NoVaS Transformations: Flexible Inference for Volatility Forecasting*

Dimitris N. Politis†
Department of Mathematics and Department of Economics
University of California, San Diego
Address: La Jolla, CA 92093-0112, USA
Telephone: (858) 534-5861; Fax: (858) 534-5273
Email: dpolitis@ucsd.edu

&

Dimitrios D. Thomakos
Department of Economics
University of Peloponnese, Greece
Rimini Center for Economic Analysis, Italy
Email: thomakos@uop.gr

April 8, 2008

*Earlier results from this research were presented at the 56th Session of the ISI (Lisbon, 2007), the Department of Economics, University of Cyprus, and the Department of Economics, University of Crete, and Department of Accounting and Finance, Athens University of Economics and Business, Greece. We would like to thank Elena Andreou and seminar participants for useful comments and suggestions. All errors are ours.

†Corresponding author.

NoVaS Transformations:
Flexible Inference for Volatility Forecasting

April 8, 2008

Abstract

In this paper we present several new findings on the NoVaS transformation approach for volatility forecasting introduced by Politis (2003a,b, 2007). In particular: (a) we present a new method for accurate volatility forecasting using NoVaS ; (b) we introduce a "time-varying" version of NoVaS and show that the NoVaS methodology is applicable in situations where (global) stationarity for returns fails such as the cases of local stationarity and/or structural breaks and/or model uncertainty; (c) we conduct an extensive simulation study on the forecasting ability of the NoVaS approach under a variety of realistic data generating processes (DGP); and (d) we illustrate the forecasting ability of NoVaS on a number of real datasets and compare it to realized and range-based volatility measures. Our empirical results show that the NoVaS -based forecasts lead to a much 'tighter' distribution of the forecasting performance measure. Perhaps our most remarkable finding is the *robustness* of the NoVaS forecasts in the context of structural breaks and/or other non-stationarities of the underlying data. Also striking is that forecasts based on NoVaS invariably outperform those based on the benchmark GARCH(1,1) even when the true DGP *is* GARCH(1,1) when the sample size is moderately large, e.g. 350 daily observations.

Keywords: ARCH, forecasting, GARCH, local stationarity, robustness, structural breaks, volatility.

1 Introduction

Accurate forecasts of the volatility of financial returns is an important part of empirical financial research. In this paper we present a number of new findings on the NoVaS transformation approach to volatility prediction. The NoVaS methodology was introduced by Politis (2003a,b, 2007) and further expanded in Politis and Thomakos (2007). The name of the method is an acronym for 'Normalizing and Variance Stabilizing' transformation. NoVaS is based on exploratory data analysis ideas, it is model-free, data-adaptive and—as the paper at hand hopes to demonstrate—especially relevant when making forecasts in the context of underlying data generating processes (DGPs) that exhibit non-stationarities (e.g. locally stationary time series, series with parameter breaks or regime switching etc.). In general, NoVaS allows for a flexible approach to inference, and is also well suited for application to short time series.

The original development of the NoVaS approach was made in Politis (2003a,b, 2007) having as its 'spring board' the popular ARCH model with normal innovations. In these papers, the main application was forecasting squared returns (as a proxy for forecasting volatility), and the evaluation of forecasting performance was addressed via the L_1-norm (instead of the usual MSE) since the case was made that financial returns might not have finite 4th moment.

In the paper at hand we further investigate the applicability of NoVaS in a forecasting context. First, we present a method for *bona fide* volatility forecasting, extending the original NoVaS notion of forecasting squared returns. Second, we conduct a very comprehensive simulation study about the relative forecasting performance of NoVaS: we consider a wide variety of volatility models as data generating processes (DGPs), and we compare the forecasting performance of NoVaS with that of a benchmark GARCH(1,1) model. We introduce the notion of a "time-varying" NoVaS approach and show that is especially relevant in these cases where the assumption of global stationarity fails. The results of our simulations show that NoVaS forecasts lead to a much 'tighter' distribution of the forecasting performance measure (mean absolute deviation of the forecast errors), when compared to the benchmark model, for all DGPs we consider. This finding is especially relevant in the context of volatility forecasting for risk management. We further illustrate the use of NoVaS for a number of real datasets and compare the forecasting performance of NoVaS-based volatility forecasts with realized and range-based volatility measures, which are frequently used in assessing the performance of volatility forecasts.

The literature on volatility modeling, forecasting and the evaluation of volatility forecasts is very large and appears to be continuously expanding. Possibly related to the paper at hand is the work by Hansen (2006) in which the problem of forming predictive intervals is addressed using a semiparametric, transformation-based approach. Hansen works with a set of (standardized)

residuals from a parametric model, and then uses the empirical distribution function of these residuals to compute conditional quantiles that can be used in forming prediction intervals. The main similarity between Hansen's work and NoVaS is that both approaches use a transformation of the original data and the empirical distribution to make forecasts. The main difference, however, is that Hansen works in the context of a (possibly misspecified) model whereas NoVaS is totally model-free.

We can only selectively mention here some recent literature related to the forecasting problems we address: Mikosch and Starica (2000) for change in structure in time series and GARCH modeling; Meddahi (2001) for an eigenfunction volatility modeling approach; Peng and Yao (2003) for robust LAD estimation of GARCH models; Poon and Granger (2003) for assessing the forecasting performance of various volatility models; Hansen, Lunde and Nason (2003) on selecting volatility models; Andersen, Bollerslev and Meddahi (2004) on analytic evaluation of volatility forecasts; Ghysels and Forsberg (2007) on the use and predictive power of absolute returns; Francq and Zakoïan (2005) on switching regime GARCH models; Hillebrand (2005) on GARCH models with structural breaks; Hansen and Lunde (2005, 2006) 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; and Ghysels, Santa Clara and Valkanov (2006) for predicting volatility by mixing data at different frequencies. The whole 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", forthcoming in the *Handbook of Economic Forecasting*, see Andersen *et al.* (2006). Fryzlewicz, Sapatinas and Subba-Rao (2006, 2007) and Dahlhaus and Subba-Rao (2006, 2007) all work in the context of local stationarity and a new class of ARCH processes with slowly varying parameters. Of course this list is by no means complete.

The rest of the paper is organized as follows: in Section 2 we briefly review the general development of the NoVaS approach; in Section 3 we present the design of our simulation study and discuss the simulation results on forecasting performance; in Section 4 we present empirical applications of NoVaS using real-world data; finally, in Section 5 we offer some concluding remarks.

2 Review of the NoVaS Methodology

In this section we present a brief overview of the NoVaS transformation, the implied NoVaS distribution, the methods for distributional matching and NoVaS forecasting. For a more comprehensive review of the NoVaS methodology see Politis and Thomakos (2007).

2.1 NoVaS transformation and implied distribution

Let us consider a zero mean, strictly stationary time series $\{X_t\}_{t\in\mathbb{Z}}$ corresponding to the returns of a financial asset. We assume that the basic properties of X_t correspond to the 'stylized facts'[1] of financial returns:

1. X_t has a non-Gaussian, approximately symmetric distribution that exhibits excess kurtosis.
2. X_t has time-varying conditional variance (volatility), denoted by $h_t^2 \stackrel{\text{def}}{=} \mathsf{E}\left[X_t^2|\mathcal{F}_{t-1}\right]$ that exhibits strong dependence, where $\mathcal{F}_{t-1} \stackrel{\text{def}}{=} \sigma(X_{t-1}, X_{t-2}, \dots)$.
3. X_t is dependent although it possibly exhibits low or no autocorrelation which suggests possible nonlinearity.

These well-established properties affect the way one models and forecasts financial returns and their volatility and form the starting point of the NoVaS methodology.

The first step in the NoVaS transformation is variance stabilization to address the time-varying conditional variance property of the returns. We construct an empirical measure of the *time-localized* variance of X_t based on the information set $\mathcal{F}_{t|t-p} \stackrel{\text{def}}{=} \{X_t, X_{t-1}, \dots, X_{t-p}\}$

$$\gamma_t \stackrel{\text{def}}{=} G(\mathcal{F}_{t|t-p}; \alpha, \boldsymbol{a}) \ , \ \gamma_t > 0 \ \forall t \tag{1}$$

where α is a scalar control parameter, $\boldsymbol{a} \stackrel{\text{def}}{=} (a_0, a_1, \dots, a_p)^\top$ is a $(p+1)\times 1$ vector of control parameters and $G(\cdot;\alpha,\boldsymbol{a})$ is to be specified.[2] The function $G(\cdot;\alpha,\boldsymbol{a})$ can be expressed in a variety of ways, using a parametric or a semiparametric specification. To keep things simple we assume that $G(\cdot;\alpha,\boldsymbol{a})$ is additive and takes the following form:

$$\begin{gathered} G(\mathcal{F}_{t|t-p}; \alpha, \boldsymbol{a}) \stackrel{\text{def}}{=} \alpha s_{t-1} + \sum_{j=0}^{p} a_j g(X_{t-j}) \\ s_{t-1} = (t-1)^{-1} \textstyle\sum_{j=1}^{t-1} g(X_j) \end{gathered} \tag{2}$$

with the implied restrictions (to maintain positivity for γ_t) that $\alpha \geq 0$, $a_i \geq 0$, $g(\cdot) > 0$ and $a_p \neq 0$ for identifiability. Although other choices are possible, the natural choices for $g(z)$ are $g(z) = z^2$ or $g(z) = |z|$. With these designations, our empirical measure of the time-localized variance becomes a combination of an unweighted, recursive estimator s_{t-1} of the unconditional variance of the returns $\sigma^2 = \mathsf{E}\left[X_1^2\right]$, or of the mean absolute deviation of the returns $\delta = \mathsf{E}|X_1|$, and a weighted average of the current[3] and the past p values of the squared or absolute returns.

[1] Departures from the assumption of these 'stylized facts' have been discussed in Politis and Thomakos (2007); in this paper, we are mostly concerned about departures/breaks in stationarity—see Section 2.4 in what follows.

[2] See the discussion about the calibration of α and $\boldsymbol{a}$ in the next section.

[3] The necessity and advantages of including the current value is elaborated upon by Politis (2003a,b,2004,2007).

Using $g(z) = z^2$ results in a measure that is reminiscent of an $ARCH(p)$ model which was employed in Politis (2003a,b, 2007). The use of absolute returns, i.e. $g(z) = |z|$ has also been advocated for volatility modeling; see e.g. Ghysels and Forsberg (2007) and the references therein. Robustness in the presence of outliers in an obvious advantage of absolute vs. squared returns. In addition, note that the mean absolute deviation is *proportional* to the standard deviation for the symmetric distributions that will be of current interest.

The second step in the NoVaS transformation is to use γ_t in constructing a studentized version of the returns, akin to the standardized innovations in the context of a parametric (e.g. GARCH-type) model. Consider the series W_t defined as:

$$W_t \equiv W_t(\alpha, \boldsymbol{a}) \stackrel{\text{def}}{=} \frac{X_t}{\phi(\gamma_t)} \tag{3}$$

where $\phi(z)$ is the time-localized standard deviation that is defined relative to our choice of $g(z)$, for example $\phi(z) = \sqrt{z}$ if $g(z) = z^2$ or $\phi(z) = z$ if $g(z) = |z|$. The aim now is to choose the NoVaS parameters in such a way as to make W_t follow as closely as possible a chosen target distribution that is easier to work with. The natural choice for such a distribution is the normal—hence the 'normalization' in the NoVaS acronym; other choices (such as the uniform) are also possible in applications, although perhaps not as intuitive—see e.g. Politis and Thomakos (2007).

Remark 1. The above distributional matching should not only focus on the first marginal distribution of the transformed series W_t. Rather, the *joint* distributions of W_t should be normalized as well; this can be accomplished by attempting to normalize linear combinations of the form $W_t + \lambda W_{t-k}$ for different values of the lag k and the weight parameter λ; see e.g. Politis (2003a,b, 2007). For practical applications it appears that the distributional matching of the first marginal distribution is quite sufficient.

A related idea is the notion of an *implied* model that is associated with the NoVaS transformation that was put forth by Politis (2004). For example, solving for X_t in eq. (3), and using the fact that γ_t depends on X_t, it follows that:

$$X_t = U_t A_{t-1} \tag{4}$$

where (corresponding to using either squared or absolute returns) the two terms on the right-hand side above are given by

$$U_t \stackrel{\text{def}}{=} \left\{ \begin{array}{lll} W_t/\sqrt{1 - a_0 W_t^2} & \text{if} & \phi(z) = \sqrt{z} \\ W_t/(1 - a_0|W_t|) & \text{if} & \phi(z) = z \end{array} \right\} \tag{5}$$

and

$$A_{t-1} \stackrel{\text{def}}{=} \left\{ \begin{array}{lll} \sqrt{\alpha s_{t-1} + \sum_{j=1}^{p} a_j X_{t-j}^2} & \text{if} & g(z) = z^2 \\ \alpha s_{t-1} + \sum_{j=1}^{p} a_j |X_{t-j}| & \text{if} & g(z) = |z| \end{array} \right\} \tag{6}$$

If one postulates that the U_t are i.i.d. according to some desired distribution, then eq. (4) becomes a *bona fide* model.[4] If the distribution of U_t is the one implied by eq. (4) with W_t having a (truncated[5]) normal distribution, then eq. (4) is the model that is 'associated' with NoVaS. With the normal target distribution and our two options for computing γ_t, the implied distribution of U_t can be derived using calculus and given below:

$$\begin{aligned} f_1(u,a_0) &= c_1(a_0)\times(1+a_0u^2)^{-1.5}\exp\left[-0.5u^2/(1+a_0u^2)\right] \quad \text{when } g(z)=z^2 \\ f_2(u,a_0) &= c_2(a_0)\times(1+a_0|u|)^{-2}\exp\left[-0.5u^2/(1+a_0|u|)^2\right] \quad \text{when } g(z)=|z| \end{aligned} \tag{7}$$

The constants $c_i(a_0)$, for $i=1,2$, ensure that the densities are proper and integrate to one.

Note that the $f_i(u,a_0)$ distributions lack moments of high orders. In particular, $f_1(u,a_0)$ has finite moments of order a if $a<2$, whereas $f_2(u,a_0)$ has finite moments of order a if $a<1$. In the terminology of Politis (2004), $f_1(u,a_0)$ has 'almost finite'[6] second moment but moments higher than two are infinite; similarly, $f_2(u,a_0)$ has 'almost' finite first moment but moments higher than one are infinite.[7]

Politis (2003b, 2004, 2007) makes the case that financial returns seem to have finite second moment but infinite 4th moments. In that case, the normal target does not seem to be compatible with the choice of absolute returns—and the same is true of the uniform target—as it seems that the case $g(z)=|z|$ might be better suited for data that do not have a finite second moment. Nevertheless, there is always the possibility of encountering such extremely heavy-tailed data, e.g. in emerging markets, for which the absolute returns might be helpful.[8]

Remark 2. The set-up of potentially infinite 4th moments has been considered by Hall and Yao (2003) and Berkes and Horvath (2004) among others, and has important implications on an issue crucial in forecasting, namely the choice of loss function for evaluating forecast performance. The most popular criterion for measuring forecasting performance is the mean-squared error (MSE) which, however, is inapplicable in forecasting squared returns (and volatility) when the 4th

[4] In particular, when $g(z)=z^2$, then (4) is tantamount to an $ARCH(p)$ model.

[5] To elaborate, recall that the range of W_t is bounded. Using eq. (3) it is straightforward to show that $|W_t|\le 1/\sqrt{a_0}$, when $g(z)=z^2$, whereas $|W_t|\le 1/a_0$, when $g(z)=|z|$. This, however, creates no practical problems. With a judicious choice for a_0 the boundedness assumption is effectively not noticeable. Take, for example, the case where the target distribution for W_t is the standard normal and $g(z)=z^2$. A simple restriction would then be $a_0\le 1/9$, which would make W_t to take values within ± 3 that cover 99.7% of the mass of the standard normal distribution. Similarly, when $g(z)=|z|$ then a_0 can be chosen as $a_0\le 1/3$.

[6] It would have *exactly* finite second moment with a more judicious (but cumbersome) choice of the truncation parameter.

[7] Similar expressions and implied distribution behavior applies when one uses a different target distribution for distributional matching, e.g. the uniform. For further details see Politis and Thomakos (2007).

[8] This might well be the case of the EFG dataset of Section 4 in what follows.

moment is infinite. In contrast, the mean absolute deviation (MAD) is as intuitive as the MSE but does not suffer from this deficiency, and can thus be used in evaluating the forecasts of either squared or absolute returns and volatility; this L_1 loss criterion will be our preferred choice in this paper.[9]

2.2 NoVaS distributional matching

2.2.1 Calibration

We next turn to the issue of optimal selection—calibration—of the NoVaS parameters. The objective is to achieve the desired distributional matching with as few parameters as possible (parsimony). The free parameters are p (the NoVaS order), and $(\alpha, \boldsymbol{a})$. The parameters α and $\boldsymbol{a}$ are constrained to be nonnegative to ensure the same for the variance. In addition, motivated by unbiasedness considerations, Politis (2003a,b, 2007) suggested the convexity condition $\alpha + \sum_{j=0}^{p} a_j = 1$. Finally, thinking of the coefficients a_i as local smoothing weights, it is intuitive to assume $a_i \geq a_j$ for $i > j$.

We now discuss in detail the case when $\alpha = 0$; see Remark 3 for the case of nonzero α. The simplest scheme that satisfies the above conditions is equal weighting, that is $a_j = 1/(p+1)$ for all $j = 0, 1, \ldots, p$. These are the 'simple' NoVaS weights proposed in Politis (2003a,b, 2007). An alternative allowing for greater weight to be placed on earlier lags is to consider exponential weights of the form:

$$a_j = \left\{ \begin{array}{lll} 1/\sum_{j=0}^{p} \exp(-bj) & \text{for} & j = 0 \\ a_0 \exp(-bj) & \text{for} & j = 1, 2, \ldots, p \end{array} \right\} \tag{8}$$

where b is the rate; these are the 'exponential' NoVaS weights proposed in Politis (2003a,b, 2007).

Both the 'simple' and 'exponential' NoVaS require the calibration of two parameters: a_0 and p for 'simple', and a_0 and b for 'exponential'.[10] Nevertheless, the exponential weighting scheme allows for greater flexibility, and will be our preferred method. In this connection, let $\boldsymbol{\theta} \stackrel{\text{def}}{=} (p, b) \mapsto (\alpha, \boldsymbol{a})$, and denote the studentized series as $W_t \equiv W_t(\boldsymbol{\theta})$ rather than $W_t \equiv W_t(\alpha, \boldsymbol{a})$. For any given value of the parameter vector $\boldsymbol{\theta}$ we need to evaluate the 'closeness' of the marginal distribution of W_t with the target distribution. To do this, an appropriately defined objective function is needed, and discussed in the next subsection.

[9] See also the recent paper by Hansen and Lunde (2006) about the relevance of MSE in evaluating volatility forecasts.

[10] Note that now p in the exponential NoVaS could, in theory, be set to infinity; it takes a finite (but large) value only for practical considerations.

2.2.2 Objective functions for optimization

To evaluate whether the objective of distributional matching has been achieved, many different objective functions could be used. For example, one could use moment-based matching (e.g. kurtosis matching as originally proposed by Politis [2003a,b, 2007]), or complete distributional matching via any goodness-of-fit statistic like the Kolmogorov-Smirnov statistic, the quantile-quantile correlation coefficient (Shapiro-Wilks statistic for testing normality) and others. All these measures are essentially distance-based and the optimization will attempt to minimize the distance between empirical (sample) and target values.[11]

Let us consider the simplest case first, i.e., moment matching. Assuming that the data are approximately symmetrically distributed and only have excess kurtosis, one first computes the sample excess kurtosis of the studentized returns as:

$$\mathcal{K}_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\sum_{t=1}^{n}(W_t - \bar{W}_n)^4}{n s_n^4} - \kappa^* \tag{9}$$

where $\bar{W}_n \stackrel{\text{def}}{=} (1/n)\sum_{t=1}^{n} W_t$ denotes the the sample mean, $s_n^2 \stackrel{\text{def}}{=} (1/n)\sum_{t=1}^{n}(W_t - \bar{W}_n)^2$ denotes the sample variance of the $W_t(\boldsymbol{\theta})$ series, and κ^* denotes the theoretical kurtosis coefficient of the target distribution. For the normal distribution $\kappa^* = 3$.

The objective function for this case can be taken to be the absolute value, i.e., $D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} |\mathcal{K}_n(\boldsymbol{\theta})|$, and one would adjust the values of $\boldsymbol{\theta}$ so as to minimize $D_n(\boldsymbol{\theta})$. As noted by Politis (2003a,b, 2007) such an optimization procedure will always have a solution in view of the intermediate value theorem. To see this, note that when $p = 0$, a_0 must equal 1, and thus $W_t = sign(X_t)$ that corresponds to $\mathcal{K}_n(\boldsymbol{\theta}) < 0$ for any choice of the target distribution. On the other hand, for large values of p we expect that $\mathcal{K}_n(\boldsymbol{\theta}) > 0$, since it is assumed that the data have large excess kurtosis. Therefore, there must be a value of $\boldsymbol{\theta}$ that will make the sample excess kurtosis approximately equal to zero. This observation motivates the following algorithm for exponential NoVaS (Politis [2003a, 2007]):

- Let p take a very high starting value, e.g., let $p_{\max} \approx n/4$.
- Let $\alpha = 0$ and consider a discrete grid of b values, say $B \stackrel{\text{def}}{=} (b_{(1)}, b_{(2)}, \ldots, b_{(M)})$, $M > 0$. Find the optimal value of b, say b^*, that minimizes $D_n(\boldsymbol{\theta})$ over $b \in B$, and compute the optimal parameter vector $\boldsymbol{a}^*$ using eq. (8).

[11] This part of the NoVaS application appears similar at the outset to the Minimum Distance Method (MDM) of Wolfowitz (1957). Nevertheless, their objectives are quite different since the latter is typically employed for parameter estimation and testing whereas in NoVaS there is little interest in parameters—the focus lying on effective forecasting.

- Trim the value of p by removing (i.e., setting to zero) the a_j parameters that do not exceed a pre-specified threshold, and re-normalize the remaining parameters so that their sum equals one.

While kurtosis matching has been found to work well with financial returns data, it is straightforward to extend the above algorithm to a variety of different objective functions. For example, one can opt for a combination of skewness and kurtosis matching[12], or for goodness-of-fit statistics such as the Shapiro-Wilks or the Kolmogorov-Smirnov statistic. One performs the same steps but simply evaluates a different objective function. Note that for any choice of the objective function $D_n(\boldsymbol{\theta}) \geq 0$ and, as noted in the algorithm above, the optimal values of the parameters are clearly determined by the condition:

$$\boldsymbol{\theta}_n^* \stackrel{\text{def}}{=} \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, D_n(\boldsymbol{\theta}) \tag{10}$$

Remark 3. The discussion so far was under the assumption that the parameter α, that controls the weight given to the recursive estimator of the unconditional variance, is zero. If desired one can select a non-zero value by doing a direct search over a discrete grid of possible values while obeying the summability condition $\alpha + \sum_{j=0}^{p} a_j = 1$. For example, one can choose the value of α that optimizes out-of-sample predictive performance; see Politis (2003a,b, 2007) for more details.

2.3 NoVaS Forecasting

Once the NoVaS parameters are calibrated one can compute volatility forecasts. In fact, as Politis (2003a,b, 2007) has shown, one can compute forecasts for different functions of the returns, including higher powers (with absolute value or not). The choice of an appropriate forecasting loss function, both for producing and for evaluating the forecasts, is crucial for maximizing forecasting performance.

Per our Remark 2, we now focus on the L_1 loss function for producing the forecasts and the mean absolute deviation (MAD) of the forecast errors for assessing forecasting performance. After optimization of the NoVaS parameters we now have both the optimal transformed series $W_t^* = W_t(\boldsymbol{\theta}_n^*)$ but also the series U_t^* that is defined as a function of W_t^* via eq. (5).

Let $\Pi_k\left[X|_Z\right]$ denote the k^{th} (regular or absolute) conditional power operator of the argument X given the argument Z. For example, $\Pi_1\left[XZ|_Z\right] = XZ$, $\Pi_2\left[XZ|_Z\right] = (X^2|_Z)\cdot Z^2$ etc. Applying the power operator in the definition of the *implied* model of eq. (4) at time $n+1$ we obtain:

$$\Pi_k\left[X_{n+1}|\mathcal{F}_n\right] = \Pi_k\left[U_{n+1}^*|\mathcal{F}_n\right]\Pi_k\left[A_n^*\right] \tag{11}$$

[12]When the target distribution is the standard normal the objective function could be similar to the well known Jarque-Bera test for assessing normality.

Depending on our choice of k, and on whether we take regular or absolute powers, we can now forecast returns ($k = 1$), absolute returns ($k = 1$ with absolute value), squared returns ($k = 2$), etc., and the task is simplified in forecasting the corresponding power of the U^*_{n+1} series. To see this note that, in the context of the L_1 forecasting loss function, the conditional median is the optimal predictor, so we have:

$$\mathsf{Med}\left[\Pi_k\left[X_{n+1}|\mathcal{F}_n\right]\right] = \mathsf{Med}\left[\Pi_k\left[U^*_{n+1}|\mathcal{F}_n\right]\right]\Pi_k\left[A^*_n\right] \tag{12}$$

where $\mathsf{Med}\left[x\right]$ stands for the median of x. Therefore, what we are after is an estimate of the conditional median of $\Pi_k\left[U^*_{n+1}|\mathcal{F}_n\right]$.

The rest of the procedure depends on the dependence properties of the studentized series W^*_n and the target distribution. From our experience, what has invariably been observed with financial returns is that their corresponding W^*_n series appears—for all practical purposes—to be uncorrelated.[13] If the target distribution is the normal then—by the approximate normality of its *joint* distributions—the W^*_n series would be independent as well. The series U^*_n would inherit the W^*_ns independence by eq. (5), and therefore the best estimate of the conditional median $\mathsf{Med}\left[\Pi_k\left[U^*_{n+1}|\mathcal{F}_n\right]\right]$ is the unconditional sample median of the appropriate power of the U^*_n series, namely $\widehat{\mathsf{Med}}\left[\Pi_k\left(U^*_n\right)\right]$.

Based on the above discussion we are now able to obtain volatility forecasts $\widehat{h}^2_{n+1}$ in a variety of ways: (a) we can use the forecasts of absolute or squared returns; (b) we can use only the component of the conditional variance A^2_n for $\phi(z) = \sqrt{z}$ or A_n for $\phi(z) = z$, akin to a GARCH approach; (c) we can combine (a) and (b) and use the forecast of the empirical measure $\widehat{\gamma}_{n+1}$.

Consider the use of squared returns first. The volatility forecast based on (a) above would be:

$$\widehat{h}^2_{n+1,1} \equiv \widehat{X}^2_{n+1} \stackrel{\text{def}}{=} \widehat{\mathsf{Med}}\left[\Pi_2\left(U^*_n\right)\right]\Pi_2\left[A^*_n\right] \tag{13}$$

When using (b) the corresponding forecast would just be the power of the A^*_n component, something very similar to an $ARCH(\infty)$ forecast:

$$\widehat{h}^2_{n+1,2} \stackrel{\text{def}}{=} \Pi_2\left[A^*_n\right] \tag{14}$$

However, the most relevant and appropriate volatility forecast in the NoVaS context should be based on (c), i.e. on a forecast of the estimate of the time-localized variance measure $\widehat{\gamma}_{n+1}$, which was originally used to initiate the NoVaS procedure in eq. (1). What is important to note is that forecasting based on $\widehat{\gamma}_{n+1}$ is neither forecasting of squared returns nor forecasting based on past

[13] This is an empirical finding; if, however, the W^*_n series is not independent then a slightly different procedure involving a (hopefully) linear predictor would be required—see Politis (2003a, 2007) and Politis and Thomakos (2007) for details.

information alone. It is, in fact, a linear combination of the two, thus incorporating elements from essentially two approaches. Using eqs. (1), (2), (5) and (6) it is straightforward to show that $\widehat{\gamma}_{n+1}$ can be expressed as:

$$\begin{aligned} \widehat{\gamma}_{n+1} \equiv \widehat{h}^2_{n+1,3} &\stackrel{\text{def}}{=} \left\{ a_0^* \widehat{\mathsf{Med}} \left[\Pi_2 \left(U_n^* \right) \right] + 1 \right\} \Pi_2 \left[A_n^* \right] \\ &= a_0^* \widehat{h}^2_{n+1,1} + \widehat{h}^2_{n+1,2} \end{aligned} \tag{15}$$

Equation (15) is our new proposal for volatility forecasting using NoVaS. In his original work, Politis (2003a) used eq. (13), and in effect conducted forecasting of the one-step-ahead squared returns via NoVaS. By contrast, eq. (15) is a *bona fide* predictor of the one-step-ahead *volatility,* i.e., the conditional variance. For this reason, eq. (15) will be the formula used in what follows, our simulations and real data examples.

Forecasts using absolute returns are constructed in a similar fashion, the only difference being that we will be forecasting *directly* standard deviations $\widehat{h}_{n+1}$ and not variances. Using again eqs. (1), (5) and (6) it is easy to show that the forecast based on (c) would be given by:

$$\begin{aligned} \widehat{\gamma}_{n+1} \equiv \widehat{h}_{n+1,3} &\stackrel{\text{def}}{=} \left\{ a_0^* \widehat{\mathsf{Med}} \left[\Pi_1 \left(U_n^* \right) \right] + 1 \right\} \Pi_1 \left[A_n^* \right] \\ &= a_0^* \widehat{h}_{n+1,1} + \widehat{h}_{n+1,2} \end{aligned} \tag{16}$$

with $\widehat{h}_{n+1,1}$ and $\widehat{h}_{n+1,2}$ being identical expressions to eqs. (13) and (14) which use the first order absolute power transformation.

2.4 Departures from the assumption of stationarity: local stationarity and structural breaks

Consider the case of a very long time series $\{X_1, \ldots, X_n\}$, e.g., a daily series of stock returns spanning a decade. It may be unrealistic to assume that the stochastic structure of the series has stayed invariant over such a long stretch of time. A more realistic model might assume a slowly-changing stochastic structure, i.e., a locally stationary model as given by Dahlhaus (1997).

Recent research has tried to address this issue by fitting time-varying GARCH models to the data but those techniques have not found global acceptance yet, in part due to their extreme computational cost. Fryzlewicz, Sapatinas and Subba-Rao (2006, 2007) and Dahlhaus and Subba-Rao (2006, 2007b) all work in the context of local stationarity for a new class of ARCH processes with slowly varying parameters.

Surprisingly, NoVaS is flexible enough to accommodate such smooth/slow changes in the stochastic structure. All that is required is a *time-varying* NoVaS fitting, i.e., selecting/calibrating the NoVaS parameters on the basis of a *rolling* window of data as opposed to using the entire available past. Interestingly, as will be apparent in our simulations, the time-varying NoVaS

method works well even in the presence of *structural breaks* that would typically cause a breakdown of traditional methods unless explicitly taken into account. The reason for this robustness is the simplicity in the NoVaS estimate of local variance: it is just a linear combination of (present and) past squared returns. Even if the coefficients of the linear combination are not optimally selected (which may happen in the neighborhood of a break), the linear combination remains a reasonable estimate of local variance.

By contrast, the presence of structural breaks can throw off the (typically nonlinear) fitting of GARCH parameters. Therefore, a GARCH practitioner must always be on the look-out for structural breaks, essentially conducting a hypothesis test before each application. While there are several change point tests available in the literature, the risk of non-detection of a change point can be a concern. Fortunately, the NoVaS practitioner does not have to worry about structural breaks because of the aforementioned *robustness* of the NoVaS approach.

3 NoVaS Forecasting Performance: A Simulation Analysis

It is of obvious interest to compare the forecasting performance of NoVaS-based volatility forecasts with the standard benchmark model, the GARCH(1,1), under a variety of different underlying DGPs. Although there are numerous models for producing volatility forecasts, including direct modeling of realized volatility series, it is not clear which of these models should be used in any particular situation, and whether they can always offer substantial improvements over the GARCH benchmark. In the context of a simulation, we will be able to better see the relative performance of NoVaS -based volatility forecasts versus GARCH-based forecasts and, in addition, we will have available the true volatility measure for forecast evaluation. This latter point, the availability of an appropriate series of true volatility, is important since in practice we do not have such a series of true volatility. The proxies range from realized volatility—generally agreed to be one of the best (if not the best) such measure—, to range-based measures, and to squared returns. We use such proxies in the empirical examples of the next section.

3.1 Simulation Design

We consider a variety of models as possible DGPs. Each model $j = 1, 2, \ldots, M(= 7)$ is simulated over the index $i = 1, 2, \ldots, N(= 500)$ with time indices $t = 1, 2, \ldots, T(= 1250)$. The sample size T amounts to about 5 years of daily data. The parameter values for the models are chosen so as to reflect annualized volatilities between about 8% to 25%, depending on the model being used. For each model we simulate a volatility series and the corresponding returns series based on the

standard representation:

$$\begin{aligned} X_{t,ij} &\stackrel{\text{def}}{=} \mu_j + h_{t,ij} Z_{t,ij} \\ h^2_{t,ij} &\stackrel{\text{def}}{=} h_j(h^2_{t-1,ij}, X^2_{t-1,ij}, \boldsymbol{\theta}_{tj}) \end{aligned} \tag{17}$$

where $h_j(\cdot)$ changes depending on the model being simulated.

The seven models simulated are: a standard GARCH, a GARCH with discrete breaks (B-GARCH), a GARCH with slowly varying parameters (TV-GARCH), a Markov switching GARCH (MS-GARCH), a smooth transition GARCH (ST-GARCH), a GARCH with an added deterministic function (D-GARCH) and a stochastic volatility model (SV-GARCH). Note that the parameter vector $\boldsymbol{\theta}_t$ will be time-varying for the Markov switching model, the smooth transition model, the time-varying parameters model and the discrete breaks model. For the simulation we set $Z_t \sim t_{(3)}$, standardized to have unit variance.[14]

We next present the volatility eqs. of the above models. For ease of notation we drop the i and j subscripts when presenting the models. The first model we simulate is a standard GARCH(1,1) with volatility equation given by:

$$h^2_t = \omega + \alpha h^2_{t-1} + \beta(X_{t-1} - \mu)^2 \tag{18}$$

The parameter values were set to $\alpha = 0.9$, $\beta = 0.07$ and $\omega = 1.2e-5$, corresponding to an annualized volatility of 10%. The mean return was set to $\mu = 2e-4$ (same for all models, except the MS-GARCH) and the volatility series was initialized with the unconditional variance.

The second model we simulate is a GARCH(1,1) with discrete changes (breaks) in the variance parameters. These breaks depend on changes in the annualized unconditional variance, ranging from about 8% to about 22% and we assume two equidistant changes per year for a total of $B = 10$ breaks. The model form is identical to the GARCH(1,1) above:

$$h^2_t = \omega_b + \alpha_b h^2_{t-1} + \beta_b(X_{t-1} - \mu)^2 \ , \ b = 1, 2, \ldots, B \tag{19}$$

The α_b parameters were drawn from a uniform distribution in the interval $[0.8, 0.99]$ and the β_b parameters were computed as $\beta_b = 1 - \alpha_b - c$, for c either 0.015 or 0.02. The ω_b parameters were computed as $\omega_b = \sigma^2_b(1 - \alpha_b - \beta_b)/250$, where σ^2_b is the annualized variance.

The third model we simulate is a GARCH(1,1) with slowly varying variance parameters, of a nature very similar to the time-varying ARCH models recently considered by Dahlhaus and Subba-Rao (2006, 2007). The model is given by:

$$h^2_t = \omega(t) + \alpha(t) h^2_{t-1} + \beta(t)(X_{t-1} - \mu)^2 \tag{20}$$

[14] We fix the degrees of freedom to their true value of 3 during estimation and forecasting, thus giving GARCH a relative advantage in estimation.

where the parameters satisfy the finite unconditional variance assumption $\alpha(t)+\beta(t) < 1$ for all t. The parameters functions $\alpha(t)$ and $\beta(t)$ are sums of sinusoidal functions of different frequencies ν_k of the form $c(t) = \sum_{k=1}^{K} \sin(2\pi\nu_k t)$, for $c(t) = \alpha(t)$ or $\beta(t)$. For $\alpha(t)$ we set $K = 4$ and $\nu_k = \{1/700, 1/500, 1/250, 1/125\}$ and for $\beta(t)$ we set $K = 2$ and $\nu_k = \{1/500, 1/250\}$. That is, we set the persistence parameter function $\alpha(t)$ to exhibit more variation than the parameter function $\beta(t)$ that controls the effect of squared returns.

The fourth model we simulate is a two-state Markov Switching GARCH(1,1) model, after Francq and Zakoian (2005). The form of the model is given by:

$$h_t^2 = \sum_{s=1}^{2} \mathbf{1}\left\{\mathsf{P}(S_t = s)\right\}\left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2\right] \tag{21}$$

In the first regime (high persistence and high volatility state) we set $\alpha_1 = 0.9$, $\beta_1 = 0.07$ and $\omega_1 = 2.4e-5$, corresponding to an annualized volatility of 20%, and $\mu_1 = 2e-4$. In the second regime (low persistence and low volatility state) we set $\alpha_2 = 0.7$, $\beta_2 = 0.22$ and $\omega_2 = 1.2e-4$ corresponding to an annualized volatility of 10%, and $\mu_2 = 0$. The transition probabilities for the first regime are $p_{11} = 0.9$ and $p_{12} = 0.1$ while for the second regime we try to alternative specifications $p_{21} = \{0.3, 0.1\}$ and $p_{22} = \{0.7, 0.9\}$.

The fifth model we simulate is a (logistic) smooth transition GARCH(1,1); see Taylor (2004) and references therein for a discussion on the use of such models. The form the model takes is given by:

$$h_t^2 = \sum_{s=1}^{2} Q_s(X_{t-1})\left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2\right] \tag{22}$$

where $Q_1(\cdot) + Q_2(\cdot) = 1$ and $Q_s = \left[1 + \exp(-\gamma_1 X_{t-1}^{\gamma_2})\right]^{-1}$ is the logistic transition function. The parameters $\alpha_s, \beta_s, \omega_s$ and μ_s are set to the same values as in the previous MS-GARCH model. The parameters of the transition function are set to $\gamma_1 = 12.3$ and $\gamma_2 = 1$.

The sixth model we simulate is a GARCH(1,1) model with an added smooth deterministic function yielding a *locally stationary* model as a result. For the convenient case of a linear function we have that the volatility equation is the same as in the standard GARCH(1,1) model in eq. (18) while the return equation takes the following form:

$$X_t = \mu + [a - b(t/T)]\, h_t Z_t \tag{23}$$

To ensure positivity of the resulting variance we require that $(a/b) > (t/T)$. Since $(t/T) \in (0, 1]$ we set $a = \alpha + \beta = 0.97$ and $b = (\beta/\alpha) \approx 0.078$ so that the positivity condition is satisfied for all t.

Finally, the last model we simulate is a stochastic volatility model with the volatility equation expressed in logarithmic terms and taking the form of an autoregression with normal innovations.

The model now takes the form:

$$\log h_t^2 = \omega + \alpha \log h_{t-1}^2 + w_t \ , \ w_t \sim \mathcal{N}(0, \sigma_w^2) \tag{24}$$

and we set the parameter values to $\alpha = 0.95$, $\omega \approx -0.4$ and $\sigma_w = 0.2$.

For each simulation run i and for each model j we split the sample into two parts $T = T_0 + T_1$, where T_0 is the estimation sample and T_1 is the forecast sample. We consider two values for T_0, namely 250 or 900, which correspond respectively to about a year and three and a half years of daily data. We roll the estimation sample T_1 times and thus generate T_1 out-of-sample forecasts. In estimation the parameters are re-estimated (for GARCH) or updated (for NoVaS) every 20 observations (about one month for daily data). We always forecast the volatility of the corresponding return series we simulate (eqs. (17) and (18)) and evaluate it with the known, one-step ahead simulated volatility. NoVaS forecasts are produced for using a normal target distribution and both squared and absolute returns. The nomenclature used in the tables is as follows:

1. SQNT, NoVaS forecasts made using squared returns and normal target.
2. ABNT, NoVaS forecasts made using absolute returns and normal target.
3. GARCH, L_2-based GARCH forecasts.
4. M-GARCH, L_1-based GARCH forecasts.

The naïve forecast benchmark is the sample variance of the rolling estimation sample. Therefore, for each model j being simulated we produce a total of $F = 4$ forecasts; the forecasts are numbered $f = 0, 1, 2, \ldots, F$ with $f = 0$ denoting the naïve forecast. We then have to analyze T_1 forecast errors $e_{t,ijf} \stackrel{\text{def}}{=} h_{t+1,ij}^2 - \widehat{h}_{t+1,ijf}^2$. Using these forecast errors we compute the mean absolute deviation for each model, each forecast method and each simulation run as:

$$m_{ijf} = MAD_{ijf} \stackrel{\text{def}}{=} \frac{1}{T_1} \sum_{t=T_0+1}^{T_1} |e_{t,ijf}| \tag{25}$$

The values $\{m_{ijf}\}_{i=1,\ldots,N;j=1,\ldots,M;f=0,\ldots,F}$ now become our data for meta-analysis. We compute various descriptive statistics about their distribution (across i, the independent simulation runs and for each f the different forecasting methods) like mean ($\bar{x}_f$ in the tables), std. deviation ($\widehat{\sigma}_f$ in the tables), min, the 10%, 25%, 50%, 75%, 90% quantiles and max (Q_p in the tables, $p = 0, 0.1, 0.25, 0.5, 0.75, 0.9, 1$). For example, we have that:

$$\bar{x}_{jf} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} m_{ijf} \tag{26}$$

We also compute the percentage of times that the relative (to the benchmark) MAD's of the NoVaS forecasts are better than the GARCH forecasts. Define $m_{ij,N} \stackrel{\text{def}}{=} m_{ijf}/m_{ij0}$, $f = 1, 2$ to be the ratio of the MAD of any of the NoVaS forecasts relative to the benchmark and $m_{ij,G} \stackrel{\text{def}}{=} m_{ijf}/m_{ij0}$, $f = 3, 4$ to be the ratio of the MAD of the two GARCH forecasts relative to the benchmark. That is, for each model j and forecasting method f we compute (dropping the j model subscript):

$$\widehat{P}_f \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1} \left(m_{ij,N} \leq m_{ij,G} \right) \tag{27}$$

Then, we consider the total number of times that any NoVaS forecasting method had a smaller relative MAD compared to the relative MAD of the GARCH forecasts and compute also $\widehat{P} \stackrel{\text{def}}{=} \cup_f \widehat{P}_f$ as the union across. So $\widehat{P}_f$, for $f = 1, 2$ corresponds to the aforementioned methods NoVaS methods SQNT and ABNT respectively and $\widehat{P}$ corresponds to their union.

3.2 Detailed discussion of Simulation Results

The simulation helps compare the NoVaS forecasts to the usual GARCH forecasts, i.e., L_2-based GARCH forecasts, and also to the M-GARCH forecasts, i.e., L_1-based GARCH forecasts, the latter being recommended by Politis (2003a, 2004, 2007). We break the discussion according to the seven DGP models:

- GARCH (Tables 1 and 8): In this situation, where the true DGP is GARCH, it would seem intuitive that GARCH forecasts would have an advantage. Thus, Table 2 may come as a surpise: *any* of the NoVaS methods (SQNT, ABNT) is seen to outperform both GARCH and M-GARCH in *all* measured areas: mean of the MAD distribution ($\bar{x}_f$, mean error), tightness of MAD distribution ($\hat{\sigma}_f$ and the related quantiles), and finally the % of times NoVaS MAD was better. Actually, in this setting, the GARCH forecasts are vastly underperforming as compared to the Naive benchmark. The best NoVaS method here is the SQNT that achieves a mean error $\bar{x}_f$ almost half of that of the benchmark, and with a much tighter MAD distribution.

  Table 9 sheds more light in this situation: it appears that a training sample of size 250 (Table 2) is too small for GARCH to work well; with a training sample of size 900 (Table 9) the performance of GARCH is greatly improved, and GARCH manages to beat the benchmark in terms of mean error (but not variance). SQNT NoVaS however is *still* the best method in terms of mean error and variance; it beats M-GARCH in terms of the $\widehat{P}_1$ percentage, and narrowly underperforms as compared to GARCH in this criterion.

  All in all, SQNT NoVaS volatility forecasting appears to beat GARCH forecasts when the

DGP *is* GARCH—a remarkable finding. Furthermore, GARCH apparently requires a very large training sample in order to work well; but with a sample spanning 3-4 years questions of non-stationarity may arise that will be addressed in what follows.

- GARCH with discrete breaks (B-GARCH) (Tables 2 and 9): It is apparent here that ignoring possible structural breaks when fitting a GARCH model can be disastrous. The GARCH forecasts vastly underperform compared to the Naive benchmark with either small (Table 3) or big training sample (Table 10). Interestingly, *both* NoVaS methods are better than the benchmark with SQNT seemingly the best again. The SQNT method is better than either GARCH method 99% of the time (Table 3) and at least 86% of the time (Table 10). It should be stressed here that NoVaS does *not* attempt to estimate any breaks; it applies totally automatically, and is seemingly unperturbed by structural breaks.

- GARCH with slowly varying parameters (TV-GARCH) (Tables 3 and 10): This situation is very similar to the previous one except that the performance of GARCH is a little better as compared to the benchmark—but only when given a big training sample (Table 11). However, still *both* NoVaS methods are better than either GARCH method. The best is again SQNT. Either of those beats either GARCH method 98% of the time (Table 4) and at least 88% of the time (Table 11).

- Markov switching GARCH (MS-GARCH)(Tables 4 and 11): We note again that ignoring possible intricacies—such as the Markov switching property—when fitting a GARCH model can be disastrous. GARCH forecasts vastly underperform the Naive benchmark with either small (Table 5) or big training sample (Table 12). Again *all* NoVaS methods are better than the benchmark with SQNT being the best.

- Smooth transition GARCH (ST-GARCH)(Tables 5 and 12): This situation is more like the first one (where the DGP is GARCH); with a large enough training sample, GARCH forecasts are able to beat the benchmark, and be competitive with NoVaS. Still, however, SQNT NoVaS is best, not only because of smallest mean error but also in terms of tightness of MAD distribution.

- GARCH with deterministic function (D-GARCH)(Tables 6 and 13): This is similar to the above ST-GARCH; when given a large training sample, GARCH forecasts are able to beat the benchmark, and be competitive with NoVaS . Again, SQNT NoVaS is best, not only because of smallest mean error but also in terms of tightness of MAD distribution.

- Stochastic volatility model (SV-GARCH) (Tables 7 and 14): Again, similar behavior to the

above. Although (with a big training sample) GARCH does well in terms of mean error, note the large spread of the MAD distribution.

3.3 Summary of Findings from the Simulation

The results from the simulations are very interesting and can be summarized as follows:

- GARCH forecasts are extremely off-the-mark when the training sample is not large (of the order of 2-3 years of daily data). Note that large training sample sizes are prone to be problematic if the stochastic structure of the returns changes over time.
- Even given a large training sample, NoVaS forecasts are best; this holds *even when the true DGP is actually GARCH!*
- Ignoring possible breaks (B-GARCH), slowly varying parameters (TV-GARCH), or a Markov switching feature (MS-GARCH) when fitting a GARCH model can be disastrous in terms of forecasts. In contrast, NoVaS forecasts seem unperturbed by such gross non-stationarities.
- Ignoring the presence of a smooth transition GARCH (ST-GARCH), a GARCH with an added deterministic function (D-GARCH), or a stochastic volatility model (SV-GARCH) does not seem as crucial at least when the the implied nonstationarity features are small and/or slowly varying.
- Overall, it seems that SQNT NoVaS is the volatility forecasting method of choice since it is the best in all examples except TV-GARCH (in which case it is a close second to NoVaS).

4 Empirical Application

In this section we provide an empirical illustration of the application and potential of the NoVaS approach using four real datasets. In judging the forecasting performance for NoVaS we consider different measures of 'true' volatility, including realized and range-based volatility.

4.1 Data and Summary Statistics

Our first dataset consists of monthly returns and associated realized volatility for the S&P500 index, with the sample extending from February 1970 to May 2007 for a total of $n = 448$ observations. The second dataset consists of monthly returns and associated realized, range-based volatility for the stock of Microsoft (MSFT). The sample period is from April 1986 to August 2007 for a total of $n = 257$ observations. For both these datasets the associated realized

volatility was constructed by summing daily squared returns (for the S&P500 data) or daily range-based volatility (for the MSFT data). Specifically, if we denote by $r_{t,i}$ the i^{th} daily return for month t then the monthly realized volatility is defined as $\sigma_t^2 \stackrel{\text{def}}{=} \sum_{i=1}^{m} r_{t,i}^2$, where m is the number of days. For the calculation of the realized range-based volatility denote by $H_{t,i}$ and $L_{t,i}$ the daily high and low prices for the i^{th} day of month t. The daily range-based volatility is defined as in Parkinson (1980) as $\sigma_{t,i}^2 \stackrel{\text{def}}{=} [\ln(H_{t,i}) - \ln(L_{t,i})]^2 / [4\ln(2)]$; then, the corresponding monthly realized measure would be defined as $\sigma_t^2 \stackrel{\text{def}}{=} \sum_{i=1}^{m} \sigma_{t,i}^2$. Our third dataset consists of daily returns and realized volatility for the US dollar/Japanese Yen exchange rate for a sample period between 1997 and 2005 for a total of $n = 2236$ observations. The realized volatility measure was constructed as above using intraday returns. The final dataset we examine is the stock of a major private bank in the Athens Stock Exchange, EFG Eurobank. The sample period is from 1999 to 2004 for a total of $n = 1403$ observations. For lack of intraday returns we use the daily range-based volatility estimator as defined before.

Descriptive statistics of the returns for all four of our datasets are given in Table 15. We are mainly interested in the kurtosis of the returns, as we will be using kurtosis-based matching in performing NoVaS . All series have unconditional means that are not statistically different from zero and no significant serial correlation, with the exception of the last series (EFG) that has a significant first order serial correlation estimate. Also, all four series have negative skewness which is, however, statistically insignificant except for the monthly S&P500 and MSFT series where it is significant at the 5% level. Finally, all series are characterized by heavy tails with kurtosis coefficients ranging from 5.04 (monthly S&P500) to 24.32 (EFG). The hypothesis of normality is strongly rejected for all series.

In Figures 1 to 8 we present graphs for the return series, the corresponding volatility and log volatility, the quantile-quantile (QQ) plot for the returns and four recursive moments. The computation of the recursive moments is useful for illustrating the potential unstable nature that may be characterizing the series. Figures 1 and 2 are for the monthly S&P500 returns, Figures 3 and 4 are for monthly MSFT returns, Figures 5 and 6 are for the daily USD/Yen returns and Figures 7 and 8 are for the daily EFG returns. Of interest are the figures that plot the estimated recursive moments. In Figure 2 we see that the mean and standard deviation of the monthly S&P500 returns are fairly stable while the skewness and kurtosis exhibit breaks. In fact, the kurtosis exhibits the tendency to rise in jolts/shocks and does not retreat to previous levels thereby indicating that there might not be an finite fourth moment for this series. Similar observations can be made for the other four series as far as recursive kurtosis goes. This is especially relevant about our argument that NoVaS can handle such possible global non-stationarities.

4.2 NoVaS Optimization and Forecasting Specifications

Our NoVaS in-sample analysis is performed for two possible combinations of target distribution and variance measures, i.e. squared and absolute returns using a normal target, as in the simulation analysis. We use the exponential NoVaS algorithm as discussed in section 2, with $\alpha = 0.0$, a trimming threshold of 0.01 and $p_{\max} = n/4$. The objective function for optimization is kurtosis-matching, i.e. $D_n(\boldsymbol{\theta}) = |\mathcal{K}_n(\boldsymbol{\theta})|$, as in eq. (9) — robustness to deviations from these baseline specification is also discussed below. The results of our in-sample analysis are given in Table 16. In the table we present the optimal values of the exponential constant b^*, the first coefficient a_0^*, the implied optimal lag length p^*, the value of the objective function $D_n(\boldsymbol{\theta}^*)$ and two measures of distributional fit. The first is the QQ correlation coefficient for the original series, QQ_X, and the second is the QQ correlation coefficient for the transformed series $W_t(\boldsymbol{\theta}^*)$ series, QQ_W. These last two measures are used to gauge the 'quality' of the attempted distributional matching before and after the application of the NoVaS transformation.

Our NoVaS out-of-sample analysis is reported in Tables 17, 18, 19 and 20. All forecasts are based on a rolling sample whose length n_0 differs according to the series examined: for the monthly S&P500 series we use $n_0 = 300$ observations; for the monthly MSFT series we use $n_0 = 157$ observations; for EFG series we use $n_0 = 900$ observations; for the daily USD/Yen series we use $n_0 = 1250$ observations. The corresponding evaluation samples are $n_1 = \{148, 100, 986, 503\}$ for the four series respectively. Note that our examples cover a variety of different lengths, ranging from 157 observations for the MSFT series to 1250 observations for the USD/Yen series. All forecasts we make are 'honest' out-of-sample forecasts: they use only observations prior to the time period to be forecasted. The NoVaS parameters are re-optimized as the window rolls over the entire evaluation sample (every month for the monthly series and every 20 observations for the daily series). We forecast volatility both by using absolute or squared returns (depending on the specification), as described in the section on NoVaS forecasting, and by using the empirical variance measure $\widehat{\gamma}_{n+1}$ - see eqs. (15) and (16).[15] To compare the performance of the NoVaS approach we estimate and forecast using a standard $GARCH(1,1)$ model for each series, assuming a $t_{(\nu)}$ distribution with degrees of freedom estimated from the data. The parameters of the model are re-estimated as the window rolls over, as described above. As noted in Politis (2003a,b, 2007), the performance of GARCH forecasts is found to be improved under an L_1 rather than L_2 loss. We therefore report standard mean forecasts as well as median forecasts from the GARCH models. We always evaluate our forecasts using the 'true' volatility measures

[15] All NoVaS forecasts were made without applying an explicit predictor as all $W_t(\boldsymbol{\theta}^*)$ series were found to be uncorrelated.

given in the previous section and report several measures of forecasting performance. This is important as a single evaluation measure may not always provide an accurate description of the performance of competing models.

We first calculate the mean absolute deviation (MAD) and root mean-squared (RMSE) of the forecast errors $e_t \stackrel{\text{def}}{=} \sigma_t^2 - \widehat{\sigma}_t^2$, given by:

$$MAD(e) \stackrel{\text{def}}{=} \frac{1}{n_1} \sum_{t=n_0+1}^{n} |e_t|, \qquad RMSE(e) \stackrel{\text{def}}{=} \sqrt{\frac{1}{n_1} \sum_{t=n_0+1}^{n} (e_t - \bar{e})^2} \tag{28}$$

where $\widehat{\sigma}_t^2$ denotes the forecast for any of the methods/models we use. As a Naive benchmark we use the (rolling) sample variance. We then calculate the Diebold and Mariano (1995) test for comparing forecasting models. We use the absolute value function in computing the relevant statistic and so we can formally compare the MAD rankings of the various models.

Finally, we calculate and report certain statistics based on the forecasting unbiasedness regression (also known as 'Mincer-Zarnowitz regression'). This regression can be expressed in several ways and we use the following representation:

$$e_t = a + b\widehat{\sigma}_t^2 + \zeta_t \tag{29}$$

where ζ_t is the regression error. Under the hypothesis of forecast unbiasedness we expect to have $\mathsf{E}\left[e_t|\mathcal{F}_{t-1}\right] = 0$ and therefore we expect $a = b = 0$ (and $\mathsf{E}\left[\zeta_t|\mathcal{F}_{t-1}\right] = 0$ as well.) Furthermore, the R^2 from the above regression is an indication as to how much of the forecast error variability can still be explained by the forecast. For any two competing forecasting models A and B we say that model A is superior to model B if $R_A^2 < R_B^2$, i.e. if we can make no further improvements in our forecast.

Our forecasting results are summarized in Tables 17 and 18 for the MAD and RMSE rankings and in Tables 19 and 20 for the Diebold-Mariano test and forecasting unbiasedness regressions. Similar results were obtained when using a recursive sample and are available on request.

4.3 Discussion of Results

We begin our discussion with the in-sample results and, in particular, the degree of normalization achieved by NoVaS . Looking at the value of the objective function in Table 16 we see that it is zero to three decimals for practically all cases. Therefore, NoVaS is very successful in reducing the excess kurtosis in the original return series. In addition, the quantile-quantile correlation coefficient is very high (in excess of 0.99 in all cases examined, frequently being practically one). One should compare the two QQ measures of *before and after* the NoVaS transformation to see the difference that the transformation has on the data. The case of the EFG series is particularly

worth mentioning as that series has the highest kurtosis: we can see from the table that we get a QQ correlation coefficient in excess of 0.998; this is a very clear indication that the desired distributional matching has been achieved for all practical purposes. A visual confirmation of the differences in the distribution of returns before and after NoVaS transformations is given in Figures 9 to 12. In these figures we have QQ plots for all the series and four combinations of return distributions, including the uniform for visual comparison. It is apparent from these figures that normalization has been achieved in all cases examined. Finally, a second noticeable in-sample result is the optimal lag length chosen by the different NoVaS specifications. In particular, we see from Table 16 that the optimal lag length is greater when using squared returns than when using absolute returns. As expected, longer lag lengths are associated with a smaller a_0^* coefficient.

We now turn to the out-of-sample results on the forecasting performance of NoVaS , which are summarized in Tables 17, 18, 19 and 20. The results are slightly different across the series we examine but the overall impression is that the NoVaS-based forecasts are superior to the GARCH forecasts, based on the combined performance of all evaluation measures. We discuss these in turn.

If we look at the MAD results in Table 17 the NoVaS forecasts outperform both the Naive benchmark and the GARCH-based forecasts. Note that the use of squared returns gives better results in the two series with the smallest sample kurtosis (S&P500 and USD/Yen series) while the use of absolute returns gives better results in the two series with the highest kurtosis (MSFT and EFG series). Its also worthwhile to note that the most drastic performance improvement, vis-a-vis the benchmark, can be seen for the MSFT series (smallest sample size) and the EFG series (highest kurtosis).[16] This is important since we expected NoVaS to perform well in both these cases: the small sample size makes inference difficult while high kurtosis can be the result of non-stationarities in the series. Finally, the results are similar if we consider the RMSE ranking in Table 18. Based on these two descriptive evaluation measures the NoVaS forecasts outperform the benchmark and GARCH models.

To examine whether there are statistically significant differences between the NoVaS and GARCH forecasts and the benchmark, we next consider the results from the application of the Diebold-Mariano (1995) test for comparing forecasting performance. Looking at Table 19 we can see that there are statistically significant differences between the NoVaS forecasts and the Naive benchmark for the S&P500 series and the MSFT series, with the NoVaS forecasts being

[16]Note also the performance improvement from the use of the median GARCH vs. the mean GARCH forecasts for the MSFT series. Recall that our simulation results showed that the performance of a GARCH model could be way off the mark if the training sample was small; here we use only 157 observations for training the MSFT series and the GARCH forecasts cannot outperform even the Naive benchmark.

significantly better.[17] For the other two series the test does not indicate a (statistically) superior performance of any of the other models compared to the benchmark.

Our empirical results so far clearly indicate that the NoVaS forecasts offer improvements in forecasting performance, both over the Naive benchmark and the GARCH models. We next discuss the results from the forecasting unbiasedness regressions of eq. (29), where we try to see whether the forecasts are correct 'on average' and whether they make any systematic mistakes. We start by noting that the estimates from a regression like eq. (29) suffer from bias since the regressor used $\widehat{\sigma}_t^2$ is estimated and not measured directly. Therefore we should be interpreting our results with some caution and connect them with our previous discussion. Looking at Table 20 we can see that in many cases the constant term a is estimated to be (numerically close to) zero, although it is statistically significant. The slope parameter b estimates show that there is still bias in the direction of the forecasts, either positive or negative, but the NoVaS estimates of b are in general much lower than those of the benchmark and the GARCH models, with the exception of the MSFT series. Furthermore, for the S&P500 and the EFG series the slope parameter is not statistically significant, at the 10% level, indicating a possibly unbiased NoVaS forecast. The R^2 values from these regressions are also supportive of the NoVaS forecasts (remember that low values are preferred over high values): the corresponding R^2 values from the NoVaS forecasts are lower than both the benchmark and the GARCH values by at least 30%. Note that for the S&P500 series where the value of the R^2 of the benchmark is lower than the corresponding NoVaS value, we also have a (numerically) large value for the slope parameter b for the benchmark compared to NoVaS . The only real problem with the R^2 from these regressions is for the MSFT series which we discuss below in Remark 4. All in all the results from Table 20 support the superior performance of NoVaS against its competitors and show that is a much less biased forecasting procedure.

Remark 4. Can we obtain further improvements using the NoVaS methodology? In particular, how does changes in the value of the α parameter affect the forecasting performance? This is an empirically interesting questions since our results can be affected both by the small sample size and the degree of kurtosis in the data. The MSFT series exhibits both these problems and is thus worthwhile to see whether we can improve our results by allowing the unconditional estimator of the variance to enter the calculations.[18] We repeated our analysis for the MSFT series using $\alpha = 0.5$ and our results improved dramatically. The MAD and RMSE values from the ABNT NoVaS method dropped from 0.551 to 0.360 and from 0.951 to 0.524 respectively,

[17] For the MSFT series the benchmark forecasts are also significantly better than the GARCH forecasts.

[18] Changing the value of α did not result in improvements in the other three series.

with the Diebold-Mariano test still indicating a statistically significant performance over the Naive benchmark. In addition, the results from the forecasting unbiasedness regression are now better than the benchmark for the ABNT NoVaS method: the estimate of the slope parameter b is -0.145 and not statistically significant while the R^2 value is 0.010 compared to 0.012 for the benchmark.

In summary, our results are especially encouraging because they reflect on the very idea of the NoVaS transformation: a model-free approach that can account for different types of potential DGPs, that include breaks, switching regimes and lack of higher moments. NoVaS is successful in overcoming the parametrization and estimation problems that one would encounter in models that have variability and uncertainty not only in their parameters but also in their functional form. Of course our results are specific to the datasets examined and, it is true, we made no attempt to consider other types of parametric volatility models. But this is one of the problems that NoVaS attempts to solve: we have no *a priori* guidance as to which parametric volatility model to choose, be it simple GARCH, exponential GARCH, asymmetric GARCH and so on. With NoVaS we face no such problem as the very concept of a model does not enter into consideration.

5 Concluding Remarks

In this paper we have presented several findings on the NoVaS transformation approach for volatility forecasting introduced by Politis (2003a,b, 2007) and extended in Politis and Thomakos (2007). It was shown that NoVaS can be a flexible method for forecasting volatility of financial returns that is simple to implement, and robust against non-stationarities.

In particular, we focused on a new method for volatility forecasting using NoVaS and conducted an extensive simulation to study its forecasting performance under different DGPs. It was shown that the NoVaS methodology remains successful in situations where (global) stationarity fails such as the cases of local stationarity and/or structural breaks, and invariably outperforms the GARCH benchmark for all non-GARCH DGPs. Remarkably, the NoVaS methodology was found to outperform the GARCH forecasts *even* when the underlying DGP *is* itself a (stationary) GARCH as long as the sample size is only moderately large. It was also found that NoVaS forecasts lead to a much 'tighter' distribution of the forecasting performance measure used (the MAD) for all DGPs considered. Our empirical illustrations using four real datasets are also very supportive of the excellent forecasting performance of NoVaS compared to the standard GARCH forecasts.

Extensions of the current work include, among others, the use of the NoVaS approach on empirical calculations of value at risk (VaR), the generalization to more than one assets and the calculation of NoVaS correlations, and further extensive testing on the out-of-sample forecasting performance of the proposed method. Some of the above are pursued by the authors.

References

[1] Andersen, T.G., Bollerslev, T., Christoffersen, P.F., and F. X. Diebold, 2006. "Volatility and Correlation Forecasting" in G. Elliott, C.W.J. Granger, and Allan Timmermann (eds.), *Handbook of Economic Forecasting*, Amsterdam: North-Holland, pp. 778-878.

[2] Andersen, T.G., Bollerslev, T. and Meddahi, N., 2004. "Analytic evaluation of volatility forecasts, *International Economic Review*, vol. 45, pp. 1079-1110.

[3] Berkes, I. and L. Horvath, 2004. "The efficiency of the estimators of the parameters in GARCH processes", *Annals of Statistics*, 32, pp. 633-655.

[4] Dahlhaus, R. (1997), "Fitting time series models to nonstationary processes", *Annals of Statistics*, 25 pp. 1-37.

[5] Dahlhaus, R. and S. Subba-Rao, 2006. "Statistical Inference for Time-Varying ARCH Processes", *Annals of Statistics*, vol. 34, pp. 1075-1114.

[6] Dahlhaus, R. and S. Subba-Rao, 2007. "A Recursive Online Algorithm for the Estimation of Time Varying ARCH Parameters", *Bernoulli*, vol 13, pp. 389-422.

[7] Diebold, F. X. and R. S. Mariano, 1995. "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, vol. 13, pp. 253.263.

[8] Francq, C. and J-M. Zakoian, 2005. "L2 Structures of Standard and Switching-Regime GARCH Models", *Stochastic Processes and Their Applications*, 115, pp. 1557-1582.

[9] Fryzlewicz, P., Sapatinas, T. and S. Subba-Rao, 2006. "A Haar-Fisz Technique for Locally Stationary Volatility Estimation", *Biometrika*, vol. 93, pp. 687-704.

[10] Fryzlewicz, P., Sapatinas, T. and S. Subba-Rao, 2007. "Normalized Least Squares Estimation in Time-Varying ARCH Models", *Annals of Statistics*, to appear.

[11] Ghysels, E. and L. Forsberg, 2007. "Why Do Absolute Returns Predict Volatility So Well?", *Journal of Financial Econometrics*, vol. 5, pp. 31-67.

[12] Ghysels, E., P. Santa-Clara, and R. Valkanov, 2006. "Predicting Volatility: How to Get Most Out of Returns Data Sampled at Different Frequencies", *Journal of Econometrics*, forthcoming.

[13] Hall, P. and Q. Yao, 2003. "Inference in ARCH and GARCH Models with heavy-tailed errors", *Econometrica*, 71, pp. 285-317

[14] Hansen, B., 2006. "Interval Forecasts and Parameter Uncertainty", *Journal of Econometrics*, vol. 127, pp. 377-398.

[15] Hansen, P. R. and A. Lunde, 2006. "Consistent ranking of volatility models", *Journal of Econometrics*, 131, pp. 97-121.

[16] 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*, vol. 65, pp. 839-861.

[17] Hillebrand, E. 2005. "Neglecting Parameter Changes in GARCH Models", *Journal of Econometrics*, 129, pp. 121-138.

[18] Lunde, A. and P. R. Hansen, 2005. "A forecast comparison of volatility models: does anything beat a GARCH(1,1)?", *Journal of Applied Econometrics*, 20(7), pp. 873-889.

[19] Meddahi, N., 2001. "An eigenfunction approach for volatility modeling", Technical report, CIRANO Working paper 2001s-70, Univiversity of Montreal.

[20] Mikosch, T. and C. Starica, 2000. "Change of Structure in Financial Time Series, Long Range Dependence and the GARCH model", CAF Working Paper Series, No. 58.

[21] Parkinson, M., 1980. "The Extreme Value Method for Estimating the Variance of the Rate of Return", *Journal of Business*, 53, pp.6168.

[22] Patton, A., 2005. "Volatility forecast evaluation and comparison using imperfect volatility proxies", mimeo.

[23] Peng, L. and Q. Yao, 2003. "Least absolute deviations estimation for ARCH and GARCH models", *Biometrika*, 90, pp. 967-975.

[24] Politis, D.N., 2003a. "Model-Free Volatility Prediction", UCSD Dept. of Economics Discussion Paper 2003-16.

[25] Politis, D.N., 2003b. "A Normalizing and Variance-Stabilizing Transformation for Financial Time Series, in *Recent Advances and Trends in Nonparametric Statistics*, M.G. Akritas and D.N. Politis, (Eds.), Elsevier: North Holland, pp. 335-347.

[26] Politis, D.N., 2004. "A heavy-tailed distribution for ARCH residuals with application to volatility prediction", *Annals of Economics and Finance*, vol. 5, pp. 283-298.

[27] Politis, D.N., 2007. "Model-free vs. model-based volatility prediction", *J. Financial Econometrics*, vol. 5, pp. 358-389.

[28] Politis, D. and D. Thomakos, 2007. "Financial Time Series and Volatility Prediction using NoVaS Transformations", forthcoming in *Forecasting in the Presence of Parameter Uncertainty and Structural Breaks*, D. E. Rapach and M. E. Wohar (Eds.), Emerald Group Publishing Ltd.

[29] Poon, S. and C. Granger, 2003. "Forecasting Volatility in Financial Markets: A Review", *Journal of Economic Literature*, 41, pp. 478539.

[30] Taylor, J., 2004. "Volatility Forecasting using Smooth Transition Exponential Smoothing", *International Journal of Forecasting*, vol. 20, pp. 273-286.

[31] Wolfowitz, A., 1957. "The Minimum Distance Method", *Annals of Mathematical Statistics*, 28, pp. 75-88.

Tables

Table 1. Simulation Results for GARCH, $T_1 = 1,000$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.24	0.33	0.06	0.09	0.11	0.15	0.24	0.45	4.80
SQNT	0.14	0.08	0.08	0.09	0.10	0.11	0.14	0.19	1.08
ABNT	0.21	0.09	0.15	0.16	0.17	0.19	0.22	0.28	1.28
GARCH	2.64	13.43	0.07	0.10	0.16	0.34	1.00	3.53	169.78
M-GARCH	1.56	7.39	0.13	0.16	0.18	0.29	0.66	2.04	93.41

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.93	0.66	0.93
M-GARCH	1.00	0.74	1.00

Notes:

1. The model being simulated is a standard GARCH(1,1) $h_t^2 = \omega + \alpha h_{t-1}^2 + \beta(X_{t-1} - \mu)^2$.
2. $T_1 = 1,000$ denotes the number of forecasts generated for computing the mean absolute deviation (MAD) in each replication.
3. The first table presents distributional statistics of the MAD of the forecast errors over 500 replications (all entries are $\times 1,000$.) The second table presents the proportion of times that the NoVaS MAD relative to the naïve benchmark was smaller than the GARCH MAD relative to the same benchmark, see eq. (27) in the main text.
4. $\bar{x}_f$ denotes the sample mean, $\widehat{\sigma}_f$ denotes the sample std. deviation and Q_p denotes the p^{th} sample quantile of the MAD distribution over 500 replications.
5. Naïve denotes forecasts based on the rolling sample variance, SQNT (ABNT) denotes NoVaS forecasts based on a normal target distribution and squared (absolute) returns, GARCH and M-GARCH denote L_2 and L_1 based forecasts from a standard GARCH model.

Table 2. Simulation Results for B-GARCH, $T_1 = 1,000$

Distributional Statistics for MAD

	$\bar{x}$	$\widehat{\sigma}$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.43	0.96	0.09	0.13	0.16	0.22	0.33	0.71	14.74
SQNT	0.17	0.47	0.09	0.10	0.11	0.12	0.15	0.21	9.27
ABNT	0.28	0.47	0.14	0.17	0.18	0.20	0.25	0.36	8.02
GARCH	29.10	385.48	0.09	0.15	0.21	0.50	1.54	4.19	7236.79
M-GARCH	16.15	212.13	0.13	0.18	0.23	0.40	0.97	2.51	3981.49

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.98	0.76	0.98
M-GARCH	0.99	0.87	0.99

Notes:

1. The model being simulated is a standard GARCH(1,1) with parameter breaks $h_t^2 = \omega_b + \alpha_b h_{t-1}^2 + \beta_b (X_{t-1} - \mu)^2$, $b = 1, 2, \ldots, B$.
2. See other notes in Table 1.

Table 3. Simulation Results for TV-GARCH, $T_1 = 1,000$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.31	0.53	0.10	0.12	0.15	0.19	0.27	0.51	5.95
SQNT	0.14	0.23	0.05	0.06	0.07	0.09	0.12	0.19	2.38
ABNT	0.15	0.16	0.08	0.09	0.10	0.11	0.14	0.20	1.67
GARCH	1.70	14.11	0.07	0.10	0.12	0.20	0.47	1.51	224.31
M-GARCH	1.02	7.78	0.08	0.11	0.12	0.17	0.36	0.91	123.71

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.98	0.85	1.00
M-GARCH	0.99	0.98	1.00

Notes:

1. The model being simulated is a GARCH(1,1) with slowly varying varying parameters $h_t^2 = \omega(t) + \alpha(t)h_{t-1}^2 + \beta(t)(X_{t-1} - \mu)^2$.
2. See other notes in Table 1.

Table 4a. Simulation Results for MS-GARCH, $T_1 = 1,000$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.36	0.42	0.11	0.15	0.18	0.24	0.40	0.61	4.97
SQNT	0.20	0.12	0.13	0.14	0.15	0.17	0.20	0.26	2.30
ABNT	0.30	0.14	0.21	0.23	0.25	0.27	0.31	0.37	1.95
GARCH	1.33	3.04	0.11	0.17	0.22	0.41	1.07	2.88	34.44
M-GARCH	0.88	1.68	0.20	0.24	0.26	0.37	0.73	1.79	19.03

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.94	0.62	0.94
M-GARCH	1.00	0.73	1.00

Notes:

1. The model being simulated is a two-state Markov switching GARCH(1,1) $h_t^2 = \sum_{s=1}^{2} \mathbf{1}\{\mathsf{P}(S_t = s)\}\left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2\right]$.
2. The transition probabilities are $p_{11} = 0.9$, $p_{12} = 0.1$, $p_{21} = 0.3$, $p_{22} = 0.7$.
3. See other notes in Table 1.

Table 4b. Simulation Results for MS-GARCH, $T_1 = 1,000$

Descriptive Statistics - all MAD's

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.48	2.34	0.10	0.13	0.16	0.23	0.37	0.62	51.10
SQNT	0.18	0.15	0.11	0.12	0.13	0.15	0.18	0.24	1.90
ABNT	0.26	0.15	0.17	0.19	0.20	0.23	0.26	0.33	1.95
GARCH	3.21	23.07	0.09	0.13	0.18	0.31	0.81	2.83	426.85
M-GARCH	1.91	12.71	0.16	0.19	0.22	0.30	0.60	1.69	235.17

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.90	0.60	0.90
M-GARCH	1.00	0.75	1.00

Notes:

1. The model being simulated is a two-state Markov switching GARCH(1,1) $h_t^2 = \sum_{s=1}^{2} \mathbf{1}\{\mathsf{P}(S_t = s)\}\left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2\right]$.
2. The transition probabilities are $p_{11} = 0.9$, $p_{12} = 0.1$, $p_{21} = 0.1$, $p_{22} = 0.9$.
3. See other notes in Table 1.

Table 5. Simulation Results for ST-GARCH, $T_1 = 1,000$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.32	0.34	0.09	0.12	0.15	0.21	0.33	0.62	3.32
SQNT	0.15	0.07	0.10	0.11	0.11	0.13	0.15	0.20	0.80
ABNT	0.24	0.10	0.17	0.19	0.20	0.22	0.25	0.32	0.98
GARCH	2.05	10.15	0.08	0.12	0.16	0.26	0.88	2.53	119.51
M-GARCH	1.25	5.60	0.15	0.18	0.21	0.26	0.62	1.53	66.29

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.91	0.55	0.91
M-GARCH	1.00	0.67	1.00

Notes:

1. The model being simulated is a smooth transition GARCH(1,1)
   $h_t^2 = \sum_{s=1}^{2} Q_s(X_{t-1}) \left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2\right]$.
2. The transition function is $Q_1(\cdot) + Q_2(\cdot) = 1$ and $Q_s = \left[1 + \exp(-\gamma_1 X_{t-1}^{\gamma_2})\right]^{-1}$.
3. See other notes in Table 1.

Table 6. Simulation Results for D-GARCH, $T_1 = 1,000$

Descriptive Statistics - all MAD's

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.16	0.17	0.07	0.08	0.09	0.10	0.15	0.28	1.69
SQNT	0.12	0.04	0.09	0.10	0.10	0.10	0.12	0.15	0.44
ABNT	0.18	0.05	0.14	0.15	0.16	0.16	0.18	0.22	0.62
GARCH	1.62	9.01	0.06	0.09	0.11	0.21	0.67	1.78	112.29
M-GARCH	0.98	4.96	0.12	0.14	0.15	0.20	0.46	1.13	61.85

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.76	0.55	0.76
M-GARCH	1.00	0.61	1.00

Notes:

1. The model being simulated is a GARCH(1,1) with an added deterministic function with a returns' equation given by $X_t = \mu + [a - b(t/T)]\, h_t Z_t$ and with a standard GARCH volatility function.
2. See other notes in Table 1.

Table 7. Simulation Results for SV-GARCH, $T_1 = 1,000$

Descriptive Statistics - all MAD's

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.26	0.16	0.14	0.17	0.20	0.23	0.27	0.32	3.05
SQNT	0.21	0.13	0.12	0.15	0.17	0.19	0.23	0.26	2.64
ABNT	0.23	0.11	0.13	0.18	0.20	0.22	0.25	0.28	2.29
GARCH	1.50	8.74	0.13	0.18	0.22	0.33	0.91	2.71	189.82
M-GARCH	0.95	4.81	0.17	0.22	0.25	0.32	0.64	1.62	104.58

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.90	0.70	0.91
M-GARCH	0.97	0.99	1.00

Notes:

1. The model being simulated is a stochastic volatility model $\log h_t^2 = \omega + \alpha \log h_{t-1}^2 + w_t$, $w_t \sim \mathcal{N}(0, \sigma_w^2)$.
2. See other notes in Table 1.

Table 8. Simulation Results for GARCH, $T_1 = 350$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.26	0.39	0.03	0.07	0.09	0.14	0.24	0.48	3.90
SQNT	0.14	0.13	0.08	0.09	0.10	0.11	0.13	0.20	1.67
ABNT	0.21	0.13	0.14	0.15	0.16	0.18	0.21	0.29	1.42
GARCH	0.22	0.75	0.01	0.04	0.06	0.10	0.16	0.35	13.01
M-GARCH	0.24	0.49	0.02	0.09	0.13	0.17	0.22	0.33	8.52

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.43	0.13	0.43
M-GARCH	0.86	0.35	0.86

Notes:

1. The model being simulated is a standard GARCH(1,1) $h_t^2 = \omega + \alpha h_{t-1}^2 + \beta(X_{t-1} - \mu)^2$.
2. $T_1 = 350$ denotes the number of forecasts generated for computing the mean absolute deviation (MAD) in each replication.
3. See other notes in Table 1.

Table 9. Simulation Results for B-GARCH, $T_1 = 350$

Distributional Statistics for MAD

	$\bar{x}$	$\widehat{\sigma}$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.39	0.87	0.08	0.12	0.15	0.21	0.34	0.56	9.95
SQNT	0.10	0.09	0.07	0.07	0.08	0.08	0.10	0.13	0.95
ABNT	0.22	0.32	0.10	0.12	0.13	0.15	0.21	0.28	3.69
GARCH	0.65	4.99	0.05	0.07	0.10	0.13	0.23	0.37	61.51
M-GARCH	0.47	2.75	0.06	0.09	0.11	0.15	0.23	0.34	33.78

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.86	0.35	0.86
M-GARCH	0.96	0.42	0.96

Notes:

1. The model being simulated is a standard GARCH(1,1) with parameter breaks $h_t^2 = \omega_b + \alpha_b h_{t-1}^2 + \beta_b (X_{t-1} - \mu)^2$, $b = 1, 2, \ldots, B$.
2. See other notes in Table 8.

Table 10. Simulation Results for TV-GARCH, $T_1 = 350$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.31	0.58	0.10	0.13	0.15	0.19	0.27	0.49	10.13
SQNT	0.13	0.30	0.05	0.06	0.07	0.08	0.11	0.19	5.87
ABNT	0.15	0.19	0.08	0.09	0.10	0.11	0.14	0.20	3.36
GARCH	0.20	0.37	0.06	0.08	0.09	0.12	0.17	0.28	5.97
M-GARCH	0.20	0.38	0.08	0.10	0.11	0.13	0.17	0.29	7.20

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.89	0.52	0.98
M-GARCH	0.96	0.91	0.99

Notes:

1. The model being simulated is a GARCH(1,1) with slowly varying varying parameters $h_t^2 = \omega(t) + \alpha(t)h_{t-1}^2 + \beta(t)(X_{t-1} - \mu)^2$.
2. See other notes in Table 8.

Table 11a. Simulation Results for MS-GARCH, $T_1 = 350$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.37	0.70	0.06	0.11	0.16	0.22	0.35	0.64	9.53
SQNT	0.20	0.16	0.11	0.13	0.14	0.16	0.19	0.27	2.42
ABNT	0.32	0.33	0.18	0.21	0.23	0.25	0.30	0.40	5.09
GARCH	2.70	42.77	0.04	0.08	0.11	0.15	0.24	0.45	918.41
M-GARCH	1.65	23.68	0.07	0.14	0.18	0.23	0.31	0.47	508.48

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.42	0.14	0.42
M-GARCH	0.85	0.30	0.86

Notes:

1. The model being simulated is a two-state Markov switching GARCH(1,1) $h_t^2 = \sum_{s=1}^{2} \mathbf{1}\{\mathsf{P}(S_t = s)\}\left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2\right]$.
2. The transition probabilities are $p_{11} = 0.9$, $p_{12} = 0.1$, $p_{21} = 0.3$, $p_{22} = 0.7$.
3. See other notes in Table 8.

Table 11b. Simulation Results for MS-GARCH, $T_1 = 350$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.47	1.95	0.06	0.11	0.14	0.20	0.35	0.67	40.34
SQNT	0.20	0.30	0.10	0.11	0.12	0.14	0.17	0.27	4.85
ABNT	0.27	0.26	0.16	0.18	0.19	0.21	0.25	0.37	4.27
GARCH	5.56	84.17	0.05	0.07	0.10	0.13	0.22	0.42	1591.98
M-GARCH	3.21	46.39	0.06	0.12	0.15	0.19	0.27	0.46	877.21

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.45	0.18	0.46
M-GARCH	0.87	0.36	0.89

Notes:

1. The model being simulated is a two-state Markov switching GARCH(1,1) $h_t^2 = \sum_{s=1}^{2} \mathbf{1}\{\mathsf{P}(S_t = s)\}\left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2\right]$.
2. The transition probabilities are $p_{11} = 0.9$, $p_{12} = 0.1$, $p_{21} = 0.1$, $p_{22} = 0.9$.
3. See other notes in Table 8.

Table 12. Simulation Results for ST-GARCH, $T_1 = 350$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.31	0.42	0.04	0.10	0.13	0.20	0.32	0.56	4.11
SQNT	0.15	0.12	0.09	0.10	0.11	0.12	0.14	0.21	1.67
ABNT	0.25	0.17	0.16	0.17	0.19	0.21	0.24	0.30	1.84
GARCH	0.19	0.31	0.03	0.06	0.08	0.12	0.19	0.34	4.26
M-GARCH	0.24	0.27	0.04	0.12	0.15	0.19	0.25	0.34	2.73

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.47	0.14	0.47
M-GARCH	0.91	0.31	0.92

Notes:

1. The model being simulated is a smooth transition GARCH(1,1)
   $h_t^2 = \sum_{s=1}^{2} Q_s(X_{t-1}) \left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2\right]$.
2. The transition function is $Q_1(\cdot) + Q_2(\cdot) = 1$ and $Q_s = \left[1 + \exp(-\gamma_1 X_{t-1}^{\gamma_2})\right]^{-1}$.
3. See other notes in Table 8.

Table 13. Simulation Results for D-GARCH, $T_1 = 350$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.13	0.19	0.02	0.04	0.06	0.08	0.13	0.24	2.29
SQNT	0.11	0.05	0.09	0.10	0.10	0.10	0.11	0.13	0.62
ABNT	0.17	0.06	0.13	0.14	0.14	0.15	0.17	0.20	0.81
GARCH	0.12	0.22	0.02	0.04	0.05	0.07	0.10	0.18	3.06
M-GARCH	0.15	0.14	0.03	0.08	0.11	0.13	0.16	0.20	1.85

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.24	0.09	0.24
M-GARCH	0.77	0.19	0.77

Notes:

1. The model being simulated is a GARCH(1,1) with an added deterministic function with a returns' equation given by $X_t = \mu + [a - b(t/T)]\, h_t Z_t$ and with a standard GARCH volatility function.
2. See other notes in Table 8.

Table 14. Simulation Results for SV-GARCH, $T_1 = 350$

Distributional Statistics for MAD

	$\bar{x}_f$	$\widehat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
Naive	0.26	0.33	0.11	0.16	0.19	0.22	0.27	0.34	7.10
SQNT	0.22	0.36	0.08	0.13	0.15	0.19	0.23	0.28	7.98
ABNT	0.24	0.28	0.10	0.16	0.18	0.21	0.25	0.30	6.04
GARCH	0.24	0.98	0.09	0.13	0.15	0.18	0.22	0.26	22.10
M-GARCH	0.27	0.58	0.09	0.16	0.18	0.23	0.29	0.34	13.02

% of times that the NoVaS MAD was better than the GARCH MAD

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}$
GARCH	0.36	0.17	0.40
M-GARCH	0.84	0.73	0.91

Notes:

1. The model being simulated is a stochastic volatility model $\log h_t^2 = \omega + \alpha \log h_{t-1}^2 + w_t$, $w_t \sim \mathcal{N}(0, \sigma_w^2)$.
2. See other notes in Table 8.

Table 15. Descriptive Statistics for Empirical Series

Series	n	$\bar{x}$	$\widehat{\sigma}$	$\mathcal{S}$	$\mathcal{K}$	$\mathcal{N}$	$\widehat{r}(1)$
S&P500, monthly	448	1.01%	4.35%	-0.37	5.04	0.00	0.00
MSFT, monthly	257	0.00%	1.53%	-1.75	9.00	0.00	-0.10
USD/Yen, daily	2236	-0.00%	0.72%	-0.70	8.52	0.00	0.00
EFG, daily	1403	-0.07%	2.11%	-1.24	24.32	0.00	0.14

Notes:

1. n denotes the number of observations, $\bar{x}$ denotes the sample mean, $\widehat{\sigma}$ denotes the sample standard deviation, $\mathcal{S}$ denotes the sample skewness, $\mathcal{K}$ denotes the sample kurtosis.
2. $\mathcal{N}$ is the p-value of the Cramer-Von Misses test for normality of the underlying series.
3. $\widehat{r}(1)$ denotes the estimate of the first order serial correlation coefficient.

Table 16. Full-sample NoVaS Summary Measures

Type	b^*	$D_n(\boldsymbol{\theta}^*)$	a_0^*	p^*	QQ_X	QQ_W
S&P500 monthly						
SQNT	0.039	0.000	0.052	34	0.989	0.996
ABNT	0.070	0.000	0.078	27	0.989	0.996
MSFT monthly						
SQNT	0.175	0.000	0.171	15	0.916	0.988
ABNT	0.251	0.000	0.231	12	0.916	0.986
USD/Yen daily						
SQNT	0.062	0.000	0.071	29	0.978	0.999
ABNT	0.121	0.000	0.124	20	0.978	0.999
EFG daily						
SQNT	0.089	0.007	0.096	24	0.943	0.999
ABNT	0.171	0.000	0.166	16	0.943	0.999

Notes:

1. SQNT, ABNT denote NoVaS made forecasts based on square and absolute returns and a normal target distribution.
2. b^*, a_0^* and p^* denote the optimal exponential constant, first coefficient and implied lag length.
3. $D_n(\boldsymbol{\theta}^*)$ is the value of the objective function based on kurtosis matching.
4. QQ_X and QQ_W denote the QQ correlation coefficient of the original series and the transformed series respectively.

Table 17. Mean Absolute Deviation (MAD)of Forecast Errors

Series	Naïve	SQNT	ABNT	Mean GARCH	Median GARCH
S&P500, monthly	0.152	0.118	0.134	0.139	0.157
MSFT, monthly	1.883	1.030	0.551	43.28	23.67
USD/Yen, daily	0.026	0.016	0.018	0.022	0.016
EFG, daily	0.251	0.143	0.120	0.225	0.141

Table 18. Root Mean-Squared (RMSE)of Forecast Errors

Series	Naïve	SQNT	ABNT	Mean GARCH	Median GARCH
S&P500, monthly	0.243	0.206	0.206	0.224	0.232
MSFT, monthly	0.530	1.552	0.951	162.0	89.17
USD/Yen, daily	0.031	0.028	0.028	0.030	0.029
EFG, daily	0.227	0.208	0.194	0.211	0.212

Notes:

1. All forecasts computed using a rolling evaluation sample.
2. The evaluation sample used for computing the entries of the tables is as follows: 148 observations for the monthly S&P500 series, 100 observations for the monthly MSFT series, 986 observations for the daily USD/Yen series and 503 observations for the daily EFG series.
3. Table entries are the values of the evaluation measure (MAD for Table 18 and RMSE for Table 19) multiplied by 100 (S&P500 and MSFT monthly series) and by 1000 (USD/Yen and EFG daily series) respectively.
4. SQNT, ABNT denote NoVaS made forecasts based on square and absolute returns and normal target distribution.
5. Mean and median GARCH forecasts denote forecasts made with a GARCH model and an underlying t error distribution with degrees of freedom estimated from the data.
6. The Naive forecast is based on the rolling sample variance.

Table 19. Diebold-Mariano Test for Difference in Forecasting Performance NoVaS and GARCH against the Naive benchmark

Series	SQNT	ABNT	Mean GARCH	Median GARCH
S&P500, monthly				
Test value	3.369	1.762	1.282	-0.414
p-value	0.000	0.078	0.200	0.679
MSFT, monthly				
Test value	2.931	7.022	-2.671	-2.559
p-value	0.003	0.000	0.007	0.010
USD/Yen, daily				
Test value	0.101	0.083	0.037	0.096
p-value	0.919	0.933	0.971	0.924
EFG, daily				
Test value	1.077	1.301	0.259	1.095
p-value	0.281	0.190	0.795	0.274

Notes:

1. See Tables 17 and 18 for column nomenclature.
2. The entries of Table 19 are the test and p-values for the Diebold-Mariano (1995) test for comparing forecasting accuracy. The tests use the absolute value function for the calculation of the statistic and are expressed relative to the Naive benchmark.
3. Positive values indicate that the competing model is superior, negative values that the Naive benchmark is superior.

Table 20. Forecast Unbiasedness Regressions

Series	Naïve	SQNT	ABNT	Mean GARCH	Median GARCH
S&P500, monthly					
Estimates	(-0.003,1.824)	(0.000,0.317)	(0.000,0.879)	(-0.002,1.685)	(-0.002,3.879)
p-values	(0.597,0.540)	(0.527,0.055)	(0.344,0.000)	(0.000,0.000)	(0.000,0.000)
R^2	0.003	0.025	0.111	0.118	0.177
MSFT, monthly					
Estimates	(-0.025,0.242)	(0.004,-0.859)	(0.004,-0.729)	(0.007,-1.000)	(0.007,-1.000)
p-values	(0.000,0.276)	(0.000,0.000)	(0.000,0.000)	(0.000,0.000)	(0.000,0.000)
R^2	0.012	0.871	0.689	1.000	1.000
USD/Yen, daily					
Estimates	(0.000,-1.099)	(0.000,-0.476)	(0.000,0.355)	(0.000,-0.803)	(0.000,0.642)
p-values	(0.000,0.000)	(0.000,0.000)	(0.000,0.000)	(0.000,0.000)	(0.000,0.000)
R^2	0.188	0.055	0.017	0.136	0.029
EFG, daily					
Estimates	(0.000,-0.767)	(0.000,-0.378)	(0.000,0.058)	(0.000,0.138)	(0.000,0.567)
p-values	(0.017,0.000)	(0.000,0.000)	(0.000,0.518)	(0.038,0.318)	(0.038,0.025)
R^2	0.072	0.062	0.001	0.002	0.002

Notes:

1. See Tables 17 and 18 for column nomenclature.
2. The entries of Table 20 are the coefficient estimates $(\widehat{a}, \widehat{b})$ (first line), corresponding p-values (second line) and R^2 (third line) from the forecast unbiasedness regression $e_t = a + b\widehat{\sigma}_t^2 + \zeta_t$.
3. Under the hypothesis of forecast unbiasedness we must have $a = b = 0$ and $R^2 \to 0$. For any two competing models A and B for which we have that $R_A^2 < R_B^2$ we say that model A is superior to model B.

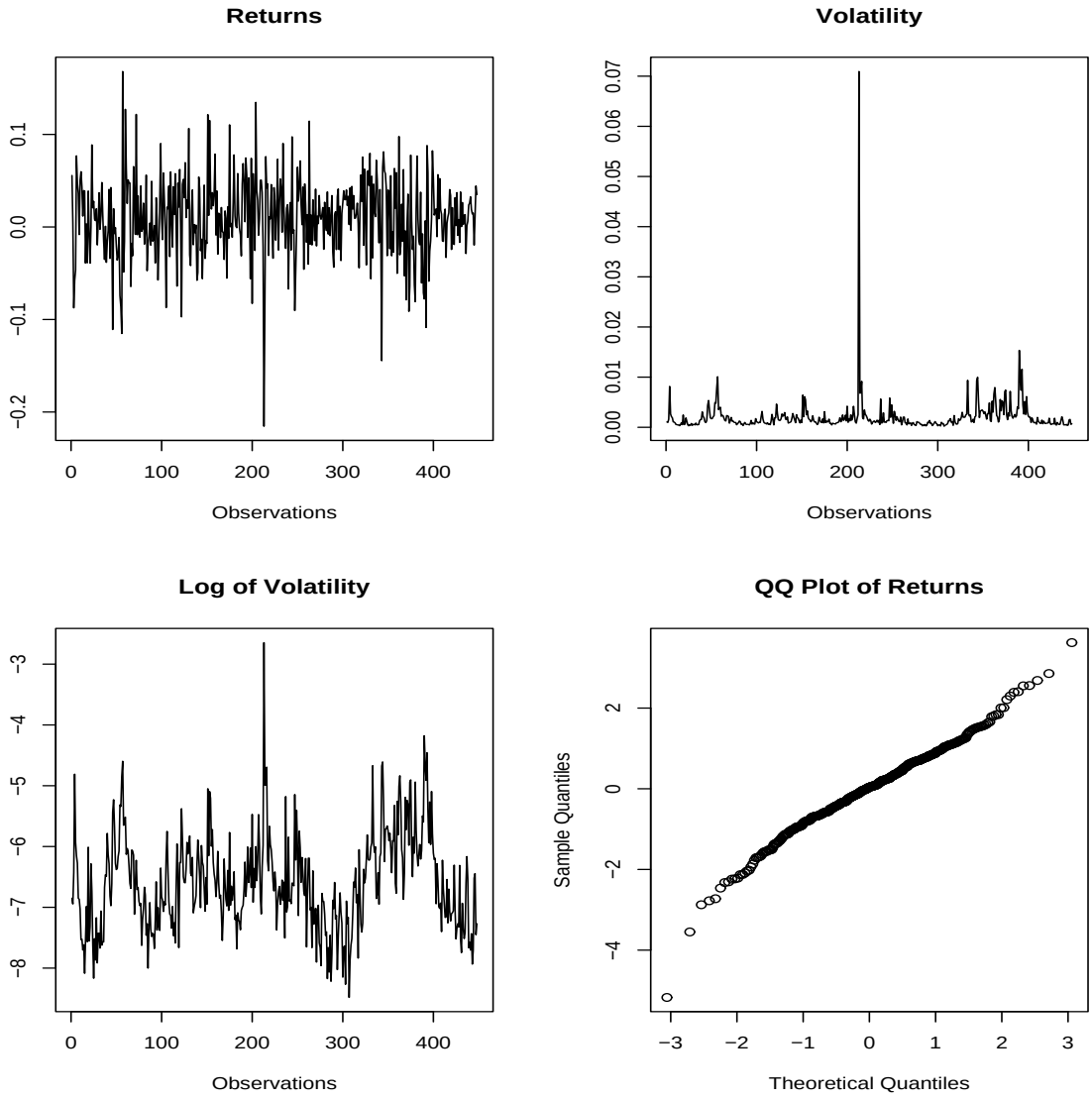

Figure 1: Return, volatility and QQ plots for the monthly S&P500 series

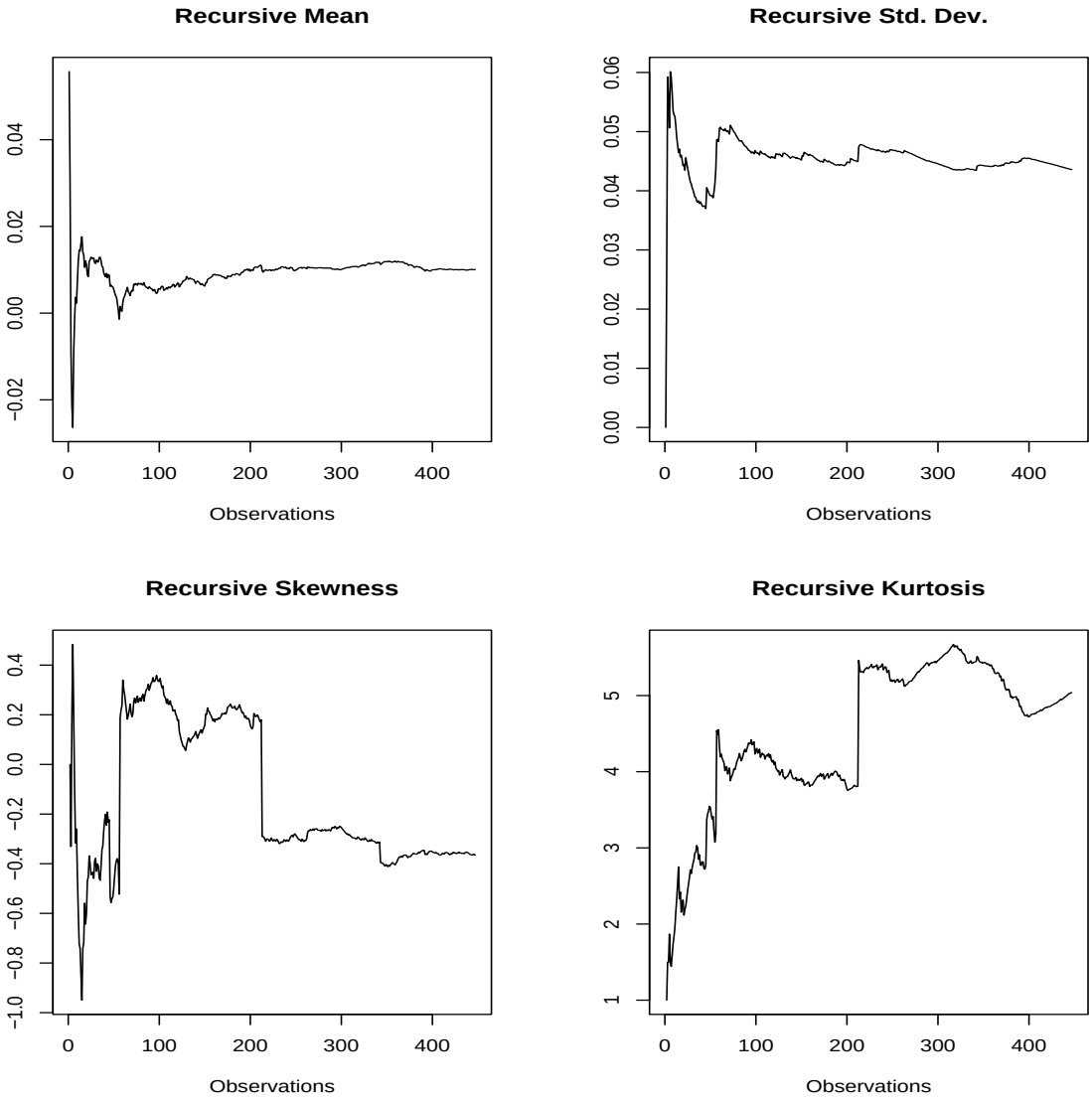

Figure 2: Recursive moments for the monthly S&P500 series

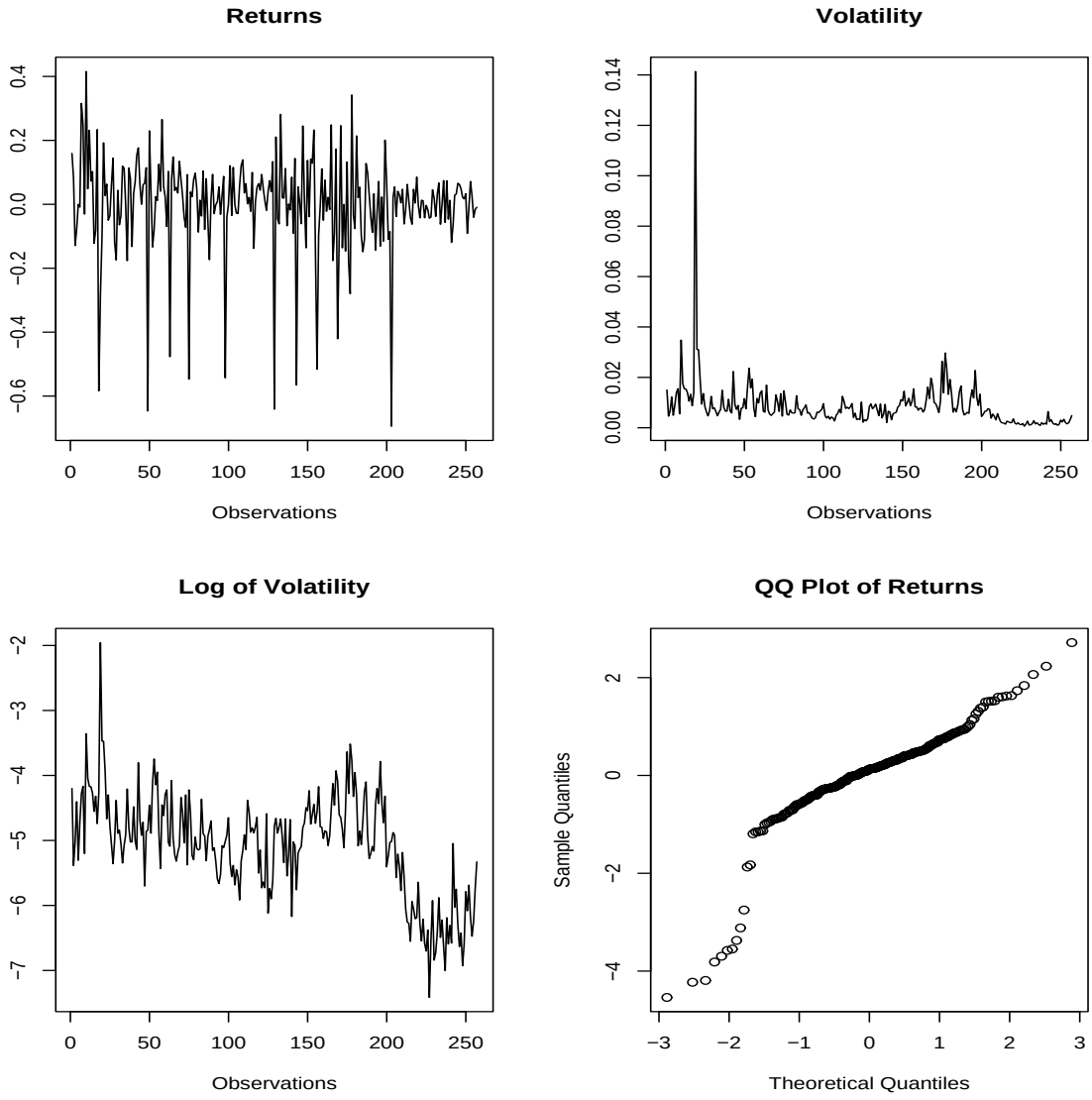

Figure 3: Return, volatility and QQ plots for the monthly MSFT series

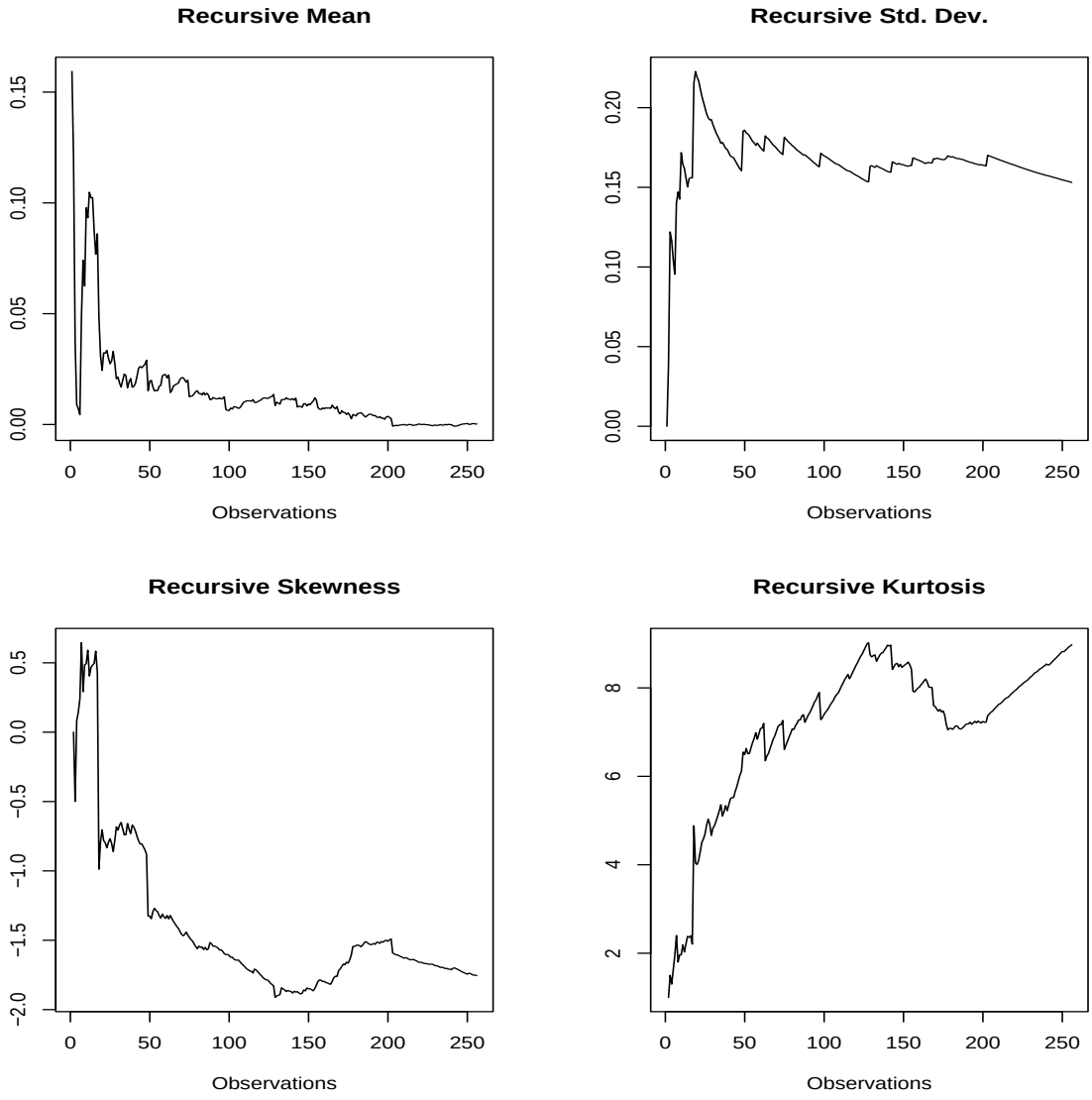

Figure 4: Recursive moments for the monthly MSFT series

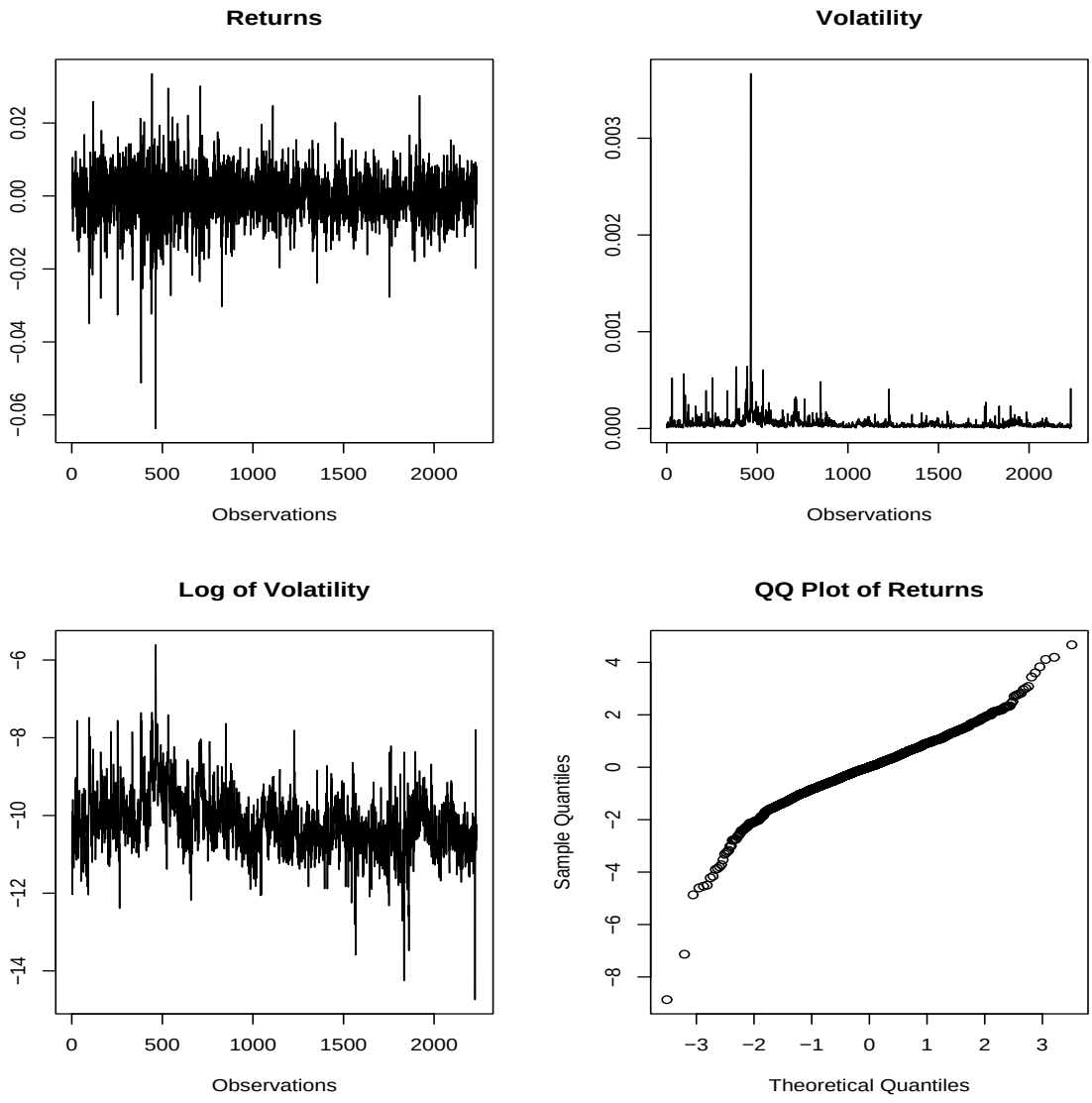

Figure 5: Return, volatility and QQ plots for the daily USD/Yen series

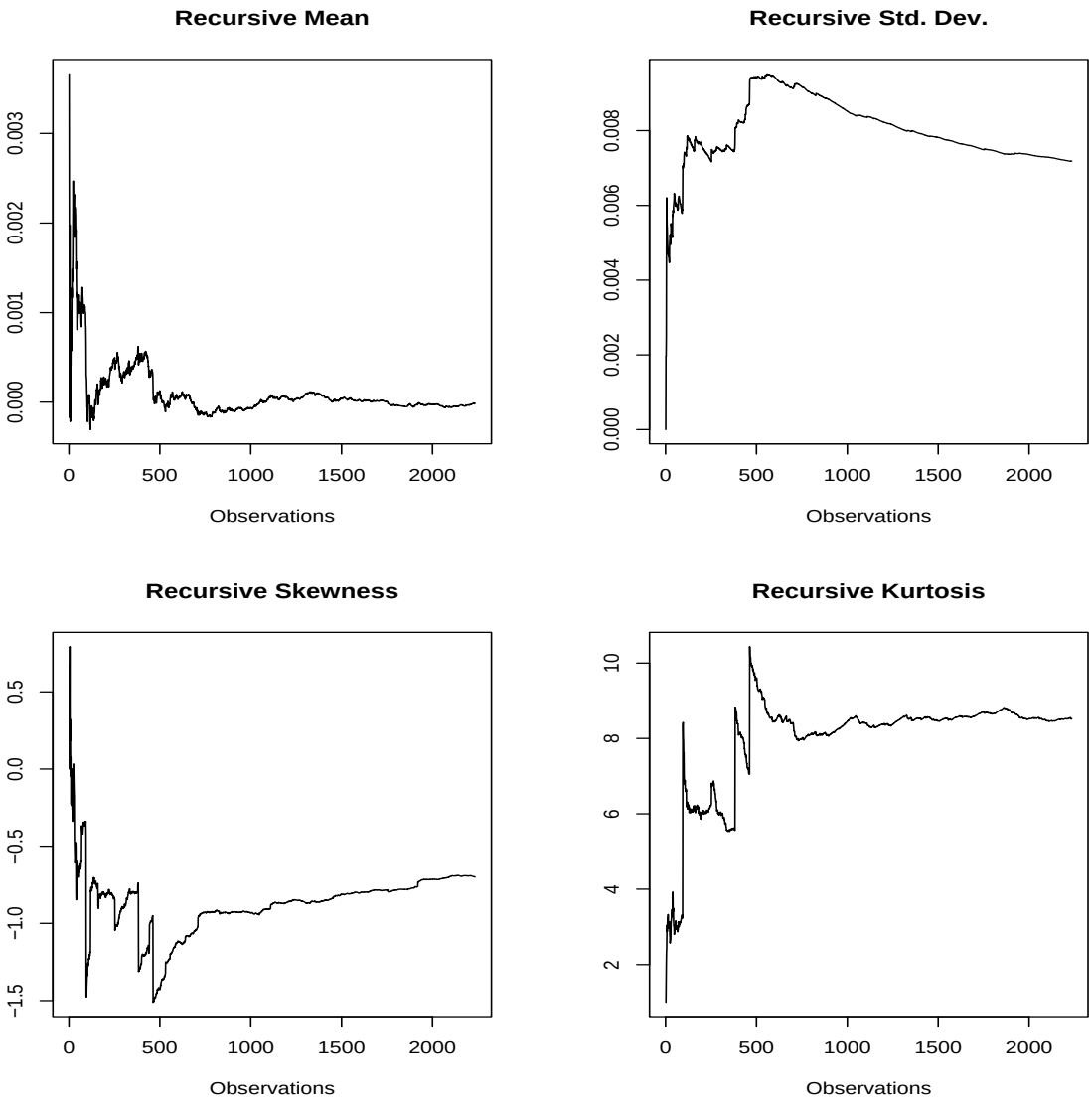

Figure 6: Recursive moments for the daily USD/Yen series

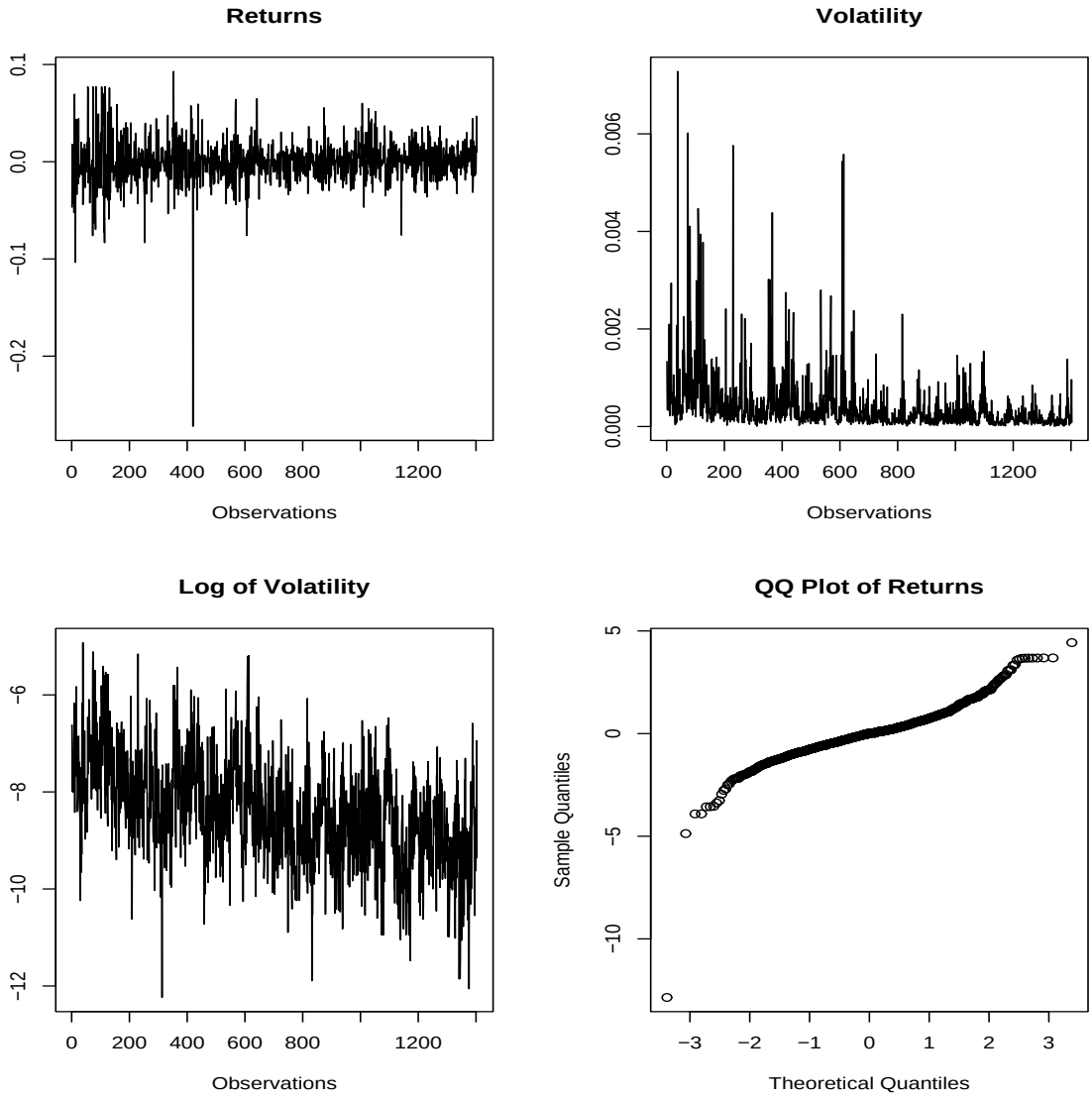

Figure 7: Return, volatility and QQ plots for the daily EFG series

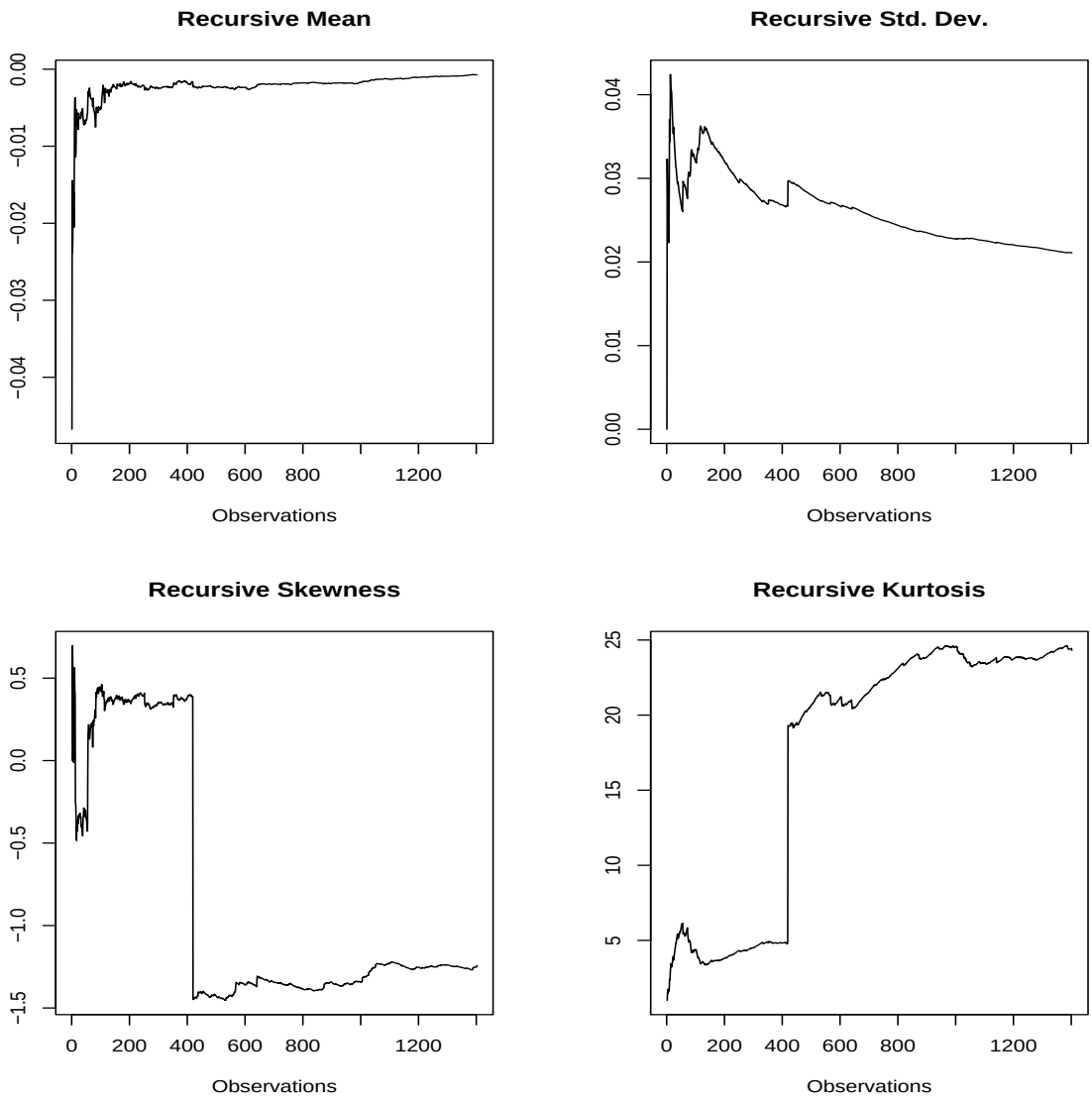

Figure 8: Recursive moments for the daily EFG series

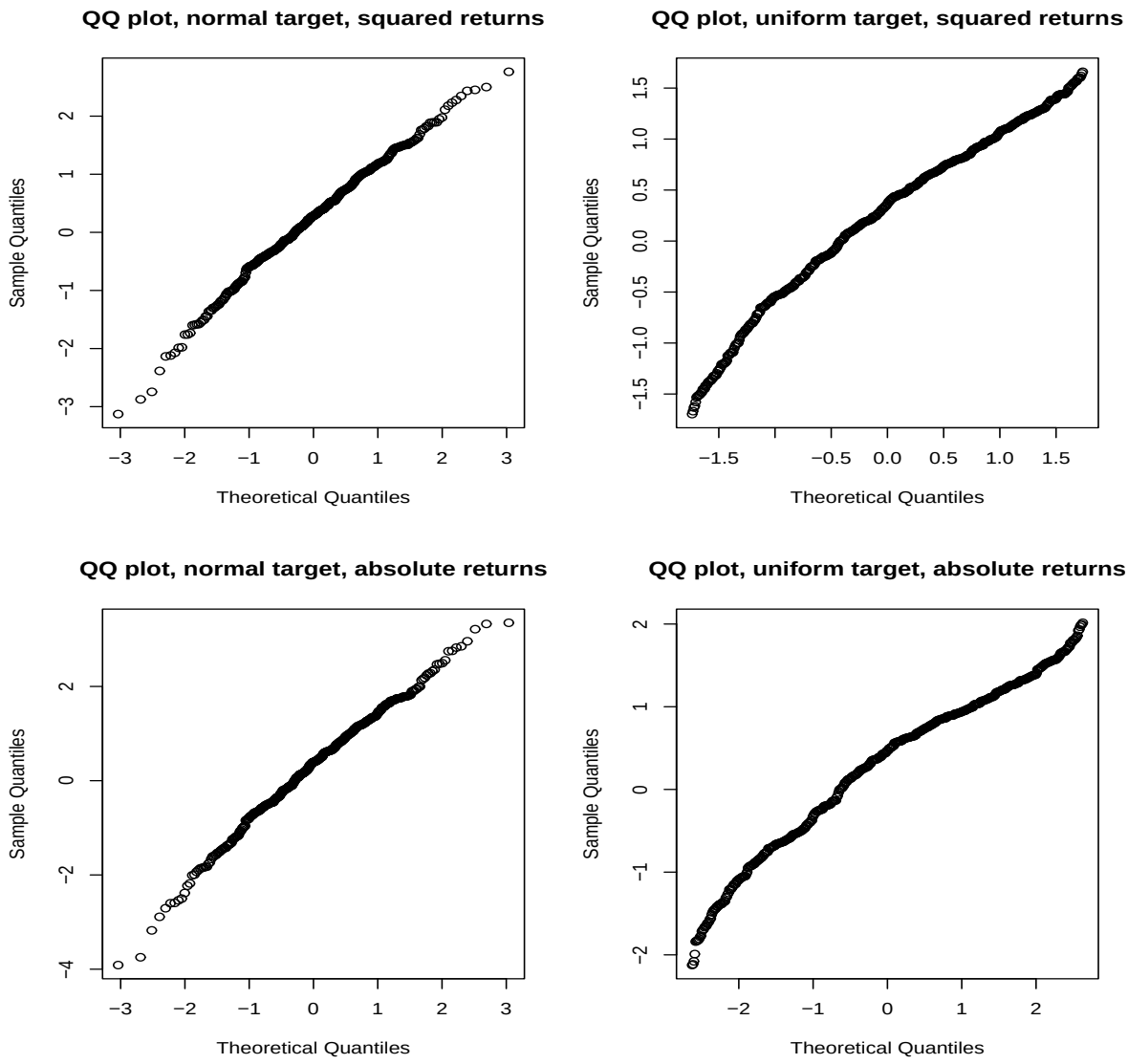

Figure 9: QQ plots of the NoVaS -transformed W series for the monthly S&P500 series

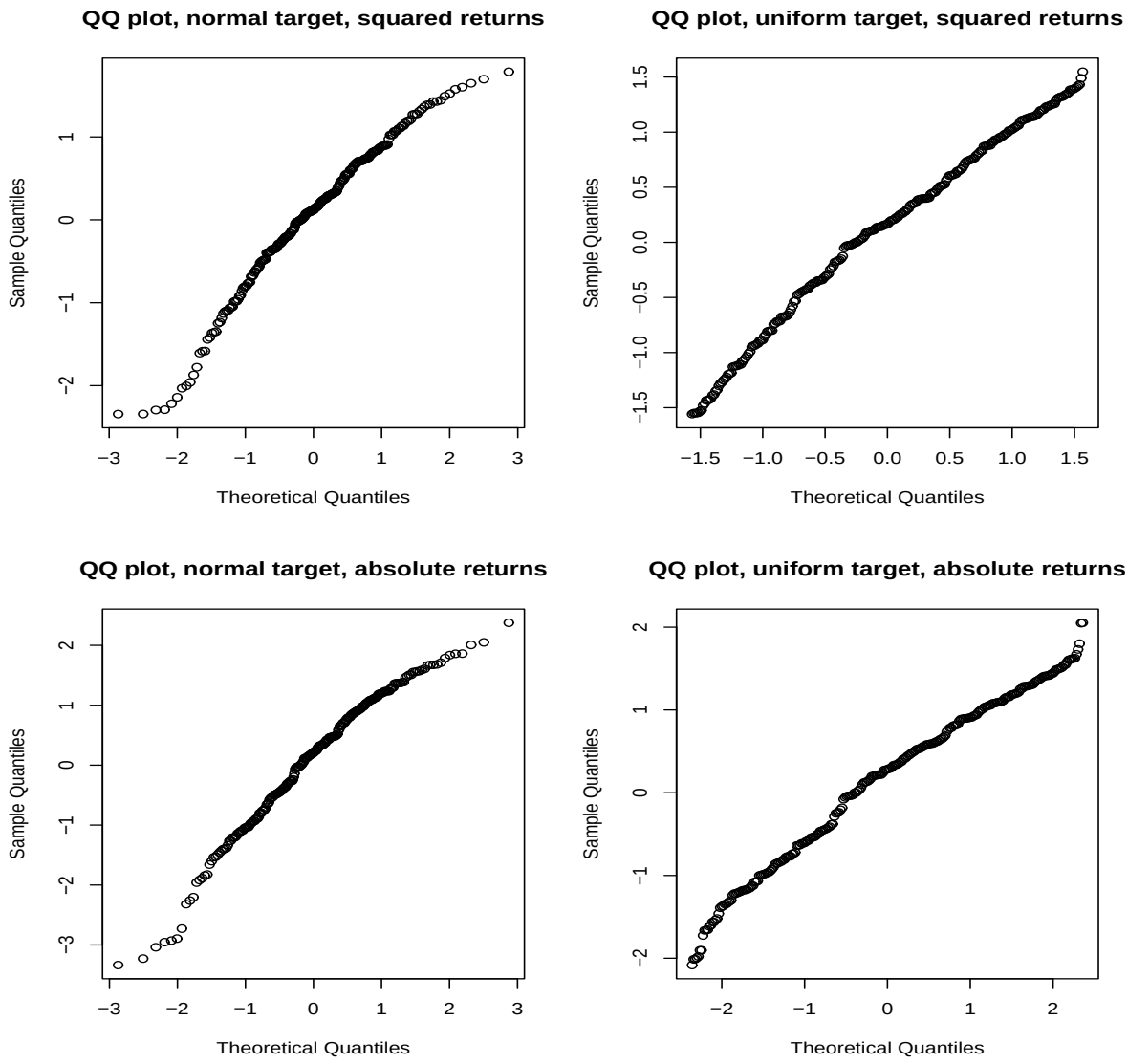

Figure 10: QQ plots of the NoVaS -transformed W series for the monthly MSFT series

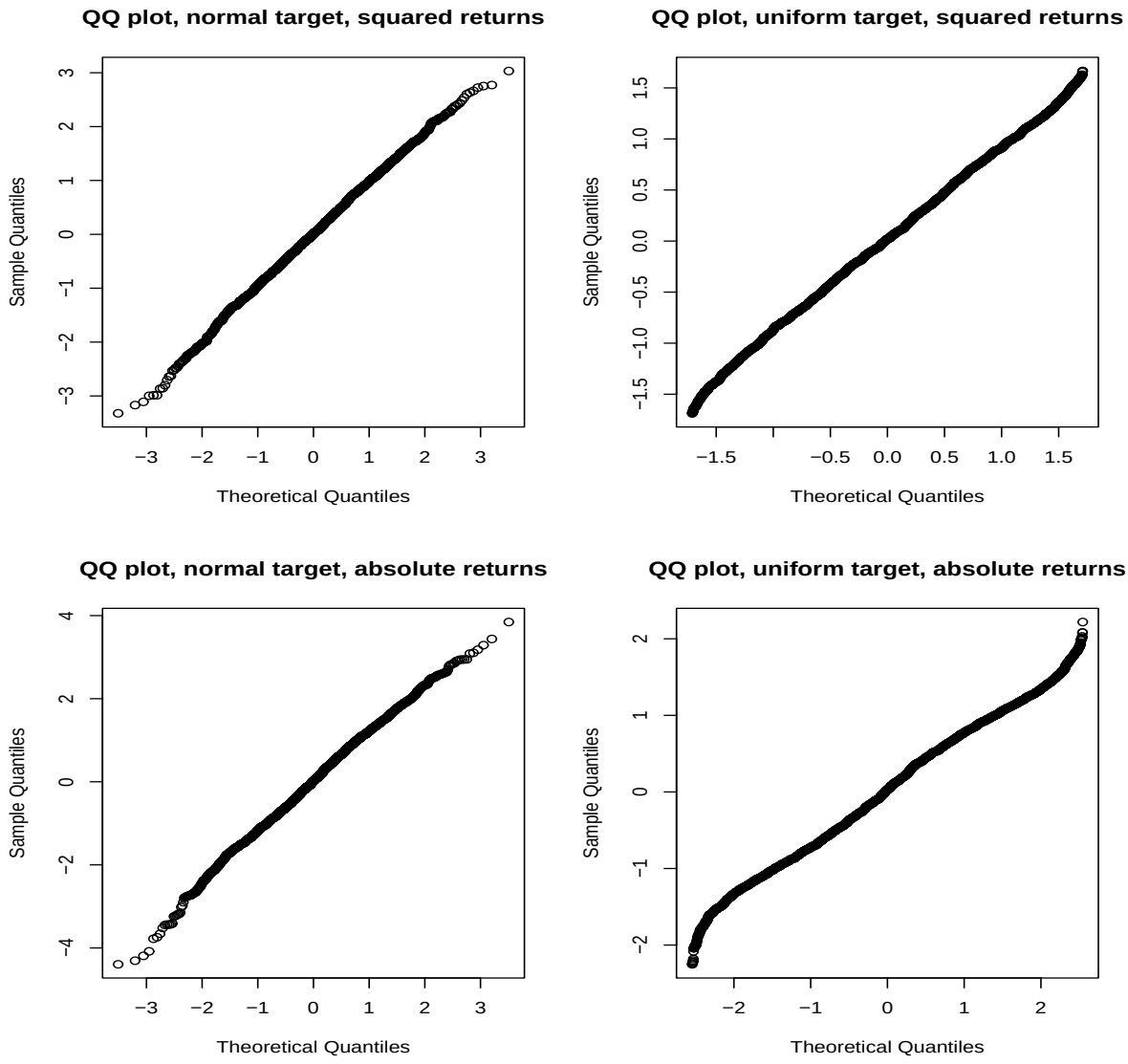

Figure 11: QQ plots of the NoVaS -transformed W series for the daily USD/Yen series

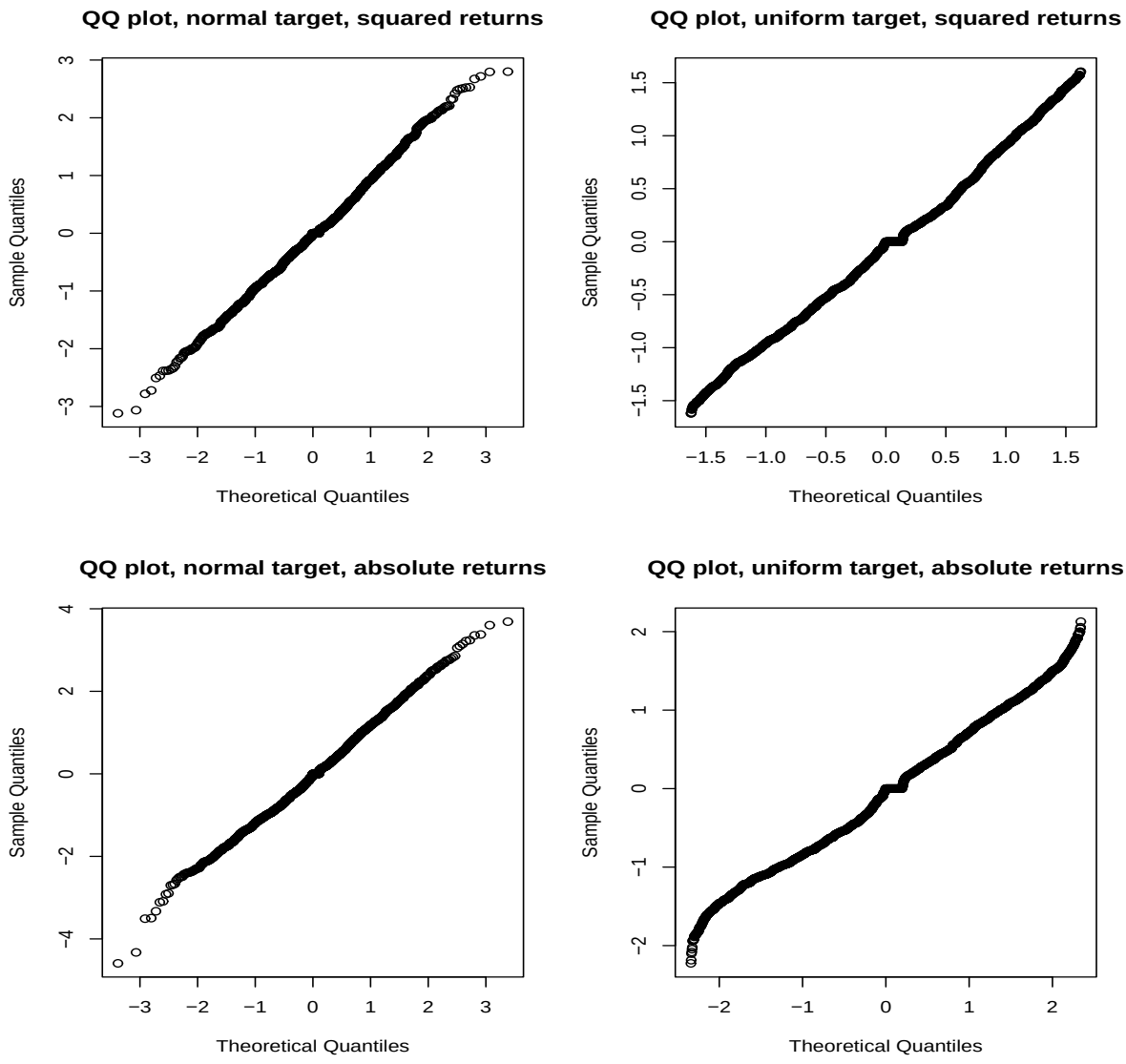

Figure 12: QQ plots of the NoVaS -transformed W series for the daily EFG series