Optimal Bayesian Hedging Strategies

26th November 2009, University of Oxford

Alok Gupta
D.Phil Mathematical Finance
alok.gupta@maths.ox.ac.uk

MCFG, Mathematical Institute, Oxford University
Nomura Bank
EPSRC

Supervised by Christoph Reisinger
Contents

• Introduction
  - Calibration Problem
  - Bayesian Estimators
  - Consistency
  - Example: Local Volatility

• Hedging in the Presence of Model Uncertainty
  - Hedging Formulation
  - Motivating Examples
  - Bayesian Hedging Strategies

• Hedging Error Loss Functions
  - Examples
  - Hedging Improvement
  - Link to Utility Functions

• Numerical Examples

• Conclusion & Extensions
Introduction
Motivation

• Since Black-Scholes model proposed in 1973, huge growth in variety of financial models to capture behaviour of different markets e.g. stochastic interest rate models, credit models, etc.

• Agent will typically want to use model to price and hedge an instrument but before she can do this she must calibrate model to observable prices to avoid introducing arbitrage.

• Calibration not straight forward: instead of Black-Scholes single parameter, now calibrate vectors and functions e.g. Levy density, local volatility.

• Perfect calibration not possible — introducing problem of uniqueness. This leads to competing hedging strategies.

• Wealth of literature on local volatility hedging e.g. McIntyre (1999), Hull & Suo (2002), Coleman et al (2003).
**Calibration Problem**

Suppose we observe a price process \( S = (S_t)_{t \geq 0} \) and model it as a function of time \( t \), some stochastic process(es) \( X = (X_t)_{t \geq 0} \), and finite dimensional parameter \( \theta \in \Theta \), i.e.

\[
S_t = S(t, (X_u)_{0 \leq u \leq t}, \theta)
\]  

(1)

by abuse of notation of \( S \). Let \( \mathcal{F} = (\mathcal{F}_t)_{t \geq 0} \) be the filtration generated by \( X \) so \( S \) is an \( \mathcal{F} \)-adapted process.

Now consider an option \( X \) over a finite time horizon \([0, T]\) written on \( S \) and with payoff function \( h \). Let the time \( t \) model value of this option be written as \( f_t(\theta) \), where we include the argument \( \theta \) to emphasise the dependence of this price on the model parameters. Explicitly,

\[
f_t(\theta) = \mathbb{E}[B^{-1}(t, T)h(S(\theta))|\mathcal{F}_t]
\]

with respect to some measure \( \mathbb{P} \) (depending on \( \theta \)) and where \( B^{-1}(t, T) \) is the discount factor, possibly stochastic.
Calibration Problem

Suppose at time $t \in [0, T]$ we observe a set of such option prices

$$\{f_t^{(i)}(\theta) : i \in I_t\}$$

possibly with noise $\{e_t^{(i)} : i \in I_t\}$. In other words, we observe

$$V_t^{(i)} = f_t^{(i)}(\theta) + e_t^{(i)}$$

for $i \in I_t$.

Then the calibration problem is to find the value of $\theta$ that best reproduces the observed prices $\{V_t^{(i)} : i \in I_t, t \in \Upsilon_n([0, T])\}$, for some measurement of best. Here

$$\Upsilon_n([0, T]) = \{t_1, \ldots, t_n : 0 = t_1 < t_2 < \ldots < t_n \leq T\}$$

is a partition of the interval $[0, T]$ into $n$ parts. We can then use this parameter $\theta$ to hedge another claim $Y$. 
Bayesian Estimators

Suppose we wish to estimate the value of some parameter $\theta$. Assume we have some prior information for $\theta$ (for example that it belongs to a particular space, or is positive, or represents a smooth function), summarised by a prior density $p(\theta)$ for $\theta$. And suppose we observe some noisy data $V = \{V_t : t \in \Upsilon_n\}$ related to $\theta$ by

$$V_t = f_t(\theta) + e_t$$

for all $t \in \Upsilon_n$ where $e_t$ is some random noise and $\Upsilon_n$ is an index set of size $n$. Then $p(V|\theta)$ is the probability of observing the data $V$ given $\theta$ and is called the likelihood function.

Application of Bayes rule gives that the posterior density of $\theta$ is given by

$$p(\theta | V) \propto p(V|\theta) p(\theta).$$

We can use the posterior to find distributions/estimates of other quantities of interest.
Loss Functions

The loss function $L(\theta, \theta')$ gives the deficit incurred by taking $\theta'$ as the estimator for $\theta$. It must satisfy

\[
\begin{align*}
L(\theta, \theta') &= 0 & \text{if } \theta' = \theta \\
L(\theta, \theta') &> 0 & \text{if } \theta' \neq \theta.
\end{align*}
\]

Given data $V$, the corresponding Bayes estimator $\theta_L(V)$ is the value of $\theta$ which minimises the expected loss with respect to the posterior i.e.

\[
\theta_L(V) = \arg \min_{\theta'} \left\{ \int L(\theta, \theta') p(\theta|V) \, d\theta \right\}.
\]

Since the loss function should penalise estimators which are further from the true value, we assume $L$ is a (not necessarily strictly) increasing function of $|\theta - \theta'|$. 
Consistency

Suppose the price noises are given by $e_t \sim N(0, \varepsilon_t^2)$ with $\varepsilon_t \in [c, C] \subseteq \mathbb{R}^+$ and are independent of each other and the underlying driving process. Take a nested sequences of partitions $\Upsilon_n \supset \Upsilon_{n-1}$ and a loss function $L$. Let the r.v. $\theta_n(V) \sim p_n(\theta|V)$ and let $\theta^*$ be the true parameter value. Define the sequence of Bayes estimators $\hat{\theta}$ by,

$$
\hat{\theta}_n(V) = \arg \min_{\theta' \in \Theta} \left\{ \int_{\Theta} L(\theta, \theta') p_n(\theta|V) \, d\theta \right\}
$$

where $\Theta$ is the support of the posterior density $p_n(\theta|V)$ which is explicitly given by

$$
p_n(\theta|V) = \prod_{t \in \Upsilon_n} \frac{1}{\sqrt{2\pi\varepsilon_t}} \exp \left\{ -\frac{1}{2\varepsilon_t^2} (V_t - f_t(\theta))^2 \right\} \frac{p(\theta)}{p_n(V)}.
$$

There only exist consistency results (e.g. Fitzpatrick (1991)) for i.i.d. observations.
Consistency

With the following assumptions we have that for all scalar and vector $\theta$ the subsequent Lemma and Theorem hold.

**Assumption.** The prior $\mathbb{P}_{prior}$ (corresponding to prior density $p(\theta)$) and its support $\Theta$ satisfy:

1. $\forall \xi > 0$, $\mathbb{P}_{prior}[\|\theta - \theta^*\| < \xi] > c \xi$ for some constant $c > 0$

2. $\Theta$ is closed and bounded

**Assumption.** For each $t$, conditional on $\mathcal{F}_t$ the function $f_t(\theta)$ (which is a mapping $f_t : \Theta \rightarrow \mathbb{R}$ where $\Theta \subseteq \mathbb{R}^m$) satisfies the following. Define

$$\Upsilon_n(\theta; k) = \left\{ t \in \Upsilon_n : \frac{1}{\varepsilon_t} \left| \frac{f_t(\theta) - f_t(\theta^*)}{\|\theta - \theta^*\|} \right| > k \right\}.$$ 

Then for each $\theta \in \Theta$ there exists $k_\theta > 0$ such that $|\Upsilon_n(\theta; k_\theta)| \rightarrow \infty$ as $n \rightarrow \infty$. 
Consistency

Lemma. For all \( V \) \( \sigma_n(V) \xrightarrow{P} \sigma^* \).

Proof. (Outline)

- Write \( p_n(\sigma|V) = q_n(V)p(\sigma)e^{-\frac{1}{2}\phi_n(\sigma,V)} \) then can show
  \[
  \left| \frac{2u(\sigma - \sigma^*)}{\phi_n(\sigma, Y)} \right| \leq \frac{1}{\alpha_n} \to 0 \quad n \to \infty
  \]

- Define the moment generating function
  \[
  \varphi_n(u) = \mathbb{E}[e^{u(\sigma_n - \sigma^*)}]
  \]
  then it follows that \( \varphi_n(u) \to 1 \) as \( n \to \infty \) i.e. Dirac density \( \delta(\sigma - \sigma^*) \).

- By Levy’s Continuity Theorem this implies that \( \sigma_n(V) \xrightarrow{D} \sigma^* \) where \( \sigma^* \) is a constant almost surely.

- Hence \( \sigma_n(V) \xrightarrow{P} \sigma^* \).
**Consistency**

**Theorem.** For all $L$ bounded and continuous on $\Sigma$ the Bayes estimator $\hat{\sigma}_n(V)$ is consistent.

**Proof.** (Outline)

- First observe that we can write $L(\sigma, \sigma') = l(\sigma - \sigma')$ for some function $l$.
- $\mathbb{P}_{\sigma^*}[|\hat{\sigma}_n(V) - \sigma^*| \geq \delta]$ 
  
  $\leq \mathbb{P}_{\sigma^*}[|\hat{\sigma}_n(V) - \sigma_n(V)| \geq \frac{1}{2}\delta] + \mathbb{P}_{\sigma^*}[|\sigma_n(V) - \sigma^*| \geq \frac{1}{2}\delta]$ 

- But $\mathbb{P}_{\sigma^*}[|\sigma_n(V) - \sigma^*| \geq \frac{1}{2}\delta] \to 0$ as $n \to \infty$ by above lemma.

- And can show $\mathbb{P}_{\sigma^*}[|\hat{\sigma}_n(V) - \sigma_n(V)| \geq \frac{1}{2}\delta] \to 0$ as $n \to \infty$ for $L$ bounded and continuous.

- Hence, for all $\delta > 0$, $\mathbb{P}_{\sigma^*}[|\hat{\sigma}_n(V) - \sigma^*| \geq \delta] \to 0$ as $n \to \infty$. 

\[\square\]
Example: Local Volatility

Corresponding to the model originally proposed by Black & Scholes, let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0 \leq t \leq T}, (Z_t)_{0 \leq t \leq T})$ be the standard Wiener space i.e. $Z_t$ is Brownian motion, $\mathcal{F}_t$ is the natural filtration of $Z_t$ over $\Omega$ and $\mathcal{F} = \mathcal{F}_T$. Then the underlying asset price $S$ is given by

$$dS_t = \mu S_t dt + \sigma S_t dZ_t$$

where $\mu$ is the drift and $\sigma$ the volatility. In the local volatility model we choose $\sigma$ to be a function of both the asset price and the time:

$$\sigma = \sigma (S, t).$$

Although Dupire found an explicit formula to calculate this function using the implied volatility surface, the resulting local volatility surface is unstable and spikey. Furthermore, the formula depends on knowledge of the prices of options for all strikes and maturities, which is usually not available in practice.
Example: Local Volatility

Instead, we identify key characteristics expected of the local volatility surface that can be recast into a Bayesian prior. There are three properties we would expect of $\sigma(S, t)$:

**Positivity:** $\sigma(S, t) > 0$ for all values of $S$ and $t$; since the price variation squared $\sigma^2 > 0$ we adopt the convention $\sigma > 0$.

**Smoothness:** there should be no sharp spikes or troughs in the surface; no reason why current prices should be able to predict abrupt changes in future volatility.

**Consistency:** for small values of $t$ especially, $\sigma$ should be close to today’s at-the-money (ATM) volatility $\sigma_{atm}$. 
The Prior (Regularisation)

For the purposes of introducing the theory we consider the simplest density - the Gaussian density. It is also the second order approximation to any density. In light of the assumptions presented earlier we take for our prior

\[ p_{lv}(\sigma) \propto \exp \left\{ -\frac{1}{2} \tilde{\lambda} \| \log(\sigma) - \log(\sigma_{atm}) \|_2^2 \right\} \]

where \( \| \cdot \|_\kappa \) is a Sobolev norm given by

\[ \| u \|_\kappa^2 = (1 - \kappa) \| u \|_2^2 + \kappa \| \nabla u \|_2^2. \]

Working in the logarithmic space guarantees \( \sigma \) is positive and the norm ensures greater prior density is attached to \( \sigma \) that are both smoother and closer to ATM volatility.

\( \tilde{\lambda} \) quantifies how strong our prior assumptions are: a higher value of \( \tilde{\lambda} \) indicating greater confidence in our assumptions.

Clearly, those \( \theta \) which better satisfy prior beliefs have greater prior density.
The Likelihood (Calibration)

Let $V_t^{(i)}$ be the market observed price at time $t$ of a European call and $f_t^{(i)}(\theta)$ the corresponding theoretical price. Then define the basis point square-error functional as

$$G_t(\theta) = \frac{10^8}{S_t^2} \sum_{i \in I} w_i |f_t^{(i)}(\theta) - V_t^{(i)}|^2$$

where the $w_i$ are weights summing to one. But only attach positive Bayesian posterior density if parameter reproduces prices to within their spreads:

$$G(\theta) \leq \delta^2$$

where $\delta^2 = \sum_{i \in I} w_i \delta_i^2$ is the pre-specified tolerance. Hence, for the Bayesian likelihood for non-parametric models we will take

$$p(V|\theta) = 1_{G(\theta) \leq \delta^2} \exp \left\{ -\frac{1}{2\delta^2} G(\theta) \right\} .$$

So those surfaces $\sigma$ which reproduce prices closest to the market observed prices $V$ have the greatest likelihood values.
The Posterior

Combining the prior and likelihood functions we get the explicit form for the posterior function \( p(\theta|V) \) as

\[
p(\theta|V) \propto 1_{G(\theta) \leq \delta^2} \exp \left\{ -\frac{1}{2\delta^2} \left[ \lambda \|\theta\|^2 + G(\theta) \right] \right\}.
\]

**Remark.** Observe that maximising the posterior is equivalent to minimising the expression

\[
\lambda \|\theta\|^2 + G(\theta)
\]

which is exactly the form of functional authors such as Lagnado & Osher (1997) and Jackson, Suli & Howison (1999) seek to minimise to find their optimal calibration parameter. This is not a coincidence but an insight into how the Bayesian approach reformats traditional Tikhonov regularisation methods into a unified and rigorous framework.
The Posterior

Priced 66 European call options (on a known local volatility surface) with 11 strikes and 6 maturities and added Gaussian noise. We calibrate a 27-node surface. 479 calibrated surfaces are sampled from the posterior:
Bayesian Pricing

Prices for a 3 month at-the-money up-and-out barrier call option with barrier $1.1S_0$. Included are the true price with its bid-ask spread, the MAP price, and the Bayes price ($\frac{1}{N} \sum_{i=1}^{N} f(\theta_i)$) with associated posterior pdf of prices.
Hedging in the Presence of Model Uncertainty
Model uncertainty results from not being able to find the correct model underlying observed data. Many authors have studied the impact on hedging:

- Branger & Schlag (2004) calculate the correction to the Black-Scholes delta hedge when true underlying is Heston stochastic volatility.
- Psychoyios & Skiadopoulos (2006) test using volatility options as hedging instruments in different models.
- Li finds analytical formulas for the sensitivity of greeks to changes in the calibration prices and sets up ‘instrumental hedges’.
- Monoyios (2007) assess the impact of drift parameter uncertainty on hedging error distributions and proposes a filtering approach with learning in order to improve the performance of the hedging strategy.
Hedging Formulation

Recall the underlying price process

\[ S_t = S(t, (X_u)_{0 \leq u \leq t}, \theta). \]

Again, consider option \( X \) with finite time horizon \([0, T]\) written on \( S \) with payoff function \( h \), and time \( t \) value \( f_t(\theta) \). Assuming market completeness,

\[ f_t(\theta) = \mathbb{E}[B^{-1}(t, T)h(S(\theta))|\mathcal{F}_t]. \]

If a calibrated parameter \( \hat{\theta} \) is chosen then the value of \( X \) at time \( t \) is

\[ f_t(\hat{\theta}) = \mathbb{E}[B^{-1}(t, T)h(S(\hat{\theta}))|\mathcal{F}_t]. \]

Furthermore, taking a portfolio \((\Delta, \Psi)\) of stock \( S \) and cash \( B \) respectively to hedge the option, the corresponding Black-Scholes delta at time \( t \) is

\[ \Delta_t(\hat{\theta}) = \frac{\partial f_t(\hat{\theta})}{\partial S_t}. \]
Hedging Formulation

In the literature, there are two frequently cited delta hedges which, in the Bayesian Gaussian framework we assume, can be referred to as the following:

1. $\Delta_t(\theta^{MLE})$ — the delta hedge corresponding to the *maximum likelihood estimator* (MLE) $\theta^{MLE}$ (see e.g. Hull & Suo, Coleman et al, McIntyre, Dumas et al). It minimises the calibration error so is given by

$$\theta^{MLE}_t = \arg \min \{p(\theta|V)\}.$$ 

2. $\Delta(\theta^{MAP})$ — the delta hedge corresponding to the *maximum a posteriori estimator* (MAP) $\theta^{MAP}$ (see e.g. Jackson et al, Lagnado & Osher, Crepey). It maximises the Bayesian posterior and is given by

$$\theta^{MAP}_t = \arg \min \{p(V|\theta)\}.$$ 

Not usually referred to as the ‘MLE’ and ‘MAP’ estimates but are equivalent to this under Gaussian distribution assumptions.
Motivating Examples

Suppose underlying $S$ follows Black-Scholes model with volatility 0.15. But we only observe spread $[V^{bid}, V^{ask}] = [23.958, 24.103]$ of a 1 year European call with strike 80 and where $S_0 = 100$ and $r = 0.05$. 

![Graph showing density and hedging profit]
Motivating Examples

Suppose the volatility is now 0.13. Again we only observe spread $[V^\text{bid}, V^\text{ask}] = [23.958, 24.103]$ of a 1 year European call with strike 80 and where $S_0 = 100$ and $r = 0.05$.

Figure 2: Density corresponding to prior view of volatility (left) and performance comparison among different hedging strategies (right).
Bayesian Hedging Strategies

Might seem intuitive to use Bayesian model averaging and take parameter

$$\bar{\theta} = \int \theta p(\theta|V) d\theta$$

and hedge or price using this value, or directly take the delta hedge (Branger & Schlag) to be

$$\bar{\Delta} = \int \Delta(\theta) p(\theta|V) d\theta.$$ 

However, no guarantee or intuition for why the above parameter or hedge would give the optimal hedging strategy. Not even sure if $\bar{\theta}$ reproduces the observed data $V$ or that $\bar{\Delta}$ corresponds to a calibrated parameter $\theta$.

**Key Idea:** Let $L(\theta, \theta')$ correspond to some measure of the hedging error caused by hedging contract $X$ using parameter $\theta'$ when the correct hedge is found using parameter $\theta$. So take the estimator

$$\hat{\theta} = \theta_L(V).$$
Hedging Error Loss Functions
Examples

Consider hedging strategy given by a portfolio with time $t$ value $\Pi_t$. $\Pi$ used to hedge an option $X$ written on $S$ with payoff $h(S)$ at maturity time $T$ and has observable market value $\Pi_0 = V_0$ at inception time $0$. The hedging error at time $t$ is

$$E_t(\theta, \theta') = \Pi_t(S(\theta), \Delta(\theta')) - V_t(S(\theta))$$

where the underlying evolves according to model $\theta$ and we hedge in model $\theta'$. Then take the loss function as

$$L^g(\theta, \theta') = \mathbb{E}^\theta [g(E_T(\theta, \theta')) | \mathcal{F}_0]$$

for some function $g$ of the random variable $E_T(\theta, \theta')$. Recall that $E_T(\theta, \theta')$ is a random variable on the set of paths $\omega$ and the expectation $\mathbb{E}^\theta$ is taken over these paths using model (measure) $\theta$. 
Examples

Different choices of $g$ give common hedging performance indicators:

1. $g_\mu(z) = -z$ gives the average hedging loss.

2. $g_\sigma(z) = |z - \mathbb{E}[z]|$ gives the absolute average hedging error.

3. $g_\eta(z) = -z 1_{z < q_z(\beta)}$ gives the expected shortfall of the hedging loss, where the quantiles $q$ are given by $\beta = \mathbb{P}[z \leq q_z(\beta)]$ for some $\beta \in (0, 1)$.

Remark. $g_\mu$ gives a loss function $L^{g_\mu}$ which is not necessarily non-negative. This violates the definition of loss functions. Can add a constant

$$\tilde{L}^{g_\mu}(\theta, \theta') = L^{g_\mu}(\theta, \theta') + K$$

but then no longer guarantee $\tilde{L}^{g_\mu}(\theta, \theta) = 0$ so violate the definition again. This is only a technical and not conceptual problem. For the case of hedging, deviations from the true parameter can be both profitable and non-profitable.
Hedging Improvement

Let $\theta^0$ be the original (e.g. MAP) parameter used for hedging. Then the improvement in hedging performance is

$$I(\theta^0, \theta_L) := L(\theta^*, \theta^0) - L(\theta^*, \theta_L)$$

(3)

The expected value (with respect to the posterior density $p(\theta|V)$) of the improvement $I(\theta^0, \theta_L)$ is

$$\mathbb{E}[I(\theta^0, \theta_L)] = \int L(\theta, \theta^0) p(\theta|V) \, d\theta - \int L(\theta, \theta_L) p(\theta|V) \, d\theta$$

(4)

$$\geq 0$$

by the definition of the Bayes estimator.

$\mathbb{E}[I(\theta^0, \theta_L)] \geq 0$ might seem a trivial (or tautologous) result but the implications are fundamental to the motivation behind the Bayesian approach.
Hedging Improvement

Furthermore, because we can actually calculate the difference \( (4) \), if it is found to be large, then there is a good chance the actual hedging improvement \( (3) \) is significant.

Of course, how close the two quantities \( (4) \) and \( (3) \) are to one another will depend on the accuracy of the posterior density function \( p(\theta|V) \).

Shown earlier that, under particular assumptions on the parameter space \( \Theta \) and pricing functions \( f \), if a true model parameter \( \theta^* \) exists then

\[
p(\theta|V) \to \delta(\theta - \theta^*)
\]

in probability as the number of observations \( V \) increases (where \( \delta(z) \) is the Dirac delta probability density — zero everywhere except at \( z = 0 \)).
Hedging Improvement

We can calculate the variance of the improvement as

$$\text{Var}[I(\theta^0, \theta_L)] = \int [L(\theta, \theta^0) - L(\theta, \theta_L)]^2 p(\theta | V) \, d\theta - \{\mathbb{E}[I(\theta^0, \theta_L)]\}^2$$  \hspace{1cm} (5)

to give estimated bounds for the actual improvement (3).

For example

$$\left[ \mathbb{E}[I(\theta^0, \theta_L)] - 2\sqrt{\text{Var}[I(\theta^0, \theta_L)]}, \mathbb{E}[I(\theta^0, \theta_L)] + 2\sqrt{\text{Var}[I(\theta^0, \theta_L)]} \right]$$  \hspace{1cm} (6)

would correspond to a 95% confidence interval around the mean (4) if we approximate the distribution of (3) as Gaussian.

If the variance (5) is low then we can get fairly tight bounds on the actual difference.
A utility function $U$ is a map from $\mathbb{R} \rightarrow [-\infty, \infty)$ representing an agent’s preferences over different contingent claims:

$$Y \text{ is preferred to } X \iff \mathbb{E}[U(Y)] \geq \mathbb{E}[U(X)]$$

where $X$ and $Y$ are contingent claims and $U$ is increasing and concave.

In the context of the optimal hedging strategies, the utility theory approach would be to maximise the expected hedging profit, i.e.

Find $\theta'$ which maximises $\mathbb{E}^{Q,\theta}[U(\Pi(\theta', \omega) - h(\theta, \omega))]$ \hspace{1cm} (7)

where the expectation is taken over the different paths $\omega$ (using measure $\theta$) of the driving process and also over the different possible models $\theta$ (using measure $Q$).
Link to Utility Functions

On the other hand, the Bayesian approach is to

Find $\theta'$ which minimises

$$
\int L(\theta, \theta') p(\theta|V) \, d\theta
$$

$$
= \int \mathbb{E}^\theta [g(\Pi(\theta', \omega) - h(\theta, \omega))] p(\theta|V) \, d\theta
$$

$$
= \mathbb{E}^{Q,\theta} [g(\Pi(\theta', \omega) - h(\theta, \omega))]
$$

i.e. Find $\theta'$ which maximises

$$
\mathbb{E}^{Q,\theta} [-g(\Pi(\theta', \omega) - h(\theta, \omega))]
$$

(8)

So we see that the utility approach (7) and Bayesian approach (8) coincide precisely when

$$
U = -g.
$$

(9)

When this equality holds, we must have then that $g$ is decreasing and convex (since $U$ is increasing and concave). Berger (1985) and Föllmer & Schied (2002) arrive at a very similar identity.
Numerical Examples
Local Volatility Model

Using this distribution of surfaces we can evaluate the performance of difference hedging strategies. First a 3 month at-the-money call option:
Local Volatility Model

And we can table the results of hedging the call:

<table>
<thead>
<tr>
<th></th>
<th>mean hedge profit</th>
<th>absolute deviation</th>
<th>5% profit shortfall</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>true</em> delta</td>
<td>0.4</td>
<td>12.0</td>
<td>-26.2</td>
</tr>
<tr>
<td><em>MAP</em> delta</td>
<td>1.3</td>
<td>12.0</td>
<td>-24.9</td>
</tr>
<tr>
<td><em>MAP</em> delta-vega</td>
<td>1.1</td>
<td>4.3</td>
<td>-10.5</td>
</tr>
<tr>
<td>$\mu$-delta</td>
<td>2.3</td>
<td>12.2</td>
<td>-22.2</td>
</tr>
<tr>
<td>$\sigma$-delta</td>
<td>1.6</td>
<td>12.1</td>
<td>-23.5</td>
</tr>
<tr>
<td>$\eta$-delta</td>
<td>2.3</td>
<td>12.2</td>
<td>-22.1</td>
</tr>
<tr>
<td>$I(\theta^{MAP}, \theta_L)$</td>
<td>+1.1</td>
<td>-0.1</td>
<td>+2.8</td>
</tr>
<tr>
<td>quasi conf. int.</td>
<td>[2.2,2.3]</td>
<td>[-0.1,0.7]</td>
<td>[2.7,5.6]</td>
</tr>
</tbody>
</table>

Optimal Bayesian Hedging Strategies ⋆ 26 Nov 2009 ⋆ alok.gupta@maths.ox.ac.uk – p. 37
Local Volatility Model

And a 3 month at-the-money up-and-out barrier call option with barrier $1.1S_0$:
Table of improvements for the barrier call option:

<table>
<thead>
<tr>
<th></th>
<th>mean hedge profit</th>
<th>absolute deviation</th>
<th>5% profit shortfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>true delta</td>
<td>-5.5</td>
<td>46.2</td>
<td>-137.4</td>
</tr>
<tr>
<td>MAP delta</td>
<td>-4.4</td>
<td>52.7</td>
<td>-178.7</td>
</tr>
<tr>
<td>MAP delta-vega</td>
<td>-5.4</td>
<td>52.0</td>
<td>-178.6</td>
</tr>
<tr>
<td>μ- delta</td>
<td>8.6</td>
<td>52.7</td>
<td>-160.7</td>
</tr>
<tr>
<td>σ- delta</td>
<td>1.6</td>
<td>52.5</td>
<td>-170.5</td>
</tr>
<tr>
<td>η- delta</td>
<td>8.6</td>
<td>52.7</td>
<td>-160.4</td>
</tr>
<tr>
<td>$I(\theta_{MAP}, \theta_L)$</td>
<td>+13.0</td>
<td>+0.1</td>
<td>+18.2</td>
</tr>
<tr>
<td>quasi conf. int.</td>
<td>[5.3,13.1]</td>
<td>[-108.4,212.0]</td>
<td>[-30.1,105.6]</td>
</tr>
</tbody>
</table>
Heston Model

Priced 70 European call options (in a known Heston stochastic volatility model) with 10 strikes and 7 maturities and added Gaussian noise. We calibrate a 32-node local volatility surface. 600 samples plotted:
Heston Model

Using this distribution of surfaces we can evaluate the performance of different hedging strategies. First a 3 month at-the-money call option:
Heston Model

And we can table the results of hedging the call:

<table>
<thead>
<tr>
<th></th>
<th>mean hedge profit</th>
<th>absolute deviation</th>
<th>5% profit shortfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MAP$ delta</td>
<td>-0.1</td>
<td>11.5</td>
<td>-27.2</td>
</tr>
<tr>
<td>$MAP$ delta-vega</td>
<td>9.6</td>
<td>8.8</td>
<td>-6.3</td>
</tr>
<tr>
<td>$\mu$- delta</td>
<td>3.2</td>
<td>11.3</td>
<td>-23.4</td>
</tr>
<tr>
<td>$\sigma$- delta</td>
<td>2.0</td>
<td>11.3</td>
<td>-24.7</td>
</tr>
<tr>
<td>$\eta$- delta</td>
<td>3.3</td>
<td>11.2</td>
<td>-23.3</td>
</tr>
<tr>
<td>$I(\theta^{MAP}, \theta_L)$</td>
<td>+3.3</td>
<td>+0.2</td>
<td>+4.0</td>
</tr>
<tr>
<td>quasi conf. int.</td>
<td>[3.3,3.5]</td>
<td>[-0.3,1.1]</td>
<td>[1.5,9.5]</td>
</tr>
</tbody>
</table>
Heston Model

And a 3 month at-the-money up-and-out barrier call option with barrier $1.1S_0$:
Table of improvements for the barrier call option:

<table>
<thead>
<tr>
<th></th>
<th>mean hedge profit</th>
<th>absolute deviation</th>
<th>5% profit shortfall</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MAP$ delta</td>
<td>-9.1</td>
<td>45.6</td>
<td>-148.5</td>
</tr>
<tr>
<td>$MAP$ delta-vega</td>
<td>-0.3</td>
<td>47.4</td>
<td>-148.2</td>
</tr>
<tr>
<td>$\mu$- delta</td>
<td>1.6</td>
<td>45.5</td>
<td>-136.0</td>
</tr>
<tr>
<td>$\sigma$- delta</td>
<td>-8.4</td>
<td>45.7</td>
<td>-148.2</td>
</tr>
<tr>
<td>$\eta$- delta</td>
<td>1.6</td>
<td>45.5</td>
<td>-136.0</td>
</tr>
<tr>
<td>$I(\theta_{MAP}, \theta_L)$</td>
<td>+10.8</td>
<td>-0.1</td>
<td>+12.5</td>
</tr>
<tr>
<td>quasi conf. int.</td>
<td>[6.4,15.1]</td>
<td>[-68.5,115]</td>
<td>[-32.6,96.3]</td>
</tr>
</tbody>
</table>
Robustness

For the local volatility example, we changed the form of the prior (by adjusting the value of the constant $\kappa$ in the formulation of the prior). Effect on hedging performance of barrier option:

![Graphs showing hedging performance for different values of $\kappa$.]

(a) $\kappa = 10^{-2.0}$  
(b) $\kappa = 10^{0.0}$
Robustness

For the local volatility example, we tested adding more noise to the market data (used $\varepsilon = 10^{-2.5}$ instead of $\varepsilon = 10^{-3.0}$). Effect on hedging performance of barrier option:

(c) Noise A

(d) Noise B
Conclusion & Extensions
Conclusion

• Introduced the Bayesian framework for calibrating the parameters of financial models to market prices.

• Described the implicit model uncertainty and designed loss functions which optimised hedging performance indicators.

• Remarked on how to estimate bounds for the improvement, and use this to decide whether or not to implement the Bayesian strategy.

• Used local volatility model and Heston model as case studies, tested hedging contracts in both models using local volatility deltas.

• Saw improvements in hedging performance when using the Bayesian hedges instead of typical MAP strategies, especially for path dependent options.
Extensions

• The methodology is very general and can be applied to any parametric or non-parametric hedging strategy model — not just delta hedging

• Can use the loss functions $L(\theta, \theta')$ to quantify measures for the model uncertainty of any contingent claim. Such measures would be important for a risk manager or agent trying to decide between different products.

• Can expand the choice of loss functions e.g. by exploiting the relationship with utility functions and/or by taking combinations of loss functions:

$$L^{g\mu} + \alpha L^{g\sigma}$$

for some risk-return tradeoff parameter $\alpha$

• Try to extend the Bayesian philosophy to portfolio optimization problems. Higher dimensionality will make it difficult but there should be considerable scope for this.
Thank you for your attention

Questions?