Preprints.org

Article

Advanced Statistical Analysis of the Predicted Volatility Levels in Crypto Markets

Vadim Azhmyakov [*,‡] , Ilya Shirokov [‡] , Luz Adriana Guzman Trujillo [‡]

Article

Advanced Statistical Analysis of the Predicted Volatility Levels in Crypto Markets

Vadim Azhmyakov [1,*,†] (ID)**, Ilya Shirokov** [1,†]**, and Luz Adriana Guzman Trujillo** [2,†] (ID)

[1] 1ex Corporation, Dubai, United Arab Emirates; vadim.azhmiakov@1ex.com; iyushirokov@1ex.com
[2] LARIS, Université d'Angers, Angers, France; luzadriguz@gmail.com
[*] Correspondence: vadim.azhmiakov@1ex.com
[†] These authors contributed equally to this work.

Abstract: Our paper deals with an advanced statistical tool for the volatility prediction problem in financial (crypto) markets. First, we consider the conventional GARCH-based volatility models. Next, we extend the corresponding GARCH-based forecasting and calculate a specific probability associated with the predicted volatility levels. As the probability evaluation is based on a stochastic model, we develop an advanced data-driven estimation of this probability. The novel statistical estimation we propose uses real market data. The obtained analytical results for the statistical probability of the levels are also discussed in the framework of the integrated volatility concept. The possible application of the established probability estimation approach to the volatility clustering problem is also mentioned. Our paper includes a concrete implementation of the proposed volatility prediction tool and considers a novel trading and volatility estimation module for crypto markets, recently developed by the 1ex Trading Board group in collaboration with GoldenGate Venture. We also briefly discuss the possible application of a model combined with the data-driven volatility prediction methodology to financial risk management.

Keywords: technical analysis; formal volatility models; volatility prediction; statistical probability of levels; trading algorithms

---

1. Introduction and Motivation

In financial engineering, volatility is usually defined as the dispersion of a return series and is computed by taking the (sample) standard deviation (see, e.g., [11,19,32]). Volatility is the most important parameter in the pricing of crypto derivatives, and the trading volume has drastically increased in recent years. To price an option, one needs to know (or estimate) the volatility of the underlying asset from the real-time instant until the option expires. Note that the probability distributions of financial returns are characterized by high volatility persistence and thick tails (see, e.g., [1,12]). In the realm of modern crypto markets, the manifestation of volatility constitutes a very important indicator of the inherent fluctuations in the main market characteristics [13,32,36]. It gives crucial information to crypto traders and constitutes a pivotal element of many effective cryptocurrency trading strategies. Volatility information is very important for assessing risk and pricing derivative products, as well as for developing trading strategies.

The volatility prediction problem is an important topic in the technical analysis of financial markets (see [9,10,20,32,37,40,41], and the references therein). It is common knowledge that various time-series models are widely used for handling the data of highly volatile financial markets. For example, the relatively simple generalized autoregressive conditional heteroskedasticity (GARCH) models can effectively be applied to the volatility forecasting problem (see [9,10,16,22,38,39,41]). On the other hand, the majority of the existing time-series techniques for volatility prediction involve the celebrated quasi-maximum likelihood (QML) estimation approach [7,10,17,18,33]. One can use this conventional estimation methodology for necessary parameter identification in a selected GARCH (p,q) model. We refer to [2,28,33] for the general identification theory.

However, the conventional QML estimation approach mentioned above was initially designed under the classic normality assumption [17,32,33,39]. Thus, it is inefficient if the volatility proxies are non-Gaussian (see [14]). Otherwise, there are several specific features of financial market volatility

that are well-documented. These "stylized facts" include the fat-tailed distributions of asset returns, volatility clustering, asymmetry, and mean reversion. It is well-known that the QML method is non-robust in the presence of outliers, even with fat-tailed and skewed distributions. Therefore, the classic (Gaussian) QML method for the GARCH parameter identification problem needs to be improved and extended with some additional analytical tools.

Although the development of mathematical models for crypto markets and new trading algorithms has been a major topic of research, forecasting financial market volatility is more difficult. Surprisingly, although high volatility can pose a considerable menace to shareholders, it can also be a source of significant financial returns. Even when stock markets oscillate, fall, or skyrocket, there is always a possibility to profit if market volatility is exploited.

In this paper, we extend the existing GARCH volatility prediction technique using an additional useful tool. We introduce predicted volatility levels and calculate the probability that financial market volatility will not fall below these (predicted) levels. This formal mathematical technique involves the well-known stochastic volatility model (see, e.g., [32]). Additionally, we use real market data and develop a constructive lower estimation for the probability evaluation mentioned above. This data-driven estimation of the probability of volatility levels integrates the model parameters and real market data. The advanced statistical analysis we propose can be considered for integrated volatility frameworks. It can also be useful in the context of the volatility clustering problem [29,30]. Our paper includes a short presentation of a practically oriented volatility analysis module, recently developed by the 1ex Trading Board group in collaboration with GoldenGate Venture. This novel tool includes the volatility prediction methodology discussed in this paper and some related trading algorithms.

Let us also note that robust volatility forecasting plays an important role in financial risk management [25,36,42]. Various hedge funds, banks, financial groups, and trading houses use the well-known value-at-risk (VaR) indicators. Modern VaR estimators essentially use some volatility predictors. Moreover, a credible volatility prediction scheme can also be applied to optimize the design of novel profitable trading algorithms for crypto markets [3–6,15]. It is well-known that the highly fluctuating crypto exchange prices and the corresponding highly frequent changes in the main market indicators make accurate price forecasting nearly impossible. In this situation, consistent volatility prediction can essentially improve concrete trading strategies.

The remainder of our paper is organized as follows: Section 2 contains the formal volatility prediction problem formulation in the framework of the general GARCH model. We also examine concrete GARCH abstractions, discuss some useful mathematical and financial facts, and introduce the concept of the predicted volatility levels. Section 3 includes a critical consideration of the conventional QLM technique for model (parameter) identification in a general GARCH model. We focus our attention on the conceptual difficulties of this widely used methodology in the case of non-Gaussian stochastic errors. The criticism in this section helps in understanding the necessity of some additional and novel predictive tools. Section 4 is devoted to the development of a novel probabilistic tool for volatility prediction. We use the well-known stochastic volatility model for this purpose. The application of an advanced mathematical technique makes it possible to calculate the characteristic probability associated with the predicted volatility levels. Concretely, we evaluate the exact probability of financial market volatility not falling below a prescribed (predicted) level. In Section 5, we perform a statistical analysis of financial market data and derive a novel, lower estimation of the probability associated with predicted volatility. This data-driven version of the probability estimation is a formal consequence of the stochastic volatility model studied in the previous section. Section 6 contains a short description of the practically oriented volatility estimation module "AI NEWS," recently developed by the 1ex Trading Board group in collaboration with GoldenGate Venture. Section 7 summarizes our paper.

2. Volatility Prediction in Financial Markets Using GARCH Models

Consider a series $\{p_s\}_{s=0,1,...,T}$ of prices of an asset at $(T+1)$ time points and introduce the corresponding logarithmic return (log return):

$$r_s := \ln(p_s/p_{s-1}),$$
$$s = 1, ..., T.$$

Using the obtained data set, $\{r_s\}_{s=1,...,t}$, we now define the (sample) volatility, σ_t^2, for the given time period:

$$\sigma_t^2 = \frac{1}{t-1} \sum_{s=1}^{t} (r_s - \bar{r}_t)^2. \tag{1}$$

Here,

$$\bar{r}_t := \frac{1}{t} \sum_{s=1}^{t} r_s$$

is a sample mean return. Evidently, (1) constitutes a method of moment unbiased estimation of the second moment (i.e., variance of return) for the observable series of returns, $\{r_s\}_{s=1,...,t}$. Here, we do not assume the covariance stationarity of $\{r_s\}_{s=1,...,t}$. The complete time period $s = 0, 1, ..., T$ can be interpreted as a full time frame associated with a complete series of historical prices.

Note that there are a number of theoretical and practical advantages to using log returns in finance (see, e.g., [11,19]). In financial engineering, volatility is often defined as the square root of (1) (the sample standard deviation); however, the square root of σ_t^2 in (1) constitutes a biased estimation of the corresponding standard deviation (see, e.g., [33,39] for details).

As mentioned in the introduction, there are various methods for estimating the volatility σ_{t+1}^2 in (1). Next, we focus our attention on a simple but effective volatility forecasting procedure that uses a relatively simple GARCH model (see [16,19], and the references therein). Recall that the generic GARCH(p,q) abstraction has the following formal expression:

$$r_t = \mu_t + \epsilon_t, \ t \in \mathbb{N}$$
$$\epsilon_t = \tilde{\sigma}_t z_t \tag{2}$$
$$\tilde{\sigma}_t^2 = \omega_t + \Sigma_{i=1}^{q} \alpha_t^i \epsilon_{t-i}^2 + \Sigma_{j=1}^{p} \beta_t^j \tilde{\sigma}_{t-j}^2.$$

Here,

$$\mu_t := E[r_t | \mathcal{F}_{t-1}]$$

is a conditional mean and $\mathcal{F}_{t-1}$ is a sigma-algebra generated by the data that are available up to the time instant $t-1$. Moreover, the possibly non-normal random variables, z_t, $t \in \mathbb{N}$, are assumed to be independent and identically distributed with

$$E[z_t | \mathcal{F}_{t-1}] = 0,$$
$$E[z_t^2 | \mathcal{F}_{t-1}] = 1.$$

We also assume that the GARCH(p,q) coefficients

$$(\alpha_t^1, ..., \alpha_t^q),$$
$$(\beta_t^1, ..., \beta_t^p)$$

in (2) are known for every $t \in \mathbb{N}$. The deterministic value, ω_t, in model (2) is sometimes called an intercept. Note that we consider a general non-stationary GARCH model here. Evidently, the above non-stationary abstractions constitute an adequate modeling framework for modern, highly volatile crypto markets.

As the statistical properties of the sample mean, $\bar{r}_t$, in (1) make it a very inaccurate approximation of a true mean, taking the necessary deviations around zero instead of $\bar{r}_t$, as in formulae (1), increases the accuracy of the volatility prediction. Therefore, we next assume that $\mu_t \equiv 0$ for all $t \in \mathbb{N}$ in (2).

The conditional variance $\tilde{\sigma}_t^2$ in the GARCH(p,q) model (2) constitutes a specific model-based volatility estimation. The main idea of the proposed GARCH model is that the conditional variance of returns has an autoregressive structure and is positively correlated to its own recent past. Note that this model also generates the volatility clustering effect.

In the case of a GARCH (1,1), we obtain

$$\tilde{\sigma}_t^2 = \omega_t + \alpha_t^1 \epsilon_{t-1}^2 + \beta_t^1 \tilde{\sigma}_{t-1}^2. \tag{3}$$

From (3), next, we derive the model-based unconditional variance estimation, $\tilde{\sigma}_{t+1}^2$, of the return r_{t+1}:

$$\tilde{\sigma}_{t+1}^2 = \omega_{t+1} + (\alpha_{t+1}^1 + \beta_{t+1}^1)\tilde{\sigma}_t^2, \ t \in \mathbb{N}. \tag{4}$$

The resulting volatility prediction expression (4) has a recursive nature. In the stationary case, namely, for

$$\alpha_t^1 \equiv \alpha > 0, \ \beta_t^1 \equiv \beta > 0,$$
$$\omega_t \equiv \omega > 0$$
$$1 - \alpha - \beta > 0$$

we evidently have the explicit time-invariant volatility prediction:

$$\tilde{\sigma}_{t+1}^2 = \tilde{\sigma}^2 = \frac{\omega}{1 - \alpha - \beta}. \tag{5}$$

The so-called volatility persistence is given here by $(\alpha + \beta)$.

Let us also present the resulting formulae for the estimated volatility, $\tilde{\sigma}_{t+1}^2$, of return associated with a stationary GARCH (2,2) predictive model

$$\tilde{\sigma}_t^2 = \omega + \alpha^1 \epsilon_{t-1}^2 + \beta^1 \tilde{\sigma}_{t-1}^2 + \alpha^2 \epsilon_{t-2}^2 + \beta^2 \tilde{\sigma}_{t-2}^2. \tag{6}$$

The corresponding volatility prediction in that case can be expressed as follows:

$$\tilde{\sigma}_{t+1}^2 = \tilde{\sigma}^2 = \frac{\omega}{1 - (\alpha^1 + \beta^1) - (\alpha^2 + \beta^2)}. \tag{7}$$

We refer to [16] for the necessary mathematical formalism. Recall that the covariance-stationarity condition for the general GARCH(p,q) process (2) has the generic form:

$$\Sigma_{i=1}^q \alpha_t^i + \Sigma_{j=1}^p \beta_t^j < 0.$$

Next, using a "predicted volatility level," we denote the value $\tilde{\sigma}_{t+1}^2$, calculated using one of the above formulae (4), (5), or (7). In parallel with the common volatility definition, we also consider the well-known "integrated volatility" concept over the period t to $t + 1$:

$$\Phi_{t+1} := \int_0^1 \sigma_{t+\tau}^2 d\tau. \tag{8}$$

The integrated volatility concept (8) is of central importance in the pricing of crypto derivatives (see, e.g., [29] for details). Let us also refer to [32] for some existing integrated volatility estimators. Evidently, the GARCH-based predicted volatility levels mentioned above naturally imply the corresponding

levels of integrated volatility in (8). For example, the combination of the simple trapezoidal rule and (4) implies the following estimation

$$\tilde{\Phi}_{t+1} \approx \frac{1}{2}(\omega_{t+1} + (1 + \alpha_{t+1}^1 + \beta_{t+1}^1)\tilde{\sigma}_t^2) \tag{9}$$

of Φ_{t+1}.

Note that the integrated volatility, Φ_{t+1}, can also be estimated using the return samplings for a time interval of the sufficiently frequent returns (see, e.g., [1])

$$\lim_{n\to\infty} P[|\Phi_{t+1} - \Sigma_{s=1}^n r_{n,(t+s)/n}^2| \leq \delta] = 0 \; \forall \delta > 0. \tag{10}$$

Here, n is the sampling frequency, $r_{n,(t+s)/n}$ denotes a compound return, and $P[\cdot]$ is the probability associated with the exchange prices (exchange rates) under consideration. The basic relation (10) also involves a useful concept from modern financial engineering; namely, the so-called "realized volatility" (see [20,32] and the references therein). Similar to the forecasting technique for the predicted volatility $\tilde{\sigma}_{t+1}^2$ discussed above, one can also estimate some of the additional important statistical characteristics of return. For example, the GARCH models presented in this section provide a consistent analytic basis for an adequate estimation of the corresponding kurtosis coefficients (see, e.g., [27] for details).

3. On the Critical Analysis of the QML Method for Parameter Identification

The general GARCH-based volatility model (2), as well as the concrete predictive relations (4), (5), (7), and (9), are derived under the assumption of the known model parameters (coefficients)

$$\{(\omega_t, \; \alpha_t^i), \; i = 1, ..., q\}$$

and

$$\{\beta_t^j, \; j = 1, ..., p\}.$$

However, practical application of these approaches involves a necessary identification procedure for defining the GARCH parameters mentioned above. Moreover, one also needs to estimate (identify) the standardized errors, z_t, in model (2).

The quasi-maximum likelihood (QML) method is widely used for the identification of the GARCH(p,q) models (2). We refer to [7,10,17,18,33] for some mathematical details and concrete applications of the QML techniques. Under the standard normality assumptions, this method provides consistent and asymptotically normal estimations in the case of strictly stationary GARCH processes. Recall that the conventional QML involves maximizing the Gaussian log-likelihood, and the resulting solution constitutes an adequate estimation of a parameter vector under the normality assumptions.

However, it is common knowledge that financial time-series (for example, the crypto exchange rates) have the characteristics of being leptokurtotic and fat-tailed with skewness. Moreover, these series usually involve the so-called volatility clustering effect. The "stylized facts" about the financial market volatility mentioned above also include asymmetry and mean reversion. Note that these properties of the volatility dynamics are now well documented (see, e.g., [12,32]).

The non-regular behavior of the modern financial markets and the corresponding stylized facts about volatility make it impossible to consider the classical Gaussian assumption for the stochastic errors in the GARCH(p,q) models. Note that the generic normal probability distributions do not involve outliers and are incompatible with the fat-tailed and skewness effects mentioned above. On the other side, it is well-known that the QML method is non-robust in the presence of data outliers generated by fat-tailed and skewed distributions. It is remarkable that in some professional publications and also in practical trading manuals, one still follows inconsistent normality assumptions. As a result, this simplified Gaussian-based modeling approach involves a deficient description and faulty forecasting of the real (crypto) market dynamics.

The above problem of an adequate modeling framework for the stochastic errors in the general GARCH(p,q) model (2) is crucially important for the resulting model-based volatility prediction. The basic QML estimation is inefficient if the volatility proxies are non-Gaussian (see, e.g., [32] and the references therein). As a consequence, one will obtain a possible inconsequential estimation, $\hat{\theta}_t$, for the parameters of the basic GARCH volatility model under consideration.

In this situation, one can consider some concrete fat-tailed and skewed probability distribution in order to examine and simulate the more realistic case studies of the modeled volatility dynamics. For example, one can use the "contaminated" and skewed normal distribution, skewed Student distribution, skewed generalized error distribution, and many others. These non-regular probability distributions generate various types of realistic additive and innovative outliers for the time-series-based modeling of financial time-series. We refer to [2,14,23] for the corresponding research and simulation results.

Let us note that the fat-tailed and skewed distributions constitute an adequate modeling framework in the case of cryptocurrency time-series (see [12]). On the other hand, we usually have no information about a concrete real probability distribution associated with these specific financial series. The same is also true with respect to the series of returns. That means that the concrete non-Gaussian (fat-tailed and skewed) probability distributions of the stochastic errors in the GARCH model (2) are generally unknown.

The above fact constitutes the main motivation for the necessary methodological extension of the existing model-based techniques for volatility prediction and for developing some additional data-driven statistical tools. Next, we propose a novel statistical metric that can be used as an auxiliary analytic tool for the GARCH-based prediction of the volatility levels. This metric involves real market data and constitutes a quantitative method for seeing how well the model-based volatility prediction would have done. The novel methodology we introduce in the next sections can also be implemented as a part of the common backtesting procedures for the design and verification of new algorithmic trading strategies.

4. Exact Probability Calculation for the Predicted Volatility Levels

This section presents a useful result that can be applied to the formal probabilistic analysis of the predicted volatility levels. By taking into consideration the conceptual difficulties of the OML method discussed in Section 3, we propose an auxiliary statistical-based predictive metric.

Consider the well-known stochastic volatility model (see, e.g., [32] and the references therein):

$$d\sigma_t^2 = (\mu_v - \beta\sigma_t^2)dt + \sigma_v\sigma_t^2 dW_v(t),\ t \in \mathbb{R}_+, \tag{11}$$

where μ_v is an average, β is the speed of the volatility process, and σ_v is called "volatility of volatility". Using $W_v(t)$, we denote a Wiener process with

$$W_v(0) = 0.$$

The above stochastic volatility model is usually considered in combination with the price dynamics:

$$dp_t = \mu_p p_t dt + \sigma p_t dW_p(t),\ t \in \mathbb{R}_+, \tag{12}$$

Similar to (11), we are dealing with an associated Wiener process here:

$$W_p(t),\ W_p(0) = 0.$$

Using μ_p in (12), we denote the mean. Many useful mathematical models of financial markets include the generic abstraction (12). Let us mention the classic Samuelson pricing model and the celebrated Black-Scholes theory (see [8,35]).

Note that the price and the volatility models (11)–(12) constitute an interconnected system of equations. This natural interconnection can be described by a specific correlation between $W_v(t)$ and $W_p(t)$. The fundamental system (11)–(12) can also be used for modeling the crypto markets. This model also generates some stylized facts about the volatility mentioned in the previous sections; namely, the Black–Scholes volatility smile and volatility clustering. We refer to [8] for further technical details.

We now consider the GARCH(1,1) and GARCH(2,2) models from Section 2 and the corresponding predicted volatility levels (4), (5), and (7). Let v_t^* be a required volatility level that is associated with the corresponding GARCH-based predictions. Roughly speaking, we have here the non-stationary level v_t^* such that

$$v_t^* < \tilde{\sigma}_t^2, \ \forall t \in \mathbb{R}_+$$

for (4), and the stationary level

$$v^* < \tilde{\sigma}^2$$

in the case of (5) and (7). The required probability associated with the predicted volatility level, v_t^*, can now be defined as follows:

$$P[\sigma_t^2 \in (v_t^*, +\infty) \mid \sigma_0^2 > v_t^*], \ \forall t \in [0, T]. \tag{13}$$

where $T \in \mathbb{N}$. Recall that $P[\cdot]$ denotes the probability measure associated with the exchange prices, p_t (see Section 2). Note that the above diffusion Markov processes—namely, processes (11) and (12)—are assumed to be defined in the same probability space.

The proposed definition (13) expresses a probability that the volatility does not fall below a specific level (a level of v_t^*) for $t \in [0, T]$. This constitutes a kind of "consistency" for the model-based volatility prediction concept determined by GARCH(1,1) or GARCH(2,2). Note that in (13), we are dealing with a conditional probability and assuming that the initial volatility, σ_0^2, is higher than v_t^*. We now calculate the required probability determined by (13) using only the stochastic volatility abstraction discussed above. Note that this calculation does not involve the GARCH(p,q) volatility model.

In parallel with the characteristic probability (13), we introduce the formal complement:

$$\rho(t, \sigma_t^2) := 1 - P[\sigma_t^2 \in (v_t^*, +\infty) \mid \sigma_0^2 > v_t^*], \ \ \forall t \in [0, T], \tag{14}$$

where the stochastic dynamics of σ_t^2 are given by (11). This complement expresses the probability of the following "complementary" event

$$\sigma_t^2 \le v_t^*$$

for $t < T$. The complementary probability given by (14) can now be evaluated. From the abstract result of [31], it follows that function $\rho(t, \sigma_t^2)$ in (14) satisfies the following boundary value problem:

$$\begin{aligned} &\frac{\partial \rho}{\partial t} = (\mu_v - \beta\sigma_t^2)\frac{\partial \rho}{\partial x} + \frac{1}{2}\sigma_v^2\sigma_t^4\frac{\partial^2 \rho}{\partial x^2}, \ \ t > 0, \ \sigma_t^2 > v_t^*, \\ &\rho(t, v_t^*) = 1, \ \ t \ge 0, \\ &\rho(0, \sigma_t^2) = 0, \ \ \sigma_t^2 > v_t^*, \end{aligned} \tag{15}$$

where σ_t^2 is a solution of (11). From [31], we also deduce that the auxiliary function

$$\xi(t, \sigma_t^2) := (1 - \rho(t, \sigma_t^2)) \exp\frac{1}{\sigma_v^2\sigma_t^4}(\mu_v - \beta\sigma_t^2)(\sigma_t^2 - v_t^*) + \frac{1}{2\sigma_v^2\sigma_t^4}(\mu_v - \beta\sigma_t^2)^2 t$$

satisfies the boundary value problem for the conventional heat equation:

$$\frac{\partial \xi}{\partial t} = \frac{1}{2}\sigma_v^2\sigma_t^4\frac{\partial^2 \xi}{\partial \sigma_t^4}, \ t > 0, \ \sigma_t^2 > v_t^*,$$

$$\xi(t, v_t^*) = 0, \ t \geq 0,$$

$$\xi(t, \sigma_t^2 - v_t^*) = \exp\frac{1}{\sigma_v^2\sigma_t^4}(\sigma_t^2 - v_t^*), \ \sigma_t^2 > v_t^*. \tag{16}$$

The solution of the boundary value problem (16) can be written as follows (see, e.g., [26]):

$$\xi(t, \sigma_t^2) = \exp\left\{\frac{1}{2\sigma_v^2\sigma_t^4}(\mu_v - \beta\sigma_t^2)^2 t\right\} \times$$

$$\left[\psi\left(\frac{(\mu_v - \beta\sigma_t^2)t + (\sigma_t^2 - v_t^*)}{\sigma_v\sigma_t^2\sqrt{t}}\right) \times \exp\left\{\frac{1}{\sigma_v^2\sigma_t^4}(\mu_v - \beta\sigma_t^2)(\sigma_t^2 - v_t^*)\right\} - \right.$$

$$\left. \psi\left(\frac{(\mu_v - \beta\sigma_t^2)t - (\sigma_t^2 - v_t^*)}{\sigma_v\sigma_t^2\sqrt{t}}\right) \times \exp\left\{-\frac{1}{\sigma_v^2\sigma_t^4}(\mu_v - \beta\sigma_t^2)(\sigma_t^2 - v_t^*)\right\}\right]. \tag{17}$$

Here, we use the following notation for the auxiliary function:

$$\psi(\zeta) := \frac{1}{2\pi}\int_{-\infty}^{\zeta}\exp\left\{-\frac{u^2}{2}\right\}du,$$

where

$$\zeta := \mu_v - \beta\sigma_t^2.$$

We now consider the definition (17) of the auxiliary function $\xi(\cdot, \cdot)$ and obtain an explicit expression for the original function $\rho(\cdot, \cdot)$ in problem (15). This expression implies the corresponding formal result for the desired probability in (13):

$$P[\sigma_t^2 \in (v_t^*, +\infty) \mid \sigma_0^2 > v_t^*] = 1 - \rho(t, \sigma_t^2) =$$

$$1 + \psi\left(\frac{(\mu_v - \beta\sigma_t^2)t - (\sigma_t^2 - v_t^*)}{\sigma_v\sigma_t^2\sqrt{t}}\right) \times exp\left\{-\frac{2}{\sigma_v^2\sigma_t^4}(\mu_v - \beta\sigma_t^2)(\sigma_t^2 - v_t^*)\right\} -$$

$$\psi\left(\frac{(\mu_v - \beta\sigma_t^2)t + (\sigma_t^2 - v_t^*)}{\sigma_v\sigma_t^2\sqrt{t}}\right), \tag{18}$$

where $t \in [0, T]$.

A direct verification shows that the obtained complementary to (18)—namely, the probability $\rho(\cdot, \cdot)$—satisfies the boundary value problem (15). Moreover, for the function $\rho(\cdot, \cdot)$ and for the desired probability in (18), we can verify the natural condition

$$0 \leq P[\sigma_t^2 \in (v_t^*, +\infty) \mid \sigma_0^2 > v_t^*] \leq 1,$$

for $t \in [0, T]$.

We now conclude that the probability of the GARCH-based predicted volatility level, v_t^*, is explicitly given by the resulting relation (18). It expresses the probability of the event

$$\sigma_t^2 \in (v_t^*, +\infty)$$

for $[0, T]$, assuming that the initial volatility (for $t = 0$) satisfies the inequality condition

$$\sigma_0^2 > v_t^*.$$

Finally, let us note that the exact probability calculus developed in this section is based on the generic stochastic modeling approach (11) for volatility dynamics.

5. Statistics of the Predicted Volatility Levels

The probabilistic analysis of the predicted volatility levels performed in the previous section is based on an abstract mathematical model; namely, on the stochastic equation (11). Next, we consider this obtained theoretical technique and use it for an applied, data-driven statistical analysis of the GARCH-based estimations of volatility levels.

Consider the resulting formulae (18) from Section 4 and put $t = T$. Our aim is to derive a lower estimation of the probability expressed in (18). We examine it for the constant predicted volatility levels, $\tilde{\sigma}^2$, in (5) (GARCH(1,1) model) and (7) (GARCH(2,2) model). For a constant volatility level, $v^* = $ const, with

$$v^* < \tilde{\sigma},$$

we obtain

$$
\begin{aligned}
&P[\sigma_t^2 \in (v^*, +\infty) \mid \sigma_0^2 > v^*] \geq \\
&1 + \psi\left(\frac{(\mu_v - \beta\tilde{\sigma}^2)T - (\tilde{\sigma}^2 - v^*)}{\sigma_v \tilde{\sigma}^2 \sqrt{T}}\right) \times exp\left\{-\frac{2}{\sigma_v^2 \tilde{\sigma}^4}(\mu_v - \beta\tilde{\sigma}^2)(\tilde{\sigma}^2 - v^*)\right\} - \\
&\psi\left(\frac{(\mu_v - \beta\tilde{\sigma}^2)T + (\tilde{\sigma}^2 - v^*)}{\sigma_v \tilde{\sigma}^2 \sqrt{T}}\right).
\end{aligned}
\tag{19}
$$

Here, $t \in [0, T]$.

We now examine the limit value of the probability expression in (19) for $T \to \infty$. We next interpret the resulting P_∞ as a probability that in the "foreseeable future", the volatility does not fall below the prescribed (constant) level, v^*. As

$$\lim_{\zeta \to \infty} \psi(\zeta) = 1,$$

we obtain

$$P_\infty \geq exp\left\{-\frac{2}{\sigma_v^2 \tilde{\sigma}^4}(\mu_v - \beta\tilde{\sigma}^2)(\tilde{\sigma}^2 - v^*)\right\}. \tag{20}$$

Coming back to a real financial (crypto) market data, we introduce the number $M \in \mathbb{N}$, defined as follows:

$$M := \sum_{t \in \mathbb{N}} \mathbf{1}(\sigma_t^2 \geq v^*), \tag{21}$$

where σ_t^2 is the real market volatility. Function $\mathbf{1}(\cdot)$ in (21) is a generic indicator function

$$\mathbf{1}(y) = 1 \text{ if } y \text{ is true};$$
$$\mathbf{1}(y) = 0 \text{ if } y \text{ is false};$$

for the discrete time $t \in \mathbb{N}$. Roughly speaking, the number $M \in \mathbb{N}$ indicates how many times the real market volatility σ_t^2 is higher as a given constant level, v^*. Using (20) and the basic properties of the exponential function, we deduce our final lower estimation of the limiting probability, P_∞:

$$P_\infty \geq M^{-2(\mu_v - \beta\tilde{\sigma}^2)(\tilde{\sigma}^2 - v^*)/\sigma_v^2 \tilde{\sigma}^4}. \tag{22}$$

Recall that the value $\tilde{\sigma}^2$ in (22) is determined by the stationary GARCH (1,1) predictive model (formulae (5)) or by the stationary GARCH(2,2) predictive model (formulae (7)).

The obtained final probability estimation (22) constitutes a data-driven (statistical) estimation of the probability that the volatility does not fall below a prescribed level v^*. The number M (determined above) describes the real behavior of the market volatility, σ_t^2. This number can be obtained from a concrete historical market data set. The same data set can also be used for the identification of the necessary parameters μ_v, β, and σ_v of the stochastic volatility model (11).

Consider now a simple example. In the simplified case

$$\mu_v = 0,$$
$$\beta = 1,$$
$$\sigma_v = 1,$$

we obtain the following illustrative version of the general estimation (22):

$$P_\infty \geq M^{2(\tilde{\sigma}^2 - v^*)/\tilde{\sigma}^2}. \tag{23}$$

Note that there is no loss of generality in example (23), due to the parametric scalability of the obtained lower estimation (22).

The historical market data set used above for the evaluation of estimation (22) can also be applied to the so-called in-sample forecasting technique. This in-sample method can now be combined with the complementary out-of-sample testing. This approach is methodologically similar to the main idea of the celebrated Monte Carlo method (see [4,21,33,34] and the references therein).

The exact probability evaluation (18) from Section 4 as well as the corresponding data-driven statistical estimations (22) and (23) can also be performed in the context of the integrated volatility (8). One can use the exact value Φ_{t+1} of the integrated volatility in (8) or the simple approximate formulae (9) for $\tilde{\Phi}_{t+1}$ and define the exact probability or the corresponding statistical estimation similar to (18) or (22), respectively. As many modern financial risk indicators, for example, the well-known value-at-risk (VaR) indicators, use volatility predictors, the probabilistic analysis of the predicted volatility levels presented in Section 4 and Section 5 can also be applied to modern risk management.

6. Some Practical Implementations

We now discuss a concrete practical implementation of the statistical analysis for the volatility forecasting methodology developed in our paper. Concretely, we present the volatility prediction module for crypto markets recently developed by the 1ex Trading Board group (see https://1ex.com/) in collaboration with GoldenGate Venture. This novel analytic tool, called "AI NEWS," has an interconnected structure that includes a generic GARCH(2,2) model and the associated statistical probability estimation (22) developed in the previous section.

The conceptual block diagram of the 1ex Trading Board volatility forecasting algorithm for the AI NEWS module mentioned above is given in Figure 1. Note that the three grouped sub-modules in Figure 1 represent a statistical block which complements the conventional GARCH-based volatility analysis and constitutes the main contribution of our paper.
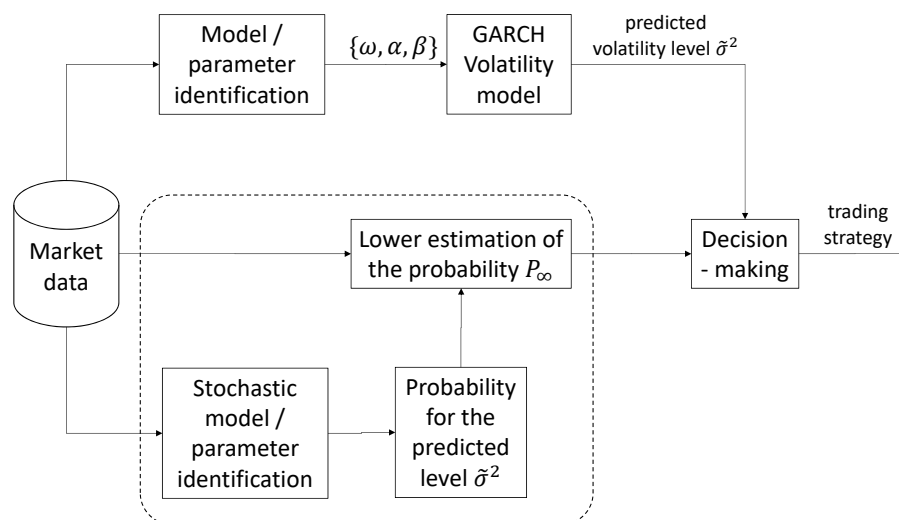
Figure 1. The conceptual block diagram of the 1ex Trading Board volatility prediction module.

This block diagram represents the necessary data and information flows, as well as the system inputs, outputs, and the operational sub-modules. The given sub-modules are necessary for the calculation of the predicted volatility levels and for the evaluation of the statistical estimation (22). The output of the presented block diagram constitutes a final trading decision. The current market data, as well as the involved volatility prediction and statistical probability of the levels, can be visualized by the designed module.

We now use the AI NEWS methodology mentioned above and depict (Figure 2) the market dynamics, the volatility prediction profile, and the corresponding statistical probability level for the concrete crypto trading pair BTC/USDT (date: 10 April 2024).

Figure 2. The 1ex Trading Board volatility prediction module: BTC/USDT.

As one can see, the estimated statistical probability associated with the selected volatility level in that example is equal to 0.98.

Let us also apply the developed predictive statistical analysis to an alternative example; namely, to the trading pair SOL/USDT (date 09/04/2024). The corresponding market dynamics, volatility prediction profile and the calculated statistical probability level for this pair are shown in Figure 3.

Figure 3. The 1ex Trading Board volatility prediction module: SOL/USDT.

The statistical analysis of the predicted volatility levels studied in this section has a potential to be applied to the risk management. The 1ex volatility tool mentioned above helps investors and portfolio managers to fix certain levels of risk which they can bear. An adequate forecast of the market volatility of asset prices over the investment holding period constitutes a very useful initial information for assessing the investment risk.

Finally note that many modern profitable cryptocurrency trading algorithms involve the adequate volatility prediction schemes. Let us refer here to the class of feedback type trading algorithms and also to the widely used family of pullback and drawdown trading strategies (see [3,4,5,6]). The novel methodology of the predicted volatility levels developed in this paper can also be used in the celebrated Avellaneda-Stoikov market making strategy.

7. Concluding Remarks

In this paper, we developed a complementary probabilistic tool for the conventional GARCH-based volatility predictor. The classic GARCH abstractions naturally generate some stationary or non-stationary predicted volatility levels. As mentioned in Introduction, the real probability distributions of financial returns involve high volatility persistence, fat tails, and some additional effects. On the other hand, the widely used QLM parameter estimation methodology is closely related to the (non-realistic) normality assumption and the robust versions of this technique involve some restrictive technical assumptions.

The deficiency of the QLM method mentioned above, as well as the general methodological difficulties of the classic parameter identification approaches, have motivated the development of some additional (complementary) volatility forecasting techniques. From a formal point of view, a GARCH-based volatility predictor constitutes a model-based approach. The main idea of the approach proposed in this paper consists of using a novel data-driven volatility prediction metric that has an auxiliary character for the basic GARCH predictor. In this study, we developed a new additional predictive metric by applying some advanced probabilistic and statistical tools. Roughly speaking, for a model-based predicted volatility level we propose to calculate a statistical probability that the financial market volatility does not fall below this level. We call it the "probability of predicted volatility levels."

As the exact result of computation of the above probability is given by a sophisticated (theoretical) expression, we next extended the proposed approach and calculate a constructive lower estimate of the "probability of predicted volatility levels." The obtained lower estimate involves the necessary real market data and constitutes an implementable robust version of the prediction metric mentioned above. In fact, we finally proposed a combined prediction approach that contains the model-based and the data-driven elements. Moreover, this approach is compatible with some techniques of the celebrated Monte Carlo methodology, in view of the given (historical) data set.

The developed data-driven volatility prediction approach can be used as an auxiliary tool in many analytical concepts of modern financial engineering; for example, it can be studied in the context of the general time-series forecasting. Robust and credible volatility prediction is also an part of many modern cryptocurrency trading algorithms. Moreover, the proposed estimation technique can be applied to the advanced characterization of the integrated volatility and to the important problem of volatility clustering in crypto markets. A robust volatility prediction technique plays a crucial role in modern financial risk management.

Let us also note that the proposed formal calculation of the statistical probability associated with the predicted volatility levels can be implemented not only for the concrete GARCH-based models studied in this paper, as the developed technique is fully compatible with some alternative volatility forecasting approaches. Moreover, the proposed probabilistic and statistical techniques have a general analytic nature and do not depend on some specific financial data or market conditions.

The probabilistic analysis studied in our paper constitutes an initial theoretical development. In our study, we concentrated on some rigorous mathematical details of the proposed prediction schemes.

However, these analytical techniques were implemented in a concrete module for the crypto market volatility forecasting developed by the 1ex Trading Board group in collaboration with GoldenGate Venture. We discussed this practical tool briefly in our paper. The paper does not compare the proposed statistical approach to the existing volatility prediction methods. Due to a very large number of modern theoretical and applied results related to the predicted volatility, we consider a necessary comparative analysis as a future work. The detailed comparative analysis constitutes an important and self-contained topic of a future paper.

Finally, note that the auxiliary statistical analysis of the predicted volatility levels developed in this study can also be considered in the context of novel forecasting approaches based on modern machine learning methodologies [24,37]. It seems to be possible to extend the proposed concepts of volatility levels and the corresponding statistical analysis developed in our paper to volatility forecasting schemes involving deep learning approaches.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. T.G. Andersen, T. Bollerslev, F.X. Diebold, H. Ebens, *The distribution of realized stockr return volatility*, Journal of Financial Economics vol. 61, 2001, pp. 43 — 76.
2. V. Azhmyakov, J. Pereira Arango, M. Bonilla, R. Juarez del Torro and St. Pickl, *Robust state estimations in controlled ARMA processes with the non-Gaussian noises: applications to the delayed dynamics*, IFAC PapersOnline, vol. 54, 2021, pp. 334 – 339.
3. V. Azhmyakov, I. Shirokov, L. A. Guzman Trujillo, *Application of a switched PIDD control strategy to the model-free algorithmic trading*, IFAC PapersOnline, vol. 55, 2022, pp. 145 – 150.
4. V. Azhmyakov, I. Shirokov, Yu Dernov, L. A. Guzman Trujillo *On a data-driven optimization approach to the PID-based algorithmic trading*, Journal of Risk and Financial Management, vol. 16, 2023, pp. 1 – 18.
5. B. R. Barmish and J. A. Primbs, *On a new paradigm for stock trading via a model-free feedback controller*, IEEE Transactions on Automatic Control, vol. 61, 2016, pp. 662 -– 676.
6. M. H. Baumann, *On stock trading via feedback control when underlying stock returns are discontinuous*, IEEE Transactions on Automatic Control, vol. 62, 2017, pp. 2987 -– 2992.
7. J.R. Birge, F. Louveaux, Introduction to Stochastic Programming, Springer, New York, USA, 2011.
8. F. Black and M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy, vol. 81, 1973, pp. 637 — 659.
9. T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics, vol. 31, 1986, pp. 307 – 327.
10. T. Bollerslev, T. Wooldridge, *Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances*, Econometric Reviews, vol. 11, 1992, pp. 143 – 172.
11. C. Brooks, Introductory Econometcs for Finance, Cambridge University Press, Glasgow, UK, 2015.
12. R. Cont, *Empirical properties of asset returns: stylized facts and statistical issues*, Quantitative Finance vol. 1, 2001, pp. 223 — 236.
13. J. Danielsson, M. Valenzuela, I. Zer, *Learning from history: volatility and financial crises*, Review of Finance, vol. 21, 2018, pp. 2774 — 2805.
14. J. Fan, L. Qi, D. Xiu, *Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods*, Journal of Business and Economic Statistics, vol. 32, 2014, pp. 178 — 191.
15. S. Formentin, F. Previdi, G. Maroni and C. Cantaro, *Stock trading via feedback control: an extremum seeking approach*, in: Proceedings of the Mediterranean Conference on Control and Automation, Zadar, Croatia, 2018, pp. 523 – 528.
16. C. Francq, J. Zakoian, GARCH Models, Wiley, Wiltshire, UK, 2010.
17. P. Franses, H. Ghijsels, *Additive outliers, GARCH and forecasting volatility*, International Journal of Forecasting , vol. 15, 1999, pp. 1 – 9.
18. R.G. Gallager, Stochastic Processes, Cambridge University Press, NY, USA, 2013.
19. H. Greene, Econometric Analysis, Pearson, UK, 2011.

20. M.G. Haas, J.P. Franziska, *Implementing intraday model-free implied volatility for individual equities to analyze the return–volatility relationship*, Journal of Risk and Financial Management, vol. 17, 2024, pp. 2 – 19.

21. C. Hammel and W. B. Paul, *Monte Carlo simulations of a trader-based market model*, Physica A, vol. 313, 2002, pp. 640 – 650.

22. D. Huang, H. Wang, Q. Yao, *Estimating GARCH models: when to use what?*, Econometric Journal , vol. 11, 2008, pp. 27 – 38.

23. P.J. Huber, E.M. Ronchetti, Robust Statistics, Wiley, New York, USA, 2005.

24. St. Jansen, Machine Learning for Algorithmic Trading, Packt, Birmingham, UK, 2020.

25. D. Kahneman, A. Tversky, *Prospect theory: An analysis of decision under risk*, in: Handbook of the Fundamentals of Financial Decision Making, World Scientific Publishing, Singapore, 2013.

26. J. Kevorkian, Partial Differential Equations: Analytical Solution Techniques. Texts in Applied Mathematics, USA, New York, Springer, 2000.

27. T.-H. Kim, H. White, *On more robust estimation of skewness and kurtosis*, Finance Research Letters, vol. 1, 2004, pp. 56 — 73.

28. F. L. Lewis, Optimal Estimation, Wiley, New York, USA, 1986.

29. T. Lux, M. Marchesi, *Volatility clustering in financial markets: a microsimulation of interacting agents*, International Journal of Theoretical and Applied Finance, 2000, vol. 3, 2000, pp. 675 -– 702.

30. V. Nikolova, J.E. Trinidad Segovia, M. Fernández-Martínez. M.A. Sánchez-Granero *A novel methodology to calculate the probability of volatility clusters in financial series: an application to cryptocurrency markets*, Mathematics, vol. 8, 2020. pp.

31. L.S. Pontryagin, A.A. Andronov, and A.A. Vitt, *On statistical analysis of dinamical systems*, Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki, vol. 3, 1933, pp. 165 – 180.

32. S.-H. Poon, C.W.J. Granger, *Forecasting volatility in financial markets: a review*, 2003, pp. 478 – 539.

33. A. Poznyak, Advanced Mathematical Tools for Automatic Control Engineers: Stochastic Tools, Elsevier, NY, USA, 2009.

34. R.Y. Rubinstein, Simulation and the Monte Carlo Method, John Wiley Inc., New York, USA, 1981. 1981).

35. P.A. Samuelson, *Rational theory of warrant pricing*, Industrial Management Review, vol. 6, 1965, pp. 13 – 31.

36. W.G. Schwert, *Stock volatility and crash of 87*, Review of Financial studies, 1990, pp. 77 – 102.

37. D. Shah, W. Campbell, F H. Zulkernin, *A comparative study of LSTM and DNN for stock market forecasting*, in: Proceedings of the IEEE International Conference on Big Data, Seattle, USA, 2018, pp 4148 – 4155.

38. J. Sen, S. Mehtab, A. Dutta *Volatility modeling of stocks from selected sectors of the Indian economy using GARCH*, in: Proceedings of the Asian Conference of Innovation in Technology, Pune, India, 2021, pp. 1 – 9.

39. S. Taylor, Modeling Financial Time Series, Wiley, Chichester, UK, 1986.

40. P.B. Verhoeven, B. Pilgram, M. McAleer, A. Mees, *Non-linear modelling and forecasting of S & P 500 volatility*, Mathematics and Computers in Simulation, vol. 59, 2002. pp. 233 – 241.

41. L. Wang, F. Ma, J. Liu, L. Yang, *Forecasting stock price volatility: new evidence from the GARCH-MIDAS model*, International Journal of Forecasting, vol 36, 2020, pp. 684 – 694.

42. W.T. Ziemba, R.G. Vickson, Stochastic Optimization Models in Finance, Academic Press, New York, USA, 1975.