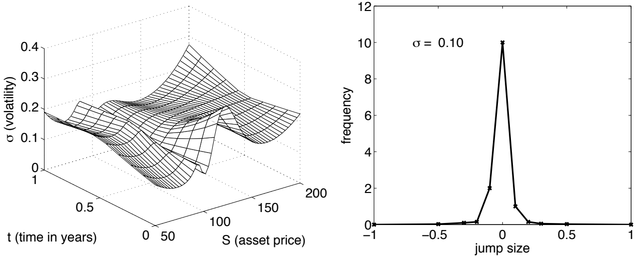
Calibrating Financial Models Using Consistent Bayesian Estimators

Christoph Reisinger
Joint work with Alok Gupta

June 25, 2010

Example – model uncertainty

- A local volatility model, jump diffusion model, and (Heston) stochastic volatility model calibrated to 60 observed European calls for different strike/maturity pairs within 3 basis points.

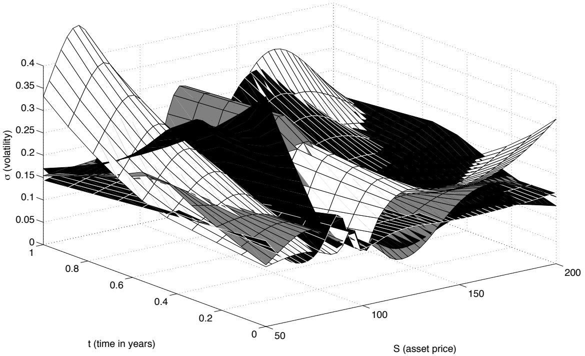


Parameter	rate of reversion	long run variance	volatility of volatility	correlation	initial variance
Value	0.0745	0.1415	0.1038	-0.2127	0.0167

- The value of an up-and-out barrier call with strike 90% and barrier 110% of the spot varies by 177 basis points.

Example – parameter uncertainty

- Three different local volatility models calibrated to 60 observed European calls for different strike/maturity pairs within 3 basis points. See also Hamida and Cont (2005).



- The value of an up-and-out barrier call with strike 90% and barrier 110% of the spot varies by 26 basis points.

Derivative pricing models

Model choice:

- Assume a model θ;
- model value of a derivative $V(\theta)$.

Calibration:

- Find θ^* s.t. $V(\theta^*) = V^*$ the market price of liquid contracts.

Pricing and hedging:

- Solve a pricing equation for a new (exotic) derivative,

$$A(\theta^*)\widehat{V}(\theta^*) = 0;$$

- hedge with sensitivities derived from $\widehat{V}(\theta^*)$.

An 'ill-posed' problem

Remidies for this model ambiguity.

- **Regularisation**:

$$\text{market fit}(\theta) \; + \; \text{regularity measure}(\theta) \;\longrightarrow\; \min_{\theta}$$

- **Worst-case replication approach**:

$$\sup_{\theta} A(\theta) V(\theta) = 0, \quad \text{s.t.} \quad V(\theta) = V^* \;\text{ for calibration products}$$

- **Bayesian framework**:
  - prior information encapsulated in $p(\theta)$
  - likelihood of market prices $p(V^*|\theta)$
  - posterior distribution $p(\theta|V^*)$

Motivation

- Model ambiguity and over-parametrisation lead to uncertainty in the pricing model and the need to quantify and risk-manage the resulting risk.
- A Bayesian perspective seems well-suited to these objectives.
- It combines prior and historical information ('regularisation') with currently observed prices ('calibration').
- Consistency guarantees that parameter estimates are not led astray by prior assumptions.

Outline

- Calibration problems in financial engineering and their ill-posedness
- Bayesian approach to the calibration problem
- **Consistency of Bayesian estimators**
- **Practical construction of posteriors and examples**
- Related work: measuring model uncertainty, robust hedging
- Conclusions

Model for underlying

- Assume price process $S = (S_t)_{t \geq 0}$ s.t. (by abuse of notation)

$$S_t = S(t, (Z_u)_{0 \leq u \leq t}, \theta)$$

  a function of
  - time t,
  - some 'standard' process $Z = (Z_t)_{t \geq 0}$, and
  - parameter(s) $\theta \in \Theta$.
- Assume henceforth that θ is a finite dimensional vector: $\Theta \subseteq \mathbb{R}^M$.
- We are specifically interested in applications where this parameter is the discretisation of a functional parameter, for example representing a local volatility function.

Model price

Now consider

- an option over a finite time horizon $[0, T]$ written on S and with payoff function h, and
- the time t value of this option written as

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

with respect to some risk-neutral measure $\mathbb{Q}$, where
- $B(t, T)$ is the discount factor for the time interval $[t, T]$.

Observations

- Denote θ^* the 'true' parameter.
- Suppose at time $t \in [0, T]$ we observe a set of such option prices $\{f_t^{(i)}(\theta) : i \in I_t\}$, with additive noise $\{e_t^{(i)} : i \in I_t\}$, i.e. we observe
$$V_t^{(i)} = f_t^{(i)}(\theta^*) + e_t^{(i)}.$$
- The calibration problem is to find the value of θ that *best* reproduces the observed prices
$$V = \{V_t^{(i)} : i \in I_t,\ t \in \Upsilon_n([0, T])\}.$$
- 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.

Bayesian framework

- Assume we have some prior information for θ, e.g. it
  - belongs to a particular subspace of the parameter space, or
  - is positive, or
  - represents a smooth function,

  summarised by a *prior* density $p(\theta)$ for θ.
- $p(V|\theta)$ is the *likelihood* of observing the data V given θ.
- Bayes rule gives the *posterior* density of θ,

$$p(\theta \mid V) = \frac{p(V|\theta)\, p(\theta)}{p(V)},$$

where $p(V)$ is given by

$$p(V) = \int p(V|\theta)\, p(\theta)\, \mathrm{d}\theta.$$

Bayesian literature

Consistency of Bayesian estimators:

- Doob (1953), Schwartz (1965)
- Le Cam (1953): relation to maximum likelihood estimators
- Fitzpatrick (1991): relation to regularisation
- Wasserman (1998), Barron, Schervish, and Wasserman (1999), Shen and Wasserman (2001), Goshal (1998), Goshal, Gosh, and van der Vaart (2000): properties, convergence rates

All assume i.i.d. data.

- Here: observations of different functions of the parameter.

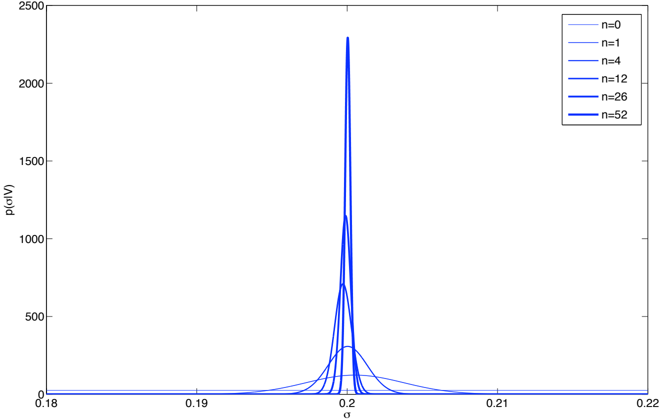
Example

- Black-Scholes model with $\sigma^* = 0.2$;
- observe prices each week for the first 52 weeks of a two year at-the-money call option;
- $S_0 = 100$ and the interest rate $r = 0.05$, s.t. $f_0(\sigma^*) = 16.13$;
- uniform prior $p(\sigma)$ on [0.18,0.22];
- mean-zero Gaussian noise e_t of standard deviation 5% of the true option price, i.e.

$$e_t \sim N(0, \tfrac{1}{20} f_t(\sigma^*)).$$

- See also Jacquier and Jarrow (2000).

Example



Posterior densities after *n* observations. Notice that most of the probability measure collects around the true value of $\sigma^* = 0.2$.

Convergence in probability

Assumptions on the **prior**:

- The prior p has compact support Θ,
- p is bounded, continuous at θ^* (true parameter) with $p(\theta^*) > 0$.

Assumptions on the **observations**:

- $\mathcal{F}_{t_n} \perp\!\!\!\perp \mathcal{G}_{t_m}$ for all (n, m), i.e. the driving process of the underlying is independent from the market noise,
- Gaussian noise with variance ϵ_t^2, and
- $\forall t,\ \theta \neq \theta' \in \Theta \qquad \frac{1}{\varepsilon_t} \frac{|f_t(\theta) - f_t(\theta')|}{|\theta - \theta'|} \geq k > 0.$

Then:

- $\theta_{\mathbf{n}}(\mathbf{V}) := \theta | \mathcal{F}_{t_n} \vee \mathcal{G}_{t_n} \xrightarrow{\mathbf{P}} \theta^*.$

Estimators

- A function $L : \mathbb{R}^{2M} \to \mathbb{R}$ is a *loss function* $L(\theta, \theta')$ iff

$$\begin{cases} L(\theta, \theta') = 0 & \text{if } \theta' = \theta \in \mathbb{R}^M \\ L(\theta, \theta') > 0 & \text{if } \theta' \neq \theta. \end{cases}$$

- The corresponding *Bayes estimator* $\theta_L(V)$ is

$$\theta_L(V) = \arg \min_{\theta' \in \Theta} \left\{ \int_\Theta L(\theta, \theta')\, p(\theta|V)\, \mathrm{d}\theta \right\}.$$

- Examples:
  - $L_1(\theta, \theta') = \|\theta - \theta'\|^2$ gives Bayes estimator $\theta_{L_1}(Y) = \mathbb{E}[\theta|V]$ (the *mean* value of θ with respect to the Bayesian posterior density $p(\theta|V)$)
  - $\theta_{MAP}(V) = \arg\max\{p(\theta|V)\}$, the *maximum a posteriori* (MAP) estimator

Consistency result

- $p(\theta_n(V))$, the posterior density of θ after n observations, is

$$\begin{aligned} p(\theta_n(V)) &= \frac{p_n(V|\theta)\, p(\theta)}{p_n(V)} = \frac{p(V_{t_1}|\theta)\cdot\ldots\cdot p(V_{t_n}|\theta)\, p(\theta)}{p_n(V)} \\ &= \prod_{t\in\Upsilon_n} \frac{1}{\sqrt{2\pi}\varepsilon_t} \exp\left\{-\frac{1}{2\varepsilon_t^2}(V_t - f_t(\theta))^2\right\} \frac{p(\theta)}{p_n(V)}. \end{aligned}$$

- Define the sequence of Bayes estimators $\hat{\theta}$ by,

$$\begin{aligned} g(\theta_n(V), \theta') &= \mathbb{E}[L(\theta_n(V), \theta')] = \int_\Theta L(\theta, \theta')\, p_n(\theta|V)\, \mathrm{d}\theta \\ \hat{\theta}_n(V) &= \arg\min_{\theta'\in\Theta} \{g(\theta_n(V), \theta')\}. \end{aligned}$$

Then, under the assumptions from earlier, and

- for L bounded and continuous on Θ, $\hat{\theta}_n(V)$ **is consistent**.

Multiple observations

- Suppose multiple observations $f_t^{(i)}$ per time, $i \in I_t$, with similar assumptions as above for all i.
- Deduce the Bayes estimator $\hat{\theta}_n(V)$ is consistent.
- Speeds up convergence.
- Taken to the extreme, can construct a consistent estimator by gathering a large number of observations of different functions (options with different strikes, maturities) of θ at time 0.
- We give an example of this later.

Non-scalar parameter

- Take the case when θ is not scalar but a **finite-dimensional parameter**, $\theta \in \mathbb{R}^M$.
- Replace the monotonicity assumption on the observations by:

$$\exists\, K > k > 0 \;\; \forall\, \theta \in \Theta \quad K^2 \geq \frac{1}{n} \sum_{t \in \Upsilon_n} \frac{1}{\varepsilon_t^2} \frac{|f_t(\theta) - f_t(\theta^*)|^2}{\|\theta - \theta^*\|^2} \geq k^2$$

- For all L bounded and continuous on θ, the non-scalar Bayes estimator $\hat{\theta}_n(V)$ **is consistent**.

Discussion of assumptions

- Let $f_t(\theta)$ be smooth in t and θ, and $\epsilon_t = \epsilon$ constant.
- Then the above assumption can only be violated if either
  1. $\exists \theta \neq \theta^* \, \forall t \quad f_t(\theta) = f_t(\theta^*)$, or
  2. $\exists \theta \neq \theta^* \, \forall t \quad (\theta - \theta^*) \cdot \nabla_\theta f_t(\theta^*) = 0$.

1. Under 1., it is clearly impossible to identify which parameter gave rise to the observations.
2. Under 2., perturbations of the parameter in directions orthogonal to the gradient are overshadowed by the noise.

This confirms an intuitive rule for a good choice of observation variables (calibration products) as those which are most sensitive to the parameters.

Local volatility

The (discretised) local volatility model is a good example:

- Complete market model.
- Used by traders in some markets.
- Large (infinite) number of parameters.
- Ill-conditioned (ill-posed) calibration.
- Dynamically inconsistent.

Inverse problems literature

Identification of local volatility:

- [Dupire (1994)]
- Lagnado and Osher (1997)
- Jackson, Süli, and Howison (1999)
- Chiarella, Craddock, and El-Hassan (2000)
- Coleman, Li, and Verma (2001)
- Berestycki, Busca, and Florent (2002)
- Egger and Engl (2005)
- Achdou and Pironneau (2004)
- Zubelli, Scherzer, and De Cezaro (2010)

Bayesian prior

- We incorporate:
  - positivity
  - the a-t-m vol
  - smoothness
- Use the natural Gaussian prior

$$p(\theta) \propto \exp\left\{-\frac{1}{2}\tilde{\lambda}\|\theta - \theta_0\|^2\right\}$$

- $1/\tilde{\lambda}$ can be thought of as the prior variance of θ
- Example:

$$p_{lv}(\sigma) \propto \exp\left\{-\frac{1}{2}\lambda_p\|\log(\sigma) - \log(\sigma_{atm})\|_\kappa^2\right\}$$

where

$$\|u\|_\kappa^2 = (1-\kappa)\|u\|_2^2 + \kappa\||\nabla u|\|_2^2$$

Likelihood

- Recall $V_t^{(i)}$ the market observed price at t of a European call with strike K_i, maturity T_i ;
- $f_t^{(i)}(\theta)$ the theoretical price when the model parameter is θ;
- define the basis point square-error function as

$$\begin{aligned} G_t(\theta) &= \frac{10^8}{S_t^2} \sum_{i \in I} w_i \, |f_t^{(i)}(\theta) - V_t^{(i)}|^2 \\ V_t^{(i)} &= \frac{1}{2}(V_t^{(i)bid} + V_t^{(i)ask}); \end{aligned}$$

- define $\delta_i = \frac{10^4}{S_0} |V_t^{(i)ask} - V_t^{(i)bid}|$ a basis point bid-ask spread.
- As in Hamida and Cont (2005) demand $G(\theta) \leq \delta^2$, then

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

Construction of posterior

- Then the posterior is

$$p(\theta|V) \propto 1_{G(\theta)\leq\delta^2} \exp\left\{-\frac{1}{2\delta^2}\left[\lambda\|\theta-\theta_0\|^2 + G(\theta)\right]\right\}.$$

Note: maximising the posterior is equivalent to specific Tikhonov regularisations (e.g. Fitzpatrick (1991)).

Two datasets

1. **Simulated data-set:**
   - We price European calls with 11 strikes and 6 maturities on the surface given in Jackson, Süli and Howison (1999).
   - Similar to there, we take $S_0 = 5000$, $r = 0.05$, $d = 0.03$.
   - To each of the prices we add Gaussian noise with mean zero and standard deviation 0.1% as in Hamida and Cont (2005) and treat these as the observed prices.
   - We take the calibration error acceptance level as $\delta = 3$ basis points following the results of Jackson et al (1999).
2. **Market data:**
   - We take real S&P 500 implied volatility data used in Coleman, Li and Verma (2001) to price corresponding European calls.
   - 70 European call prices are calculated from implied volatilities with 10 strikes and 7 maturities.
   - Spot price of the underlying at time 0 is $S_0 = \$590$, interest rate is $r = 0.060$ and dividend rate is $d = 0.026$.

Parameter discretisation

1. For the first example, we take grid nodes

$$\begin{aligned} s &= 2500, 4500, 4750, 5000, 5250, 5500, 7000, 10000, \\ t &= 0.0, 0.5, 1.0, \end{aligned}$$

   so a total of $M = 27$ parameters (cf 66 calibration prices).

2. For the second example,

$$\begin{aligned} s &= 300, 500, 560, 590, 620, 670, 800, 1200, \\ t &= 0.0, 0.5, 1.0, 2.0, \end{aligned}$$

   so a total of $M = 32$ parameters (cf 70 calibration prices).

Interpolate with cubic splines in S, linear in t.

Computational issues

Sample from the posterior using *Markov Chain Monte Carlo*, see e.g. Beskos and Stuart (2009):

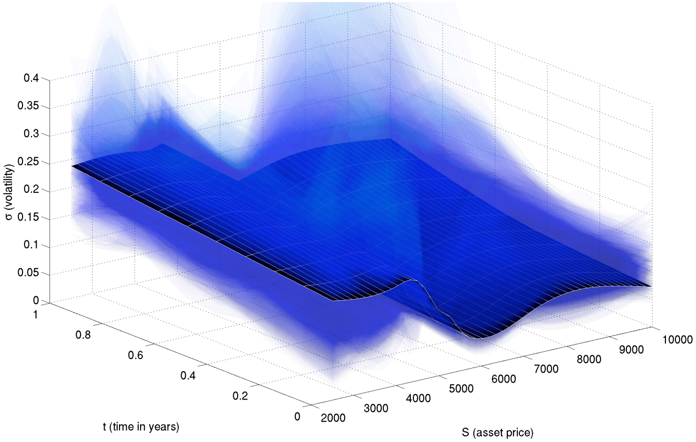
1. Select a starting point θ_0 for which $g(\theta_0|V) > 0$.
2. For $r = 1, \ldots, n$, sample a proposal $\theta^{\#}$ from a *symmetric jumping distribution* $J(\theta^{\#}|\theta_{r-1})$ and set

$$\theta_r = \begin{cases} \theta^{\#} & \text{with probability } \min\left\{\frac{g(\theta^{\#}|V)}{g(\theta_{r-1}|V)}, 1\right\} \\ \theta_{r-1} & \text{otherwise.} \end{cases}$$

Then the sequence of iterations $\theta_1, \ldots, \theta_n$ converges to the target distribution $g(\theta|V)$.

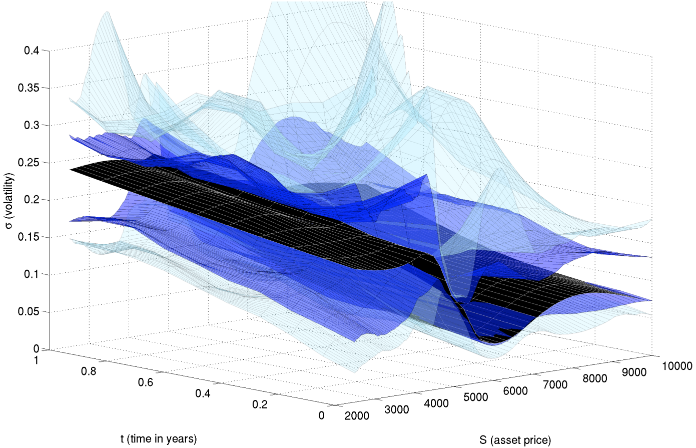
- Speed up by *thinning*, and eliminate *burn-in*.
- Monitor *potential scale reduction factor* for convergence.

Sampling the posterior



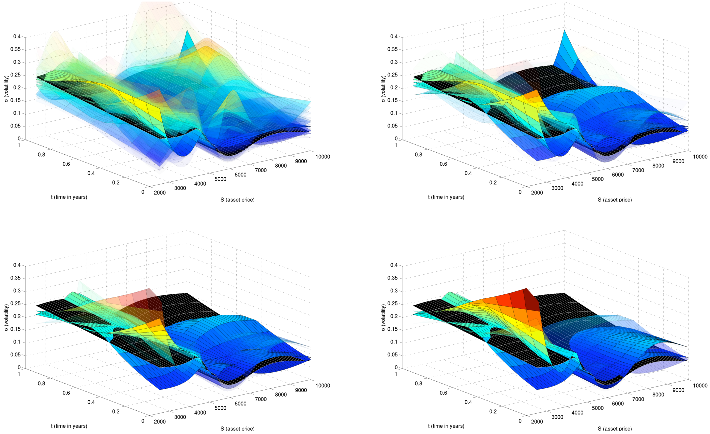
For the simulated dataset: 479 surfaces sampled from the posterior distribution, the true surface in opaque black.

Pointwise confidence intervals



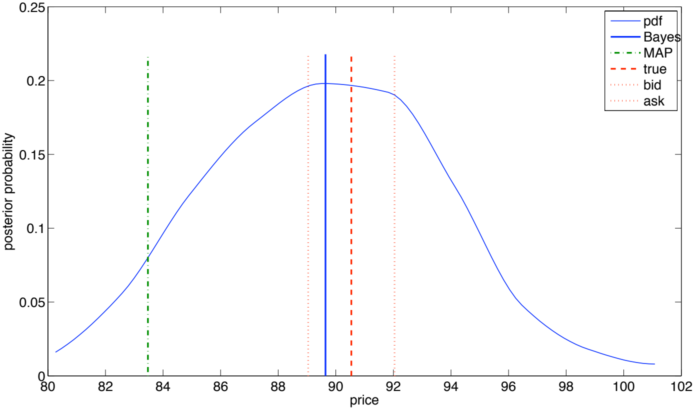
For the simulated dataset: 95% and 68% pointwise confidence intervals for volatility of paths, the true surface in opaque black.

Re-calibration



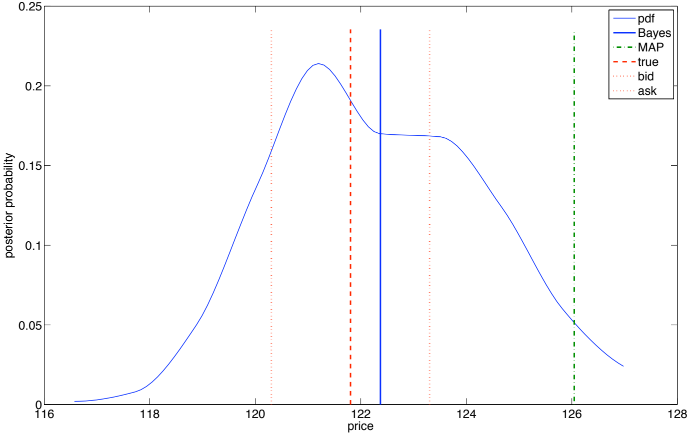
Now a path is simulated on the true local volatility surface and the Bayesian posterior is updated using the newly observed prices each week for 12 weeks (plotted: weeks 3,6,9,12). The transparency of each surface reflects the Bayesian weight of the surface.

Pricing a barrier option



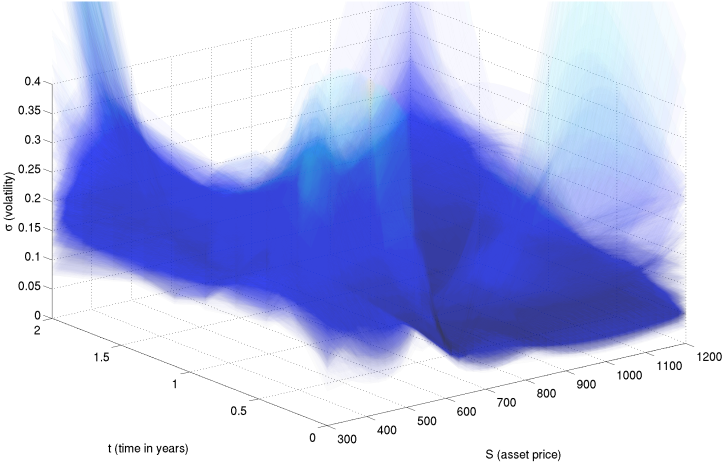
For simulated dataset: prices for up-and-out barrier calls with strike 5000 ($S_0 = 5000$), barrier 5500, maturity 3 months. Included are the 'true' price with its bid-ask spread, the MAP price, and the Bayes price with its associated posterior pdf.

Pricing an American option



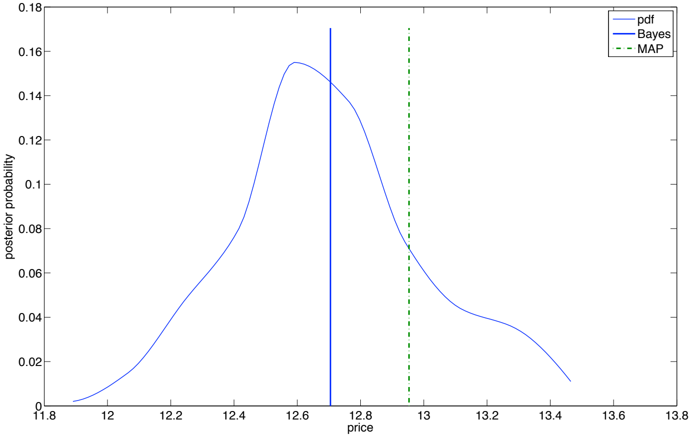
For the simulated dataset: prices for American puts with strike 5000 ($S_0 = 5000$) and maturity 1 year. Included are the 'true' price with its bid-ask spread, the MAP price, and the Bayes price with its associated posterior pdf.

Market data



For S&P 500 dataset: using Metropolis sampling, 600 surfaces from the posterior distribution.

Pricing an American option



For S&P 500 dataset: prices for American put option with strike \$590 ($S_0 = \$590$) and maturity 1 year. Included are the MAP price and the Bayes price with its associated posterior pdf of prices.

Model uncertainty measures

'Bayesian' *model uncertainty measures*:

- Branger and Schlag (2004)
- Gupta and R. (2010)

This is in contrast to 'worst-case' measures:

- 'Price-based': Cont (2006)
- 'Risk-differencing': Kerkhof, Melenberg, Schumacher (2002)
- 'Hedging-based': uncertain parameter models, e.g. Avellaneda, Lévy, and Paras (1995)

Discussion and extensions

- Construction of Bayesian posteriors using prior information and market data
- Consistency – would also like ‘negative’ result
- Gives model uncertainty measures
- Potentially useful for robust hedging