Convergence of the ARMA-GARCH Implied Calibration Algorithm

P. Arbuzov

Abstract — **This paper explores the convergence properties of the randomized stochastic projected gradient-free (RSPGF) algorithm for calibrating the ARMA-GARCH model using market option prices. The calibration problem is framed as a stochastic optimization task within the risk-neutral probability measure, addressing the discrepancy between theoretical and market-implied option prices. The ARMA-GARCH model, combining autoregressive moving average and generalized autoregressive conditional heteroskedasticity components, captures the volatility clustering and dynamics of financial asset returns. The proposed RSPGF algorithm integrates gradient-free optimization with random smoothing techniques to handle the nonlinearity and complexity of the model, where analytical gradient computation is infeasible. A Monte Carlo method estimates the loss function, ensuring unbiased estimates with bounded variance under time series stationarity. The paper proves a theorem establishing the Lipschitz continuity and boundedness properties of the loss function, providing theoretical guarantees for the algorithm's convergence to an ε-stationary point at a rate of O(1/√N). These findings confirm the algorithm's applicability for robust calibration of ARMA-GARCH models, offering practical insights for financial modeling and option pricing.**

Keywords — **ARMA-GARCH models, Convergence analysis, Option pricing, Stochastic optimization.**

I. INTRODUCTION

In this paper the problem of calibrating the parameters of the ARMA-GARCH time series model for the returns of the base asset using market option quotes is considered [1]. Market option quotes often differ from the prices suggested by theoretical models [2], [11]. This means that market participants have their own view on the possible future dynamics of the underlying asset.

The paper considers the application of the ARMA-GARCH model – a combination of the ARMA model, which is a model of autoregression and moving average often used in modeling many discrete random processes; and the GARCH model, which is a generalized autoregressive conditional heteroskedasticity model that allows modeling the volatility clustering effect often observed in financial markets [9], [10].

In the second section of the article, the formulation of the problem of calibrating the parameters of the ARMA-GARCH model for the returns of the base asset in the risk-neutral probability measure is presented as a stochastic optimization problem [5]. The third section provides an overview of the random smoothing method and formulates the randomized stochastic projected gradient free algorithm that solves the posed stochastic optimization problem.

In the concluding section of the work, a theorem about the properties of the considered loss function is proven, a conclusion of which is the convergence of the presented algorithm in the calibration of the ARMA-GARCH model using market option prices.

II. PROBLEM STATEMENT

The $\mathrm{ARMA}(p,q)-\mathrm{GARCH}(P,Q)$ model has the form:

$$\tilde{Y}_t = \tilde{m}_t + \tilde{\epsilon}_t; \tag{1}$$

$$\tilde{\epsilon}_t = \sqrt{\tilde{h}_t}\varepsilon_t, \quad \varepsilon_t \sim \mathrm{iid}(0,1); \tag{2}$$

$$\tilde{m}_t = \phi_0 + \phi_1\tilde{Y}_{t-1} + \dots + \phi_p\tilde{Y}_{t-p} + \theta_1\tilde{\epsilon}_{t-1} + \dots + \theta_q\tilde{\epsilon}_{t-q}; \tag{3}$$

$$\tilde{h}_t = \alpha_0 + \alpha_1\tilde{h}_{t-1} + \dots + \alpha_P\tilde{h}_{t-P} + \beta_1\tilde{\epsilon}^2_{t-1} + \dots + \beta_Q\tilde{\epsilon}^2_{t-Q}; \tag{4}$$

$$\alpha_0 > 0, \alpha_1, \dots, \alpha_P, \beta_1, \dots, \beta_Q \geq 0 \tag{5}$$

$$\tilde{Y}_0 = \tilde{h}_0 = \tilde{\epsilon}_0 = 0. \tag{6}$$

Note that $\tilde{m}_t$ and $\tilde{h}_t$ are also the conditional expectation and variance, respectively:

$$\tilde{m}_t = \mathbb{E}[\tilde{Y}_t \mid F_{t-1}] \tag{7}$$

$$\tilde{h}_t = \mathrm{Var}[\tilde{Y}_t \mid F_{t-1}] \tag{8}$$

where F_{t-1} is the natural filtration of the random process Y_t in the probability space $(\Omega, F, \mathbb{P})$.

We consider the $\mathrm{ARMA}(1,1)-\mathrm{GARCH}(1,1)$ model with parameters $(\phi_0, \phi_1, \theta_1, \alpha_0, \alpha_1, \beta_1)$, for which the sufficient conditions for the stationarity of the modeled time series are known [8]:

$$\begin{cases} |\phi_1 + \theta_1| < 1 \\ \alpha_1 + \beta_1 < 1 \end{cases} \tag{9}$$

Manuscript received May 26, 2025.

Petr Arbuzov, Ph.D student at Lomonosov Moscow State University (e-mail: arbuzov.parb@gmail.com).

It is known that when modeling the logarithmic returns of the underlying asset $Y_t = \ln\left(\frac{S_t}{S_{t-1}}\right)$, where S_t is the asset price at time t, in the risk-neutral probability measure, the model takes the form [5]:

$$Y_t = e^{r\tau} - 1 + \sqrt{h_t}\left(\frac{e^{r\tau}}{1+m_t}\right)\varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0,1) \tag{10}$$

where Y_t is the linear approximation of the underlying asset return, τ is the length of the time interval between consecutive observations of the time series expressed in years [1]. Since Y_t represent the daily logarithmic returns of the underlying asset, the price of the underlying asset at the final time T is given by

$$S_T = S_0 e^{Y_1+Y_2+\cdots+Y_t} \tag{11}$$

where S_0 is the price of the underlying asset at the initial time.

The problem of selecting the parameters of the underlying asset model can be formulated as a conditional optimization problem. We define the loss function as the sum of the expected relative errors in the estimation of the option prices – the expectation of the absolute value of the ratio of the difference between the discounted option payoff and its market price to the market price of the option.

For the case where call options are considered, the loss function f can be represented as follows:

$$f(x) = \sum_{Calls} \mathbb{E}\frac{|e^{-rT}max(S_T - X, 0) - C_{obs}|}{C_{obs}} \tag{12}$$

where x is the model parameters vector $(\phi_0, \phi_1, \theta_1, \alpha_0, \alpha_1, \beta_1)$, S_T is the price of the underlying asset at the expiration time of the option, X is the options strike, r is the risk-free interest rate, C_{obs} are the observed market call options quotes, $Calls$ is the set of all considered call options.

Due to the linearity of expectation, the loss function (12) can be represented as the expectation of the sum of the absolute values of the ratios of the differences between the discounted option payoffs and their market prices to the corresponding market prices:

$$f(x) = \mathbb{E}\left[\sum_{Calls}\frac{|e^{-rT}\max(S_T - X, 0) - C_{\text{obs}}|}{C_{\text{obs}}}\right] \tag{13}$$

Let ξ describe the realizations of the random trajectories of the underlying asset, and $f(x,\xi)$ be the value of the loss function corresponding to these trajectories. Then

$$f(x) = \mathbb{E}_\xi[f(x,\xi)] \tag{14}$$

which means that this problem can be represented as a classical stochastic optimization problem:

$$\min_{x \in \Omega} f(x) \tag{15}$$

For the problem of calibrating the ARMA-GARCH model for the returns of the underlying asset, additional conditions (9) are imposed on the solution to ensure the stationarity of the time series of the underlying asset returns.

Thus, the problem takes the form of a conditional stochastic optimization problem with a convex set of admissible values for the controlled parameters.

To ensure that the domain Ω is compact, it is necessary to impose additional restrictions on the possible values of the model parameters and to strengthen the sufficient conditions for stationarity:

$$|\phi_1 + \theta_1| \le 1 - \delta_A, \tag{16}$$

$$\alpha_1 + \beta_1 \le 1 - \delta_G, \tag{17}$$

$$\alpha_0 \ge \delta_\alpha, \tag{18}$$

$$\alpha_1 \ge 0, \tag{19}$$

$$\beta_1 \ge 0, \tag{20}$$

$$|\phi_0| \le C, \tag{21}$$

$$|\phi_1| \le C, \tag{22}$$

$$|\theta_1| \le C. \tag{23}$$

where $\delta_A > 0$, $\delta_G > 0$ are small constants strengthening the stationarity conditions for the ARMA and GARCH models, respectively, δ_α is a small constant bounding the parameter α from below, $C > 0$ is a constant bounding the model parameters from above.

III. Randomized Smoothing Method and Randomized Stochastic Projected Gradient Free Algorithm

Due to the nonlinearity and overall complexity of the used time series model for the underlying asset returns, there is no confirmation of the convexity of this problem, and the analytical calculation of the gradient of the loss function is not feasible in the case of a large number of considered options and a long planning horizon T. In such a case, the randomized stochastic projected gradient free algorithm [3] may be suitable for solving this problem, which is a modification of the classical gradient projection method and belongs to the class of stochastic approximation methods.

For this algorithm at each iteration, it is necessary to compute the current estimate of the loss function value, which should be unbiased and have bounded variance. The estimation of the current loss function value can be performed using the Monte Carlo method. It is necessary to generate a large number of trajectories of the underlying asset, calculate the loss function value for each trajectory, and average the result. Let $f(x)$ be the Monte Carlo estimate of the loss function, i.e.,

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} f(x,\xi_i) \tag{24}$$

where N is the number of trajectories of the underlying asset, ξ_i are the random daily returns of the underlying asset along the trajectory number i.

It is known that the estimate obtained by the Monte Carlo method is unbiased:

$$\mathbb{E}[f(x)] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} f(x,\xi_i)\right] = f(x) \tag{25}$$

and the variance of such an estimate is bounded in the case of stationarity of the studied time series: the stationarity of the time series guarantees the boundedness of the variances of the underlying asset returns, which in turn implies the boundedness of the variance of the underlying asset price at the final time and the boundedness of the variance of the loss function itself.

At the beginning of each iteration, it is necessary to approximate the gradient of the loss function with respect to the model parameter vector based on the values of the loss function. The approximation of the gradient of the loss function can be performed using the random smoothing method proposed by Nesterov [4]. This method allows constructing a smooth approximation of the stochastic gradient. Let ν be a random vector in the space R^n with density distribution ρ, then the smooth approximation of the function f is given by

$$f_\mu(x) = \int f(x + \mu v)\rho(v)\ dv, \qquad (26)$$

where $\mu > 0$ is the smoothing parameter. One of the frequently used methods for solving similar problems is the smoothing method using the multidimensional normal distribution. If ν is a vector drawn from an n-dimensional standard normal distribution, the function admits the following approximation:

$$f_\mu(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int f(x + \mu v)\, e^{-\frac{1}{2}|v|^2}\ dv \qquad (27)$$

which equals to the expectation $\mathbb{E}_v[f(x + \mu v)]$.

For the domain $\Omega \subseteq \mathbb{R}^n$ let us define $f \in \mathbb{C}_L^{1,1}(\Omega)$ if the gradient of the function f satisfies the Lipschitz condition in this area with constant L, that is for any $x, y \in \Omega$:

$$\|\nabla f(y) - \nabla f(x)\| \le L||y - x||. \qquad (28)$$

It has been proven that for $f \in \mathbb{C}_L^{1,1}(\Omega)$, its random smooth approximation using the normal distribution f_μ has the following properties [4]:

1. f_μ is also continuously differentiable with Lipschitz constant $L_\mu < L$ and its gradient is given by:

$$\frac{1}{(2\pi)^{\frac{n}{2}}} \int \frac{f(x + \mu v) - f(x)}{\mu} v e^{-\frac{1}{2}|v|^2}\ dv \qquad (29)$$

2. For any $x \in \mathbb{R}^n$ the following inequalities hold:

$$|f_\mu(x) - f(x)| \le \frac{\mu^2}{2} Ln, \qquad (30)$$

$$\|\nabla f_\mu(x) - \nabla f(x)\| \le \frac{\mu}{2} L(n+3)^{3/2}, \qquad (31)$$

$$\mathbb{E}_v\left[\left\|\frac{f(x + \mu v) - f(x)}{\mu} v\right\|^2\right] \le \le 2(n+4) + \frac{\mu^2}{2} L^2 (n+6)^3. \qquad (32)$$

Let us represent the approximation of the stochastic gradient of the function f at the parameter values x_k, where k is the iteration number of the randomized stochastic projected gradient free algorithm, in the following form:

$$G_\mu(x_k, \xi_k, v) = \frac{f(x_k + \mu v, \xi_k) - f(x_k, \xi_k)}{\mu} v, \qquad (33)$$

where ξ_k are the realizations of the random daily returns of the underlying asset for all generated trajectories at the k-th iteration of the algorithm. In this case, it has been proven [4] that

$$\mathbb{E}_{v,\xi_k}\left[G_\mu(x_k, \xi_k, v)\right] = \nabla f_\mu(x_k). \qquad (34)$$

Additionally, Nesterov has proven that to improve the convergence rate of the algorithm, it is advisable to use the average of several stochastic estimates of the gradient of the loss function at each iteration:

$$G_{\mu,k} = \frac{1}{m_k} \sum_{i=1}^{m_k} G_\mu(x_k, \xi_{k,i}, v_{k,i}), \qquad (35)$$

where $G_{\mu,k}$ is the approximation of the gradient of the loss function, in the direction opposite to which the algorithm step will be taken at the k-th iteration.

The randomized stochastic projected gradient free (RSPGF) algorithm consists of several steps and is as follows:

1. At the beginning of each iteration of the algorithm, it is necessary to determine the current approximation of the gradient of the loss function $G_{\mu,k}$.
2. Then, it is necessary to calculate the new parameter values – to take a step in the direction of the negative gradient of the loss function.
3. If the new parameter values fall outside the domain, it is necessary to find the projection of the parameter vector onto the domain and use the projection result as the next considered set of parameters.
4. Based on the new parameter values of the model, a new approximated value of the loss function $f(x_{k+1})$ can be calculated.
5. Next, it is necessary to check whether the chosen stopping condition of the method is satisfied. If for the new parameter vector or the new loss function value the stopping condition is not met, it is necessary to proceed to the next iteration of the method and return to step 1.

Formally, the iteration number k of the algorithm can be described as follows:

$$x_{k+1} = \pi_\Omega(x_k - \lambda_k \cdot G_{\mu,k}), \qquad (36)$$

where λ_k is the step size at the $k - th$ iteration, π_Ω is the projection onto the domain of the loss function. The stopping condition of the method can be, for example, that the norm of the gradient of the loss function does not exceed a certain value:

$$||\nabla f(x)|| \le \epsilon. \qquad (37)$$

It has been proven that the randomized stochastic projected gradient free algorithm converges on average at a rate of $O(n\sigma^2/\epsilon^2)$ to an ϵ −stationary point $\bar{x}$ [3]:

$$\mathbb{E}[||g_\Omega(\bar{x})||^2] \le \epsilon, \qquad (38)$$

where n is the number of model parameters, $g_{\Omega}(\bar{x})$ is the generalized projection of the gradient of the function f at the point $\bar{x}$, σ is a constant such that its square bounds the variance of the stochastic approximation of the gradient:

$$\mathbb{E}\left[||G_{\mu} - \nabla f(\mathrm{x})||^2\right] \leq \sigma^2. \quad (39)$$

Moreover, the sufficient conditions for the convergence of the algorithm when using the above method of stochastic gradient approximation are [3]:

1. Lipschitz continuity of the loss function.
2. $L(\xi)$- Lipschitz continuity of the loss function estimate, where ξ are all random daily returns of the asset in all Monte Carlo trajectories.
3. Boundedness of the second moment of the Lipschitz constant of the loss function approximation: $\mathbb{E}[L^2(\xi)] < G^2$.
4. Lipschitz continuity of the gradient of the loss function.
5. Boundedness of the gradient of the loss function.

Additionally, it is necessary that the step size λ_k satisfies the condition

$$\lambda_k \leq \frac{1}{2L\sqrt{(n+4)}}, \quad (40)$$

where L is the Lipschitz constant of the loss function f.

IV. Convergence of the RSPGF Algorithm for the ARMA-GARCH Model Calibration Problem

Theorem. In the case of stationarity of the studied time series in both the original and risk-neutral probability measures, the loss function $f(x)$ in the calibration problem of the risk-neutral ARMA-GARCH time series model for the underlying asset returns using market option prices satisfies the following conditions:

1. The loss function satisfies the Lipschitz condition:
$$|f(y) - f(x)| \leq L||y - x||; \quad (41)$$
2. For fixed values of the random returns, the loss function satisfies the Lipschitz condition:
$$|f(y,\xi) - f(x,\xi)| \leq L(\xi)||y - x||; \quad (42)$$
3. The second moment of the Lipschitz constant $L(\xi)$ is bounded:
$$\mathbb{E}[L^2(\xi)] < G^2; \quad (43)$$
4. The gradient of the loss function satisfies the Lipschitz condition:
$$||\nabla f(y) - \nabla f(x)|| \leq L_{\nabla}||y - x||; \quad (44)$$
5. The gradient of the loss function is bounded:
$$||\nabla f(x)|| \leq M_{\nabla}, \quad \forall x, y \in \Omega, \xi \in \mathcal{Z}, \quad (45)$$

where $\mathcal{Z}$ is the probability space of the random components of the daily returns over the entire considered time interval for all considered trajectories of the underlying asset.

Proof. From the general form (13) of the loss function $f(x)$, taking into account the properties of Lipschitz functions, it is easy to see that a sufficient condition for the validity of the above properties of $f(x)$ is their validity for $S_T(x)$, $S_T(x,\xi)$ and $\nabla S_T(x)$ respectively.

First, let us prove that since $f(x) = \mathbb{E}_{\xi}[f(x,\xi)]$, if $f(x,\xi)$ is Lipschitz continuous with constant $L(\xi)$, then $\mathrm{f}(x)$ satisfies the Lipschitz condition with constant $L = \mathbb{E}_{\xi}[L(\xi)]$. To do this, let us estimate $|f(y) - f(x)|$:

$$|f(y) - f(x)| = \left|\mathbb{E}_{\xi}[f(y,\xi)] - \mathbb{E}_{\xi}[f(x,\xi)]\right|. \quad (46)$$

By the property of expectation, the difference of expectations is equal to the expectation of the difference:

$$\left|\mathbb{E}_{\xi}[f(y,\xi)] - \mathbb{E}_{\xi}[f(x,\xi)]\right| = \left|\mathbb{E}_{\xi}[f(y,\xi) - f(x,\xi)]\right| \leq \mathbb{E}_{\xi}[|f(y,\xi) - f(x,\xi)|].$$

Since $f(x,\xi)$ is Lipschitz continuous with constant $L(\xi)$, it holds that

$$\mathbb{E}_{\xi}[|f(y,\xi) - f(x,\xi)|] \leq \mathbb{E}_{\xi}[L(\xi)|y - x|]. \quad (47)$$

Since $|y - x|$ does not depend on ξ, it can be taken out of the expectation:

$$|f(y) - f(x)| \leq \mathbb{E}_{\xi}[|f(y,\xi) - f(x,\xi)|] \leq \mathbb{E}_{\xi}[L(\xi)] \cdot |y - x|,$$

thus, by definition, $f(x)$ satisfies the Lipschitz condition (41) with constant $\mathbb{E}_{\xi}[L(\xi)]$.

From this, it also follows that, if for fixed values of the random returns the gradient of the loss function satisfies the Lipschitz condition with constant $L_{\nabla}(\xi)$:

$$||\nabla f(y,\xi) - \nabla f(x,\xi)|| \leq L_{\nabla}(\xi)||y - x||, \quad (48)$$

then the gradient of the loss function satisfies the Lipschitz condition (44) with constant $L_{\nabla} = \mathbb{E}_{\xi}[L_{\nabla}(\xi)]$:

$$||\nabla f(y) - \nabla f(x)|| \leq \mathbb{E}_{\xi}[L_{\nabla}(\xi)]||y - x||. \quad (49)$$

Here and below, for any of the model parameters ψ we introduce the notations:

$$Y_t \coloneqq Y_t(\psi,\xi); \widehat{Y_t} \coloneqq Y_t(\widehat{\psi},\xi), \quad (50)$$
$$h_t \coloneqq h_t(\psi,\xi); \widehat{h_t} \coloneqq h_t(\widehat{\psi},\xi), \quad (51)$$
$$m_t \coloneqq \mathrm{m}_t(\psi,\xi); \widehat{m_t} \coloneqq m_t(\widehat{\psi},\xi). \quad (52)$$

Also, note that due to the triangle inequality, it is clear that the above properties hold for $S_T(x)$ if they hold for each of the model parameters individually. To prove the theorem, we will use the following lemmas [6].

Lemma 1. If the functions $f(x)$ and $g(x)$ are bounded by constants M_f and M_g and satisfy the Lipschitz condition with constants L_f and L_g respectively in some domain Ω, then function $f(x) \cdot g(x)$ also satisfies the Lipschitz condition with constant $M_f L_g + M_g L_f$ on that domain Ω.

Lemma 2. If the partial derivative of the function $\frac{\partial f(x,\xi)}{\partial x}$ is bounded by constant $L(\xi)$ on the interval $[a; b]$, then function $f(x,\xi)$ satisfies the Lipschitz condition with constant $L(\xi)$ on this interval.

Corollary. The function $f(x) = e^x$ is Lipschitz continuous on any finite interval $[a; b]$, since its derivative $\frac{df(x)}{dx} = e^x$ is bounded on any finite interval $[a; b]$.

Lemma 3. If functions $f(x)$ and $g(x)$ are bounded by constants M_f and M_g and are Lipsitz continuous with constants L_f and L_g respectively, and the function $g(x)$ is bounded from below by constant m_g, then the function $\frac{f(x)}{g(x)}$ is Lipsitz continuous with constant $\frac{M_f L_g + M_g L_f}{m_g^2}$.

Let us prove the fulfillment of the necessary conditions of the theorem for the model parameter ϕ_0. The proof for the other model parameters is constructed similarly.

First, let us prove the fulfillment of the Lipschitz condition for S_T for any fixed $\varepsilon = (\varepsilon_1, \dots, \varepsilon_T)$. Note that $h_t = \alpha_0 + \alpha_1 h_{t-1} + \beta_1 \epsilon_{t-1}^2$ is bounded from above due to the boundedness of the model parameters. Also h_t is bounded from below by α_0, since the parameters $(\phi_0, \phi_1, \theta_1, \alpha_0, \alpha_1, \beta_1)$ must be non-negative. Also note that $\frac{e^{r\tau} h_t}{(1+m_t)^2}$ is the conditional variance of Y_t:

$$\frac{e^{r\tau} h_t}{(1+m_t)^2} = \mathrm{Var}[Y_t | F_{t-1}], \quad (53)$$

which is bounded due to the stationarity of the studied time series of returns in the risk-neutral probability measure by the condition of the theorem. Let us denote the constant that bounds the conditional variance of the time series Y_t by D_t.

Let us show that the return Y_t satisfies the Lipschitz condition by the mathematical induction method. The Lipschitz condition holds at $t = 0$, since $Y_0 = \widehat{Y}_0 = 0$. Then

$$|Y_t - \widehat{Y}_t| = \left| \varepsilon_t \sqrt{h_t} e^{r\tau} \left(\frac{1}{1+m_t} - \frac{1}{1+\widehat{m}_t} \right) \right|$$

$$= \left| (\widehat{m}_t - m_t) \frac{\varepsilon_t \sqrt{h_t} e^{r\tau}}{(1+m_t)(1+\widehat{m}_t)} \right|$$

The term $\frac{\sqrt{h_t} e^{r\tau}}{(1+m_t)(1+\widehat{m}_t)}$ is bounded, since it represents the product of bounded conditional standard deviations σ_t and $\widehat{\sigma_t}$ of models with parameters ϕ_0 and $\widehat{\phi_0}$ respectively, to the $\sqrt{h_t}$, which is bounded from below:

$$\frac{\sqrt{h_t} e^{r\tau}}{(1+m_t)(1+\widehat{m}_t)} = \frac{\sigma_t \cdot \widehat{\sigma_t}}{\sqrt{h_t}}. \quad (54)$$

Let this term be bounded by some constant $\widehat{D_t}$. Then

$$|Y_t - \widehat{Y}_t| < \varepsilon_t \widehat{D_t} |\widehat{m_t} - m_t| =$$
$$= \varepsilon_t \widehat{D_t} \phi_0 - \widehat{\phi_0} + \phi_1 (Y_{t-1} - \widehat{Y}_{t-1}) <$$
$$< \varepsilon_t \widehat{D_t} |\phi_0 - \widehat{\phi_0}| + \phi_1 \varepsilon_t \widehat{D_t} |Y_{t-1} - \widehat{Y}_{t-1}|.$$

Since $\phi_1 < 1$ and the Lipsitz condition holds for Y_{t-1} by the induction hypothesis with some constant $L_{Y_{t-1}}(\varepsilon_{t-1})$, it also holds for Y_t with constant $L_{Y_t}(\varepsilon_t) = \varepsilon_t \widehat{D_t} \left(1 + \phi_1 L_{Y_{t-1}}(\varepsilon)\right)$.

It is also easy to see that the return Y_t is bounded by some constant $M_{Y_t}(\varepsilon_t)$ because it is given by the sum of a constant $e^{r\tau} - 1$ and the product of the conditional variance, which is bounded due to the stationarity of the considered time series, and the random return $\sqrt{h_t} \left(\frac{e^{r\tau}}{1+m_t} \right) \varepsilon_t$:

$$M_{Y_t}(\varepsilon_t) = e^{r\tau} - 1 + D_t \varepsilon_t. \quad (55)$$

Let us now show the fulfillment of the Lipschitz condition for the loss function with respect to the parameter ϕ_0 and the boundedness of the second moment of the corresponding Lipschitz constant. As mentioned earlier, for this, it is sufficient to show that these properties hold for the price of the base asset at the final time T. Since $S_T = \exp(Y_1 + Y_2 + \dots + Y_T)$, by induction it is easy to see that it is sufficient to show these properties for the function e^{Y_t}. Let us use Lemma 2 and its corollary:

$$|e^{Y_t} - e^{\widehat{Y}_t}| < e^{M_{Y_t}(\varepsilon_t)} |Y_t - \widehat{Y}_t|$$
$$< e^{M_{Y_t}(\varepsilon_t)} L_{Y_t}(\varepsilon_t) |\phi_0 - \widehat{\phi_0}|.$$

Thus, e^{Y_t} satisfies the Lipschitz condition with constant $e^{M_{Y_t}(\varepsilon_t)} \widehat{D_t} \left(1 + \phi_1 L_{Y_{t-1}}(\varepsilon)\right)$. Let us show the boundedness of the second moment of this constant. We will use the Cauchy-Bunyakovsky inequality for expectations [7]:

$$\mathbb{E}\left[\left(e^{M_{Y_t}(\varepsilon_t)} L_{Y_t}(\varepsilon_t)\right)^2\right] = \mathbb{E}\left[e^{2M_{Y_t}(\varepsilon_t)} L_{Y_t}^2(\varepsilon_t)\right] \le$$
$$\le \sqrt{\mathbb{E}\left[e^{4M_{Y_t}(\varepsilon_t)}\right]} \cdot \sqrt{\mathbb{E}\left[L_{Y_t}^4(\varepsilon_t)\right]}.$$

Therefore, it is sufficient to show the boundedness of the fourth moments of $e^{M_{Y_t}(\varepsilon_t)}$ and $L_{Y_t}(\varepsilon_t)$. As mentioned earlier, $M_{Y_t}(\varepsilon_t) = e^{r\tau} - 1 + D_t \varepsilon_t$, which means that $M_{Y_t}(\varepsilon_t)$ is normally distributed with parameters $e^{r\tau} - 1$ and D_t^2. Thus, the fourth moment of $M_{Y_t}(\varepsilon_t)$ is nothing but the value of the moment-generating function of the normal distribution at the point $t = 4$, or which there is an analytical formula:

$$\mathbb{E}\left[e^{4M_{Y_t}(\varepsilon_t)}\right] = e^{4(e^{r\tau}-1) + \frac{16 D_t^2}{2}}, \quad (56)$$

which means that the fourth moment $M_{Y_t}(\varepsilon_t)$ is finite.

Let us now prove the boundedness of the fourth moment of $L_{Y_t}(\varepsilon_t)$ by the method of induction. Since $Y_0 = 0$ for any parameter value ϕ_0, we can take $L_{Y_0} = 1$. Recall that $L_{Y_t}(\varepsilon_t) = \varepsilon_t \cdot L_{Y_{t-1}}(\varepsilon_{t-1})$ up to a constant multiplier, and therefore, by the Cauchy-Bunyakovsky inequality:

$$\mathbb{E}\left[L_{Y_t}^4(\varepsilon_t)\right] = \mathbb{E}\left[\left(\varepsilon_t \cdot L_{Y_{t-1}}(\varepsilon_{t-1})\right)^4\right]$$
$$\le \sqrt{\mathbb{E}[\varepsilon_t^8] \cdot \mathbb{E}\left[\left(L_{Y_{t-1}}(\varepsilon_{t-1})\right)^8\right]},$$

From which we can see that $L_{Y_t}(\varepsilon_t)$ is proportional to the moment of the standard normal distribution of order $8t$. Therefore, the fourth moment of $L_{Y_t}(\varepsilon_t)$ since all moments of the standard normal distribution are bounded.

Thus, it has been shown that the second and fourth moments of the Lipschitz constant for S_T are bounded, which implies that the second moment of the Lipschitz constant of the loss function with respect to the parameter ϕ_0 is also bounded.

Let us show the validity of the Lipschitz condition for the partial derivative $\frac{\partial S_T}{\partial \phi_0}$, which can be defined recursively as follows, using the method of mathematical induction:

$$\frac{\partial S_T}{\partial \phi_0} = \exp(Y_1 + \cdots + Y_t) \cdot \left(\frac{\partial Y_1}{\partial \phi_0} + \cdots + \frac{\partial Y_t}{\partial \phi_0}\right) \quad (57)$$

$$\frac{\partial Y_t}{\partial \phi_0} = \frac{\partial m_t}{\partial \phi_0} \cdot (-1) \frac{e^{r\tau}\sqrt{h_t}}{(1+m_t)^2} \varepsilon_t \quad (58)$$

$$\frac{\partial m_t}{\partial \phi_0} = 1 + \phi_1 \frac{\partial Y_{t-1}}{\partial \phi_0} \quad (59)$$

Let $\frac{\partial Y_{t-1}}{\partial \phi_0}$ satisfy the Lipsitz condition with constant $L_{Y_{t-1}}(\varepsilon_{t-1})$ by the induction hypothesis, then the Lipschitz condition holds for $\frac{\partial m_t}{\partial \phi_0}$ with constant $\phi_1 L_{Y_{t-1}}(\varepsilon_{t-1})$.

Let us show the validity of the Lipschitz condition for the term $\frac{e^{r\tau}\sqrt{h_t}}{(1+m_t)^2}$. First note that $\left|\sqrt{h_t} - \sqrt{\widehat{h_t}}\right| = 0$, since h_t does not depend on ϕ_0, which means that $\sqrt{h_t}$ satisfies the Lipsitz condition. As mentioned earlier, $\sqrt{h_t}$ is bounded from below, $\frac{e^{r\tau}\sqrt{h_t}}{(1+m_t)^2}$ is bounded from above, which means that $\frac{1}{(1+m_t)^2}$ is also bounded from below. Then, by Lemmas 2 and 3, the function $\frac{e^{r\tau}\sqrt{h_t}\varepsilon_t}{(1+m_t)^2}$ is Lipschitz continuous if $(1+m_t)^2$ satisfies the Lipschitz condition. First note that:

$$\begin{aligned} &|(1+m_t)^2 - (1+\widehat{m_t})^2| = \\ &= |(m_t - \widehat{m_t})(2 + m_t + \widehat{m_t})|. \end{aligned} \quad (60)$$

Then since $(1+m_t)^2$ is bounded by some constant $M_m^2(\varepsilon)$ and it has been shown that Y_t satisfies the Lipsitz condition, then

$$\begin{aligned} |(m_t - \widehat{m_t})(2 + m_t + \widehat{m_t})| &\le 2M_m(\varepsilon)|m_t - \widehat{m_t}| = \\ &= 2M_m(\varepsilon)\left|\phi_0 - \widehat{\phi_0} + \phi_1\left(Y_{t-1} - \hat{Y}_{t-1}\right)\right| \le \\ &\le 2M_m(\varepsilon)\left(1 + \phi_1 L_{Y_{t-1}}(\varepsilon)\right)|\phi_0 - \widehat{\phi_0}|. \end{aligned}$$

Which means that $(1+m_t)^2$ is Lipsitz continuous with constant $2M_m(\varepsilon)\left(1 + \phi_1 L_{Y_{t-1}}(\varepsilon)\right)$. Therefore, by Lemma 1, $\frac{\partial Y_t}{\partial \phi_0}$ is also Lipsitz continuous. Thus, using Lemmas 1 and 2, we obtain that $\frac{\partial S_T}{\partial \phi_0}$ is Lipschitz continuous. Consequently, conditions 1–4 of the theorem hold for the loss function $f(x)$ with respect to the parameter ϕ_0.

As mentioned earlier, the proof of the validity of conditions 1–4 for the other model parameters is constructed similarly by induction. Therefore, due to the triangle inequality, since conditions 1–4 of the theorem hold for each of the model parameters, they also hold for the loss function $f(x)$.

Since the domain Ω is compact and the gradient of the loss function is Lipschitz continuous, it is also bounded on this set, which means that condition 5 of the theorem is also satisfied. This completes the proof.

Corollary. The randomized stochastic projected gradient-free algorithm, applied to calibrate the risk-neutral ARMA-GARCH time series model for base asset returns using market option prices, converges on average to an ϵ- stationary point at a rate of $(6\sigma^2/\epsilon^2)$.

V. Conclusion

In this paper, we formulated and proved a theorem on the properties of the loss function in the ARMA-GARCH model calibration problem using market option prices. As a corollary, we established that the randomized stochastic projected gradient-free algorithm converges on average to an ϵ-stationary point at a rate of $\mathcal{O}(6\sigma^2/\epsilon^2)$ for this problem. This result provides a theoretical guarantee for the applicability of the proposed algorithm to such calibration tasks.

References

[1] P. Arbuzov, D. Golembiovskiy, "Calibration of ARIMA-GARCH model of base asset price based on market option quotes", *Economic and Mathematical Methods*, to be published.

[2] J. C. Hull, Options, *Futures and Other Derivative Securities*, 8th ed., Englewood Cliffs, NJ: Prentice Hall, 2011.

[3] Ghadimi, Lan, Zhang, "Mini-batch Stochastic Approximation Methods for Nonconvex Stochastic Composite Optimization", *Mathematical Programming*, vol. 155, pp. 267-305, 2016.

[4] Y. Nesterov, V. Spokoiny, "Random Gradient-Free Minimization of Convex Functions", *Foundations of Computational Mathematics*, vol. 17, pp. 527-566, 2015.

[5] A. R. Danilishin, "Approximation of the Girsanov measure with logarithmic returns in the case of heavy-tailed distributions" in *Proc. ISA RAS*, vol. 73, pp. 21–30, 2023.

[6] V. A. Zorich, *Mathematical Analysis-2*, 9th ed., Moscow: MCNMO Publishing, 2019.

[7] A. N. Shiryaev, *Probability-2*, 3rd ed., Moscow: MCNMO Publishing, 2004.

[8] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed., New York: Springer, 2002.

[9] O. Y. Grachev, "Application of Time Series Models (ARIMA, GARCH, and ARMA-GARCH) for Stock Market Forecasting" Honors Capstone, Dept., 2017, no. 177.

[10] C. Francq, J.-M. Zakoian, *GARCH Models: Structure, Statistical Inference and Financial Applications*, Chichester: John Wiley & Sons Ltd, 2010.

[11] A¨ıt-Sahalia, Y. and Lo, A.W., "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices." *Journal of Finance*, vol. 53, pp. 499-547, 1998.