Journal of Mathematical Sciences, Vol. 153, No. 3, 2008

OPTIMAL ROBUST MEAN-VARIANCE HEDGING IN INCOMPLETE FINANCIAL MARKETS

N. Lazrieva and T. Toronjadze UDC 519.2

Abstract. An optimal B-robust estimate is constructed for the multidimensional parameter in the drift coefficient of a diffusion-type process with a small noise. The optimal mean-variance robust (optimal V-robust) trading strategy is to hedge (in the mean-variance sense) the contingent claim in an incomplete financial market with an arbitrary information structure and a misspecified volatility of the asset price, which is modelled by a multidimensional continuous semimartingale. The obtained results are applied to the stochastic volatility model, where the model of the latent volatility process contains the unknown multidimensional parameter in the drift coefficient and a small parameter in the diffusion term.

CONTENTS

1. Introduction, Motivation, and Results

The hedging and pricing of contingent claims in incomplete financial markets and dynamic portfolio selection problems are important issues in the modern theory of finance. These problems are associated with the so-called mean-variance approach.

For determining a "good" hedging strategy in an incomplete market with an arbitrary information structure $F = (\mathcal{F})_{0\le t\le T}$, one riskless asset and d, $d \ge 1$, risky assets whose price process is a semimartingale X, the mean-variance approach suggests using the quadratic criterion to measure the hedging error, i.e. to solve the mean-variance hedging problem introduced by Föllmer and Sondermann [10]:

$$\text{minimize}\quad E\left(H - x - \int_0^T \theta_t dX_t\right)^2 \quad \text{over all}\quad \theta\in\Theta, \tag{1.1}$$

where the contingent claim H is a $\mathcal{F}_T$-measurable square-integrable random variable (r.v.), x is the initial investment, Θ is the class of admissible trading strategies, and T is the investment horizon.

The mean-variance formulation by Markowitz [26] provides a foundation for a single period portfolio selection (see also [27]). In [22], the concept of Markowitz's mean-variance formulation for finding the

Translated from Sovremennaya Matematika i Ee Prilozheniya (Contemporary Mathematics and Its Applications), Vol. 45, Martingale Theory and Its Application, 2007.

 1072–3374/08/1533–0262

optimal portfolio policy and determining the efficient frontier in analytical form was extended to the multiperiod portfolio selection.

As was pointed out in [22], the results in the multiperiod mean-variance formulation with one riskless asset can be derived by using the results of the mean-variance hedging formulation.

Therefore, mean-variance hedging is a powerful approach for both major problems mentioned above.

Problem (1.1) was intensively investigated in the last decade (see, e.g., [8, 9, 11, 18, 28, 31, 33, 36–38]).

The stochastic volatility model proposed by Hull and White [13] and Scott [39], in which the stock price volatility is a random process, is a popular model of an incomplete market, where the mean-variance hedging approach can be used (see, e.g., [13, 18, 24, 31]).

Consider the stochastic volatility model described by the following system of SDE:

$$\begin{gathered} dX_t = X_t\,dR_t, \quad X_0 > 0, \\ dR_t = \mu_t(R_t, Y_t)\,dt + \sigma_. dw_t^R, \quad R_0 = 0, \\ \sigma_t^2 = f(Y_t), \\ dY_t = a(t, Y_t; \alpha)\,dt + \varepsilon\,dw_t^\sigma, \quad Y_0 = 0, \end{gathered} \tag{1.2}$$

where $w = (w^R, w^\sigma)$ is the standard two-dimensional Wiener process defined on a complete probability space $(\Omega, \mathcal{F}, P)$, $F^w = (\mathcal{F}_t^w)_{0\le t\le T}$ is the P-augmentation of the natural filtration $\mathcal{F}_t^w = \sigma(w_s, 0 \le s \le t)$, $0 \le t \le T$, generated by w, $f(\cdot)$ is a continuous one-to-one positive locally bounded function (e.g., $f(x) = e^x$), $\alpha = (\alpha_1, \dots, \alpha_m)$, $m \ge 1$, is the vector of unknown parameters, and ε, $0 < \varepsilon \ll 1$, is a small number. Assume that system (1.2) has a unique strong solution.

This model is analogous to the model proposed by Renault and Touzi [32]. The principal difference is the presence of a small parameter ε in our model, which is due to the assumption that the volatility of the randomly fluctuated volatility process is small (see also [40]). Thus, the assumption enables us to use the prices of trading options with short, nearest to the current time value maturities for volatility process filtration and parameter estimation purposes (see below). In contrast, the model [32] needs to assume that there exist trading derivatives with any (up to infinity) maturities.

An important feature of the stochastic volatility models is that the volatility process Y is an unobservable (latent) process. To obtain an explicit form of the optimal trading strategy, a full knowledge of the model of the process Y is necessary, and hence one needs to estimate the unknown parameter $\alpha = (\alpha_1, \dots, \alpha_m)$, $m \ge 1$.

A variety of estimation procedures are used, which involve either a direct statistical analysis of the historical data or the use of implied volatilities extracted from prices of existing traded derivatives.

For example, one can use the following method based on historical data.

Fix the time variable t. From observations $X_{t_0^{(n)}}, \dots, X_{t_n^{(n)}}$, $0 = t_0^{(n)} < \dots < t_n^{(n)} = t$, $\max_j [t_{j+1}^{(n)} - t_j^{(n)}] \to 0$ as $n \to 0$, we calculate the realization of yield process $R_t = \int_0^t \frac{dX_s}{X_s}$, and then calculate the sum

$$S_n(t) = \sum_{j=0}^{n-1} |R_{t_{j+1}^{(n)}} - R_{t_j^{(n)}}|^2.$$

It is well known (see, e.g., [23]) that

$$S_n(t) \xrightarrow{P} \int_0^t \sigma_s^2\,ds \quad \text{as} \quad n \to \infty.$$

Since $\sigma_t^2(\omega) = f(Y_t)$ is a continuous process, we obtain

$$\sigma_t^2(\omega) = \lim_{\Delta \downarrow 0} \frac{F(t+\Delta, \omega) - F(t, \omega)}{\Delta},$$

where

$$F(t, \omega) = \int_0^t \sigma_s^2(\omega) ds.$$

Hence the realization $(y_t)_{0 \le t \le T}$ of the process Y can be found by the formula $y_t = f^{-1}(\sigma_t^2)$, $0 \le t \le T$.

More sophisticated methods using the same idea can be found, e.g., in [5, 30].

We can use the reconstructed sample path (y_t), $0 \le t \le T$, to estimate the unknown parameter α in the drift coefficient of the diffusion process Y.

The second, market price adjusted procedure for reconstructing the sample path of the volatility process Y and the parameter estimate was suggested in [32], where the authors used implied volatility data.

We present a brief review of this method adapted to our model (1.2).

Suppose that the volatility risk premium $\lambda^\sigma \equiv 0$, which means that the risk from the volatility process is noncompensated (or can be diversified away). Then the price $C_t(\sigma)$ of the European call option can be calculated by the Hull and White formula (see, e.g., [32]), and the Black–Scholes (BS) implied volatility $\sigma^i(\sigma)$ can be found as a unique solution of the equation

$$C_t(\sigma) = C_t^{BS}(\sigma^i(\sigma)),$$

where $C^{BS}(\sigma)$ denotes the standard BS formula written as a function of the volatility parameter σ.

Here (for further estimational purposes), only in-the-money options are used.

Under some technical assumptions (see [32, Proposition 5.1] and [23] for the general diffusion of the volatility process)

$$\frac{\partial \sigma_t^i(\sigma, \alpha)}{\partial \sigma_t} > 0 \tag{1.3}$$

(recall that the drift coefficient of the process Y depends on the unknown parameter α).

Fix the current value of the time parameter t, $0 \le t \le T$, and let $0 < T_1 < T_2 < \cdots < T_{k-1} < t < T_k$ be the maturity times of some traded in-the-money options.

Let $\sigma_{t_j^\varepsilon}^{i^*}$ be the observations of the implied volatility at the time instants $0 = t_0^\varepsilon < t_1^\varepsilon < \cdots < t_{[t/\varepsilon]} = t$, $\max_j [t_{j+1}^\varepsilon - t_j^\varepsilon] \to 0$ as $\varepsilon \to 0$.

Then, using (1.3) and solving the equation

$$\sigma_{t_j^\varepsilon}^i(\sigma_{t_j^\varepsilon}, \alpha) = \sigma_{t_j^\varepsilon}^{i^*},$$

one can obtain the realization $\{\widetilde{\sigma}_{t_j^\varepsilon}\}$ of the volatility (σ_t), and thus, using the formula $y_{t_j^\varepsilon} = f^{-1}(\widetilde{\sigma}_{t_j^\varepsilon}^2)$, the realization $\{y_{t_j^\varepsilon}\}$ of the volatility process (Y_t), which can be viewed as the realization of the nonlinear AR(1) process:

$$Y_{t_{j+1}^\varepsilon} - Y_{t_j^\varepsilon} = a(t_j^\varepsilon, Y_{t_j^\varepsilon}; \alpha)(t_{j+1}^\varepsilon - t_j^\varepsilon) + \varepsilon(w_{t_{j+1}^\varepsilon}^\sigma - w_{t_j^\varepsilon}^\sigma).$$

Using the data $\{y_{t_j^\varepsilon}\}$, one can construct the MLE $\widehat{\alpha}_t^\varepsilon$ of the parameter α (see, e.g., [19, 25, 26].

Recall the scheme for constructing MLE. Using obvious simple notation, rewrite the previous AR(1) process in the form

$$Y_{j+1} - Y_j = a(t_j, Y_j; \alpha)\Delta + \varepsilon \Delta w_j^\sigma.$$

Then

$$\frac{\partial}{\partial y} P\{Y_{j+1} \le y \mid Y_j\} = \frac{1}{\sqrt{2\pi\Delta\varepsilon}} \exp\left(-\frac{(y - Y_j - a(t_j, Y_j; \alpha)\Delta)^2}{2\varepsilon^2\Delta}\right) =: \varphi_{j+1}(y, Y_j; \alpha),$$

and the log-derivative of the likelihood process $\ell_t = (\ell_t^{(1)}, \dots, \ell_t^{(m)})$ is given by the relation

$$\ell_t^{(i)} = \sum_j \ell_{j+1}^{(i)}, \quad i = \overline{1, m},$$

where

$$\ell_{j+1}^{(i)}(y; \alpha) = \frac{\partial}{\partial \alpha_i} \ln \varphi_{j+1}(y, Y_j; \alpha) = \frac{1}{\varepsilon^2\Delta}(y - Y_j - a(t_j, Y_j; \alpha)\Delta)\dot{a}^{(i)}(t_j, Y_j; \alpha)\Delta.$$

Hence the MLE is a solution (under some conditions) of the system of equations

$$\frac{1}{\varepsilon^2\Delta} \sum_j (y_{j+1} - y_j - a(t_j, y_j; \alpha)\Delta)\dot{a}^{(i)}(t_j, y_j; \alpha)\Delta = 0, \quad i = \overline{1, m},$$

where the reconstructed data $\{y_j\} = \{y_{t_j^\varepsilon}\}$ are substituted.

Following [32], let us introduce the functionals

$$HW_\varepsilon^{-1} : \widehat{\alpha}_t^\varepsilon(p) \to \left(y_{t_j^\varepsilon}^{(p+1)}, \ 0 \le j \le \left[\frac{t}{\varepsilon}\right]\right),$$

$$MLE_\varepsilon : \left(y_{t_j^\varepsilon}^{(p+1)}, \quad 0 \le j \le \left[\frac{t}{\varepsilon}\right]\right) \to \widehat{\alpha}_t^\varepsilon(p+1),$$

$$\phi_\varepsilon = MLE_\varepsilon \circ HW_\varepsilon^{-1}.$$

Starting from some constant initial value (or the preliminary estimate obtained, e.g., from the historical data) one can compute a sequence of estimates

$$\widehat{\alpha}_t^\varepsilon(p+1) = \phi_\varepsilon(\widetilde{\alpha}_t^\varepsilon(p)), \quad p \ge 1.$$

If the operator ϕ_ε is a strong contraction in a neighborhood of the true value of the parameter α^0 for a sufficiently small ε, then one can define the estimate $\widehat{\alpha}_t^\varepsilon$ as the limits of the sequence $\{\widehat{\alpha}_t^\varepsilon(p)\}_{p\ge1}$. It was proved in [32] that $\widehat{\alpha}_t^\varepsilon$ is a strong consistent estimate of the parameter α.

Let us return to our consideration.

Interpolating in some way the corresponding (to the estimate $\widehat{\alpha}_t^\varepsilon$) realization $\{y_{t_j^\varepsilon}\}$, we obtain the reconstructed continuous sample path $(y_s)_{0\le s\le t}$ of the latent process Y, which can be used for further analysis.

Unfortunately, both described statistical procedures are highly sensitive to errors at all steps of the parameter identification process.

Hence this is a natural place for introducing the robust procedure of parameter estimates.

Suppose that the sample path $(y_s)_{0\le s\le t}$ comes from the observation of process $(\widetilde{Y}_s)_{0\le s\le t}$ with distribution $\widetilde{P}_\alpha^\varepsilon$ from the shrinking contamination neighborhood of the distribution P_α^ε of the basic process $Y = (Y_s)_{0\le s\le t}$. That is,

$$\frac{d\widetilde{P}_\alpha^\varepsilon}{dP_\alpha^\varepsilon}\bigg| \mathcal{F}_t^w = \mathcal{E}_t(\varepsilon N^\varepsilon), \tag{1.4}$$

where $N^\varepsilon = (N_s^\varepsilon)_{0\le s\le t}$ is a P_α^ε-square integrable martingale and $\mathcal{E}_t(M)$ is the Dolean exponential of martingale M.

In the diffusion-type framework, (1.4) represents the Huber gross error model (as is explained in Remark 2.2). The model of type (1.4) of contamination of measures for statistical models with filtration was suggested by Lazrieva and Toronjadze [20, 21].

In Sec. 2, we study the problem of constructing the robust estimates for the contamination model (1.4).

In Sec. 2.1, we give a description of the basic model and the definition of consistent uniformly linear asymptotically normal (CULAN) estimates connected with the basic model (Definition 2.1).

In Sec. 2.2, we introduce the notion of shrinking contamination neighborhood described in terms of the contamination of the nominal distribution, which naturally leads to the class of alternative measures (see (2.18) and (2.19)).

In Sec. 2.3, we study the asymptotic behavior of CULAN estimates under alternative measures (Proposition 2.2), which is a basis for the formulation of the optimization problem. The optimization problem is solved, which leads to the construction of the optimal B-robust estimate (Theorem 2.1).

Based on the limit theorem (Sec. 2.1), one can construct the asymptotic confidence region of level γ for unknown parameter α:

$$\lim_{\varepsilon\to\infty} P_\alpha^\varepsilon\left(\varepsilon^{-2}(\alpha-\alpha_t^{*,\varepsilon})'V^{-1}(\psi^*;\alpha_t^{*,\varepsilon})(\alpha-\alpha_t^{*,\varepsilon})\le\chi_\gamma^2\right)=1-\gamma,$$

where χ_γ^2 is a quantile of order $1-\gamma$ of the χ^2-distribution with m degrees of freedom and $V(\psi^*;\alpha)$ is given by (2.17).

This region shrinks to the estimate $\alpha_t^{*,0}$ as $\varepsilon\to 0$.

Now if the coefficient $a(t,y;\alpha)$ in (1.2) is such that the solution $Y_t^\varepsilon(\alpha)$ of the SDE (1.2) is continuous with respect to the parameter α (see, e.g., [16]), then the confidence region of parameter α is mapped into the confidence interval for $Y_t^\varepsilon(\alpha)$, which shrinks to $Y_t^*=Y_t^0(\alpha_t^{*,0})$. Furthermore, by the function f, the latter interval is mapped into the confidence interval for σ_t, which shrinks to $\sigma_t^*=f^{1/2}(Y_t^0(\alpha_t^{*,0}))$. Denote by σ_t^0 the center of this interval. Then the interval can be written in the form

$$\sigma_t=\sigma_t^0+\delta(\varepsilon)h_t,$$

where $\delta(\varepsilon)\to 0$ as $\varepsilon\to 0$ and $h\in\mathcal{H}$ (see (3.19)).

Thus, we arrive at the asset price model (1.2) with misspecified volatility, and it is natural to consider the problem of constructing the robust trading strategy to hedge a contingent claim H.

We investigate this problem in the mean-variance setting in Sec. 3. We consider the general situation where the asset price is modelled by a d-dimensional continuous semimartingale and the information structure is given by some general filtration.

In Sec. 3.1, we give a description of the financial market model.

In Sec. 3.2, we collect the facts concerning the variance-optimal equivalent local martingale measure, which plays a key role in the mean-variance hedging approach.

In Sec. 3.3, we construct the "optimal robust hedging strategy" (Theorem 3.1) by approximating the optimization problem (3.26) by problem (3.28). As is mentioned in Remark 3.2, such an approach and term are common in robust statistical theory. In contact to the optimal B-robustness (see Sec. 2), here, we develop the approach known in robust statistics as the optimal V-robustness (see [12]).

Note that our approach allows incorporating current information on the underlying model and hence is adaptive. Precisely, passing from a time value t to $t+\tau$, $\tau>0$, when more information on the market prices are available, the asymptotic variance-covariance of the constructed estimate $\alpha_t^{*,\varepsilon}$ becomes smaller, and hence the estimation procedure becomes more precise.

In [35], the adaptive approach to risk management under a general uncertainty (restricted information) was developed. As is mentioned in this paper, there exists a series of investigations dealing with various type of adaptive approaches (see the list of references in [35]). But in all these papers (except for [35]), the uncertainty is only in the stock appreciation rate in contrast to our consideration where the model misspecification is due to the volatility parameter.

The consideration of misspecified asset price models was initiated in [1, 2].

Various authors in different settings attacked the robustness problem. The method used in Sec. 3 was suggested by Toronjadze [41] for an asset price process modelled by a one-dimensional process. As will be shown in Remark 3.2 below, in the simplest case where the asset price process is a martingale with respect to the initial measure P, and it is possible to find the solution of the "exact" optimization problem (3.26), this solution coincides with the solution of an approximating optimization problem

(3.28). In a more general situation (when the asset price process is no longer the P-martingale), the investigation of problem (3.26) by, e.g., control theory methods appears to be difficult. Anyway, we do not know the solution of problem (3.26).

Let us return to the stochastic volatility model (1.2) and describe the successive steps of our approach:

(1) For each current time value t, $0 < t < T$, reconstruct the sample path $(y_s)_{0\le s\le t}$ using the historical data or the tradable derivatives prices.

(2) Using the approach developed in Sec. 2, calculate the value $\alpha_t^{*,\varepsilon}$ of the robust estimate of the parameter α (i.e., construct the deterministic function $t \to \alpha_t^{*,\varepsilon} \in \mathbb{R}^m$), and then find the confidence region for the parameter α.

(3) Based on the volatility process model, find the confidence interval for $Y_t(\alpha)$.

(4) Denoting $a^*(t,y) = a(t,y;\alpha_t^{*,\varepsilon})$, where $a(t,y;\alpha)$ is the drift coefficient of the volatility process, consider the stochastic volatility model with misspecified asset price model and fully specified volatility process model:

$$dX_t = X_t\,dR_t, \quad X_0 > 0,$$
$$dR_t = (\sigma_t^0 + \delta(\varepsilon)h_t)dM_t^0, \quad R_0 = 0,$$
$$dY_t = a^*(t,Y_t)\,dt + \varepsilon\,dw_t^\sigma, \quad Y_0 = 0, \quad 0 \le t \le T,$$

where

$$dM_t^0 = k_t\,dt + dw_t^R,$$

$h \in \mathcal{H}$ and σ_t^0 is the center of the confidence interval of volatility.

Using Theorem 3.1, construct the optimal robust hedging strategy by the formula (3.45):

$$\theta_t^* = \frac{1}{\sigma_t^0}\left[\psi_t^{1,H} + \zeta_t(V_t^* - (\psi_t^H)'U_t\right],$$

where all objects are defined in Theorem 3.1.

It should be noted that if one constructs the hedging strategy $\widetilde{\theta}_t^*$ by the above-given formula with $\sigma_t^{*,\varepsilon} = f^{1/2}(Y_t^\varepsilon(\alpha_t^{*,\varepsilon}))$ instead of σ_t^0, then the strategies $\widetilde{\theta}_t^*$ and θ_t^* are different, since $\sigma_t^{*,\varepsilon} \ne \sigma_t^0$, in general. Hence the value $\Delta_t = |\sigma_t^{*,\varepsilon} - \sigma_t^0|$ defines the correction term between the robust, θ_t^*, and non-robust, $\widetilde{\theta}^*$, strategies.

In the nontrivial case where $k_t = k(Y_t)$, the variance-optimal martingale measure $\widetilde{P}$ is given by (3.17), $\zeta_t = -k_t\mathcal{E}_t(-k\cdot M^0)$ (see Sec. 3.2), and the process $(X_t,Y_t)_{0\le t\le T}$ is a Markov process. If $H = h(X_T,Y_T)$ ($h(x,y)$ is some function), then $\widetilde{V}_t^H = E^{\widetilde{P}}(H|\mathcal{F}_t^w) = E^{\widetilde{P}}(h(X_T,Y_T)|\mathcal{F}_t^w) = v(t,X_T,Y_T)$ and if, e.g., $v(t,x,y) \in C^{1,2,2}$, then v is a unique solution of the following partial differential equation:

$$\frac{\partial v}{\partial t} + a^*\frac{\partial v}{\partial y} + \frac{1}{2}\left(\varepsilon^2\frac{\partial^2 v}{\partial y^2} + x^2v^2\frac{\partial^2 v}{\partial x^2}\right) = 0,$$

with the boundary condition $v(T,x,t) = h(x,y)$. A more general situation with a nonsmooth v is considered in [18, 24].

Further, one can find the Galtchouck–Kunita–Watanabe decomposition of r.v. H (see, e.g., [31]) setting

$$\xi_t^H = \frac{\partial v(t,X_t,Y_t)}{\partial x}, \quad L_T^H = \varepsilon\int_0^T \frac{\partial v}{\partial y}(t,X_t,Y_t)\,dw_t^\sigma,$$

and calculate ψ_t^H, L_T, and V_t^* using formulas (4.13) and (4.14) of [33].

Thus one obtains the explicit solution of the mean-variance hedging problem.

Finally, below is a short summary of the approach:

(a) Incorporate the robust procedure in the statistical analysis of the volatility process. That is, construct and use the optimal B-robust estimate of the unknown parameter in the drift coefficient of the volatility process in the model. The parameter estimation naturally leads to the asset price model misspecification.

(b) Incorporate the second robust procedure in the financial analysis of the contingent claim hedging. That is, construct and use the optimal V-robust trading strategy for hedging purposes.

In our opinion, this "double robust" strategy is more attractive to protect the hedger against the possible errors.

The general asymptotic theory of estimation can be found in [14]; the theory of robust statistics is developed in [12, 34]; the theory of the trend parameter estimates for a diffusion process with small noise is developed in [17]; the book of Musiela and Rutkowsky [29] is devoted to the mathematical theory of finance, and, finally, the general theory of martingales can be found in [15].

2. Optimal B-Robust Estimates

2.1. Basic model. CULAN estimates. The basic model of observations is described by the SDE

$$dY_s = a(s, Y; \alpha)\, ds + \varepsilon\, dw_s, \quad Y_0 = 0, \quad 0 \le s \le t, \tag{2.1}$$

where t is a fixed number, $w = (w_s)_{0\le s\le t}$ is the standard Wiener process defined on the filtered probability space $(\Omega, \mathcal{F}, F = (\mathcal{F}_s)_{0\le s\le t}, P)$ satisfying the usual conditions, $\alpha = (\alpha_1, \dots, \alpha_m)$, $m \ge 1$, is the unknown parameter to be estimated, $\alpha \in \mathcal{A} \subset \mathbb{R}^m$, $\mathcal{A}$ is an open subset of $\mathbb{R}^m$, and ε, $0 < \varepsilon \ll 1$, is a small parameter (index of series). In our further considerations all limits correspond to $\varepsilon \to 0$.

Denote by $(C_t, \mathcal{B}_t)$ the measurable space of functions $x = (x_s)_{0\le s\le t}$ continuous on $[0, t]$ with σ-algebra $\mathcal{B}_t = \sigma(x : x_s, s \le t)$. We set $\mathcal{B}_s = \sigma(x : x_u, u \le s)$.

Assume that for each $\alpha \in \mathcal{A}$, the drift coefficient $a(s, x; \alpha)$, $0 \le s \le t$, $x \in C_t$ is a known nonanticipative (i.e., $\mathcal{B}_s$-measurable for each s, $0 \le s \le t$) functional satisfying the functional Lipshitz and linear growth conditions **L**:

$$|a(s, x^1; \alpha) - a(s, x^2; \alpha)| \le L_1 \int_0^s |x_u^1 - x_u^2|\, dk_u + L_2 |x_s^1 - x_s^2|,$$

$$|a(s, x; \alpha)| \le L_1 \int_0^s (1 + |x_u|)\, dk_u + L_2 (1 + |x_s|),$$

where L_1 and L_2 are constants independent of α and $k = (k(s))_{0\le s\le t}$ is a nondecreasing right-continuous function, $0 \le k(s) \le k_0$, $0 : k_0 < \infty$, $x^1, x^2 \in C_t$.

Then, as is well known (see, e.g., [23]), for each $\alpha \in \mathcal{A}$, Eq. (2.1) has a unique strong solution $Y^\varepsilon(\alpha) = (Y_s^\varepsilon(\alpha))_{0\le s\le t}$, and, in addition (see [17]),

$$\sup_{0\le s\le t} |Y_s^\varepsilon(\alpha) - Y_s^0(\alpha)| \le C\varepsilon \sup_{0\le s\le t} |w_s| \quad P\text{-}a.s.,$$

with some constant $C = C(L_1, L_2, k_0, t)$, where $Y^0(\alpha) = (Y_s^0(\alpha))_{0\le s\le t}$ is a solution of the following nonperturbated differential equation:

$$dY_s = a(s, Y; \alpha)\, ds, \quad Y_0 = 0. \tag{2.2}$$

Change the initial problem of estimation of the parameter α by the equivalent one, in which the observations are modelled according to the following SDE:

$$dX_s = a_\varepsilon(s, X; \alpha)\, ds + dw_s, \quad X_0 = 0, \tag{2.3}$$

where $a_\varepsilon(s, x; \alpha) = \frac{1}{\varepsilon} a(s, \varepsilon x; \alpha)$, $0 \le s \le t$, $x \in C_t$, $\alpha \in \mathcal{A}$.

It is clear that if $X^\varepsilon(\alpha) = (X_s^\varepsilon(\alpha))_{0\le s\le t}$ is a solution of SDE (2.3), then for each $s \in [0, t]$, $\varepsilon X_s^\varepsilon(\alpha) = Y_s^\varepsilon(\alpha)$.

Denote by P_α^ε the distribution of process $X^\varepsilon(\alpha)$ on the space $(C_t, \mathcal{B}_t)$, i.e., P_α^ε is the probability measure on $(C_t, \mathcal{B}_t)$ induced by the process $X^\varepsilon(\alpha)$. Let P^w be a Wiener measure on $(C_t, \mathcal{B}_t)$. Denote by $X = (X_s)_{0\le s\le t}$ the coordinate process on $(C_t, \mathcal{B}_t)$, i.e., $X_s(x) = x_s$, $x \in C_t$.

The conditions **L** guarantee that for each $\alpha \in \mathcal{A}$, the measures P_α^ε and P^w are equivalent ($P_\alpha^\varepsilon \sim P^w$), and if we denote by $z_s^{\alpha,\varepsilon} = \frac{dP_\alpha^\varepsilon}{dP^w}|\mathcal{B}_s$ the density process (likelihood ratio process), then

$$z_s^{\alpha,\varepsilon}(X) = \mathcal{E}_s(a_\varepsilon(\alpha)\cdot X) := \exp\left\{\int_0^s a_\varepsilon(u, X; \alpha)\,dX_u - \frac{1}{2}\int_0^s a_\varepsilon^2(u, X; \alpha)\,du\right\}.$$

Introduce the class Ψ of $\mathbb{R}^m$-valued nonanticipative functionals ψ, $\psi : [0, t] \times C_t \times \mathcal{A} \to \mathbb{R}^m$ such that for each $\alpha \in \mathcal{A}$ and $\varepsilon > 0$,

$$(1)\quad E_\alpha^\varepsilon \int_0^t |\psi(s, X; \alpha)|^2 ds < \infty, \tag{2.4}$$

$$(2)\quad \int_0^t |\psi(s, Y^0(\alpha); \alpha)|^2 ds < \infty, \tag{2.5}$$

(3) uniformly in α on each compact set $K \subset \mathcal{A}$,

$$P_\alpha^\varepsilon - \lim_{\varepsilon\to 0} \int_0^t |\psi(s, \varepsilon X; \alpha) - \psi(s, Y^0(\alpha); \alpha)|^2 ds = 0, \tag{2.6}$$

where $|\cdot|$ is an Euclidean norm in $\mathbb{R}^m$ and $P_\alpha^\varepsilon - \lim_{\varepsilon\to 0} \zeta_\varepsilon = \zeta$ denotes the convergence $P_\alpha^\varepsilon\{|\zeta_\varepsilon - \zeta| > \rho\} \to 0$ as $\varepsilon \to 0$ for all ρ, $\rho > 0$.

Assume that for each $s \in [0, t]$ and $x \in C_t$, the functional $a(s, x; a)$ is differentiable in α and the gradient $\dot{a} = \left(\frac{\partial}{\partial\alpha_1} a, \dots, \frac{\partial}{\partial\alpha_m} a\right)'$ belongs to Ψ ($\dot{a} \in \Psi$), where the prime denotes the transposition.

Then the Fisher information process

$$I_s^\varepsilon(X; \alpha) := \int_0^s \dot{a}_\varepsilon(u, X; \alpha)[\dot{a}_\varepsilon(u, X; \alpha)]' du, \quad 0 \le s \le t,$$

is well defined, and, moreover, uniformly in α on each compact set,

$$P_\alpha^\varepsilon - \lim_{\varepsilon\to 0} \varepsilon^2 I_t^\varepsilon(\alpha) = I^0(\alpha), \tag{2.7}$$

where

$$I^0(\alpha) := \int_0^t \dot{a}(s, Y^0(\alpha); \alpha)[\dot{a}(s, Y^0(\alpha); \alpha]' ds.$$

For each $\psi \in \Psi$, introduce the functional $\psi_\varepsilon(s, x; \alpha) := \frac{1}{\varepsilon}\psi(s, \varepsilon x; \alpha)$ and the matrices $\Gamma_\varepsilon^\psi(\alpha)$ and $\gamma_\varepsilon^\psi \alpha$ as follows:

$$\Gamma_\varepsilon^\psi(X; \alpha) := \int_0^t \psi_\varepsilon(s, X; \alpha)[\psi_\varepsilon(s, X; \alpha)]' ds, \tag{2.8}$$

$$\gamma_\varepsilon^\psi(X; \alpha) := \int_0^t \psi_\varepsilon(s, X; \alpha)[\dot{a}_\varepsilon(s, X; \alpha)]' ds. \tag{2.9}$$

Then from (2.6), it follows that uniformly in α on each compact set,

$$P^{\varepsilon}_{\alpha} - \lim_{\varepsilon\to 0} \varepsilon^2 \Gamma^{\psi}_{\varepsilon}(\alpha) = \Gamma^{\psi}_{0}(\alpha), \tag{2.10}$$

$$P^{\varepsilon}_{\alpha} - \lim_{\varepsilon\to 0} \varepsilon^2 \gamma^{\psi}_{\varepsilon}(\alpha) = \gamma^{\psi}_{0}(\alpha), \tag{2.11}$$

where the matrices $\Gamma^{\psi}_0(\alpha)$ and $\gamma^{\psi}_0(\alpha)$ are defined as follows:

$$\Gamma^{\psi}_0(\alpha) := \int_0^t \psi(s, Y^0(\alpha); \alpha)[\psi(s, Y^0(\alpha); \alpha)]' ds, \tag{2.12}$$

$$\gamma^{\psi}_0(\alpha) := \int_0^t \psi(s, Y^0(\alpha); \alpha)[\dot{a}(s, Y^0(\alpha); \alpha)]' ds. \tag{2.13}$$

Note that by virtue of (2.4), (2.5), and $\dot{a} \in \Psi$, the matrices given by (2.8), (2.9), (2.12), and (2.13) are well defined.

Denote by Ψ_0 the subset of Ψ such that for each $\psi \in \Psi_0$ and $\alpha \in \mathcal{A}$, $\operatorname{rank} \Gamma^{\psi}_0(\alpha) = m$ and $\operatorname{rank} \gamma^{\psi}_0(\alpha) = m$.

Assume that $\dot{a} \in \Psi_0$. For each $\psi \in \Psi_0$, define the P^{ε}_{α}-square integrable martingale $L^{\psi,\varepsilon}(\alpha) = (L^{\psi,\varepsilon}_s(\alpha))_{0\le s\le t}$ as follows:

$$L^{\psi,\varepsilon}_s(X; \alpha) = \int_0^s \psi_{\varepsilon}(u, X; \alpha)(dX_u - \alpha_{\varepsilon}(u, X; \alpha)\, du). \tag{2.14}$$

Now we give a definition of CULAN M-estimates.

Definition 2.1. An estimate $(\alpha^{\psi,\varepsilon}_t)_{\varepsilon>0} = (\alpha^{\psi,\varepsilon}_{1,t}, \dots, \alpha^{\psi,\varepsilon}_{m,t})'_{\varepsilon>0}$, $\psi \in \Psi_0$, is said to be consistent uniformly lineal asymptotically normal (CULAN) if it admits the expansion

$$\alpha^{\psi,\varepsilon}_t = \alpha + [\gamma^{\psi}_0(\alpha)]^{-1} \varepsilon^2 L^{\psi,\varepsilon}_t(\alpha) + r_{\psi,\varepsilon}(\alpha), \tag{2.15}$$

where uniformly in α on each compact set,

$$P^{\varepsilon}_{\alpha} - \lim_{\varepsilon\to 0} \varepsilon^{-1} r_{\psi,\varepsilon}(\alpha) = 0. \tag{2.16}$$

It is well known (see [19]) that under the above conditions, uniformly in α on each compact set,

$$\mathcal{L}\{\varepsilon^{-1}(\alpha^{\psi,\varepsilon}_t - \alpha) \mid P^{\varepsilon}_{\alpha}\} \xrightarrow{w} N(0, V(\psi; \alpha)),$$

with

$$V(\psi; \alpha) := [\gamma^{\psi}_0(\alpha)]^{-1} \Gamma^{\psi}_0(\alpha)([\gamma^{\psi}_0(\alpha)]^{-1})', \tag{2.17}$$

where $\mathcal{L}(\zeta|P)$ denotes the distribution of the random vector ζ calculated under the measure P, the symbol "$\xrightarrow{w}$" denotes the weak convergence of measures, and $N(0, V(\psi; \alpha))$ is the distribution of Gaussian vectors with zero mean and covariance matrix $V(\psi; \alpha)$.

Remark 2.1. In context of diffusion type processes, the M-estimate $(\alpha^{\psi,\varepsilon}_t)_{\varepsilon>0}$ is defined as a solution of the stochastic equation

$$L^{\psi,\varepsilon}_t(X; \alpha) = 0,$$

where $L^{\psi,\varepsilon}_t(X; \alpha)$ is defined by (2.14) and $\psi \in \Psi_0$.

The asymptotic theory of M-estimates for general statistical models with filtration is developed in [7]. Precisely, the problem of existence and global behavior of solutions is studied. In particular, the conditions of regularity and ergodicity type are established under which M-estimates have the CULAN property.

For our model, in the case where $\mathcal{A} = \mathbb{R}^m$, the sufficient conditions for CULAN property take the form:

(1) for all s, $0 \le s \le t$, and $x \in C_t$, the functionals $\psi(s, x; \alpha)$ and $\dot{a}(s, x; \alpha)$ are twice continuously differentiable in α with bounded derivatives satisfying the functional Lipshitz conditions with constants that do not depend on α;

(2) the equation (with respect to y)

$$\Delta(\alpha, y) := \int_0^t \psi(s, Y^0(\alpha); y)(a(s, Y^0(\alpha); \alpha) - a(s, Y^0(\alpha); y))\, ds = 0,$$

has a unique solution $y = \alpha$.

The MLE is a special case of M-estimates when $\psi = \dot{a}$.

Remark 2.2. According to (2.7), the asymptotic covariance matrix of the MLE $(\widehat{\alpha}_t^\varepsilon)_{\varepsilon>0}$ is $[I_0(\alpha)]^{-1}$. Using the usual technique, one can show that for each $\alpha \in \mathcal{A}$ and $\psi \in \Psi_0$, $I_0^{-1}(\alpha) \le V(\psi, \alpha)$ (see (2.17)), where for two symmetric matrices B and C the relation $B \le C$ means that the matrix $C - B$ is positive semidefinite.

Thus, the MLE has the minimal covariance matrix among all M-estimates.

2.2. Shrinking contamination neighborhoods. In this section, we give the notion of contamination of the basic model (2.3) described in terms of shrinking neighborhoods of basic measures $\{P_\alpha^\varepsilon$, $\alpha \in \mathcal{A}$, $\varepsilon > 0\}$, which is an analog of the Huber gross error model (see, e.g., [12] and also Remark 2.3 below).

Let $\mathcal{H}$ be a family of bounded nonanticipative functionals $h : [0, t] \times C_t \times \mathcal{A} \to \mathbb{R}^1$ such that for all $s \in [0, t]$ and $\alpha \in \mathcal{A}$, the functional $h(s, x; \alpha)$ is continuous at the point $x_0 = Y^0(\alpha)$.

For each $h \in \mathcal{H}$, $\alpha \in \mathcal{A}$, and $\varepsilon > 0$, let $P_\alpha^{\varepsilon,h}$ be a measure on $(C_t, \mathcal{B}_t)$ such that

$$(1) \qquad P_\alpha^{\varepsilon,h} \sim P_\alpha^\varepsilon,$$

$$(2) \qquad \frac{dP_\alpha^{\varepsilon,h}}{dP_\alpha^\varepsilon} = \mathcal{E}_t(\varepsilon N_\alpha^{\varepsilon,h}), \tag{2.18}$$

where

$$(3) \qquad N_{\alpha,s}^{\varepsilon,h} := \int_0^s h_s(u, X; \alpha)(dX_u - a_\varepsilon(u, X; \alpha)\, du), \tag{2.19}$$

with $h_\varepsilon(s, x; \alpha) := \frac{1}{\varepsilon} h(s, \varepsilon x; \alpha)$, $0 \le s \le t$, $x \in C_t$.

Denote by $\mathbf{P}_\alpha^{\varepsilon,\mathcal{H}}$ the class of measures $P_\alpha^{\varepsilon,h}$, $h \in \mathcal{H}$, i.e.,

$$\mathbf{P}_\alpha^{\varepsilon,\mathcal{H}} = \{P_\alpha^{\varepsilon,h};\ h \in \mathcal{H}\}.$$

We call $(\mathbf{P}_\alpha^{\varepsilon,\mathcal{H}})_{\varepsilon>0}$ shrinking contamination neighborhoods of the basic measures $(P_\alpha^\varepsilon)_{\varepsilon>0}$, and the element $(P_\alpha^{\varepsilon,h})_{\varepsilon>0}$ of these neighborhoods is called the alternative measure (or simply alternative).

Obviously, for each $h \in \mathcal{H}$ and $\alpha \in \mathcal{A}$, the process $N_\alpha^{\varepsilon,h} = (N_{\alpha,s}^{\varepsilon,h})_{0\le s\le t}$ defined by (2.19) is a P_α^ε-square integrable martingale. Since under measure P_α^ε, the process $\overline{w} = (\overline{w}_s)_{0\le s\le t}$ defined as

$$\overline{w}_s := X_s - \int_0^s a_\varepsilon(u, X; \alpha)\, du, \quad 0 \le s \le t,$$

is a Wiener process, by virtue of the Girsanov theorem, the process $\widetilde{w} := \overline{w} + \langle \overline{w}, \varepsilon N_\alpha^{\varepsilon,h} \rangle$ is a Wiener process under changed measure $P_\alpha^{\varepsilon,h}$. But, by definition,

$$\widetilde{w}_s = X_s - \int_0^s (a_\varepsilon(u, X; \alpha) + \varepsilon h_\varepsilon(u, X; \alpha))\, du,$$

and hence one can conclude that $P_\alpha^{\varepsilon,h}$ is a weak solution of the SDE

$$dX_s = (a_\varepsilon(s, X; \alpha) + \varepsilon h_\varepsilon(s, X; \alpha))\, ds + dw_s, \quad X_0 = 0.$$

This SDE can be viewed as a "small" perturbation of the basic model (2.3).

Remark 2.3. 1. In the case of i.i.d. observations $X_1, X_2, \dots, X_n$, $n \ge 1$, the Huber gross error model in shrinking setting is defined as follows:

$$f^{n,h}(x; \alpha) := (1 - \varepsilon_n) f(x; \alpha) + \varepsilon_n h(x; \alpha),$$

where $f(x; \alpha)$ is a basic (core) density of distribution of r.v. X_i (with respect to some dominating measure μ), $h(x; \alpha)$ is a contaminating density, $f^{n,h}(x; \alpha)$ is a contaminated density, and $\varepsilon_n = O(n^{-1/2})$. If we denote by P_α^n and $P_\alpha^{n,h}$ the measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ generated by $f(x; \alpha)$ and $f^{n,h}(x; \alpha)$, respectively, then

$$\frac{dP_\alpha^{n,h}}{dP_\alpha^n} = \prod_{i=1}^n \frac{f^{n,h}(X_i; \alpha)}{f(X_i; \alpha)} = \prod_{i=1}^n (1 + \varepsilon_n H(X_i; \alpha)) = \mathcal{E}_n(\varepsilon_n \cdot N_\alpha^{n,h}),$$

where $H = \frac{h-f}{f}$, $N_\alpha^{n,h} = (N_{\alpha,m}^{n,h})_{1 \le m \le n}$, $N_{\alpha,m}^{n,h} = \sum_{i=1}^m H(X_i; \alpha)$, $N_\alpha^{n,h}$ is a P_α^n-martingale, and

$$\mathcal{E}_n(\varepsilon_n N_\alpha^{n,h}) = \prod_{i=1}^n (1 + \varepsilon_n \Delta N_{\alpha,i}^{n,h})$$

is the Dolean exponential in the discrete time case.

Thus,

$$\frac{dP_\alpha^{n,h}}{dP_\alpha^n} = \mathcal{E}(\varepsilon_n \cdot N_\alpha^{n,n}), \tag{2.20}$$

and relation (2.18) is a direct analog of (2.20).

2. The concept of shrinking contamination neighborhoods expressed in the form of (2.18) was proposed in [20] for a more general situation concerning the contamination areas for semimartingale statistical models with filtration.

Note here that the degree of the small parameter ε is crucial. We cannot consider the perturbation of the measure with a different power of ε if we wish to obtain nontrivial results.

In the remainder of this section, we study the asymptotic properties of CULAN estimates under alternatives.

For this purpose, we first consider the problem of contiguity of the measures $(P_\alpha^{\varepsilon,h})_{\varepsilon>0}$ to $(P_\alpha^\varepsilon)_{\varepsilon>0}$.

Let $(\varepsilon_n)_{n\ge1}$, $\varepsilon_n \downarrow 0$, and $(\alpha_n)_{n\ge1}$, where $\alpha_n \in K$, $K \subset \mathcal{A}$, is a compact set, be arbitrary sequences.

Proposition 2.1. *For each $h \in \mathcal{H}$, the sequence of measures $(P_{\alpha_n}^{\varepsilon_n,h})$ is contiguous to sequence of measures $(P_{\alpha_n}^{\varepsilon_n})$, i.e.,*

$$(P_{\alpha_n}^{\varepsilon_n,h}) \lhd (P_{\alpha_n}^{\varepsilon_n}).$$

Proof. From the predictable criteria of contiguity (see, e.g., [15]), it follows that we need to verify the relation

$$\lim_{N\to\infty} \limsup_{n\to\infty} P_{\alpha_n}^{\varepsilon_n,h} \left\{ h_t^n\left(\frac{1}{2}\right) > N \right\} = 0, \tag{2.21}$$

where $h^n(\frac{1}{2}) = (h^n_s(\frac{1}{2}))_{0\le s\le t}$ is the Hellinger process of order $\frac{1}{2}$.

By the definition of Hellinger process (see, e.g., [15]), we have

$$h^n_t\left(\frac{1}{2}\right) = h^n_t\left(\frac{1}{2}, P^{\varepsilon_n,h}_{\alpha_n}, P^{\varepsilon_n}_{\alpha_n}\right) = \frac{1}{8}\int_0^t [h(s,\varepsilon_n X;\alpha_n)]^2\,ds,$$

and since $h \in \mathcal{H}$ and hence is bounded, $h^n_t(\frac{1}{2})$ is also bounded, which provides (2.21). □

Proposition 2.2. *For each estimate $(\alpha^{\varepsilon,\psi}_t)_{\varepsilon>0}$ with $\psi \in \Psi_0$ and each alternative*

$$(P^{\varepsilon,h}_\alpha)_{\varepsilon>0} \in (\mathbf{P}^{\varepsilon,\mathcal{H}}_\alpha)_{\varepsilon>0},$$

the following relation holds:

$$\mathcal{L}\left\{\varepsilon^{-1}(\alpha^{\psi,\varepsilon}_t - \alpha) \mid P^{\varepsilon,h}_\alpha\right\} \xrightarrow{w} N\left([\gamma^\psi_0(\alpha)]^{-1} b(\psi,h;\alpha), V(\psi,\alpha)\right),$$

where

$$b(\psi,h;\alpha) := \int_0^t \psi(s,Y^0(\alpha);\alpha)h(s,Y^0(\alpha);\alpha)\,ds.$$

Proof. Proposition 2.1, together with (2.16), provides that uniformly in α, on each compact set

$$P^{\varepsilon,h}_\alpha - \lim_{\varepsilon\to 0}\varepsilon^{-1} r_{\psi,\varepsilon}(\alpha) = 0,$$

and, therefore, we need to establish the limit distribution of the random vector $[\gamma^\psi_0(\alpha)]^{-1}\varepsilon L^{\psi,\varepsilon}_t$ under the measures $(P^{\varepsilon,h}_\alpha)_{\varepsilon>0}$.

By the Girsanov theorem, the process $L^{\psi,\varepsilon}(\alpha) = (L^{\psi,\varepsilon}_s(\alpha))_{0\le s\le t}$ is a semimartingale with the canonical decomposition

$$L^{\psi,\varepsilon}_s(\alpha) = \widetilde{L}^{\psi,\varepsilon}_s(\alpha) + b_{\varepsilon,s}(\psi,h;\alpha), \quad 0 \le s \le t, \tag{2.22}$$

where $\widetilde{L}^{\psi,\varepsilon}(\alpha) = (\widetilde{L}^{\psi,\varepsilon}_s(\alpha))_{0\le s\le t}$ is a $P^{\varepsilon,h}_\alpha$-square integrable martingales defined as follows:

$$\widetilde{L}^{\psi,\varepsilon}_s(X;\alpha) := \int_0^s \psi_\varepsilon(u,X;\alpha)\ (dX_u - (a_\varepsilon(u,X;\alpha) + \varepsilon h_\varepsilon(u,X;\alpha))\ du,$$

$$b_{\varepsilon,s}(\psi,h;\alpha) := \varepsilon\int_0^s \psi_\varepsilon(u,X;\alpha)h_\varepsilon(u,X;\alpha)\,du.$$

But $\langle\widetilde{L}^{\psi,\varepsilon}(\alpha)\rangle_t = \Gamma^\psi_\varepsilon(\alpha)$, where $\Gamma^\psi_\varepsilon(\alpha)$ is defined by (2.8). On the other hand, from Proposition 2.1 and (2.10), it follows that

$$P^{\varepsilon,h}_\alpha - \lim_{\varepsilon\to 0}\langle\varepsilon\widetilde{L}^{\psi,\varepsilon}(\alpha)\rangle_t = P^{\varepsilon,h}_\alpha - \lim_{\varepsilon\to 0}\varepsilon^2\Gamma^\psi_\varepsilon(\alpha) = P^\varepsilon_\alpha - \lim_{\varepsilon\to 0}\varepsilon^2\Gamma^\psi_\varepsilon(\alpha) = \Gamma^\psi_0(\alpha)$$

uniformly in α on each compact set, and hence

$$\mathcal{L}\left\{[\gamma^\psi_0(\alpha)]^{-1}\varepsilon\widetilde{L}^{\psi,\varepsilon}_t(\alpha) \mid P^{\varepsilon,h}_\alpha\right\} \xrightarrow{w} N(0, V(\psi;\alpha)). \tag{2.23}$$

Finally, relation (2.23), together with (2.22) and the relation

$$P^{\varepsilon,h}_\theta - \lim_{\varepsilon\to 0}\varepsilon b_{\varepsilon,t}(\psi,h;\alpha) = \int_0^t \psi(s,Y^0(\alpha);\alpha)h(s,Y^0(\alpha);\alpha)\,ds = b(\psi,h;\alpha),$$

provides the desirable results. □

2.3. Optimization criteria. Construction of optimal B-robust estimates. In this section, we state and solve the optimization problem, which results in constructing the optimal B-robust estimate.

Initially, it should be noted that the bias vector $\widetilde{b}(\psi,h;\alpha) := [\gamma_0^{\psi}(\alpha)]^{-1}b(\psi,h;\alpha)$ can be viewed as the influence functional of the estimate $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$ with respect to the alternative $(P_\alpha^{\psi,h})_{\varepsilon>0}$.

Indeed, expansion (2.15), together with (2.22) and (2.23), allows us to conclude that

$$\mathcal{L}\left\{\varepsilon^{-1}(\alpha_t^{\psi,\varepsilon} - \alpha - \varepsilon^2[\gamma_0^{\psi}(\alpha)]^{-1}b_\varepsilon(\psi,h;\alpha)) \mid P_\alpha^{\varepsilon,h}\right\} \xrightarrow{w} N(0,V(\psi,\alpha)),$$

and hence the expression

$$\alpha + \varepsilon^2[\gamma_0^{\psi}(\alpha)]^{-1}b_\varepsilon(\psi,h;\alpha) - \alpha = \varepsilon^2[\gamma_0^{\psi}(\alpha)]^{-1}b_\varepsilon(\psi,h;\alpha),$$

plays the role of bias on the "fixed step ε," and it seems natural to interpret the limit

$$P_\alpha^{\varepsilon,h} - \lim_{\varepsilon\to 0}\frac{\alpha + \varepsilon^2[\gamma_0^{\psi}(\alpha)]^{-1}b_\varepsilon(\psi,h;\alpha) - \alpha}{\varepsilon} = [\gamma_0^{\psi}(\alpha)]^{-1}b(\psi,h;\alpha),$$

as the influence functional.

For each estimate $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$, $\psi\in\Psi_0$, define the risk functional with respect to alternative $(P_\alpha^{\varepsilon,h})_{\varepsilon>0}$, $h\in\mathcal{H}$, as follows:

$$D(\psi,h;\alpha) = \lim_{a\to\infty}\lim_{\varepsilon\to 0}E_\alpha^{\varepsilon,h}\left((\varepsilon^{-2}|\alpha_t^{\psi\varepsilon} - \alpha|^2)\wedge a\right),$$

where $x\wedge\alpha = \min(x,a)$, $a>0$, and $E_\alpha^{\varepsilon,h}$ is an expectation with respect to the measure $P_\alpha^{\varepsilon,h}$.

Using Proposition 2.2, it is easy to verify that

$$D(\psi,h;\alpha) = |\widetilde{b}(\psi,h;\alpha)|^2 + \operatorname{tr}V(\psi,\alpha),$$

where $\operatorname{tr}A$ denotes the trace of the matrix A.

By Proposition 2.2,

$$\varepsilon^{-1}(\alpha_t^{\psi,\varepsilon} - \alpha) \xrightarrow{d} N\big(\widetilde{b}(\psi,h;\alpha),V(\psi;\alpha)\big),$$

where $\xrightarrow{d}$ denotes the convergence in distribution (by distribution $P_\alpha^{\varepsilon,h}$ in our case) and $N(\widetilde{b},V)$ is a Gaussian random vector with mean $\widetilde{b}$ and covariation matrix V.

But, if $\xi = (\xi_1,\dots,\xi_m)'$ is a Gaussian vector with parameters (μ,σ^2), then

$$E|\xi|^2 = \sum_{i=1}^m E\xi_i^2 = \sum_{i=1}^m (E\xi_i)^2 + \sum_{i=1}^m D\xi_i = |\mu|^2 + \operatorname{tr}\sigma^2,$$

as was required.

With each $\psi\in\Psi_0$, let us associate the function $\widetilde{\psi}$ as follows:

$$\widetilde{\psi}(s,x;\alpha) = [\gamma_0^{\psi}(\alpha)]^{-1}\psi(s,x;\alpha),\quad 0\le s\le t,\quad x\in C_t,\quad \alpha\in\mathcal{A}.$$

Then $\widetilde{\psi}\in\Psi_0$ and

$$\gamma_0^{\widetilde{\psi}}(\alpha) = \mathrm{Id},$$

where Id is the identity matrix,

$$V(\psi;\alpha) = V(\widetilde{\psi};\alpha) = \Gamma_0^{\widetilde{\psi}}(\alpha),\quad \widetilde{b}(\psi,h;\alpha) = \widetilde{b}(\widetilde{\psi},h;\alpha) = b(\widetilde{\psi},h;\alpha).$$

Therefore,

$$D(\psi,h;\alpha) = D(\widetilde{\psi},h;\alpha) = |b(\widetilde{\psi},h;\alpha)|^2 + \operatorname{tr}\Gamma_0^{\widetilde{\psi}}(\alpha). \tag{2.24}$$

Denote by $\mathcal{H}_r$ the set of functions $h\in\mathcal{H}$ such that for each $\alpha\in\mathcal{A}$,

$$\int_0^t |h(s,Y^0(\alpha);\alpha)|\,ds \le r,$$

where r, $r > 0$, is a constant.

Since, for each $r > 0$,

$$\sup_{h\in\mathcal{H}_r} |b(\widetilde{\psi}, h; \alpha)| \leq \mathrm{const}(r) \sup_{0\leq s\leq t} |\widetilde{\psi}(s, Y^0(\alpha); \alpha)|,$$

where the constant depends on r, we call the function $\widetilde{\psi}$ the influence function of estimate $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$, and the quantity

$$\gamma_\psi^*(\alpha) = \sup_{0\leq s\leq t} |\widetilde{\psi}(s, Y^0(\alpha); \alpha)|$$

is called the (unstandardized) gross error sensitivity at point α of the estimate $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$.

Define

$$\Psi_{0,c} = \Big\{ \psi \in \Psi_0 \;:\; \int_0^t \psi(s, Y^0(\alpha); \alpha)[\dot{a}(s, Y^0(\alpha); \alpha)]' ds = \mathrm{Id}, \tag{2.25}$$

$$\gamma_\psi^*(\alpha) \leq c \Big\}, \tag{2.26}$$

where $c \in [0, \infty)$ is a generic constant.

Taking into account expression (2.24) for the risk functional, we come to the following optimization problem known in robust estimation theory as the Hampel optimization problem: minimize the trace of the asymptotic covariance matrix of the estimate $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$ over the class $\Psi_{0,c}$, i.e.,

$$\text{minimize} \quad \int_0^t \psi(s, Y^0(\alpha); \alpha)[\psi(s, Y^0(\alpha); \alpha)]' ds \tag{2.27}$$

under the conditions (2.25) and (2.26).

Define the Huber function $h_c(z)$, $z \in \mathbb{R}^m$, $c > 0$, as follows:

$$h_c(z) := z \min\left(1, \frac{c}{|z|}\right).$$

For an arbitrary nondegenerate matrix A, we denote $\psi_c^A = h_c(A\dot{a})$.

Theorem 2.1. *Assume that for given constant c there exists a nonsingular $(m \times m)$-matrix $A_c^*(\alpha)$, which solves the equation (with respect to the matrix A)*

$$\int_0^t \psi_c^A(s, Y^0(\alpha); \alpha)[\dot{a}(s, Y^0(\alpha); \alpha)]' ds = \mathrm{Id}. \tag{2.28}$$

Then the function $\psi_c^{A_c^*(\alpha)} = h_c(A_c^*(\alpha)\dot{a})$ solves the optimization problem (2.27).

Proof. We follow Hampel et al. [12]. Let A be an arbitrary $(m \times m)$-matrix. Since for each $\psi \in \Psi_{0,c}$,

$$\int \psi(\dot{a})' = \mathrm{Id}, \quad \int \dot{a}[\dot{a}]' = I^0(\alpha)$$

(see (2.7)), we have

$$\int (\psi - A\dot{a})(\psi - A\dot{a})' = \int \psi\psi' - A - A' + AI^0(\alpha)A'$$

(here and below, we use simple and obvious notation for integrals).

Therefore, since the trace is an additive functional, instead of minimizing $\mathrm{tr} \int \psi\psi'$, we can minimize

$$\mathrm{tr} \int (\psi - A\dot{a})(\psi - A\dot{a})' = \int |\psi - A\dot{a}|^2.$$

Note that for each z,

$$\arg\min_{|y|\le c} |z-y|^2 = h_c(z).$$

Indeed, it is obvious that minimizer y has the form $y=\beta z$, where β, $0\le\beta\le 1$, is constant. Then

$$\min_{|y|\le c} |z-y|^2 = \min_{\beta\le \frac{c}{|z|}} (1-\beta)^2|z|^2.$$

Thus, we need to find

$$\arg\min_{\beta\le \frac{c}{|z|}} (1-\beta)^2 = \min\Big(1, \frac{c}{|z|}\Big).$$

But the last relation is trivially satisfied. Hence the minimizer $y^* = z\min(1, \frac{c}{|z|})$, and

$$\arg\min_{|\psi|\le c} |\psi - A\dot{a}|^2 = h_c(A\dot{a}).$$

On the other hand,

$$|h_c(z)|^2 = |z|^2 I_{\{|z|\le c\}} + \frac{|z|^2}{|z|^2}\, c^2\, I_{(|z|\ge c)} \le c^2.$$

Hence

$$|h_c(z)| \le c \quad \text{for all } z$$

and, therefore, $h_c(A\dot{a})$ satisfies condition (2.26) for each A.

Now it is obvious that the function $h_c(A\dot{a})$ minimizes the expression under the sign of the integral, and hence the integral itself over all functions $\psi\in\Psi_0$ satisfying (2.26).

At the same time, condition (2.25), generally speaking, can be violated. But, since the matrix A is arbitrary, we can choose $A = A_c^*(\alpha)$ from (2.28) which, of course, guarantees the fulfillment of (2.25) for $\psi_c^* = \psi_c^{A_c^*(\alpha)}$. □

As was seen, the resulting optimal influence function ψ_c^* is defined along the process $Y^0(\alpha) = (Y_s^0(\alpha))_{0\le s\le t}$, which is a solution of Eq. (2.2).

But for constructing optimal estimate, we need a function $\psi_c^*(s,x;\alpha)$ defined on the whole space $[0,t]\times C_t\times\mathcal{A}$.

For this purpose, define $\psi_c^*(s,x;\alpha)$ as follows:

$$\psi_c^*(s,x;\alpha) = \psi_c^{A_c^*(\alpha)}(s,x;\alpha) = h_\varepsilon(A_c^*(\alpha)\dot{a}(s,x;\alpha)), \tag{2.29}$$

and, as usual, $\psi_{c,\varepsilon}^*(s,x;\alpha) = \frac{1}{\varepsilon}\psi_c^*(s,\varepsilon x;\alpha)$, $0\le s\le t$, $x\in C_t$, $\alpha\in\mathcal{A}$.

Definition 2.2. We say that $\psi_c^*(s,x;\alpha)$, $0\le s\le t$, $x\in C_t$, $\alpha\in\mathcal{A}$, is the influence function of the optimal B-robust estimate $(\alpha_t^{*,\varepsilon})_{\varepsilon>0} = (\alpha_t^{\psi_c^*,\varepsilon})_{\varepsilon>0}$ over the class of CULAN estimates $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$, $\psi\in\Psi_{0,c}$, if the matrix $A^*(\alpha)$ is differentiable in α.

From (2.9), (2.11), (2.28), and (2.29), it directly follows that

$$\gamma_0^{\psi_c^*}(\alpha) = P_\alpha^\varepsilon - \lim_{\varepsilon to 0} \varepsilon^2\gamma_\varepsilon^{\psi_c^*}(\alpha) = \int_0^t \psi_c^*(s,Y^0(\alpha);\alpha)(\dot{a}(s,Y^0(\alpha);\alpha))' ds = \mathrm{Id}\,.$$

Moreover, for each alternative $(P_\alpha^{\varepsilon,h})_{\varepsilon>0}$, $h\in\mathcal{H}$, according to Proposition 2.2, we have

$$\mathcal{L}\Big\{\varepsilon^{-1}(\alpha_t^{*,\varepsilon}-\alpha) \mid P_\alpha^{\varepsilon,h}\Big\} \xrightarrow{w} N(b(\psi_c^*,h;\alpha), V(\psi_c^*;\alpha)) \quad \text{as } \varepsilon\to 0,$$

where

$$b(\psi_c^*,h;\alpha) = \int_0^t \psi_c^*(s,Y^0(\alpha);\alpha)h(s,Y^0(\alpha);\alpha)\,ds,$$

and $V(\psi_c^*;\alpha) = \Gamma_0^{\psi_c^*}(\alpha)$.

Hence the risk functional for estimate $(\alpha_t^{*,\varepsilon})_{\varepsilon>0}$ is

$$D(\psi_c^*, h; \alpha) = |b(\psi_c^*, h; \alpha)|^2 + \operatorname{tr} \Gamma_0^{\psi_c^*}, \quad h \in \mathcal{H},$$

and the (unstandardized) gross error sensitivity of $(\alpha_t^{*,\varepsilon})_{\varepsilon>0}$ is

$$\gamma_{\psi_c^*}(\alpha) = \sup_{0\le s\le t} |\psi_c^*(s, Y^0(\alpha); \alpha)| \le c.$$

From the above arguments, we may conclude that $(\alpha_t^{*,\varepsilon})_{\varepsilon>0}$ is the optimal B-robust estimate over the class of estimates $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$, $\psi \in \Psi_{0,c}$ in the following sense: the trace of the asymptotic covariance matrix of $(\alpha_t^{*,\varepsilon})_{\varepsilon>0}$ is minimal among all estimates $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$ with sensitivity bounded by constant gross error, i.e.,

$$\Gamma_0^{\psi_c^*}(\alpha) \le \Gamma_0^{\psi}(\alpha) \quad \text{for all} \quad \psi \in \Psi_{0,c}.$$

Note that for each estimate $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$ and alternatives $(P_\alpha^{\varepsilon,h})_{\varepsilon>0}$, $h \in \mathcal{H}$, the influence functional is bounded by $\operatorname{const}(r) \cdot c$. Indeed, for $\psi \in \Psi_{0,c}$, we have

$$\sup_{h\in\mathcal{H}_r} |b(\psi, h; \alpha)| \le \operatorname{const}(r) \cdot c := C(r, c),$$

and since from (2.24)

$$\inf_{\psi\in\Psi_{0,c}} \sup_{h\in\mathcal{H}_r} D(\psi, h; \alpha) \le C^2(r, c) + \operatorname{tr} \Gamma_0^{\psi_c^*}(\alpha),$$

we can choose the "optimal level" of truncation minimizing the expression

$$C^2(r, c) + \operatorname{tr} \Gamma_0^{\psi_c^*}(\alpha)$$

over all constants c for which Eq. (2.28) has a solution $A_c^*(\alpha)$. This can be done by using the numerical methods.

For the problem of existence and uniqueness of solution of Eq. (2.28), we refer the reader to [34].

In the case of a one-dimensional parameter α (i.e., $m = 1$), the optimal level c^* of truncation is given as a unique solution of the following equation (see [20, 21]):

$$r^2c^2 = \int_0^t [\dot a(s, Y^0(\alpha); \alpha)]_{-c}^{c} \dot a(s, Y^0(\alpha); \alpha)\, ds - \int_0^t ([\dot a(s, Y^0(\alpha); \alpha)]_{-c}^{c})^2\, ds,$$

where $[x]_a^b = (x \wedge b) \vee a$ and the resulting function

$$\psi^*(s, x; \alpha) = [\dot a(s, x; \alpha)]_{-c^*}^{c^*}, \quad 0 \le s \le t, \quad x \in C_t,$$

is $(\Psi_0, \mathcal{H}_r)$ optimal in the following minimax sense:

$$\sup_{h\in\mathcal{H}_r} D(\psi^*, h; \alpha) = \inf_{\psi\in\Psi} \sup_{h\in\mathcal{H}_r} D(\psi, h; \alpha).$$

3. Optimal Mean-Variance Robust Hedging

3.1. A financial market model. Let $(\Omega, \mathcal{F}, F = (\mathcal{F}_t)_{0\le t\le T}, P)$ be a filtered probability space with filtration F satisfying the usual conditions, where $T \in (0, \infty]$ is a fixed time horizon. Assume that $\mathcal{F}_0$ is trivial and $\mathcal{F}_T = \mathcal{F}$.

There exist $d + 1$, $d \ge 1$, primitive assets: one bound whose price process is assumed to be 1 at all times, and d risky assets (stocks) whose $\mathbb{R}^d$-valued price process $X = (X_t)_{0\le t\le T}$ is a continuous semimartingale given by the relation

$$dX_t = \operatorname{diag}(X_t)\, dR_t, \quad X_0 > 0, \tag{3.1}$$

where diag(X) denotes the diagonal $(d \times d)$-matrix with diagonal entries $X^1, \dots, X^d$, and the yield process $R = (R_t)_{0 \le t \le T}$ is an $\mathbb{R}^d$-valued continuous semimartingale satisfying the stricture condition (SC). That is (see [37]),

$$dR_t = d\langle \widetilde{M} \rangle_t \lambda_t + d\widetilde{M}_t, \quad R_0 = 0, \tag{3.2}$$

where $\widetilde{M} = (\widetilde{M}_t)_{0 \le t \le T}$ is an $\mathbb{R}^d$-valued continuous martingale, $\widetilde{M} \in \mathcal{M}^2_{0,\mathrm{loc}}(P)$, $\lambda = (\lambda_t)_{0 \le t \le T}$ is an F-predictable $\mathbb{R}^d$-valued process, and the mean-variance tradeoff (MVT) process $\widetilde{\mathcal{K}} = (\widetilde{\mathcal{K}}_t)_{0 \le t \le T}$ of the process R,

$$\widetilde{\mathcal{K}}_t := \int_0^t \lambda_s' d\langle \widetilde{M} \rangle_s \lambda_s = \langle \lambda' \cdot \widetilde{M} \rangle_t < \infty \quad P\text{-}a.s., \quad t \in [0, T]. \tag{3.3}$$

Remark 3.1. Recall that all vectors are assumed to be column vectors.

Suppose that the martingale $\widetilde{M}$ has the form

$$\widetilde{M} = \sigma \cdot M, \tag{3.4}$$

where $M = (M_t)_{0 \le t \le T}$ is an $\mathbb{R}^d$-valued continuous martingale, $M \in \mathcal{M}^2_{0,\mathrm{loc}}(P)$ with $d\langle M^i, M^j \rangle_t = I^{d\times d}_{i_j} dC_t$, $I^{d \times d}$ is the identity matrix, and $C = (C_t)_{0 \le t \le T}$ is a continuous increasing bounded process with $C_0 = 0$.

Further, let $\sigma = (\sigma_t)_{0 \le t \le T}$ be a $(d \times d)$-matrix valued, F-predictable process with rank$(\sigma_t) = d$ for any t, P-a.s., the process $(\sigma_t^{-1})_{0 \le t \le T}$ is locally bounded, and

$$\int_0^T \sigma_t \, d\langle M \rangle_t \sigma_t' < \infty \quad P\text{-}a.s. \tag{3.5}$$

Assume that the following condition is satisfied.

There exists a fixed $\mathbb{R}^d$-valued, F-predictable process $k = (k_t)_{0 \le t \le T}$ such that

$$\lambda = \lambda(\sigma) = (\sigma')^{-1} k. \tag{3.6}$$

In this case, from (3.2), we obtain

$$dR_t = d\langle \widetilde{M} \rangle_t \lambda_t + d\widetilde{M}_t = \sigma_t d\langle M \rangle_t \sigma_t' (\sigma_t')^{-1} k_t + \sigma_t dM_t = \sigma_t (d\langle M \rangle_t k_t + dM_t) \tag{3.7}$$

and

$$\widetilde{\mathcal{K}}_t = \int_0^t \lambda_s' d\langle \widetilde{M} \rangle_s \lambda_s = \int_0^t k_t'((\sigma_t')^{-1})' \sigma_t d\langle \widetilde{M} \rangle_t \sigma_t' (\sigma_t')^{-1} k_t = \int_0^t k_t' d\langle M \rangle_t k_t = \langle k \cdot M \rangle_t := \mathcal{K}_t.$$

From (3.3), we have

$$\mathcal{K}_t < \infty \quad P\text{-}a.s. \quad \text{for all } t \in [0, T]. \tag{3.8}$$

Thus, if we introduce the process $M^0 = (M^0_t)_{0 \le t \le T}$ by the relation

$$dM^0_t = d\langle M \rangle_t k_t + dM_t, \quad M^0_0 = 0, \tag{3.9}$$

then the MVT process $\mathcal{K} = (\mathcal{K}_t)_{0 \le t \le T}$ of $\mathbb{R}^d$-valued semimartingale M^0 is finite, and hence M^0 satisfies SC.

Finally, the scheme (3.1), (3.2), (3.4), (3.6), and (3.9) can be rewritten in the following form:

$$\begin{aligned} dX_t &= \mathrm{diag}(X_t)\, dR_t, & X_0 &> 0, \\ dR_t &= \sigma_t \, dM^0_t, & R_0 &= 0, \\ dM^0_t &= d\langle M \rangle_t k_t + dM_t, & M_0 &= 0, \end{aligned} \tag{3.10}$$

where σ and k satisfy (3.5) and (3.8), respectively.

This is our financial market model.

3.2. Characterization of variance-optimal ELMM (equivalent local martingale measure). A key role in mean-variance hedging is played by variance-optimal ELMM (see, e.g., [11, 33]). Here, we collect some facts characterizing this measure.

We start from the remark that the sets ELMMs for processes X, R, and M^0 of the form (3.10) coincide. Hence we can consider the simplest process M^0.

Introduce the notation

$$\mathcal{M}_2^e := \left\{ Q \sim P \,:\, \frac{dQ}{dP} \in L^2(P), \quad M^0 \text{ is a } Q\text{-local martingale} \right\}$$

and assume that

(c1) $\mathcal{M}_2^e \neq \varnothing$.

A solution $\widetilde{P}$ of the optimization problem

$$E\mathcal{E}_T^2(\mathcal{M}^Q) \to \inf_{Q \in \mathcal{M}_2^e}$$

is called a variance-optimal ELMM. Here,

$$\frac{dQ}{dP}\Big|_{\mathcal{F}_T} = \mathcal{E}_T(M^Q),$$

and $(\mathcal{E}_t(M^Q))_{0 \le t \le T}$ is the Dolean exponential of martingale M^Q.

It is well known (see, e.g., [37, 38]) that under condition (c1), a variance-optimal ELMM $\widetilde{P}$ exists.

Denote

$$\widetilde{z}_T := \frac{d\widetilde{P}}{dP}\Big|_{\mathcal{F}_T},$$

and introduce the RCLL process $\widetilde{z} = (\widetilde{z}_t)_{0 \le t \le T}$ by the relation

$$\widetilde{z}_t = E^{\widetilde{P}}(\widetilde{z}_T / \mathcal{F}_T), \quad 0 \le t \le T.$$

Then, by [37, 38],

$$\widetilde{z}_T = \widetilde{z}_0 + \int_0^T \zeta_t' \, dM_t^0, \tag{3.11}$$

where $\zeta = (\zeta_t)_{0 \le t \le T}$ is the $\mathbb{R}^d$-valued F-predictable process with

$$\int_0^T \zeta_t' \, d\langle M \rangle_t \zeta_t < \infty$$

and the process $\left(\int_0^t \zeta_s' dM_s^0 \right)_{0 \le t \le T}$ is a $\widetilde{P}$-martingale.

Relation (3.11) easily implies that the process $\widetilde{z}$ is actually continuous.

Suppose, in addition to (c1), that the following condition is satisfied:

(c*) all P-local martingales are continuous.

This technical assumption is satisfied in stochastic volatility models, where $F = F^w$ is the natural filtration generated by the Wiener process.

It was shown in [25, 34] that under conditions (c1) and (c*), the density $\widetilde{z}_T$ of the variance optimal ELMM is uniquely characterized by the relation

$$\widetilde{z}_T = \frac{\mathcal{E}_T((\varphi - k)' \cdot M^0)}{E\mathcal{E}_T((\varphi - k)' \cdot M^0)}, \tag{3.12}$$

where φ, together with the pair (L, c), is a unique solution of the equation

$$\frac{\mathcal{E}_T((\varphi - 2k)' \cdot M)}{\mathcal{E}_T(L)} = c\mathcal{E}_T^2(-k' \cdot M), \tag{3.13}$$

where $L \in M^2_{0,\text{loc}}(P)$, $\langle L, M\rangle = 0$, and c is a constant.

Moreover, the process $\zeta = (\zeta_t)_{0\le t\le T}$ from (3.11) has the form

$$\zeta_t = (\varphi_t - k_t)\mathcal{E}_t((\varphi - k)' \cdot M^0). \tag{3.14}$$

Here, $\varphi = (\varphi_t)_{0\le t\le T}$ is an $\mathbb{R}^d$-valued, F-predictable process with

$$\int_0^T \varphi_t' \, d\langle M\rangle_t \varphi_t < \infty.$$

Let τ be the F-stopping time and $\langle k' \cdot M\rangle_{T\tau} = \langle k' \cdot M\rangle_T - \langle k' \cdot M\rangle_\tau$.

Proposition 3.1 (see [3, 18]). (1) *Equation* (3.13) *is equivalent to the equation*

$$\frac{\mathcal{E}_T(\varphi' \cdot M^*)}{\mathcal{E}_T(L)} = ce^{\langle k' \cdot M\rangle_T}, \tag{3.15}$$

where the $\mathbb{R}^d$*-valued process* $M^* = (M_t^*)_{0\le t\le T}$ *is given by the relation*

$$dM_t^* = 2d\langle M\rangle_t k_t + dM_t, \quad M_0^* = 0.$$

(2) (a) *If there exists a martingale* $m = (m_t)_{0\le t\le T}$, $m \in \mathcal{M}^2_{0,\text{loc}}(P)$, *such that*

$$e^{-\langle k' \cdot M\rangle_T} = c + m_T, \quad \langle m, M\rangle = 0, \tag{3.16}$$

then $\varphi \equiv 0$ *and* $L_T = \int_0^T \frac{1}{c+m}\, dm_s$ *solve Eq.* (3.15). *In this case,*

$$\widetilde{z}_T = \frac{\mathcal{E}_T(-k' \cdot M^0)}{E\mathcal{E}_T(-k' \cdot M^0)}, \tag{3.17}$$

the process $\zeta = (\zeta_t)_{0\le t\le T}$ *from* (3.11) *is equal to*

$$\zeta_t = -k_t\mathcal{E}_t(-k' \cdot M^0),$$

and

$$E\left[\left(\frac{\widetilde{z}_T}{\widetilde{z}_\tau}\right)^2 \Big/ \mathcal{F}_\tau\right] = \frac{1}{E(e^{-\langle k' \cdot M\rangle_{T\tau}}/\mathcal{F}_\tau)}.$$

(b) *If there exist an* $\mathbb{R}^d$*-valued* F*-predictable process* $\ell = (\ell_t)_{0\le t\le T}$, $\int_0^T \ell_t' d\langle M\rangle \ell_t < \infty$, *and*

$$e^{\langle k' \cdot M\rangle_T} = c + \int_0^T \ell_t' \, dM_t^*,$$

then $L \equiv 0$ *and*

$$\varphi_t = \frac{\ell_t}{c + \int_0^t \ell_s' dM_s^*}$$

solve Eq. (3.15). *In this case,*

$$\widetilde{z}_T = \mathcal{E}_T(-k' \cdot M) \quad (:= \widehat{z}_T, \quad \textit{the density of the minimal martingale measure } \widehat{P}),$$

and

$$E\left(\left(\frac{\widetilde{z}_T}{\widetilde{z}_\tau}\right)^2 \Big/ \mathcal{F}_\tau\right) = E^{P^*}(e^{\langle k'\cdot M\rangle_{T\tau}}/\mathcal{F}_\tau),$$

where $dP^* = \mathcal{E}_T(-2k'\cdot M)dP$.

Proof. (1) By the Yor formula,

$$\begin{aligned}\mathcal{E}_T\left((\varphi-2k)'\cdot M\right) &= \mathcal{E}_T(\psi'\cdot M - 2k'\cdot M)\\ &= \mathcal{E}_T\left(\varphi'\cdot\left(M + 2\int_0^\cdot d\langle M\rangle_t k_t\right) - 2\int_0^\cdot \psi_t' d\langle M\rangle_t k_t - 2k'\cdot M\right)\\ &= \mathcal{E}_T(\varphi'\cdot M^*)\mathcal{E}_T(-2k'\cdot M),\end{aligned}$$

and

$$\mathcal{E}_T^2(-k'\cdot M) = \mathcal{E}_T(-2k'\cdot M)e^{\langle k'\cdot M\rangle_T}.$$

The assertion follows.

(2) (a) First, note that $\langle L, M\rangle = 0$. Further, by this formula, we can write

$$\ln(c+m_t) - \ln c = \int_0^t \frac{1}{c+m_s}\,dm_s - \frac{1}{2}\int_0^t \frac{1}{(c+m_s)^2}\,d\langle m\rangle_s.$$

Hence

$$e^{\ln(c+m_T)-\ln c} = \mathcal{E}_T(L),$$

and thus,

$$\mathcal{E}_T(L) = \frac{c+m_T}{c} = \frac{e^{-\langle k'\cdot M\rangle_T}}{c}.$$

Finally, by the Bayes rule and the Girasnov theorem,

$$\begin{aligned}E\left(\left(\frac{\widetilde{z}_T}{\widetilde{z}_\tau}\right)^2 \Big/ \mathcal{F}_\tau\right) &= \frac{E(\mathcal{E}_T(-2k'\cdot M)e^{-\langle k'\cdot M\rangle_T}/\mathcal{F}_\tau)}{E^2(\mathcal{E}_T(-k'\cdot M)e^{-\langle k'\cdot M\rangle_T}/\mathcal{F}_\tau)}\\ &= \frac{E^*(c+m_T/\mathcal{F}_\tau)\mathcal{E}_T^2(-k'\cdot M)}{(\widehat{E}(c+m_\tau/\mathcal{F}_\tau))^2\mathcal{E}_T^2(-2k'\cdot M)} = \frac{c+m_\tau}{(c+m_\tau)^2}\cdot e^{\langle k'\cdot M\rangle_\tau} = \frac{1}{E(e^{\langle k'\cdot M\rangle_{T\tau}}/\mathcal{F}_\tau)}.\end{aligned}$$

The proof of case (2)(b) is similar. The proposition is proved. ☐

3.3. Misspecified asset price model and robust hedging. Denote by $\mathrm{Ball}_L(0,r)$, $r\in[0,\infty)$, the closed r-radius ball in the space $L = L_\infty(dt\times dP)$, centered at the origin, and let

$$\mathcal{H} := \big\{h = \{h_{ij}\},\ i,j = \widehat{1,d} : h \text{ is an } F\text{-predictable } (d\times d)\text{-matrix-valued process,} \tag{3.18}$$

$$\operatorname{rank}(h) = d,\quad h_{ij}\in \mathrm{Ball}_L(0,r),\quad r\in[0,\infty)\big\}. \tag{3.19}$$

The class $\mathcal{H}$ is called the class of alternatives.

Fix the value of the small parameter $\delta > 0$, as well as the $(d\times d)$-matrix-valued, F-predictable process $\sigma^0 = (\sigma_t^0)_{0\le t\le T} = (\{\sigma_{ij,t}^0\},\ 1\le i,j\le d)_t$ such that $|\sigma_{ij,t}^0| \le \mathrm{const}$, $\forall i,j,t$, the matrix $(\sigma^0)^2 = \sigma^0(\sigma^0)'$ is uniformly elliptic, i.e., for each vector $v_t = (v_t', \dots, v_t^d)$ with probability 1

$$\sum_{i,j=1}^d (\sigma^0)^2_{ij,t} v_t^i v_t^j \ge c\sum_{i=1}^d |v_t^i|^2,\quad c>0,\quad 0\le t\le T, \tag{3.20}$$

and denote

$$A_\delta = \{\sigma : \sigma = \sigma^0 + \delta h,\ \ h\in\mathcal{H}\}. \tag{3.21}$$

Proposition 3.2. *Every σ from the class A_δ for a sufficiently small δ is an F-predictable $(d \times d)$-valued process with bounded elements and the matrix $\sigma^2 = \sigma\sigma'$ is uniformly elliptic.*

Proof. The process σ is F-predictable as a linear combination of F-predictable processes. Further,

$$|\sigma_{ij,t}| = |\sigma^0_{ij,t} + \delta h_{ij,t}| \leq \text{const} + \delta r, \quad 0 < \delta \ll 1.$$

From (3.20) and (3.21), for each vector $\nu_t = (\nu_t^1, \dots, \nu_t^d)$, we have

$$\sum_{i,j=1}^d (\sigma^2)_{ij,t}\nu_t^i\nu_t^j = \sum_{i,j=1}^d (\sigma^0 + \delta h)(\sigma^0 + \delta h)'_{ij,t}\nu_t^i\nu_t^j$$
$$= \sum_{i,j=1}^d (\sigma^0(\sigma^0)')_{ij,t}\nu_t^i\nu_t^j + \delta \sum_{i,j=1}^d (\sigma^0 h')_{ij,t}\nu_t^i\nu_t^j + \delta \sum_{i,j=1}^d (h(\sigma^0)')_{ij,t}\nu_t^i\nu_t^j + \delta^2 \sum_{i,j=1}^d (hh')_{ij,t}\nu_t^i\nu_t^j. \quad (3.22)$$

Note now that the entries of the matrices σ^0 and h are bounded. Hence, choosing δ sufficiently small, we obtain

$$\max\left(\delta|(\sigma^0 h')_{ij,t}|, \delta|(h(\sigma^0))_{ij,t}|, \delta^2|(hh')_{ij,t}|\right) \leq \frac{\varepsilon}{3}.$$

Therefore, from (3.20) and (3.22), we obtain

$$\sum_{i,j=1}^d \sigma^2_{ij,t}\nu_t^i\nu_t^j \geq (c - \text{const} \cdot \varepsilon) \sum_{i,j=1}^d |\nu_t^i|^2 \quad \text{for each } \varepsilon > 0.$$

The proposition is proved. □

Consider the set of processes $\{R^\sigma (\text{or } X^\sigma),\ \sigma \in A_\delta\}$, which represents the misspecified asset price model.

Define the class of admissible trading strategies $\Theta = \Theta(\sigma^0)$.

Proposition 3.3. *For each $\mathbb{R}^d$-valued F-predictable process $\theta = (\theta_t)_{0\leq t\leq T}$ and for each $\sigma \in A_\delta$, $\delta > 0$,*

$$aE\int_0^T |\theta_t|^2 dC_t \leq E\int_0^T \theta_t'\sigma_t d\langle M\rangle_t \sigma_t'\theta_t = E\int_0^T \theta_t'\sigma_t\sigma_t'\theta_t dC_t \leq AE\int_0^T |\theta_t|^2 dC_t,$$

where the constants a and A are such that $0 < a \leq A < \infty$ and the parameter $\delta > 0$ is sufficiently small.

Proof. Recall that $d\langle M\rangle_t = d\langle M^i, M^j\rangle_t = I_{ij}^{d\times d} dC_t$. Hence

$$E\int_0^T \theta_t'\sigma_t d\langle M\rangle_t \sigma_t'\theta_t = E\int_0^T \theta_t'\sigma_t\sigma_t'\theta_t\, dC_t.$$

Further, since $\sigma = \sigma^0 + \delta h$ and entries of the matrices σ^0 and h are bounded, the same is true for the entries of the matrix σ with $0 \leq \delta \leq \text{const}$. Thus, using the inequality $ab \leq 2(a^2 + b^2)$, we obtain

$$E\int_0^T \theta_t'\sigma_t\sigma_t'\theta_t\, dC_t \leq AE\int_0^T |\theta_t|^2 dC_t.$$

On the other hand, by Proposition 3.2, the matrix $\sigma^2 = \sigma\sigma'$ is uniformly elliptic for a sufficiently small δ, which yields the first inequality. □

Definition 3.1. The class $\Theta = \Theta(\sigma^0)$ is a class of $\mathbb{R}^d$-valued F-predictable processes $\theta = (\theta_t)_{0\le t\le T}$ such that

$$E\int_0^T |\theta_t|^2 dC_t < \infty. \tag{3.23}$$

Let $\theta \in \Theta$ be the dollar amount (rather than the number of shares) invested in the stock X^σ, $\sigma \in A_\delta$. Then for each $\sigma \in A_\delta$, the trading gains induced by the self-financing portfolio strategy associated with θ have the form

$$G_t(\sigma,\theta) = \int_0^t \theta'_s\, dR^\sigma_s, \quad 0 \le t \le T, \tag{3.24}$$

where $R^d = (R^d_t)_{0\le t\le T}$ is the yield process given by (3.10).

Introduce the following condition:

(c2) There exists ELMM $\overline{Q}$ such that the density process $z = z^{\overline{Q}}$ satisfies the reverse Hölder inequality $R_2(P)$; see the definition in [33].

It is well known that under conditions (c1) and (c2), the density process $\widetilde{z} = (\widetilde{z}_t)_{\le t\le T}$ of the variance-optimal ELMM also satisfies $R_2(P)$ (see [8]).

Now under conditions (c1) and (c2), the r.v. $G_T(\sigma,\theta) \in L^2(P)$ for all $\sigma \in A_\delta$, and the space $G_T(\sigma,\Theta)$ is closed in $L^2(P)$, $\forall\sigma \in A_\delta$ (see, e.g., [33, Theorem 2]).

A contingent claim is an $\mathcal{F}_T$-measurable square-integrable r.v. H, which models the payoff from a financial product at the maturity date T.

The problem we are interested in is to find the robust hedging strategy for a contingent claim H in the incomplete financial market model described above with misspecified asset price process X^σ, $\sigma \in A_\delta$, by using the mean-variance approach.

For each $\sigma \in A_\delta$, the total loss of a hedger, who starts from the initial capital x, uses the strategy θ, believes that the stock price process follows X^σ, and has to pay a random amount H at the date T, is H-x-$G_T(\sigma,\theta)$.

Denote

$$\mathcal{J}(\sigma,\theta) := E(H - x - G_T(\sigma,\theta))^2. \tag{3.25}$$

One setting of the robust mean-variance hedging problem consist in solving the optimization problem

$$\text{minimize} \quad \sup_{\sigma\in A_\delta} \mathcal{J}(\sigma,\theta) \quad \text{over all strategies} \quad \theta \in \Theta. \tag{3.26}$$

We "slightly" change this problem using the approach developed in [41], which is based on the following approximation:

$$\sup_{\sigma\in A_\delta} \mathcal{J}(\sigma,\theta) = \exp\big\{\sup_{h\in\mathcal{H}} \ln \mathcal{J}(\sigma^0+\delta h,\theta)\big\} \simeq \exp\left\{\sup_{h\in\mathcal{H}}\left[\ln \mathcal{J}(\sigma^0,\theta) + \delta\,\frac{D\mathcal{J}(\sigma^0,h,\theta)}{\mathcal{J}(\sigma^0,\theta)}\right]\right\}$$
$$= \mathcal{J}(\sigma^0,\theta)\exp\left\{\delta \sup_{h\in\mathcal{H}} \frac{D\mathcal{J}(\sigma^0,h,\theta)}{\mathcal{J}(\sigma^0,\theta)}\right\},$$

where

$$D\mathcal{J}(\sigma^0,h,\theta) := \frac{d}{d\delta}\,\mathcal{J}(\sigma^0+\delta h,\theta)|_{\delta=0} = \lim_{\delta\to 0}\frac{\mathcal{J}(\sigma^0+\delta h,\theta) - \mathcal{J}(\sigma^0,\theta)}{\delta},$$

is the Gateaux differential of the functional $\mathcal{J}$ at the point σ^0 in the direction h.

Approximate (in leading order δ) the optimization problem (3.26) by the problem

$$\text{minimize } \mathcal{J}(\sigma^0,\theta)\exp\left\{\delta \sup_{h\in\mathcal{H}} \frac{D\mathcal{J}(\sigma^0,h,\theta)}{\mathcal{J}(\sigma^0,\theta)}\right\} \text{ over all strategies } \theta \in \Theta. \tag{3.27}$$

Note that each solution θ^* of problem (3.27) minimizes $\mathcal{J}(\sigma^0, \theta)$ under the constraint

$$\sup_{h \in \mathcal{H}} \frac{D\mathcal{J}(\sigma^0, h, \theta)}{\mathcal{J}(\sigma^0, \theta)} \le c := \sup_{h \in \mathcal{H}} \frac{D\mathcal{J}(\sigma^0, h, \theta^*)}{\mathcal{J}(\sigma^0, \theta^*)}.$$

This characterization of an optimal strategy θ^* of problem (3.27) leads to the following definition.

Definition 3.2. A trading strategy $\theta^* \in \Theta$ is called the optimal mean-variance robust trading strategy against the class of alternatives $\mathcal{H}$ if it is a solution of the optimization problem

$$\text{minimize } \mathcal{J}(\sigma^0, \theta) \text{ over all strategies } \theta \in \Theta, \text{ subject to constraint } \sup_{h \in \mathcal{H}} \frac{D\mathcal{J}(\sigma^0, h, \theta)}{\mathcal{J}(\sigma^0, \theta)} \le c, \qquad (3.28)$$

where c is some generic constant.

Remark 3.2. In contrast to the "mean-variance robust" trading strategy which associates with the optimization problem (3.26) and control theory, we find the "optimal mean-variance robust" strategy in the sense of Definition 3.2. Such an approach and term are common in robust statistics theory (see, e.g., [12, 34]).

Does the suggested approach provide "good" approximation? Consider the following case.

Diffusion model with zero drift. Let the standard Wiener process $w = (w_t)_{0 \le t \le T}$ be given on the complete probability space $(\Omega, \mathcal{F}, P)$. Denote by $F^w = (\mathcal{F}_t^w,\ 0 \le t \le T)$ the P-augmentation of the natural filtration $\mathcal{F}_t^w = \sigma(w_s,\ 0 \le s \le t)$, $0 \le t \le T$, generated by w.

Let the stock price process be modeled by the equation

$$dX_t^\sigma = X_t^\sigma \cdot \sigma_t \, dw_t, \quad X_0^\sigma > 0, \quad 0 \le t \le T,$$

where $\sigma \in A_\delta$ with

$$\int_0^T (\sigma_t^0)^2 \, dt < \infty$$

and $h \in \mathrm{Ball}_{L_\infty(dt \times dP)}(0, r)$, $0 < r < \infty$. All considered processes are real-valued.

Denote by R^σ the yield process, i.e.,

$$dR_t^\sigma = \sigma_t \, dw_t, \quad R_0^\sigma = 0, \quad 0 \le t \le T.$$

The wealth at maturity T with the initial endowment x is equal to

$$V_T^{x,\theta}(\sigma) = x + \int_0^T \theta_t \, dR_t^\sigma.$$

Further, let the contingent claim H be the $\mathcal{F}_T^w$-measurable P-square-integrable r.v.

Consider the optimization problem (3.26). It is easy to see that if $\sigma \in A_\delta$, then

$$\sigma_t^0 - \delta r \le \sigma_t \le \sigma_t^0 + \delta r, \quad 0 \le t \le T, \quad P\text{-a.s.}$$

By the martingale representation theorem,

$$H = EH + \int_0^T \varphi_t^H \, dw_t, \quad P\text{-a.s.},$$

where φ^H is the F^w-predictable process with

$$E \int_0^T (\varphi_t^H)^2 \, dt < \infty. \qquad (3.29)$$

Hence

$$E\big(H - V_T^{x,\theta}(\sigma)\big)^2 = (EH - x)^2 + E\int_0^T (\varphi_t^H - \sigma_t\theta_t)^2\,dt.$$

From this, it directly follows that the process

$$\sigma_t^*(\theta) = (\sigma_t^0 - \delta r) I_{\{\frac{\varphi_t^H}{\theta_t} \ge \sigma_t^0\}} I_{\{\theta_t \ne 0\}} + (\sigma_t^0 + \delta r) I_{\{\frac{\varphi_t^H}{\theta_t} < \sigma_t^0\}} I_{\{\theta_t \ne 0\}}, \quad 0 \le t \le T, \tag{3.30}$$

is a solution of the optimization problem

$$\text{maximize } E\big(H - V_T^{x,\theta}(\sigma)\big)^2 \text{ over all } \sigma \in A_\delta \text{ with a given } \theta \in \Theta.$$

It remains to minimize (with respect to θ) the expression

$$E\int_0^T \big(\varphi_t^H - \sigma_t^*(\theta)\theta_t\big)^2\,dt.$$

From (3.30), it easily follows that the equation (with respect to θ)

$$\varphi_t^H - \sigma_t^*(\theta)\theta_t = 0$$

has no solution, but

$$\theta_t^* = \frac{\varphi_t^H}{\sigma_t^0} I_{\{\sigma_t^0 \ne 0\}}, \quad 0 \le t \le T, \tag{3.31}$$

solves problem. We assume that $0/0 := 0$.

Now we consider the optimization problem (3.28).

For each fixed h,

$$\begin{aligned} J(\sigma,\theta) &= E\Big(H - x - \int_0^T \theta_t\,dR_t^\sigma\Big)^2 = E\bigg(H - x - \int_0^T \theta_t\sigma_t^0\,dw_t - \delta\int_0^T \theta_t h_t\,dw_t\bigg)^2 \\ &= J(\sigma^0,\theta) - 2\delta E\bigg[\Big(EH - x + \int_0^T \big(\varphi_t^H - \theta_t\sigma_t^0\big)\,dw_t\Big)\int_0^T \theta_t h_t\,dw_t\bigg] + \delta^2 E\int_0^T \theta_t^2 h_t^2\,dt, \end{aligned}$$

and hence

$$DJ(\sigma^0,h;\theta) = 2E\int_0^T \big(\theta_t\sigma_t^0 - \varphi_t^H\big)\theta_t h_t\,dt, \tag{3.32}$$

as follows from (3.29), the definition of the class $\mathcal{H}$, and the estimate

$$\begin{aligned} \Big(E\int_0^T \big(\theta_t\sigma_t^0 - \varphi_t^H\big)\theta_t h_t\,dt\Big)^2 &\le E\int_0^T \big(\theta_t\sigma_t^0 - \theta_t^H\big)^2\,dt\; E\int_0^T \theta_t^2 h_t^2\,dt \\ &\le \text{const}\cdot r^2\bigg(E\int_0^T \theta_t^2(\sigma_t^0)^2\,dt + E\int_0^T (\varphi_t^H)^2\,dt\bigg)E\int_0^T \theta_t^2\,dt < \infty. \end{aligned} \tag{3.33}$$

Since, further, $DJ(\sigma^0,h;\theta) = 0$ for $h \equiv 0$, using (3.33), we obtain

$$0 \le \sup_{h\in\mathcal{H}} DJ(\sigma^0,h;\theta) < \infty.$$

Hence we can take $0 \le c < \infty$. Now if we substitute θ^* from (3.31) into (3.32), then $DJ(\sigma^0, h; \theta^*) = 0$ for each h, and thus,

$$\frac{\sup\limits_{h \in \mathcal{H}} DJ(\sigma^0, h; \theta^*)}{J(\sigma^0, \theta^*)} = 0.$$

Recall that $\theta^* = \arg \min\limits_{\theta \in \Theta_{A_\delta}} J(\sigma^0, \theta)$; we obtain that θ^* defined by (3.31) is a solution of this optimization problem as well.

Thus, we prove that

(a) the mean-variance robust trading strategy $\theta^* = (\theta^*_t)_{0 \le t \le T}$ for the optimization problem (3.26) is given by the formula

$$\theta^*_t = \frac{\varphi^H_t}{\sigma^0_t} I_{\{\sigma^0_t \ne 0\}};$$

(b) at the same time, this strategy is an optimal mean-variance robust trading strategy for the optimization problem (3.28).

Hence, in this case, the suggested approach leads to the exact solution of the initial problem (3.26).

To solve problem (3.28) in general case, we need to calculate $D\mathcal{J}(\sigma^0, h, \theta)$. Assume that $k = (k_t)_{0 \le t \le T} = (k_{i,t},\ 1 \le i \le d)_{0 \le t \le T}$ from (3.10) is such that $|k_{i,t}| \le \text{const}$ for all i and t.

Following [11, 33], we introduce the probability measure $\widetilde{Q} \sim P$ on $\mathcal{F}_T$ by the relation

$$d\widetilde{Q} = \frac{\widetilde{z}_T}{\widetilde{z}_0}\, d\widetilde{P} \quad \Big(\text{and hence} \quad d\widetilde{Q} = \frac{\widetilde{z}^2_T}{\widetilde{z}_0}\, dP\Big). \tag{3.34}$$

Using [11, Proposition 5.1], we can write

$$\begin{aligned}\mathcal{J}(\sigma, \theta) &= E\, \frac{\widetilde{z}^2_T}{\widetilde{z}^2_0} \frac{\widetilde{z}^2_0}{\widetilde{z}^2_T} \left(H - x - \int_0^T \theta'_t\, dR^\sigma_t \right)^2 = \widetilde{z}_0^{-1} E^{\widetilde{Q}}\, \frac{\widetilde{z}^2_0}{\widetilde{z}^2_T} \left(H - x - \int_0^T \theta'_t \sigma_t\, dM^0_t \right)^2 \\ &= \widetilde{z}_0^{-1} E^{\widetilde{Q}} \left(\frac{H \widetilde{z}_0}{\widetilde{z}_T} - x - \int_0^T \psi^0_t(\sigma)\, d\,\frac{\widetilde{z}^2_0}{\widetilde{z}^2_t} - \int_0^T \big(\psi^1_t(\sigma)\big)'\, d\,\frac{M^0_t}{\widetilde{z}_t}\, \widetilde{z}_0 \right)^2 := \overline{\mathcal{J}}(\sigma, \psi^0, \psi^1) \end{aligned} \tag{3.35}$$

(or $\overline{\mathcal{J}}(\sigma, \psi)$ with $\psi = (\psi^0, \psi^1)'$), where

$$\psi^1_t = \psi^1_t(\sigma) = \sigma'_t \theta_t, \quad \psi^0_t = \psi^0_t(\sigma) = \int_0^t \theta'_s \sigma_s dM^0_s - \theta'_t \sigma_t M^0_t, \quad 0 \le t \le T. \tag{3.36}$$

Thus,

$$\psi^1_t(\sigma) = \psi^1_t(\sigma^0) + \delta\psi^1_t(h), \quad \psi^0_t(\sigma) = \psi^0_t(\sigma^0) + \delta\psi^0_t(h).$$

Let (following [33])

$$\frac{H}{\widetilde{z}_T}\, \widetilde{z}_0 = E\left(\frac{H}{\widetilde{z}_T}\, \widetilde{z}_0 \right) + \int_0^T (\psi^H_t)' dU_t + L_T, \tag{3.37}$$

be the Galtchouk–Kunita–Watanabe decomposition of the r.v. $\frac{H}{\widetilde{z}_T}\widetilde{z}_0$ with respect to the $\mathbb{R}^{(d+1)}$-valued $\widetilde{Q}$-local martingale $U = \big(\frac{\widetilde{z}_0}{\widetilde{z}}, \frac{M^0}{\widetilde{z}}\, \widetilde{z}_0\big)'$, where $\psi^H = (\psi^{0,H}, \psi^{1,H})' \in L^2(U, \widetilde{Q})$, the space of F-predictable processes ψ such that $\int \psi' dU \in \mathcal{M}^2(\widetilde{Q})$ of the martingale, and $L \in \mathcal{M}^2_{0,\text{loc}}(\widetilde{Q})$, L is $\widetilde{Q}$-strongly orthogonal to U.

Recall that

$$\psi = (\psi^0, \psi^1)'. \tag{3.38}$$

Then, using (3.35), (3.36), and (3.37), for each h, we can write

$$\mathcal{J}(\sigma^0+\delta h,\psi)=\mathcal{J}(\sigma^0,\psi)+\delta\cdot 2\widetilde{z}_0^{-1}E^{\widetilde{Q}}\Bigg\{\Bigg[\bigg(x-E^{\widetilde{Q}}\frac{H}{\widetilde{z}_T}\,\widetilde{z}_0\bigg)-L_T+\int_0^T(\psi_t(\sigma^0)-\psi_t^H)'dU_t\Bigg]$$

$$\times\int_0^T(\overline{\psi}_t(h))'dU_t\Bigg\}+\delta^2\widetilde{z}_0^{-1}E^{\widetilde{Q}}\Bigg[\int_0^T(\overline{\psi}_t(h))'dU_t\Bigg]^2$$

$$=\mathcal{J}(\sigma^0,\psi)+\delta\cdot 2\widetilde{z}_0^{-1}E^{\widetilde{Q}}\Bigg[\int_0^T(\psi_t(\sigma^0)-\psi_t^H)'dU_t\int_0^T(\overline{\psi}_t(h))'dU_t\Bigg]+\delta^2\widetilde{z}_0^{-1}E^{\widetilde{Q}}\Bigg[\int_0^T(\overline{\psi}_t(h))'dU_t\Bigg]^2. \quad (3.39)$$

Using [33, Proposition 8], for each h, we have

$$\frac{\widetilde{z}_0}{\widetilde{z}_T}\,G_r(h,\Theta)=\Bigg\{\int_0^T(\psi(h))'dU_t:\psi(h)\in L^2(U,\widetilde{Q})\Bigg\},$$

and hence, by (3.24),

$$E^{\widetilde{Q}}\Bigg(\int_0^T(\psi_t(h))'dU_t\Bigg)^2=E^{\widetilde{Q}}\frac{\widetilde{z}_0^2}{\widetilde{z}_T^2}\,G_T^2(h,\theta)=\widetilde{z}_0EG_T^2(h,\theta)=\widetilde{z}_0E\Bigg(\int_0^T\theta_t'\,dR_t^h\Bigg)^2$$

$$=\widetilde{z}_0E\Bigg(\int_0^T\theta_t'h_tdM_t^0\Bigg)^2=\widetilde{z}_0E\Bigg(\int_0^T\theta_t'h_td\langle M\rangle_tk_t+\int_0^T\theta_t'h_tdM_t\Bigg)^2$$

$$\le\text{const}\Bigg[E\Bigg(\int_0^T|\theta_t'h_td\langle M\rangle_tk_t|\Bigg)^2+E\Bigg(\int_0^T\theta_t'h_tdM_t\Bigg)^2\Bigg]\le\text{const }r^2E\int_0^T|\theta_t|^2dC_t<\infty. \quad (3.40)$$

Further,

$$\Bigg(E^{\widetilde{Q}}\Bigg[\int_0^T(\psi_t(\sigma^0)-\psi_t^H)'dU_t\int_0^T(\psi_t(h))'dU_t\Bigg]\Bigg)^2$$

$$\le E^{\widetilde{Q}}\Bigg(\int_0^T(\psi_t(\sigma^0)-\psi_t^H)'dU_t\Bigg)^2E^{\widetilde{Q}}\Bigg(\int_0^T(\psi_t(h))'dU_t\Bigg)^2<\infty. \quad (3.41)$$

From these estimates, we conclude that

(1)

$$D\overline{\mathcal{J}}(\sigma^0,h,\psi)=2\widetilde{z}_0^{-1}E^{\widetilde{Q}}\int_0^T(\psi_t(\sigma^0)-\psi_t^H)'d\langle U\rangle_t\psi_t(h)<\infty, \quad (3.42)$$

owing to (3.40).

(2) $D\overline{\mathcal{J}}(\sigma^0,h,\psi)|_{h\equiv 0}=0$, since $\psi(0)=0$ by (3.38) and (3.36). Thus,

$$\sup_{h\in\mathcal{H}}D\overline{\mathcal{J}}(\sigma^0,h,\psi)\ge 0. \quad (3.43)$$

(3) From (3.41) and (3.40), we obtain

$$(D\overline{\mathcal{J}}(\sigma^0,h,\psi))^2 \le \text{const}\, \widetilde{z}_0^{-2} r^2 E^{\widetilde{Q}} \int_0^T (\psi_t(\sigma^0)-\psi_t^H)' d\langle U\rangle_t (\psi_t(\sigma^0)-\psi_t^H) E \int_0^T |\theta_t|^2 dC_t < \infty.$$

Thus, $|D\overline{\mathcal{J}}(\sigma^0,h,\psi)|$ is estimated by an expression independent of h and is equal to zero if we substitute $\psi_t(\sigma^0) \equiv \psi_t^H$, $0 \le t \le T$.

Hence, by (3.43),

$$0 \le \sup_{h\in\mathcal{H}} D\overline{\mathcal{J}}(\sigma^0,h,\psi)|_{\psi\equiv\psi^H} \le \sup_{h\in\mathcal{H}} |D\overline{\mathcal{J}}(\sigma^0,h,\psi)|\big|_{\psi\equiv\psi^H} = 0. \tag{3.44}$$

Further, from (3.43), it follows that we can take $c \in [0,\infty)$ in (3.28).

Now substituting $\psi \equiv \psi^H$ in $\overline{\mathcal{J}}(\sigma^0,\psi)$ and $D\overline{\mathcal{J}}(\sigma^0,h,\psi)$, we obtain

$$\overline{\mathcal{J}}(\sigma^0,\psi^H) = \min_{\psi} \overline{\mathcal{J}}(\sigma^0,\psi) = \widetilde{z}_0^{-1}(E^{\widetilde{P}}H - x)^2 + \widetilde{z}_0^{-1} E^{\widetilde{Q}} L_T^2$$

(see [11, Lemma 5.1]) and

$$\sup_{h\in\mathcal{H}} \frac{D\overline{\mathcal{J}}(\sigma^0,h,\psi^H)}{\overline{\mathcal{J}}(\sigma^0,\psi^H)} = 0.$$

Hence the constraint of problem (3.28) is satisfied.

Remark 3.3. If $x = E^{\widetilde{P}}H$ and $L_T \equiv 0$, then we obtain

$$\frac{D\overline{\mathcal{J}}(\sigma^0,h,\psi^H)}{\overline{\mathcal{J}}(\sigma^0,\psi^H)} = \frac{0}{0}$$

which is assumed to be zero, since if we consider the shifted risk functional $\widetilde{\mathcal{J}} = \overline{\mathcal{J}} + 1$, the optimization problem and the optimal trading strategy are not changed, but $D\widetilde{\mathcal{J}}(\sigma^0,h,\psi^H) = D\overline{\mathcal{J}}(\sigma^0,h,\psi^H) = 0$ and $\widetilde{\mathcal{J}}(\sigma^0,\psi^H) = 1$.

Finally, using [33, Proposition 8] we obtain the following theorem.

Theorem 3.1. *In model* (3.10) *under conditions* (c1) *and* (c2), *the optimal mean-variance robust trading strategy* (*in the sense of Definition* 3.1) *is given by the formula*

$$\theta_t^* = ((\sigma_t^0)')^{-1}[\psi_t^{1,H} + \zeta_t(V_t^* - (\psi_t^H)'U_t)], \quad 0 \le t \le T, \tag{3.45}$$

where

$$\psi_t^H = (\psi_t^{0,H}, \psi_t^{1,H})', \quad U_t = \left(\frac{\widetilde{z}_0}{\widetilde{z}_t}, \frac{M_t^0}{\widetilde{z}_t}\widetilde{z}_0\right)', \quad V_t^* = \frac{\widetilde{z}_0}{\widetilde{z}_t}\left(x + \int_0^t (\psi_t^H)' dU_t\right),$$

ψ_t^H *and* ζ_t *are given by the relations* (3.37) *and* (3.11), *respectively, and* $\widetilde{z}_t$ *is defined in* (3.11).

REFERENCES

1. M. Avellaneda, A. Levy, and A. Paras, "Pricing and hedging derivative securities in markets with uncertain volatilities," *Appl. Math. Finance*, **2**, 73–88 (1995).
2. M. Avellaneda and A. Paras, "Managing the volatility risk of portfolios of derivative securities: The Lagrangian uncertain volatility model," *Appl. Math. Finance*, **3**, 21–52 (1996).
3. F. Biagini, P. Guasoni, and M. Pratelli, "Mean-variance hedging for stochastic volatility models," *Math. Finance*, **10**, No. 2, 109–123 (2000).
4. I. Bajeux-Besnainou and J. Rochet, "Dynamic spanning: Are options an appropriate instrument?" *Math. Finance*, **6**, No. 1, 1–16 (1996).

5. M. Chesney, R. J. Elliott, D. Madan, and H. Yang, "Diffusion coefficient estimation and asset pricing when risk premia and sensitivities are time varying," *Math. Finance*, **3**, No. 2, 85–99 (1993).
6. R. Chitashvili, N. Lazrieva, and T. Toronjadze, *Asymptotic theory of M-estimators in general statistical models. Part* I. *On asymptotic behaviour of estimators under model perturbance*, Report BS-R9019, Centrum voor Woskunde en Informatica Amsterdam (1990).
7. R. Chitashvili, N. Lazrieva and T. Toronjadze, *Asymptotic theory of M-estimators in general statistical models. Part* II. *On asymptotic behaviour of estimators in the presence of nuisance parameter*, Report BS-R9020, Centrum voor Woskunde en Informatica Amsterdam (1990).
8. F. Delbaen, P. Monat, W. Schachermayer, M. Schweizer, and C. Stricker, "Weighted norm inequalities and hedging in incomplete markets," *Finance Stoch.* **1**, No. 3, 181–227 (1997).
9. D. Duffie and H. R. Richardson, "Mean-variance hedging in continuous time," *Ann. Appl. Probab.* **1**, No. 1, 1–15 (1991).
10. H. Föllmer and D. Sondermann, "Hedging of nonredundant contingent claims," in: *Contributions to Mathematical Economics*, North-Holland, Amsterdam (1986), pp. 205–223.
11. C. Gourieroux, J. P. Laurent, and H. Pham, "Mean-variance hedging and numéraire," *Math. Finance*, **8**, No. 3, 179–200 (1998).
12. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics. The Approach Based on Influence Functions*, John Wiley & Sons, New York (1986).
13. J. Hull and A. White, "The pricing of options on assets with stochastic voltilities," *J. Finance*, **3**, 281–300 (1987).
14. I. A. Ibragimov and R. Z. Has'minski, *Statistical Estimation. Asymptotic theory*, Springer-Verlag, New York–Berlin (1981).
15. J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin (1987).
16. N. V. Krylov, *Controlled Diffusion Processes*, Springer-Verlag, New York–Berlin (1980).
17. Yu. Kutoyants, *Identification of Dynamical Systems with Small Noise*, Kluwer Academic Publ., Dordrecht (1994).
18. J. P. Laurent and H. Pham, "Dynamic programming and mean-variance hedging," *Finance Stoch.*, **3**, No. 1, 83–110 (1999).
19. N. L. Lazrieva and T. A. Toronjadze, "Asymptotic properties of an M-estimate in a general statistical experiment scheme," in: *Statistics and Control of Random Processes* [in Russian], Nauka, Moscow (1989), pp. 105–112.
20. N. Lazrieva and T. Toronjadze, "Robust estimators in statistical models with filtration. Shrinking neighbourhoods," in: *Seminarberichte, Fachbereich Mathematik, Fernuniversität, Hagen, Germany*, **48** (1994), pp. 50–68.
21. N. Lazrieva and T. Toronjadze, "Robust estimators in statistical models associated with semimartingales," *Proc. Math. Inst. Razmadze*, **118**, 73–100 (1998).
22. D. Li and W.-L. Ng, "Optimal dynamic portfolio selection: multiperiod mean-variance formulation," *Math. Finance*, **10**, No. 3, 387–406 (2000).
23. R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes,* I. *General theory*, Springer-Verlag, New York–Heidelberg (1977).
24. M. Mania and R. Tevzadze, "A semimartingale Bellman equation and the variance-optimal martingale measure," *Georgian Math. J.* **7**, No. 4, 765–792 (2000).
25. M. Mania, M. Santacroce and R. Tevzadze, "A semimartingale backward equation related to the p-optimal martingale measure and the lower price of a contingent claim," in: *Stochastic Processes and Related Topics*, Stochastics Monogr., **12**, Taylor & Francis, London (2002), pp. 189–212.
26. H. M. Markowitz, *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Basil Blackwell, Oxford (1987).

27. R. C. Merton, "An analytic derivation of the efficient portfolio frontier," *J. Financ. Qualitat. Anal.*, **7**, 1851–1872 (1972).
28. P. Monat and C. Stricker, "Föllmer–Schweizer decomposition and mean-variance hedging for general claims," *Ann. Probab.*, **23**, No. 2, 605–628 (1995).
29. M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modelling*, Springer-Verlag, Berlin (1997).
30. S. Pastorello, "Diffusion coefficient estimation and asset pricing when risk premia and sensitivities are time varying: A comment," *Math. Finance*, **6**, No. 1, 111–117 (1996).
31. H. Pham, T. Rheinländer, and M. Schweizer, "Mean-variance hedging for continuous processes: new proofs and examples," *Finance Stoch.*, **2**, No. 2, 173–198 (1998).
32. E. Renault and N. Touzi, "Option hedging and implied volatilities in a stochastic volatility model," *Math. Finance*, **6**, No. 3, 279–302 (1996).
33. T. Rheinländer and M. Schweizer, "On L^2-projections on a space of stochastic integrals," *Ann. Probab.*, **25**, No. 4, 1810–1831 (1997).
34. H. Rieder, *Robust Asymptotic Statistics*, Springer-Verlag, New York (1994).
35. W. J. Runggaldier and A. Zaccaria, "A stochastic control approach to risk management under restricted information," *Math. Finance*, **10**, No. 2, 277–288 (2000).
36. M. Schweizer, "Mean-variance hedging for general claims," *Ann. Appl. Probab.*, **2**, No. 1, 171–179 (1992).
37. M. Schweizer, "Approximating random variables by stochastic integrals," *Ann. Probab.*, **22**, No. 3, 1536–1575 (1994).
38. M. Schweizer, "Approximation pricing and the variance-optimal martingale measure," *Ann. Probab.*, **24**, No. 1, 206–236 (1996).
39. L. Scott, "Option pricing when the variance changes randomly: theory, estimation and application," *J. Financ. Qualitat. Anal.*, **22**, 419–438 (1987).
40. K. R. Sircar and G. G. Papanicolaou, "Stochastic volatility, smile, and asymptotics," *Appl. Math. Finance*, **6**, No. 2, 107–145 (1999).
41. T. Toronjadze, "Optimal mean-variance robust hedging under asset price model misspecification," *Georgian Math. J.*, **8**, No. 1, 189–199 (2001).

N. Lazrieva
Georgian–American University, Business School, Tbilisi, Georgia;
A. Razmadze Mathematical Institute, Tbilisi, Georgia
T. Toronjadze
Georgian–American University, Business School, Tbilisi, Georgia;
A. Razmadze Mathematical Institute, Tbilisi, Georgia
E-mail: toronj333@yahoo.com