Bayesian Quadrature: State-Price Density Estimation

Huei-Wen Teng, The Pennsylvania State University, Rao Prize Conference 2009

Abstract

Asset pricing theories give the option price as the discounted expected payoff function under the risk-neutral probability. Such a pricing density is called the state-price density. Extracting the state-price density using options prices and/or historic data of underlying assets have been widely developed over the past decade. A nonparametric approach is preferable in that it is free of the joint-hypothesis problem, which says that any test of the economic theory is a joint test of the theory and the assumed option pricing model. We propose a quadrature method to make inference for the state-price density from a Bayesian perspective. For numerical purposes, a Gibbs sampler with slice sampling is proposed. Simulation studies and studies based on real data using S&P 500 index options show that our approach produces good model fit.

Overview

$\wp_{ij}(x)$ denotes the payoff function:

$$\wp_{ij}(x) = \left((-1)^i(x - c_{ij})\right)^+ .$$

Theoretic options price is

$$e^{-rT} E[\wp_{ij}(X)]$$

where X has a state-price density $f(x)$ under the expectation operator E.

Our focus

Let y_{ijk} be the observed option prices, which follow

$$y_{ijk} = e^{-rT} E[\wp_{ij}(X)]\, e^{\varepsilon_{ijk}},$$

for $k = 1, \dots, N_{ij}$ with errors ε_{ijk} iid $N(0, \sigma_\varepsilon^2)$.

Quadrature: (w,x)

Consider

$$\hat{f}(\cdot|w, x) = w_1 \delta_{x_1}(\cdot) + \cdots w_M \delta_{x_M}(\cdot).$$

The theoretical price under $\hat{f}$ is

$$G_{ij}(w, x) = e^{-rT} \sum_{m=1}^{M} w_m \wp_{ij}(x_m).$$

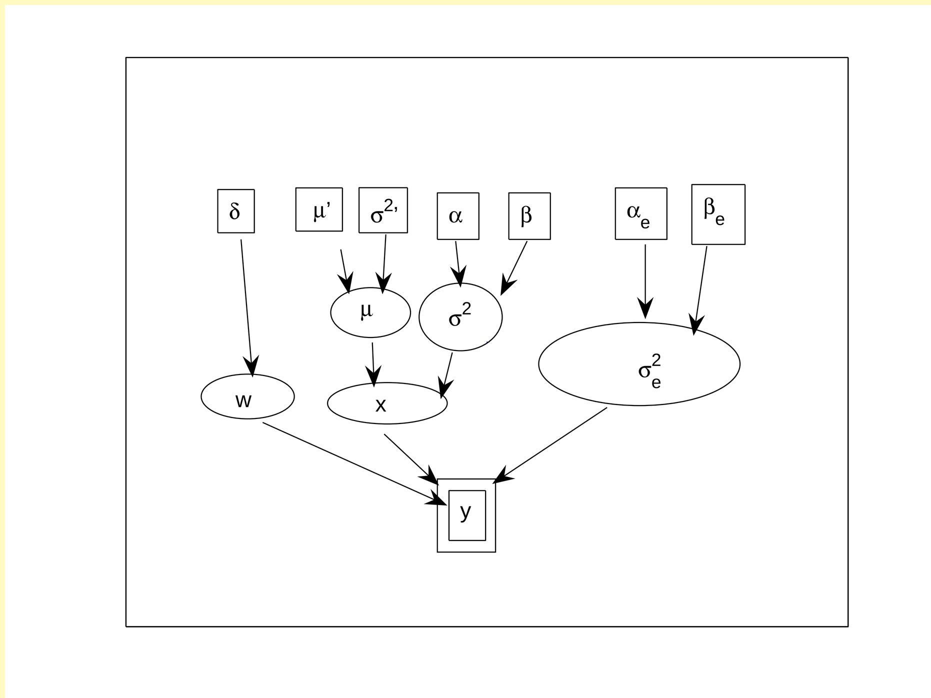
Model specifications

The likelihood, $L(y|w, x, \sigma_\varepsilon^2)$, is

$$\prod_{i=1}^{2} \prod_{j=1}^{N_i} \prod_{k=1}^{N_{ij}} \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} e^{-\frac{(\log y_{ijk} - \log G_{ij}(w,x))^2}{2\sigma_\varepsilon^2}} .$$

Prior distributions are assigned as follows:

1. $w \sim D(\delta, \dots, \delta)$.
2. $x_m \sim LN(\mu, \sigma^2)$.
   (a) $\mu \sim N(\bar{\mu}, \bar{\sigma}^2)$.
   (b) $\sigma^2 \sim IG(\alpha, \beta)$.
3. $\sigma_\varepsilon^2 \sim IG(\alpha_\varepsilon, \beta_\varepsilon)$.



MCMC algorithm

1. $\mu | \cdots \sim$
$$N\left(\frac{\sum_{m=1}^{M} \log(x_m)/\sigma^2 + \bar{\mu}/\bar{\sigma}^2}{M/\sigma^2 + 1/\bar{\sigma}^2}, \frac{1}{M/\sigma^2 + 1/\bar{\sigma}^2}\right).$$
2. $\sigma^2 | \cdots \sim$
$$IG\left(M/2 + \alpha, \sum_{m=1}^{M} (\log(x_m) - \mu)^2/2 + \beta\right).$$
3. $w_m | \cdots \sim U(T_m)$, where T_m is an open interval, and $w_M = 1 - w_1 - \cdots - w_{M-1}$, for $m = 1, \dots, M-1$.
4. $x_m | \cdots \sim U(S_m)$, where S_m is an open interval, for $m = 1, \dots, M$.
5. $\sigma_\varepsilon^2 | \cdots \sim IG(\alpha_\varepsilon^*, \beta_\varepsilon^*)$, where
$$\alpha_\varepsilon^* = \alpha_\varepsilon + \sum_{i=1}^{2} \sum_{j=1}^{N_i} \sum_{k=1}^{N_{ij}} v_{ijk}/2,$$
$$\beta_\varepsilon^* = \beta_\varepsilon + \sum_{i=1}^{2} \sum_{j=1}^{N_i} \sum_{k=1}^{N_{ij}} v_{ijk}(\log y_{ijk} - \log G_{ij})^2/2.$$

Study design

M is fixed at 5, 10, 20, 40, and 80 to evaluate model fit of our quadrature method. With random start, the first 600 burn-in samples are discarded, and the following 200 samples are collected.

Example 1

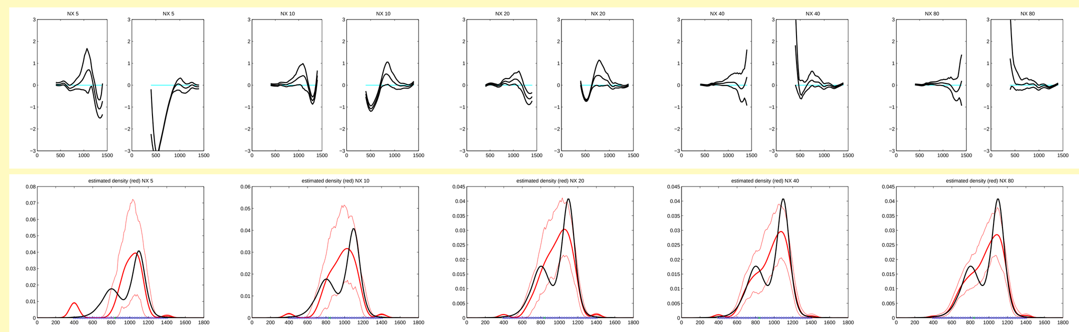
X follows

$$0.4t_5(800, 100^2) + 0.6t_5(1100, 60^2).$$

We summarize means of posterior distributions of μ, σ^2, and σ_ε^2, log-likelihood (LL), root mean square errors (RMSE), and elapsed time in seconds (eT).

M	μ	σ^2	σ_ε^2	LL	RMSE	eT
5	6.75	0.92	1.27	-255.58	1.13	2
10	6.78	0.76	0.43	-242.48	0.63	4
20	6.83	0.57	0.32	-230.56	0.53	10
40	6.81	0.39	0.30	-221.01	0.50	28
80	6.78	0.26	0.27	-212.78	0.46	97

Plots of residuals and estimated densities.



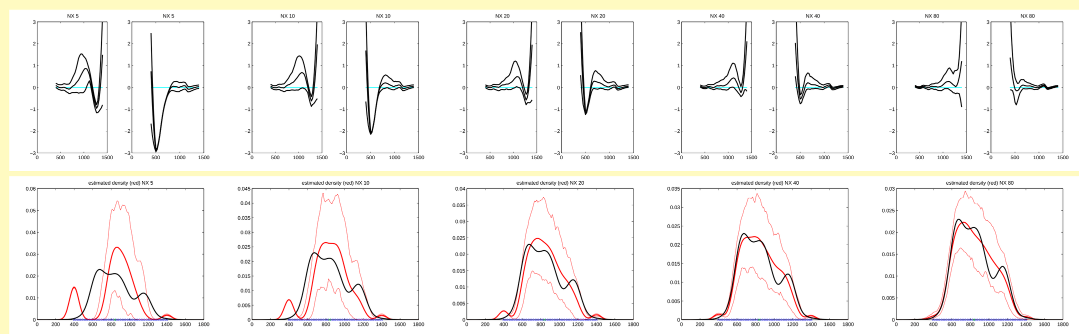
Example 2

X follows

$$0.3t_5(650, 60^2) + 0.5t_5(850, 100^2) + 0.2t_5(1150, 70^2).$$

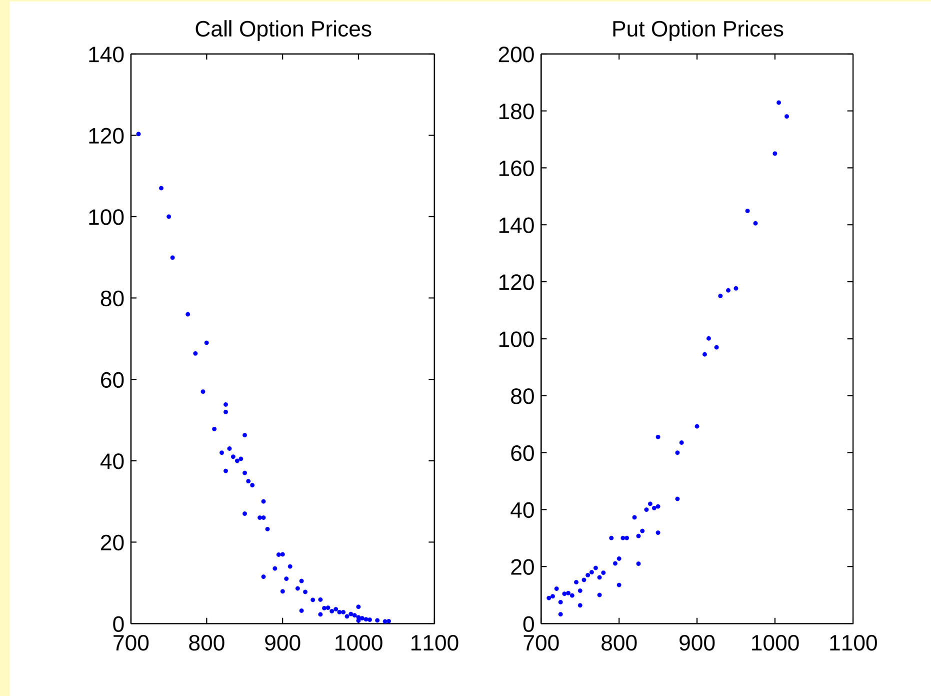
M	μ	σ^2	σ_ε^2	LL	RMSE	eT
5	6.73	0.88	1.07	-252.97	1.03	2
10	6.72	0.78	0.68	-248.59	0.81	4
20	6.73	0.58	0.51	-242.12	0.69	10
40	6.70	0.40	0.40	-234.87	0.59	28
80	6.70	0.26	0.28	-215.24	0.47	90

Plots of residuals and estimated densities.

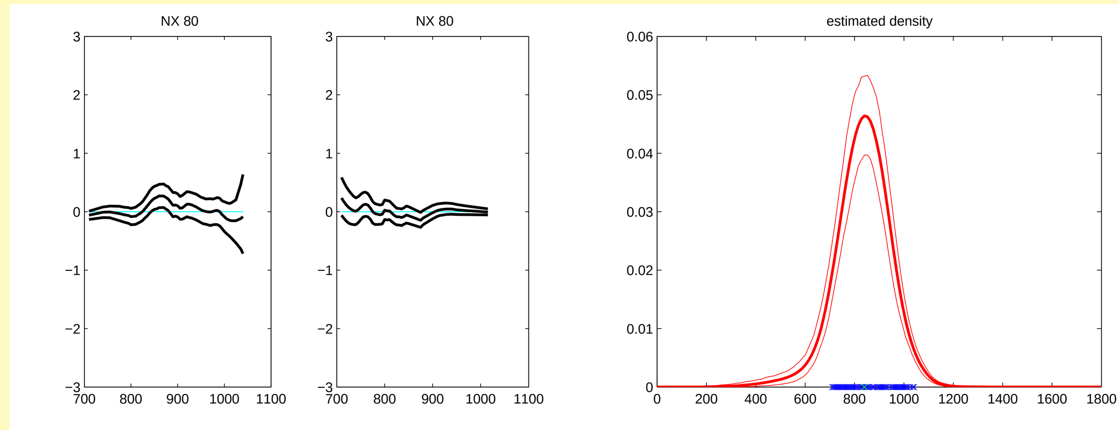


Real dataset

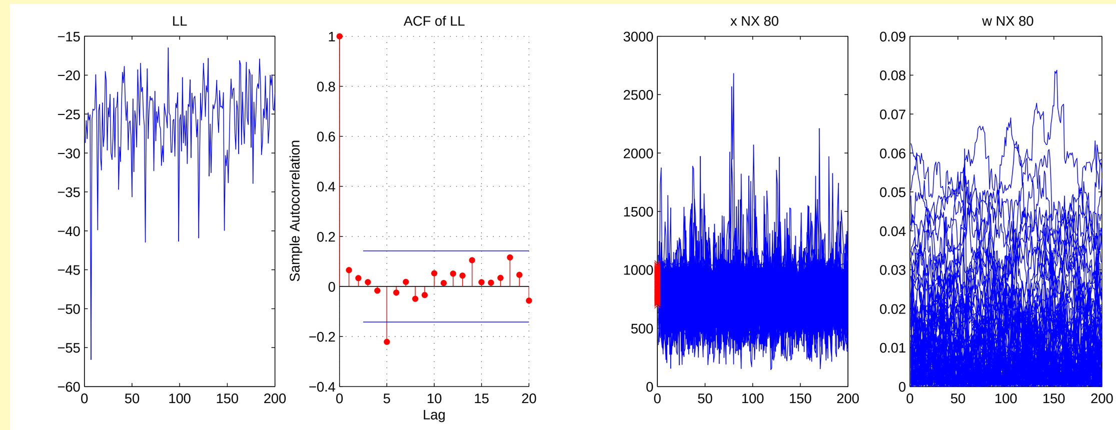
The scatter plots of S&P 500 index options (closing at 840.95) on 2008/12/16.



Plots of residuals and estimated densities.



LL and its ACF plot, and traces of x's and w's.



Conclusions and future work

Our method provides good model fit with a moderate M. When M increases, RMSE decreases, and the estimated densities seem to be closer to the true density. However, the simulation time increases linearly in M. Hence, selection of M needs more study.

The proposed method also forms a framework to solve a general integral equation with measurement errors:

$$h(k) = \int p(k, x) f(x) dx + \varepsilon,$$

where $h(k)$ is the observed, $p(k, x)$ is a known function, ε is the error, and $f(x)$ needs to be estimated and may be of high dimension.