

Time-varying NoVaS vs. GARCH: point prediction, volatility estimation and prediction intervals

Jie Chen

Dimitris N. Politis*

Abstract

The NoVaS methodology for prediction of stationary financial returns is reviewed, and the applicability of the NoVaS transformation for volatility estimation is illustrated using realized volatility as a proxy. The realm of applicability of the NoVaS methodology is then extended to non-stationary data (involving local stationarity and/or structural breaks) for one-step ahead point prediction of squared returns. In addition, a NoVaS-based algorithm is proposed for the construction of bootstrap prediction intervals for one-step ahead squared returns for both stationary and non-stationary data. It is shown that the ‘Time-varying’ NoVaS is robust against possible nonstationarities in the data; this is true in terms of locally (but not globally) financial returns but also in change point problems where the NoVaS methodology adapts fast to the new regime that occurs after an unknown/undetected change point. Extensive empirical work shows that the NoVaS methodology generally outperforms the GARCH benchmark for (i) point prediction of squared returns, (ii) interval prediction of squared returns, and (iii) volatility estimation. With regards to target (i), earlier work had shown little advantage of using a nonzero α in the NoVaS transformation. However, in terms of targets (ii) and (iii), it appears that using the Generalized version of NoVaS—either Simple or Exponential—can be quite beneficial and well-worth the associated computational cost.

Keywords: Time-varying data, Non-stationarity, Structural breaks, Realized volatility, Interval prediction, Locally stationary data

*Department of Mathematics and Halicioglu Data Science Institute, University of California at San Diego, La Jolla, CA 92093 USA.

1 Introduction

In applied econometrics and finance, accurate volatility forecasting is of great importance. Auto-Regressive Conditional Heteroscedasticity models (ARCH) and Generalized Auto-Regressive Conditional Heteroscedasticity models (GARCH) have gained prominence and are widely used in financial engineering since they were introduced by Engle (1982) and Bollerslev (1986) respectively. The simple GARCH(1,1) model emerged early on as the most popular and has been the benchmark in applied work for the last 30 years; see Francq and Zakoian (2011) and the references therein.

As noted early on, ARCH/GARCH models with respect to normal errors can account only partly for the degree of heavy tails empirically found in the distribution of returns. Consequently, researchers and practitioners have been resorting to ARCH/GARCH models with heavy-tailed errors, e.g., the t -distribution with degrees of freedom empirically chosen to match the apparent degree of heavy tails in the residuals; see Shephard (1996) and the references therein. However, this situation is not entirely satisfactory since the choice of a t -distribution seems quite arbitrary. Perhaps the real issue is that a simple and neat parametric model such as ARCH/GARCH could not be expected to perfectly capture the behavior of a complicated real-world phenomenon such as the evolution of financial returns that—almost by definition of market *efficiency*—ranks at the top in terms of difficulty of modeling/prediction.

The Model-free prediction principle was first introduced in Politis (2003) to understand such complex types of data. The *Normalizing and Variance-Stabilizing transformation* (NoVaS, for short) is a straightforward application of the Model-free principle in prediction of squared financial returns. The original development of the NoVaS approach was made in Politis (2003, 2007) having as its ‘spring board’ the popular ARCH model with normal innovations. Politis (2007) and Politis and Thomakos (2013) showed that NoVaS methods outperform the benchmark GARCH(1,1) model in prediction of squared returns under the assumption of stationarity.

Furthermore, a crucial problem in GARCH modeling is that the GARCH models are not robust with respect to violations from the stationarity assumption, e.g. in the presence of local stationarity and/or structural breaks; see Mikosch and Starica (2004); Polzehl and Spokoiny (2006). Subsequently, the theory of time-varying ARCH/GARCH process was developed for locally stationary time series;

see Dahlhaus and Rao (2006). The work of Polzehl and Spokoiny (2006) indicates that time-varying GARCH(1,1) models demonstrates a relatively good predicting performance as far as the short term forecasting horizon is considered. To deal with a time-varying stochastic structure of the data, we hereby propose a time-varying version of the NoVaS transformation. Our interest in this paper is to apply different time-varying NoVaS methods to predicting squared financial returns in situations where (global) stationarity for returns fails such as the aforementioned cases of local stationarity and/or structural breaks.

In this paper, we focus on studying the predictive power of different NoVaS methods to non-stationary data in point prediction of squared returns as well as in volatility estimation. In addition, we use the NoVaS methodology to construct prediction intervals for squared returns for both stationary and non-stationary time series. A comprehensive simulation and real world data analysis are conducted to study the relative forecasting performance of time-varying (TV) NoVaS methods compared to that of the benchmark time-varying GARCH(1,1) model. The evaluation of forecasting performance for the NoVaS transformation and the benchmark GARCH(1,1) models is addressed via the L_1 -norm instead of the usual mean squared error (MSE) since the case has been made that financial returns might not have a finite 4th moment; see e.g. Politis (2007, 2015).

The literature on volatility modeling, prediction and the evaluation of difference volatility forecasts is huge. Here, we selectively review some recent literature related to the volatility prediction: Mikosch and Starica (2004) for structural changes in volatility time series and GARCH modeling; Peng and Yao (2003) for robust least absolute deviations estimation of GARCH models; Poon and Granger (2003) for assessing the forecasting performance of various volatility models; Hansen et al. (2003) on selecting volatility models; Hansen and Lunde (2006a) for using a semi-parametric, transformation-based approach to forming predictive intervals; Ghysels et al. (2006) on the use and predictive power of absolute returns; Francq et al. (2005), Lux and Morales-Arias (2010) and Choi et al. (2010) on switching regime GARCH models, structural breaks and long memory in volatility; Hillebrand (2005) on GARCH models with structural breaks; Hansen and Lunde (2005, 2006b) for comparing forecasts of volatility models against the standard GARCH(1,1) model and for consistent ranking of volatility models and the use of an appropriate series as the ‘true’ volatility; Ghysels et al. (2006) for predicting volatility by mixing data at different frequencies,

and Ghysels and Sohn (2009) for the type of power variation that is successful in predicting volatility in the context of mixed data frequencies. Chen et al. (2008) for volatility forecasting in the context of threshold models coupled with volatility measurement based on intraday range. Bandi et al. (2008) discuss the selection of optimal sampling frequency in realized volatility estimation and forecasting; Patton and Sheppard (2009) present results on optimal combinations of realized volatility estimators in the context of volatility forecasting while Patton and Sheppard (2015) discuss the signed jumps and the persistence of volatility.

Andersen et al. (2004, 2005) examine the analytic evaluation of volatility forecasts, and Andersen et al. (2007) consider modeling realized volatility when jump components are included. The long line of work of Andersen, Bollerslev, Diebold and their various co-authors on realized volatility and volatility forecasting is nicely summarized in their review article “Volatility and Correlation Forecasting” in the Handbook of Economic Forecasting, see Andersen et al. (2006). Furthermore, the excellent book by Francq and Zakoian (2011) provides a comprehensive and systematic approach to understanding GARCH time series models and their applications whilst presenting the most advanced results concerning the theory and practical aspects of GARCH.

With regards time series that are locally but not globally stationary, Priestley (1965, 1988) and Dahlhaus (1997) were the pioneering works; see Dahlhaus (2012) for a review. In terms of applications to ARCH/GARCH modeling, Fryzlewicz et al. (2006, 2008) and Dahlhaus and Rao (2006, 2007) all work in a context of local stationarity, and develop ARCH processes with time varying parameters.

The paper is organized as follows: Section 2 reviews the work of Politis (2007) on the NoVaS transformation for point prediction of stationary financial returns, and compares the performance of NoVaS to GARCH(1,1) in one-step-ahead point prediction of squared returns under different assumptions (including non-stationarities). Section 3 addresses the comparison of the performance of GARCH(1,1) and NoVaS in volatility estimation; interestingly, NoVaS methods have not been used previously for this purpose. Section 4 presents the novel Model-Free algorithms for interval prediction of squared returns, and illustrates their numerical performance by means of both simulated examples and real world data. Concluding remarks are provided in Section 5.

2 Point Prediction

In this section we compare the performance of NoVaS transformation and GARCH(1,1) in one-step-ahead point prediction of squared returns under different assumptions (including non-stationarities).

2.1 Review of GARCH(1,1) and NoVaS Transformation for Stationary Data

2.1.1 The Benchmark: GARCH(1,1)

Let Y_1, \dots, Y_n be an observed stretch from a financial returns time series $\{Y_t, t \in \mathbf{Z}\}$. To start with, assume that $\{Y_t\}$ is (strictly) stationary with mean zero, which implies that trends and other non-stationarities have been successfully removed.

The standard ARCH(p) model is described by an equation of the type:

$$Y_t = Z_t \sqrt{a + \sum_{i=1}^p a_i Y_{t-i}^2} \quad (1)$$

where the series $\{Z_t\}$ is assumed to be independent and identically distributed (i.i.d.), and p is a positive integer indicating the order of the model. Engle (1982) assumed that the errors Z_t have a $N(0, 1)$ distribution; later on other researchers considered instead a t -distribution (standardized to unit variance) for the errors—see e.g. Shephard (1996).

Let \mathcal{F}_n be short-hand for the observed information set, i.e., $\mathcal{F}_n = \{Y_t, 1 \leq t \leq n\}$. Under the above ARCH(p) model, the L_2 optimal predictor of Y_{n+1}^2 based on \mathcal{F}_n is

$$E(Y_{n+1}^2 | \mathcal{F}_n) = a + \sum_{i=1}^p a_i Y_{n+1-i}^2. \quad (2)$$

This conditional expectation $E(Y_{n+1}^2 | \mathcal{F}_n)$ is commonly referred to as the *volatility*, although the same term is sometimes also used for its square root.

A standard GARCH(1,1) model is described by the equation:

$$Y_t = h_t Z_t \quad \text{with} \quad h_t^2 = C + AY_{t-1}^2 + Bh_{t-1}^2 \quad (3)$$

where the series $\{Z_t\}$ is *i.i.d.* $(0, 1)$, and the parameters A, B, C are assumed non-negative. The quantity $h_t^2 = E(Y_t^2 | \mathcal{F}_{t-1})$ is the *volatility* of the GARCH model; it is also the L_2 optimal predictor of Y_t^2 given \mathcal{F}_{t-1} .

Back-solving in the right-hand-side of Eq. (3), it is easy to show that the GARCH(1,1) model is tantamount to the ARCH model of Eq. (1) with $p = \infty$ and the following identifications:

$$a = \frac{C}{1 - B}, \quad \text{and} \quad a_i = AB^{i-1} \quad \text{for} \quad i = 1, 2, \dots \quad (4)$$

Under the objective of L_1 optimal prediction, the optimal predictor is the conditional median—not the conditional expectation. For an ARCH(p) process, the L_1 optimal predictor of Y_{n+1}^2 given \mathcal{F}_n is

$$\text{Median}(Y_{n+1}^2 | \mathcal{F}_n) = \left(a + \sum_{i=1}^p a_i Y_{n+1-i}^2 \right) \cdot \text{Median}(Z_{n+1}^2 | \mathcal{F}_n). \quad (5)$$

Under the usual causality assumption, the error Z_t is independent of \mathcal{F}_{t-1} for all t . Hence, $\text{Median}(Z_{n+1}^2 | \mathcal{F}_n) = \text{Median}(Z_{n+1}^2)$.

Furthermore, the aforementioned equivalence of GARCH(1,1) to a particular ARCH(∞) model implies that Eq. (5) would also give the L_1 optimal GARCH(1,1) predictor of Y_{n+1}^2 by allowing $p = \infty$ with the ARCH coefficients a, a_1, a_2, \dots having the structure given by Eq. (4). Alternatively, under the GARCH(1,1) model we can simply write

$$\text{Median}(Y_{n+1}^2 | \mathcal{F}_n) = h_{n+1}^2 \cdot \text{Median}(Z_{n+1}^2). \quad (6)$$

Note that if Z_t has an (approximately) symmetric unimodal distribution, then Z_t^2 is a positively skewed random variable. It follows that $\text{Median}(Z_{n+1}^2) \leq E(Z_{n+1}^2) = 1$. For example, $\text{Median}(Z_{n+1}^2) \simeq 0.45$ if Z_t is $N(0, 1)$, whereas $\text{Median}(Z_{n+1}^2) \simeq 0.53$ if Z_t has a t distribution with 5 degrees of freedom. Hence, in both ARCH and GARCH cases, the L_1 optimal predictor of Y_{n+1}^2 is quite smaller than its L_2 analog;

The GARCH(1,1) is by far the most popular of the ARCH/GARCH models, and typically forms the benchmark for modeling financial returns. That is why we compare the predictive ability of NoVaS methodology to that of GARCH(1,1) in what follows.

Remark. If the objective is point prediction of Y_t^2 given \mathcal{F}_{t-1} , should one choose and L_2 or L_1 optimal predictor? The crux here is that financial returns are governed by

a fat-tailed distribution; it is generally accepted that they have a finite variance but it is not clear that they have a finite fourth moment. An L_2 measure of prediction of the squares is essentially a fourth moment; if it is infinite, it does not make sense to try to minimize it! The paper by Politis (2007) brought attention to this problem and advocated adopting an L_1 measure of prediction of the squares, i.e. prediction by conditional median (not mean); see also Ch. 10.4.1 of Politis (2015).

2.1.2 NoVaS Methodology

Let us continue considering a zero mean and (strictly) stationary financial return time series $\{Y_t, t \in \mathbf{Z}\}$. The NoVaS methodology attempts to map the dataset Y_1, \dots, Y_n to a sample from a Gaussian stationary series.

The starting point is the ARCH model of Eq. (1) that can be re-written as

$$Z_t = \frac{Y_t}{\sqrt{a + \sum_{i=1}^p a_i Y_{t-i}^2}}. \quad (7)$$

Engle (1982) assumed that the errors Z_t are i.i.d. $N(0, 1)$; this assumption was later challenged by empirical researchers but we re-visit it here. Note that the ratio given in Eq. (7) can be interpreted as an attempt to ‘studentize’ the return Y_t by dividing with a time-localized measure of the standard deviation of Y_t . However, there seems to be no reason to exclude the value of Y_t from an empirical, causal estimate of the standard deviation of Y_t ; recall that a causal estimate is one involving present and past data only, i.e., the data $\{Y_s, s \leq t\}$.

Under this rationale, Politis (2003) defined a new ‘studentized’ quantity as follows:

$$W_{t,a} := \frac{Y_t}{\sqrt{a s_{t-1}^2 + a_0 Y_t^2 + \sum_{i=1}^p a_i Y_{t-i}^2}} \quad \text{for } t = p+1, p+2, \dots, n. \quad (8)$$

In the above, s_{t-1}^2 is an estimator of the unconditional variance $\sigma_Y^2 = \text{Var}(Y_t)$ based on the data up to (but not including¹) time t ; under the zero mean assumption for Y_t , the natural estimator is $s_{t-1}^2 = (t-1)^{-1} \sum_{k=1}^{t-1} Y_k^2$.

The definition in Eq. (8) describes the basic normalizing and variance-stabilizing transformation (NoVaS) under which the data series $\{Y_t\}$ is mapped to the new series

¹The reason for not including time t in the variance estimator is for purposes of notational clarity as well as the easy identifiability of the effect of the coefficient a_0 associated with Y_t^2 in the denominator of Eq. (8).

$\{W_{t,a}\}$. The order p (≥ 0) and the vector of nonnegative parameters $(\alpha, a_0, \dots, a_p)$ are chosen by the practitioner with the twin goals of normalization and variance stabilization.

The NoVaS transformation of Eq. (8) can be re-arranged to yield:

$$Y_t = W_{t,a} \sqrt{\alpha s_{t-1}^2 + a_0 Y_t^2 + \sum_{i=1}^p a_i Y_{t-i}^2}. \quad (9)$$

The only essential difference between the NoVaS of Eq. (9) and the ARCH of Eq. (1) is the presence of the term Y_t^2 in the right-hand-side paired with the coefficient a_0 . Replacing the term a in Eq. (1) by the term αs_{t-1}^2 in Eq. (9) is natural since the former has—by necessity—units of variance; in other words, the term a in Eq. (1) is not scale invariant, whereas the term α in Eq. (9) is.

The target of variance stabilization for the new series $\{W_{t,a}\}$ relies on the construction of a local estimator of scale for studentization, and requires the ‘unbiasedness’ condition:

$$\alpha + \sum_{i=0}^p a_i = 1 \quad \text{where } \alpha \geq 0, \quad a_i \geq 0 \quad \text{for all } i \geq 0. \quad (10)$$

Eq. (10) has the interesting implication that the $\{W_{t,a}\}$ series can be assumed to have an (unconditional) variance that is (approximately) unity. Nevertheless, note that p and α, a_0, \dots, a_p must be carefully chosen to achieve a degree of conditional homoscedasticity as well; to do this, one must necessarily take p small enough—as well as α small enough or even equal to zero—so that a local (as opposed to global) estimator of scale is obtained.

Politis (2003) provided two structures for the a_i coefficients satisfying Eq. (10). One is to let $\alpha = 0$ and $a_i = 1/(p+1)$ for all $0 \leq i \leq p$; this specification is called the *simple* NoVaS transformation. The other one is given by the *exponential* decay NoVaS where $\alpha = 0$ and $a_i = c^i e^{-ci}$ for all $0 \leq i \leq p$. The simple and exponential NoVaS schemes are very intuitive as they correspond to the two popular time series methods of obtaining a ‘local’, one-sided average, namely a moving average (of the last $p+1$ values) and ‘exponential smoothing’; see e.g. Hamilton (1994).

With regards the target of normalizing the distribution of the series $\{W_{t,a}\}$, note

that

$$\frac{1}{W_{t,a}^2} = \frac{\alpha s_{t-1}^2 + a_0 Y_t^2 + \sum_{i=1}^p a_i Y_{t-i}^2}{Y_t^2} \geq a_0$$

since all the parameters are nonnegative; therefore, the random variable $W_{t,a}$ is bounded:

$$|W_{t,a}| \leq 1/\sqrt{a_0} \quad (11)$$

So one must be careful to ensure that the $\{W_{t,a}\}$ variables have a large enough range such that the boundedness is not seen as spoiling the normality. Thus, we also require

$$\frac{1}{\sqrt{a_0}} \geq C \quad \text{i.e.,} \quad a_0 \leq 1/C^2 \quad (12)$$

for some appropriate C of the practitioner's choice. Recalling that 99.7% of the mass of the $N(0, 1)$ distribution is found in the range ± 3 , the simple choice $C = 3$ can be suggested; this choice seems to work reasonably well for the usual samples sizes.

2.1.3 Basic NoVaS schemes

We then proceed to choose p and $\alpha, a_0, a_1, \dots, a_p$ (the parameters needed to identify) with the optimization goal of making the $\{W_{t,a}\}$ series as close to normal as possible. To quantify this goal, one can attempt to minimize a (pseudo)distance measuring departure of the transformed data from normality. Recall that it is a matter of common practice to assume that the distribution of financial returns is *symmetric* (at least to a first approximation) and therefore, the skewness of financial returns is often ignored. In contrast, the kurtosis is typically quite large, indicating a heavy tailed distribution. Hence, the kurtosis can serve as a simple measure of the departure of a (non-skewed) dataset from normality. Other measures of non-normality can be used, e.g., the Shapiro-Wilk statistic or Kolmogorov distance; see Politis and Thomakos (2008), Politis and Thomakos (2013). In practice, all these measures give similar results; hence, we focus on the kurtosis measure which is also easy to interpret.

Let $KURT_n(Y)$ denote the empirical kurtosis of a dataset $\{Y_t, t = 1, \dots, n\}$, i.e.,

$$KURT_n(Y) = \frac{n^{-1} \sum_{t=1}^n (Y_t - \bar{Y})^4}{(n^{-1} \sum_{t=1}^n (Y_t - \bar{Y})^2)^2}$$

where $\bar{Y} = n^{-1} \sum_{t=1}^n Y_t$ denotes the sample mean.

The following two algorithms can be used to select the optimal parameters for NoVaS; see Politis (2003). Note that the only free parameter in Simple NoVaS is the order p ; therefore, the Simple NoVaS transformation will be denoted by $W_{t,p}^S$. In the Exponential NoVaS, to specify all the a_i s, one just needs to specify the two parameters p and $c > 0$, in view of Eq. (10). However, because of the exponential decay, the parameter p is now of secondary significance; thus, we concisely denote the exponential NoVaS transformation by $W_{t,c}^E$.

ALGORITHM 2.1. SIMPLE NOVAS

1. Let $\alpha = 0$ and $a_i = 1/(p + 1)$ for all $0 \leq i \leq p$.
2. Pick p such that $|KURT_n(W_{t,p}^S) - 3|$ is minimized.

In the above, the value 3 as a target kurtosis corresponds to the kurtosis of the Gaussian distribution.

ALGORITHM 2.2. EXPONENTIAL NOVAS (EXP-NOVAS)

1. Let p take a very high starting value, e.g., let $p \simeq n/4$ or $n/5$. Then, let $\alpha = 0$ and $a_i = c'e^{-ci}$ for all $0 \leq i \leq p$, where $c' = 1/\sum_{i=0}^p e^{-ci}$ by Eq. (10).
2. Pick $c > 0$ in such a way that $|KURT_n(W_{t,c}^E) - 3|$ is minimized.

Step 2 in the above could be better understood as a moment matching criterion, i.e., replace Step 2 by Step 2' below:

- 2'. Pick $c > 0$ such that $KURT_n(W_{t,p}^S) = 3$.

This is true because it is generally possible to choose the NoVaS parameters (in this case the value of c) to match the Gaussian kurtosis exactly; see Politis (2007) for the argument based on the intermediate value theorem.

The search in Algorithm 2 is over $c \in (0, \infty)$ which appears formidable; what makes this optimization problem well-behaved is that we know that high values of c can not plausibly be solutions. To see why, note that if c is large, then $a_i \approx 0$ for $i = 1, 2, \dots$ and hence $W_{t,c}^E \approx Y_t$ which has kurtosis much larger than 3 by assumption. It is apparent that the search for the optimal c will be practically conducted over a discrete grid of c -values spanning an interval of the type $(0, s]$ for some s of the

order of one. A practical way to narrow in on the optimal c value is to run two grid searches, one coarse followed by a fine one: (i) use a coarse grid search over the whole interval $(0, s]$, and denote \tilde{c}_0 the minimizer over the coarse grid search; and (ii) run a fine grid search over a neighborhood of \tilde{c}_0 . Let c_0 denote the resulting minimizer from the above algorithm. If needed, the following range-adjustment step may be added.

3. If c_0 as found above is such that Eq. (12) is not satisfied, then decrease c stepwise (starting from c_0) over the discrete grid until Eq. (12) is satisfied.

Finally, the value of p must be trimmed for efficiency of usage of the available sample; to do this we can simply discard the a_i coefficients that are close to zero, i.e., those that fall below a certain threshold/tolerance level ϵ which is the practitioner's choice. A threshold value of $\epsilon = 0.01$ is reasonable in connection with the a_i which—as should be stressed—are normalized to sum to one.

4. Trim the value of p by a criterion of the type: if $a_i < \epsilon$, then let $a_i = 0$. If i_0 is the smallest integer such that $a_i < \epsilon$ for all $i \geq i_0$, then let $p = i_0$ and *re-normalize* the a_i s so that their sum (for $i = 0, 1, \dots, i_0$) equals one.

Remark. Note that the NoVaS approach is not model-based, and therefore Maximum Likelihood Estimation (MLE) or Quasi-MLE is not advisable. In fact, there is a different philosophy at work here. The traditional philosophy starts with assuming a model (typically driven by i.i.d. errors), and using MLE (or related techniques) to fit it. In a model-free approach such as NoVaS, we do not assume any model as being true; rather, we use the structure of the data at hand in order to devise a transformation that maps our dataset $\{Y_t\}$ to a new dataset $\{W_{t,a}\}$ with i.i.d. entries. The criterion of success of the model-free approach is the success of the transformation; this is a different criterion than MLE.

2.1.4 Generalized NoVaS schemes

Simple NoVaS has only one parameter, and the same is (effectively) true for Exponential NoVaS. Although many different multi-parameter NoVaS schemes can be devised, we now elaborate on the possibility of a nonzero value for the parameter

α (say $\alpha_1, \alpha_2, \dots, \alpha_K$, that span a subset of the interval $[0,1]$) in Eq. (8) in connection with the Simple and Exponential NoVaS. We thus define the Generalized Simple (GS) and Generalized Exponential (GE) NoVaS denoted by $W_{t;p,\alpha}^{GS}$ and $W_{t;c,\alpha}^{GE}$ indicating their respective two free parameters; both are based on Eq. (8).

ALGORITHM 2.3. GENERALIZED SIMPLE NOVAS (GS NOVAS)

A. For $k = 1, \dots, K$ perform the following steps.

1. Let $\alpha = \alpha_k$ and $a_i = (1 - \alpha_k)/(p + 1)$ for all $0 \leq i \leq p$ so that eq. (10) is satisfied while all the coefficients a_0, a_1, \dots, a_p are the same.
2. Denote by p_k the minimizer of $|KURT_n(W_{t,p}^{GS}) - 3|$ over values of $p = 1, 2, \dots$
3. If p_k (and a_0) as found above are such that eq. (12) is not satisfied, then increase p_k accordingly, i.e., re-define $p_k = \lfloor 1 + C^2(1 - \alpha_k) \rfloor$, and let $a_i = (1 - \alpha_k)/(p_k + 1)$ for all $0 \leq i \leq p_k$ by Eq. (10); here, $\lfloor x \rfloor$ denotes the integer part of x .

B. Finally, compare the transformations $\{W_{t;p_k,\alpha_k}^{GS}, k = 1, \dots, K\}$ in terms of their performance in a desired task, e.g., volatility prediction, and pick the value of k leading to optimal performance.

ALGORITHM 2.4. GENERALIZED EXPONENTIAL NOVAS (GE NOVAS)

A. For $k = 1, \dots, K$ perform the following steps.

1. Let p take a very high starting value, e.g., let $p \simeq n/4$ or $n/5$. Then, let $\alpha = \alpha_k$ and $a_i = c' e^{-ci}$ for all $0 \leq i \leq p$, where $c' = (1 - \alpha_k) / \sum_{i=0}^p e^{-ci}$ by Eq. (10).
2. Pick c in such a way that $|KURT_n(W_{t;c,\alpha_k}^{GE}) - 3|$ is minimized, and denote by c_k the minimizing value.²
3. Trim the value of p to some value p_k as before: if $a_i < \epsilon$, then set $a_i = 0$. Thus, if $a_i < \epsilon$, for all $i \geq i_k$, then let $p_k = i_k$, and re-normalize the a_i s so that their sum (for $i = 0, 1, \dots, p_k$) equals $1 - \alpha_k$ by Eq. (10).

B. Finally, compare the transformations $\{W_{t;c_k,\alpha_k}^{GE}, k = 1, \dots, K\}$ in terms of their performance in a desired task, e.g., volatility prediction, and pick the value of k leading to optimal performance.

Remark. For all empirical work in this paper, we chose $K = 8$ and $\alpha_k = 0, 0.1, 0.2, \dots, 0.7$ in connection with Generalized NoVaS—either GS or GE. For the Exponential NoVaS and GE NoVaS algorithms we employed the choices $p = n/4$ and $\epsilon = 0.01$.

2.1.5 NoVaS prediction equation

Suppose that the NoVaS parameters, i.e., the order $p(\geq 0)$ and the parameters α, a_0, \dots, a_p have already been chosen. Re-arrange the NoVaS Eq. (8) to yield:

$$Y_t^2 = \frac{W_{t,a}^2}{1 - a_0 W_{t,a}^2} \left(\alpha s_{t-1}^2 + \sum_{i=1}^p a_i Y_{t-i}^2 \right) \quad \text{for } t = p+1, \dots, n \quad (13)$$

and

$$Y_t = \frac{W_{t,a}}{\sqrt{1 - a_0 W_{t,a}^2}} \sqrt{\alpha s_{t-1}^2 + \sum_{i=1}^p a_i Y_{t-i}^2} \quad \text{for } t = p+1, \dots, n. \quad (14)$$

²As before, if c_k is such that Eq. (12) is not satisfied, then decrease it stepwise over its discrete grid until Eq. (12) is satisfied.

Let $g(\cdot)$ be some (measurable) function of interest; examples include $g_0(x) = x$, $g_1(x) = |x|$, and $g_2(x) = x^2$, the latter being the function of interest for volatility prediction. From Eq.(14) it follows that the predictive (given \mathcal{F}_n) distribution of $g(Y_{n+1})$ is identical to the distribution of the random variable

$$g\left(A_n \frac{W}{\sqrt{1 - a_0 W^2}}\right) \quad (15)$$

where $A_n = \sqrt{\alpha s_n^2 + \sum_{i=1}^p a_i Y_{n+1-i}^2}$ is treated as a constant given the past \mathcal{F}_n , and the random variable W has the same distribution as the conditional (on \mathcal{F}_n) distribution of the random variable $W_{n+1,a}$.

Therefore, the L_1 optimal prediction of $g(Y_{n+1})$ given \mathcal{F}_n is given by the median of the conditional (given \mathcal{F}_n) distribution of $g(Y_{n+1})$, i.e.,

$$g(\widehat{Y}_{n+1}) := \text{Median} \left(g \left(A_n \frac{W_{n+1,a}}{\sqrt{1 - a_0 W_{n+1,a}^2}} \right) \middle| \mathcal{F}_n \right) \quad (16)$$

Specializing to the case of our interest, i.e., volatility prediction and the function $g_2(x) = x^2$ yields the NoVaS predictor:

$$\widehat{Y}_{n+1}^2 = \mu_2 A_n^2 \quad \text{where} \quad \mu_2 = \text{Median} \left(\frac{W_{n+1,a}^2}{1 - a_0 W_{n+1,a}^2} \middle| \mathcal{F}_n \right). \quad (17)$$

Remark. By construction, the series $\{W_{t,a}\}$ is approximately Gaussian but can be correlated. Nevertheless, applied work involving numerous real datasets has invariably shown that the series $\{W_{t,a}\}$ is effectively white noise, and hence it is i.i.d. due to Gaussianity; see Politis (2015) and the references therein. As a result, the conditioning on \mathcal{F}_n can be dropped in both Eq. (16) and (17). In this case, a quantity such as μ_2 of Eq. (17) is the unconditional median, and can readily be estimated by the sample median of $\{W_{t,a}$ for $t = p + 1, \dots, n\}$.

The above remark helps clarify the philosophical distinction between model-based and model-free inference. For example, defining $U_{t,a} = W_{t,a} / \sqrt{1 - a_0 W_{t,a}^2}$ we can rewrite eq. (14) as

$$Y_t = U_{t,a} \sqrt{\alpha s_{t-1}^2 + \sum_{i=1}^p a_i Y_{t-i}^2} \quad (18)$$

which formally looks like the defining equation of an ARCH(p) model. The difference is that in the model-based approach, one assumes that the $U_{t,a}$ are i.i.d. in an equation such as (18), and then proceeds to estimate the coefficients a_i by (say) Maximum Likelihood. By contrast, in the model-free setting of NoVaS, the $W_{t,a}$ (and therefore the $U_{t,a}$ as well) are i.i.d. *by construction*, not by assumption. Furthermore, this construction has already taken place before being able to even write down eq. (18). In other words, the predictive equation (18) is the *conclusion* of the model-free fitting of NoVaS, where in a model-based setup such an equation would be the *starting point*.

2.2 Local Stationarity

When we consider time series data Y_1, \dots, Y_n spanning a long time interval, e.g., annual rainfall measurements spanning over 100 years or daily financial returns spanning several decades, it may be unrealistic to assume that the stochastic structure of time series $\{Y_t, t \in \mathbf{Z}\}$ has stayed invariant over such a long stretch of time, i.e., we can not assume that $\{Y_t, t \in \mathbf{Z}\}$ is stationary. It may be more plausible to assume a slowly-changing stochastic structure, i.e., a *locally stationary model*; see Priestley (1965, 1988), Dahlhaus (1997) and the review of Dahlhaus (2012).

2.2.1 Time-varying GARCH and Time-varying NoVaS

Intuitively, a locally stationary process can be considered as approximately stationary over a short time window which has led to the notion of Time-varying ARCH/GARCH models. In particular, the theory of Time-varying ARCH (TV-ARCH) processes was developed, and the asymptotic properties of weighted quasi-likelihood estimators were studied in detail by Dahlhaus and Rao (2006).

The analysis of a Time-varying ARCH/GARCH model can be based on the premise of local stationarity. For example, in order to predict $g(Y_{t+1})$ based on \mathcal{F}_t via a Time-varying GARCH(1,1) model, we can simply fit the GARCH(1,1) model of Eq. (3) using the ‘windowed’ data Y_{t-b+1}, \dots, Y_t ; as a result, the coefficients of the fitted GARCH(1,1) model are varying with time. Here, the window size b should be large enough so that accurate estimation of the GARCH parameters is possible based on the subseries Y_{t-b+1}, \dots, Y_t but small enough so that such a subseries can plausibly be considered stationary.

In a similar vein, we can predict $g(Y_{t+1})$ by fitting one of the NoVaS algorithms (Simple vs. Exponential, Generalized or not) just using the ‘windowed’ data Y_{t-b+1}, \dots, Y_t . In so doing, we are constructing a **Time-varying NoVaS** (TV-NoVaS) transformation as discussed in Politis (2015). Recall that in extensive numerical work, Politis and Thomakos (2008), Politis and Thomakos (2013) showed that NoVaS fitting can be done more efficiently than GARCH fitting by (numerical) MLE. Thus, it is conjectured that TV-NoVaS may be able to capture a slowly changing stochastic structure in a more flexible manner; stated in different term, the window size b required for accurate NoVaS fitting could be smaller than the one required for accurate GARCH fitting, thus allowing us to capture the time-varying stochastic structure more accurately.

2.2.2 Structural Breaks and Change Points

A different form of non-stationarity is due to the possible presence of structural breaks, i.e., change points, occurring at some isolated time points. Kokoszka and Leipus (2000), and Berkes et al. (2004) have studied the detection/estimation of change points in ARCH/GARCH modeling. Mikosch and Starica (2004) and Stărică and Granger (2005) show the interesting effects that an undetected change point may have on our interpretation and analysis of ARCH/GARCH modeling. Polzehl and Spokoiny (2006) show that time-varying models can give a better predictive power for time series with change points.

Fitting ARCH/GARCH models with change points can be quite cumbersome from a practical point of view. So it is of interest to investigate the performance of the aforementioned time-varying GARCH and NoVaS methods given data that are stationary except for breaks. If data over short windows are used, then TV-GARCH and TV-NoVaS should perform generally well except in the immediate neighborhood of the change points. Hence, in the simulations that follow, we also include a structural break model in order to see the effect of an undetected change point on the performance of TV-GARCH and TV-NoVaS predictors for squared returns.

2.3 Simulations and Results

Using simulated data, we can illustrate and compare the one-step ahead prediction performance of NoVaS with that of the benchmark GARCH(1,1) model. Since the

stationary case was dealt with in detail by Politis (2007, 2015), in what follows we focus on the non-stationary case using either local stationary models or stationary models with breaks as a Data Generating Process (DGP).

2.3.1 Simulation Design

For the simulation, 500 datasets $\underline{Y}_n = (Y_1, \dots, Y_n)'$ were constructed using either a TV-GARCH or a change point GARCH (CP-GARCH) model; these were defined using the standard GARCH model of Eq. (3) as building block with $C = 10^{-5}$. The *i.i.d.* errors Z_t associated with a GARCH model are often assumed to have a Student t_5 distribution. Instead, we use the simple assignment $Z_t \sim \text{i.i.d. } N(0, 1)$ in the simulation in order to give the GARCH fitting a “head-start”; using a GARCH with normal errors as DGP should facilitate the convergence of the numerical (Gaussian) MLE employed in fitting the TV-GARCH model.

The exact models used to generate simulated data are as follows:

CP-GARCH: For $t \leq n/2$, let $A = 0.10$ and $B = 0.73$; for $t > n/2$, let $A = 0.05$ and $B = 0.93$. These values are close to the ones used by Mikosch and Starica (2004).

TV-GARCH: The value of A decreases as a linear function of t , starting at $A = 0.10$ for $t = 1$, and ending at $A = 0.05$ for $t = n$. At the same time, the value of B increases as a linear function of t , starting at $B = 0.73$ for $t = 1$, and ending at $B = 0.93$ for $t = n$.

The difference between the CP-GARCH and the TV-GARCH model, is an abrupt vs. a smooth transition of the parameters spanning the same values. Some more information on the simulation follows.

- The prediction method employed was the conditional median obtained either from a TV-GARCH model with normal errors fitted by windowed Gaussian MLE, or via TV-NoVaS (Simple or Exponential); in either case, two window sizes were tried out, namely $b = 125$ or 250 .
- The sample size was $n = 1001$ corresponding to about 4 years of daily data; so the choices $b = 125$ and 250 correspond to 6 and 12 months respectively.

- Training period for all methods was 250, i.e., the experiment amounted to predicting Y_{t+1}^2 from the ‘windowed’ data Y_{t-b+1}, \dots, Y_t for $t = 250, 251, \dots, 1000$.
- Updating (re-estimation) of all methods would ideally be for each $t = 250, 251, \dots, 1000$. To save computing time, updating in the simulation was only performed for t being an integer multiple of 50. In fairness, the performance of predictions was recorded and compared *only* at the moment of updating the model, i.e., at time points 250, 300, 350, \dots , 1000. For each point, we give the mean absolute deviation (MAD) of the prediction error at the update time point averaged over the 500 replications.

2.3.2 Results and Conclusions

Figure 1 shows the MAD of volatility prediction errors of TV-GARCH as compared to that of TV-NoVaS (Simple or Exponential) with data from model CP-GARCH for the 16 time points where the updating and prediction occurred, i.e., the time points 250, 300, 350, \dots , 1000. The left panel depicts the case $b = 125$ while the right panel depicts the case $b = 250$. Figure 2 is similar but using data generated by a TV-GARCH model instead.

Some conclusions are as follows:

- The CP-GARCH model of Figure 1 is stationary up to time point 500 where the break occurs. Hence, time points 250, 300, 350, 400, 450, and 500 in the left panel of Figure 1 corroborate the aforementioned fact that NoVaS (Simple or Exponential) beats GARCH for prediction of squared returns even if the data generating model is (stationary) GARCH as long as the sample size available for model-fitting is small—equal to 125 in this case which is the size of the local window. The corresponding points in the right panel of Figure 1 indicate that GARCH manages to do as well as (or better than)³ NoVaS when the effective sample size is increased to 250.

³Note that here GARCH is fitted by Gaussian MLE with only three free parameters; in the more realistic case of four parameter MLE using the t distribution—the fourth parameter being the degrees of freedom—GARCH may need a sample size of at least 350 in order to outperform NoVaS even though the true underlying model is GARCH indeed; see Politis and Thomakos (2008), Politis and Thomakos (2013).

- Figure 1 shows that the change point at $t = 500$ wreaks havoc in TV-GARCH model fitting and the associated predictions; this adds another dimension to the observations of Mikosch and Starica (2004). By contrast, both NoVaS methods seem to adapt immediately to the new regime that occurs after the unknown/undetected change point.
- Figure 2 shows that TV-NoVaS (Simple or Exponential) beats TV-GARCH for prediction of squared returns even when the data generating model is indeed TV-GARCH. Not only is the MAD of prediction of TV-NoVaS just a small fraction of that of TV-GARCH, but the wild swings associated with the latter indicate the inherent instability of GARCH model-fitting; this instability is prominent even in this simplistic case where the errors have a true Gaussian distribution, and Gaussian MLE is used for estimating just the three GARCH parameters.
- As seen in both Figures 1 and 2, the performance of Simple NoVaS is practically indistinguishable from that of Exponential NoVaS although upon closer look the latter appears to be marginally better.
- It is possible to also consider Time-varying version of Generalized NoVaS both Simple and Exponential. However, in the context of the present simulation the results of TV GS NoVaS and TV GE NoVaS are practically indistinguishable from the results of the Simple and Exp-Novas; they were not added to Figures 1 and 2 in order not to clutter the pictures and maintain their readability.

Remark. GARCH model fitting via (numerical) MLE is well-known to be problematic unless the sample size is quite large, e.g. more than 300. However, as discussed in the Remark given at the end of subsection 2.1.3, NoVaS fitting is straightforward and does not employ numerical MLE. This is what makes NoVaS outperform GARCH even when the model is GARCH with a moderate sample size; see the left panel of Figure 1 where for the first 500 entries the model is pure GARCH. In this sense, is it expected that TV-GARCH is not performing well with locally stationary data (see Figure 2), since the MLE fitting has to be done over short windows.

3 Estimation of Realized Volatility

In this section, we compare the performance of NoVaS methods vs. GARCH models in the estimation of daily realized volatility with real world time series. Based on the empirical work of Politis (2003, 2007), the NoVaS series $\{W_{t,a}\}$ appears to be uncorrelated; this is true for a variety of daily returns including stock prices, stock indices, and foreign exchange. Since $\{W_{t,a}\}$ is (approximately) Gaussian, then we can infer that the series $\{W_{t,a}\}$ is not only uncorrelated but also independent; therefore, it is straightforward to construct a Model-free estimate of the conditional expectation $E(Y_{n+1}^2|\mathcal{F}_n)$. In this case, Eq. (13) implies that $E(Y_{n+1}^2|\mathcal{F}_n) = A_n^2 E\left(\frac{W_{t,a}^2}{1-a_0 W_{t,a}^2}\right)$; the NoVaS estimate of realized volatility therefore is

$$\frac{A_n^2}{n-p} \sum_{t=p+1}^n \left(\frac{W_{t,a}^2}{1-a_0 W_{t,a}^2} \right) \quad (19)$$

which has validity, e.g. consistency, under the sole assumption that Y_t has a finite second moment conditionally on \mathcal{F}_n (and therefore unconditionally as well). To examine the performance of this estimator, we conduct a real-data empirical analysis of estimating the daily realized volatility either using the formula (19) or fitting a GARCH(1,1) model.

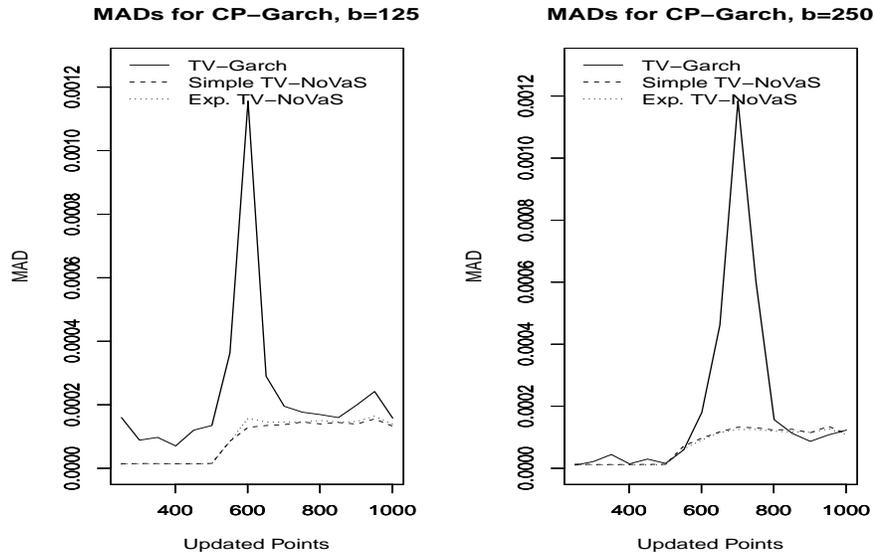


Figure 1: MAD of prediction of squared returns obtained by fitting TV-GARCH vs. TV-NoVaS; data from CP-GARCH model.

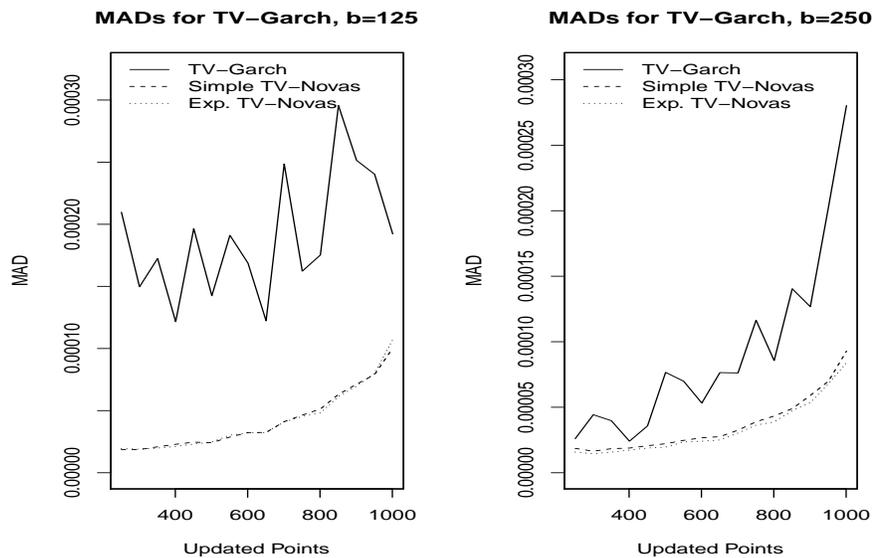


Figure 2: MAD of prediction of squared returns obtained by fitting TV-GARCH vs. TV-NoVaS; data from TV-GARCH model.

3.1 Real Data and Summary Statistics

We consider two real world time series; both are intraday 30-minute IBM stock returns over different time periods and associated with the daily realized volatility. The sample period of the first dataset is from 07-01-2010 to 07-31-2012 for a total of $n = 526$ days. The sample period of the second dataset is from 01-02-2013 to 11-10-2016 for a total of $n = 975$ days. Weekends and holidays are excluded for both series. In Figures 3, we present graphs for the daily realized volatility of the two datasets. The associated daily realized volatility was constructed by summing all 30-minute squared returns of one day for IBM stock. Specifically, if we denote by $r_{t,i}$, the i^{th} daily return for day t , then the daily realized volatility is defined as $h_t^2 \stackrel{\text{def}}{=} \sum_{i=1}^{n_t} r_{t,i}^2$, where n_t is the total number of 30-minutes intervals during day t of open market.

3.2 NoVaS and GARCH(1,1): Optimization and Estimating Specifications

To estimate the realized volatility for our two series, we continue using the TV-NoVaS transformation and TV-GARCH(1,1) models. All forecasts we make are ‘*honest*’, i.e., we use only observations prior to the time period to be forecasted. The parameters of NoVaS approach and GARCH(1,1) models are re-estimated as the window rolls over the entire evaluation sample. The window sizes we choose are $m = 126$ (six months) and $m = 252$ (one year). We compare the performance of the NoVaS approach to the standard GARCH(1,1) with normal errors, and to GARCH(1,1) with errors following a $t_{(v)}$ distribution with degree of freedom estimated from the data. For completeness, we employ all NoVaS algorithms: Simple-NoVaS, Exp-NoVaS, GS NoVaS and GE NoVaS. For Exp-NoVaS and GE NoVaS algorithms, we set $p_{max} = n/4$ and the trimming threshold of 0.01 as in Section 2.

In the process of analysis, we always evaluate our estimation using the ‘true’ realized volatility measure given in the previous subsection and report the mean absolute deviation (MAD) and root mean-squared error (RMSE) of the estimation

errors $e_t \stackrel{\text{def}}{=} h_t^2 - \widehat{h}_t^2$, given by:

$$MAD(e) \stackrel{\text{def}}{=} \frac{1}{n-m} \sum_{t=m+1}^n |e_t - \bar{e}|, \text{ and } RMSE(e) \stackrel{\text{def}}{=} \sqrt{\frac{1}{n-m} \sum_{t=m+1}^n (e_t - \bar{e})^2}$$

where \widehat{h}_t^2 denotes the estimation of the daily realized volatility for any of the methods used, and $\bar{e} \stackrel{\text{def}}{=} \frac{1}{n-m} \sum_{t=m+1}^n e_t$.

3.3 Results and Conclusions

Our estimation results are summarized in Tables 1 and 2. Table 1 gives the results for the series from 07-01-2010 to 07-31-2012, and Table 2 for the series from 01-02-2013 to 11-10-2016. In the second columns of each table are the MADs and RMSEs of TV-NoVaS and TV-GARCH(1,1) with window size 126. The third columns in both tables record the MADs and RMSEs with window size 252. Some general comments:

- For all methodologies, the MADs and RMSEs of the estimation errors of daily realized volatility are decreasing when the window size is smaller, i.e., 126; this may indicate a non-stationarity in the data.
- The TV-GARCH(1,1) models with student t innovations outperform the TV-GARCH(1,1) models with Normal innovations for both series; this should not be surprising in view of the mainstream GARCH literature.
- The performance of TV-Exp-NoVaS is always better than that of TV-Simple-NoVaS. Similarly, TV-GE-NoVaS also performs better than TV-GS-NoVaS. Hence, using exponential weights may be worth it; see also Politis (2007).
- TV-Exp-NoVaS and TV-Simple-NoVaS both underperform compared to TV Generalized NoVaS, and sometimes even compared to TV-GARCH(1,1) models. Interestingly, TV Generalized NoVaS (both simple and exponential—but especially TV-GE-NoVaS) performs best, giving smaller MADs and RMSEs of estimation errors among all methods used. Therefore, choosing an appropriate non-zero α in Eq. (8) of the NoVaS transformation is crucial for good performance when the goal is to estimate realized volatility although it was not deemed important for the prediction of squared returns discussed in Section 2.

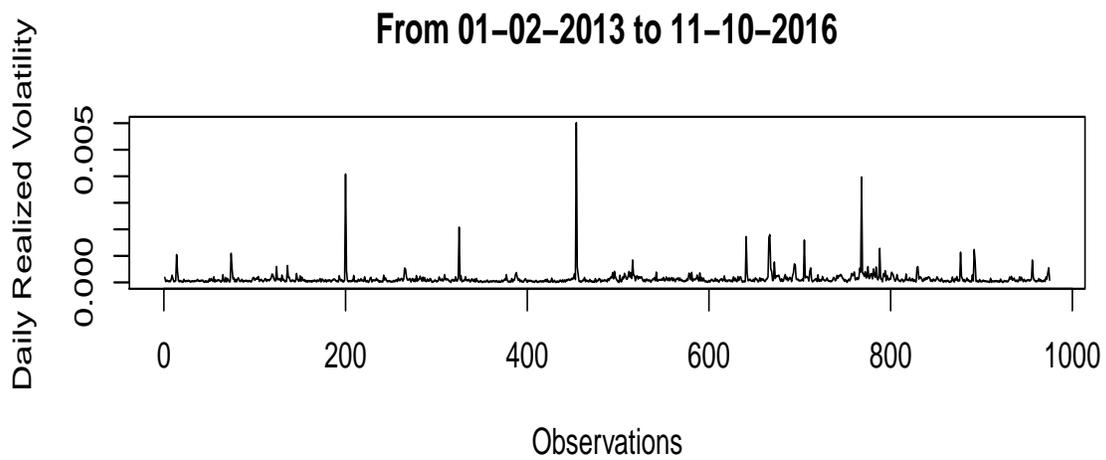
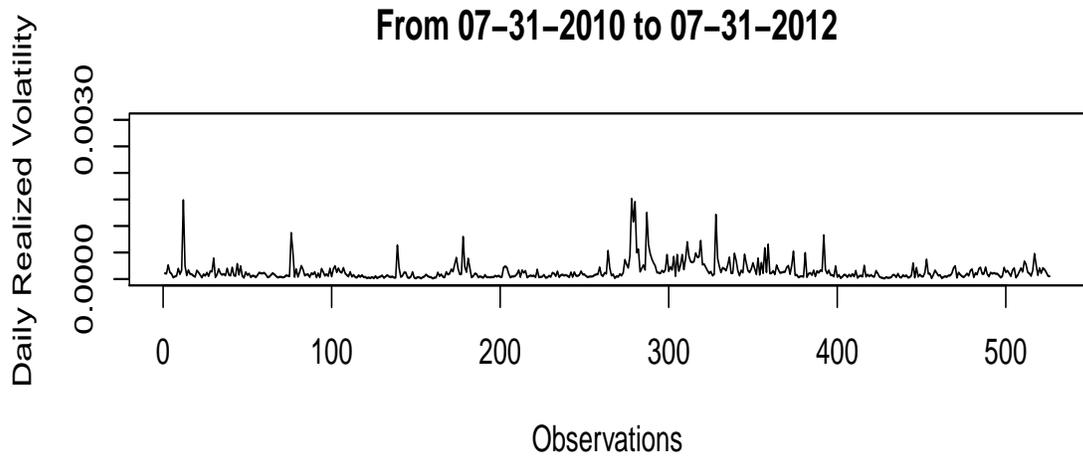


Figure 3: Daily realized volatility for IBM stock; the top one is daily realized volatility for data from 07-01-2010 to 07-31-2012 and the bottom is for the series from 01-02-2013 to 11-10-2016.

Series 2010-2012	Window size = 126		Window size = 252	
Methods	MAD	RMSE	MAD	RMSE
GARCH(1,1) with normal error	1.418E-04	2.112E-04	1.732E-04	2.292E-04
GARCH(1,1) with t -error	1.397E-04	2.049E-04	1.706E-04	2.233E-04
Simple-NoVaS	1.450E-04	1.941E-04	1.793E-04	2.231E-04
Exp-NoVaS	1.356E-04	1.698E-04	1.589E-04	1.922E-04
GS NoVaS	1.199E-04	1.689E-04	1.439E-04	1.901E-04
GE NoVaS	1.115E-04	1.684E-04	1.343E-04	1.908E-04

Table 1: The mean absolute deviation (MAD) and root mean-squared error (RMSE) of estimation errors for the daily realized volatility of IBM stock from July 1, 2010 to July 31, 2012.

Series 2013-2016	Window size = 126		Window size = 252	
Methods	MAD	RMSE	MAD	RMSE
GARCH(1,1) with normal error	1.362E-04	3.692E-04	1.446E-04	3.617E-04
GARCH(1,1) with t -error	1.343E-04	3.682E-04	1.390E-04	3.592E-04
Simple-NoVaS	1.655E-04	3.398E-04	1.664E-04	3.319E-04
Exp-NoVaS	1.546E-04	3.291E-04	1.531E-04	3.147E-04
GS NoVaS	1.175E-04	3.252E-04	1.191E-04	3.179E-04
GE NoVaS	1.114E-04	3.253E-04	1.138E-04	3.175E-04

Table 2: The mean absolute deviation (MAD) and root mean-squared error (RMSE) of estimation errors for the daily realized volatility of IBM stock from January 2, 2013 to November 10, 2016.

4 Bootstrap Prediction Intervals

Coming back to the problem of prediction of the one-step ahead squared return Y_{n+1}^2 , it is generally desirable to go beyond a simple point predictor, and instead try to construct a prediction *interval* that will contain Y_{n+1}^2 with (conditional) probability $1 - \alpha$ approximately, i.e., asymptotically over large samples. It is well known that in order to capture the variability of estimation/model-fitting, a prediction interval typically involves some form of resampling; see Politis (2015) for a discussion. The construction of bootstrap prediction intervals based on ARCH/GARCH models has been addressed in the literature; see Pan and Politis (2014, 2016) and the references therein. It is of interest to see if prediction intervals can be constructed in a Model-Free fashion, e.g. based on NoVaS transformation, and compare their performance with the model-based ones.

To fix ideas, we will return to the setup of a stationary series $\{Y_t\}$; extensions to local stationarity using TV-GARCH and TV-NoVaS are immediate in view of the last section. Recall that the key idea of Model-Free Prediction is to transform a given complex dataset into one that is *i.i.d.*, and therefore easier to handle. In the NoVaS transformation, we transform our dataset to an *i.i.d.* data vector that is standard normal. In Model-Based Bootstrap, confidence and prediction intervals are based on generating one-step ahead pseudo-data using the fitted model equation, and some approximation for the distribution of the errors, e.g., the empirical distribution of the residuals. By contrast, the Model-Free Bootstrap re-samples the transformed dataset that is *i.i.d.*, and then transforms the resampled vector back to obtain pseudo-data in the original domain.

4.1 Description of Interval Prediction Algorithms

Based on the NoVaS transformation, the optimal (in an L_1 sense) predictor of $g(Y_{n+1})$ given \mathcal{F}_n was given in eq. (16), i.e.,

$$g(\widehat{Y}_{n+1}) = \text{Median} \left(g \left(A_n \frac{W_{n+1,a}}{\sqrt{1 - a_0 W_{n+1,a}^2}} \right) \mid \mathcal{F}_n \right)$$

$$= \text{Median} \left(g \left(A_n \frac{W_{n+1,a}}{\sqrt{1 - a_0 W_{n+1,a}^2}} \right) \right)$$

where the second equality is true when the quasi-Gaussian series $W_{t,a}$ is uncorrelated—and therefore independent—as has been the empirical finding with various types of daily financial returns. In addition, a preliminary approximation to the predictive distribution of $g(Y_{n+1})$ can be given by the empirical distribution of the random variables $\{g \left(A_n \frac{W_{t,a}}{\sqrt{1 - a_0 W_{t,a}^2}} \right)$ for $t = p + 1, \dots, n\}$. However, as Politis (2015) remarked, such an approximation treats the NoVaS parameters as known, and ignores the variability associated with estimating them; to incorporate this variability, some kind of resampling—in this case the Model-free bootstrap—is inevitable.

Note that the point predictor $g(\widehat{Y}_{n+1})$ is a function only⁴ of Y_n, \dots, Y_{n-p+1} , i.e., is a predictor of the type of a nonlinear autoregressive (AR) model or a Markov process of order p . Hence, to develop the relevant resampling algorithms, we can borrow some ideas from the work of Pan and Politis (2014, 2016); in particular, we will adopt the ‘*forward*’ bootstrap methodology, i.e., generate bootstrap series forward in time but also ensure that Y_{n+1}^* is constructed correctly; see also Politis (2015).

The basic Model-free (MF) bootstrap algorithm for prediction intervals in the setting of financial returns goes as follows. For concreteness, our construction of prediction intervals below is based on the L_1 -optimal predictor, i.e., the conditional median, because of its robustness; construction of prediction intervals based on other types of predictors can be done in an analogous manner if so desired.

⁴In the case of Generalized NoVaS (Simple or Exponential), $g(\widehat{Y}_{n+1})$ is also a function of s_n^2 which, however, converges to EY_t^2 for large n ; hence, it can be treated as a constant for the purposes of this discussion.

ALGORITHM 4.1. MF BOOTSTRAP PREDICTION INTERVALS FOR $g(Y_{n+1})$

1. Use one of the NoVaS algorithms (Simple vs. Exponential, Generalized or not, etc.) to create the transformed data $\{W_{t,a}$ for $t = p + 1, \dots, n\}$ that are assumed to be approximately i.i.d. Let p , α and a_i denote the fitted NoVaS parameters.
2. Calculate $g(\widehat{Y}_{n+1})$, the point predictor of $g(Y_{n+1})$, as the median of the set $\{g\left(A_n \frac{W_{t,a}}{\sqrt{1-a_0 W_{t,a}^2}}\right)$ for $t = p + 1, \dots, n\}$; recall $A_n = \sqrt{\alpha s_n^2 + \sum_{i=1}^p a_i Y_{n+1-i}^2}$.
3. (a) Sample randomly (with replacement) the transformed variables $\{W_{t,a}$ for $t = p + 1, \dots, n\}$ to create the pseudo-data $W_{p+1}^*, \dots, W_{n-1}^*, W_n^*$ and W_{n+1}^* .
 (b) Let $(Y_1^*, \dots, Y_p^*)' = (Y_{1+I}, \dots, Y_{p+I})'$ where I is generated as a discrete random variable uniform on the values $0, 1, \dots, n - p$.
 (c) Generate the bootstrap pseudo-data Y_t^* for $t = p + 1, \dots, n$ using the following equations, namely:

$$Y_t^* = \frac{W_t^*}{\sqrt{1 - a_0 W_t^{*2}}} \sqrt{\alpha s_{t-1}^{*2} + \sum_{i=1}^p a_i Y_{t-i}^{*2}} \quad \text{for } t = p + 1, \dots, n \quad (20)$$

$$\text{where } s_{t-1}^{*2} = (t - 1)^{-1} \sum_{k=1}^{t-1} Y_k^{*2}.$$

- (d) Re-estimate the NoVaS transformation based on the bootstrap data Y_1^*, \dots, Y_n^* , yielding bootstrap parameters p^* , α^* , a_0^* , a_1^*, \dots, a_p^* .

Let $A_n^* = \sqrt{\alpha^* s_n^{*2} + \sum_{i=1}^{p^*} a_i^* Y_{n+1-i}^{*2}}$, and calculate the bootstrap predictor $g(\widehat{Y}_{n+1}^*)$ as the median of the set

$$\left\{ g \left(A_n^* \frac{W_{t,a}}{\sqrt{1 - a_0^* W_{t,a}^2}} \right) \quad \text{for } t = p + 1, \dots, n \right\} \quad (21)$$

using the convention⁵ that when $1 - a_0^* W_{t,a}^2 \leq 0$, we assign $\frac{1}{\sqrt{1 - a_0^* W_{t,a}^2}} = \infty$.

- (e) Generate the bootstrap future value Y_{n+1}^* as

$$Y_{n+1}^* = \frac{W_{n+1}^*}{\sqrt{1 - a_0 W_{n+1}^{*2}}} \sqrt{\alpha s_n^{*2} + \sum_{i=1}^p a_i Y_{n-i+1}^{*2}}. \quad (22)$$

- (f) Calculate the bootstrap ‘root’: $g(Y_{n+1}^*) - g(\widehat{Y}_{n+1}^*)$.

4. Repeat step 3 above B times; the B bootstrap root replicates are collected in the form of an empirical distribution whose α -quantile is denoted $q(\alpha)$.
5. The $(1 - \alpha)100\%$ equal-tailed bootstrap prediction interval for $g(Y_{n+1})$ is given by

$$[g(\widehat{Y}_{n+1}) + q(\alpha/2), g(\widehat{Y}_{n+1}) + q(1 - \alpha/2)].$$

Note that the last p values from the *original* data, i.e., Y_{n-p+1}, \dots, Y_n , are used in both the creation of the bootstrap predictor in Eq.(21) and bootstrap future value in Eq.(22); this is in accordance with the ‘forward’ bootstrap methodology in the work of Pan and Politis (2014) but also with the general Model-free Bootstrap described in Algorithm 2.4.1 in Politis (2015).

Another version of Algorithm 4.1 can also be devised in the spirit of the Limit Model-Free(LMF) Bootstrap of Politis (2015); it would amount to replacing Step 3 (a) above by:

(a') Generate $W_{p+1}^*, \dots, W_{n-1}^*, W_n^*$ and W_{n+1}^* as i.i.d. from a $N(0, 1)$ distribution truncated to $\pm 1/\sqrt{a_0}$.

4.2 Local Stationarity

The construction of prediction intervals given in Algorithm 4.1 was based on the assumption of stationarity. However, if the returns $\{Y_t\}$ are only assumed to be locally stationary, Algorithm 5 should be applied to each of the windowed (local) datasets, thus yielding Time-varying NoVaS bootstrap prediction intervals. Similarly, the forward bootstrap algorithm for prediction intervals in stationary ARCH/GARCH models was detailed in Pan and Politis (2014); if the returns are only locally stationary, the model-based bootstrap algorithm should be applied to each of the windowed (local) datasets, in effect using the TV ARCH/GARCH model instead.

⁵This is because the original NoVaS data satisfies $|W_{t,a}| \leq 1/\sqrt{a_0}$ but a_0^* might turn out bigger (or smaller) than a_0 . Alternatively, one can base Eq. (21) on the NoVaS transformed series $W_{t,a}^*$ that corresponds to the bootstrap data Y_1^*, \dots, Y_n^* , or on a Monte Carlo experiment using a $N(0, 1)$ distribution truncated to $\pm 1/\sqrt{a_0^*}$; this is the case of step (a') in the LMF Algorithm that follows. All these options are practically indistinguishable as far as taking the median is concerned, and Eq. (21) is the most straightforward.

4.3 Finite-Sample Performance of Model-Free and Model-Based Prediction Intervals

In what follows, we will use real and simulated data to compare the performance of the proposed bootstrap prediction intervals, model-free (i.e., NoVaS) and model-based.

4.3.1 Three Illustrative Datasets

In the work of prediction intervals, we also focus on another three real world datasets of daily returns taken from a foreign exchange rate, a stock price, and a stock index. A brief description of these datasets is as follows; for more details, see Politis (2015).

- **Example 1: Foreign exchange rate.** Daily returns from the Yen vs. Dollar exchange rate from January 1, 1988 to August 1, 2002; the data were downloaded from Datastream. The sample size is $n = 3600$ (weekends and holidays are excluded).
- **Example 2: Stock index.** Daily returns of the S&P500 stock index from October 1, 1983 to August 30, 1991; the data are available as part of the `GARCH` module in `Splus`. The sample size is $n = 2000$.
- **Example 3: Stock price.** Daily returns of the IBM stock price from February 1, 1984 to December 31, 1991; the data are again available as part of the `GARCH` module in `Splus`. The sample size is $n = 2000$.

4.3.2 Simulation

In the following simulation work, we will compare the performance in interval prediction of squared returns by using Simple-NoVaS, Exp-NoVaS, Limit Model-Free with Simple-NoVaS method (LMF Simple-NoVaS), Limit Model-Free Exp-NoVaS method (LMF Exp-NoVaS), Generalized Simple-NoVaS method (GS-NoVaS), Generalized Exp-NoVaS method (GE-NoVaS), Limit Model-Free with Generalized Simple-NoVaS method (LMF GS-NoVaS) and Limit Model-Free with Generalized Exp-NoVaS method (LMF GE-NoVaS) method with that by using `GARCH(1,1)` models with normal errors. Each interval is constructed based on the windowed data series.

So all the methods and models here are still time-varying models. For the window size, we will use 125 and 250, as are worked in the point predictions. For prediction intervals using a GARCH(1,1) model, we use the Algorithm 9.2.1 of Model-Based prediction intervals for Y_{n+1} from Politis (2015); see also Pan and Politis (2016).

As in Section 2 , we employ the same TV-GARCH(1,1) and CP-GARCH(1,1) models to generate the simulated datasets. For consistency, two stationary processes are generated by the following two standard GARCH(1,1) models:

- **Model 1.** $Y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = .00001 + .93\sigma_{t-1}^2 + .05Y_{t-1}^2$, $\{\epsilon_t\} \sim i.i.d. N(0, 1)$.
- **Model 2.** $Y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = .00001 + .73\sigma_{t-1}^2 + .10Y_{t-1}^2$, $\{\epsilon_t\} \sim i.i.d. N(0, 1)$.

Each simulated dataset has size $n = 1000$, and the window sizes are $b = 125$ and $b = 250$, similar to those used in Section 2 of point predictions. For the three real world data discussed in Section 4.3.1, we use window sizes $b = 250$ and 500 because their full sample size is bigger. For computational reasons, we chose $B = 500$ for bootstrap resampling throughout; a practitioner that has to deal with a single dataset could well afford to use a larger B , maybe even 10 times larger.

For each dataset, $n - b$ windowed datasets of size b are extracted. For each windowed dataset, one of the bootstrap methods—based on GARCH(1,1) or on different NoVaS algorithms—was used to create B bootstrap sample paths and B one-step ahead “future” values denoted by $Y_{(b+1,j)}$ for $j = 1, 2, \dots, B$; The bootstrap prediction interval (L_i, U_i) was constructed for the “future” value $Y_{(b+1)}$ of windowed dataset i , here $i = b + 1, b + 2, \dots, n$. The corresponding empirical average coverage level (CVR) and the average length (LEN) of the constructed intervals, and the standard error (St.err) associated with each length of the constructed intervals are calculated as

$$CVR = \frac{1}{n - b} \sum_{i=1}^{n-b} CVR_i$$

$$LEN = \frac{1}{n - b} \sum_{i=1}^{n-b} LEN_i \text{ and } St.err = \sqrt{\frac{1}{n - b} \sum_{i=1}^{n-b} (LEN_i - LEN)^2}$$

where

$$CVR_i = \frac{1}{B} \sum_{j=1}^B \mathbf{1}_{[L_j, U_j]} Y_{(b+1,i)} \text{ and } LEN_i = U_i - L_i.$$

4.3.3 Discussion of results

Our results are summarized in Tables 3 and 4 for simulated stationary datasets, in Tables 5 and 6 for simulated datasets generated by CP-GARCH(1,1) and TV-GARCH(1,1) models respectively and in Tables 7, 8 and 9 for the three real world datasets. Each table has two subtables with different window sizes. The first two lines of each subtable are the results using the Simple NoVaS (Simple-NoVaS) and Exponential NoVaS (Exp-NoVaS) methods. Lines 3 and 4 of each subtable are the results using Limit Model-Free with Simple-NoVaS (LMF Simple-NoVaS) and Limit Model-Free Exp-NoVaS (LMF Exp-NoVaS) methods. Lines 5 and 6 of each subtable are the results of Generalized Simple-NoVaS (GS-NoVaS) and Generalized Exp-NoVaS (GE-NoVaS) methods. Lines 7 and 8 of each subtable are for Limit Model-Free with Generalized Simple-NoVaS method (LMF GS-NoVaS) and Limit Model-Free with Generalized Exp-NoVaS (LMF GE-NoVaS) method. The last line of each subtable gives the results by using GARCH(1,1) models with normal errors.

For simulated stationary time series data by GARCH(1,1) models, Tables 3 and 4 show that the performance of different NoVaS methods is better than that of GARCH(1,1), that is, NoVaS methods have average coverages very close to or equal to the nominal ones. Remarkably, the Generalized Simple and/or Exponential NoVaS methods beats Simple/Exp-NoVaS as well as LMF Simple/Exp-NoVaS. In addition, there are not significant differences between Limit Model-Free NoVaS and Generalized NoVaS methods when the data is stationary. It is also shown that the window size has little effect—the coverages are still close to nominal ones—for the performance of NoVaS methodology, while the reaction of GARCH(1,1) to the change of the window size is relatively large. We can see that the average coverages get much closer to the nominal ones as the window size is increased from 125 to 250.

Tables 6 and 7 give the results for data generated from time-varying GARCH(1,1) and CP-GARCH(1,1) models with normal errors respectively. It is shown that in each table, all NoVaS methods have the average coverages much closer to the nominal ones than GARCH(1,1) in both subtables for each table. When the window size is increased to 250, GARCH(1,1) models perform better as expected from earlier work of Politis and Thomakos (2008, 2013). Meanwhile, the performance of NoVaS methods is not sensitive to changes in the sample size; this finding is in accordance with the results for point prediction of squared returns of Section 2. However, this

effect of window size on GARCH(1,1) under data with nonstationarity is bigger than when the data are stationary; compare to Tables 3 and 4.

It is also worthwhile to note that Limit Model-Free NoVaS and Generalized NoVaS methods can sometimes capture the nominal coverages exactly with smaller average lengths and standard deviations for the predicted intervals. Also, we can find that Limit Model-Free NoVaS methods perform better than Simple-NoVaS and Exponential NoVaS, while the performance of the Limit Model-Free Generalized NoVaS methods are not as good as the Generalized NoVaS under both simulated datasets.

If we look at the results for the three real world financial series in Tables 7, 8 and 9, the NoVaS still outperforms the benchmark GARCH(1,1) models. Looking at Tables 5 and 6, GARCH(1,1) performs worse when the window size is increasing from 250 to 500, that is, the average coverages are smaller and the average length and standard errors of the predicted intervals are larger. The possible reason might be that the assumption of stationarity might be broken when the window size is too large, like 500. It is also apparent that the performance of Limit Model-Free Generalized NoVaS methods are indistinguishable from that of Generalized NoVaS although the latter appears to be marginally better upon closer comparison. All in all, when the data is non-stationary, the results from Tables 6–9 are supporting the superior performance of time-varying NoVaS in interval prediction against the benchmark time-varying GARCH(1,1). These results are not only consistent with the results of point predictions in section 2 and section 3, but also consistent with the findings of the performance of NoVaS without using time-varying processes; see more in Politis (2015), Chapter 10.

5 Concluding Remarks

The NoVaS methodology for prediction of stationary financial returns was revisited, and its applicability for estimating realized volatility was demonstrated. Working with high-frequency real data, the performance of NoVaS was compared to the benchmark GARCH(1,1) model for volatility estimation. The NoVaS—in particular the so-called Generalized NoVaS—empirically outperformed GARCH (1,1) model with either normal or t errors; this is a new finding.

Moving on to the setting of non-stationary data, e.g., locally stationary time series, a Time-varying (TV) version of NoVaS was constructed and compared to fitting TV GARCH models in the context of one-step ahead point predictions. All types of NoVaS methodologies were included in the comparison, including the Generalized NoVaS which again showed definite advantages. Using DGPs that included change-point and TV GARCH(1,1) models, a comprehensive simulation confirmed that the NoVaS methodology remains successful with nonstationary data, and generally outperforms the benchmark GARCH(1,1) model.

In a remarkable finding, it appears that the TV NoVaS approach for prediction outperforms the fitting of a TV GARCH(1,1) model even when the DGP is indeed a TV GARCH(1,1) model. The reason for this can be attributed to the large sample size needed in order to reliably estimate the GARCH parameters via numerical MLE. With locally stationary data, it is the window size that dictates the effective sample size for model fitting; this window size can not be too large if the practitioner wants to capture the underlying nonstationarity.

Last but not least, a Model-Free algorithm for bootstrap prediction intervals was proposed based on the NoVaS transformation. The applicability of the methodology extends to locally stationary time series using the TV NoVaS notion, and can be compared to model-based prediction intervals based on the GARCH(1,1) model. In extensive empirical work on interval prediction of squared returns using both real world and simulated non-stationary series, it was found that NoVaS gives a higher coverage level (on the average) than the GARCH(1,1) method. In contrast to the latter, the performance of NoVaS does not depend much on the window size which is reassuring.

In conclusion, the TV NoVaS methodology is robust against possible nonstationarities in the data, and generally outperforms the GARCH benchmark for (i) point prediction of squared returns, (ii) interval prediction of squared returns, and (iii) volatility estimation. With regards to target (i), earlier work had shown little advantage of using a nonzero α in the NoVaS transformation. However, in terms of targets (ii) and (iii), it appears that using the Generalized version of NoVaS—either Simple or Exponential—can be beneficial and well-worth the associated computational cost.

Acknowledgement. Many thanks are due to the Editor and an anonymous reviewer for their thoughtful comments. Research supported by NSF grant DMS 19-14556.

Window size 125	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.948	6.17E-04	5.84E-04	0.901	3.89E-04	2.25E-04
Exp-NoVaS	0.963	3.96E-03	1.25E-03	0.912	3.52E-04	2.01E-04
Limit Simple-NoVaS	0.951	4.00E-04	2.34E-04	0.899	2.73E-04	1.54E-04
Limit Exp-NoVaS	0.954	4.89E-04	2.05E-04	0.893	3.12E-04	1.22E-04
GS NoVaS	0.947	2.82E-04	9.98E-05	0.896	2.27E-04	1.08E-04
GE NoVaS	0.950	3.57E-04	2.59E-04	0.899	2.55E-04	1.54E-04
LMF GS-NoVaS	0.946	2.52E-04	1.14E-04	0.893	2.01E-04	8.70E-05
LMF GE-NoVaS	0.944	3.06E-04	8.44E-05	0.893	2.43E-04	6.47E-05
GARCH(1,1)	0.926	2.74E-04	1.25E-04	0.879	2.08E-04	9.40E-05
Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.949	5.17E-04	3.20E-04	0.895	3.79E-04	2.16E-04
Exp-NoVaS	0.953	4.73E-04	2.37E-04	0.896	2.92E-04	1.12E-04
Limit Simple-NoVaS	0.931	4.10E-04	2.31E-04	0.886	2.89E-04	1.60E-04
Limit Exp-NoVaS	0.943	4.59E-04	1.70E-04	0.894	3.05E-04	1.10E-04
GS NoVaS	0.945	2.86E-04	1.10E-04	0.896	2.10E-04	1.21E-04
GE NoVaS	0.948	5.92E-04	5.63E-04	0.900	2.81E-04	1.57E-04
LMF GS-NoVaS	0.943	2.73E-04	1.48E-04	0.892	2.09E-04	1.10E-04
LMF GE-NoVaS	0.948	2.99E-04	8.76E-05	0.893	3.11E-04	1.75E-04
GARCH(1,1)	0.931	2.52E-04	9.64E-05	0.883	2.53E-04	7.28E-05

Table 3: A stationary process generated by GARCH(1,1) with $C = 10^{-5}$, $B = 0.73$, $A = 0.10$

Window size 125	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.950	4.05E-03	4.71E-03	0.901	2.42E-03	1.38E-03
Exp-NoVaS	0.960	7.41E-03	3.86E-02	0.907	2.60E-03	1.60E-03
Limit Simple-NoVaS	0.951	3.06E-03	1.88E-03	0.893	2.04E-03	1.18E-03
Limit Exp-NoVaS	0.949	3.88E-03	2.02E-03	0.886	2.49E-03	1.30E-03
GS-NoVaS	0.952	1.87E-03	6.74E-04	0.899	2.28E-03	1.78E-03
GE-NoVaS	0.947	2.67E-04	1.02E-04	0.898	2.78E-04	1.68E-04
LIMIT GS-NOVAS	0.947	3.13E-03	1.76E-03	0.893	2.48E-03	1.27E-03
LIMIT GE-NOVAS	0.949	3.30E-03	1.60E-03	0.894	2.81E-03	1.57E-03
GARCH(1,1)	0.914	2.37E-03	1.03E-03	0.866	2.09E-03	8.27E-03
Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.958	4.06E-03	2.25E-03	0.906	2.81E-03	1.54E-03
Exp-NoVaS	0.960	3.14E-03	2.09E-03	0.903	2.06E-03	1.19E-03
Limit Simple-NoVaS	0.949	5.62E-03	3.81E-03	0.899	3.94E-03	2.66E-03
Limit Exp-NoVaS	0.938	3.64E-03	1.81E-03	0.888	2.42E-03	1.22E-03
GS-NoVaS	0.951	2.14E-03	9.14E-04	0.900	2.48E-03	3.71E-04
GE-NoVaS	0.949	2.71E-04	1.17E-04	0.900	2.19E-04	1.19E-04
LIMIT GS-NOVAS	0.946	2.67E-03	1.44E-03	0.894	2.09E-03	7.86E-04
LIMIT GE-NOVAS	0.949	3.33E-03	1.31E-03	0.899	2.71E-03	1.41E-03
GARCH(1,1)	0.919	2.16E-03	1.09E-03	0.873	1.87E-03	5.81E-03

Table 4: A stationary process generated by GARCH(1,1) with $C = 10^{-5}$, $B = 0.93$, $A = 0.05$

Window size 125	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.954	1.24E-02	1.87E-03	0.900	2.05E-03	1.86E-03
Exp-NoVaS	0.965	2.63E-02	5.14E-03	0.918	8.87E-02	6.41E-03
LMF Simple-NoVaS	0.951	1.07E-03	7.81E-04	0.906	7.43E-04	5.29E-04
LMF Exp-NoVaS	0.945	1.09E-03	7.49E-04	0.890	7.12E-04	4.74E-04
GS-NoVaS	0.951	8.84E-04	5.78E-04	0.901	6.27E-04	3.99E-04
GE-NoVaS	0.950	9.25E-04	8.72E-04	0.892	6.66E-04	6.26E-04
LMF GS-NoVaS	0.944	8.10E-04	6.39E-04	0.897	6.83E-04	5.92E-04
LMF GE-NoVaS	0.947	9.25E-04	6.34E-04	0.887	5.15E-04	3.99E-04
GARCH(1,1)	0.885	5.03E-04	4.14E-04	0.842	4.07E-04	5.29E-04
Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.939	2.61E-03	3.43E-03	0.884	1.22E-03	8.47E-03
Exp-NoVaS	0.968	9.69E-04	6.66E-04	0.911	6.85E-04	4.42E-04
LMF Simple-NoVaS	0.944	9.27E-04	7.90E-04	0.893	6.54E-04	4.96E-04
LMF Exp-NoVaS	0.924	1.12E-03	7.41E-04	0.881	7.51E-04	4.97E-04
GS-NoVaS	0.947	9.82E-04	6.39E-04	0.891	5.52E-04	2.97E-04
GE-NoVaS	0.948	9.25E-04	6.70E-04	0.888	5.25E-04	2.91E-04
LMF GS-NoVaS	0.940	8.71E-04	6.71E-04	0.891	1.02E-03	1.00E-03
LMF GE-NoVaS	0.940	8.11E-04	5.68E-04	0.895	6.36E-04	5.84E-04
GARCH(1,1)	0.909	4.19E-04	4.03E-04	0.854	3.28E-04	5.22E-04

Table 5: Data generated by TV-GARCH models

Window size 125	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.960	3.96E-03	4.21E-02	0.908	2.21E-03	2.36E-02
Exp-NoVaS	0.963	1.07E-02	1.92E-02	0.918	1.61E-03	1.80E-03
LMF Simple-NoVaS	0.958	1.73E-03	1.79E-03	0.897	1.18E-03	1.21E-03
LMF Exp-NoVaS	0.946	2.40E-03	2.58E-03	0.894	1.58E-03	1.70E-03
GS-NoVaS	0.950	1.10E-03	8.86E-04	0.901	8.10E-04	6.33E-04
GE-NoVaS	0.950	2.39E-03	2.35E-03	0.897	1.79E-03	1.79E-03
LMF GS-NoVaS	0.938	1.78E-03	1.72E-03	0.896	1.37E-03	1.36E-03
LMF GE-NoVaS	0.949	1.70E-03	2.03E-03	0.889	9.42E-04	9.35E-04
GARCH(1,1)	0.901	3.42E-03	1.52E-03	0.833	2.60E-03	5.74E-03
Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.949	3.04E-03	5.25E-03	0.907	2.10E-03	3.19E-03
Exp-NoVaS	0.957	2.84E-03	3.01E-03	0.900	1.82E-03	1.78E-03
LMF Simple-NoVaS	0.952	1.80E-03	1.42E-03	0.892	1.25E-03	9.79E-04
LMF Exp-NoVaS	0.941	2.46E-03	1.85E-03	0.893	1.66E-03	1.23E-03
GS-NoVaS	0.949	3.86E-03	3.09E-03	0.906	4.84E-03	4.15E-03
GE-NoVaS	0.944	1.89E-03	1.24E-03	0.890	1.75E-03	1.36E-03
LMF GS-NoVaS	0.956	4.08E-03	4.86E-03	0.911	2.84E-03	3.03E-03
LMF GE-NoVaS	0.943	1.96E-03	1.73E-03	0.880	1.28E-03	1.00E-03
GARCH(1,1)	0.916	1.86E-03	1.47E-03	0.871	1.34E-03	7.87E-03

Table 6: Data generated by CP-GARCH models

Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.97	6.64E-04	6.25E-04	0.922	3.86E-04	3.32E-04
Exp-NoVaS	0.968	6.86E-04	5.63E-04	0.908	3.35E-04	2.33E-04
LMF Simple-NoVaS	0.958	4.71E-04	3.97E-04	0.898	3.02E-04	2.39E-04
LMF Exp-NoVaS	0.964	5.23E-04	3.47E-04	0.924	3.15E-04	2.14E-04
GS-NoVaS	0.950	4.45E-04	3.72E-04	0.906	3.09E-04	2.02E-04
GE-NoVaS	0.950	4.51E-04	2.78E-04	0.896	2.71E-04	1.72E-04
LMF GS-NoVaS	0.946	4.14E-04	2.85E-04	0.9	2.81E-04	1.54E-04
LMF GE-NoVaS	0.950	4.20E-04	1.82E-04	0.886	2.67E-04	9.95E-05
GARCH(1,1)	0.928	3.60E-04	1.73E-04	0.87	2.25E-04	1.37E-04
Window size 500	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.945	6.17E-04	8.48E-04	0.895	3.79E-04	5.07E-04
Exp-NoVaS	0.953	5.00E-04	5.27E-04	0.888	3.09E-04	3.16E-04
LMF Simple-NoVaS	0.918	4.86E-04	3.03E-04	0.862	6.09E-04	3.69E-04
LMF Exp-NoVaS	0.954	5.13E-04	5.18E-04	0.914	3.14E-04	3.10E-04
GS-NoVaS	0.949	3.05E-04	1.55E-04	0.906	2.08E-04	8.80E-05
GE-NoVaS	0.949	3.56E-04	1.90E-04	0.902	2.05E-04	9.05E-05
LMF GS-NoVaS	0.946	4.10E-04	1.98E-04	0.890	2.66E-04	1.21E-04
LMF GE-NoVaS	0.950	3.89E-04	1.20E-04	0.912	2.56E-04	6.66E-05
GARCH(1,1)	0.928	1.60E-04	1.73E-04	0.870	2.25E-04	1.37E-04

Table 7: Foreign exchange rate

Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.946	6.04E-04	8.10E-04	0.896	3.78E-04	4.11E-04
Exp-NoVaS	0.940	5.51E-04	4.37E-04	0.880	3.42E-04	2.35E-04
LMF Simple-NoVaS	0.934	5.02E-04	5.69E-04	0.880	3.24E-04	3.20E-04
LMF Exp-NoVaS	0.956	5.32E-04	4.19E-04	0.894	3.31E-04	2.56E-04
GS-NoVaS	0.948	4.12E-04	3.66E-04	0.896	2.99E-04	2.54E-04
GE-NoVaS	0.950	4.20E-04	2.76E-04	0.898	2.95E-04	1.70E-04
LMF GS-NoVaS	0.946	4.38E-04	3.47E-04	0.892	2.86E-04	2.14E-04
LMF GE-NoVaS	0.948	4.25E-04	2.93E-04	0.900	2.87E-04	1.99E-04
GARCH(1,1)	0.938	3.25E-04	2.27E-04	0.862	1.92E-04	1.67E-04
Window size 500	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.961	1.59E-03	5.35E-03	0.918	1.06E-03	3.84E-03
Exp-NoVaS	0.948	1.16E-03	8.49E-04	0.878	7.77E-04	5.68E-04
LMF Simple-NoVaS	0.929	1.49E-03	5.13E-03	0.872	9.68E-04	3.44E-03
LMF Exp-NoVaS	0.952	1.61E-03	4.67E-03	0.897	9.92E-04	2.81E-03
GS-NoVaS	0.949	5.26E-04	2.71E-04	0.894	3.43E-04	1.91E-04
GE-NoVaS	0.949	3.56E-04	1.90E-04	0.892	3.36E-04	1.51E-04
LMF GS-NoVaS	0.944	6.78E-04	4.87E-04	0.894	4.45E-04	3.18E-04
LMF GE-NoVaS	0.936	5.86E-04	3.01E-04	0.884	4.58E-04	2.75E-04
GARCH(1,1)	0.927	4.58E-04	3.04E-03	0.863	3.60E-04	1.82E-03

Table 8: S&P500 Stock index

Window size 250	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.964	1.42E-03	7.51E-04	0.918	7.67E-04	4.71E-04
Exp-NoVaS	0.954	1.31E-03	1.28E-03	0.894	7.75E-04	4.21E-04
LMF Simple-NoVaS	0.958	1.02E-03	5.73E-04	0.918	7.83E-04	4.42E-04
LMF Exp-NoVaS	0.956	1.08E-03	6.64E-04	0.914	7.48E-04	3.81E-04
GS-NoVaS	0.954	9.16E-04	4.45E-04	0.894	7.09E-04	4.37E-04
GE-NoVaS	0.946	8.50E-04	3.48E-04	0.896	6.99E-04	4.28E-04
LMF GS-NoVaS	0.950	9.17E-04	4.81E-04	0.890	6.88E-04	3.89E-04
LMF GE-NoVaS	0.952	1.01E-03	5.85E-04	0.898	6.67E-04	3.43E-04
GARCH(1,1)	0.940	8.99E-04	4.01E-04	0.878	5.60E-04	1.75E-04
Window size 500	nominal coverage 95%			nominal coverage 90%		
Methods	CVR	LEN	St.err	CVR	LEN	St.err
Simple-NoVaS	0.957	1.26E-03	7.64E-03	0.910	8.47E-04	7.74E-04
Exp-NoVaS	0.949	1.16E-03	8.24E-03	0.879	7.74E-04	5.51E-04
LMF Simple-NoVaS	0.955	1.25E-03	1.16E-03	0.906	8.39E-04	7.67E-04
LMF Exp-NoVaS	0.957	1.33E-03	9.84E-04	0.879	8.67E-04	5.51E-04
GS-NoVaS	0.948	1.14E-03	1.91E-04	0.902	1.13E-04	2.61E-04
GE-NoVaS	0.949	2.13E-03	3.75E-04	0.898	1.71E-04	3.54E-04
LMF GS-NoVaS	0.948	1.68E-03	4.55E-04	0.904	1.33E-04	2.97E-04
LMF GE-NoVaS	0.950	1.93E-03	3.07E-03	0.896	1.30E-04	2.34E-03
GARCH(1,1)	0.934	5.22E-04	8.06E-04	0.878	1.56E-04	5.56E-04

Table 9: Stock price series (IBM)

References

- Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). *Volatility and Correlation Forecasting*, pages 777–878. Amsterdam: North-Holland.
- Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720.
- Andersen, T. G., Bollerslev, T., and Meddahi, N. (2004). Analytical evaluation of volatility forecasts. *International Economic Review*, 45(4):1079–1110.
- Andersen, T. G., Bollerslev, T., and Meddahi, N. (2005). Correcting the errors: volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73(1):279–296.
- Bandi, F. M., Russell, J. R., and Yang, C. (2008). Realized volatility forecasting and option pricing. *Journal of Econometrics*, 147(1):34–46.
- Berkes, I., Horváth, L., and Kokoszka, P. (2004). Testing for parameter constancy in GARCH(p, q) models. *Statistics & Probability letters*, 70(4):263–273.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Chen, C. W., Gerlach, R., and Lin, E. M. (2008). Volatility forecasting using threshold heteroskedastic models of the intra-day range. *Computational Statistics & Data Analysis*, 52(6):2990–3010.
- Choi, K., Yu, W.-C., and Zivot, E. (2010). Long memory versus structural breaks in modeling and forecasting realized volatility. *Journal of International Money and Finance*, 29(5):857–875.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37.
- Dahlhaus, R. (2012). Locally stationary processes. *Handbook of Statistics: Time Series Analysis: Methods and Applications*, 30:351.

- Dahlhaus, R. and Rao, S. S. (2006). Statistical inference for time-varying arch processes. *The Annals of Statistics*, 34(3):1075–1114.
- Dahlhaus, R. and Rao, S. S. (2007). A recursive online algorithm for the estimation of time-varying arch parameters. *Bernoulli*, 13(2):389–422.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.
- Francq, C., Roy, R., and Zakoïan, J.-M. (2005). Diagnostic checking in arma models with uncorrelated errors. *Journal of the American Statistical Association*, 100(470):532–544.
- Francq, C. and Zakoïan, J.-M. (2011). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.
- Fryzlewicz, P., Sapatinas, T., and Rao, S. S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93(3):687–704.
- Fryzlewicz, P., Sapatinas, T., and Rao, S. S. (2008). Normalized least-squares estimation in time-varying arch models. *The Annals of Statistics*, 36(2):742–786.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1):59–95.
- Ghysels, E. and Sohn, B. (2009). Which power variation predicts volatility well? *Journal of Empirical Finance*, 16(4):686–700.
- Hansen, P. R. and Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3(4):525–554.
- Hansen, P. R. and Lunde, A. (2006a). Consistent ranking of volatility models. *Journal of Econometrics*, 131(1):97–121.
- Hansen, P. R. and Lunde, A. (2006b). Realized variance and market micro-structure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

- Hansen, P. R., Lunde, A., and Nason, J. M. (2003). Choosing the best volatility models: the model confidence set approach. *Oxford Bulletin of Economics and Statistics*, 65(s1):839–861.
- Hillebrand, E. (2005). Neglecting parameter changes in GARCH models. *Journal of Econometrics*, 129(1):121–138.
- Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli*, 6(3):513–539.
- Lux, T. and Morales-Arias, L. (2010). Forecasting volatility under fractality, regime-switching, long memory and student-t innovations. *Computational Statistics & Data Analysis*, 54(11):2676–2692.
- Mikosch, T. and Starica, C. (2004). Changes of structure in financial time series and the GARCH model. *Revstat Statistical Journal*, 2(1):41–73.
- Pan, L. and Politis, D. N. (2014). Bootstrap prediction intervals for markov processes. *Computational Statistics & Data Analysis*, 100:467–494.
- Pan, L. and Politis, D. N. (2016). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference*, 177:1–27.
- Patton, A. and Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics*, 97(3):683–697.
- Patton, A. J. and Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, 25(2):218–238.
- Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, 90(4):967–975.
- Politis, D. N. (2003). A normalizing and variance-stabilizing transformation for financial time series. In Akritas, M. G. and Politis, D. N., editors, *Recent Advances and Trends in Nonparametric Statistics*, pages 335 – 347. Elsevier, Amsterdam.

- Politis, D. N. (2007). Model-free versus model-based volatility prediction. *Journal of Financial Econometrics*, 5(3):358–389.
- Politis, D. N. (2015). *Model-Free Prediction and Regression: A Transformation-Based Approach to Inference*. Springer, New York.
- Politis, D. N. and Thomakos, D. D. (2008). Financial time series and volatility prediction using novas transformations. *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, (David Rapach and Mark Wohar, Eds.), pages 417–447, Emerald Group Publishing, United Kingdom.
- Politis, D. N. and Thomakos, D. D. (2013). NoVaS transformations: flexible inference for volatility forecasting. In *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr.*, (X. Chen and N. Swanson, Eds.), pages 489–528. Springer, New York.
- Polzehl, J. and Spokoiny, V. (2006). Varying coefficient GARCH versus local constant volatility modeling: comparison of the predictive power. Technical report, SFB 649 discussion paper.
- Poon, S.-H. and Granger, C. W. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539.
- Priestley, M. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 204–237.
- Priestley, M. (1988). Non-linear and non-stationary time series analysis. *Academic Press, London*.
- Shephard, N. (1996). *Statistical Aspects of ARCH and Stochastic Volatility Time Series in Econometrics, Finance and Other Fields*. Chapman and Hall, London.
- Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *Review of Economics and Statistics*, 87(3):503–522.