Available online at www.sciencedirect.com

ScienceDirect

Computational Statistics & Data Analysis 51 (2006) 2339 – 2349

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Fast algorithm for nonparametric arbitrage-free SPD estimation ☆

Zdeněk Hlávka*

Department of Statistics, Charles University in Prague, Sokolovská 83, 18675 Praha, Czech Republic

Available online 24 August 2006

Abstract

State price density (SPD) contains important information concerning market expectations. An estimator of the SPD based on observed European option prices, taking into account the time of the trade, has been previously considered. Financial markets produce huge amounts of data and, due to time constraints, it is not always possible to calculate the estimator using all available data. Using a model for the covariance structure of the observed option prices, the algorithm identifies observations with little importance to the estimator. Dropping these observations increases the speed of computation and allows frequenter updating of the estimator. The algorithms efficiently use indices that combine information contained in the data. Fast algorithms are proposed and their properties are investigated using both simulated and real data sets.

Keywords: Option pricing; Nonlinear least squares; Constrained estimation

1. Motivation

The price $C_t(K, T)$ of a European Call option with payoff $(S_T - K)_+ = \max(S_T - K, 0)$, where S_T denotes the price of the underlying stock at time T, t the current time, K the strike price, and r the risk free interest rate, can be written as the discounted expected value of the payoff,

$$C_t(K, T) = \exp\{-r(T-t)\} \int_0^{+\infty} (S_T - K)_+ f(S_T)\, \mathrm{d}S_T, \tag{1}$$

with respect to the so-called state price density (SPD) $f(.)$ containing information on the expectations of the market. Knowing SPD, complicated (exotic) options may be simply priced by evaluating their expected payoff. The existence of a unique risk neutral SPD $f(.)$ implies the absence of arbitrage (Harrison and Pliska, 1981).

The estimation procedure described in the following sections takes advantage of the covariance structure of the observations. The side-effect is the dependency of the speed of the calculation on the number of observations some of which contribute only very little information to the estimator. The purpose of this paper is to investigate possibilities of increasing the speed of the calculation while maintaining as much precision as possible.

In the rest of this section, we describe the estimation of the SPDs suggested in Härdle and Hlávka (2006). In Section 2, we suggest two approaches to reducing the time of calculation for large data sets. In Section 3, these approaches are applied to observations with the same covariance structure as the observed option prices used in the

☆ The research was supported by MSM 0021620839 and 1K04018.
* Tel.: +420 221 913 284; fax: +420 222 323 316.
E-mail address: hlavka@karlin.mff.cuni.cz.

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.08.006

SPD estimation. Based on our observations on the behavior of the suggested methods, we investigate properties of various estimation strategies suitable for real data in Section 4.

1.1. SPD estimation

The observed prices of European Call options allow the deduction of the SPD in the following form (Breeden and Litzenberger, 1978):

$$f(K) = \exp\{r(T-t)\}\frac{\partial^2 C_t(K,T)}{\partial K^2}. \tag{2}$$

Eq. (2) is often used to estimate the state price density by means of nonparametric regression (Aït-Sahalia and Lo, 2000; Aït-Sahalia and Duarte, 2003). Considering the observed intraday option prices, Härdle and Hlávka (2006) propose a constrained nonparametric estimator of the SPD.

Let $\mathscr{C} = (C_1, \ldots, C_n)^\top$ be the vector of the observed option prices at day t. The corresponding vector of the strike prices has the following structure:

$$\mathscr{K} = \begin{pmatrix} K_1 \\ \vdots \\ K_n \end{pmatrix} = \begin{pmatrix} k_1 \mathbf{1}_{n_1} \\ \vdots \\ k_p \mathbf{1}_{n_p} \end{pmatrix},$$

where $k_1 < k_2 < \cdots < k_p$, $n_j = \sum_{i=1}^{n} \boldsymbol{I}\left(K_i = k_j\right)$ with $\boldsymbol{I}(.)$ denoting the indicator function and $\mathbf{1}_n$ the vector of ones of length n.

For fixed time t and time to maturity $\tau = T - t$, the ith observed option price (corresponding to strike price K_i) follows the model

$$C_{t,i}(K_i, T) = \mu(K_i) + \varepsilon_i, \tag{3}$$

where ε_i are (possibly correlated) random errors, $\operatorname{Var} \varepsilon = V\sigma^2 > 0$.

1.2. Constrained linear model

From (1) and properties of a probability density it follows that the function of the true conditional means $\mu(.)$ in (3) has to satisfy no-arbitrage constraints: (A) positivity, (B) monotonicity, (C) convexity, (D) second derivative of $\mu()$ is a probability density. In Härdle and Hlávka (2006), the linear model (3) was reparametrized in terms of parameters $\beta = \left(\beta_0, \ldots, \beta_p\right)^\top$ so that an estimator of β can be interpreted as an estimator of the SPD. This is achieved by modeling the vector of conditional means $\mu = \left(\mu_1, \ldots, \mu_p\right)^\top$ as $\mu = \Delta\beta$, with

$$\Delta = \begin{pmatrix} 1 & \Delta_p^1 & \Delta_{p-1}^1 & \Delta_{p-2}^1 & \cdots & \Delta_3^1 & \Delta_2^1 \\ 1 & \Delta_p^2 & \Delta_{p-1}^2 & \Delta_{p-2}^2 & \cdots & \Delta_3^2 & 0 \\ \vdots & & & & & & \\ 1 & \Delta_p^{p-1} & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \tag{4}$$

where $\Delta_j^i = \max\left(k_j - k_i, 0\right)$ denotes the positive part of the distance between k_i and k_j, the ith and the jth $(1 \leqslant i \leqslant j \leqslant p)$ sorted distinct strike prices.

Denote as X_Δ a matrix in which ith row of matrix Δ is repeated n_i times, $i = 1, \ldots, p$ and consider the model $\mathscr{C} = X_\Delta \beta + \varepsilon$, where $\operatorname{Var} \varepsilon = V\sigma^2$.

Let us assume that the observed SPD develops in time: $\log \widetilde{\beta}_{t_i} = \log \widetilde{\beta}_{t_{i-1}} + (t_i - t_{i-1})^{1/2} u_i$, where u_i are iid random errors with $Eu_i = 0$, $\widetilde{\beta}_{t_i}$ is the vector of unknown parameters at time t_i, and $t_i \in (0, 1)$ denotes the time of the ith observation. Inserting this expression into the model for the ith option price: $C_i(k_j) = \Delta_j \widetilde{\beta}_{t_n}$, where Δ_j is the j-th row