



Article

Advanced Statistical Analysis of the Predicted Volatility Levels in Crypto Markets

Vadim Azhmyakov ^{1,*}, Ilya Shirokov ^{1,†} and Luz Adriana Guzman Trujillo ^{2,†}

¹ 1ex Corporation, Dubai P.O. Box 9305, United Arab Emirates; iyushirokov@1ex.com

² LARIS, Université d'Angers, 49000 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



Citation: Azhmyakov, Vadim, Ilya Shirokov, and Luz Adriana Guzman Trujillo. 2024. Advanced Statistical Analysis of the Predicted Volatility Levels in Crypto Markets. *Journal of Risk and Financial Management* 17: 279. <https://doi.org/10.3390/jrfm17070279>

Academic Editors: Svetlozar (Zari) Rachev and W. Brent Lindquist

Received: 12 February 2024

Revised: 17 April 2024

Accepted: 19 May 2024

Published: 3 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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., Brooks 2015; Greene 2011; Poon and Granger 2003). 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., Andersen et al. 2001; Cont 2001). 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 (Danielsson et al. 2018; Poon and Granger 2003; Schwert 1990). 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 Bollerslev 1986; Bollerslev and Wooldridge 1992; Haas and Peter 2024; Poon and Granger 2003; Shah et al. 2018; Verhoeven et al. 2002; Wang et al. 2020, 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 Bollerslev 1986;

Bollerslev and Wooldridge 1992; Francq and Zakoian 2010; Huang et al. 2008; Sen et al. 2021; Taylor 1986; Wang et al. 2020). On the other hand, the majority of the existing time-series techniques for volatility prediction involve the celebrated quasi-maximum likelihood (QML) estimation approach (Birge and Louveaux 2011; Bollerslev and Wooldridge 1992; Franses and Ghijssels 1999; Gallager 2013; Poznyak 2009). One can use this conventional estimation methodology for necessary parameter identification in a selected GARCH(p,q) model. We refer to Azhmyakov et al. (2021), Lewis (1986), and Poznyak (2009) for the general identification theory.

However, the conventional QML estimation approach mentioned above was initially designed under the classic normality assumption (Franses and Ghijssels 1999; Poon and Granger 2003; Poznyak 2009; Taylor 1986). Thus, it is inefficient if the volatility proxies are non-Gaussian (see Fan et al. 2014). 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., Poon and Granger 2003). 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 (Lux and Marchesi 2000; Nikolova et al. 2020). 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 (Kahneman and Tversky 2013; Schwert 1990; Ziemba and Vickson 1975). 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 (Azhmyakov et al. 2022, 2023; Barmish and Primbs 2016; Baumann 2017; Formentin et al. 2018). It is well known that the highly fluctuating crypto exchange prices and the corresponding very 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 $\{p_s\}_{s=0,1,\dots,T}$ of prices of an asset at $(T + 1)$ time points and introduce the corresponding logarithmic return (log return):

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

$$s = 1, \dots, T.$$

Using the obtained data set, $\{r_s\}_{s=1,\dots,t}$, 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, $\{r_s\}_{s=1,\dots,t}$. Here, we do not assume the covariance stationarity of $\{r_s\}_{s=1,\dots,t}$. The complete time period $s = 0, 1, \dots, T$ 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., Brooks 2015; Greene 2011). 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., Poznyak 2009; Taylor 1986 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 Francq and Zakoian 2010; Greene 2011, and the references therein). Recall that the generic GARCH(p,q) abstraction has the following formal expression:

$$r_t = \mu_t + \epsilon_t, \quad t \in \mathbb{N}$$

$$\epsilon_t = \tilde{\sigma}_t z_t$$

$$\tilde{\sigma}_t^2 = \omega_t + \sum_{i=1}^q \alpha_t^i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_t^j \tilde{\sigma}_{t-j}^2. \tag{2}$$

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

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

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} :

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

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

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

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 [Francq and Zakoian \(2010\)](#) for the necessary mathematical formalism. Recall that the covariance-stationarity condition for the general GARCH(p,q) process (2) has the generic form

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

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., [Lux and Marchesi 2000](#) for details). Let us also refer to [Poon and Granger \(2003\)](#) 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 of Φ_{t+1} :

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

Note that the integrated volatility, Φ_{t+1} , can also be estimated using the return samplings for a time interval of sufficiently frequent returns (see, e.g., [Andersen et al. 2001](#)):

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

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 [Haas and Peter 2024](#); [Poon and Granger 2003](#), 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., [Kim and White 2004](#) 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)

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

and

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

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) model (2). We refer to [Birge and Louveaux \(2011\)](#); [Bollerslev and Wooldridge \(1992\)](#); [Franses and Ghijssels \(1999\)](#); [Gallager \(2013\)](#), and [Poznyak \(2009\)](#) 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., [Cont 2001](#); [Poon and Granger 2003](#)).

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 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, inconsistent normality assumptions are still followed. 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., [Poon and Granger 2003](#) 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 [Azhmyakov et al. \(2021\)](#), [Fan et al. \(2014\)](#), and [Huber and Ronchetti \(2005\)](#) 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 [Cont 2001](#)). 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 performed. 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., [Poon and Granger 2003](#) and the references therein):

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

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:

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

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

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

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 [Samuelson 1965](#); [Black and Scholes 1973](#)).

Note that the price and the volatility models (11) and (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) and (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 [Black and Scholes \(1973\)](#) 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

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

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:

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

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

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

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

$$\sigma_t^2 \leq v_t^*$$

for $t < T$. The complementary probability given by (14) can now be evaluated. From the abstract result of Pontryagin et al. (1933), it follows that function $\rho(t, \sigma_t^2)$ in (14) satisfies the boundary value problem

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

where σ_t^2 is a solution of (11). From Pontryagin et al. (1933), we also deduce that the auxiliary function

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

satisfies the boundary value problem for the conventional heat equation:

$$\begin{aligned} \frac{\partial \zeta}{\partial t} &= \frac{1}{2} \sigma_v^2 \sigma_t^4 \frac{\partial^2 \zeta}{\partial \sigma_t^4}, \quad t > 0, \sigma_t^2 > v_t^*, \\ \zeta(t, v_t^*) &= 0, \quad t \geq 0, \\ \zeta(t, \sigma_t^2 - v_t^*) &= \exp \frac{1}{\sigma_v^2 \sigma_t^4} (\sigma_t^2 - v_t^*), \quad \sigma_t^2 > v_t^*. \end{aligned} \tag{16}$$

The solution of the boundary value problem (16) can be written as follows (see, e.g., Kevorkian 2000):

$$\begin{aligned} \zeta(t, \sigma_t^2) &= \exp \left\{ \frac{1}{2 \sigma_v^2 \sigma_t^4} (\mu_v - \beta \sigma_t^2)^2 t \right\} \times \\ & \left[\psi \left(\frac{(\mu_v - \beta \sigma_t^2) t + (\sigma_t^2 - v_t^*)}{\sigma_v \sigma_t^2 \sqrt{t}} \right) \times \exp \left\{ \frac{1}{\sigma_v^2 \sigma_t^4} (\mu_v - \beta \sigma_t^2) (\sigma_t^2 - v_t^*) \right\} - \right. \\ & \left. \psi \left(\frac{(\mu_v - \beta \sigma_t^2) t - (\sigma_t^2 - v_t^*)}{\sigma_v \sigma_t^2 \sqrt{t}} \right) \times \exp \left\{ -\frac{1}{\sigma_v^2 \sigma_t^4} (\mu_v - \beta \sigma_t^2) (\sigma_t^2 - v_t^*) \right\} \right]. \end{aligned} \tag{17}$$

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 $\zeta(\cdot, \cdot)$ 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):

$$\begin{aligned} P[\sigma_t^2 \in (v_t^*, +\infty) \mid \sigma_0^2 > v_t^*] &= 1 - \rho(t, \sigma_t^2) = \\ & 1 + \psi \left(\frac{(\mu_v - \beta \sigma_t^2) t - (\sigma_t^2 - v_t^*)}{\sigma_v \sigma_t^2 \sqrt{t}} \right) \times \exp \left\{ -\frac{2}{\sigma_v^2 \sigma_t^4} (\mu_v - \beta \sigma_t^2) (\sigma_t^2 - v_t^*) \right\} - \\ & \psi \left(\frac{(\mu_v - \beta \sigma_t^2) t + (\sigma_t^2 - v_t^*)}{\sigma_v \sigma_t^2 \sqrt{t}} \right), \end{aligned} \tag{18}$$

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, $v^* = \text{const}$, with

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

we obtain

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

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

We now examine the limit value of the probability expression in (19) for $T \rightarrow \infty$. 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

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

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,

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

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_∞ :

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

Recall that the value $\bar{\sigma}^2$ 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,

$$\begin{aligned} \mu_v &= 0, \\ \beta &= 1, \\ \sigma_v &= 1, \end{aligned}$$

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

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

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 Azhmyakov et al. 2023; Hammel and Paul 2002; Poznyak 2009; Rubinstein 1981, 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 Sections 4 and 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 (<https://1ex.com/> (accessed on 11 February 2024)) 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.

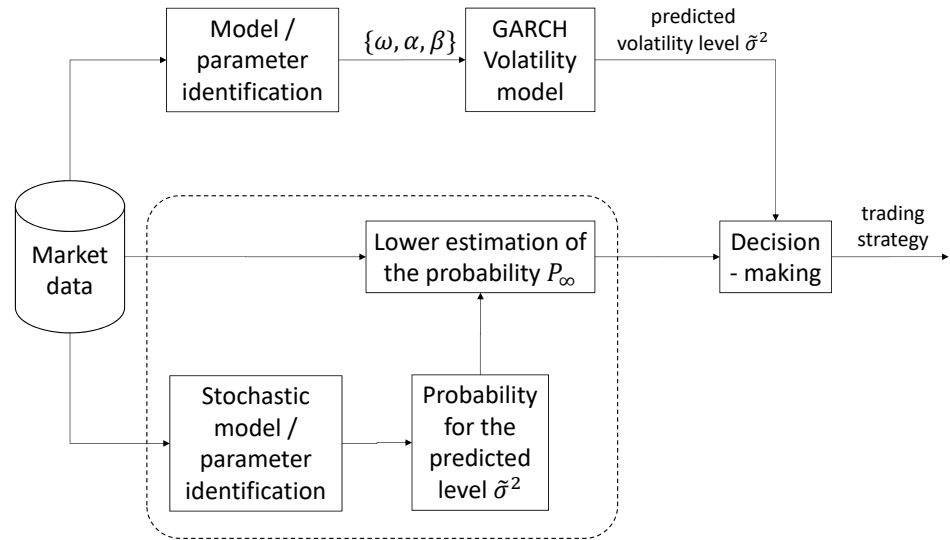


Figure 1. The conceptual block diagram of the 1ex Trading Board volatility prediction module.

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

The statistical analysis of the predicted volatility levels studied in this section has the potential to be applied to 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 very useful initial information for assessing the investment risk.

Finally, note that many modern profitable cryptocurrency trading algorithms involve 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 [Azhmyakov et al. 2022, 2023](#); [Barmish and Primbs 2016](#); [Baumann 2017](#)). 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 the 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 calculating 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 calculated a constructive lower estimate of the “probability of predicted volatility levels”. The obtained lower estimate involved the necessary real market data and constituted 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 general time-series forecasting. Robust and credible volatility prediction is also a 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 with models other than 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 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 predicted volatility, we consider a necessary comparative analysis as a future work. Such a 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 [Jansen \(2020\)](#); [Shah et al. \(2018\)](#). It seems 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.

Author Contributions: Conceptualization, V.A.; Methodology, L.A.G.T.; Software, I.S.; Validation, I.S.; Formal analysis, I.S.; Investigation, L.A.G.T.; Data curation, L.A.G.T.; Writing—review & editing, V.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: Authors Vadim Azhmyakov and Ilya Shirokov are employed by the company 1ex Corporation. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

References

- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61: 43–76. [[CrossRef](#)]
- Azhmyakov, Vadim, Ilya Shirokov, and Luz Adriana Guzman Trujillo. 2022. Application of a switched PID control strategy to the model-free algorithmic trading. *IFAC PapersOnline* 55: 145–50. [[CrossRef](#)]

- Azhmyakov, Vadim, Ilya Shirokov, Yu Dernov, and Luz Adriana Guzman Trujillo. 2023. On a data-driven optimization approach to the PID-based algorithmic trading. *Journal of Risk and Financial Management* 16: 1–18. [\[CrossRef\]](#)
- Azhmyakov, Vadim, Jose Pereira Arango, Moises Bonilla, Raymundo Juarez del Torro, and Stefan Pickl. 2021. Robust state estimations in controlled ARMA processes with the non-Gaussian noises: Applications to the delayed dynamics. *IFAC PapersOnline* 54: 334–39. [\[CrossRef\]](#)
- Barmish, B. Ross, and James A. Primbs. 2016. On a new paradigm for stock trading via a model-free feedback controller. *IEEE Transactions on Automatic Control* 61: 662–76. [\[CrossRef\]](#)
- Baumann, Michael H. 2017. On stock trading via feedback control when underlying stock returns are discontinuous. *IEEE Transactions on Automatic Control* 62: 2987–92. [\[CrossRef\]](#)
- Birge, John R., and François Louveaux. 2011. *Introduction to Stochastic Programming*. New York: Springer.
- Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–59. [\[CrossRef\]](#)
- Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [\[CrossRef\]](#)
- Bollerslev, Tim, and Jeffrey M. Wooldridge. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11: 143–72. [\[CrossRef\]](#)
- Brooks, Chris. 2015. *Introductory Econometrics for Finance*. Glasgow: Cambridge University Press.
- Cont, Rama. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223–36. [\[CrossRef\]](#)
- Danielsson, Jon, Marcela Valenzuela, and Ilknur Zer. 2018. Learning from history: Volatility and financial crises. *Review of Finance* 21: 2774–805.
- Fan, Jianqing, Lei Qi, and Dacheng Xiu. 2014. Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business and Economic Statistics* 32: 178–91. [\[CrossRef\]](#)
- Formentin, Simone, Fabio Previdi, Gabriele Maroni, and Claudio Cantaro. 2018. Stock trading via feedback control: An extremum seeking approach. Paper presented at the Mediterranean Conference on Control and Automation, Zadar, Croatia, June 19–22; pp. 523–528.
- Francq, Christian, and Jean-Michel Zakoian. 2010. *GARCH Models*. Wiltshire: Wiley.
- Franses, Philip, and Hendrik Ghijssels. 1999. Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting* 15: 1–9. [\[CrossRef\]](#)
- Gallager, Robert G. 2013. *Stochastic Processes*. New York: Cambridge University Press.
- Greene, William H. 2011. *Econometric Analysis*. London: Pearson.
- Haas, Martin G., and Franziska J. Peter. 2024. Implementing intraday model-free implied volatility for individual equities to analyze the return–volatility relationship. *Journal of Risk and Financial Management* 17: 2–19. [\[CrossRef\]](#)
- Hammel, C., and Wolfgang B. Paul. 2002. Monte Carlo simulations of a trader-based market model. *Physica A* 313: 640–50. [\[CrossRef\]](#)
- Huang, Da, Hansheng Wang, and Qiwei Yao. 2008. Estimating GARCH models: When to use what? *Econometric Journal* 11: 27–38. [\[CrossRef\]](#)
- Huber, Peter J., and Elvezio M. Ronchetti. 2005. *Robust Statistics*. New York: Wiley.
- Jansen, Stefan. 2020. *Machine Learning for Algorithmic Trading*. Birmingham: Packt.
- Kahneman, Daniel, and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making*. Singapore: World Scientific Publishing.
- Kevorkian, Jirair. 2000. *Partial Differential Equations: Analytical Solution Techniques*. Texts in Applied Mathematics; New York: Springer.
- Kim, Tae-Hwan, and Halbert White. 2004. On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1: 56–73. [\[CrossRef\]](#)
- Lewis, Frank L. 1986. *Optimal Estimation*. New York: Wiley.
- Lux, Thomas, and Michele Marchesi. 2000. Volatility clustering in financial markets: A microsimulation of interacting agents. *International Journal of Theoretical and Applied Finance* 3: 675–702. [\[CrossRef\]](#)
- Nikolova, Venelina, Juan E. Trinidad Segovia, Manuel Fernández-Martínez, and Miguel A. Sánchez-Granero. 2020. A novel methodology to calculate the probability of volatility clusters in financial series: An application to cryptocurrency markets. *Mathematics* 8: 1216. [\[CrossRef\]](#)
- Pontryagin, L. S., A. A. Andronov, and A. A. Vitt. 1933. On statistical analysis of dynamical systems. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki* 3: 165–80.
- Poon, Ser-Huang, and Clive W.J. Granger. 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* 41: 478–539. [\[CrossRef\]](#)
- Poznyak, Alexander. 2009. *Advanced Mathematical Tools for Automatic Control Engineers: Stochastic Tools*. New York: Elsevier.
- Rubinstein, Reuven Y. 1981. *Simulation and the Monte Carlo Method*. New York: John Wiley Inc.
- Samuelson, Paul A. 1965. Rational theory of warrant pricing. *Industrial Management Review* 6: 13–31.
- Schwert, G. William. 1990. Stock volatility and crash of 87. *Review of Financial Studies* 3: 77–102. [\[CrossRef\]](#)
- Sen, Jaydip, Sidra Mehtab, and Abhishek Dutta. 2021. Volatility modeling of stocks from selected sectors of the Indian economy using GARCH. Paper presented at the Asian Conference of Innovation in Technology, Pune, India, August 27–29; pp. 1–9.
- Shah, Dev, Wesley Campbell, and Farhana H. Zulkernin. 2018. A comparative study of LSTM and DNN for stock market forecasting. Paper presented at the IEEE International Conference on Big Data, Seattle, WA, USA, December 10–13; pp. 4148–55.

- Taylor, Stephen. 1986. *Modeling Financial Time Series*. Chichester: Wiley.
- Verhoeven, Peter B., Berndt Pilgram, Michael McAleer, and Alistair Mees. 2002. Non-linear modelling and forecasting of S & P 500 volatility. *Mathematics and Computers in Simulation* 59: 233–41.
- Wang, Lu, Feng Ma, Jing Liu, and Lin Yang. 2020. Forecasting stock price volatility: New evidence from the GARCH-MIDAS model. *International Journal of Forecasting* 36: 684–94. [[CrossRef](#)]
- Ziemba, William T., and Raymond G. Vickson. 1975. *Stochastic Optimization Models in Finance*. New York: Academic Press.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.