GARCHX-NoVaS: A Bootstrap-based approach of forecasting for GARCHX models

Kejin Wu[1], Sayar Karmakar[2], and Rangan Gupta[3]

[1]Department of Mathematics, University of California San Diego
[2]Department of Statistics, University of Florida
[3]Department of Economics, University of Pretoria

Abstract

In this work, we explore the forecasting ability of a recently proposed normalizing and variance-stabilizing (NoVaS) transformation with the possible inclusion of exogenous variables in GARCH volatility specification. From an applied point-of-view, extra knowledge such as fundamentals- and sentiments-based information could be beneficial to improve the prediction accuracy of market volatility if they are incorporated into the forecasting process. In the classical approach, these models including exogenous variables are typically termed GARCHX-type models. Inspired by a Model-free prediction principle, NoVaS has generally shown more accurate, stable and robust (to misspecifications) performance than that compared to classical GARCH-type methods. This motivates us to extend this framework to the GARCHX forecasting as it enlarges the scope of application especially where one conjectures the improvement in prediction performance by including covariates. We derive the NoVaS transformation needed to include exogenous covariates and then construct the corresponding prediction procedure. We show through extensive simulation studies that bolster our claim that the NoVaS method outperforms traditional ones, especially for long-term time aggregated predictions. We also provide an interesting data analysis to exhibit how our method could possibly shed light on the role of geopolitical risks in forecasting volatility in national stock market indices for three different countries in Europe.

Keywords: Volatility forecasting; Bootstrap; GARCH; GARCHX
JEL Codes: C32; C53; C63; Q54

1 Introduction

In the long history of time series econometrics literature, accurate forecasting has always stood out as a fundamental and important problem. It has a range of applications in various industries, e.g., weather forecasting, climate forecasting, and

economic forecasting. Discrete-time series data, e.g., heights of ocean tides, and temperature of a city, is the realization of a stochastic process $\{X_t, t \in \mathbb{Z}\}$. The earliest modern time series analysis could be traced back to the work of Yule (1927) where the pattern of the sunspots number was studied. Unlike the prediction of independent data, the prediction of time series gets more complicated due to the inherent data dependence. To get accurate predictions and inferences, it is crucial to model the dependent relationship within the data. Usually, very generally speaking, the time series data is assumed to be generated by some underlying mechanism as follows:

$$X_t = G(\boldsymbol{X}_{t-p}, \epsilon_t); \tag{1}$$

$G(\cdot, \cdot)$ could be any suitable function; ϵ_t is called innovation and assumed to be $i.i.d.$ with appropriate moments and independent with X_{t-i}, $i \geq 1$; $\boldsymbol{X}_{t-p}$ represents $\{X_{t-1}, \ldots, X_{t-p}\}$ and stands for the historical information. To further simplify the forecasting problem, participators focus on some standard formats of $G(\cdot, \cdot)$, e.g., linear or non-linear. For linear models, such as linear AR, MA and ARMA models, we can apply the Box-Jenkins method of identifying, fitting, checking and predicting models systematically (Box et al., 2015). However, the prediction of non-linear models is not as trivial as the case of linear models since the innovation must be appropriately included in the prediction process, especially for the multi-step ahead predictions; see Wu and Politis (2023) for more related discussions.

In this paper, we are exclusively interested in one non-linear type of Eq. (1) which is the so-called Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) model proposed by Bollerslev (1986) and has a form below:

$$\begin{aligned} Y_t &= \sigma_t W_t, \\ \sigma_t^2 &= a + a_1 Y_{t-1}^2 + b_1 \sigma_{t-1}^2; \end{aligned} \tag{2}$$

where, $a \geq 0$, $a_1 > 0$, $b_1 > 0$, and W_t are usually assumed to be $i.i.d.$ with standard normal distribution in practice. The GARCH model is a generalization of the famous Autoregressive Conditional Heteroskedasticity (ARCH) model proposed by Engle (1982). Its ability to forecast the absolute magnitude and quantiles or entire density of squared financial log-returns (i.e., equivalent to volatility forecasting to some extent) was shown by Engle and Patton (2001) using the Dow Jones Industrial Index. Later, many studies to investigate the performance of different GARCH-type models in predicting volatility of financial series were conducted; see following references (Peters, 2001; González-Rivera et al., 2004; Lim and Sek, 2013; Herrera et al., 2018; Karmakar and Roy, 2021). For ARCH/GARCH-type models, it is usual practice to identify and fit models based on quasi-maximum likelihood inference. However, the distributional assumption on W_t brings unavoidable model misspecification when it is hard to assume the normality of the unobservable innovation process. Later, we will show one advantage of our method is that such a restrictive assumption can be weakened.

Traditionally, economists primarily utilize univariate GARCH-family models to understand dynamics of econometric data such as stock/ index/ price, etc observed for a long time. However, one of the key focuses of financial econometrics is to understand how extra knowledge such as fundamentals- and sentiments-based information could be beneficial to improve the prediction accuracy of market volatility if they are incorporated into the forecasting process; see more discussion from Engle and Patton (2007) and the references therein. When the predictor is incorporated into the GARCH

volatility specification, the model is called GARCHX (X standing for covariate). The estimation methodology of GARCHX models was discussed thoroughly in the work of Francq and Thieu (2019); see more details about the GARCHX model in Section 2.1.

Rather than taking the traditional approach (i.e., specifying and fitting a model and then predicting), we utilize a Model-free prediction principle which was proposed by Politis (2015). In short, the Model-free prediction principle hinges on the idea of applying an inverse transformation function to bridge two equivalent probability spaces. For example, if we observe a univariate time series $\{Y_1, \ldots, Y_T\}$, we can try to find a transformation to map $\{Y_1, \ldots, Y_T\}$ to an *i.i.d.* series $\{Z_1, \ldots, Z_T\}$. Since the prediction of *i.i.d.* data is trivial, we can then transform the prediction of *i.i.d.* data back to the prediction of the original data; see more details about the Model-free prediction principle in Section 2.2. Guided by such a prediction principle, the normalizing and variance-stabilizing transformation (NoVaS transformation) method was developed and showed superior performance compared to other classical counterparts in the context of volatility forecasting. This NoVaS method was initially developed by Politis (2003) and then well discussed under the framework of the Model-free prediction principle in Politis (2015).

In the literature this NoVaS method is usually called Model-free method but a discussion about its nature is in order. Usually, for both estimation and prediction of various types of time-series models, one assumes certain distributions of the *i.i.d.* innovation and often this distribution is taken to be standard normal to facilitate computations for both the steps. Comparatively, our NoVaS based method in this paper only depends on the *i.i.d.* nature of the innovations and relaxes any specific distributional assumption on the innovations. However, we use the specific model structure of GARCHX process to arrive at a suitable transformation that leads to obtaining the i.i.d. transformed variables from the observed data. Therefore, the NoVaS method concerned in this paper is at an intermediate stage between purely Model-free and Model-based approaches and probably a somewhat clearer description is to call it distribution-free; see more discussions in Remark 2.1. To clear up any confusion, we call the NoVaS a *Bootstrap-based* method since its cornerstone is about determining a transformation function and applying the Bootstrap technique to mimic the distribution of future values.

In the huge literature of applied econometrics, while analyzing data observed over a long time that can show signs of heteroscedasticity, the usual practice is to pick some specific GARCH-type model. Next an estimation of that model is carried out accordingly and subsequently the forecast of future volatility will be made based on the estimated model. Apparently, there is no universal rule for the choice of the specific GARCH model. In other words, a specific GARCH model can not work uniformly well across different datasets compared to other variants. On the other hand, the NoVaS prediction method could work well for any scenario as long as a suitable transformation function can be found. Therefore, the model-selection stage is shunned and the NoVaS prediction approach is more robust against the model misspecification. Moreover, the standard GARCH methods require a relatively large sample size to be estimated well. For the NoVaS method, it tends to work stably even with short data. The existence of such transformation function in the context of predicting with exogenous variables will be analyzed theoretically in Section 2.2.

Empirically speaking, NoVaS methods were mainly applied to forecast volatility in financial econometrics in the past few years. Gulay and Emec (2018) showed that the NoVaS method could beat GARCH-type models (GARCH, EGARCH and

GJR-GARCH) with generalized error distributions by comparing the pseudo-out of sample (POOS) forecasting of volatility. Here the POOS forecasting analysis means using data up to and including the current time to predict future values. Later, Chen and Politis (2019) extended the NoVaS method to do multi-step ahead predictions. Wu and Karmakar (2021) further substantiated the great performance of NoVaS methods on time-aggregated long-term (30-steps ahead) predictions. Wang and Politis (2022) applied the Model-free idea to provide estimation and prediction inference for a general class of time series. Our present work is motivated by the Wu and Karmakar (2023) work where the authors recommended a so-called GARCH-NoVaS (GA-NoVaS) transformation structure inspired by the development of GARCH from ARCH. This NoVaS method is significantly robust against different model misspecification. Given this, it was a natural and probably quite an important question to see if such a robust forecasting framework can be built where exogenous covariates can be included and thus improve forecasting accuracy.

In this work, we explore the new methodology of forecasting stock market volatility with additional covariates being available to be included in the volatility dynamics. As far as we know, the NoVaS prediction idea has not been studied when the exogenous variables are featured even in modeling the mean or average let alone the more complicated variance or volatility dynamics of a time-series. Due to the superior performance of the GARCH-NoVaS method in volatility forecasting, for this paper, we stick to the variance part and attempt to further boost the ability of the GA-NoVaS method with the help of exogenous covariate information. Towards this, we propose a so-called GARCHX-NoVaS (abbreviated as GAX-NoVaS henceforth) method which takes the GARCHX model as the starting step to build transformation. To obtain the inference about the future situation at an overall level, we choose the time-aggregated prediction metric. This aggregated metric has been applied to evaluate future predictions of electricity price or financial data (Fryzlewicz et al., 2008; Chudý et al., 2020; Karmakar et al., 2022); see the formal definition in Section 4.1. We wanted to check if the NoVaS prediction method can incorporate exogenous variables. More importantly, we hope the GAX-NoVaS method can sustain its great performance compared to the GARCHX method.

In addition to comparing our NoVaS method and the classical GARCHX model with several simulated datasets, we also provide an interesting real data analysis. Our goal is to exhibit how our method could possibly shed light on the role of geopolitical risks, which are currently engulfing the global economy with multiple wars taking place, in forecasting the volatility of three stock markets of Europe: Germany and its two neighbors (Austria and Switzerland), based on a daily index of uncertainty associated with the Russia-Ukraine war as perceived by German Twitter activity. We also use newspaper-based metrics of global geopolitical risks due to acts and threats, as developed by Caldara and Iacoviello (2022), to check for the robustness of our result covering a longer data sample. Hence, we add from a methodological perspective to the existing literature on forecasting international stock returns volatility using the information contained in geopolitical events and threats that basically rely on GARCH-type models (see, for example, Salisu et al. (2022); Zhang et al. (2023) for details discussion of this literature). In this regard, note that, Caldara and Iacoviello (2022) pointed out that entrepreneurs, market participants, and central bank officials view geopolitical risks as key determinants of investment decisions and stock market dynamics, with such risks, along with economic and policy uncertainties, forming an "uncertainty trinity" that would adversely impact the economy and the financial

sector, as has been traditionally reported in the large existing literature on the impact of terror attacks and threats Mehmet et al. (2018); Christos et al. (2019); Bouri et al. (2020). In our empirical and simulation exercises, we measure the performance of GARCHX and GAX-NoVaS methods by the standard mean square prediction error. Moreover, we apply the forecast comparison tests to compare the two methods in a statistical way.

Our main contributions are summarized as follows:

- We propose a new methodology– namely GAX-NoVaS–to do the volatility forecasting with exogenous variables. This method depends on a transformation function to connect two equivalent probability spaces instead of relying on any restrictive model assumption. The idea behind the GAX-NoVaS method hinges on the Model-free prediction principle.

- Due to the manner of the transformation idea, the standard normality assumption of the GARCH model can be avoided with our method. As a result, our method is more robust against model misspecification. In other words, instead of estimating a GARCH model with the maximum likelihood technique which relies on distributional assumption inescapably, the NoVaS method aims to solve an optimization problem that yields good forecasting results even when particular model assumptions fail.

- To make volatility forecasting with (multiple) covariates X under the framework of the NoVaS method, we show the existence of a transformation function that maps $\{X_i\}_{i=1}^T$ and $\{Y_i\}_{i=1}^T$ to $i.i.d.$ $\{Z_i\}_{i=1}^T$ which possesses a simple distribution under some mild conditions. This serves as the theoretical foundation of our method. Our NoVaS method is further a practically useful approximation to this oracle transformation function; see more discussions in Section 2.2.

- We apply our new method and standard GARCHX model to investigate the role of geopolitical risks in forecasting volatility. It turns out that our new method can be significantly more accurate, especially for long-horizon time aggregated predictions.

We organize the remainder of this article as follows. In Section 2, we review the classical forecasting model, namely GARCHX, which is used as the starting point to propose the GAX-NoVaS method. Also, we present more details of the Model-free prediction principle and prove the existence of a transformation function with some exogenous variables existing. In Section 3, we delineate the details of proposed GAX-NoVaS method. Then, some simulation studies and model evaluation criteria are collated in Section 4. Next, we contrast our methods to existing classical ones on three empirical datasets in Section 5. Finally, in Section 6, we conclude by discussing the implications of our findings and some future directions.

2 GARCHX estimation and Model-free prediction principle

Before introducing our GAX-NoVaS method, we first explain the GARCHX model since the transformation of GAX-NoVaS is based on the GARCHX model. In addition, we give more details on the Model-free prediction principle and we prove the existence of a transformation function to achieve the Model-free prediction goal.

2.1 GARCHX model

In a seminal work, ARCH was proposed by Engle (1982) to model volatility or σ_t^2 for a time-series in a dynamic way. Following this, many different variants were developed in the econometrics literature. The GARCH model, especially the GARCH(1,1) expressed in Eq. (2), stands as possibly the most popular one. When additional information is available, people would like to utilize this extra knowledge to improve prediction accuracy. Subsequently, the so-called GARCHX model enters the public eye; see Francq and Thieu (2019) for discussions on the quasi-maximum likelihood estimation inference of GARCHX models. To simplify the analysis, participators usually assume the normality of W_t. As mentioned before, the normality assumption is not necessary with the NoVaS method. After taking a vector of exogenous covariates $\boldsymbol{X} = (X_1, \ldots, X_m)$ into account, we can wrap the exogenous covariates into the prediction process by turning the GARCH(1,1) model into the following GARCHX(1,1,1) model:

$$\begin{aligned} Y_t &= \sigma_t W_t, \\ \sigma_t^2 &= a + a_1 Y_{t-1}^2 + b_1 \sigma_{t-1}^2 + \boldsymbol{c}^T \boldsymbol{X}_{t-1}. \end{aligned} \tag{3}$$

To guarantee the non-negativity of σ_t^2, we define $\boldsymbol{X}_{t-1}$ as $(|X_{1,t-1}|, \ldots, |X_{m,t-1}|)$ and restrict the coefficients of these exogenous variables to be positive, i.e., $\boldsymbol{c} \geq 0$. To perform a moving-window out-of-sample prediction experiment, we first need to estimate the GARCH(1,1) and GARCHX(1,1,1) models[1], and then we compute predictions iteratively; see Section 4 for details. In this process, we assume that we know the true exogenous variables, which is feasible because we generate out-of-sample predictions. For practical applications, if needed, the future exogenous information can be estimated separately.

2.2 Model-free prediction principle

The model-free prediction principle was initially well developed by Politis (2015). Later, Chen and Politis (2019) applied this idea to multi-step ahead predictions of financial returns in the context of an ARCH-model structure. In short, the main idea behind the model-free prediction is to apply an invertible transformation function, H_T, that can map a non-*i.i.d.* vector, $\{Y_t \,; t = 1, \ldots, T\}$, to a vector, $\{\epsilon_t;\ t = 1, \ldots, T\}$, with *i.i.d.* components (chosen as standard normal in this work, but we should notice that other simple distributions can also work). Due to the invertibility of the function, H_T, it is possible to construct a one-to-one relationship between a future value, Y_{T+1}, and ϵ_{T+1}, i.e.,

$$Y_{T+1} = f_{T+1}(\boldsymbol{Y}_T, \boldsymbol{X}_{T+1}, \epsilon_{T+1}); \tag{4}$$

where $\boldsymbol{Y}_T$ denotes all historical data $\{Y_t;\ t = 1, \ldots, T\}$; $\boldsymbol{X}_{T+1}$ is the collection of all absolute predictors, and it also contains the value of a future predictor $|X_{T+1}|$; the form of $f_{T+1}(\cdot)$ depends on H_T^{-1}. This relationship implies that we can also transform the prediction of ϵ_{T+1} to the prediction of Y_{T+1}. Assume we have $\hat{\epsilon}_{T+1}$ to the be the predictor of ϵ_{T+1}, we can express the predictor of Y_{T+1} as

$$\widehat{Y}_{T+1} = f_{T+1}(\boldsymbol{Y}_T, \boldsymbol{X}_{T+1}, \hat{\epsilon}_{T+1}). \tag{5}$$

[1]For estimation of the GARCH and GARCHX models, we use the *fGarch* (Wuertz et al., 2013) and *garchx* packages (Sucarrat, 2020) in the *R* language and environment (R Core Team, 2023).

Because the prediction of *i.i.d.* data is standard, the L_1 (Mean Absolute Deviation), L_2 (Mean Squared Error), or another optimal quantile predictor of ϵ_{T+1} can easily be found. We, thus, can easily obtain the corresponding optimal predictor of Y_{T+1}.

For multi-step (h-step) ahead prediction, we simply repeat this prediction process, i.e., we express Y_{T+h} through a function w.r.t. $\boldsymbol{Y}_T$, $\boldsymbol{X}_{T+1}$ and $\{\epsilon_{T+1}, \ldots, \epsilon_{T+h}\}$:

$$Y_{T+h} = f_{T+h}(\boldsymbol{Y}_T, \boldsymbol{X}_{T+1}, \epsilon_{T+1}, \ldots, \epsilon_{T+h}). \tag{6}$$

In order to compute the prediction of Y_{T+h}, we take a distribution-match approach to approximate the distribution of Y_{T+h}. Ideally, when we know the exact distribution of the *i.i.d.* ϵ, we can use a Monte Carlo simulation to approximate the distribution of Y_{T+h} based on Eq. (6). Practically speaking, when we just have the empirical transformation results, i.e., the observed sample $\{\epsilon_t\}_{t=1}^T$, bootstrap is an appropriate approach. Moreover, we can even predict $g(Y_{T+h})$, where $g(\cdot)$ is a general continuous function. For example, we can compute the L_1 and L_2 optimal predictors of $g(Y_{T+h})$ as below:

$$\begin{aligned} g(Y_{T+h})_{L_2} &= \frac{1}{M}\sum_{m=1}^{M} g(f_{T+h}(\boldsymbol{Y}_T, \boldsymbol{X}_{T+1}, \hat{\epsilon}_{T+1,m}, \ldots, \hat{\epsilon}_{T+h,m})), \\ g(Y_{T+h})_{L_1} &= \text{Median of } \{g(f_{T+h}(\boldsymbol{Y}_T, \boldsymbol{X}_{T+1}, \hat{\epsilon}_{T+1,m}, \ldots, \hat{\epsilon}_{T+h,m})); m = 1, \ldots, M\}; \end{aligned} \tag{7}$$

where $g(Y_{T+h})_{L_2}$ and $g(Y_{T+h})_{L_1}$ represent the optimal L_2 and L_1 predictor of $g(Y_{T+1})$, the $\{\hat{\epsilon}_{T+1,m}\}_{m=1}^M$ are generated by bootstrap or Monte Carlo simulation, and M is some large number (2000 in our empirical analysis). For further discussion, see Politis (2015).

To the best of our knowledge, the model-free prediction idea has not been studied when the model features exogenous variables. However, it may be beneficial to take into account in the prediction process the additional information embedded in such exogenous variables. To show the NoVaS approach is still applicable, we need a transformation function that maps the targeted variables and exogenous predictors together into some simple *i.i.d.* random variables. Under some mild conditions, we show the existence of such a transformation function based on the probability integral transform. We assume:

- A1. The joint density of $\{Y_1, \cdots, Y_T\}$ exists for any $T \geq 1$.
- A2. For exogenous random vector $\boldsymbol{X} := \{X_1, \cdots, X_m\}$, the joint density $\{Y_1, \cdots, Y_T, X_1, \cdots, X_m\}$ exists for any $m \geq 1$.

Then, the feasibility of NoVaS transformation with exogenous variables existing is guaranteed by Theorem 2.1 shown below:

Theorem 2.1. *Under A1 and A2, there exists a function* $\boldsymbol{g}$ *such that* $\boldsymbol{Z} = \boldsymbol{g}((\boldsymbol{Y}, \boldsymbol{X}))$ *and the corresponding inverse function* $\boldsymbol{h}$ *such that* $(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{X}}) = \boldsymbol{h}(\boldsymbol{Z})$*;* $\boldsymbol{Z} \sim N(0, \boldsymbol{I}_{T+m})$*;* $\boldsymbol{Y} = (Y_1, \cdots, Y_T)$ *and* $\boldsymbol{X} = (X_1, \cdots, X_m)$ *are any two random vectors;* $(\widetilde{\boldsymbol{Y}}, \widetilde{\boldsymbol{X}})$ *have the same joint distribution of* $(\boldsymbol{Y}, \boldsymbol{X})$.

Proof. The proof of Theorem 2.1 is based on the probability integral transform; see Angus (1994) for a review. Without loss of generality, we start from Y_1 to determine the transformation function $\boldsymbol{g}$. Let $U_1 := \tilde{g}_1(Y) = F(Y_1)$; $F(Y_1)$ is the distribution

of Y_1. According to the probability integral transform, we know U_1 has a uniform distribution on $[0,1]$. Then, we make

$$U_2 := \tilde{g}_2(Y_1, Y_2) = F(Y_2|Y_1). \tag{8}$$

$F(Y_2|Y_1)$ is the conditional distribution of Y_2. Eq. (8) implies that Z_2 is Uniform$(0,1)$ conditional on Y_1. Thus, the unconditional (marginal) distribution of U_2 is still Uniform$(0,1)$, and U_2 and Y_1 are independent so that U_2 is also independent with U_1. This can be seen from the equation below:

$$p_{U_2,Y_1}(u_2, y_1) = p_{U_2|Y_1}(u_2|y_1)p_{Y_1}(y_1). \tag{9}$$

Integrating both sides w.r.t. y_1, we can find $p_{U_2}(u_2) = 1$ on the region $[0,1]$, since Z_2 is Uniform$(0,1)$ conditional on $Y_1 = y_1$ for any y_1. We can repeat this process as a Gram–Schmidt-like recursion, i.e., we let $U_3 := \tilde{g}_3(Y_1, Y_2, Y_3) = F(Y_2|Y_1, Y_2)$ and so on. In total, we need $\{\tilde{g}_1, \cdots, \tilde{g}_{T+m}\}$ and they are functions of $\boldsymbol{Y}$. Thus, there exists a function $\tilde{\boldsymbol{g}}$ which maps $(\boldsymbol{Y}, \boldsymbol{X})$ to $\boldsymbol{U}$ which has *i.i.d.* uniform components $\{U_1, \cdots, U_{m+T}\}$, i.e., $\boldsymbol{U} = \tilde{\boldsymbol{g}}((\boldsymbol{Y}, \boldsymbol{X}))$. Then, $\boldsymbol{Z} = \Phi^{-1} \circ \tilde{\boldsymbol{g}}((\boldsymbol{Y}, \boldsymbol{X}))$ has multivariate normal distribution $N(0, \boldsymbol{I}_{T+m})$; Φ^{-1} is the quantile function of $N(0, \boldsymbol{I}_{T+m})$. Finally, we can take $\boldsymbol{g} = \Phi^{-1} \circ \tilde{\boldsymbol{g}}$.

On the other hand, if $F_{Y_1}^{-1} := \inf\{x : F_{Y_1}(x) \geq y\}, 0 \leq y \leq 1$, then $F_{Y_1}^{-1}(U_1)$ has the distribution as the same as Y_1. Similarly, we can get the conditional distribution of Y_2 on Y_1 by taking $F_{Y_2|Y_1}^{-1}(U_2)$. By repeating this process, we can recover the joint distribution of $(\boldsymbol{Y}, \boldsymbol{X})$ by chain rule. In other words, there exists a $\boldsymbol{h}$ such that $(\boldsymbol{Y}, \boldsymbol{X}) \overset{d}{=} \boldsymbol{h}(\boldsymbol{Z})$. □

The direct implication of Theorem 2.1 is that the Model-free prediction principle is feasible even when exogenous variables are included in the dependence dynamics. Moreover, our theorem is more general than Lemma 3.1 from Wang and Politis (2022) where the time series model must satisfy some strict conditions.

Remark 2.1. *To go one step further, Theorem 2.1 implies the fact that there are two oracle functions $\boldsymbol{g}$ and $\boldsymbol{h}$ which can serve as our desired transformation functions so that the prediction can be made without any restrictive model assumption. Unfortunately, $\boldsymbol{g}$ and $\boldsymbol{h}$ are unknown and must be determined by the data at hand. Moreover, $\boldsymbol{g}$ and $\boldsymbol{h}$ can be pretty complicated so the non-parametric estimation may be needed to achieve a purely Model-free prediction, but the curse of dimensionality impairs its feasibility when the dimension of X and Y are high. Thus, the NoVaS method, which is more parsimonious, can be thought of as a compromise between applying a naive model-based and thoroughly Model-free prediction approaches, i.e., the terminology distribution-free put forward in Section 1. Also, this distribution-free property is implied by Theorem 2.1 since no requirement about $\boldsymbol{X}$ and $\boldsymbol{Y}$ to possess some specific distribution. In other words, the NoVaS method attempts to estimate $\boldsymbol{g}$ and $\boldsymbol{h}$ motivated by some specific model structure.*

3 GAX-NoVaS prediction method

We first present the state-of-the-art GARCH-NoVaS method which is based on a so-called NoVaS transformation. Then, we extend the GARCH-NoVaS method to a GAX-NoVaS method, which features the exogenous variables.

3.1 NoVaS transformation

For the sake of completeness, we first give a brief introduction to the NoVaS transformation (model) which is a direct application of the Model-free prediction idea explained in Section 2.2. Initially, the NoVaS transformation is developed from the ARCH model:

$$Y_t = W_t \sqrt{a + \sum_{i=1}^{p} a_i Y_{t-i}^2}; \tag{10}$$

here, these parameters satisfy $a \geq 0$, $a_i \geq 0$, for all $i = 1, \ldots, p$; W_t can be i.i.d. standard normal in the Model-based approach. In other words, the structure of the ARCH model gives us a ready-made H_T^{-1}. We can express W_t in Eq. (10) using other terms to get a potential H_T :

$$W_t = \frac{Y_t}{\sqrt{a + \sum_{i=1}^{p} a_i Y_{t-i}^2}} \; ; \text{ for } t = p+1, \ldots, T. \tag{11}$$

Politis (2003) further modified Eq. (11) as follows:

$$W_t = \frac{Y_t}{\sqrt{\alpha s_{t-1}^2 + \beta Y_t^2 + \sum_{i=1}^{p} a_i Y_{t-i}^2}} \; ; \text{ for } t = p+1, \ldots, T; \tag{12}$$

here, $\{Y_t;\ t = 1, \ldots, T\}$ is the sample data; $\{W_t;\ t = p+1, \ldots, T\}$ should not be understood as the innovation term anymore. It is more appropriate to treat it as the transformed vector which mimics some specifically simple distribution, e.g., standard normal in this paper; α is a fixed scale invariant constant; s_{t-1}^2 is an estimator of the variance of $\{Y_i;\ i = 1, \ldots, t-1\}$ and can be calculated by $(t-1)^{-1} \sum_{i=1}^{t-1} (Y_i - \overline{Y})^2$, where $\overline{Y}$ is the sample mean of $\{Y_i;\ i = 1, \ldots, t-1\}$. For making Eq. (12) be a qualified function H_T, i.e., making $\{W_t\}_{t=p+1}^{T}$ obey *i.i.d.* standard normal distribution, we need to impose some restrictions on α and $\beta, a_1, \ldots, a_p$. We first stabilize the variance by requiring:

$$\alpha \geq 0, \beta \geq 0, a_i \geq 0 \; ; \text{ for all } i \geq 1, \alpha + \beta + \sum_{i=1}^{p} a_i = 1. \tag{13}$$

In application, $\{W_t\}_{t=p+1}^{T}$ transformed from financial log-returns by NoVaS transformation are usually uncorrelated. Therefore, if we make $\{W_t\}_{t=p+1}^{T}$ close to a Gaussian series i.e., normalizing $\{W_t\}_{t=p+1}^{T}$, we can get the desired *i.i.d.* property. This is why this transformation is called NoVaS.

There are many criteria to measure the normality of a series. Under the observation that the distribution of financial log-returns is usually symmetric, we choose the kurtosis to be a simple distance to measure the departure of a non-skewed dataset from that of the standard normal distribution (Politis, 2015). Besides, matching marginal distribution seems sufficient to normalize the joint distribution of $\{W_t\}_{t=p+1}^{T}$ for practical purposes based on empirical results. If we denote the marginal distribution of $\{W_t\}_{t=p+1}^{T}$ and the corresponding kurtosis by $\widehat{F}_w$ and $\text{KURT}(W_t)$, respectively, we then attempt to minimize $|\text{KURT}(W_t) - 3|$ to obtain the optimal combination of $\alpha, \beta, a_1, \ldots, a_p$ such that $\widehat{F}_w$ is as close to standard normal distribution as possible. Subsequently, the NoVaS transformation can be determined.

The remaining difficulty is how to finish this optimization step to get optimal coefficients $\alpha, \beta, a_1, \ldots, a_p$, especially when p is large. To simplify this problem, Politis (2015) defined an exponentially decayed form of $\{a_i\}_{i=1}^p$:

$$\alpha \neq 0, \beta = c', a_i = c' e^{-ci} \ ; \ \text{for all } 1 \leq i \leq p, c' = \frac{1-\alpha}{\sum_{j=0}^{p} e^{-cj}}. \tag{14}$$

The NoVaS transformation based on coefficients defined in Eq. (14) is called Generalized Exponential NoVaS (GE-NoVaS). In other words, we can represent the $p+2$ number of coefficients by two parameters c and α, which relieve the optimization burden, but with a sacrifice that the coefficients are fixed in a decayed form. To achieve a balance between the relief of the optimization dilemma and the freedom of coefficients, inspired by the development of GARCH from ARCH, Wu and Karmakar (2023) built a NoVaS transformation according to the GARCH model, namely GARCH-NoVaS which was shown to be more stable and accurate. Later, we specify the GAX-NoVaS transformation in detail.

3.2 GAX-NoVaS transformation method

Starting from Eq. (3), we take similar steps of building GA-NoVaS to find the transformation function of the GAX-NoVaS method. To simplify the notation, we consider the case of only one exogenous covariate X_t. The case of multiple exogenous covariates can be analyzed analogously. First, we notice that we can rewrite the Eq. (3) as

$$W_t = \frac{Y_t}{\sqrt{a_0 + a_1 Y_{t-1}^2 + b_1 \sigma_{t-1}^2 + c_1 X_{t-1}}}. \tag{15}$$

We also have

$$\begin{aligned} \sigma_{t-1}^2 &= a_0 + a_1 Y_{t-2}^2 + b_1 \sigma_{t-2}^2 + c_1 X_{t-2}, \\ \sigma_{t-2}^2 &= a_0 + a_1 Y_{t-3}^2 + b_1 \sigma_{t-3}^2 + c_1 X_{t-3}, \\ &\vdots \end{aligned} \tag{16}$$

so that we can substitute these terms into Eq. (15). We then get

$$W_t = \frac{Y_t}{\sqrt{\frac{a_0}{1-b_1} + \sum_{i=1}^{p} a_1 b_1^{i-1} Y_{t-i}^2 + \sum_{i=1}^{p} c_1 b_1^{i-1} X_{t-i}}}; \tag{17}$$

where p is a large constant that is used to truncate the infinite summation. This fits with common intuition since the effects of previous data on prediction decrease as the time lag increases. In line with the NoVaS transformation, we finally write the transformation function as follows:

$$W_t = \frac{Y_t}{\sqrt{\alpha s_{t-1,Y} + \beta s_{t-1,X} + \sum_{i=1}^{p} a_1 b_1^{i-1} Y_{t-i}^2 + \sum_{i=1}^{p} c_1 b_1^{i-1} Y_{t-i}}}, \text{ for } t = p{+}1, \ldots, T; \tag{18}$$

where $s_{t-1,Y}^2$ and $s_{t-1,X}^2$ are the sample variance of $\{Y_1, \ldots, Y_{t-1}\}$ and $\{X_1, \ldots, X_{t-1}\}$, respectively. Thus, we can use Eq. (18) as the transformation function for the GAX-NoVaS method, where $\{W_t\}_{t=p+1}^T$ is the transformed series. Consequently, the

one-step (conditional) prediction $\widehat{Y}_{T+1}$ can be expressed as

$$\widehat{Y}_{T+1} = \widehat{W}_{T+1}\sqrt{\alpha s_{T,Y}^2 + \beta s_{T,X}^2 + \sum_{i=1}^{p} a_1 b_1^{i-1} Y_{T+1-i}^2 + \sum_{i=1}^{p} c_1 b_1^{i-1} X_{T+1-i}}; \quad (19)$$

where $\widehat{W}_{T+1}$ is the optimal point prediction of W_{T+1}. If $\{W_t\}$ are indeed *i.i.d.* normal, $\widehat{W}_{T+1}$ can be taken as the sample mean or sample median of $\{W_{p+1}, \ldots, W_T\}$ which leads to optimal L_2 or L_1 prediction of Y_{T+1}, respectively. Multi-step-ahead predictions can be computed as explained in Section 2.2 with the help of Bootstrap. To illustrate a little bit further, we take the procedure for 2-step ahead optimal point prediction as an example. Based on Eq. (18), we can express Y_{T+2} by W_{T+2}, W_{T+1} and all other historical information we are conditional on:

$$Y_{T+2} = W_{T+2}\sqrt{\alpha s_{T+1,Y} + \beta s_{T+1,X} + \sum_{i=1}^{p} a_1 b_1^{i-1} Y_{T+2-i}^2 + \sum_{i=1}^{p} c_1 b_1^{i-1} Y_{T+2-i}}; \quad (20)$$

Y_{T+1} is involved in Eq. (20) which depends on W_{T+1} implicitly. To derive the optimal point prediction of Y_{T+2}, we can bootstrap pseudo values $\{W_{T+1,m}W_{T+2,m}\}_{m=1}^{M}$ from $\{W_{p+1}, \ldots, W_T\}$ with a large M and then determine the optimal L_1 or L_2 prediction of Y_{T+2} as the method indicated in formula Eq. (7).

The final question left now is how to find a transformation function that indeed makes $\{W_t\}$ *i.i.d.* normal. While we have made some brief remarks on this question in Section 3.1, we next provide a full explanation with a focus on the GAX-NoVaS method. Our goal is to determine the coefficients, $\alpha, \beta, a_1, b_1, c_1$, of Eq. (18) to obtain the desired transformation. The most important step is to minimize $|\text{KURT}(W_t) - 3|$, where 3 is the kurtosis of normal distribution. For this optimization, we use the numerical technique to find the optimal coefficients.[2] In operation, one may obtain some extremely large values from the transformed series $\{W_t\}$, and such outliers may spoil the normality of the transformed series and may also influence prediction performance. Thus, before moving on to the prediction step, we truncate the transformed series by the 0.99 and 0.01 quantile values of a normal distribution with mean and standard deviation given by the sample mean and sample standard deviation of $\{W_t\}$.

Remark 3.1. *It is not difficult to perceive that the transformed series $\{W_t\}$ may be correlated, such that the decorrelation step is beneficial and necessary. One way to carry out this step is by fitting an AR(p) model on the $\{W_t\}$ series. Then, we record the residuals of the AR fit as $\{\hat{\epsilon}_t\}$. Also, we can approximate the one-step-ahead value $\widehat{W}_{T+1}$ with a fitted AR model. Then, we can create the new series as $\{\epsilon_t + \widehat{W}_{T+1}\}$. We use the empirical distribution of this new series to approximate the distribution of W_{T+1}. Eq. (19) can be used with the optimal prediction $\widehat{W}_{T+1}$ derived from this empirical distribution. It is still an open question, however, how to extend this decorrelation step to multi-step-ahead predictions.*

4 Simulations

In this section, we deploy several simulations to check the performance of GARCH, GARCHX and GAX-NoVaS methods. Before presenting the data-generating model

[2] We use the *nloptr* package (Ypma et al., 2014) for *R*.

used to do simulations, we explain the procedure of the moving-window time-aggregated predictions and give the model evaluation metrics to measure the performance of different methods.

4.1 Moving-window time-aggregated prediction

If we have sample $\{Y_1, \ldots, Y_N\}$ at hand, in order to fully exhaust the dataset, we can focus on moving-window out-of-sample predictions, i.e., we use $\{Y_1, \cdots, Y_T\}$ to predict $\{Y^2_{T+1}, \cdots, Y^2_{T+h}\}$, then we use $\{Y_2, \cdots, Y_{T+1}\}$ to predict $\{Y^2_{T+2}, \cdots, Y^2_{T+h+1}\}$, and so on until we reach the end of the sample (that is, until we use $\{Y_{N-T+h+1}, \cdots, Y_{N-h}\}$ to predict $\{Y^2_{N-h+1}, \ldots, Y^2_N\}$). Here, T denotes the moving window size; we fix its size as 250; h is the prediction horizon, i.e., 1, 5, 20 in our setting. Sometimes, we may not have enough data available to perform predictions. Thus, the window size $T = 250$ is designed to see if this method is stable and can still return accurate predictions even with short data. A 250-size moving window is in line with around one year of daily financial data. In this perspective, it is important to keep in mind that a 250-size moving window is practically meaningful since the time series may not be stationary on a wider span.

In addition, we are not only interested in the one-step-ahead prediction $h = 1$ but also multi-step-ahead prediction $h > 1$. From a practical aspect of forecasting volatility, as mentioned above in our introduction Section 1, the long-term prediction (h takes a large value) is important and can guide future strategic decisions. Before this evaluation, we start by writing time-aggregated predictions as follows:

$$\overline{\widehat{Y}}^2_{T,h} = \sum_{k=1}^{h} \widehat{Y}^2_{T+k}/h; \tag{21}$$

where $\overline{\widehat{Y}}^2_{T,h}$ is the h-step ahead time-aggregated volatility prediction starting from Y_T. For example, if the total number of data $N = 1000$ and we consider the 6-step-ahead moving window time-aggregated predictions with $T = 500$, we need to find predictions $\overline{\widehat{Y}}^2_{T,6}$ for $T = 500, \ldots, 994$.

We hope the time-aggregated volatility prediction is close to the true aggregated value calculated from the realized average squared log-returns $\overline{Y}^2_{l,h} = \sum_{k=1}^{h}(Y^2_{T+k}/h)$. To evaluate the accuracy, we can consider the specific mean of squared prediction errors (MSPE) shown below, with this statistic aiming to compare the prediction performance in an absolute way:

$$P = \frac{1}{N-h-T+1} \sum_{l=T}^{N-h} (\overline{\widehat{Y}}^2_{l,h} - \overline{Y}^2_{l,h})^2; \tag{22}$$

where $\overline{\widehat{Y}}^2_{l,h}$ and $\overline{Y}^2_{l,h}$ denote the predicted and true time-aggregated values for each moving-window forecasting, respectively.

4.2 Simulation setting

In this part, we present three data-generating models to simulate data and then evaluate the performance of various methods considered in this paper. Due to the true data-generating process being known to us, we can simulate any size of the

sample and compare predictions from various methods with oracle values. Besides, we remove the variance term $\beta s^2_{T-1,X}$ of Eq. (18) when we do the transformations since the prediction performance hardly changes with or without $\beta s^2_{T-1,X}$. Three true underlying models are presented below:

Model 1: Standard GARCH(1,1):
$Y_t = \sigma_t \epsilon_t,\ \ \sigma_t^2 = 0.00001 + 0.73\sigma_{t-1}^2 + 0.1Y_{t-1}^2 + c|X_{t-1}|,$
$X_{t-1} \sim i.i.d.\ N(0,1)$; $\{\epsilon_t\} \sim i.i.d.$ t distribution with four degrees of freedom
$c = 1$.

Model 2: Standard GARCH(1,1) with another set of coefficients:
$Y_t = \sigma_t \epsilon_t,\ \ \sigma_t^2 = 0.00001 + 0.8895\sigma_{t-1}^2 + 0.1Y_{t-1}^2 + c|X_{t-1}|,$
$X_{t-1} \sim i.i.d.\ N(0,1)$; $\{\epsilon_t\} \sim i.i.d.$ t distribution with four degrees of freedom
$c = 1$.

Model 3: Time-varying GARCHX(1,1):
$Y_t = \sigma_t \epsilon_t,\ \ \sigma_t^2 = b_t\sigma_{t-1}^2 + a_t Y_{t-1}^2 + c|X_{t-1}|,$
$X_{t-1} \sim i.i.d.\ N(0,1)$; $\{\epsilon_t\} \sim i.i.d.$ t distribution with five degrees freedom;
$c = 1; g_t = t/n;\ a_t = 0.1 - 0.05g_t;\ b_t = 0.7 + 0.2g_t,$ n is the total length of the time series.

To check the robustness of methods on model misspecification, we intend to take the innovation distribution of simulation models as the t-distribution. Besides this purpose, we argue that the t distribution as the innovation to mimic the real-world cases is more appropriate since real data usually show the heavy tail phenomenon. Models 1 and 2 are from a standard GARCH where in Model 2 we intended to explore a scenario that $\alpha_1 + \beta_1$ is very close to 1 and thus mimics what would happen for the iGARCH situation. In addition, we make the coefficients of the GARCHX model change linearly in Mode-3 so that we can observe the ability of different methods to handle the data generated from a time-varying model which is more coherent to the real-world situation. To sync with the empirical studies later, we simulate a time series with a length $T = 4694$. We take the moving-window size $T = 250$. MSPE of GARCH, GARCHX and GAX-NoVaS methods with three simulation settings are presented below:

Table 1: MSPE ratios of different methods on three simulated datasets.

	Model-1			**Model-2**			**Model-3**		
Prediction steps	1	5	20	1	5	20	1	5	20
GARCH	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GARCHX	0.985	1.007	0.854	1.014	0.957	0.701	0.921	0.956	2.680
GAX-NoVaS	0.842	0.566	0.222	0.798	0.818	0.085	0.870	1.276	0.006

Note: To simplify the presentation, we compute the ratio of MSPE of different methods, i.e., we divide all MSPE for each prediction horizon and models by the MSPE of the GARCH (benchmark) method. In addition, 0.000 value in the above table is a rounding number.

From Table 1, it is clear that the GAX-NoVaS is much better than GARCH and even the classical GARCHX models according to the MSPE criterion. Interestingly, the GARCHX model is even much worse than the GARCH model for some specific cases, e.g., the 20-step-ahead prediction of Model-3. By taking a deeper analysis, we find the terrible performance GARCHX method is due to some extremely large

predictions. On the other hand, the prediction returned by GAX-NoVaS is more stable. The superiority is further shown in Section 5 with three real datasets and various exogenous predictors.

Since the focus of this paper is exploring a new approach to incorporate exogenous variables, we take the DM test to evaluate the performance of GARCHX and GAX-NoVaS methods more formally; see Diebold and Mariano (2002) for the technical details of the DM-test[3]. The DM-test results on comparing GARCHX and GAX-NoVaS for forecasting three simulated datasets are tabularized in Table 2. These tests further verify the advantage of our methods on forecasting with exogenous variables, especially for a long-prediction horizon.

Table 2: DM-test results on simulated datasets.

	Model-1			**Model-2**			**Model-3**		
Prediction steps	1	5	20	1	5	20	1	5	20
GARCHX									
GAX-NoVaS	0.009	0.020	0.000	0.038	0.107	0.000	0.172	1.000	0.102

Note: The values in the different rows are one-sided p-values of the DM-test on the prediction error of GAX-NoVaS and GARCHX. For the DM-test here, the alternative hypothesis is that the former method GAX-NoVaS is more accurate than the latter one, GARCHX.

5 Empirical analyses with real data

A summarizing note of our findings in Section 4 reads that the GAX-NoVaS method performs better than standard GARCH-type methods, especially for long-term time aggregated predictions. In this section, we deploy an interesting data analysis to exhibit how our method could shed light on the role of geopolitical risks in forecasting volatility with real-world data. We start by describing the data below.

5.1 Data description

The ongoing Ukraine-Russia has led many countries in Europe, particularly Germany, to adjust their military and security, as well as energy policies in light of new geopolitical risks, with such adjustments entailing large costs to the macroeconomy and financial markets, as depicted by Grebe et al. (2024). In this regard, these authors, first, assemble a data set of more than eight million German Twitter posts related to the war in Ukraine to construct a daily index of uncertainty about the war as perceived by German Twitter based on using state-of-the-art methods of textual analysis. Grebe et al. (2024) show that an increase in uncertainty has strong effects on financial markets, associated with a significant decline in economic activity as well as an increase in expected inflation. We utilize this index (Ukraine)[4] in our empirical analysis to forecast stock market volatility of not only Germany, but two of its neighbors namely, Austria and Switzerland, over the daily period of 1st January, 2021 to 28th February, 2023. The national stock market indexes (ATX

[3]We perform the DM-test with the function *DM-test* in the *R* package *multDM*.

[4]The data is available for download from: https://www.uni-giessen.de/de/fbz/fb02/fb/professuren/vwl/tillmann/forschung/ukraine-uncertainty-index.

(Austria), DAX (Germany), SMI (Switzerland)) of these three countries, for which we compute log-returns to feed into our volatility models were derived from the Bloomberg terminal. With the focus being on geopolitical risks, we also utilized the daily newspapers-based geopolitical risks index (GPRD) of Caldara and Iacoviello (2022)[5], which, in turn, allowed us to analyze a longer data sample covering 2nd January, 2006 to 10th August, 2023. The starting date of this longer sample, and the choice of these three countries, were also motivated by the availability of Google searches-based daily data on economic activity (Trend) for all three countries, and inflation (Inflation) for Germany and Switzerland,[6] which are used as additional predictors to ensure that our results are not only limited to geopolitical risks.

We present three log-return series of Germany, Switzerland and Austria from 2nd January, 2006 to 10th August, 2023 in Fig. 1. The volatility clustering phenomena observed in all plots reveals the heteroskedasticity within these three series. To investigate the property of three long return series, we provide the summary statistics, e.g., the mean, skewness, and kurtosis in Table 3. To verify the heteroskedasticity with all series more directly, we split the whole time period into four equal-length sub-periods and denote the sample variance of all four sub-periods by V_i, $i = 1, \ldots, 4$. These statistics are also provided. Towards statistical tests, we also perform modified Ljung-Box (m-LB) and ARCH Lagrange Multiplier (ALM) tests to check the autocorrelation and ARCH effects of squared return series. For the m-LB test, we consider the lag order 20. For the ALM test, we consider the maximum lag order 10. These two tests are performed in *R* with functions *lbtest* and *Lm.test*, respectively. The p-values of tests are presented. Summarizing Table 3, the large kurtosis values indicate the heavy-tailed property for all three log-return series. The variance of return series in different time regions changes notably, indicating the heteroskedasticity. The ALM test with a pretty small p-value also confirms the heteroskedasticity for all return series. The m-LB test shows strong evidence of autocorrelation within all squared return series.

Table 3: Summary statistics of three long return series.

Returns series	Mean	Skew.	Kurt.	V_1	V_2	V_3	V_4	m-LB	ALM
Germany	0.01	-0.24	11.37	2.51	1.52	1.22	1.79	0.00	0.00
Switzerland	0.01	-0.43	12.59	1.83	0.83	0.88	0.95	0.00	0.00
Austria	-0.01	-0.41	10.26	4.78	2.15	1.81	3.27	0.00	0.00

Note: columns V_i, $i = 1, \ldots, 4$ represent the sample variance of each long series on four equal-length sub-periods splitted from the whole period 01/02/2006 to 08/10/2023. The column ALM represents the p-value of the ALM test with the maximum lag order being 10; the column m-LB represents the p-value of the m-LB test at the lag order being 20; 0.00 indicates the p-value is less than 3×10^{-16} for these two tests. Skew. and Kurt. represent the skewness and kurtosis respectively.

To show the fluctuations behind the Ukraine index and GPRD index, we present two plots in Fig. 2. As one can see from there, these two indices fluctuate severely around the beginning of 2022 which corresponds with the real-world event. Later, we attempt to use this information to forecast the stock market volatility of three countries.

[5]The data can be accessed from: https://www.matteoiacoviello.com/gpr.htm.
[6]The data can be downloaded from: https://www.trendecon.org/.

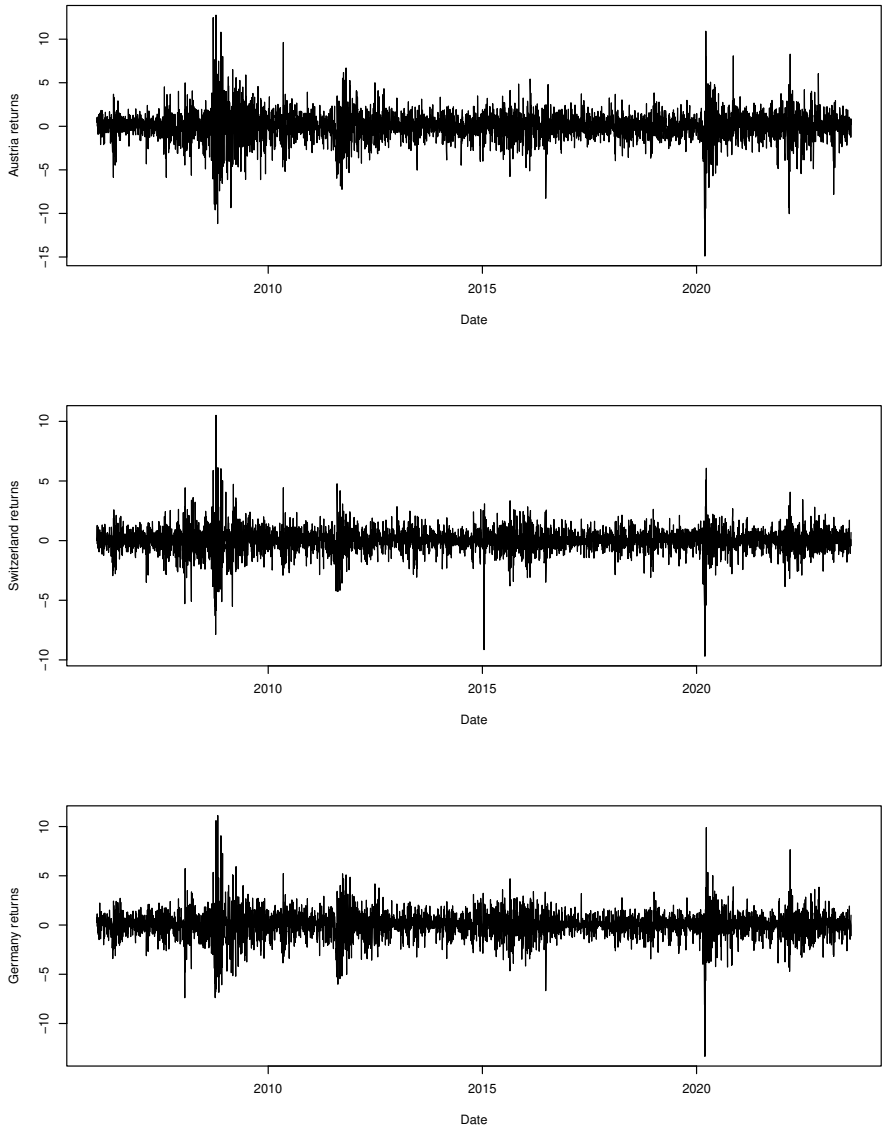

Figure 1: Top to bottom: The log-returns of Austria's, Switzerland's, and Germany's national stock market index from 01/02/2006 to 08/10/2023.

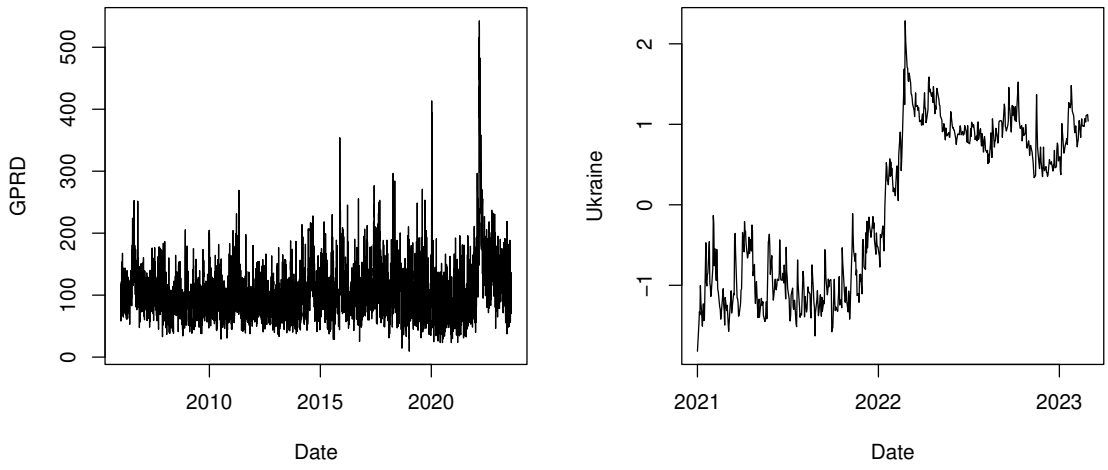

Figure 2: Left figure: The exogenous variable GPRD index from 01/02/2006 to 08/10/2023. Right figure: The exogenous variable Ukraine index from 01/01/2021 to 28/02/2023 which corresponds to the time period of our short data.

5.2 Empirical results

We first consider the forecasting exercise with the short period (1st January, 2021 to 28th February, 2023) data described in Section 5.1. Then, the analysis of three long returns series is given in Section 5.2.2.

5.2.1 Short data

To compare the performance of GARCH, GARCHX and GAX-NoVaS methods, we still apply the time aggregated prediction metric described in Section 4.1. We consider $h = 1, 5, 20$ and use a 250-size moving window, which is about 1 year of daily data. We start the empirical analysis with the short period of data and we take the Ukraine index as the exogenous predictor. The MSPE results are summarized in Table 4.

Table 4: MSPE ratios of different methods on three short datasets.

	S-Germany			**S-Austria**			**S-Switzerland**		
Prediction steps	1	5	20	1	5	20	1	5	20
GA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GAX-Ukraine	1.011	1.023	0.989	1.002	1.003	0.994	1.003	1.008	1.005
GAX-NoVaS-Ukraine	1.038	1.141	0.898	1.017	0.984	0.822	1.010	1.100	0.879

Note: "S" represents "short". To simplify the presentation, we compute the ratio of MSPE of different methods, i.e., we divide all MSPE for each prediction horizon and model by the MSPE of the GARCH (benchmark) method.

From Table 4, we can see that the GAX-NoVaS method can bring some large improvements, especially for long-horizon aggregated predictions. Meanwhile, it seems that the GARCHX and GARCH models have indistinguishable performance. However, the GAX-NoVaS method is generally better than both GARCH-type methods. The

DM-test results on comparing GARCHX and GAX-NoVaS for forecasting short real-world data are tabularized in Table 5, which reveals the significant advantage of GAX-NoVaS for long-horizon predictions, especially for the 20-step-ahead predictions of Short Austria data.

Table 5: DM-test results on predictions of short datasets.

	S-Germany			**S-Austria**			**S-Switzerland**		
Prediction steps	1	5	20	1	5	20	1	5	20
GARCHX-Ukraine									
GAX-NoVaS-Ukraine	0.616	0.727	0.306	0.578	0.410	0.030	0.549	0.860	0.118

Note: The values in the different rows are one-sided p-values of the DM-test on the prediction error of GAX-NoVaS and GARCHX. For the DM-test here, the alternative hypothesis is that the former method GAX-NoVaS is more accurate than the latter one GARCHX.

5.2.2 Long data

We continue our real data analysis with long datasets (2nd January, 2006 to 10th August, 2023). We also apply more exogenous variables. Similar to the analysis procedure for short data, we present MSPE ratios and corresponding DM-test results of GARCHX and GAX-NoVaS in Tables 6 and 7. Generally speaking, the GAX-NoVaS method still dominates the other two GARCH-type methods, and this superiority is verified to be significant by the DM-test.

Table 6: MSPE ratios of different methods on three long datasets.

	L-Germany			**L-Austria**			**L-Switzerland**		
Prediction steps	1	5	20	1	5	20	1	5	20
GA	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GA-NoVaS	0.990	0.922	0.518	1.062	1.041	0.715	1.010	0.786	0.128
GAX-GPRD	1.010	1.024	0.998	0.999	0.990	0.964	1.003	0.994	0.797
GAX-NoVaS-GPRD	1.000	0.953	0.557	1.076	1.121	0.831	1.005	0.799	0.133
GAX-Inflation	1.003	1.016	0.974				0.996	0.965	0.858
GAX-NoVaS-Inflation	1.001	0.949	0.551				1.010	0.801	0.131
GAX-Trend	0.987	0.979	1.062	0.982	0.912	0.654	0.996	0.937	0.594
GAX-NoVaS-Trend	0.998	0.949	0.540	1.077	1.169	0.866	1.019	0.828	0.138

Note: "L" represents "long". To simplify the presentation, we compute the ratio of MSPE of different methods, i.e., we divide all MSPE for each prediction horizon and model by the MSPE of the GARCH (benchmark) method.

Table 7: DM-test results on predictions of long datasets.

	L-Germany			L-Austria			L-Switzerland		
Prediction steps	1	5	20	1	5	20	1	5	20
GARCHX-GPRD									
GAX-NoVaS-GPRD	0.427	0.306	0.012	0.950	0.957	0.068	0.516	0.055	0.008
GARCHX-Inflation									
GAX-NoVaS-Inflation	0.484	0.298	0.007				0.595	0.082	0.029
GARCHX-Trend									
GAX-NoVaS-Trend	0.606	0.420	0.017	0.995	1.000	1.000	0.659	0.122	0.027

Note: The values in the different rows are one-sided p-values of the DM-test on the prediction error of GAX-NoVaS and GARCHX. For the DM-test here, the alternative hypothesis is that the former method GAX-NoVaS is more accurate than the latter one GARCHX.

6 Conclusion

We extend the current NoVaS prediction method to the realm of prediction with exogenous variables. We provide the theoretical foundation to guarantee the feasibility of applying Model-free prediction. Inspired by the GARCHX model, we propose a specifically designed Bootstrap-based method namely GAX-NoVaS prediction. The dominance of GAX-NoVaS on the classical GARCH-type methods is verified by simulation and empirical datasets. Also, such an advantage is not only exhibited by the MSPE metric but we also show some statistical significance through the parlance of classical DM tests.

We should mention that going far beyond GAX-NoVaS method might have limitations if the model becomes increasingly complex. Recall that the great performance of the GAX-NoVaS method relies on a successful transformation, the satisfied transformation may not be achievable if the underlying time series is very complicated. However, there is a growing literature on forecasting using more non-parametric neural network based models. It will be an interesting future work that combines the idea of model-free prediction with the state-of-the-art machine learning method, such as Deep neural network (DNN), convolutional neural network (CNN) or LSTM etc. Finally, in the field of binary/categorical/count data INGARCH models have recently garnered significant attention both from theoretical and applied researchers. One could potentially also think of extending the NoVaS method based on INGARCHX-type model to integrate the exogenous variables into prediction and challenge the existing methods. Beyond these extensions inspired by different prior model structures, performing NoVaS prediction with mixed frequency covariates is also appealing. In short, our paper remains the first paper to propose the idea of transformation-based forecasting focused on GARCHX-type models but the scope of extending this to several directions is ample.

Appendices

A The GARCHX(p,q,1)-NoVaS transformation function

Here, we give the formula to perform NoVaS transformation based on the GARCHX(p,q,1) model. This is the extension of the GARCHX(1,1,1) presented in the main text; p and q represent the order lag order of Y_t^2 and σ_t^2 terms, respectively. Moreover, we consider the situation in which s number exogenous variables exist in the context of making predictions. This general GARCHX model can be defined as follows:

$$\begin{aligned} Y_t &= \sigma_t W_t, \\ \sigma_t^2 &= a + \sum_{i=1}^{p} a_i Y_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-1}^2 + \sum_{k=1}^{s} c_k X_{k,t-1}. \end{aligned} \tag{23}$$

To initiate the corresponding transformation function, we rely on the expressions of σ_{t-i}^2 for $i = 1, \ldots, \infty$, i.e., we have

$$\begin{aligned} \sigma_{t-1}^2 &= a + \sum_{i=1}^{p} a_i Y_{t-1-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-1-j}^2 + \sum_{k=1}^{s} c_k X_{k,t-2}; \\ \sigma_{t-2}^2 &= a + \sum_{i=1}^{p} a_i Y_{t-2-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-2-j}^2 + \sum_{k=1}^{s} c_k X_{k,t-3}; \\ &\vdots \end{aligned} \tag{24}$$

Plug all terms in Eq. (24) into Eq. (23) iteratively, we can get:

$$\begin{aligned} Y_t &= W_t \sqrt{a + \sum_{i=1}^{p} a_i Y_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-1}^2 + \sum_{k=1}^{s} c_k X_{k,t-1}} \\ &= W_t \left(a + \sum_{i=1}^{p} a_i Y_{t-i}^2 + b_1 \left[a + \sum_{i=1}^{p} a_i Y_{t-1-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-1-j}^2 + \sum_{k=1}^{s} c_k X_{k,t-2} \right] \right. \\ &= \left. + b_2 \left[a + \sum_{i=1}^{p} a_i Y_{t-2-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-2-j}^2 + \sum_{k=1}^{s} c_k X_{k,t-3} \right] + \cdots + \sum_{k=1}^{s} c_k X_{k,t-1} \right)^{\frac{1}{2}} \\ &= W_t \left(I(a) + \sum_{i=1}^{\infty} I(Y_{t-i}^2) + \sum_{k=1}^{s} \sum_{j=1}^{\infty} I(X_{k,t-j}) \right)^{\frac{1}{2}} \end{aligned} \tag{25}$$

To simplify the expression of Eq. (25), we first consider the part $I(a)$ involved in Eq. (25). It is not hard to find that all terms including a consist of a geometric series, i.e.,

$$I(a) = \sum_{i=0}^{\infty} a_0 (b_1 + \cdots + b_q)^i = \frac{a_0}{1 - (b_1 + \cdots + b_q)}, \tag{26}$$

since we have the condition that $b_1 + \cdots + b_q < 1$ to satisfy the stationary requirement of the time series. Then, for terms $\sum_{i=1}^{\infty} I(Y_{t-i}^2)$, it can be simplied as:

$$\sum_{i=1}^{\infty} I(Y_{t-i}^2) = \sum_{i=1}^{\infty} \Xi_i Y_{t-i}^2; \tag{27}$$

where Ξ_i is the coefficient of Y_{t-i}^2 which can be defined as follows:

$$\Xi_i = \sum_{\substack{l_1,\ldots,l_q, j \in \{1,\ldots,p\} \\ s.t.,\ \sum_{v=1}^{q} v \cdot l_v + j = i}} \left(a_j \cdot N_{l_v \neq 0; v=1\ldots,q} \cdot b_1^{l_1} b_2^{l_2} \cdots b_q^{l_q} \right);$$

where $N_{l_v \neq 0; v=1\ldots,q}$ is the number of non-zero terms of l_v for $v = 1, \ldots, q$. For example, it is easy to see that:

$$\begin{aligned}
I(Y_{t-1}^2) &= a_1 Y_{t-1}^2; \\
I(Y_{t-2}^2) &= (a_1 b_1 + a_2) Y_{t-2}^2 \\
I(Y_{t-3}^2) &= (a_1 b_1^2 + a_1 b_2 + a_2 b_1 + a_3) Y_{t-3}^2 \\
I(Y_{t-4}^2) &= (a_1 b_1^3 + a_1 b_1 b_2 + a_1 b_2 b_1 + a_2 b_1^2 + a_2 b_2 + a_3 b_1 + a_4) Y_{t-3}^2
\end{aligned}$$

Similarly, we have

$$\sum_{k=1}^{s} \sum_{j=1}^{\infty} I(X_{k,t-j}) = \sum_{k=1}^{s} \sum_{i=1}^{\infty} \Lambda_{k,i} X_{k,t-i}^2; \tag{28}$$

where

$$\Lambda_{k,i} = \sum_{\substack{l_1,\ldots,l_q \\ s.t.,\ \sum_{v=1}^{q} v \cdot l_v = i-1}} \left(c_k \cdot N_{l_v \neq 0; v=1\ldots,q} \cdot b_1^{l_1} b_2^{l_2} \cdots b_q^{l_q} \right).$$

Combine all pieces in Eqs. (26) to (28), we can get

$$Y_t = W_t \left(\frac{a_0}{1 - (b_1 + \cdots + b_q)} + \sum_{i=1}^{\infty} \Xi_i Y_{t-i}^2 + \sum_{k=1}^{s} \sum_{i=1}^{\infty} \Lambda_{k,i} X_{k,t-i}^2 \right)^{\frac{1}{2}}. \tag{29}$$

Since $\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j$ is required to be less than 1, we can truncate the two infinity sums in Eq. (29) to order r which is an appropriately large constant, i.e., we can approximate Eq. (29) by a truncated summation as follows:

$$Y_t \approx W_t \left(\frac{a_0}{1 - (b_1 + \cdots + b_q)} + \sum_{i=1}^{r} \Xi_i Y_{t-i}^2 + \sum_{k=1}^{s} \sum_{i=1}^{r} \Lambda_{k,i} X_{k,t-i}^2 \right)^{\frac{1}{2}}. \tag{30}$$

Then, Eq. (30) can be thought of as a starting point to build the transformation function for the NoVaS prediction purposes. We can develop GARCHX(p,q,l) with s number of exogenous variables with a similar simplification procedure.

References

Angus, J. E. (1994). The probability integral transform and related results. SIAM review, 36(4):652–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3):307–327.

Bouri, E., Cepni, O., Gupta, R., and Jalkh, N. (2020). Geopolitical risks and stock market volatility in the g7 countries: A century of evidence from a time-varying nonparametric panel data model. Edited by Atin Basuchoudhary and Gunther G. Schulze.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

Caldara, D. and Iacoviello, M. (2022). Measuring geopolitical risk. American Economic Review, 112(4):1194–1225.

Chen, J. and Politis, D. N. (2019). Optimal Multi-step-ahead Prediction of ARCH/GARCH Models and NoVaS Transformation. Econometrics, 7(3):1–23.

Christos, B., Christina, C., Rangan, G., and Tahir, S. (2019). Geopolitical risks, returns, and volatility in emerging stock markets: Evidence from a panel garch model. Emerging Markets Finance and Trade, 55(8):1841–1856.

Chudý, M., Karmakar, S., and Wu, W. B. (2020). Long-term prediction intervals of economic time series. Empirical Economics, 58(1):191–222.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. Journal of Business & Economic Statistics, 20(1):134–144.

Engle, R. and Patton, A. (2001). What good is a volatility model? Quantitative Finance, 1:47–63.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the econometric society, pages 987–1007.

Engle, R. F. and Patton, A. J. (2007). What good is a volatility model? In Forecasting Volatility in the Financial Markets, pages 47–63. Elsevier.

Francq, C. and Thieu, L. Q. (2019). Qml inference for volatility models with covariates. Econometric Theory, 35(1):37–72.

Fryzlewicz, P., Sapatinas, T., Rao, S. S., et al. (2008). Normalized least-squares estimation in time-varying ARCH models. The Annals of Statistics, 36(2):742–786.

González-Rivera, G., Lee, T.-H., and Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. International Journal of Forecasting, 20(4):629–645.

Grebe, M., Kandemir, S., and Tillmann, P. (2024). Uncertainty about the war in ukraine: Measurement and effects on the german economy. Journal of Economic Behavior & Organization, 217(C):493–506.

Gulay, E. and Emec, H. (2018). Comparison of forecasting performances: Does normalization and variance stabilization method beat garch (1, 1)-type models? empirical evidence from the stock markets. Journal of Forecasting, 37(2):133–150.

Herrera, A. M., Hu, L., and Pastor, D. (2018). Forecasting crude oil price volatility. International Journal of Forecasting, 34(4):622–635.

Karmakar, S., Chudý, M., and Wu, W. B. (2022). Long-term prediction intervals with many covariates. Journal of Time Series Analysis, 43(4):587–609.

Karmakar, S. and Roy, A. (2021). Bayesian Modelling of Time-Varying Conditional Heteroscedasticity. Bayesian Analysis, 1(1):1–29.

Lim, C. M. and Sek, S. K. (2013). Comparing the Performances of GARCH-type Models in Capturing the Stock Market Volatility in Malaysia. Procedia Economics and Finance, 5:478–487.

Mehmet, B., Rangan, G., Christian, P., and E, W. M. (2018). Terror attacks and stock-market fluctuations: Evidence based on a nonparametric causality-in-quantiles test for the g7 countries. The European Journal of Finance, 24(4):333–346.

Peters, J.-P. (2001). Estimating and forecasting volatility of stock indices using asymmetric GARCH models and (Skewed) Student-t densities. Preprint, University of Liege, Belgium, 3:19–34.

Politis, D. N. (2003). A normalizing and variance-stabilizing transformation for financial time series. In Recent Advances and Trends in Nonparametric Statistics, pages 335–347. Elsevier Inc.

Politis, D. N. (2015). The model-free prediction principle. In Model-Free Prediction and Regression, pages 13–30. Springer.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Salisu, A. A., Ogbonna, A. E., Lasisi, L., and Olaniran, A. (2022). Geopolitical risk and stock market volatility in emerging markets: A garch–midas approach. The North American Journal of Economics and Finance, 62:101755.

Sucarrat, G. (2020). garchx: Flexible and robust garch-x modelling.

Wang, Y. and Politis, D. N. (2022). Model-free bootstrap for a general class of stationary time series. Bernoulli, 28(2):744–770.

Wu, K. and Karmakar, S. (2021). Model-free time-aggregated predictions for econometric datasets. Forecasting, 3(4):920–933.

Wu, K. and Karmakar, S. (2023). A model-free approach to do long-term volatility forecasting and its variants. Financial Innovation, 9(1):1–38.

Wu, K. and Politis, D. N. (2023). Bootstrap prediction inference of non-linear autoregressive models. arXiv preprint arXiv:2306.04126.

Wuertz, D., Runit, S., and Chalabi, M. Y. (2013). Package 'fgarch'. Technical report, Technical report, working paper/manual, 09.11. 2009. URL http://cran. r ....

Ypma, J., Borchers, H. W., and Eddelbuettel, D. (2014). nloptr: R interface to nlopt. R package version, 1(4).

Yule, G. U. (1927). Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer's sunspot numbers. Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 226(636-646):267–298.

Zhang, Y., He, J., He, M., and Li, S. (2023). Geopolitical risk and stock market volatility: A global perspective. Finance Research Letters, 53:103620.