PART 1
A PROBABILISTIC APPROACH TO LOCAL VOLATILITY
Call option: the right to buy a “stock” at future maturity date for strike-price agreed today (put—sell)
4 January 2010: 473 strike-maturity pairs
LOCAL VOLATILITY
Prices at all strike-maturities $\Leftrightarrow$ all one-dim. marginals of $(S_t)$

$\Leftrightarrow$ unique $(S_t)$ in \{continuous diffusion processes\}
Local volatility

• Stock-price model, $Q$-dynamics

\[ dS_t = rS_t dt + \sigma(t, S_t) S_t dW_t, \quad t \in [0, T^*] \]

• The local volatility function—our modelling target

\[ \sigma : [0, T^*] \times \mathbb{R}_+ \to \mathbb{R} \]
Local volatility

• Dupire’s equation defines operator, $\sigma(\cdot) \leftrightarrow C(\cdot)$

$$
\frac{\partial C}{\partial T} + r K \frac{\partial C}{\partial K} - \frac{K^2 \sigma^2(T, K)}{2} \frac{\partial^2 C}{\partial K^2} = 0
$$

$$
C(0, K) = (s - K)^+
$$

• Dupire’s formula its inverse, $C(\cdot) \leftrightarrow \sigma(\cdot)$

$$
\sigma(T, K) = \sqrt{\frac{\frac{\partial C}{\partial T} - r (C - K \frac{\partial C}{\partial K})}{\frac{1}{2} K^2 \frac{\partial^2 C}{\partial K^2}}}
$$

• Explicit: all call prices $\iff$ unique diffusion process $(S_t)$
Why local volatility?

- A model can reproduce a given set of option prices if within range attainable by model.
- For local volatility, any set of call prices attainable as long as consistent in the sense of no static arbitrage.
- Dupire’s formula gives recipe for calibration:

\[
\sigma(T, K) = \sqrt{\frac{\frac{\partial C}{\partial T} - r(C - K \frac{\partial C}{\partial K})}{\frac{1}{2} K^2 \frac{\partial^2 C}{\partial K^2}}}.
\]
Why local volatility?

- A model can reproduce a given set of option prices if within range attainable by model

- For local volatility, *any set* of call prices attainable as long as consistent in the sense of no static arbitrage

- Dupire’s formula gives recipe for calibration

\[
\sigma(T, K) = \sqrt{\frac{\partial C}{\partial T} - r(C - K \frac{\partial C}{\partial K}) \frac{1}{2 \frac{\partial^2 C}{\partial K^2}}}
\]

- However, *uniqueness* only in the limit of infinite data
In practice
Probabilistic Approach
• We use a *Gaussian process* to define a prior distribution over local volatility

• Likelihood: data is *noisy observations* of true fair price

• Infer *posterior distribution* over local volatility from observed data
Gaussian process

- A distribution over functions

\[ f : \mathcal{X} \mapsto \mathbb{R} \]

- Specified by mean- and covariance functions

\[ \mu : \mathcal{X} \rightarrow \mathbb{R} \]
\[ k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} \]

- Notation

\[ f \sim \mathcal{GP}(\mu, k) \]
Gaussian process

• For any finite set of inputs, induces a Gaussian distribution over function values

\[ \mathbf{x} = \{x_i\}, \quad x_i \in \mathcal{X} \]
\[ \mathbf{f} = f(\mathbf{x}), \quad \mathbf{f} \sim \mathcal{N}(\mu, \mathbf{K}) \]

• Mean vector and covariance matrix given by

\[ \mu_i = \mu(x_i) \]
\[ K_{i,j} = k(x_i, x_j) \]
Gaussian process

Given data \((x, y)\)

1. train (hyper)parameters by likelihood (Type-II)

\[
\log p(y|\theta) = -\frac{1}{2} (y - \mu_\theta)^\top K_\theta^{-1} (y - \mu_\theta) - \frac{1}{2} \log |2\pi K_\theta|
\]

2. predict \(f\) at unseen inputs(s)

\[
f^* \sim \mathcal{GP}(\mu^*(x), k^*(x, x')) \\
\mu^*(x) = \mu(x) + k(x, x)^\top K^{-1} (y - \mu) \\
k^*(x, x') = k(x, x') - k(x, x)^\top K^{-1} k(x', x)
\]
Gaussian process

Cost \( O(N^3) \)

\[
\log p(y) = -\frac{1}{2} (y - \mu_\theta)^\top K_\theta^{-1} (y - \mu_\theta) - \frac{1}{2} \log |2\pi K_\theta|
\]

\[
f^* \sim \mathcal{GP}(\mu^*(x), k^*(x, x'))
\]

\[
\mu^*(x) = \mu(x) + k(x, x)^\top K^{-1} (y - \mu)
\]

\[
k^*(x, x') = k(x, x') - k(x, x)^\top K^{-1} k(x', x)
\]
Set-up

\[ f(x) \sim \mathcal{GP}(\mu, k(x, x')) \]
\[ \sigma(x) = \phi(f(x)) \]
\[ \phi(f) = \log(1 + \exp(f)) > 0 \]

\[ \mathcal{X} = \{\text{strikes, maturities}\} \]
Set-up

Pricing: (latent) function to function mapping

\[ f(\cdot) \xrightarrow{\phi} \sigma(\cdot) \xrightarrow{\text{PDE}} C(\cdot, f) \]
Set-up

Pricing: discrete surface to surface mapping

\[ \mathbf{f} \xrightarrow{\phi} \sigma \xrightarrow{\text{FD}} C(x, \mathbf{f}) \]
Set-up

- Observations $\mathbf{c} = \{ c_i \}$ (mid of bid-ask) are fair prices with noise (~market frictions)

$$c_i = C(x_i, \mathbf{f}) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2_\epsilon)$$

- Likelihood

$$\log p(\mathbf{c}|\mathbf{f}, \sigma_\epsilon) = -\frac{1}{2\sigma^2_\epsilon} \sum_{i=1}^{N} (C(x_i, \mathbf{f}) - c_i)^2 - \frac{N}{2} \log(2\pi \sigma^2_\epsilon)$$

- Remark: *neither factorising nor Gaussian in $\mathbf{f}$*
Set-up

- GP prior: smooth local volatility surfaces

\[ k_K(x, x') = \sigma_f^2 \exp \left( -\frac{(T-T')^2}{2l_T^2} \right) \exp \left( -\frac{(K-K')^2}{2l_K^2} \right) \]

- Hyperprior: scaled sigmoid Gaussian

\[ \theta = \theta_{\text{min}} + \frac{\theta_{\text{max}} - \theta_{\text{min}}}{1 + \exp(-z)}, \quad z \sim \mathcal{N}(m, K) \]
Inference

• Sampling from posterior

\[ p(f, \kappa, \sigma_\epsilon, \mu | c) = \frac{1}{Z} p(c | f, \sigma_\epsilon) p(f | \kappa, \mu) p_h(\kappa, \sigma_\epsilon, \mu) \]

• Markov chain Monte Carlo (blocked Gibbs)

Update 1:

\[ f | \kappa, \sigma_\epsilon, \mu \sim p(c | f, \sigma_\epsilon, \mu) p(f | \kappa) \]

Update 2:

\[ \kappa, f | \sigma_\epsilon, \mu \sim p(c | f, \sigma_\epsilon, \mu) p(f | \kappa) p_h(\kappa) \]

Update 3:

\[ \sigma_\epsilon, \mu | \kappa, f \sim p(c | f, \sigma_\epsilon, \mu) p_h(\sigma_\epsilon, \mu) \]
Inference

Update 1: \( f | \kappa, \sigma_\epsilon, \mu \sim \mathcal{L}(f) \mathcal{N}(f; 0, K_\kappa) \)

\[ \theta \sim \text{Unif}[0, 2\pi], \quad \nu \sim \mathcal{N}(0, K_\kappa) \]

\[ f' = f \cos(\theta) + \nu \sin(\theta) \]

\[ f' \rightarrow f \quad \text{by slice sampling } \theta \]

Cost: \( 1 \times O(N^3) \) for covariance (\( O(N^{1.5}) \) w. Kronecker), \( O(N) \) for likelihood
Inference

Update 2: \( \kappa \mid \nu, \sigma, \mu \sim \mathcal{L}(f(\kappa, \nu)) \, \rho_h(\kappa) \)

\[ f = L_\kappa \nu, \quad \nu \sim \mathcal{N}(0, I), \quad L_\kappa = \text{chol}(K_\kappa) \]

\( \kappa \rightarrow \kappa'; \quad f' = L_{\kappa'} \nu \quad (f \rightarrow f', \ \nu \ \text{fixed}) \)

Mixing depends on strength of likelihood relative prior; cost \( k \times O(N^3) \)
RESULTS
Data
117 quotes over 20x8 grid (of 473 quotes over 103x14)
Markov chain Monte Carlo
Markov chain Monte Carlo

- Blunt strategy, robust to tuning parameters (if any!)
- Sum of squared errors for assessing convergence

\[
\log p(c|f, \sigma_\epsilon) = -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^{N} (C(x_i, f) - c_i)^2 - \frac{N}{2} \log(2\pi\sigma_\epsilon^2) \]

\[\underbrace{\sum_{i=1}^{N} (C(x_i, f) - c_i)^2}_{SSE(f)}\]

- Need long run for chain to reach stationarity
- Improved convergence by informed priors
Markov chain Monte Carlo
Posterior local volatility

- Strike: USD 1000, 1500
- Maturity: 0.5, 1.0, 1.5, 2.0, 2.5
- Local volatility: 0.0, 0.2, 0.4, 0.6, 0.8

±2SD
MAP
data loc.

Maturity: 0.13 year
Posterior local volatility

![3D plot of posterior local volatility with strike in USD and maturity in years, and a 2D plot showing the local volatility and strike with map and data locations. The 2D plot indicates a maturity of 0.99 year.]
Posterior local volatility

![3D plot of posterior local volatility with strike in USD and maturity in years. The plot shows a decrease in local volatility with increasing strike and maturity.]

![2D plot of local volatility against strike with maturity 3 years. The plot includes a shaded area representing ±2SD and a line for the MAP.]

maturity: 3 year
Posterior option prices

![Call Price Graph](image)

- Strike: 600 to 1800
- Call Price: 0 to 500
- Maturity: 0.13 year

![Implied Volatility Graph](image)

- Strike: 600 to 1800
- Implied Volatility: 0.0 to 0.6
- Maturity: 0.13 year

- ±2SD
- MAP
- Data
Posterior option prices

maturity: 0.48 year
Posterior option prices

- **Call Price**
  - **Strike** vs. **Call Price**
  - **Maturity**: 0.99 year

- **Implied Volatility**
  - **Strike** vs. **Implied Volatility**
  - **Maturity**: 0.99 year

- **MAP Data**
  - Includes ±2SD range
Posterior option prices

Call price

0 100 200 300 400 500
600 800 1000 1200 1400 1600 1800
strike

maturity: 3 year

±2SD
MAP
data

Implied volatility

0 0.1 0.2 0.3 0.4 0.5 0.6
600 800 1000 1200 1400 1600 1800
strike

maturity: 3 year

±2SD
MAP
data
Posterior delta

delta  vs  strike [USD] for different maturities:

- Maturity: 0.13 year
- Maturity: 0.28 year
- Maturity: 0.99 year
- Maturity: 3 years
LOCAL VOLATILITY TIME SERIES
Market dynamics
Volatility dynamics

- Local volatility model known to provide good static fit of option market; need to re-calibrate on daily (say) basis

- GP framework provide way of inferencing dynamical behaviour by introducing prior temporal dependency
Volatility dynamics

- Augmented input space

\[ x_t \equiv (t, x) = (t, T, K) \]
\[ f : x_t \mapsto f(x_t) \]
\[ f_t \equiv f(x_t), \quad f \equiv \{ f_t \}_{t=1}^{M} \]

- Prices conditionally independent over time and internally across surfaces

\[ c \equiv \{ c_t \}_{t=1}^{M}, \quad p(c|f + \mu, \sigma_\epsilon) = \prod_{t=1}^{M} p(c_t|f_t + \mu, \sigma_\epsilon) \]
Volatility dynamics

- Unchanged generative model
- Joint posterior

\[
p(f_{1:t}, \kappa, \mu, \sigma_\epsilon | c_{1:t}) = \frac{1}{Z} \mathcal{L}_{\mu,\sigma_\epsilon}(f_{1:t}) p(f_{1:t} | \kappa) p_h(\kappa, \mu, \sigma_\epsilon)
\]

- Sequential MCMC on approximative posterior
RESULTS
The Gaussian process provides predictive distribution over local volatility (price, implied volatility etc.)

Since model dependency in time, may predict forward in time

We make one-week ahead predictions
Prediction
Prediction
VIX-index: 30-day forward volatility measure implied by market—1-week ahead predictions
VIX-index: 30-day forward volatility measure implied by market—1-week ahead predictions
Concluding remarks

- Approach for probabilistic modelling of local volatility
- Prior domain knowledge straightforward to encode
- Gives means for quantifying uncertainty
Concluding remarks

• Connection with “classical” calibration (regularised optimisation) but gives formal link to probabilistic representation

• Include time-dimension in input space, gives means for inferring behaviour of (local) volatility over time, and for predicting future surfaces

• Computational method (MCMC) expensive!
PART 2
A PROBABILISTIC APPROACH TO REALISED VOLATILITY
Motivation
Volatility is rough

- Take log-volatility to follow (fOU)

\[ dX^H = -\alpha X^H dt + \sigma dW^H, \quad H \in (0, 1) \]

Hurst parameter \( H < 1/2 \) for rough process

- “behaves locally”, \( \alpha \to 0 \), as fractional BM

\[ E[|W^H_{t+\Delta} - W^H_t|^q] = c_q \Delta^{qH} \]
Volatility is rough

3962 daily 'observations' from Oxford-Man’s realised library
Volatility is rough

- From S&P 500 realised variance, $H = 0.144$
Simulation study

• Take a classic Ornstein-Uhlenbeck and estimate parameters from S&P 500 log-variance

\[(\alpha, m, \sigma) = (55.8, -4.2, 11.3)\]

• Use to simulate an OU-path and re-estimate \(H\)
Simulation study

- From simulated ($H = 0.5$) data, estimate $H = 0.14$
Hurst estimate

- For any Gaussian process with stationary increments

\[ E[|X_{t+\Delta} - X_t|^q] = c_q \sigma(\Delta)^q, \quad q \geq 0 \]

- The first step imposes a linear fit to the log-RHS

\[ f(\log \Delta) = \tilde{c}_q + q \left( \frac{\log \sigma(\Delta)}{\log \Delta} \right) \log \Delta \]

- It assumes \( q \beta(\Delta) = qH \) from the fBM with \( \sigma(\Delta) = \Delta^H \)
Hurst estimate

• Estimate $f(\log \Delta)/q = \tilde{c}_q + \beta(\Delta) \log \Delta$ from simulated OU

• Linear, $\beta(\Delta) = 0.5$, for $\Delta \to 0$, levels off for $\Delta \to T$
Hurst estimate

- Estimate from S&P 500 volatility

- Linear, $\beta(\Delta) \approx 0.14$, for $\Delta \to 0$, levels off for $\Delta \to T$?
Hurst estimate

- Simulate with *huge* rate of mean reversion

\[ \delta_m(q, \delta) \]

\[ m(q, \delta)^{1/q} \]

- S&P 500

- \( q = 0.25 \)
- \( q = 2 \)
- \( q = 6 \)
Probabilistic Approach
Linear filter

• Linear filter representation of fOU process; $X \sim \text{OU}$

$$X \xrightarrow{g} X^H$$

$$X_t^H = \int g(\omega) e^{i\omega t} dZ_X(\omega)$$

• The frequency response

$$|g(\omega)|^2 = c_H |\omega|^{1-2H}$$

shifts spectral mass to high frequencies if $H < 1/2$
Linear filter

• In time-domain, corresponds to convolution

\[ X_t^H = \int h(t - u)X_u \, du \]

\[ h(u) = \frac{1}{2\pi} \int g(\omega) e^{i\omega u} \, d\omega \]
Gaussian process model of linear filter

• Generative model

\[
\begin{align*}
\theta & \sim p(\theta) \\
x & \sim \mathcal{GP}(0, k_x(t, t')) \\
h & \sim \mathcal{GP}(0, \delta_{t'}(t)) \\
f(t) &= \int_{-\infty}^{t} e^{-\alpha(t-u)} h(t-u) x(u) du \\
y(t) | f(t) & \sim \mathcal{N}(f(t), \sigma^2)
\end{align*}
\]
Inference

• Given data \((t, y)\)

\[
p(x, h, \theta | y) = \frac{1}{p(y)} p(y | x, h, \sigma) p(x | \kappa_x) p(h | \kappa_h) p(\theta)
\]

- likelihood
- prior
- hyperprior

• Approach with variational inference

\[
q(x, h, \theta) \approx p(x, h, \theta | y)
\]
Inference

- Inducing variables: summarise $h$ and $x$ into finite dim. variables $u$ and $z$

$$u(t) = \int_{-\infty}^{t} e^{-\gamma(t-u)} h(u) du$$

$$z(m) = \langle x, \cos(2\pi m \cdot) \rangle_H$$

$$p(x, h, z, u, \theta | y) = \frac{1}{p(y)} p(y | x, h, \sigma) \underbrace{p(x | z, \kappa_x)}_{x \sim \mathcal{GP}} p(h | u, \kappa_h) \underbrace{p(z | \kappa_x)}_{\mathcal{N}(0, K_z)} p(u | \kappa_h) p(\theta)$$
Inference

- Inducing variables: summarise $h$ and $x$ into finite dim. variables $u$ and $z$

\[
\begin{align*}
u(t) &= \int_{-\infty}^{t} e^{-\gamma(t-u)} h(u) du \\
z(m) &= \langle x, \cos(2\pi m \cdot) \rangle_{\mathcal{H}}
\end{align*}
\]

\[
q(x, h, z, u, \theta) = \frac{1}{p(y)} p(y|x, h, \sigma) p(x|z, \kappa_x) p(h|u, \kappa_h) q(z)q(u)q(\theta)
\]

\[
\underbrace{x|z \sim \mathcal{GP}}_{\text{free dist.}}
\]
Inference

- Jensen’s inequality yields evidence lower bound

\[
\log p(y) = \log \int \frac{p(y|f)p(f, h, z)}{q(f, h, z)} q(f, h, z) df\, dh\, dz
\]

\[
\geq E_{q(f, h, z)} \left[ \log \left( \frac{p(y|f)p(f, h, z)}{q(f, h, z)} \right) \right] = \mathcal{F}
\]

\[
= \log p(y) - KL \left[ q(f, h, z) \mid \mid p(f, h, z|y) \right]
\]
Inference

- Optimising $\mathcal{F}$ with respect to $q$-coordinates yields

\[
\begin{align*}
\log \hat{q}(u) &= \mathbb{E}_{q(x,h,z,\theta|u)} [\log p(y|x, h, \sigma_\epsilon)] + \mathbb{E}_{q(\theta)} [\log p(u|\kappa_h)] + c \\
\log \hat{q}(z) &= \mathbb{E}_{q(x,h,u,\theta|z)} [\log p(y|x, h, \sigma_\epsilon)] + \mathbb{E}_{q(\theta)} [\log p(z|\kappa_x)] + c \\
\log \hat{q}(\theta) &= \mathbb{E}_{q(x,h,u,z|\theta)} [\log p(y|x, h, \sigma_\epsilon)] + \mathbb{E}_{q(u,z)} [\log p(u, z|\theta)] + \log p(\theta)
\end{align*}
\]
Inference

- Key: expected log-likelihood is quadratic in $u$ and $z$

\[
\log \hat{q}(u) = \log \mathcal{N}(u; \mu_u, \Sigma_u) \\
\mu_u = \Sigma_u \mathbb{E}_{\hat{q}(\theta)}[X] \mu_z \\
\Sigma_u = \left( \mathbb{E}_{\hat{q}(\theta)} \left[ -2Y \Sigma_z + \mu_z \mu_z^\top + K_u^{-1} \right] \right)^{-1}
\]
Inference

- Key: expected log-likelihood is quadratic in $u$ and $z$

$$\log \hat{q}(u) = \log \mathcal{N}(u; \mu_u, \Sigma_u)$$

$$\mu_u = \Sigma_u \mathbb{E}_{\hat{q}(\theta)}[X] \mu_z$$

$$\Sigma_u = \left( \mathbb{E}_{\hat{q}(\theta)} \left[ -2Y\Sigma_z + \mu_z\mu_z^\top + K_u^{-1} \right] \right)^{-1}$$

- Iterate $q(u)$ and $q(z)$, substitute in $\mathcal{F}$ and optimise $q(\theta)$
Results

- S&P 500 realised volatility, 14 months of data
Results
Concluding remarks

• Sampling frequency (~daily) puts restrictions to what can be inferred in higher frequencies

• Prior model dominates $\Delta \rightarrow 0$ scenario

• Model acknowledges noise; absorbs variation in data, and puts more spectral mass over low frequencies

• Potential extension: model an integrated variance process (noise $\sim$ RV estimation)
Thank you for your attention!