Boundary	Observation	Lower Boundary	Upper Boundary	Observation	Lower Boundary
4.0960	-0.1253	-0.0500	1.7966	-0.2803	-0.2231
1.4038	0.3108	-0.8727	2.2708	0.6168	-1.1672
-0.0615	-0.2441	-0.4023	0.4234	-0.0286	-0.4702
3.2179	1.6361	-0.1370	2.7574	0.9385	-0.9590
2.7685	0.8818	-1.3389	2.1867	0.7031	-1.2120
1.2525	0.5067	0.0483	2.3867	0.6779	-0.5442
2.4903	0.7203	-1.2574	2.9818	1.1174	-0.9577
3.0642	1.8490	0.2862	2.7892	2.1576	0.9789

Table: Sharpe ratio boundaries (estimated via Quantile GPR), at 16 representative points.

Estimation of Boundaries of the set $\tilde{K}$

Upper Boundary	Observation	Lower Boundary	Upper Boundary	Observation	Lower Boundary
0.151	-0.003	-0.003	0.065	-0.009	-0.009
0.051	0.011	-0.031	0.082	0.022	-0.041
-0.015	-0.008	-0.002	0.015	-0.001	-0.017
0.117	0.059	-0.006	0.100	0.034	-0.034
0.101	0.032	-0.049	0.079	0.026	-0.044
0.045	0.019	0.001	0.087	0.025	-0.020
0.090	0.026	0.046	0.109	0.040	-0.035
0.112	0.067	0.009	0.102	0.078	0.034

Table: Acceptability index boundaries (estimated via Quantile GPR), at 16 representative points. Negative signs represent acceptability indices of short positions.

$\Rightarrow$ Upper and lower performances consistent with empirical observations

Uncertainty Quantification

The uncertainty around μ_p given $(\sigma_p, \mu_n, \sigma_n)$ can be quantified by a dimensional analysis of the manifold $\tilde{K}$:

	PCA	cumulative weight (in %)	Diffusion Map	cumulative weight (in %)
λ_1	2.7529	68.82	0.0113	70.27
λ_2	1.1778	98.27	0.0045	98.58
λ_3	0.0685	99.98	0.0002	99.64
λ_4	0.0009	100.0	0.0001	100.0

Table: Eigenvalues's weights for PCA and diffusion map on the quantized dataset.

Observations

- $\Rightarrow$ Upper and lower boundaries for μ_p are relatively close;
- $\Rightarrow$ Boundaries are well approximated by linear functions

A Modified Lucas Tree Economy

- Is prospects theory consistent with the risk seeking behaviors in losses?
- Consider the following variation of a Lucas tree economy:
  - Two periods $i = 0, 1$
  - Each agent endowed with a single risky asset (a tree) with payoff S_i, $i = 0, 1$
  - Assume: S_0 known, $S_1 = S_0 e^{G-L}$, with G and L independent gamma variates
  - There is a risk free asset in zero net supply with risk free rate r_f
  - Consumption is determined by borrowing/lending ℓ at time zero:

$$C_0 = S_0 + \ell, \; C_1 = S_0 e^{G-L} - \ell e^{r_f}.$$

  - Preferences: let $X = s_0 + G - L$, $0 < \beta, \rho < 1$, and (logarithms in lower case)

$$U(C_0, C_1) = u(c_0) + e^{-\beta} \mathbb{E}\left[u(c_1)\mathbb{1}_{\{X \geq 0\}} - u(-c_1)\mathbb{1}_{\{X \leq 0\}}\right]$$

- Equilibrium condition $\ell = 0$ gives

$$r_f^e = \beta - \rho \log(s_0) - \log\left(\mathbb{E}[(X)^{-\rho} e^{-X} \mathbb{1}_{\{X \geq 0\}}] - \mathbb{E}[(-X))^{-\rho} e^{-X} \mathbb{1}_{\{X \leq 0\}}]\right).$$

A Modified Lucas Three Economy

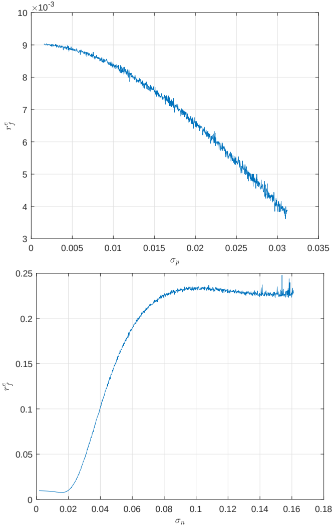

Figure: Equilibrium rate as function of σ_p, μ_p, σ_n, μ_n.

Estimating $\hat{K}$

Dataset and Methods

- **Dataset**: (b_p, c_p, b_n, c_n) calibrated every 10 days for 10 sector ETFs for the period between 1/1/2015 to 31/12/2020 for each of the four middle maturities traded $\Rightarrow$ 4812 observations;
- Bounds for c_p are estimated utilizing:
  - **quantile regression**: Quantile loss function replaced with

$$S(x) = \tau x + \alpha \log(1 - e^{-x/\alpha}), \ \alpha = 10^{-4}$$

  - **distorted least squares**: Objective function:

$$\min_{f \in \mathcal{F}} \sum_i r_i^2 \left(\Psi(q_i) - \Psi\left(q_i - \frac{1}{n}\right)\right),$$

    where r_i is residual, Ψ is MINMAXVAR distortion, q_i is the i-th empirical quantile of the residual's empirical distribution

Visualization of Speed Bounds

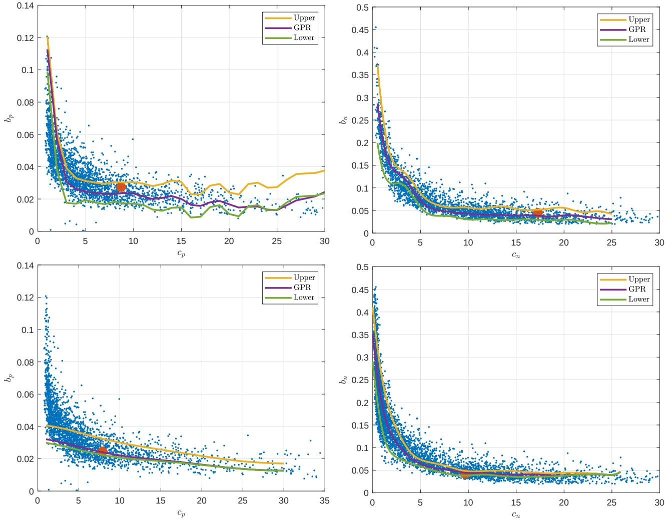

Figure: Visualization of quantile (upper pictures) and distorted (lower pictures) GPR boundaries around a randomly selected point (in red).

Uncertainty Quantification

	PCA	cumulative weight (in %)	Diffusion Map	cumulative weight (in %)
λ_1	1.2234	30.58	0.9999	50.06
λ_2	0.9949	55.46	0.9155	95.89
λ_3	0.9574	79.39	0.0575	98.77
λ_4	0.8244	100.0	0.0092	100.0

Table: Eigenvalues's weights for PCA and diffusion map on the risk neutral dataset.

$\Rightarrow$ Embedding and boundaries are nonlinear

$\Rightarrow$ Variance of speed well explained by that of scale

Options Implied Bid-Ask Prices to Unwind a 1$ Position

Upper Valuation	Lower Valuation	% of Points Represented	Upper Valuation	Lower Valuation	% of Points Represented
0.9806	1.0364	0.1641	0.9113	1.1205	0.0347
0.9712	1.0210	0.1610	0.9319	1.0358	0.0339
0.9630	1.0359	0.1240	0.9250	1.1468	0.0265
0.9646	1.0306	0.0943	0.9759	1.2024	0.0253
0.9538	1.0321	0.0920	0.9497	1.0631	0.0214
0.9586	1.0586	0.0799	0.9089	1.1389	0.0164
0.9638	1.0666	0.0608	0.8442	1.1997	0.0109
0.9286	1.0911	0.0452	0.8946	1.2626	0.0094

Work in Progress

Comparison with Model-Free Options Implied Prices

The set K

Constructing the Pricing Measures

- From options mid prices, estimate:
  - risk neutral $(\hat{b}_p, \hat{c}_p, \hat{b}_n, \hat{c}_n)$;
  - range $\hat{C}_p = (\hat{c}_{p,m}, \hat{c}_{p,M})$ for c_p given $(\hat{b}_p, \hat{b}_n, \hat{c}_n)$;
- From equity prices, estimate
  - statistical $(\tilde{b}_p, \tilde{c}_p, \tilde{b}_n, \tilde{c}_n)$;
  - range $\tilde{C}_p = (\tilde{c}_{p,m}, \tilde{c}_{p,M})$ for c_p given $(\tilde{b}_p, \tilde{b}_n, \tilde{c}_n)$;
- Set $C = \tilde{C}_p \times \tilde{c}_n \cup \hat{C}_p \times \hat{c}_n$;
- For each pair $(c_p, c_n) \in C$, estimate (b_p, b_n) that match best option prices.
- The pricing measures consists of resulting BG laws.

Constructing the Pricing Measures

Example with data on SPY as of October 8 2020.[1]

$$(\hat{b}_p, \hat{c}_p, \hat{b}_n, \hat{c}_n) = (0.0175, 24.1090, 0.0262, 42.9922)$$
$$\Rightarrow \hat{C}_p = (22.1135, 27.5439)$$
$$(\tilde{b}_p, \tilde{c}_p, \tilde{b}_n, \tilde{c}_n) = (0.0082, 0.1802, 0.0224, 0.4165)$$
$$\Rightarrow \tilde{C}_p = (0.6950, 1.3105)$$

Set $C = \hat{C}_p \cup \tilde{C}_p$, and

⇒ for each $(c_p, c_n) \in C$, compute (b_p, b_n) by matching first and second moment of options implied risk neutral distribution:

$$\varphi(-i; c_p, c_n, b_p, b_n) = \varphi(-i; \hat{c}_p, \hat{c}_n, \hat{b}_p, \hat{b}_n)$$
$$\varphi(-2i; c_p, c_n, b_p, b_n) = \varphi(-2i; \hat{c}_p, \hat{c}_n, \hat{b}_p, \hat{b}_n)$$

⇒ set ask price operator $a(\cdot) = \sup_{c_p, c_n \in C} \mathbb{E}^{\mathbb{Q}_{\kappa(c_p, c_n)}}[\cdot]$

[1] Upper and lower bounds on c_p are estimated via quantile GPR and quantile regression.

Valuation of Long Term Options

- Speed bounds control **short term** uncertainty;
- For long maturities, evolution of risk neutral/statistical BG parameters specified by 4D Markov chain $\{\mathbf{K}^{t_j}\}_{j=1,\dots,N}$:
  - Quantize the dataset of options implied BG parameters into S representative points $\{\mathbf{k}_s\}_{s=1,\dots,S}$, each representing a fraction $\mathbf{p}_s$ of points;
  - Define

$$\hat{q}_{s,r} := \frac{1}{\|F_{\mathbf{k}_s} - F_{\mathbf{k}_r}\|_W}, \quad Q_{s,r} := \underset{Q\in\mathcal{S},\mathbf{p}Q=0}{\arg\min} \sum_{s\neq r} \|\hat{q}_{s,r} - Q_{s,r}\|^2$$

$$(P)^j := e^{Qt_j}$$

$\Rightarrow$ Stationary distribution $\mathbf{p}$ and transition probability to close states maximized;

$\Rightarrow$ Simulate $\mathbf{K}^t$ and along each path compute bid ask prices backward.

Conclusions

Summary of Results

- Speed uncertainty allows to **consistently** model equity and option prices;
- This seems fundamental as quotes must depend on both **market expectations** and range of **risks** deemed acceptable by operators/regulators;
- Focus on Speed is needed to reflect **biases** in the financial markets introduced by risk averse/seeking behaviors;
- Quantile/Distorted GPR show certain promise in capturing risk neutral and statistical features of equity and option prices, such as:
  - Increasing utility of **variance of losses**;
  - Sharpe ratios and other **performance** measures;
  - **Forward** prices to unwind $1\$$ valuations.
- Potential Application: pricing **combos and reversals**, for which fast and good quotes need to be provided not to lose market shares.
- Future work include:
  - Additional statistical studies to compare forward prices with **model-free** prices;
  - Development of **statistical methods** on implied volatility surface to generate prices of portfolios of options.

Thank you!