DOI 10.62232/barp.8.2023.7434

Stochastic Volatility Model with Small Randomness. Construction of CULAN Estimators

T. Toronjadze[1,2]

[1]Georgian American University, Business School, 10 Merab Aleksidze Str., 0160, Tbilisi, Georgia;
[2]A. Razmadze Mathematical Institute of I. Javakhishvili Tbilisi State University, 2 Merab Alekside II Lane, 0193 Tbilisi, Georgia

Abstract

CULAN (consistent uniformly linear asymptotically normal) estimators is one of the most important class of estimators in robust statistics. Construction of such estimators for stochastic volatility model with small randomness is a goal of the present paper.

Key words and phrases: Stochastic volatility, small randomness, CULAN estimators.

MSC 2010: 62F12, 62F35.

1 Introduction

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

$$\begin{gathered} dX_t = X_t\, dR_t, \quad X_0 > 0, \\ dR_t = \mu_t(R_t, Y_t)\, dt + \sigma_t\, dw_t^R, \quad R_0 = 0, \\ \sigma_t^2 = f(Y_t), \\ dY_t = a(t, Y_t; \alpha)\, dt + \varepsilon\, dw_t^\sigma, \quad Y_0 = 0, \end{gathered} \tag{1.1}$$

where $w = (w^R, w^\sigma)$ is a standard two-dimensional Wiener process, defined on 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 a vector of unknown parameters, and ε, $0 < \varepsilon < 1$, is a small number. Assume that the system (1.1) has an unique strong solution.

Suppose that the sample path $(y_s)_{0\le s\le t}$ comes from the observations 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,

$$\frac{d\widetilde{P}_\alpha^\varepsilon}{dP_\alpha^\varepsilon} \mid \mathcal{F}_t^w = \mathcal{E}_t(\varepsilon N^\varepsilon), \tag{1.2}$$

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

In the diffusion-type processes framework (1.2) represents the Huber gross error model (as it explain in Remark 2.3). The model of type (1.2) of contamination of measures for statistical models with filtration was suggested by Lazrieva and Toronjadze [1].

In Section 2, we study the problem of construction of robust estimators for contamination model (1.2).

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

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

In the same subsection, we study the asymptotic behaviour of CULAN estimators under alternative measures (Proposition 2.2), which is the basis for the formulation of the optimization problem.

In subsection 2.3, the optimization problem is solved which leads to construction of optimal B-robust estimator (Theorem 2.1).

2 Construction of CULAN estimators

2.1 Basic model

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 a 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 an unknown parameter to be estimated, $\alpha \in \mathcal{A} \subset R^m$, $\mathcal{A}$ is an open subset of R^m, ε, $0 < \varepsilon \ll 1$, is small parameter (index of series). In our further considerations all limits correspond to $\varepsilon \to 0$.

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

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**:

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, which do not depend on α, $k = (k(s))_{0 \le s \le t}$ is a non-decreasing right-continuous function, $0 \le k(s) \le k_0$, $0 < k_0 < \infty$, $x^1, x^2 \in C_t$.

Then, as it is well known (see, e.g., Lipster and Shiryaev [2]), for each $\alpha \in \mathcal{A}$, the equation (2.1) has an unique strong solution $Y^\varepsilon(\alpha) = (Y_s^\varepsilon(\alpha))_{0 \le s \le t}$ and, in addition (see Kutoyants [3]),

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

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 the solution of the following nonperturbated differential equation

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

Change of initial problem of estimation of parameter α by the equivalent one, when 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 it clear that if $X^\varepsilon(\alpha) = (X^\varepsilon_s(\alpha))_{0 \le s \le t}$ is the solution os SDE (2.3), then for each $s \in [0, t]$, $\varepsilon X^\varepsilon_s(\alpha) = Y^\varepsilon_s(\alpha)$.

Denote by P^ε_α the distribution of process $X^\varepsilon(\alpha)$ on the space $(C_t, \mathcal{B}_t)$,i.e., P^ε_α 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 $X = (X_s)_{0 \le s \le t}$ a coordinate process on $(C_t, \mathcal{B}_t)$, that is, $X_s(x) = x_s$, $x \in C_t$.

The conditions **L** guarantee that for each $\alpha \in \mathcal{A}$, the measures P^ε_α and P^w are equivalent ($P^\varepsilon_\alpha \sim P^w$), and if we denote $z^{\alpha,\varepsilon}_s = \frac{dP^\varepsilon_\alpha}{dP^w} \mid \mathcal{B}_s$ the density process (likelihood ratio process), then

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

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

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

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

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

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

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

Assume that for each $s \in [0, t]$ and $x \in C_t$, the functional $a(s, x; \alpha)$is differentiable in α and 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 sign "$'$" denotes a transposition.

Then the Fisher information process

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

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

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

where

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

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

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

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

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

$$P_\alpha^\varepsilon - \lim_{\varepsilon\to 0} \varepsilon^2 \Gamma_{t\varepsilon}^\psi(\alpha) = \Gamma_{t0}^\psi(\alpha), \tag{2.10}$$

$$P_\alpha^\varepsilon - \lim_{\varepsilon\to 0} \varepsilon^2 \gamma_{t\varepsilon}^\psi(\alpha) = \gamma_{t0}^\psi(\alpha), \tag{2.11}$$

where the matrices $\Gamma_{t0}^\psi(\alpha)$ and $\gamma_{t0}^\psi(\alpha)$ are defined as follows:

$$\Gamma_{t0}^\psi(\alpha) = \int_0^t \psi(s, Y^0(\alpha); \alpha)[\psi(s, Y^0(\alpha); \alpha)]'\, ds, \tag{2.12}$$

$$\gamma_{t0}^\psi(\alpha) = \int_0^t \psi(s, Y^0(\alpha); \alpha)[\dot a(s, Y^0(\alpha); \alpha)]'\, ds. \tag{2.13}$$

Note that, by virtue of (2.4), (2.5) and $\dot a \in \Psi$, 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 $\alpha \in \mathcal{A}$, $\operatorname{rank} \Gamma_{t0}^\psi(\alpha) = m$ and $\operatorname{rank} \gamma_{t0}^\psi(\alpha) = m$.

Assume that $\dot a \in \Psi_0$.

For each $\psi \in \Psi_0$, define a P_α^ε-square integrable martingale $L_t^{\psi,\varepsilon}(\alpha)$ as follows:

$$L_t^{\psi,\varepsilon}(X; \alpha) := \int_0^t \psi_\varepsilon(u, X; \alpha)(dX_u - a_\varepsilon(u, X; \alpha)du). \tag{2.14}$$

Now we give a definition of CULAN M-estimators.

Definition 2.1. An estimator $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0} = (\alpha_{1,t}^{\psi,\varepsilon}, \dots, \alpha_{m,t}^{\psi,\varepsilon})'_{\varepsilon>0}$, $\psi \in \Psi_0$, is called consistent uniformly linear asymptotically normal (CULAN) if it admits the following expansion:

$$\alpha_t^{\psi,\varepsilon} = \alpha + [\gamma_{t0}^{\psi}(\alpha)]^{-1}\varepsilon^2 L_t^{\psi,\varepsilon}(\alpha) + r_{t\varepsilon}^{\psi}(\alpha), \tag{2.15}$$

where uniformly in α on each compact,

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

It is well known (see Kutoyants [3]) that under the above conditions, uniformly in α on each compact,

$$\mathcal{L}\{\varepsilon^{-1}(\alpha_t^{\psi,\varepsilon} - \alpha) \mid P_\alpha^\varepsilon\} \xrightarrow{w} N(0, V_t(\psi;\alpha)),$$

with

$$V_t(\psi;\alpha) := [\gamma_{t0}^{\psi}(\alpha)]^{-1}\Gamma_{t0}^{\psi}(\alpha)([\gamma_{t0}^{\psi}(\alpha)]^{-1})', \tag{2.17}$$

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

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

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

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

The asymptotic theory of M-estimators for general statistical models with filtration is developed in Toronjadze [4]. Namely, the problem of existence and global asymptotic behaviour of solutions is studied. In particular, the conditions of regularity and ergodicity type are established under which M-estimators have a CULAN property.

For our model, in case when $\mathcal{A} = R^m$, 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, which do not depend on α.

(2) the equation (w.r.t. y)

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

has a unique solution $y = \alpha$.

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

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

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

2.2 Shrinking contamination neighborhoods

In this subsection, we give a notion of a 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., Hampel et al. [5] and, also, Remark 2.3 below).

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

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

$$\begin{aligned} &1) \quad P_\alpha^{\varepsilon,h} \sim P_\alpha^\varepsilon, \\ &2) \quad \frac{dP_\alpha^{\varepsilon,h}}{dP_\alpha^\varepsilon} = \mathcal{E}_t(\varepsilon N_\alpha^{\varepsilon,h}) \end{aligned} \tag{2.18}$$

where

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

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 $\mathbf{P}_\alpha^{\varepsilon,\mathcal{H}}$ a class of measures $P_\alpha^{\varepsilon,h}$, $h \in \mathcal{H}$, that is,

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

We call $(\mathbf{P}_\alpha^{\varepsilon,\mathcal{H}})_{\varepsilon>0}$ a shrinking contamination neighborhoods of the basic measures $(P_\alpha^\varepsilon)_{\varepsilon>0}$, and the element $(P_\alpha^{\varepsilon,h})_{\varepsilon>0}$ of these neighborhoods are called alternative measures (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 the 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 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 $X_1, X_2, \ldots, X_n$, $n \ge 1$, 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 (w.r.t. some dominating measure μ), $h(x, \alpha)$ is a contaminating density, $f^{n,h}(x; \alpha)$ is a contaminated density, $\varepsilon_n = O(n^{-1/2})$. If we denote by P_α^n and $P_\alpha^{n,h}$ the measures on $(R^n, \mathcal{B}(R^n))$, 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, $\mathcal{E}_n(\varepsilon_n N_\alpha^{n,h}) = \prod_{i=1}^n (1 + \varepsilon_n \Delta N_{\alpha,i}^{n,h})$ is the Dolean exponential in discrete time case.

Thus

$$\frac{dP^{n,h}_{\alpha}}{dP^{n}_{\alpha}} = \mathcal{E}_n(\varepsilon_n \cdot N^{n,n}_{\alpha}) \tag{2.20}$$

and the 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 Lazrieva and Toronjadze [1] for more general situation, concerning with the contamination areas for semimartingale statistical models with filtration. □

In the remainder of this subsection, we study the asymptotic properties of CULAN estimators under alternatives.

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

Let $(\varepsilon_n)_{n\geq 1}$, $\varepsilon_n \downarrow 0$, and $(\alpha_n)_{n\geq 1}$, $\alpha_n \in K$, $K \subset \mathcal{A}$ is a compact, be arbitrary sequences.

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

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

Proof. From the predictable criteria of contiguity (see, e.g., Jacod and Shiryaev [6]), it follows that we have to verify the relation

$$\lim_{N\to\infty} \limsup_{n\to\infty} P^{\varepsilon_n,h}_{\alpha_n}\Big\{h^n_t\Big(\frac{1}{2}\Big) > N\Big\} = 0, \tag{2.21}$$

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

By the definition of Hellinger process (see, e.g., Jacod and Shiryaev [6]), we have

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

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

Proposition 2.2. *For each estimator $(\alpha^{\varepsilon,\psi}_t)_{\varepsilon>0}$ with $\psi \in \Psi_0$ and each alternative $(P^{\varepsilon,h}_{\alpha})_{\varepsilon>0} \in (\mathbf{P}^{\varepsilon,h}_{\alpha})_{\varepsilon>0}$, the following relation holds true:*

$$\mathcal{L}\{\varepsilon^{-1}(\alpha^{\psi,\varepsilon}_t - \alpha) \mid P^{\varepsilon,h}_{\alpha}\} \xrightarrow{w} N([\gamma^{\psi}_{t0}(\alpha)]^{-1} b_t(\psi,h;\alpha), V_t(\psi;\alpha)),$$

where

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

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

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

and therefore we have to establish the limit distribution of random vector $[\gamma_{t0}^{\psi}(\alpha)]^{-1}\varepsilon L_t^{\psi,\varepsilon}$ under the measures $(P_\alpha^{\varepsilon,h})_{\varepsilon>0}$.

By virtue of the Girsanov Theorem, the process $L_t^{\psi,\varepsilon}(\alpha) = (L_s^{\psi,\varepsilon}(\alpha))_{0\le s\le t}$ is a semimartingale with canonical decomposition

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

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

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

and

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

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

$$P_\alpha^{\varepsilon,h} - \lim_{\varepsilon\to 0} \langle \varepsilon \widetilde{L}^{\psi,\varepsilon}(\alpha)\rangle_t = P_\alpha^{\varepsilon,h} - \lim_{\varepsilon\to 0} \varepsilon^2 \Gamma_{t\varepsilon}^{\psi}(\alpha) = P_\alpha^{\varepsilon} - \lim_{\varepsilon\to 0} \varepsilon^2 \Gamma_{t\varepsilon}^{\psi}(\alpha) = \Gamma_{t0}^{\psi}(\alpha)$$

uniformly in α on each compact, and hence

$$\mathcal{L}\{([\gamma_{t0}^{\psi}(\alpha)]^{-1}\varepsilon \widetilde{L}_t^{\psi,\varepsilon} \mid P_\alpha^{\varepsilon,h}\} \xrightarrow{w} N(0, V_t(\psi;\alpha)). \tag{2.23}$$

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

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

provides the desirable result. □

2.3 Optimization criteria. Construction of optimal B-robust estimators

In this subsection, we state and solve an optimization problem, which results in construction of optimal B-robust estimator.

Initially, it should be stressed that the bias vector $\widetilde{b}_t(\psi, h; \alpha) := [\gamma_{t0}^{\psi}(\alpha)]^{-1}$ $b_t(\psi, h; \alpha)$ can be viewed as the influence functional of the estimator $(\alpha_t^{\psi,\varepsilon})_{\varepsilon>0}$ w.r.t. alternative $(P_\alpha^{\psi,h})_{\varepsilon>0}$.

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

$$\mathcal{L}\{\varepsilon^{-1}(a_t^{\psi,\varepsilon} - \alpha - \varepsilon^2[\gamma_{t0}^{\psi}(\alpha)]^{-1} b_{\varepsilon t}(\psi, h; \alpha)) \mid P_\alpha^{\varepsilon,h}\} \xrightarrow{w} N(0, V_t(\psi; \alpha)),$$

and hence, the expression

$$\alpha + \varepsilon^2[\gamma_{t0}^{\psi}(\alpha)]^{-1} b_{\varepsilon t}(\psi, h; \alpha) - \alpha = \varepsilon^2[\gamma_{t0}^{\psi}(\varepsilon)]^{-1} b_{\varepsilon t}(\psi, h; \alpha)$$

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

$$P_\alpha^{\varepsilon,h} - \lim_{\varepsilon\to 0} \frac{\alpha + \varepsilon^2[\gamma_{t0}^{\psi}(\alpha)]^{-1} b_{\varepsilon t}(\psi, h; \alpha) - \alpha}{\varepsilon} = [\gamma_{t0}^{\psi}(\varepsilon)]^{-1} b_{\varepsilon t}(\psi, h; \alpha)$$

as the influence functional.

For each estimator $(a_t^{\psi,\varepsilon})_{\varepsilon>0}$, $\psi \in \Psi_0$, defined the risk functional w.r.t. alternative $(P_\alpha^{\varepsilon,h})_{\varepsilon>0}$, $h \in \mathcal{H}$, as follows:

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

where $x \wedge a = \min(x, a)$, $a > 0$, $E_\alpha^{\varepsilon,h}$ is an expectation w.r.t. measures $P_\alpha^{\varepsilon,h}$.

Using Proposition 2.2 it is not hard to verify that

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

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

Connect with each $\psi \in \Psi_0$ the function $\widetilde{\psi}$ as follows:

$$\widetilde{\psi}(t, x; \alpha) = [\gamma_{t0}^{\psi}(\alpha)]^{-1} \psi(t, x; \alpha).$$

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

$$\gamma_{t0}^{\widetilde{\psi}}(\alpha) = Id,$$

where Id is a unit matrix,

$$V_t(\psi; \alpha) = V_t(\widetilde{\psi}; \alpha) = \Gamma_{t0}^{\widetilde{\psi}}(\alpha), \quad \widetilde{b}_t(\psi, h; \alpha) = \widetilde{b}_t(\widetilde{\psi}, h; \alpha) = b_t(\widetilde{\psi}, h; \alpha).$$

Therefore

$$D_t(\psi, h; \alpha) = D_t(\widetilde{\psi}, h; \alpha) = |b_t(\widetilde{\psi}, h; \alpha)|^2 + \operatorname{tr} \Gamma_{t0}^{\widetilde{\psi}}(\alpha). \tag{2.24}$$

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

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

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

Since, for each $r > 0$,

$$\sup_{h \in \mathcal{H}_r} |b_t(\widetilde{\psi}, h; \alpha)| \le \text{const.}(r) \sup_{0 \le s \le t} |\widetilde{\psi}_t(s, Y^0(\alpha); \alpha)|,$$

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

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

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

Define

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

$$\gamma_{t\psi}^*(\alpha) \le c \bigg\}, \tag{2.26}$$

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

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

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

under the side conditions (2.25) and (2.26).

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

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

For arbitrary nondegenerate matrix A denote $\psi_c^A = h_c(Aa)$.

Theorem 2.1. *Assume that for given constant c there exists a nondegenerate $m \times m$ matrix $A_c^*(\alpha)$, which solves the equation (w.r.t. matrix A)*

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

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

Proof. (See, e.g., Hampel et al. [5].)

Let A be an arbitrary $m \times m$ matrix.

Since for each $\psi \in \Psi_0$, $\int \psi(\dot{a})' = Id$, $\int \dot{a}[\dot{a}]' = I^0(\alpha)$ (see (2.7)) and the trace is an additive functional, 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 evident notations for integrals).

Therefore instead of minimizing of $\mathrm{tr} \int \psi\psi'$ we can minimize

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

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

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

As we have 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 equation (2.2).

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

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

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

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

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 an influence function of optimal B-robust estimator $(\alpha_t^{*,\varepsilon})_{\varepsilon > 0} = (\alpha_t^{\psi_c^*,\varepsilon})_{\varepsilon > 0}$ over the class of CULAN estimators $(\alpha_t^{\psi,\varepsilon})_{\varepsilon > 0}$, $\psi \in \Psi_{0,c}$, if the matrix $A^*(\alpha)$ is differentiable in α.

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

$$\gamma_{t0}^{\psi_c^*}(\alpha) = P_\alpha^\varepsilon - \lim_{\varepsilon\to 0} \varepsilon^2 \gamma_{t\varepsilon}^{\psi_c^*}(\alpha) = \int_0^t \psi_c^*(s, Y^0(\alpha); \alpha)(\dot{a}(s, Y^0(\alpha); \alpha))' ds = Id.$$

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

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

where

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

and $V_t(\psi_c^*; \alpha) = \Gamma_{t0}^{\psi_c^*}(\alpha)$.

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

$$D_t(\psi_c^*, h; \alpha) = |b_t(\psi_c^*, h; \alpha)|^2 + \operatorname{tr} \Gamma_{t0}^{\psi_c^*}, \quad h \in \mathcal{H},$$

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

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

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

$$\Gamma_{t0}^{\psi_c^*}(\alpha) \le \Gamma_{t0}^{\psi}(\alpha) \quad \text{for all } \psi \in \Psi_{0,c}.$$

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

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

and since from (2.24)

$$\inf_{\psi\in\Psi_{0,c}} \sup_{h\in\mathcal{H}_r} D_t(\psi, h; \alpha) \le C^2(r; c) + \operatorname{tr} \Gamma_{t0}^{\psi_c^*}(\alpha),$$

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

$$C^2(r; c) + \operatorname{tr} \Gamma_{t0}^{\psi_c^*}(\alpha)$$

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

For the problem of existence and uniqueness of solution of equation (2.28), we address to Rieder [7].

In the case of 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 Lazrieva and Toronjadze [1])

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

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:

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

Appendix

Important feature of the stochastic volatility model is that volatility process Y is unobservable (latent) process. Clear that 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 direct statistical analysis of the historical data or the use of implied volatilities extracted from prices of existing traded derivatives.

Consider the method based on historical data.

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

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

It is well known (see, e.g., Lipster and Shiryaev [2]) that

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

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

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

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 $y_t = f^{-1}(\sigma_t^2)$, $0 \le t \le T$.

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 diffusion process Y.

The second market price adjusted procedure of reconstruction the sample path of volatility process Y and parameter estimate was suggested by Renault and Touzi [8], where they used implied volatility data.

We present a quick review of this method, adapted to our model (1.1).

Suppose that the volatility risk premium $\lambda^\sigma \equiv 0$, meaning that the risk from the volatility process is non-compensated (or can be diversified away). Then the price $C_t(\sigma)$ of European call option can be calculated by Hull and White formula (see, e.g., Renault and Touzi [8]), and Black–Scholes (BS) implied volatility $\sigma^i(\sigma)$ can be found as an 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 at-the-money options are used.

Under some technical assumptions (see Proposition 5.1 of Renault and Touzi [8]),

$$\frac{\partial \sigma_t^i(\sigma,\alpha)}{\partial \sigma_t} > 0 \tag{2.30}$$

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

Fix current value of 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 at-the-money options.

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

Then, using (2.30) and solving the equation

$$\sigma^i_{t^\varepsilon_j}(\sigma_{t^\varepsilon_j},\alpha) = \sigma^{i^*}_{t^\varepsilon_j},$$

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

$$Y_{t^\varepsilon_{j+1}} - Y_{t^\varepsilon_j} = a(t^\varepsilon_j, Y_{t^\varepsilon_j};\alpha)(t^\varepsilon_{j+1} - t^\varepsilon_j) + \varepsilon(w^\sigma_{t^\varepsilon_{j+1}} - w^\sigma_{t^\varepsilon_j}).$$

Using the data $\{y_{t^\varepsilon_j}\}$ one can construct the MLE $\widehat{\alpha}^\varepsilon_t$ of parameter α, see, e.g., Chitashvili et al. [9].

(Remember the scheme of construction of MLE. Rewrite the previous AR(1) process, using obvious simple notations, in form

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

Then

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

and the "likelihood" process $\ell_t = (\ell^{(1)}_t, \dots, \ell^{(m)}_t)$ is given by the relation

$$\ell^{(i)}_t = \sum_j \ell^{(i)}_{j+1}, \quad i = \overline{1,m},$$

where

$$\begin{aligned}\ell^{(i)}_{j+1}(y;\alpha) &= \frac{\partial}{\partial\alpha_i} \ln\varphi_{j+1}(y, Y_j;\alpha) \\ &= \frac{1}{\varepsilon^2\Delta}\big(y - Y_j - a(t_j, Y_j;\alpha)\Delta\big)\dot{a}^{(i)}(t_j, Y_j;\alpha)\Delta.\end{aligned}$$

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

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

where the reconstructed data $\{y_j\} = \{y_{t^\varepsilon_j}\}$ are substituted.)

Let us introduce the functionals

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

and

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

Starting with some constant initial value (or preliminary estimator obtained, e.g., from historical data), one can compute a sequence of estimators

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

If the operator ϕ_ε is a strong contraction in the neighborhood of the true value of the parameter, α^0, for a small enough ε, then one can define the estimator $\widehat{\alpha}_t^\varepsilon$ as the limit of the sequence $\{\widehat{\alpha}_t^\varepsilon(p)\}_{p\ge 1}$ which is a strong consistent estimator of the parameter α.

References

[1] N. Lazrieva and T. Toronjadze, Robust estimators in statistical models associated with semimartingales. *Proc. A. Razmadze Math. Inst.* **118** (1998), 73–100.

[2] R. S. Liptser and A. N. Shiryaev, *Statistics of Stochastic Processes*. Springer, Berlin, Heidelberg, New York, 1977.

[3] Yu. Kutoyants, *Identification of Dynamical Systems with Small Noise*. Mathematics and its Applications, 300. Kluwer Academic Publishers Group, Dordrecht, 1994.

[4] T. Toronjadze, Construction of identifying and real M-estimators in general statistical model with filtration. *Business Administration Research Papers*, Dec. 2022, `https://doi.org/10.48614/bara.7.2022.6042`.

[5] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust Statistics. The Approach Based on Influence Functions*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.

[6] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 288. Springer-Verlag, Berlin, 1987.

[7] H. Rieder, *Robust Asymptotic Statistics*. Springer Series in Statistics. Springer-Verlag, New York, 1994.

[8] E. Renault and N. Touzi, Option hedging and implied volatilities in a stochastic volatility model. *Math. Finance* **6** (1996), no. 3, 279–302.

[9] 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*. Centrum voor Wiskunde en Informatica Amsterdarn, Report BS - R9019, 1990.