Entropies, divergences and optimal statistical decisions about some financial models

Wolfgang Stummer

Abstract: We study Bayesian decision making based on observations of the price dynamics $(X_t : t \in [0,T])$ of a financial asset, when the hypothesis is the classical geometric Brownian motion and the alternative is a more general random diffusion process. We obtain exact formulae – respectively bounds – for the minimal mean decision loss (Bayes risk), and also for some generalized relative entropies ($I_\alpha-$divergences).

MSC 2000: 62C10, 60H10, 91B28
Key words: Bayesian statistical decision theory, generalized Black-Scholes-Merton models, generalized relative entropies, $I_\alpha-$divergences, Hellinger integrals.

1 Introduction

Let X_t be the value of a financial asset (for instance, a stock index) at time $t \geq 0$, and suppose that for the description of X_t we have the choice between the following two models $(\mathcal{H})$ and $(\mathcal{A})$:

$(\mathcal{H})$ the geometric Brownian motion X_t of Samuelson (1965), which is for example used in the Black-Scholes-Merton option pricing framework, and which is the (strong, hence weak) solution of the stochastic differential equation (SDE)

$$dX_t \;=\; c\; X_t\; dt \;+\; \sigma\; X_t\; dW_t\;, \qquad X_0 \;= x \;>\; 0\;, \tag{1}$$

with a given growth rate *constant* $c \in \mathbb{R}$ and volatility $\sigma > 0$, and with a standard Brownian motion W. The corresponding probability law of (1) (for infinite time horizon) is denoted by P_x;

$(\mathcal{A})$ the diffusion process X_t which is the (weak) solution of the SDE

$$dX_t \;=\; g(X_t)\; X_t\; dt \;+\; \sigma\; X_t\; d\overline{W}_t\;\;, \qquad X_0 \;= x \;>\; 0\;, \tag{2}$$

with a given growth rate *function* g, with the same volatility $\sigma > 0$ as in $(\mathcal{H})$, and with a standard Brownian motion $\overline{W}$. The corresponding probability law of (2) (for infinite time horizon) is denoted by Q_x. We suppose that the equation (2) differs from (1) in the sense that $Q_x \neq P_x$ for all $x > 0$.

In the hypothetic model $\mathcal{H}$, the random process $(X_t : t \geq 0)$ is a geometric Brownian motion, i.e.

$$X_t \;=\; x\; \exp\Big(\sigma W_t + \Big(c - \frac{\sigma^2}{2}\Big)\, t\Big)\;, \qquad t \geq 0. \tag{3}$$

We thank for the partial support by the the ASCR grant 1075104 and the MSMTV grant 1M 0572.

At each fixed time $t \geq 0$, X_t is (in $\mathcal{H}$, i.e. under P_x) lognormally distributed. In the alternative model ($\mathcal{A}$), the random process $(X_t : t \geq 0)$ need not (under Q_x) be lognormally distributed.

Within the framework introduced above, we study Bayes decisions with binary decision spaces $\mathcal{D} = \{d_{\mathcal{H}}, d_{\mathcal{A}}\}$ and loss functions

$$\begin{pmatrix} L\Big(d_{\mathcal{H}}, \mathcal{H}\Big) & L\Big(d_{\mathcal{H}}, \mathcal{A}\Big) \\ L\Big(d_{\mathcal{A}}, \mathcal{H}\Big) & L\Big(d_{\mathcal{A}}, \mathcal{A}\Big) \end{pmatrix} = \begin{pmatrix} 0 & L_{\mathcal{A}} \\ L_{\mathcal{H}} & 0 \end{pmatrix}, \tag{4}$$

with losses $L_{\mathcal{H}} > 0$ and $L_{\mathcal{A}} > 0$. Thus, $d_{\mathcal{H}}$ is assumed to be a zero-loss decision under $\mathcal{H}$ and $d_{\mathcal{A}}$ is a zero-loss decision under $\mathcal{A}$.

We are interested in Bayes decisions of the hypothesis $\mathcal{H}$ against the alternative $\mathcal{A}$ based on the random asset value observations $\mathcal{X}^T = (X_t : t \in [0,T])$. Formally, the Bayes decisions can be considered as functions $\delta = \delta(\mathcal{X}^T)$ of random paths $\mathcal{X}^T$ into $\{d_{\mathcal{H}}, d_{\mathcal{A}}\}$. The Bayes decision function minimizes the risk (average loss)

$$p_{\mathcal{H}}\, L_{\mathcal{H}}\, Pr[\, \delta(\mathcal{X}^T) = d_{\mathcal{A}} \mid \mathcal{H}\,] \; + \; p_{\mathcal{A}}\, L_{\mathcal{A}}\, Pr[\, \delta(\mathcal{X}^T) = d_{\mathcal{H}} \mid \mathcal{A}\,] \tag{5}$$

for given prior probabilities $p_{\mathcal{H}} = Pr[\mathcal{H}]$ for $\mathcal{H}$ and $p_{\mathcal{A}} = Pr[\mathcal{A}] = 1 - p_{\mathcal{H}}$ for $\mathcal{A}$ (which describe the model risk knowledge at time $t = 0$, prior to the random asset value observations $\mathcal{X}^T$). If $L_{\mathcal{H}} = L_{\mathcal{A}} = 1$ then the Bayes risk is the minimal average probability error (Bayes error) of the decision about $\mathcal{H}$ against $\mathcal{A}$, and the corresponding Bayes decision function represents a Bayes test of $\mathcal{H}$ against $\mathcal{A}$. Thus we have included into our considerations Bayes tests, and also the more general Bayes decisions corresponding to unequal $L_{\mathcal{H}}$ and $L_{\mathcal{A}}$.

Within the general statistical decision framework described above, the solution X_t of the modelling SDE (2) is typically not explicitly available, implying that it is usually hard to obtain explicit values for the corresponding decision-theoretic characteristics, such as the Bayes factor and the Bayes risk. Thus, it makes sense to find some *bounds* on these characteristics, especially for a large observation duration T. In order to do so, we use the following technical

Assumption A1. The non-stochastic condition

$$\sup_{a \in \mathbb{R}} \int_{a-1}^{a+1} \Big(g(e^{\varsigma})\Big)^2 \, d\varsigma \; < \; \infty \tag{6}$$

holds.

It was proven in [3] that this assumption guarantees the existence and uniqueness of a (weak) solution (X_t, Q_x) of (2) for all starting values $x > 0$; furthermore, one has $X_t > 0$ for all $t \geq 0$ (Q_x−almost surely).

2 Bounds on prior- and loss-independent quantities

In a straightforward way, formally (and under Assumption A1 also technically correctly) one can obtain from the prior binomial probabilities $p_{\mathcal{H}}$ for $\mathcal{H}$ and $p_{\mathcal{A}} = 1 - p_{\mathcal{H}}$ for $\mathcal{A}$ the posterior probabilities

$$p_{\mathcal{H}}^{post,T} = \frac{p_{\mathcal{H}}}{(1-p_{\mathcal{H}})\ Z_T\ +\ p_{\mathcal{H}}} \qquad \text{for } \mathcal{H}\ , \tag{7}$$

$$p_{\mathcal{A}}^{post,T} = \frac{(1-p_{\mathcal{H}})\ Z_T}{(1-p_{\mathcal{H}})\ Z_T\ +\ p_{\mathcal{H}}} \qquad \text{for } \mathcal{A}\ , \tag{8}$$

with Girsanov-type density

$$\frac{dQ_x}{dP_x}\bigg|_{[0,T]} = Z_T = \exp\left(\int_0^T \frac{g(X_v)-c}{\sigma}\, dW_v \ - \ \frac{1}{2} \int_0^T \frac{\big(g(X_v)-c\big)^2}{\sigma^2}\, dv \right) . \tag{9}$$

From this, one can compute the posterior odds ratio of $\mathcal{A}$ to $\mathcal{H}$ as

$$\frac{p_{\mathcal{A}}^{post,T}}{p_{\mathcal{H}}^{post,T}} = \frac{1-p_{\mathcal{H}}}{p_{\mathcal{H}}}\ Z_T\ . \tag{10}$$

Thus, the corresponding Bayes factor $\mathcal{BF} = \mathcal{BF}_T$ of the decision in favour of $\mathcal{A}$ at time T is given by

$$\mathcal{BF}_T = \frac{\text{posterior odds ratio of } \mathcal{A} \text{ to } \mathcal{H}}{\text{prior odds ratio of } \mathcal{A} \text{ to } \mathcal{H}} = Z_T\ . \tag{11}$$

As usual, the Bayes factor $\mathcal{BF}_T$ can be interpreted as the odds for $\mathcal{A}$ against $\mathcal{H}$ that are given by the data (here, the observed asset-value sample paths $\mathcal{X}^T$ in the period $[0,T]$). One can give the following bounds on the moments of $\mathcal{BF}_T$ with respect to the law P_x:

Theorem 2.1. *Let the Assumption A1 be satisfied. Then the following assertions hold:*
(a) For every real number $\alpha \in]0,1[$ there exist constants $c_1 > 0$ and $c_2 > 0$ such that for all observation durations $T > 0$ and all starting asset values $x > 0$

$$EP_x\big[\,(\mathcal{BF}_T)^\alpha\big] \ \geq\ \exp\Big(- c_1 - c_2\ T\Big)\ . \tag{12}$$

(b) For all observation durations $T > 0$ and all starting asset values $x > 0$

$$EP_x\big[\,\mathcal{BF}_T\big] \ =\ 1\ . \tag{13}$$

(c) For every real number $\alpha \notin [0,1]$ there exist constants $c_3 > 0$ and $c_4 > 0$ such that for all observation durations $T > 0$ and all starting asset values $x > 0$

$$EP_x\big[\,(\mathcal{BF}_T)^\alpha\big] \ \leq\ \exp\Big(c_3 + c_4\ T\Big)\ . \tag{14}$$

An analoguous result can be obtained for the Bayes factor moments with respect to the law Q_x. Apart from the important Bayes factor, it is also useful to study the "distance" between the two models (2) and (1), especially for large durations T of the observation periods. For instance, one can investigate the I_α^T−divergences between the two corresponding probability laws Q_x and P_x, defined by

$$I_\alpha^T(Q_x||P_x) \;=\; \int f_\alpha\left(\frac{dQ_x}{dP_x}\Big|_{[0,T]}\right) dP_x \;,$$

with the nonnegative functions $f_\alpha : [0,\infty[\to [0,\infty[$ defined by

$$f_\alpha(\rho) \;=\; \begin{cases} -\log\rho+\rho-1, & \text{if } \alpha=0,\\ \frac{\alpha\rho+1-\alpha-\rho^\alpha}{\alpha(1-\alpha)}, & \text{if } \alpha\in\mathbb{R}\backslash\{0,1\},\\ \rho\,\log\rho+1-\rho, & \text{if } \alpha=1. \end{cases}$$

Basic facts on I_α^T−divergences of general measures can be found e.g. in [1] and on I_α^T−divergences of discrete measures in [2]. Of course, $I_1^T(Q_x||P_x)$ is nothing but the diffusion version of the relative entropy (Kullback-Leibler information measure). Again, because of the complexity of the models considered here, one usually has to be satisfied by bounds which give a "rough, but definite" idea on what can happen in the worst resp. best case. Such bounds on $I_\alpha^T(Q_x||P_x)$ can be obtained by using the bounds on the Bayes factor given in Theorem 2.1. In fact, one gets the following

Theorem 2.2. *Let the Assumption A1 be satisfied. Then the following assertions hold:*
(a) For every real number $\alpha\in]0,1[$ *there exist constants* $c_1>0$ *and* $c_2>0$ *such that for all observation durations* $T>0$ *and all starting asset values* $x>0$

$$I_\alpha^T(Q_x||P_x) \;\le\; \frac{1}{\alpha\,(1-\alpha)}\left\{1-\exp\Big(-c_1-c_2\,T\Big)\right\}. \tag{15}$$

(b) For every real number $\alpha\notin[0,1]$ *there exist constants* $c_3>0$ *and* $c_4>0$ *such that for all observation durations* $T>0$ *and all starting asset values* $x>0$

$$I_\alpha^T(Q_x||P_x) \;\le\; \frac{1}{\alpha\,(\alpha-1)}\left\{\exp\Big(c_3+c_4\,T\Big)\;-\;1\right\}. \tag{16}$$

3 Bounds on the Bayes risk

In formula (5) of Section 1, we introduced the risk (average loss) of a decision function $\delta_T(\mathcal{X}^T)=\delta(\mathcal{X}^T)$ taking values in $\mathcal{D}=\{d_\mathcal{H},d_\mathcal{A}\}$. If we reject the hypothesis $\mathcal{H}$ whenever the observed asset value sample path $\mathcal{X}^T=(X_t:t\in[0,T])$ lies within a critical region $G^T=\delta_T^{-1}(d_\mathcal{A})$, we can rewrite this risk in the form

$$\mathcal{R}_x(G^T) \;=\; p_\mathcal{H}\,L_\mathcal{H}\,P_x[G^T] \;+\; p_\mathcal{A}\,L_\mathcal{A}\,Q_x[\Omega^T-G^T]\;.$$

By means of the parameters

$$\lambda_{\mathcal{H}} \ = \ p_{\mathcal{H}}\, L_{\mathcal{H}} \qquad \text{and} \qquad \lambda_{\mathcal{A}} \ = \ p_{\mathcal{A}} L_{\mathcal{A}}\ ,$$

which carry combined prior and loss information, we obtain the formula

$$\mathcal{R}_x(G^T) \ = \ \lambda_{\mathcal{H}}\ P_x[G^T\,] \ + \ \lambda_{\mathcal{A}}\ \Big(1 \ - \ Q_x[G^T\,]\Big)\ . \tag{17}$$

By definition, the Bayes risk $\mathcal{R}_x^T$ minimizes the risk, i.e.

$$\mathcal{R}_x^T \ = \ \min \mathcal{R}_x(G^T)\ , \tag{18}$$

where the minimum is taken over all measurable sets $G^T \subset \Omega^T$ of sample paths. By (17),

$$\begin{aligned} \mathcal{R}_x(G^T) \ &= \ \lambda_{\mathcal{A}} \ + \ \int_{G^T} (\lambda_{\mathcal{H}} - \lambda_{\mathcal{A}}\, \mathcal{BF}_T)\ dP_x \\ &\geq \ \lambda_{\mathcal{A}} \ + \int_{G^T \cap \{\lambda_{\mathcal{H}} \leq \lambda_{\mathcal{A}} \mathcal{BF}_T\}} (\lambda_{\mathcal{H}} \ - \lambda_{\mathcal{A}}\, \mathcal{BF}_T)\ dP_x \\ &\geq \ \lambda_{\mathcal{A}} \ + \int_{\{\lambda_{\mathcal{H}} \leq \lambda_{\mathcal{A}} \mathcal{BF}_T\}} (\lambda_{\mathcal{H}} \ - \lambda_{\mathcal{A}}\, \mathcal{BF}_T)\ dP_x . \end{aligned} \tag{19}$$

Therefore, the Bayes risk is achieved by the decision rule $\delta(\mathcal{X}^T)$ which rejects $\mathcal{H}$ (decides for $\mathcal{A}$) if the observed path $\mathcal{X}^T$ is contained in the sample path set

$$G^T_{min} \ = \ \{\lambda_{\mathcal{H}} \leq \lambda_{\mathcal{A}}\, \mathcal{BF}_T\}\ , \tag{20}$$

and rejects $\mathcal{A}$ (decides for $\mathcal{H}$) if $\mathcal{X}^T$ is contained on the complement of this set. Accordingly, with the help of (17) we obtain the Bayes risk as

$$\begin{aligned} \mathcal{R}_x^T \ &= \ \mathcal{R}_x(G^T_{min}) \ = \ \int_{G^T_{min}} \lambda_{\mathcal{H}}\ dP_x \ + \int_{\Omega^T \setminus G^T_{min}} \lambda_{\mathcal{A}}\ \mathcal{BF}_T\ dP_x \\ &= \int_{\Omega^T} \min\{\lambda_{\mathcal{H}}, \lambda_{\mathcal{A}}\, \mathcal{BF}_T\}\ dP_x\ . \end{aligned} \tag{21}$$

By using the same kind of argumentation as in Section 2, one can see that a direct explicit formula for the Bayes risk is hard to obtain in our context, since the Bayes factor involves the advanced concept of stochastic integrals. Thus, one aims at least for some bounds on the Bayes risk. Applying part (a) of the Bayes-factor-treating Theorem 2.1 at the formula (21) in a non-straightforward way, one arrives at the following result concerning the Bayes risk $\mathcal{R}_x^T$, especially in terms of the sample path observation duration T:

Theorem 3.1. *If Assumption A1 is satisfied, then there exist constants* $c_5 > 0$ *and* $c_6 > 0$ *such that for all prior-loss-information parameters* $\lambda_{\mathcal{A}} \geq 0$*,* $\lambda_{\mathcal{H}} \geq 0$ *as well as for all observation durations* $T > 0$ *and all starting asset values* $x > 0$

$$\mathcal{R}_x^T \;\geq\; \frac{\lambda_{\mathcal{A}}\, \lambda_{\mathcal{H}}}{\lambda_{\mathcal{A}} + \lambda_{\mathcal{H}}}\; \exp\Big(- c_5 - c_6\; T \Big)\,. \tag{22}$$

Theorem 3.1 states that the Bayes risk under hypothesis and alternative given by (1) and (2), *cannot* go to zero faster than exponentially in T.

In this talk, we shall also present (i) more precise bounds for $\mathcal{R}_x^T$ under more restrictive assumptions, (ii) an explicit formula for $\mathcal{R}_x^T$ in the special case where $\mathcal{A}$ is also a geometric Brownian motion (different from $\mathcal{H}$), (iii) some sketches of the underlying proofs, as well as (iv) illuminations on the special case of a "classical" Bayesian *testing* set-up.

The full details of the abovementioned investigations will appear in [4].

Furthermore, we indicate how to generalize this framework to the time-inhomogeneous case.

References

[1] F. Liese, F. and I. Vajda. *Convex Statistical Distances.* Teubner, Leipzig, 1987.

[2] T.R.C. Read and N.A.C. Cressie. *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Springer, New York, 1988.

[3] W. Stummer. On a statistical information measure for a generalized Samuelson-Black-Scholes model. *Statistics & Decisions*, 19:289–314, 2001.

[4] W. Stummer and I. Vajda. Optimal Statistical Decisions About Some Alternative Financial Models. *Journal of Econometrics*, to appear, 2006.

Wolfgang Stummer: University of Erlangen-Nürnberg, Department of Mathematics, Bismarckstrasse $1\frac{1}{2}$, Erlangen, 91054, Germany, *stummer@mi.uni-erlangen.de*