Advanced Derivative Models (MSCF 46915). Stochastic Volatility Models

Sergey Nadtochiy

MSCF Program, Carnegie Mellon University

Fall 2025, Mini 2

Lecture outline

1. Stochastic volatility models
2. Autonomous volatility models
3. Heston model
4. GARCH model
5. SABR model
6. Incompleteness
7. Calibration
8. PDE pricing
9. Fourier pricing
10. Monte Carlo pricing
11. Hedging in Stochastic Vol models
12. Static hedging of European options

Stochastic volatility models

- General stochastic volatility (SV) models have the form

$$dS_t/S_t = r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}},$$

  where (v_t) a positive stochastic process called the spot variance or instantaneous variance.

- A concrete specification of (v_t) is needed and is typically given as another SDE.

- Model is specified directly under a risk-neutral measure $\mathbb{Q}$, with $W^{\mathbb{Q}}$ being a $\mathbb{Q}$-Brownian motion.

- This is because SV models are usually incomplete. Thus, the risk-neutral measure is not unique, but its existence guarantees no arbitrage.

Example: local vol models

- Setting the spot variance to the square of a local volatility function,

$$v_t = \sigma^2(t, S_t),$$

  shows that any local volatility model is a special case of SV model.
- Often not considered a proper SV model since spot variance is deterministically dependent on the risky asset.

Autonomous volatility

- SV model with spot variance following the SDE

$$dv_t = a(v_t)\,dt + b(v_t)\,dZ_t,$$

driven by a separate correlated Brownian motion Z.

- The functions a and b are chosen to ensure $v_t > 0$.
- Most SV models are of this type.

Heston model

- Heston model is the most prominent example of a SV model with spot variance given by

$$dv_t = \lambda(\bar{v} - v_t)\, dt + \eta \sqrt{v_t}\, dZ_t,$$
$$dZ_t\, dW_t^{\mathbb{Q}} = \rho\, dt,$$

  where Z is another $\mathbb{Q}$-Brownian motion correlated with $W^{\mathbb{Q}}$.

- Follows a square-root process (*Feller 1951*) also known as a CIR process (*Cox-Ingersoll-Ross 1985*).
- Mean-reverts at rate $\lambda > 0$ toward long-term variance $\bar{v} > 0$.
- Volatility of volatility parameter $\eta > 0$.
- Correlation $\rho \in (-1, 1)$ reproduces the leverage effect (aka implied skew) when $\rho < 0$.
- Feller condition: $2\lambda\, \bar{v} \geq \eta^2$ and $v_0 > 0$ ensure $v_t > 0$ for all $t \geq 0$.

Simulated Heston underlying price vs. real

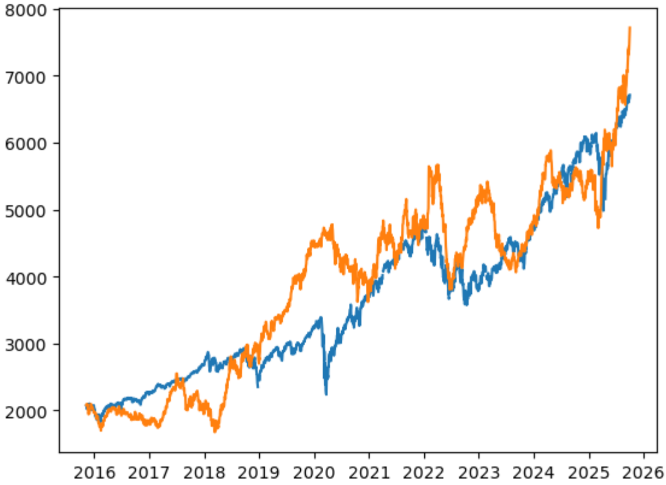

Figure: Historical price of SP500 (blue) and simulated Heston price with matching parameters (orange), over 2015 – 2025.

Realized vs. simulated returns

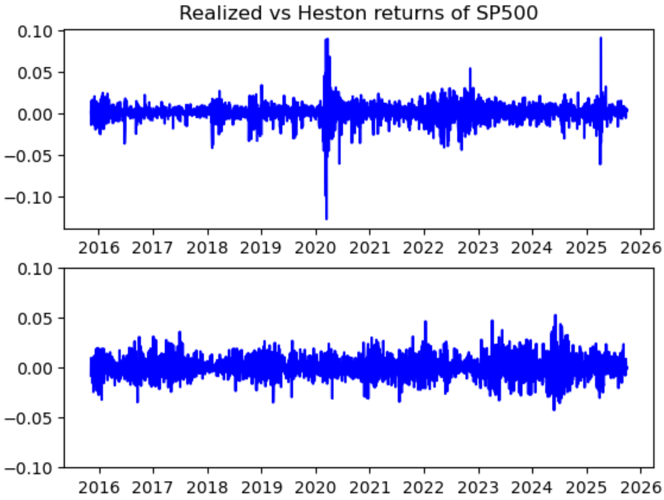

Figure: Realized daily returns of SP500 (top) and simulated Heston returns with matching parameters (bottom).

Realized vs. simulated volatility

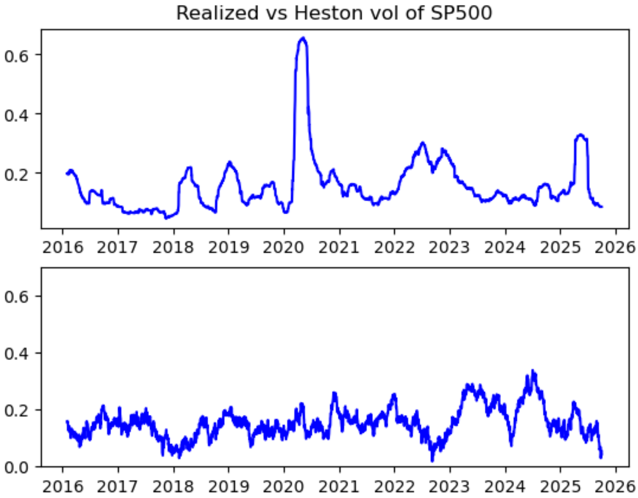

Figure: Realized volatility of SP500 (top) and simulated Heston volatility with matching parameters (bottom).

GARCH model

- GARCH: general auto-regressive conditional heteroscedastic.
- Similar to Heston model except has linear rather than square-root scaling on the innovation term

$$dv_t = \lambda(\bar{v} - v_t)\, dt + \eta\, v_t\, dZ_t.$$

- Volatility of volatility is monotone increasing in v_t.
- Heavier tailed returns than Heston that typically better fit market data.
- Less tractable than Heston.

SABR model

- SABR: stochastic alpha, beta, rho.
- The SABR model follows the SDE system

$$dS_t/S_t = r\,dt + S_t^{\beta-1}\,Y_t\,dW_t^{\mathbb{Q}},$$
$$dY_t = \alpha\,Y_t\,dZ_t,$$

  where $\alpha \geq 0$, $\beta \in (0,2)$ and $dW_t^{\mathbb{Q}}\,dZ_t = \rho\,dt$ (*Hagan-Kumar-Lesniewski-Woodward 2002*).
- Spot variance is

$$v_t = S_t^{2\beta-2}\,Y_t^2.$$

- Does not mean-revert and is typically only suitable for short time scales.
- Accurate short-term asymptotics available.
- Spot variance depends on an autonomous SDE and on S_t itself.
- Leverage/skew is captured by either $\rho < 0$ or $\beta < 1$.
- This is a form of local-stochastic volatility model, which we will explore at the end of this class.

Incompleteness

- Stochastic volatility models are typically incomplete.
- Recall that completeness is equivalent to uniqueness of RNM $\mathbb{Q}$.
- It is easy to see that a model of the form

$$dS_t/S_t = r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}},$$
$$dv_t = a(v_t)\,dt + b(v_t)\,dZ_t, \quad dW_t^{\mathbb{Q}}\,dZ_t = \rho\,dt, \quad \rho \in (-1,1),$$

has infinitely many RNMs.

- Indeed, the fact that two Brownian motions with zero covariation are independent, along with $\rho^2 \neq 1$, allows us to deduce the existence of a $\mathbb{Q}$-Brownian motion $\tilde{Z}$, independent of $W^{\mathbb{Q}}$, such that

$$Z_t = \rho\,W_t^{\mathbb{Q}} + \sqrt{1-\rho^2}\,\tilde{Z}_t.$$

- Girsanov's theorem tells us that, for any bounded process (ξ_t), there exists $\hat{\mathbb{Q}} \sim \mathbb{Q}$ s.t. $W^{\mathbb{Q}}$ and $\hat{Z}$ are independent $\hat{\mathbb{Q}}$-Brownian motions, where

$$d\hat{Z}_t = \xi_t\,dt + d\tilde{Z}_t.$$

Incompleteness

- We have

$$
\begin{aligned}
dS_t/S_t &= r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}},\\
dv_t &= a(v_t)\,dt + b(v_t)\,dZ_t\\
&= a(v_t)\,dt + b(v_t)\,(\rho\,dW_t^{\mathbb{Q}} + \sqrt{1-\rho^2}\,d\tilde{Z}_t)\\
&= a(v_t)\,dt + b(v_t)\,(\rho\,dW_t^{\mathbb{Q}} + \sqrt{1-\rho^2}\,d\hat{Z}_t - \xi_t\,\sqrt{1-\rho^2}\,dt)\\
&= [a(v_t) - \xi_t\,b(v_t)\,\sqrt{1-\rho^2}]\,dt\\
&\quad + b(v_t)\,\rho\,dW_t^{\mathbb{Q}} + b(v_t)\,\sqrt{1-\rho^2}\,d\hat{Z}_t.
\end{aligned}
$$

- If b is not identically zero and $\rho^2 \neq 1$, the choice of ξ affects the distribution of v and, in turn, S.
- Thus, we have infinitely many different RNM's $\hat{\mathbb{Q}}$, which means that the model is incomplete.
- Completeness can be restored if $\rho^2 = 1$ (and if $v_t > 0$ for all $t \geq 0$).

Calibration

- Calibration can be accomplished via an optimization problem, for example, a least squares problem:

$$\min_{\Theta} \sum_{i,j} \left(\mathrm{C}^{\Theta}(T_j, K_i) - \mathrm{C}^{\mathrm{mrkt}}(T_j, K_i) \right)^2 ,$$

  where Θ is the vector of model parameters.

- The above easy to implement, if the pricing mapping $\Theta \mapsto \mathrm{C}^{\Theta}$ is computationally tractable.
- Challenges:
  - No guarantee that there exists Θ which can match market price with desired accuracy.
  - No guarantee that the algorithm will find the desired Θ even if such exists (the optimization problem is non-convex).

PDE pricing in autonomous SV models

$$dS_t/S_t = r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}},$$
$$dv_t = a(v_t)\,dt + b(v_t)\,dZ_t, \quad dW_t^{\mathbb{Q}}\,dZ_t = \rho\,dt, \quad \rho \in (-1,1).$$

- Feynman-Kac allows prices to be recovered as solutions to PDEs, even for incomplete models without replication.
- Recall that, for a European option with payoff $X = H(S_T)$, its arbitrage-free price is given by

$$\pi_t(X) = B_t\,\mathbb{E}^{\mathbb{Q}}[X/B_T \mid \mathcal{F}_t] = B_t\,\mathbb{E}^{\mathbb{Q}}[H(S_T)/B_T \mid S_t,\, v_t],$$

  where we used the Markov property of (S_t, v_t).
- Recalling that r is constant,

$$\pi_t(X) = \mathbb{E}^{\mathbb{Q}}[H(S_T)\,e^{-r(T-t)} \mid S_t,\, v_t] =: V(t, S_t, v_t).$$

PDE pricing in autonomous SV models

$$dS_t/S_t = r\,dt + \sqrt{v_t}\,dW^{\mathbb{Q}}_t,$$
$$dv_t = a(v_t)\,dt + b(v_t)\,dZ_t, \quad dW^{\mathbb{Q}}_t\,dZ_t = \rho\,dt, \quad \rho \in (-1,1).$$

- Recalling that $(V(t,S_t,v_t)\,e^{-rt})$ is a $\mathbb{Q}$-martingale, we apply Itô's formula and equate the drift of the latter process to zero, to obtain

$$\partial_t V + \frac{1}{2}x^2\,y\,\partial^2_{xx}V + \rho\,x\sqrt{y}\,b\,\partial^2_{xy}V + \frac{1}{2}b^2\,\partial^2_{yy}V + r\,x\,\partial_x V + a\,\partial_y V = rV,$$
$$V(T,x,y) = H(x),\ x,y>0,$$

where (x,y) play the role of (S,v).

Basis expansion

- Alternative to PDE pricing works for some models.
- Decompose a European payoff $X = H(S_T)$ as a linear combination of basis payoffs

$$H(x) = \sum_{j=1}^{n} c_j \, h_j(x),$$

where each $\mathbb{E}^{\mathbb{Q}} h_j(S_T)$ can be more easily computed.

- Then by linearity of the expectation operator

$$\mathbb{E}^{\mathbb{Q}} H(S_T) = \sum_{j=1}^{n} c_j \, \mathbb{E}^{\mathbb{Q}} h_j(S_T).$$

Fourier transform and expectations

- Fourier pricing extends this idea to infinite sums and integrals.
- Fourier transform of a function $G : \mathbb{R} \to \mathbb{R}$ is

$$\hat{G}(\omega) = \int_{\mathbb{R}} G(x)\, e^{-\mathrm{i}\omega x}\, dx.$$

- Fourier inversion formula is

$$G(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{G}(\omega)\, e^{\mathrm{i}\omega x}\, d\omega.$$

- Linearity of integral and expectation (Fubini's theorem) leads to the pricing formula:

$$\mathbb{E}^{\mathbb{Q}} H(S_T) = \mathbb{E}^{\mathbb{Q}} G(\log S_T) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{G}(\omega)\, \mathbb{E}^{\mathbb{Q}}[e^{\mathrm{i}\omega \log S_T}]\, d\omega,$$

where $G(x) := H(e^x)$.

Fourier transform of European call prices

- Numerical computation of the integral is reasonable provided $\hat{G}(\omega)$ are known and $\mathbb{E}^{\mathbb{Q}}[e^{\mathrm{i}\,\omega\,\log S_T}]$ can be computed efficiently for all $\omega \in \mathbb{R}$.
- For a call payoff, we have $G(x) = H(e^x) = (e^x - K)^+$ and

$$\hat{G}(\omega) = \int_{\mathbb{R}} (e^x - K)^+ \, e^{-\mathrm{i}\omega\,x} \, dx = \int_{\log K}^{\infty} (e^{(1-\mathrm{i}\omega)\,x} - K\, e^{-\mathrm{i}\omega\,x}) \, dx$$

$$= \left[\frac{e^{(1-\mathrm{i}\omega)\,x}}{1-\mathrm{i}\omega} - K \frac{e^{\mathrm{i}\omega\,x}}{-\mathrm{i}\omega} \right]_{x=\log K}^{\infty} = \frac{K^{1-\mathrm{i}\omega}}{\mathrm{i}\omega\,(\mathrm{i}\omega - 1)},$$

which holds for all $\omega \in \mathbb{C}$ with $\mathrm{Im}(\omega) < -1$.

- Bad news: the inversion formula

$$G(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{G}(\omega)\, e^{\mathrm{i}\omega\,x} \, d\omega.$$

cannot be used directly since $\hat{G}(\omega)$ is not well defined for $\omega \in \mathbb{R}$.

European call prices via Fourier transform inversion

- To fix this issue, we consider $\omega \in \mathbb{R}$, $\beta > 1$ and

$$\hat{G}(\omega - \mathrm{i}\beta) = \int_{\mathbb{R}} G(x)\, e^{-\mathrm{i}(\omega - \mathrm{i}\beta)\, x}\, dx = \int_{\mathbb{R}} (G(x)\, e^{-\beta\, x})\, e^{-\mathrm{i}\omega\, x}\, dx.$$

- Recognizing the above as a Fourier transform of $G(x)\, e^{-\beta\, x}$, we apply the inversion formula:

$$G(x)\, e^{-\beta\, x} = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{G}(\omega - \mathrm{i}\beta)\, e^{\mathrm{i}\omega\, x}\, d\omega.$$

- From the above,

$$\mathbb{E}^{\mathbb{Q}} H(S_T) = \mathbb{E}^{\mathbb{Q}} G(\log S_T) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{G}(\omega - \mathrm{i}\beta)\, \mathbb{E}^{\mathbb{Q}}[e^{(\beta + \mathrm{i}\omega)\, \log S_T}]\, d\omega.$$

European call prices via Fourier in Black-Scholes

$$\mathbb{E}^{\mathbb{Q}} H(S_T) = \mathbb{E}^{\mathbb{Q}} G(\log S_T) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{G}(\omega - \mathrm{i}\beta)\, \mathbb{E}^{\mathbb{Q}}[e^{(\beta+\mathrm{i}\omega)\ \log S_T}]\, d\omega.$$

- We have a formula for $\hat{G}$.
- But we still need to compute $\mathbb{E}^{\mathbb{Q}}[e^{z\ \log S_T}]$ for $z \in \mathbb{C}$. This may or may not be easy - depends on a model.
- In Black-Scholes model, under RNM $\mathbb{Q}$,

  $$\log S_T = \log S_0 + (r - \sigma^2/2)\, T + \sigma\, W_T^{\mathbb{Q}} \sim N(\log S_0 + (r - \sigma^2/2)\, T,\, \sigma^2\, T).$$
- Hence

  $$\mathbb{E}^{\mathbb{Q}}[e^{z\ \log S_T}] = S_0^z\, e^{rTz + \frac{1}{2}\sigma^2\ T\, z(z-1)}.$$
- Combining above results gives the pricing formula

  $$\mathbb{E}^{\mathbb{Q}}(S_T - K)^+ = \frac{K}{2\pi} \int_{\mathbb{R}} \frac{e^{(\beta+\mathrm{i}\omega)\left(\log(S_0/K) + (r + \sigma^2(\beta - 1 + \mathrm{i}\omega)/2)\ T\right)}}{(\beta + \mathrm{i}\omega)(\beta - 1 + \mathrm{i}\omega)}\, d\omega,$$

  which has a similar computational complexity to the one of normal cdf.

European call prices via Fourier in Heston

$$
\begin{aligned}
dv_t &= \lambda(\bar{v} - v_t)\, dt + \eta \sqrt{v_t}\, dZ_t, \\
dZ_t\, dW_t^{\mathbb{Q}} &= \rho\, dt.
\end{aligned}
$$

- Remarkably, $\mathbb{E}^{\mathbb{Q}}[e^{z \log S_T}]$ has an explicit form in Heston model (because it belongs to the class of so-called affine models):

$$
\mathbb{E}^{\mathbb{Q}}[e^{z \log S_T}] = S_0^z\, e^{rTz + C(T,z)\,\bar{v} + D(T,z)\, v_0},
$$

where

$$
\begin{aligned}
C(T,z) &= \frac{\lambda}{\eta^2}\left((\lambda - \rho\eta\, z - d(z))\, T - 2\log \frac{g(z)e^{-d(z)T} - 1}{g(z) - 1} \right), \\
D(T,z) &= \frac{\lambda - \rho\eta z - d(z)}{\eta^2}\, \frac{1 - e^{-d(z)T}}{1 - g(z)e^{-d(z)T}}, \\
g(z) &= \frac{\lambda - \rho\eta z - d(z)}{\lambda - \rho\eta z + d(z)}, \quad d(z) = \sqrt{(\lambda - \rho\eta z)^2 - \eta^2 z(z-1)}.
\end{aligned}
$$

- Resulting pricing formula allows efficient calibration.

Heston implied volatility fit (Gatheral 2006)

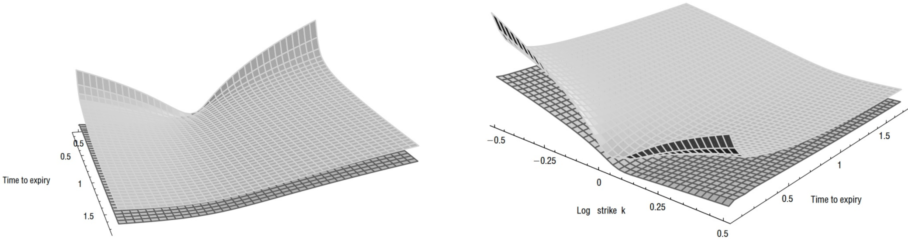

Figure: Heston fit to market implied volatility surface (parallel shift for better visibility, market above model) from two perspectives.

Monte Carlo Pricing

- Law of large numbers (LLN) ensures the validity of the approximation

$$\mathbb{E}^{\mathbb{Q}} H(S_T) \approx \frac{1}{N} \sum_{i=1}^{N} H(S_T^{(i)})$$

for large N, where $\{S_T^{(i)}\}$ is an i.i.d. sample form the distribution of S_T under $\mathbb{Q}$.

- Holds for any payoff function H and even for non-European payoffs: e.g., path-dependent derivatives.
- This makes Monte Carlo (MC) more useful for one-off pricing of exotics rather than for calibration purposes. Although, in the absence of other methods, MC is sometimes used for calibration as well.
- Question: how to simulate i.i.d. copies of S_T?

Euler scheme

- Assume a general autonomous SV model

$$dS_t/S_t = r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}},$$
$$dv_t = a(v_t)\,dt + b(v_t)\,dZ_t, \quad dW_t^{\mathbb{Q}}\,dZ_t = \rho\,dt.$$

- (Explicit) Euler discretization is the simplest numerical scheme:

$$S_{t+\Delta t} = S_t + r\,S_t\,\Delta t + \sqrt{v_t}\,S_t\,\sqrt{\Delta t}\,\epsilon^1_{t+\Delta t},$$
$$v_{t+\Delta t} = v_t + a(v_t)\,\Delta t + b(v_t)\,\sqrt{\Delta t}\,\epsilon^2_{t+\Delta t},$$

  where $\{\epsilon_t = (\epsilon^1_t, \epsilon^2_t)\}$ are i.i.d. 2-dim normal vectors with unit variance and covariance ρ.

- Under natural assumptions on the coefficients, Euler scheme converges in probability to the true solution (S_t, v_t) of the SDE, as $\Delta t \to 0$.

Euler scheme: ensuring nonnegativity of v_t

$$S_{t+\Delta t} = S_t + r\, S_t\, \Delta t + \sqrt{v_t}\, S_t \sqrt{\Delta t}\, \epsilon^1_{t+\Delta t},$$
$$v_{t+\Delta t} = v_t + a(v_t)\, \Delta t + b(v_t) \sqrt{\Delta t}\, \epsilon^2_{t+\Delta t},$$

- The above scheme will produce $v_t < 0$ for some $t > 0$.
- How to modify the scheme to ensure that (v_t) remains nonnegative?
- If we know that the true solution (v_t) stays positive (e.g., in Heston with $2\lambda\, \bar{v} \geq \eta^2$), we can use the reflecting boundary condition

$$v_{t+\Delta t} = \left| v_t + a(v_t)\, \Delta t + b(v_t) \sqrt{\Delta t}\, \epsilon^2_{t+\Delta t} \right|.$$

- If the true solution (v_t) may hit zero, we typically define it to stay at zero after that hitting time (i.e., it is absorbed at zero). In both cases, we can use the absorbing boundary condition

$$v_{t+\Delta t} = \begin{cases} \left(v_t + a(v_t)\, \Delta t + b(v_t) \sqrt{\Delta t}\, \epsilon^2_{t+\Delta t} \right)^+, & \text{if } v_t > 0, \\ 0, & \text{if } v_t = 0. \end{cases}$$

Euler scheme: accuracy

- Denote by $\hat{S}_T$ the Euler approximation, viewed as a random variable.
- Denote by $\{\hat{S}_T^{(i)}\}$ a sample of i.i.d. copies of $\hat{S}_T$.
- The accuracy of MC method is measured by

$$|\mathbb{E}H(S_T) - \frac{1}{N}\sum_{i=1}^{N} H(\hat{S}_T^{(i)})|$$

$$\leq |\mathbb{E}H(S_T) - \mathbb{E}H(\hat{S}_T)| + |\mathbb{E}H(\hat{S}_T) - \frac{1}{N}\sum_{i=1}^{N} H(\hat{S}_T^{(i)})|.$$

- For Lipschitz H (even if path-dependent), the discretization error is

$$|\mathbb{E}H(S_T) - \mathbb{E}H(\hat{S}_T)| = O(\Delta t^{1/2}).$$

- The second term is the MC (LLN) error:

$$|\mathbb{E}H(\hat{S}_T) - \frac{1}{N}\sum_{i=1}^{N} H(\hat{S}_T^{(i)})| = O(N^{-1/2}).$$

- Hence, it is reasonable to choose $\Delta t \sim N^{-1}$.

Exact scheme

- The square-root (Heston) process

$$dv_t = \lambda(\bar{v} - v_t)\,dt + \eta\sqrt{v_t}\,dZ_t$$

  has a known transition probability – i.e., the conditional distribution of $v_{t+\Delta t}$ given v_t – it is given by the non-central chi-square distribution (*Broadie-Kaya 2006*).
- Thus, we can simulate $v_{t+\Delta}$ directly from its correct distribution at each step of the scheme, which reduces the error produced by time discretization.
- This is known as the exact numerical scheme (for the associated SDE).
- Challenges:
  - we only have an exact scheme for v, but not for (S, v) – the SDE for S still needs to be approximated via Euler scheme,
  - simulation from non-central chi-square is relatively slow.
- Alternative ways to improve the performance are provided by higher-order discretization schemes.

Milstein scheme

- In the discretization error,
$$|\mathbb{E}H(S_T) - \mathbb{E}H(\hat{S}_T)| = O(\Delta t^{1/2}),$$
- the error coming from the Euler approximation of the drift term $\int a(v_t)\,dt$ is of the order $O(\Delta t)$,
- while the error coming from the Euler approximation of the diffusion term $\int b(v_t)\,dZ_t$ is of the order $O(\Delta t^{1/2})$.
- Main idea is to find a higher-order approximation of the diffusion term:
$$\begin{aligned} b(v_s) &\approx b(v_t) + \partial_v b(v_t)[v_s - v_t] \\ &\approx b(v_t) + b(v_t)\,\partial_v b(v_t)\,[Z_s - Z_t]. \end{aligned}$$
- Using the above, we obtain
$$v_{t+\Delta t} - v_t \; = \; \int_t^{t+\Delta t} dv_s \; \approx \; a(v_t)\Delta t + \int_t^{t+\Delta t} b(v_s)dZ_s$$
$$\approx a(v_t)\Delta t + b(v_t)[Z_{t+\Delta t} - Z_t] + b(v_t)\,\partial_v b(v_t) \int_t^{t+\Delta t} [Z_s - Z_t]dZ_s.$$

Milstein scheme

$$
\begin{aligned}
v_{t+\Delta t} - v_t &= \int_t^{t+\Delta t} dv_s \approx a(v_t)\Delta t + \int_t^{t+\Delta t} b(v_s) dZ_s \\
&\approx a(v_t)\Delta t + b(v_t)[Z_{t+\Delta t} - Z_t] + b(v_t)\,\partial_v b(v_t) \int_t^{t+\Delta t} [Z_s - Z_t] dZ_s.
\end{aligned}
$$

- Using Itô's formula, we have

$$\int_t^{t+\Delta t} [Z_s - Z_t] dZ_s = \frac{1}{2}[Z_{t+\Delta t} - Z_t]^2 - \frac{\Delta t}{2}.$$

- Denote $Z_{t+\Delta t} - Z_t = \sqrt{\Delta t}\,\epsilon^2_{t+\Delta t}$, where $\{\epsilon^2_t\}$ are i.i.d. standard normals.
- Then the above approximation becomes:

$$v_{t+\Delta t} = v_t + a(v_t)\Delta t + b(v_t)\sqrt{\Delta t}\,\epsilon^2_{t+\Delta t} + b(v_t)\,\partial_v b(v_t)\,\frac{\Delta t}{2}[(\epsilon^2_{t+\Delta t})^2 - 1].$$

- Discretization error of Milstein's scheme is $O(\Delta t)$. So, $\Delta t \sim N^{-1/2}$.

Milstein scheme: implementation details

- Simulated paths of (v_t) can still become negative, so we apply the same boundary conditions to address that.
- To simulate (S_t, v_t) jointly, we need to generate i.i.d. normal noise vectors $\{\epsilon_t = (\epsilon_t^1, \epsilon_t^2)\}$, with $\mathrm{Var}(\epsilon_t^i) = 1$ and $\mathrm{Cov}(\epsilon_t^1, \epsilon_t^2) = \rho$.
- To generate the desired $\epsilon_t = (\epsilon_t^1, \epsilon_t^2)$, we generate independent standard normals ζ_t^1 and ζ_t^2 and define

$$\epsilon_t^1 := \zeta_t^1, \quad \epsilon_t^2 := \rho\,\zeta_t^1 + \sqrt{1-\rho^2}\,\zeta_t^2.$$

- Instead of approximating (S_t) from its SDE directly, we approximate $(X_t := \log S_t)$:

$$X_{t+\Delta t} = X_t + (r - v_t/2)\,\Delta t + \sqrt{v_t}\,\sqrt{\Delta t}\,\epsilon_{t+\Delta t}^1.$$

- Discretization error of this scheme is still $O(\Delta t)$. So, $\Delta t \sim N^{-1/2}$.

Superhedging

- Assume that we can hedge an option with the underlying S.
- A superhedge for an option with payoff $X \in \mathcal{F}_T$ is any self-financing strategy that prescribes to hold Δ_t units of underlying at time t such that $V_T^\Delta \geq X$ a.s., where V^Δ is the wealth generated by this strategy:

$$dV_t^\Delta = \Delta_t \, dS_t + r(V_t^\Delta - \Delta_t \, S_t) \, dt.$$

- The minimal superhedging price is the smallest initial capital needed to fund a superhedge:

$$\overline{V}(X) := \min\{V_0^\Delta : \; V_T^\Delta \geq X \text{ a.s.}\}.$$

- Selling an option at $\overline{V}(X)$ allows the seller to offset the risk of that position fully – this is a "seller's price".

Superhedging price is too large

- The minimal superhedging price, in principle, depends on the model – i.e., on the distribution/dynamics of (S, v).
- However, the model-dependent superhedging price is often so large that it coincides with its model-independent version: i.e., it succeeds for any distribution/dynamics of (S, v), provided the other basic assumptions hold (such as linear pricing, negligible transaction costs, zero-dividend and constant interest rate).
- For example, $\overline{V}((K - S_T)^+) = e^{-rT} K$ in Heston model, which succeeds as a superhedging price in any model satisfying the above basic assumptions.
- Thus, $\overline{V}(X)$ is typically too large to use in practice (at least, if the hedging instruments are restricted to the underlying).

Superhedging price of a put in Heston model

$$
\begin{aligned}
&dS_t/S_t = r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}},\\
&dv_t = \lambda(\bar{v} - v_t)\,dt + \eta\sqrt{v_t}\,dZ_t, \quad dZ_t\,dW_t^{\mathbb{Q}} = 0,\\
&\overline{V}((K - S_T)^+) = e^{-rT}K.
\end{aligned}
$$

- Notice that, under $\mathbb{Q}$, (v_t) is independent of $W^{\mathbb{Q}}$ and

$$X_T := \log S_T = rT + \int_0^T \sqrt{v_t}\,dW_t^{\mathbb{Q}}, \quad X_T\,|\,(v_t)_{t\in[0,T]} \sim N\left(rT, \int_0^T v_t\,dt\right).$$

- Then, the time-zero put price in Heston model is

$$\mathrm{P}^{\mathrm{h}}(S_0, v_0, T, K) = \mathbb{E}^{\mathbb{Q}}\mathrm{P}^{\mathrm{BS}}(S_0, \sqrt{\int_0^T v_t\,dt}, T, K).$$

Superhedging price of a put in Heston model

$$\mathrm{P}^{\mathrm{h}}(S_0, v_0, T, K) = \mathbb{E}^{\mathbb{Q}} \mathrm{P}^{\mathrm{BS}}\left(S_0, \sqrt{\int_0^T v_t \, dt}, T, K\right).$$

- From the Black-Scholes formulas, we can see that

$$\mathrm{P}^{\mathrm{BS}}(S_0, \sigma, T, K) \to e^{-rT} K \text{ as } \sigma \to \infty.$$

- It is easy to see that

$$\sqrt{\int_0^T v_t \, dt} \to \infty \text{ as } v_0 \to \infty.$$

- Recall that $\mathrm{P}^{\mathrm{h}}(S_0, v_0, T, K) = V(0, S_0, v_0; T, K)$.
- Thus, $V(0, S_0, v_0; T, K) \to e^{-rT} K$ as $v_0 \to \infty$.

Superhedging price of a put in Heston model

$$V(0, S_0, v; T, K) \to e^{-rT} K \text{ as } v \to \infty,$$

$$S_t = S_0 \exp(rt + \int_0^t \sqrt{v_u}\, dW_u^{\mathbb{Q}}).$$

- It is easy to see (from the Markov property) that

$$V(\varepsilon, S_0, v; T, K) = V(0, S_0, v; T - \varepsilon, K).$$

- Thus,

$$V(\varepsilon, S_0, v; T, K) \to e^{-r(T-\varepsilon)} K \text{ as } v \to \infty.$$

- For any $\delta > 0$, with a strictly positive probability, $(v_t, W_t^{\mathbb{Q}})_{t\in[0,\varepsilon]}$ take paths such that

$$\sup_{t\in[0,\varepsilon]} \left| \int_0^t \sqrt{v_u}\, dW_u^{\mathbb{Q}} \right| \leq \delta$$

  and such that v_ε is very large.

Superhedging price of a put in Heston model

$$V(\varepsilon, S_0, v; T, K) \to e^{-r(T-\varepsilon)} K \text{ as } v \to \infty,$$
$$S_t = S_0 \exp(rt + \int_0^t \sqrt{v_u}\, dW_u^{\mathbb{Q}}).$$

- On such paths, $S_t \approx S_0 e^{rt \pm \delta}$ for $t \in [0, \varepsilon]$, and $v_\varepsilon \approx \infty$, which gives
$$V(\varepsilon, S_\varepsilon, v_\varepsilon; T, K) \approx V(\varepsilon, S_0 e^{r\varepsilon \pm \delta}, v_\varepsilon; T, K) \approx e^{-r(T-\varepsilon)} K.$$
- On the other hand, any hedge has a wealth process that depends only on S:
$$V_\varepsilon^\Delta = F_\varepsilon((S_t)_{t \in [0,\varepsilon]}).$$
- If F is continuous (in appropriate sense), then, choosing $\varepsilon, \delta > 0$ small enough:
$$V_\varepsilon^\Delta \approx F_\varepsilon((S_0 e^{rt \pm \delta})_{t \in [0,\varepsilon]}) \approx F_0(S_0) = V_0^\Delta,$$
$$V(\varepsilon, S_\varepsilon, v_\varepsilon; T, K) \approx e^{-rT} K.$$
- Since $V_\varepsilon^\Delta \geq V(\varepsilon, S_\varepsilon, v_\varepsilon; T, K)$, we conclude $V_0^\Delta \geq e^{-rT} K$.

The range of arbitrage-free prices

- Recall that SV models are typically incomplete. This issue is partially resolved by building a model under RNM $\mathbb{Q}$: e.g., we assume that Heston dynamics for (S_t, v_t) hold directly under $\mathbb{Q}$.
- This approach has some limitations: e.g., the model would not give the dynamics of (S_t, v_t) under the physical measure (we can only deduce the dynamics under $\mathbb{Q}$, by calibrating the parameters to options' prices).
- Alternatively, we can choose the type of SV model under $\mathbb{P}$, estimate its parameters, and consider the range of all arbitrage-free prices:

$$\left[B_0 \inf_{\mathbb{Q}} \mathbb{E}^{\mathbb{Q}}(X/B_T),\, B_0 \sup_{\mathbb{Q}} \mathbb{E}^{\mathbb{Q}}(X/B_T) \right],$$

  where X is the payoff of the option and $\mathbb{Q}$ ranges over all RNMs.
- Unfortunately, the above range is typically too large for practical purposes.
- For example, repeating the arguments on previous slides, one can show that, in Heston model,

$$B_0 \sup_{\mathbb{Q}} \mathbb{E}^{\mathbb{Q}}((K - S_T)^+/B_T) = \overline{V}((K - S_T)^+) = e^{-rT} K.$$

(Partial) Dynamic hedging

- Because of incompleteness and the unrealistically large superhedging price (which occurs when underlying is the only hedging instrument), we cannot expect to fully offset the risk of an open position in an option.
- Given a payoff $X \in \mathcal{F}_T$, find a self-financing strategy (Δ_t) (i.e., hold Δ_t units of underlying at each time t) with value (V_t), which solves
  - minimum squared error: $\min_\Delta \mathbb{E}^{\mathbb{P}/\mathbb{Q}}(V_T - X)^2$,
  - minimum expected downside: $\min_\Delta \mathbb{E}^{\mathbb{P}/\mathbb{Q}}(V_T - X)^-$,
  - minimum variation (aka quadratic hedging): $\min_\Delta [V - \pi(X)]_T$.
- The above seems like a complex optimization process, because its variable is an adapted process $\Delta = (\Delta_t)$.
- If the model and the payoff allow for a Markovian formulation of the above problems, the Dynamic Programming Principle (DPP) reduces their complexity significantly.

Quadratic hedging

- Consider an autonomous SV model
$$dS_t/S_t = r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}},$$
$$dv_t = a(v_t)\,dt + b(v_t)\,dZ_t, \quad dW_t^{\mathbb{Q}}\,dZ_t = \rho\,dt,$$
which is written under $\mathbb{Q}$ (this method remains unchanged if we work under $\mathbb{P}$).
- Consider the associated arbitrage-free price of an option with payoff X:
$$\pi(X)_t = \mathbb{E}^{\mathbb{Q}}[X\,e^{-r(T-t)}\,|\,\mathcal{F}_t].$$
- Assume that we can determine/compute the covariation between $\pi(X)$ and $W^{\mathbb{Q}}, Z$:
$$d\pi(X)_t = r\,\pi(X)_t\,dt + \xi_t\,dW_t^{\mathbb{Q}} + \zeta_t\,dZ_t.$$
- For example, if $\pi(X)_t$ can be written as $V^X(t, S_t, v_t)$,
$$\xi_t = S_t\sqrt{v_t}\,\partial_s V^X(t, S_t, v_t), \quad \zeta_t = b(v_t)\,\partial_v V^X(t, S_t, v_t).$$
- The above Markovian representation of (ξ, ζ) works for European options, but also for barrier and asian options, and some others.

Quadratic hedging

$$dS_t/S_t = r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}}, \quad dW_t^{\mathbb{Q}}\,dZ_t = \rho\,dt,$$
$$d\pi(X)_t = r\,\pi(X)_t\,dt + \xi_t\,dW_t^{\mathbb{Q}} + \zeta_t\,dZ_t.$$

- We aim to minimize the total quadratic variation $[V - \pi(X)]_T$ by choosing a strategy (Δ_t):

$$dV_t = \Delta_t\,dS_t + r\,(V_t - \Delta_t\,S_t)\,dt = \Delta_t\,S_t\,\sqrt{v_t}\,dW_t^{\mathbb{Q}} + r\,V_t\,dt,$$
$$d[V - \pi(X)]_t = \left((\Delta_t\,S_t\,\sqrt{v_t} - \xi_t)\,dW_t^{\mathbb{Q}} - \zeta_t\,dZ_t\right)^2$$
$$= \left((\Delta_t\,S_t\,\sqrt{v_t} - \xi_t)^2 - 2\rho\,(\Delta_t\,S_t\,\sqrt{v_t} - \xi_t)\,\zeta_t + \zeta_t^2\right)\,dt.$$

- Minimizing the above quadratic function over Δ_t, we obtain the optimal quadratic hedge:

$$\Delta_t = \frac{\xi_t + \rho\,\zeta_t}{S_t\,\sqrt{v_t}}.$$

Market completion with options

- Main idea: we can improve our hedge by introducing additional hedging instruments.
- Liquid European options are perfect candidates for this.
- The analysis of quadratic hedge shows that, typically, we need one hedging instrument per driving Brownian motion.

Sigma-hedge in Heston model

- Consider the Heston model

$$dS_t/S_t = r\,dt + \sqrt{v_t}\,dW_t^{\mathbb{Q}},$$
$$dv_t = \lambda\,(\bar{v} - v_t)\,dt + \eta\,\sqrt{v_t}\,dZ_t, \quad dW_t^{\mathbb{Q}}\,dZ_t = \rho\,dt,\ |\rho| \neq 1.$$

- We need to hedge an option with payoff X and price

$$\pi(X)_t = \mathbb{E}^{\mathbb{Q}}[X\,e^{-r(T-t)}\,|\,\mathcal{F}_t],$$
$$d\pi(X)_t = r\,\pi(X)_t\,dt + \xi_t\,dW_t^{\mathbb{Q}} + \zeta_t\,dZ_t.$$

- Consider a liquid European call with maturity $M \geq T$:

$$C_t = \mathbb{E}^{\mathbb{Q}}[(S_M - K)^+\,e^{-r(M-t)}\,|\,\mathcal{F}_t] = C(t, S_t, v_t),$$
$$dC_t = r\,C_t\,dt + S_t\,\sqrt{v_t}\,\partial_s C\,dW_t^{\mathbb{Q}} + \eta\,\sqrt{v_t}\,\partial_v C\,dZ_t.$$

Sigma-hedge in Heston model

$$d\pi(X)_t = r\,\pi(X)_t\,dt + \xi_t\,dW_t^{\mathbb{Q}} + \zeta_t\,dZ_t,$$
$$dC_t = r\,C_t\,dt + S_t\sqrt{v_t}\,\partial_s C\,dW_t^{\mathbb{Q}} + \eta\sqrt{v_t}\,\partial_v C\,dZ_t.$$

- A strategy (Δ, Σ) trading in (B, S, C) has the value (V_t):

$$\begin{aligned} dV_t &= r\left(V_t - \Delta_t S_t - \Sigma_t C_t\right)dt + \Delta_t\,dS_t + \Sigma_t\,dC_t \\ &= rV_t\,dt + \Delta_t\,S_t\sqrt{v_t}\,dW_t^{\mathbb{Q}} + \Sigma_t\left(S_t\sqrt{v_t}\,\partial_s C\,dW_t^{\mathbb{Q}} + \eta\sqrt{v_t}\,\partial_v C\,dZ_t\right). \end{aligned}$$

- Assuming that $V_0 = \pi(X)_0$ and choosing

$$\Sigma_t = \frac{\zeta_t}{\eta\sqrt{v_t}\,\partial_v C}, \quad \Delta_t = \frac{\xi_t}{S_t\sqrt{v_t}} - \Sigma_t\partial_s C,$$

we conclude that the strategy replicates the claim, $V_t = \pi(X)_t$ for all t.

Sigma-hedge in Heston model

$$\Sigma_t = \frac{\zeta_t}{\eta\sqrt{v_t}\,\partial_v C}, \quad \Delta_t = \frac{\xi_t}{S_t\sqrt{v_t}} - \Sigma_t \partial_s C.$$

- If $\pi(X)_t = V^X(t, S_t, v_t)$, then

$$\xi_t = S_t\sqrt{v_t}\,\partial_s V^X(t, S_t, v_t), \quad \zeta_t = \eta\sqrt{v_t}\,\partial_v V^X(t, S_t, v_t),$$
$$\Sigma_t = \frac{\partial_v V^X}{\partial_v C}, \quad \Delta_t = \partial_s V^X - \Sigma_t \partial_s C.$$

- Liquid option position Σ_t eliminates the sensitivity of the overall portfolio (the replicating strategy and a short position in claim X) to changes in volatility.
- Spot position Δ_t eliminates the sensitivity to changes in spot.
- This is known as "Sigma-hedge" (*Hull-White 1987, Scot 1988*).
- Note that we need $\partial_v C \neq 0$ – i.e., we need convexity in the additional hedging instrument. In particular, a linear instrument (spot or futures) would not work.

Market completion in general

- For models including additional sources of randomness (e.g., stochastic interest rates):
  - Add a hedging instrument for each additional source of randomness (a.k.a. Brownian motion).
  - It is crucial that the hedging instruments are sensitive to the associated risk factors (and of full rank).
  - Replication argument still works.
- Approach breaks down for most models with (unpredictable) jumps.
- Dynamic hedging with options can be expensive
  - Limited liquidity away from at-the-money (ATM) strikes.
  - Even ATM bid-ask spreads tend to be much larger than for stocks.

Vega immunization

- Recall the Sigma-hedge:

$$\Sigma_t = \frac{\partial_v V^X}{\partial_v C}, \quad \Delta_t = \partial_s V^X - \Sigma_t \partial_s C.$$

- If the hedged option is European, i.e., $X = H(S_T)$, then it can be written as a (possibly infinite) linear combination of calls or puts:

$$H(S_T) = \sum_j h_j \, (S_T - K_j)^+$$

- Then, a popular approach is to represent the prices of associated options via Black-Scholes formulas,

$$C_t = \mathrm{C}^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{imp}}(M,K), M, K), \;\; \pi(X)_t = \sum_j h_j \, \mathrm{C}^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{imp}}(T,K_j), T, K_j).$$

- and to replace $\partial_v C$ and $\partial_v V^X$ by

$$\partial_{\sigma^{\mathrm{imp}}} C = \partial_{\sigma^{\mathrm{imp}}} \mathrm{C}^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{imp}}(M,K), M, K),$$
$$\partial_{\sigma^{\mathrm{imp}}} V^X = \sum_j h_j \, \partial_{\sigma^{\mathrm{imp}}} \mathrm{C}^{\mathrm{BS}}(S_t, \sigma_t^{\mathrm{imp}}(T,K_j), T, K_j).$$

Vega immunization

- Recall the Sigma-hedge

$$\Sigma_t = \frac{\partial_v V^X}{\partial_v C}, \quad \Delta_t = \partial_s V^X - \Sigma_t \partial_s C.$$

- Replacing $\partial_v C$ and $\partial_v V^X$ by $\partial_{\sigma^{\mathrm{imp}}} C$ and $\partial_{\sigma^{\mathrm{imp}}} V^X$, we obtain the Vega immunization strategy (Δ_t, Σ_t):

$$\Sigma_t = \frac{\partial_{\sigma^{\mathrm{imp}}} V^X}{\partial_{\sigma^{\mathrm{imp}}} C}, \quad \Delta_t = \partial_s V^X - \Sigma_t \partial_s C.$$

- Note that the above hedge can be computed without prescribing any particular model for (S, v)! It is very popular among practitioners.
- In a two-factor SV model, where Sigma-hedge replicates the claim X perfectly, Vega immunization gives exactly the same strategy.
- Strictly speaking, Vega immunization is limited to European claims, but one can imagine less rigorous extensions of this approach to other claims, e.g., via regression.

Static hedging of European options with calls and puts

- *Carr-Madan 1988*: for any twice differentiable $H : [0, \infty) \to \mathbb{R}$ and any $x, a \geq 0$,

$$H(x) = H(a) + H'(a)(x - a) + \int_0^a H''(K)(K - x)^+ \, dK + \int_a^\infty H''(K)(x - K)^+ \, dK.$$

- As a consequence, the payoff $H(S_T)$ is exactly replicated by a static (buy and hold) position consisting of
  - a bond position, worth initially $e^{-rT}(H(a) - H'(a)a)$,
  - $H'(a)$ shares of the risky asset,
  - and an infinite combination of vanilla puts and calls, with the same expiration T and with all possible strikes K, taken with weights $H''(K)$.
- In practice, integrals must be truncated and discretized. Thus, real-world static hedging is approximate.

Valuation

- Apply the Carr-Madan formula with $a = e^{rT} S_0 = F_{0,T}$, multiply by the discount factor e^{-rT}, and take expectation, to obtain:

$$\mathbb{E}^{\mathbb{Q}}[e^{-rT} H(S_T)] = e^{-rT} H(F_{0,T})$$
$$+ \int_0^{F_{0,T}} H''(K)\, P(T,K)\, dK + \int_{F_{0,T}}^{\infty} H''(K)\, C(T,K)\, dK.$$

- Thus, we can compute the price of any European-type claim via vanilla puts and calls.
- This representation is model-independent in the sense that it does not depend on a specific choice of the distribution/dynamics of (S, v) (though it still depends on our assumption of linear pricing and ignores the transaction costs).
- If the underlying pays dividends, the formula $e^{rT} S_0 = F_{0,T}$ fails, but the Carr-Madan formula still holds.

Log contract

- A very important application of the Carr-Madan formula is the pricing of a log-contract:

$$
\begin{aligned}
& H(S_T) = \log S_T, \\
& H''(K) = -\frac{1}{K^2}, \\
& \mathbb{E}^{\mathbb{Q}}[e^{-rT} \log S_T] = e^{-rT} \log F_{0,T} \\
& - \int_0^{F_{0,T}} \frac{1}{K^2} P(T,K)\, dK - \int_{F_{0,T}}^{\infty} \frac{1}{K^2} C(T,K)\, dK.
\end{aligned}
$$

- This formula is used for variance swap replication/valuation.
- A particularly important special case of the above is the construction of VIX index.