WP 44-07

DIMITRIS POLITIS
University of California, San Diego, USA

DIMITRIOS THOMAKOS
University of Peloponnese, Greece
and
The Rimini Center for Economic Analysis, Italy

"NOVAS TRANSFORMATIONS: FLEXIBLE INFERENCE FOR VOLATILITY FORECASTING"

NoVaS Transformations:
Flexible Inference for Volatility Forecasting[*]

Dimitris N. Politis[†]   Dimitrios D. Thomakos[‡]

October 4, 2007

Abstract

In this paper we contribute several new results on the NoVaS transformation approach for volatility forecasting introduced by Politis (2003a,b, 2007). In particular: (a) we introduce an alternative target distribution (uniform); (b) we present a new method for volatility forecasting using NoVaS ; (c) we show that the NoVaS methodology is applicable in situations where (global) stationarity fails such as the cases of local stationarity and/or structural breaks; (d) we show how to apply the NoVaS ideas in the case of returns with asymmetric distribution; and finally (e) we discuss the application of NoVaS to the problem of estimating value at risk (VaR). The NoVaS methodology allows for a flexible approach to inference and has immediate applications in the context of short time series and series that exhibit local behavior (e.g. breaks, regime switching etc.) We conduct an extensive simulation study on the predictive ability of the NoVaS approach and find that NoVaS forecasts lead to a much 'tighter' distribution of the forecasting performance measure for all data generating processes. This is especially relevant in the context of volatility predictions 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.

Keywords: ARCH, GARCH, local stationarity, structural breaks, VaR, volatility.

[†]Department of Mathematics and Department of Economics, University of California, San Diego, USA. Email: politis@math.ucsd.edu

[‡]Department of Economics, University of Peloponnese, Greece. Email: thomakos@uop.gr

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 results on the NoVaS transformation approach to volatility prediction. The NoVaS methodology was introduced by Politis (2003a,b, 2007): its name is an acronym for 'Normalizing and Variance Stabilizing' transformation. NoVaS is based on exploratory data analysis ideas, it is model-free and especially relevant when making forecasts in the context of underlying data generating processes that exhibit local behavior (e.g. locally stationary time series, series with parameter breaks or regime switching etc.). It allows for a flexible approach to inference and is also well suited for application to short time series.

NoVaS is completely data-adaptive in the sense that, for its application, one does not need to assume parametric functional expressions for the conditional mean (which is taken to be zero in most financial returns) or the conditional variance (volatility) of the series under study. In addition, NoVaS is a model-free and distribution-free approach; hence its usefulness under a variety of contexts where we do not know *a priori* which (parametric or not) family of models is appropriate for our data. Because of its flexibility, the NoVaS approach can easily handle arbitrary forms of nonlinearity in returns and volatility. Finally, an important point to note is the relative computational ease required for NoVaS , which is in marked contrast to many methods and models for volatility.

The original development of the NoVaS approach was made in Politis (2003a,b, 2007) in the context of prediction of squared returns having as its 'springing board' the popular ARCH model with normal innovations. In these papers the problem of prediction in a NoVaS context was addressed using the L_1-norm to quantify the prediction error in the special case of a single, parametric expression for the dispersion of the returns (a modified ARCH equation) and transformation to normality.

In the paper at hand we present a number of new results on NoVaS. First, we present three theoretical contributions to the NoVaS approach: (a) we introduce an alternative target distribution (uniform) for performing NoVaS and compare it to the original target distribution (standard normal); (b) we introduce a new method for *bona fide* volatility forecasting, extending the original NoVaS notion of prediction of squared returns; finally, (c) we show how the NoVaS methodology can be used in a Value-at-Risk (VaR) context. Secondly, we conduct a comprehensive simulation study about the relative forecasting performance of NoVaS: we consider a wide variety of volatility models and we compare the forecasting performance of NoVaS with that of a benchmark GARCH

model. 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 data generating processes (DGP) we consider. This finding is especially relevant in the context of volatility predictions 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 forecasting performance of volatility predictions.

To the best of our knowledge no other work has considered the volatility prediction problem in a similar fashion. Possibly related to our work is a recent paper by Hansen (2006) that has considered the problem of forming prediction intervals using a semiparametric approach. Hansen works with a set of (possibly 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 this work is that both approaches use a transformation of the original data and the empirical distribution to make predictions. The main difference, however, is that Hansen does work in the context of a (possibly misspecified) model whereas we work in a model-free context.

The literature on volatility modeling, prediction and the evaluation of volatility forecasts is very large and appears to be continuously expanding. We can only selectively mention certain relatively recent papers that are related to the 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 (2004) 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

3

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 review the general development of the NoVaS approach; in section 3 we outline the use of NoVaS for VaR applications; in section 4 we present the design of our simulation study and discuss the simulation results; in section 5 we present empirical applications of NoVaS using real-world data; finally, in section 6 we offer some concluding remarks.

2   The NoVaS Methodology

In this section we present an overview of the NoVaS methodology that includes the NoVaS transformation, the implied NoVaS distributions, the methods for distributional matching and NoVaS forecasting.

2.1   NoVaS transformation and implied distributions

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

1. X_t has a non-Gaussian, approximately symmetric distribution that exhibits excess kurtosis; alternatively, X_t may not have finite moments of order greater than 1, that is $\mathsf{E}|X_t|^{1+\delta} < \infty$ for some $\delta > 0$.

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 suggest 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. As its acronym possibly suggests, the application of the NoVaS approach aims at making the inference problem 'simpler' by applying a suitable transformation that reduces or eliminates the modeling problems created

---

[1]Departures from the assumption of stationarity and/or some of the other 'stylized facts' listed here will be discussed in what follows.

by non-Gaussianity, i.e. high volatility and nonlinearity; it attempts to transform the (marginal) distribution of X_t to a more 'manageable' one, to account for the presence of high volatility, and to reduce dependence. It is important to stress from the outset that the NoVaS transformation is not a model but a 'model-free' approach with an exploratory data analysis flavor: it requires no structural assumptions and does not estimate any constant, unknown parameters. It is closely related to the idea of 'distributional goodness of fit' as it attempts to transform the original return series X_t into another series, say W_t, whose properties will match those of a known target distribution.

The first step in the NoVaS transformation is variance stabilization that takes care of 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 $\mathcal{F}_{t|t-p} \stackrel{\text{def}}{=} \{X_t, X_{t-1}, \ldots, X_{t-p}\}$

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

where α is a scalar control parameter, $\boldsymbol{a} \stackrel{\text{def}}{=} (a_0, a_1, \ldots, a_p)^\top$ is a $(p+1) \times 1$ vector of control parameters and $G(\cdot; \alpha, \boldsymbol{a})$ is to be specified.[2] Note that the first novel element here is the introduction of the current value of X_t in constructing γ_t; this is a small but crucial difference in the NoVaS approach which is fully explained below when we discuss the implied distributions obtained under NoVaS. 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:

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

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. The obvious choices for $g(z)$ now become $g(z) = z^2$ 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 and the past p values of the squared or absolute returns.

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. $g(z) = |z|$ is popular

---

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

recently for volatility modeling; see e.g. Ghysels and Forsberg (2004) and the references therein. Robustness in the presence of outliers in an obvious advantage of absolute vs. squared returns. In addition, the mean absolute deviation is proportional to the standard deviation for the symmetric distributions that will be of current interest.

Remark 1. One of the proposed 'stylized facts' concerns the (approximate) symmetry of the distribution of returns. Asymmetries can arise for a variety of reasons and, in the end, affect the way that volatility responds to past positive and negative returns. If volatility responds in a differentiated fashion to past positive and negative returns then we should be taking this into account when computing our localized variance estimator in (1). If, therefore, we have reason to believe (or have tested and found) that asymmetries are present we can modify equation (1) as follows:

$$G(\mathcal{F}_{t|t-p}; \alpha, \boldsymbol{a}, \boldsymbol{b}) \stackrel{\text{def}}{=} \alpha s_{t-1} + \sum_{j=0}^{p} a_j g(X_{t-j}) + \sum_{k=1}^{p} b_k g(X_{t-k}) \mathbf{1}\{X_{t-k} < 0\} \tag{3}$$

where $\boldsymbol{b} \stackrel{\text{def}}{=} (b_1, \ldots, b_p)^{\top}$ and $\mathbf{1}\{A\}$ is the indicator function of the set A. As we show in the next section that there is no problem in handling asymmetries of this form within the NoVaS context.

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:

$$W_t \equiv W_t(\alpha, \boldsymbol{a}) \stackrel{\text{def}}{=} \frac{X_t}{\phi(\gamma_t)} \tag{4}$$

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 make W_t follow as closely as possible a known, target distribution that is symmetric, easier to work with and that explains the presence of excess kurtosis in X_t. The obvious choice for such a distribution is the standard normal, hence the normalization in the NoVaS method.[3] However, we are not constrained to use only the standard normal as the target distribution. A simple alternative would be the uniform distribution. We will use both the standard normal and the uniform distribution in illustrating the way the NoVaS transformation works. Matching the target distribution with the studentized return series W_t is the 'distributional goodness of fit' component of NoVaS.

Remark 2. The distributional matching noted above focuses on the marginal distribution of the transformed series W_t. Although for all practical purposes this seems sufficient, one can

---

[3]The standard normal has an added advantage that comes useful in prediction, namely that it implies optimal *linear* predictors.

also consider distributional matching for *joint* distributions of W_t. It is shown in Politis (2003a,b, 2007) that the distributional matching procedure described in the next section can be applied to a linear combination of the form $W_t + \lambda W_{t-k}$ for some value of lag k and several different values of the weight parameter λ.

Let us assume, for the moment, that such a distributional matching is feasible and that the distribution of W_t can be made statistically indistinguishable from the target distribution. What can we infer from the studentization about the conditional distribution of the returns? To answer this we need to consider the *implied* model that is a by-product of the NoVaS transformation. If we were to solve with respect to X_t in equation (4), using the fact that γ_t depends on X_t, we would obtain that:

$$X_t = U_t A_{t-1} \tag{5}$$

where the two terms on the right-hand side are given by:

$$U_t \stackrel{\text{def}}{=} \left\{ \begin{array}{ll} W_t/\sqrt{1 - a_0 W_t^2} & \text{if} \quad \phi(z) = \sqrt{z} \\ W_t/(1 - a_0|W_t|) & \text{if} \quad \phi(z) = z \end{array} \right\} \tag{6}$$

for U_t and by:

$$A_{t-1} \stackrel{\text{def}}{=} \left\{ \begin{array}{ll} \sqrt{\alpha s_{t-1} + \sum_{j=1}^p a_j X_{t-j}^2} & \text{if} \quad g(z) = z^2 \\ \alpha s_{t-1} + \sum_{j=1}^p a_j |X_{t-j}| & \text{if} \quad g(z) = |z| \end{array} \right\} \tag{7}$$

for A_{t-1} that depends on $\mathcal{F}_{t-1|t-p}$. Note that the implied model of equation (5) is similar to an $ARCH(p)$ model, when $g(z) = z^2$, with the distribution of U_t being known (e.g. the standard normal). For any given target distribution for W_t we can find the distribution of U_t that will correspond to the conditional distribution of the returns. The case where $g(z) = z^2$ and where the distribution of W_t is taken to be the standard normal was extensively analyzed in Politis (2003a,b, 2007). The case where $g(z) = |z|$ as well as the case where the distribution of W_t is taken to be the uniform is a new contribution that extends the applicability and robustness of the NoVaS approach. We further discuss their use below.

To understand the implied distribution of U_t first note that the range of W_t is bounded. Using equation (4) it is straighforward to show that $|W_t| \leq 1/\sqrt{a_0}$, when $g(z) = z^2$, whereas $|W_t| \leq 1/a_0$, 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 $a_0 \leq 1/9$, 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 $a_0 \leq 1/3$. On the other hand, if the target distribution for W_t is the uniform then our choice of a_0 determines the length of the interval on which W_t would be defined: different choices of a_0 would imply different intervals of the form $\left[-1/\sqrt{a_0}, +1/\sqrt{a_0}\right]$, for $g(z) = z^2$, and $[-1/a_0, +1/a_0]$, for $g(z) = |z|$. Notice that the use of the uniform target distribution is, in this respect, less restrictive than the use of the standard normal distribution: we do not have to impose any constraints in a_0 for using the uniform distribution as we have to do when using the standard normal.

Taking into account the boundedness in W_t the implied distribution of U_t can be derived using standard methods. With two target distributions and two options for computing γ_t we obtain four different implied densities that should be more than adequate to cover problems of practical interest. For the case where the target distribution is the standard normal we have the following implied distributions for U_t:

$$
\begin{aligned}
f_1(u, a_0) &= c_1(a_0) \times (1 + a_0 u^2)^{-1.5} \exp\left[-0.5u^2/(1 + a_0 u^2)\right] & \text{when } g(z) = z^2 \\
f_2(u, a_0) &= c_2(a_0) \times (1 + a_0|u|)^{-2} \exp\left[-0.5u^2/(1 + a_0|u|)^2\right] & \text{when } g(z) = |z|
\end{aligned}
\tag{8}
$$

whereas for the case where the target distribution is the uniform we have:

$$
\begin{aligned}
f_3(u, a_0) &= c_3(a_0) \times (1 + a_0 u^2)^{-1.5} & \text{when } g(z) = z^2 \\
f_4(u, a_0) &= c_4(a_0) \times (1 + a_0|u|)^{-2} & \text{when } g(z) = |z|
\end{aligned}
\tag{9}
$$

The densities from the use of the uniform target distribution are new, in this and related contexts. Below we discuss their similarities with the densities from equation (8).

The constants $c_i(a_0)$, for $i = 1, 2, 3, 4$, ensure that the densities are proper and integrate to one. As was noted in Politis (2004), the rate at which $f_1(u, a_0)$ tends to zero is the same as in the $t_{(2)}$ distribution, although it has practically lighter tails.[4] Also note that the use of the uniform as the target distribution gives us two densities that have the *limiting* form (for large u) of the densities that use the standard normal as the target distribution - this affects the tail behavior of $f_3(u, a_0)$ and $f_4(u, a_0)$ compared to the tail behavior of $f_1(u, a_0)$ and $f_2(u, a_0)$.

We graphically illustrate the differences among the implied densities in equations (8) and (9) and compare them with the standard normal and $t_{(2)}$ densities. In Figure 1 we plot, on four panels, the standard normal density, the $t_{(2)}$ density and the four implied NoVaS densities. We choose the parameter a_0 so as to show the flexibility of these new distributions. On the top left panel of Figure 1 we compare the standard normal and $t_{(2)}$ density with $f_1(u, 0.1)$ and we see

---

[4]Basically, $f_1(u, a_0)$ looks like a $\mathcal{N}(0, 1)$ distribution for small u but has a $t_{(2)}$-type tail.

that its tails are in-between the tails of the normal and the t distributions. On the top right panel of Figure 1 we make the same comparison with $f_2(u, 0.3)$ and we can clearly see that this NoVaS distribution approximately matches the tail behavior of the $t_{(2)}$ distribution, although it appears that the $f_2(u, 0.3)$ distribution has slightly fatter tails. On the bottom left panel of Figure 1 we plot the $f_3(u, 0.55)$ distribution and now we see an almost complete match with the almost the whole of the $t_{(2)}$ distribution - this was to be expected as $a_0 = 0.55$ matches the inverse of the degrees of freedom of the $t_{(2)}$ distribution. Finally, on the bottom right panel of Figure 1 we plot the $f_4(u, 0.75)$ distribution, which exhibits the most 'extreme' behavior being much more concentrated around zero and with substantially fatter tails than the $t_{(2)}$ distribution.

Note that all $f_i(u, a_0)$ distributions lack moments of high order. In particular, $f_1(u, a_0)$ and $f_3(u, a_0)$ have finite moments of order a if $a < 2$, whereas $f_2(u, a_0)$ and $f_4(u, a_0)$ have finite moments of order a if $a < 1$. In the terminology of Politis (2004), $f_1(u, a_0)$ and $f_3(u, a_0)$ have 'almost' finite second moments, and $f_2(u, a_0)$ and $f_4(u, a_0)$ have 'almost' finite first moments. To illustrate this point, and to see how the $f_i(u, a_0)$ distributions compare with the standard normal and the $t_{(2)}$ distributions, we report in Table 1 the absolute moments of orders 1 through 4, using the same values for a_0 as in Figure 1. We take a finite but large range to perform the integration so as to clearly show the differences among the distributions. The results in Table 1 tell the same story as Figure 1, although the points made for Figure 1 are now abundantly evident: the use of the uniform target distribution $f_4(u, a_0)$ has the most 'extreme' behavior, as noted above, and can considered to be the most flexible when one has to deal with a 'difficult' time series that does not possess finite moments. The novelty of NoVaS in introducing X_t in the time-localized measure of variance used in studentizing the returns allows us a great deal of flexibility in accounting for any degree of not only tail heaviness but also for the possible non-existence of second (or higher) moments. Therefore, the potential of NoVaS is not restricted to applications using financial returns but also to applications where the time series may have infinite variance.

Remark 3. These results affords us the opportunity to make a preliminary remark on an issue that we will have to deal with in forecasting, namely the choice of loss function for generating forecasts. The most popular criterion for measuring forecasting performance is the mean-squared error (MSE) criterion. When forecasting returns the MSE corresponds to the (conditional) variance of the forecast errors; when forecasting squared returns (equiv. volatility) the MSE corresponds to the (conditional) fourth order moment of the forecast errors. However, a lot of literature these days concerns returns that lack a finite 4th moment, see for example, Hall and Yao (2003), Politis (2003b, 2004), Berkes and Horvath (2004) and others. Of course, a lack of finite fourth moments

9

renders the use of the MSE invalid in measuring forecasting performance. In contrast, the mean absolute deviation (MAD) of the forecast errors, that corresponds to the first absolute moment, appears to be a preferred choice for comparing the forecasting performance of returns, squared returns and volatility.[5]

2.2 NoVaS distributional matching

2.2.1 Parametrization

We next turn to the issue of parameter selection or calibration. Since NoVaS does not impose a structural model on the data we would like to have a flexible, parsimonious parameter structure that would be relatively easy to adjust so as to achieve the desired distributional matching. The parameters that are free to vary are p, the NoVaS order, and $(\alpha, \boldsymbol{a})$ or $(\alpha, \boldsymbol{a}, \boldsymbol{b})$ if we want to account for possible asymmetries. The rest of the discussion will be in terms of p, α and $\boldsymbol{a}$. See Remark 4 below for the case where $\boldsymbol{b}$ is also present. The parameters α and $\boldsymbol{a}$ obey certain restrictions to ensure positivity for the variance. In addition, it is convenient to assume that the parameters act as filter weights on squared or absolute X_t's, obey a summability condition of the form $\alpha + \sum_{j=0}^{p} a_j = 1$ and they decline in magnitude $a_i \geq a_j$ for $i > j$.

We first consider the case when $\alpha = 0$. The simplest parametric 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:

$$a_j = \left\{ \begin{array}{lll} 1/\sum_{j=0}^{p} \exp(-bj) & \text{for} & j = 0 \\ a_0 \exp(-bj) & \text{for} & j = 1, 2, \ldots, p \end{array} \right\} \tag{10}$$

where b is a control parameter. These are the exponential NoVaS weights proposed in Politis (2003a,b, 2007).

The exponential weighting scheme allows for greater flexibility, without imposing an additional cost[6] in terms of the control parameter b, and is our preferred method in applications. Given a choice for the weighting scheme one needs to calibrate the parameters p, the lag length, and b so as to achieve distributional matching for the studentized series W_t. Note that since we have a direct mapping $\boldsymbol{\theta} \stackrel{\text{def}}{=} (p, b) \mapsto (\alpha, \boldsymbol{a})$ it will be convenient in what follows to denote the studentized

---

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

[6]Note that now p is effectively infinity and need not me selected per se.

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 we need an appropriately defined objective function. We discuss the possible choices of objective functions in the next subsection.

Remark 4. It is straightforward to modify equation (10) to allow for the presence of asymmetries. Allowing for exponential weights with a different control parameter, say c, for the parameters in $\boldsymbol{b}$ we get the following parameter representation:

$$\begin{cases} a_0 = \left[\sum_{j=0}^{p} \exp(-bj) + \sum_{k=1}^{p} \exp(-ck)\right]^{-1} & \text{for} \quad j = 0 \\ a_j = a_0 \exp(-bj) & \text{for} \quad j = 1, 2, \ldots, p \\ b_k = a_0 \exp(-ck) & \text{for} \quad k = 1, 2, \ldots, p \end{cases} \tag{11}$$

that obeys all restrictions discussed above. The parameter vector now becomes $\boldsymbol{\theta} \stackrel{\text{def}}{=} (p, b, c)$ mapping to $(\alpha, \boldsymbol{a}, \boldsymbol{b})$.

2.2.2 Objective functions for optimization

Natural candidates for objective functions to be used for achieving distributional matching are all smooth functions that assess one or more of the salient features of the target distribution. 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) and others. All these measures are essentially distance-based and the optimization will attempt to minimize the distance between the sample statistics and the theoretical ones.[7]

Let us consider the simplest case first–one easily used in applications–, 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:

$$\mathcal{K}_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\sum_{t=1}^{n}(W_t - \bar{W}_n)^4}{n s_n^4} - \kappa^* \tag{12}$$

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

---

[7]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 prediction.

distribution. For the standard normal distribution we have that $\kappa^* = 3$ while for the uniform distribution we have that $\kappa^* = 1.85$. The objective function for this case can be taken to be the absolute value of the sample excess kurtosis, that is $D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} |\mathcal{K}_n(\boldsymbol{\theta})|$ 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 a procedure will work in view of the intermediate value theorem. For $p = 0$ we have that $a_0 = 1$ and thus $W_t = sign(X_t)$ for which we have that $\mathcal{K}_n(\boldsymbol{\theta}) < 0$, 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 p in between $[0, p_{\max}]$ that will make the sample excess kurtosis approximately equal to zero. This is what happens in practice. This observation motivates the following algorithm, applied to the exponential weighting scheme (Politis [2003a]):

- 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 equation (10).

- Trim the value of p, if desired, by removing those parameters that do not exceed a pre-specified threshold. For example, if $a_\ell^* \leq 0.01$ then set $a_m^* = 0$ for all $m \geq \ell$, and re-normalize the remaining parameters so that they sum up to one.

The above algorithm is easily adapted for use with the exponential weighting scheme in equation (11) that accounts for asymmetries by doing a two-dimensional search over two discrete grids of values, say B as above and $C \stackrel{\text{def}}{=} (c_{(1)}, c_{(2)}, \ldots, c_{(M)})$.

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[8], or for goodness-of-fit statistics such as the quantile-quantile correlation coefficient or the Kolmogorov-Smirnov statistic, as noted above. One performs the same steps but simply evaluates a different objective function. In practice it turns out that, in the context of financial returns data that are approximately symmetric, in most cases is sufficient to consider kurtosis matching. Note that for any choice of the objective function we have that $D_n(\boldsymbol{\theta}) \geq 0$ and, as noted in the algorithm above,

---

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

the optimal values of the parameters are clearly determined by the condition:

$$\boldsymbol{\theta}_n^* \stackrel{\text{def}}{=} \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \, D_n(\boldsymbol{\theta}) \tag{13}$$

Remark 5. 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 choose the value of α from the grid that optimizes out-of-sample predictive performance; see Politis (2003a,b, 2007) for more details.

Remark 6. It is important to stress that the specialized form that $G(\cdot; \alpha, \boldsymbol{a})$ takes in equation (2) is mostly for convenience, computational tractability and for allowing us to directly compare results with various GARCH-type specifications. What makes the difference in the approach is the inclusion of the term X_t^2 or $|X_t|$ - the rest of the terms can be modeled in alternative ways. To illustrate this, and to indicate the broad scope of the NoVaS methodology, consider the following semiparametric specification with $\alpha = 0$ and $g(z) = z^2$:

$$G(X_t, X_{t-1}, \ldots, X_{t-p}; 0, a_0) \stackrel{\text{def}}{=} a_0 X_t^2 + \xi_{t-1}(\boldsymbol{x}_{t-p}^2) \tag{14}$$

where $\boldsymbol{x}_{t-p}^2 \stackrel{\text{def}}{=} (X_{t-1}^2, \ldots, X_{t-p}^2)^\top$ is and $\xi_{t-1}(\cdot)$ is an arbitrary function to be estimated nonparametrically. Let h denote a bandwidth value and denote by $K_h(\cdot/h)$ any suitable kernel function. For any given value of a_0 one can calculate $\xi_{t-1}(\cdot)$ recursively as:

$$\widehat{\xi}_t(\boldsymbol{x}_{t-p}^2) \stackrel{\text{def}}{=} \sum_{j=1}^{p} a_j X_{t-j}^2 \tag{15}$$

where the new parameters a_j are now time-varying and given by:

$$a_j \stackrel{\text{def}}{=} \frac{K_h(X_{t-j}^2 - \boldsymbol{x}_{t-p}^2)}{\sum_{j=1}^{p} K_h(X_{t-j}^2 - \boldsymbol{x}_{t-p}^2)} - \frac{a_0}{p} \tag{16}$$

The new parameter vector in this case would be $\boldsymbol{\theta} = (p, a_0, h)$. Finally note that, except for the condition $a_i \geq a_j$ for $i \geq j$, all other conditions for the NoVaS parameters are satisfied.

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 any function of the returns, including

higher regular and absolute moments. Forecasting in NoVaS is performed while keeping in mind the possible non-existence of higher moments in the implied NoVaS distributions (see Remark 3). The choice of an appropriate forecasting norm, both for producing and for evaluating the forecasts, is crucial for maximizing forecasting performance.

In what follows we outline the forecasting method used after completing the NoVaS transformation concentrating on the L_1 norm for producing the forecasts and the mean absolute deviation (MAD) of the forecast errors for assessing forecasting ability. After optimization of the NoVaS parameters we will have available the transformed series $U_n^* \stackrel{\text{def}}{=} \{U_1(\boldsymbol{\theta}_n^*), \ldots, U_n(\boldsymbol{\theta}_n^*)\}$, which is the main 'ingredient' in performing forecasting for either returns or squared returns or any other moment of our choice: the U_t series appears in the *implied* model of equation (5), $X_t = U_t A_{t-1}$. It is important to keep in mind that the U_n^* series is a function of the W_n^* series for which we have performed distributional matching.

Let $\Pi_k[X|Z]$ denote the kth (regular or absolute) conditional power operator of the argument X given the argument Z. For example, $\Pi_1[XZ|Z] = XZ$, $\Pi_2[XZ|Z] = (X^2|Z) \cdot Z^2$ etc. Applying the power operator in the definition of the *implied* model of equation (5) at time $n+1$ we obtain:

$$\Pi_k[X_{n+1}|\mathcal{F}_n] = \Pi_k[U_{n+1}^*|\mathcal{F}_n]\,\Pi_k[A_n^*] \tag{17}$$

Depending on our choice of k and 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 power of the U_{n+1}^* series. To see this note that, in the context of the L_1 forecasting norm, the conditional median is the optimal predictor, so we have:

$$\mathsf{Med}\left[\Pi_k[X_{n+1}|\mathcal{F}_n]\right] = \mathsf{Med}\left[\Pi_k[U_{n+1}^*|\mathcal{F}_n]\right]\Pi_k[A_n^*] \tag{18}$$

where $\mathsf{Med}[x]$ stands for the median of x. Therefore, what we are after is an estimate of the conditional median[9] of $\Pi_k[U_{n+1}^*|\mathcal{F}_n]$.

The rest of the procedure depends on the temporal properties of the studentized series W_n^* and the target distribution. Consider first the case where observations for the W_n^* series are uncorrelated (which is what we expect in practice for financial returns). If the target distribution is the standard normal then, by the approximate normality of its marginal distribution, the W_n^* series is also independent and therefore the best estimate of the conditional median $\mathsf{Med}\left[\Pi_k[U_{n+1}^*|\mathcal{F}_n]\right]$ is the unconditional sample median of the appropriate power of the U_n^* series,

---

[9]It should be apparent that, in principle, one can obtain median forecasts for any measurable function of the returns.

namely $\widehat{\mathsf{Med}}\left[\Pi_k\left[U_n^*|\mathcal{F}_n\right]\right]$. The same result should also hold approximately for the case where the target distribution is the uniform: if the marginal and joint distributions of the W_n^* series are uniform then the series should be independent and the use of the unconditional sample median $\widehat{\mathsf{Med}}\left[\Pi_k\left[U_n^*|\mathcal{F}_n\right]\right]$ is still the best estimate of the conditional $\mathsf{Med}\left[\Pi_k\left[U_{n+1}^*|\mathcal{F}_n\right]\right]$.

When the observations for the W_n^* series are correlated then a slightly different procedure is suggested. If the target distribution is the standard normal then the optimal predictors are linear and one proceeds as follows. First, a suitable $AR(q)$ model is estimated (using any order selection criteria) for the W_n^* series and the forecast $\widehat{W}_{n+1}^*$ and forecast errors e_t, for $t = \max(p,q)+1,\ldots,n$ are retained. The conditional distribution of W_{n+1}^* can now be approximated using the distribution of the forecast errors shifted so that they have mean equal to $\widehat{W}_{n+1}^*$, i.e. using $\tilde{W}_t^* \overset{\text{def}}{=} e_t + \widehat{W}_{n+1}^*$. Then, letting $\widehat{U}_{n+1}^*$ denote the series constructed using these shifted forecast errors, e.g. $\widehat{U}_{t+1}^* \overset{\text{def}}{=} \tilde{W}_t^*/\sqrt{1 - a_0 \tilde{W}_t^*}$ when using squared returns, we have that 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 $\widehat{U}_{n+1}^*$ series, namely $\widehat{\mathsf{Med}}\left[\Pi_k\left[\widehat{U}_{n+1}^*|\mathcal{F}_n\right]\right]$.

If the target distribution is the uniform one cannot, in principle, use a linear model for prediction of the W_n^* series. An option is to ignore the sub-optimality of linear prediction and proceed exactly as above. Another option would be to directly forecast the conditional median of the U_n^* series using a variety of available nonprametric methods, see for example Cai (2002), Gannoun, Sarraco and Yu (2003).

Based on the above discussion we are able to obtain volatility forecasts $\widehat{h}_{n+1}^2$ 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_n^2 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:

$$\widehat{h}_{n+1,1}^2 \equiv \widehat{X}_{n+1}^2 \overset{\text{def}}{=} \widehat{\mathsf{Med}}\left[\Pi_2\left[U_n^*|\mathcal{F}_n\right]\right]\Pi_2\left[A_n^*\right] \tag{19}$$

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

$$\widehat{h}_{n+1,2}^2 \overset{\text{def}}{=} \Pi_2\left[A_n^*\right] \tag{20}$$

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 equation (1). What is important

15

to note is that forecasting based on $\widehat{\gamma}_{n+1}$ is neither forecasting of squared returns nor forecasting based on past information alone. Is, in fact, a linear combination of the two forecasts above thus incorporating elements from essentially two approaches. Using equations (1), (2), (6) and (7) it is straightforward to show that $\widehat{\gamma}_{n+1}$ can be expressed as:

$$
\begin{aligned}
\widehat{\gamma}_{n+1} \equiv \widehat{h}_{n+1,3}^2 \quad &\overset{\text{def}}{=} \quad \left\{ a_0^* \widehat{\text{Med}} \left[\Pi_2 \left[U_n^* | \mathcal{F}_n \right] \right] + 1 \right\} \Pi_2 \left[A_n^* \right] \\
&= \quad a_0^* \widehat{h}_{n+1,1}^2 + \widehat{h}_{n+1,2}^2
\end{aligned}
\tag{21}
$$

The above equation (21) is our new proposal for volatility forecasting using NoVaS. In his original work Politis (2003b) invariably used equation (19), and in effect conducted prediction of the one-step-ahead squared return via NoVaS. By contrast, equation (21) is a *bona fide* predictor of the one-step-ahead *volatility,* i.e., the conditional variance. For this reason, equation (21) will be the formula used in what follows, and in 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 equations (1), (6) and (7) it is easy to show that the forecast based on (c) would be given by:

$$
\begin{aligned}
\widehat{\gamma}_{n+1} \equiv \widehat{h}_{n+1,3} \quad &\overset{\text{def}}{=} \quad \left\{ a_0^* \widehat{\text{Med}} \left[\Pi_1 \left[U_n^* | \mathcal{F}_n \right] \right] + 1 \right\} \Pi_1 \left[A_n^* \right] \\
&= \quad a_0^* \widehat{h}_{n+1,1} + \widehat{h}_{n+1,2}
\end{aligned}
\tag{22}
$$

with $\widehat{h}_{n+1,1}$ and $\widehat{h}_{n+1,2}$ being identical expressions to equations (19) and (20) which use the first order absolute power transformation.

2.4 Departures from the assumption of stationarity

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 plague traditional methods. 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.

2.5 NoVaS and VaR

An important practical application of volatility forecasting is calculating value at risk (VaR) and a large literature is associated with VaR calculations and evaluating VaR performance.[10] Although there exist different approaches in calculating VaR that do not depend on volatility forecasts (e.g. methods that are based on extreme values) it is common to use the predictive distribution of returns to calculate VaR. Such a practice is susceptible to possible misspecification both in the assumed underlying distribution of the returns and in the model used to make volatility forecasts. The NoVaS application to VaR easily avoids both these problems as both the returns' distribution and the volatility forecast are data-adaptable and essentially model-free.

To make a VaR calculation using NoVaS we require the predictive distribution of the returns and an associated quantile $x_n(p)$. Considering a given probability level p we have that $p = \mathsf{P}\left[X_{n+1} \leq x_n(p)|\mathcal{F}_n\right]$ and:

$$\mathsf{P}\left[X_{n+1} \leq x_n(p)|\mathcal{F}_n\right] = \mathsf{P}\left[U_{n+1}A_n \leq x_n(p)|\mathcal{F}_n\right] = \mathsf{P}\left[U_{n+1} \leq x_n(p)A_n^{-1}|\mathcal{F}_n\right] \tag{23}$$

Since A_n is known when we condition on $\mathcal{F}_n$ we immediately have, for $u_n(p) \stackrel{\text{def}}{=} x_n(p)A_n^{-1}$ that the required quantile of the returns' distribution is given by a scaled quantile of the distribution of U_{n+1} which is explicitly available: its one of the distribution functions associated with the densities in equations (8) and (9) and can be computed numerically by inversion. Alternatively, the required quantile can be estimated using the empirical distribution of the U_t's, which can be much faster computationally.

To see the above, note first that we have that $\mathsf{P}\left[U_{n+1} \leq u_n(p)|\mathcal{F}_n\right] = p$ and therefore:

$$u_n(p) = F^{-1}(p|\mathcal{F}_n) \Rightarrow x_n(p) = A_nu_n(p) \tag{24}$$

---

[10]We cannot attempt a review on the VaR literature here. Book length summaries can be found in Embrechts, Kuppelberg and Mikosch (1997) and Jorion (1997).

where $F^{-1}(p|\mathcal{F}_n)$ is the appropriate inverse distribution function of U_{n+1}. Using the notation on the last section, the inverse distribution function will be computed, using the optimum parameter values $\boldsymbol{\theta}^*$, as:

$$u_n^*(p) \stackrel{\text{def}}{=} \inf \left\{ u \in \mathbb{D}_i : \int_{u_{i,L}}^{u_n^*(p)} f_i(u, a_0^*) du \geq p \right\} \tag{25}$$

where $\mathbb{D}_i$ is the domain of the $f_i(u, a_0^*)$ and where $u_{i,L} \stackrel{\text{def}}{=} \inf_u \mathbb{D}_i$ is the lower value of the domain, determined by the boundedness conditions imposed by the target distribution (i.e. $u_{1,L} = u_{3,L} = -1/\sqrt{a_0^*}$ and $u_{2,L} = u_{4,L} = -1/a_0^*$). Thereafter, the required quantile for the predictive distribution of the returns will be given as $x_n^*(p) \stackrel{\text{def}}{=} A_n^* u_n^*(p)$.

In a similar, but simpler fashion, we can obtain the required quantile from the empirical distribution function of the U_t's as:

$$\widehat{u}_n^*(p) \stackrel{\text{def}}{=} \inf \left\{ u \in \mathbb{D}_i : \frac{1}{n} \sum_{t=1}^n \mathbf{1} \left\{ U_t^* \leq u \right\} \geq p \right\} \tag{26}$$

Note that, again, what makes the difference with other similar approaches for VaR calculation that are based on the predictive distribution of the returns, is the presence of the term a_0^* that determines the tail behavior of the distribution of U_{n+1}. In addition, note that we have not made used of the NoVaS volatility forecasts (as defined in the previous section) but only of the term A_n^*.

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, under a variety of different underlying data generating processes (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. Working 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 really do not have such a series of true volatility. The proxies range from realized volatility, generally agreed to be one (if not the best) such measure, to range-based measures 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:

$$
\begin{aligned}
X_{t,ij} &\stackrel{\text{def}}{=} \mu_j + h_{t,ij} Z_{t,ij} \\
h_{t,ij}^2 &\stackrel{\text{def}}{=} h_j(h_{t-1,ij}^2, X_{t-1,ij}^2, \boldsymbol{\theta}_{tj})
\end{aligned}
\tag{27}
$$

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.[11]

We next present the volatility equations 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:

$$
h_t^2 = \omega + \alpha h_{t-1}^2 + \beta (X_{t-1} - \mu)^2
\tag{28}
$$

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:

$$
h_t^2 = \omega_b + \alpha_b h_{t-1}^2 + \beta_b (X_{t-1} - \mu)^2 \ , \ b = 1, 2, \ldots, B
\tag{29}
$$

---

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

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 $\omega_b = \sigma_b^2(1 - \alpha_b - \beta_b)/250$, where σ_b^2 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:

$$h_t^2 = \omega(t) + \alpha(t)h_{t-1}^2 + \beta(t)(X_{t-1} - \mu)^2 \tag{30}$$

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:

$$h_t^2 = \sum_{s=1}^{2} \mathbf{1}\{\mathsf{P}(S_t = s)\} \left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s(X_{t-1} - \mu_s)^2\right] \tag{31}$$

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:

$$h_t^2 = \sum_{s=1}^{2} Q_s(X_{t-1}) \left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s(X_{t-1} - \mu_s)^2\right] \tag{32}$$

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 $\gamma_2 = 1$.

20

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 equation (28) while the return equation takes the following form:

$$X_t = \mu + [a - b(t/T)] h_t Z_t \tag{33}$$

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:

$$\log h_t^2 = \omega + \alpha \log h_{t-1}^2 + w_t \ , \ w_t \sim \mathcal{N}(0, \sigma_w^2) \tag{34}$$

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 and evaluate it with the known, one-step ahead simulated volatility. NoVaS forecasts are produced for both a normal and uniform target distribution and using 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. SQUT, NoVaS forecasts made using squared returns and uniform target.

3. ABNT, NoVaS forecasts made using absolute returns and normal target.

4. ABUT, NoVaS forecasts made using absolute returns and uniform target.

5. GARCH, L_2-based GARCH forecasts.

6. 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 $F = 6$ 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 $e_{t,ijf} \stackrel{\text{def}}{=} h^2_{t+1,ij} - \widehat{h}^2_{t+1,ijf}$. Using these forecast errors we compute the mean absolute deviation for each model, each forecast method and each simulation run as:

$$m_{ijf} = MAD_{ijf} \stackrel{\text{def}}{=} \frac{1}{T_1} \sum_{t=T_0+1}^{T_1} |e_{t,ijf}| \tag{35}$$

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:

$$\bar{x}_{jf} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} m_{ijf} \tag{36}$$

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 $m_{ij,N} \stackrel{\text{def}}{=} m_{ijf}/m_{ij0}$, $f = 1, 2, 3, 4$ to be the ratio of the MAD of any of the NoVaS forecasts relative to the benchmark and $m_{ij,G} \stackrel{\text{def}}{=} m_{ijf}/m_{ij0}$, $f = 5, 6$ 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):

$$\widehat{P}_f \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\left(m_{ij,N} \leq m_{ij,G}\right) \tag{37}$$

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 $j = 1, \ldots, 4$ corresponds to the aforementioned methods NoVaS methods SQNT, SQUT, ABNT, and ABUT respectively and $\widehat{P}$ corresponds to their union.

3.2   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 2 and 9): 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 surprise: *any* of the NoVaS methods (SQNT, SQUT, ABNT, ABUT) 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 350 (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 nonstationarity may arise that will be addressed in what follows.

- GARCH with discrete breaks (B-GARCH) (Tables 3 and 10): 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, *all* 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 4 and 11): 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 *all* NoVaS methods are better than either GARCH method. The best now is SQUT with SQNT a close second. 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 5 and 12): We note again that 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 6 and 13): 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 7 and 14): 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 8 and 15): 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.

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 nonstationarities.

- 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)

24

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).

Remark 8. We should note that the top performance of the SQNT forecasting method can be partly attributed to the structure of the simulation experiments: all DGPs depend on squared and not absolute returns and the underlying error distribution $(t_{(3)})$ has finite second moments. If these features are not present in real data it may well be the case that one of the other NoVaS forecasting methods (SQUT, ABNT and ABUT) is found to exhibit better performance.

4   Empirical Examples

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, DGP 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 ith daily return for month t then the monthly realized volatility is defined as $\sigma_t^2 \overset{\text{def}}{=} \sum_{i=1}^{m} r_{t,i}^2$, 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 ith day of month t. The daily range-based volatility is defined as in Parkinson (1980) as $\sigma_{t,i}^2 \overset{\text{def}}{=} [\ln(H_{t,i}) - \ln(L_{t,i})]^2 / [4\ln(2)]$; then, the corresponding monthly realized measure would be defined as $\sigma_t^2 \overset{\text{def}}{=} \sum_{i=1}^{m} \sigma_{t,i}^2$. 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 16. 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 2 to 9 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 2 and 3 are for the monthly S&P500 returns, Figures 4 and 5 are for monthly MSFT returns, Figures 6 and 7 are for the daily USD/Yen returns and Figures 8 and 9 are for the daily EFG returns. Of interest are the figures that plot the estimated recursive moments. In Figure 3 we see that the mean and standard deviation of the monthly S&P500 returns are fairly stable while the skewness and kurtosis exhibit breaks and kurtosis is rising - a possible indication that there is no finite fourth moment for this series. Similar observations can be made for the other four series as far as recursive kurtosis goes.

4.2   NoVaS Optimization and Forecasting Specifications

Our NoVaS in-sample analysis is performed for all four possible combinations of target distributions and variance measures, i.e. squared and absolute returns using a normal distribution and squared and absolute returns using a uniform distribution. We use the exponential NoVaS algorithm as discussed in section 2, with $\alpha = 0$, 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 equation (12). The results of our in-sample analysis are given in Table 17. 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

26

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 also performed for all eight possible configurations of target distributions and variance measures - we report all of them in Tables 18 and 19. 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 predictions 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 predict 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 equations (21) and (22).[12] 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), GARCH-type forecasts can be improved if done using an L_1 rather than L_2 norm. 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 given in the previous section and report the mean absolute deviation (MAD) and root mean-squared (RMSE) of the forecast errors $e_t \overset{\text{def}}{=} \sigma_t^2 - \widehat{\sigma}_t^2$, given by:

$$ MAD(e) \overset{\text{def}}{=} \frac{1}{n_1} \sum_{t=n_0+1}^{n} |e_t|, \qquad RMSE(e) \overset{\text{def}}{=} \sqrt{\frac{1}{n_1} \sum_{t=n_0+1}^{n} (e_t - \bar{e})^2} \qquad (38) $$

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. Our forecasting results are summarized in Tables 18 and 19. Similar results were obtained when using a recursive sample and are available on request.

---

[12] All NoVaS predictions were made without applying an autoregressive filter as all $W_t(\boldsymbol{\theta}^*)$ series were uncorrelated.

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 17 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 (computed using the appropriate target in each case) 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 for all four combinations of target distributions 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 is given in Figures 10 to 13. In these figures we have QQ plots for all the series and all four combinations of return distributions. It is apparent from these figures that normalization has been achieved in all cases examined.

A second noticeable result is the optimal lag length chosen by the different NoVaS specifications. In particular, we see from Table 17 that the optimal lag length is general much greater when using the normal distribution as a target and about three to four times the optimal lag length when using the uniform distribution as a target. In addition, 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 when using a normal distribution. The optimal value of a_0^* when using the uniform distribution as a target is about one-third to one-half which implies an approximate interval for the uniform distribution in the range of about ± 2.5 to about ± 2.8.

We now turn to the out-of-sample results on the forecasting performance of NoVaS , which are summarized in Tables 18 and 19. Both the NoVaS made forecasts and the GARCH-made forecasts easily outperform the simple MAD and RMSE benchmarks, with the exception of the MSFT series where the GARCH forecasts perform extremely poorly.[13] However, there are marked differences between the NoVaS forecasts and the GARCH forecasts: the NoVaS forecasts outperform the GARCH forecasts seven times. If one looks only at the mean GARCH forecasts then the NoVaS forecasts always outperform them. Its worth noting the the median GARCH forecasts offer

---

[13]Remember that our simulation results showed that the performance of a GARCH model could be way of 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.

substantial improvements over the mean GARCH forecasts, which supports the earlier claim in Politis (2003a,b, 2007) that the use of L_1-based forecasts can improve the forecasting performance of GARCH-type models - however, the median GARCH forecasts are still beaten by the NoVaS forecasts. Another interesting result is that in the case of the EFG series the best performing among the NoVaS forecasts is the one using the uniform target distribution and absolute returns. Notably, the EFG series has the highest kurtosis among the four series examined. Similarly good performance using absolute returns and the normal or uniform distribution we have for the MSFT series. We also note that the improvements over the mean and median GARCH forecasts that NoVaS offer can be quite substantial: excluding the MSFT series where the GARCH model does not perform well, the ratio of the evaluation measures of the best NoVaS forecast relative to the best GARCH forecast ranges across series from 0.79 to 1. Overall, our results suggest not only that NoVaS does outperform the GARCH forecasts but also that the standard approach of GARCH forecasting (based on the conditional mean) appears inferior to the median-based GARCH forecast.

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. All in all, the NoVaS -made forecasts are better than the GARCH-made forecasts. 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 contribute several new methodological approaches on the NoVaS transformation approach for volatility forecasting introduced by Politis (2003a,b, 2007) and show that it can be a flexible method for inference and prediction of volatility of financial returns. In particular: (a) we introduce an alternative target distribution (uniform); (b) we present a new method for volatility forecasting using NoVaS ; (c) we show that the NoVaS methodology is applicable in situations where (global) stationarity fails such as the cases of local stationarity and/or structural breaks;

(d) we show how to apply the NoVaS ideas in the case of returns with asymmetric distribution; and finally (e) we discuss the application of NoVaS to the problem of estimating value at risk (VaR). The NoVaS methodology allows for a flexible approach to inference and has immediate applications in the context of short time series and series that exhibit local behavior (e.g. breaks, regime switching etc.) We conduct an extensive simulation study on the predictive ability of the NoVaS approach and find that NoVaS forecasts lead to a much 'tighter' distribution of the forecasting performance measure for all data generating processes. Our empirical illustrations using four real datasets are also 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] Cai, Z., 2002. "Regression Quantiles for Time Series", *Econometric Theory*, 18, pp. 169-192.

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

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

[7] 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.

[8] Embrechts, P., Kuppelberg, C. and T. Mikosch, 1997. *Modeling Extremal Events*, Springer Verlag: Berlin.

[9] 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.

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

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

[12] Gannoun, A., Saracco, J. and K. Yu, 2003. "Nonparametric Prediction by Conditional Median and Quantiles", *Journal of Statistical Planning and Inference*, vol. 117, pp. 207 223.

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

[14] 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.

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

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

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

[18] 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.

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

[20] Jorion, P., 1997. *Value at Risk: The New Benchmark for Controlling Market Risk*, McGraw-Hill: Chicago.

[21] 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.

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

[23] 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.

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

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

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

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

[28] 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.

[29] 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.

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

[31] Politis, D. and D. Thomakos, 2006. "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.), Elsevier Publishers: Amsterdam.

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

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

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

Table 1. Absolute Moments of Implicit NoVaS Distributions

$$\mathsf{E}_j |u|^a \approx \int_{-100}^{100} |u|^a f_j(u, a_0) du \text{ for } j = 1, 2, 3, 4$$

	$a = 1$	$a = 2$	$a = 3$	$a = 4$
$\mathcal{N}(0, 1)$	0.80	1.00	1.59	3.00
$t_{(2)}$	1.39	7.90	194.4	9975.3
$f_1(u, 0.1)$	0.92	1.98	20.27	875.5
$f_2(u, 0.3)$	1.50	10.08	302.8	17559.4
$f_3(u, 0.55)$	1.33	7.27	176.96	9070.2
$f_4(u, 0.75)$	4.46	119.7	6339.6	427326.1

Notes: $f_i(u, a_0)$ correspond to the implied NoVaS distributions of equations (8) and (9).

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

Distributional Statistics for MAD

	$\bar{x}_f$	$\hat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
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
SQUT	0.24	0.15	0.14	0.16	0.18	0.20	0.25	0.36	2.01
ABNT	0.21	0.09	0.15	0.16	0.17	0.19	0.22	0.28	1.28
ABUT	0.25	0.13	0.15	0.17	0.19	0.21	0.25	0.35	1.86
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

	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}_3$	$\hat{P}_4$	$\hat{P}$
GARCH	0.93	0.59	0.66	0.58	0.93
M-GARCH	1.00	0.63	0.74	0.59	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 equation (37) in the main text.

4. $\bar{x}_f$ denotes the sample mean, $\hat{\sigma}_f$ denotes the sample std. deviation and Q_p denotes the pth 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, SQUT (ABUT) denotes NoVaS forecasts based on a uniform target distribution and squared (absolute) returns, GARCH and M-GARCH denote L_2 and L_1 based forecasts from a standard GARCH model.

Table 3. 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
SQUT	0.31	0.63	0.13	0.16	0.18	0.21	0.27	0.40	11.56
ABNT	0.28	0.47	0.14	0.17	0.18	0.20	0.25	0.36	8.02
ABUT	0.31	0.56	0.15	0.18	0.20	0.22	0.28	0.40	9.84
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.98	0.74	0.76	0.70	0.98
M-GARCH	0.99	0.84	0.87	0.75	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 2.

Table 4. 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
SQUT	0.10	0.09	0.06	0.07	0.08	0.08	0.10	0.13	1.28
ABNT	0.15	0.16	0.08	0.09	0.10	0.11	0.14	0.20	1.67
ABUT	0.13	0.08	0.08	0.09	0.09	0.11	0.13	0.18	0.82
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.98	1.00	0.85	0.91	1.00
M-GARCH	0.99	1.00	0.98	0.99	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 2.

Table 5a. 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
SQUT	0.31	0.16	0.19	0.22	0.24	0.27	0.33	0.40	2.27
ABNT	0.30	0.14	0.21	0.23	0.25	0.27	0.31	0.37	1.95
ABUT	0.33	0.15	0.21	0.24	0.26	0.29	0.35	0.42	1.79
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.94	0.62	0.62	0.58	0.94
M-GARCH	1.00	0.75	0.73	0.60	1.00

Notes:

1. The model being simulated is a two-state Markov switching GARCH(1,1)
$h_t^2 = \sum_{s=1}^2 \mathbf{1}\{P(S_t = s)\}\left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s(X_{t-1} - \mu_s)^2\right]$.

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 2.

Table 5b. 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
SQUT	0.24	0.13	0.15	0.17	0.18	0.21	0.25	0.32	1.89
ABNT	0.26	0.15	0.17	0.19	0.20	0.23	0.26	0.33	1.95
ABUT	0.27	0.15	0.17	0.20	0.21	0.23	0.28	0.35	2.20
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.90	0.67	0.60	0.59	0.90
M-GARCH	1.00	0.92	0.75	0.69	1.00

Notes:

1. The model being simulated is a two-state Markov switching GARCH(1,1)
   $h_t^2 = \sum_{s=1}^{2} \mathbf{1}\{P(S_t = s)\} \left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s (X_{t-1} - \mu_s)^2 \right].$

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 2.

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

Distributional Statistics for MAD

	$\bar{x}_f$	$\hat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
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
SQUT	0.22	0.08	0.14	0.16	0.17	0.19	0.23	0.30	0.76
ABNT	0.24	0.10	0.17	0.19	0.20	0.22	0.25	0.32	0.98
ABUT	0.25	0.10	0.17	0.19	0.20	0.22	0.26	0.34	0.89
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

	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}_3$	$\hat{P}_4$	$\hat{P}$
GARCH	0.91	0.60	0.55	0.53	0.91
M-GARCH	1.00	0.93	0.67	0.62	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 2.

Table 7. 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
SQUT	0.19	0.07	0.12	0.14	0.15	0.17	0.19	0.26	0.86
ABNT	0.18	0.05	0.14	0.15	0.16	0.16	0.18	0.22	0.62
ABUT	0.20	0.06	0.14	0.16	0.17	0.18	0.20	0.26	0.82
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.76	0.54	0.55	0.52	0.76
M-GARCH	1.00	0.59	0.61	0.54	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 2.

Table 8. 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
SQUT	0.24	0.08	0.13	0.19	0.21	0.23	0.26	0.29	1.70
ABNT	0.23	0.11	0.13	0.18	0.20	0.22	0.25	0.28	2.29
ABUT	0.27	0.09	0.15	0.22	0.24	0.26	0.29	0.32	1.84
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.90	0.64	0.70	0.57	0.91
M-GARCH	0.97	0.88	0.99	0.64	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 2.

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

Distributional Statistics for MAD

	$\bar{x}_f$	$\hat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
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
SQUT	0.26	0.26	0.12	0.14	0.16	0.19	0.25	0.38	3.54
ABNT	0.21	0.13	0.14	0.15	0.16	0.18	0.21	0.29	1.42
ABUT	0.26	0.21	0.14	0.15	0.17	0.20	0.26	0.35	2.23
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

	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}_3$	$\hat{P}_4$	$\hat{P}$
GARCH	0.43	0.08	0.13	0.09	0.43
M-GARCH	0.86	0.14	0.35	0.10	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 2.

Table 10. 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
SQUT	0.22	0.32	0.09	0.11	0.12	0.15	0.22	0.29	3.32
ABNT	0.22	0.32	0.10	0.12	0.13	0.15	0.21	0.28	3.69
ABUT	0.24	0.38	0.10	0.12	0.13	0.16	0.24	0.31	4.26
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.86	0.34	0.35	0.27	0.86
M-GARCH	0.96	0.47	0.42	0.26	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 9.

Table 11. 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
SQUT	0.11	0.08	0.07	0.08	0.08	0.09	0.11	0.14	1.06
ABNT	0.15	0.19	0.08	0.09	0.10	0.11	0.14	0.20	3.36
ABUT	0.14	0.12	0.08	0.09	0.10	0.11	0.13	0.18	1.85
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.89	0.88	0.52	0.60	0.98
M-GARCH	0.96	0.96	0.91	0.93	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 9.

Table 12a. 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
SQUT	0.32	0.30	0.16	0.20	0.22	0.25	0.31	0.43	4.24
ABNT	0.32	0.33	0.18	0.21	0.23	0.25	0.30	0.40	5.09
ABUT	0.34	0.35	0.18	0.22	0.24	0.28	0.33	0.46	4.68
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.42	0.11	0.14	0.10	0.42
M-GARCH	0.85	0.29	0.30	0.15	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 9.

Table 12b. 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
SQUT	0.26	0.26	0.13	0.16	0.17	0.19	0.24	0.37	2.96
ABNT	0.27	0.26	0.16	0.18	0.19	0.21	0.25	0.37	4.27
ABUT	0.28	0.25	0.15	0.18	0.19	0.22	0.27	0.40	3.31
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.45	0.21	0.18	0.15	0.46
M-GARCH	0.87	0.50	0.36	0.28	0.89

Notes:

1. The model being simulated is a two-state Markov switching GARCH(1,1)
   $h_t^2 = \sum_{s=1}^{2} \mathbf{1}\left\{\mathsf{P}(S_t = s)\right\}\left[\omega_s + \alpha_s h_{t-1}^2 + \beta_s(X_{t-1} - \mu_s)^2\right].$

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 9.

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

Distributional Statistics for MAD

	$\bar{x}_f$	$\hat{\sigma}_f$	$Q_{0.00}$	$Q_{0.10}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.90}$	$Q_{1.00}$
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
SQUT	0.22	0.15	0.13	0.15	0.16	0.18	0.22	0.32	1.68
ABNT	0.25	0.17	0.16	0.17	0.19	0.21	0.24	0.30	1.84
ABUT	0.25	0.18	0.16	0.17	0.19	0.21	0.26	0.33	1.92
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

	$\hat{P}_1$	$\hat{P}_2$	$\hat{P}_3$	$\hat{P}_4$	$\hat{P}$
GARCH	0.47	0.17	0.14	0.13	0.47
M-GARCH	0.91	0.52	0.31	0.24	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 9.

Table 14. 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
SQUT	0.18	0.10	0.12	0.13	0.14	0.15	0.18	0.23	1.26
ABNT	0.17	0.06	0.13	0.14	0.14	0.15	0.17	0.20	0.81
ABUT	0.18	0.09	0.13	0.14	0.15	0.16	0.19	0.23	1.20
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.24	0.08	0.09	0.08	0.24
M-GARCH	0.77	0.15	0.19	0.09	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 9.

Table 15. 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
SQUT	0.25	0.21	0.11	0.17	0.19	0.23	0.27	0.32	4.57
ABNT	0.24	0.28	0.10	0.16	0.18	0.21	0.25	0.30	6.04
ABUT	0.28	0.22	0.13	0.19	0.22	0.25	0.30	0.34	4.84
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

	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}$
GARCH	0.36	0.08	0.17	0.05	0.40
M-GARCH	0.84	0.52	0.73	0.18	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 9.

Table 16. 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 17. 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
SQUT	0.385	0.001	0.330	8	0.944	0.991
ABNT	0.070	0.000	0.078	27	0.989	0.996
ABUT	0.462	0.000	0.380	7	0.944	0.988
MSFT monthly						
SQNT	0.175	0.000	0.171	15	0.916	0.988
AQUT	0.506	0.000	0.404	7	0.841	0.995
ABNT	0.251	0.000	0.231	12	0.916	0.986
ABUT	0.531	0.000	0.422	6	0.841	0.993
USD/Yen daily						
SQNT	0.062	0.000	0.071	29	0.978	0.999
SQUT	0.404	0.000	0.341	8	0.926	1.000
ABNT	0.121	0.000	0.124	20	0.978	0.999
ABUT	0.486	0.000	0.393	7	0.926	0.998
EFG daily						
SQNT	0.089	0.007	0.096	24	0.943	0.999
SQUT	0.460	0.000	0.378	7	0.872	0.999
ABNT	0.171	0.000	0.166	16	0.943	0.999
ABUT	0.540	0.000	0.427	6	0.872	0.998

Notes:

1. b^*, a_0^* and p^* denote the optimal exponential constant, first coefficient and implied lag length.

2. $D_n(\boldsymbol{\theta}^*)$ is the value of the objective function based on kurtosis matching.

3. QQ_X and QQ_W denote the QQ correlation coefficient of the original series and the transformed series respectively.

Table 18. Mean Absolute Deviation (MAD)of Forecast Errors

Series	Naïve	SQNT	SQUT	ABNT	ABUT	Mean GARCH	Median GARCH
S&P500, monthly	0.152	0.118	0.120	0.134	0.136	0.139	0.157
MSF, monthly	1.883	1.030	0.913	0.551	0.661	43.28	23.67
USD/Yen, daily	0.026	0.016	0.020	0.018	0.021	0.022	0.016
EFG, daily	0.251	0.143	0.117	0.120	0.112	0.225	0.141

Table 19. Root Mean-Squared (RMSE)of Forecast Errors

Series	Naïve	SQNT	SQUT	ABNT	ABUT	Mean GARCH	Median GARCH
S&P500, monthly	0.243	0.206	0.183	0.206	0.186	0.224	0.232
MSFT, monthly	0.530	1.552	2.096	0.951	1.505	162.0	89.17
USD/Yen, daily	0.031	0.028	0.033	0.028	0.031	0.030	0.029
EFG, daily	0.227	0.208	0.196	0.194	0.175	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 monthly series) and by 1000 (USD/Yen and EFG series) respectively.

4. SQNT, SQUT, ABNT and ABUT denote NoVaS made forecasts based on square returns and normal/uniform target and based on absolute returns and normal/uniform target.

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.
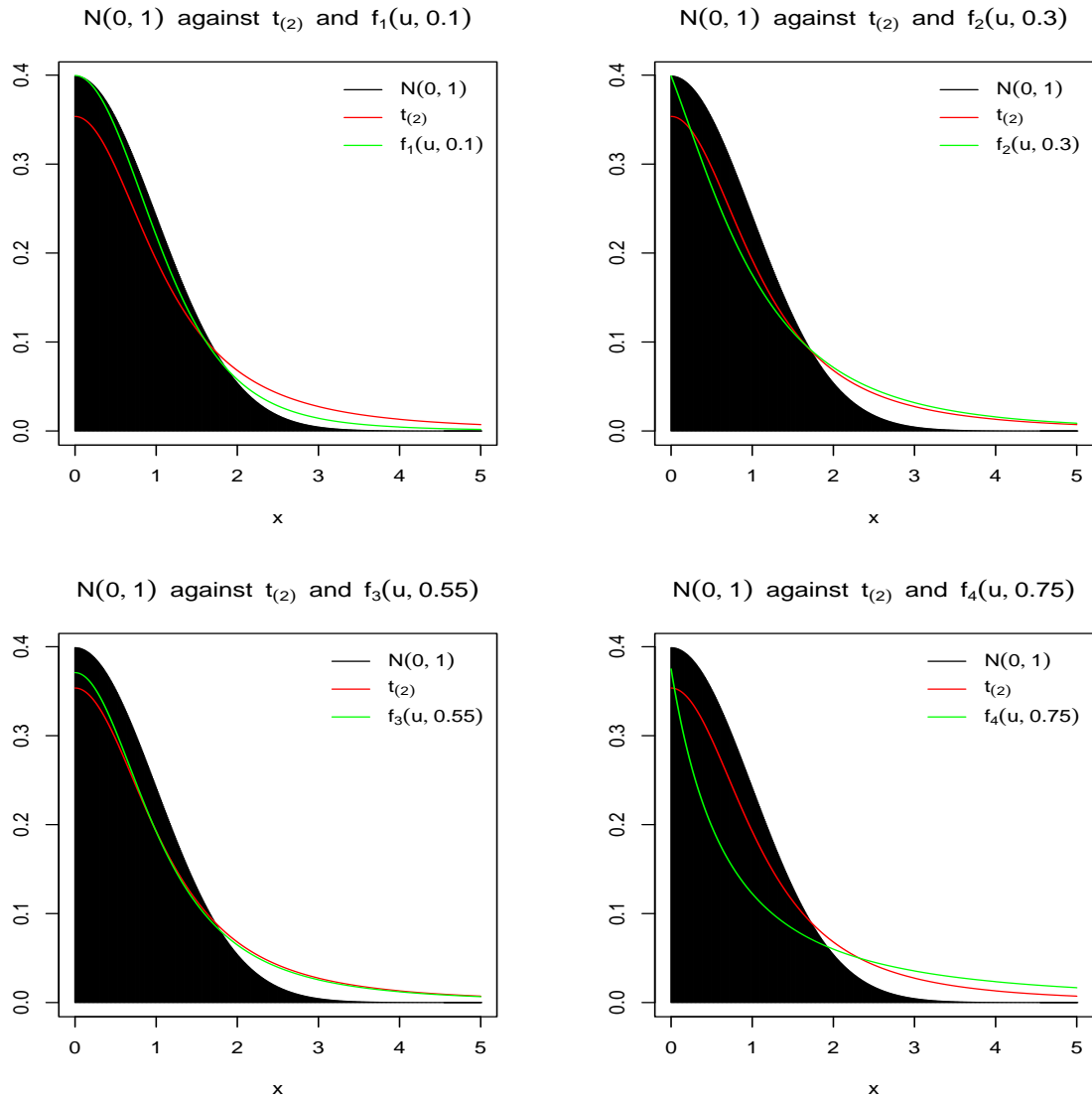
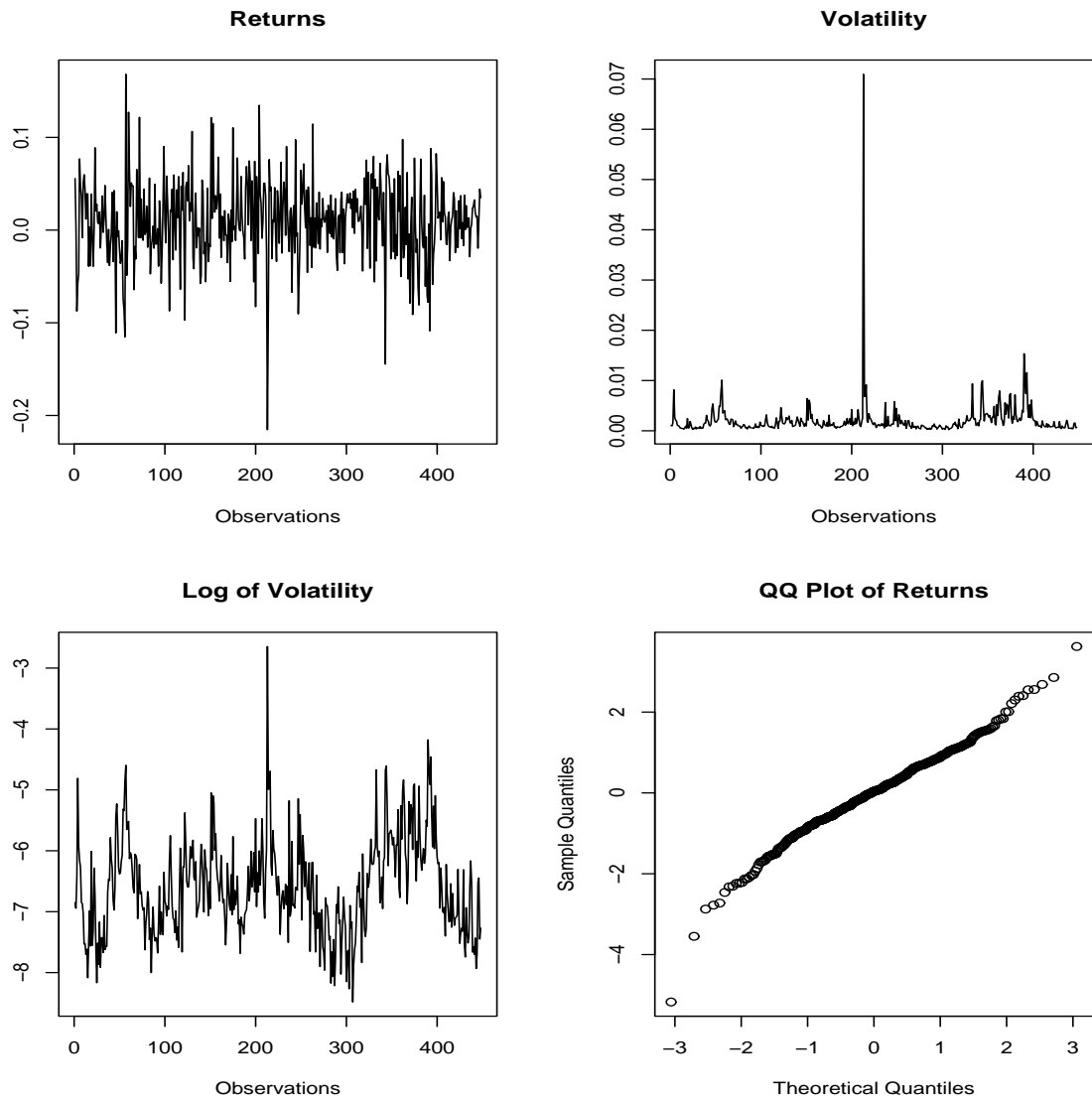Figure 1: Implied NoVaS distributions compared to the $\mathcal{N}(0,1)$ and the $t_{(2)}$ distributions

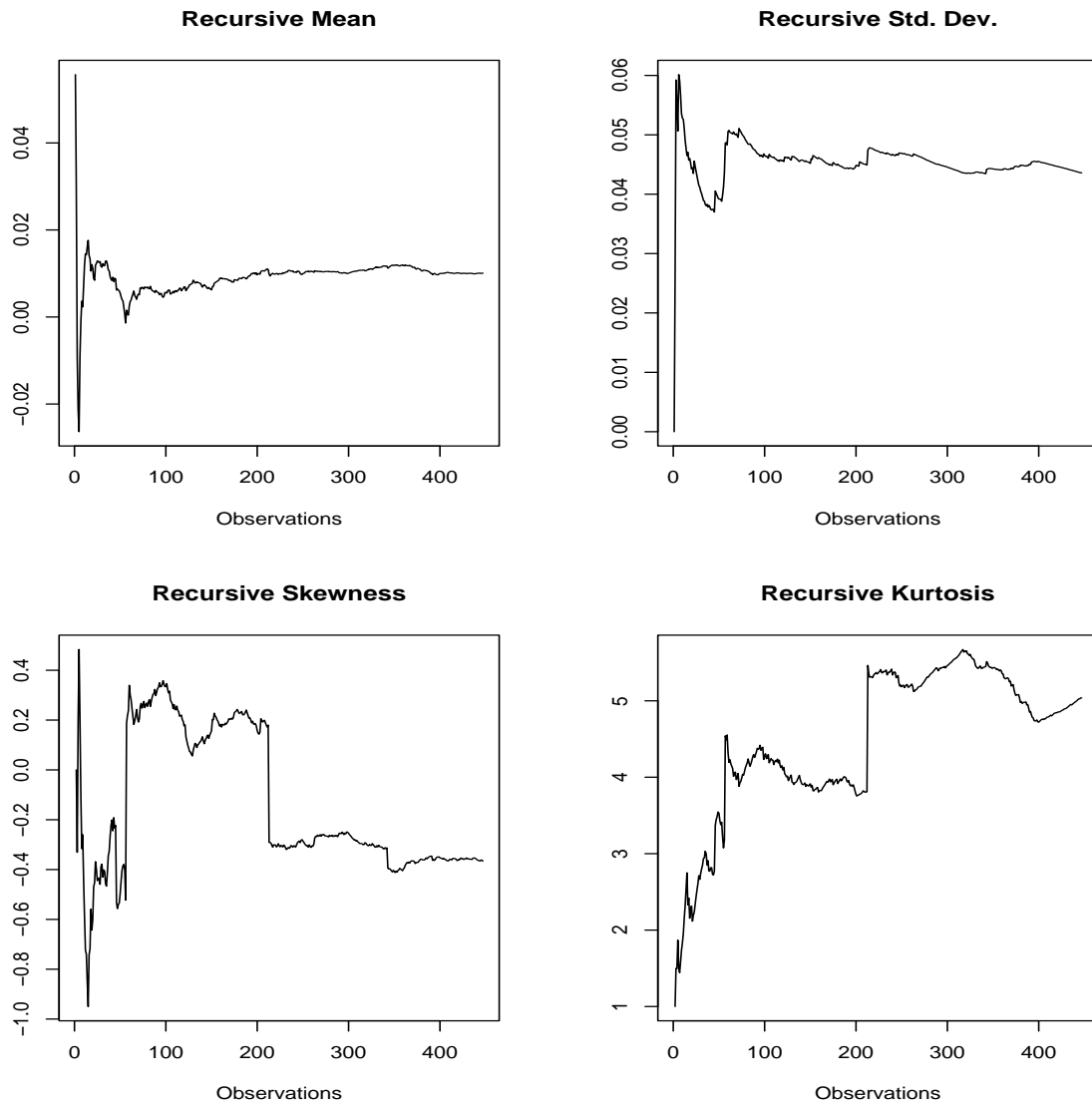Figure 2: Return, volatility and QQ plots for the monthly S&P500 series

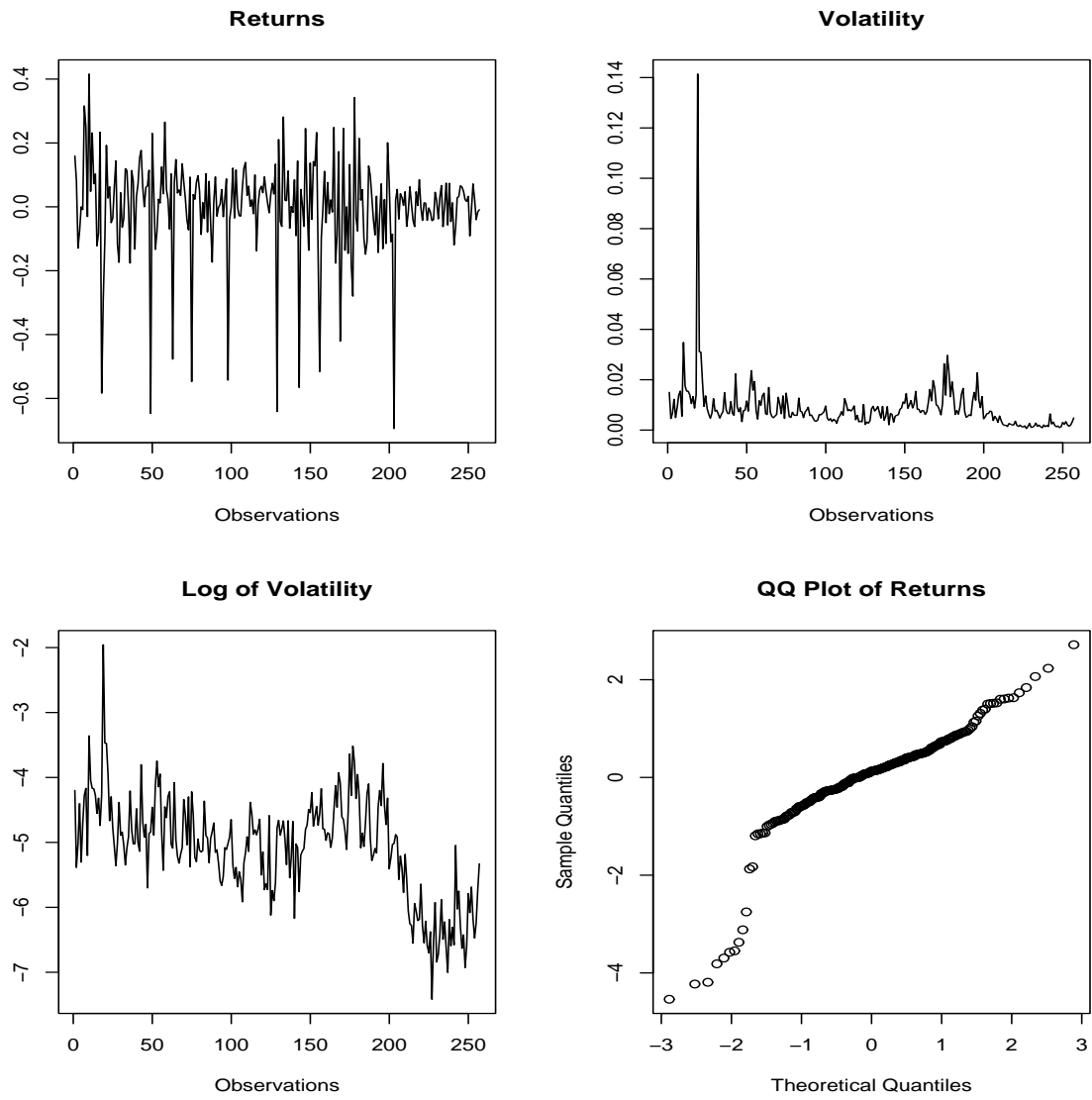Figure 3: Recursive moments for the monthly S&P500 series

56

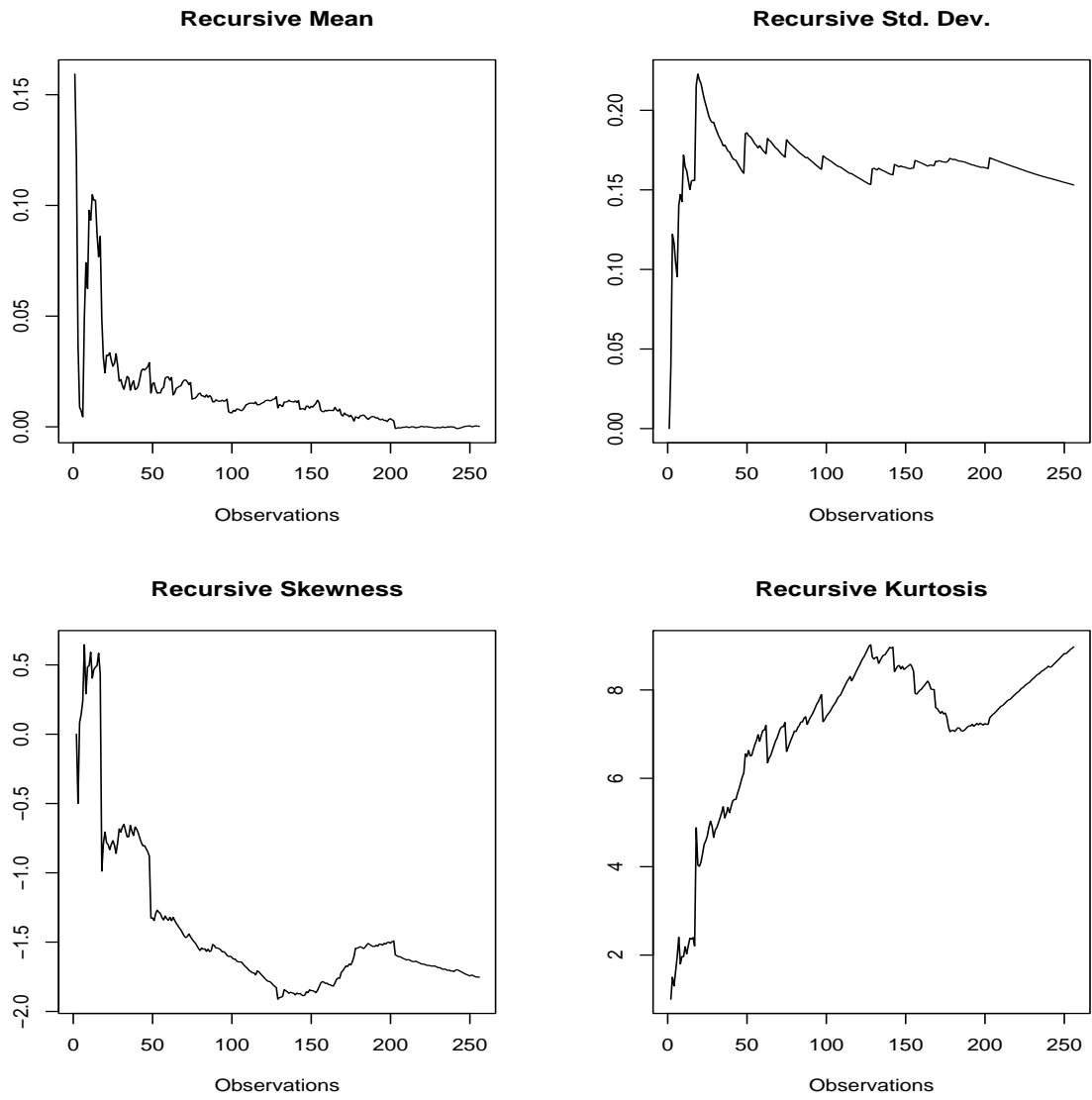Figure 4: Return, volatility and QQ plots for the monthly MSFT series

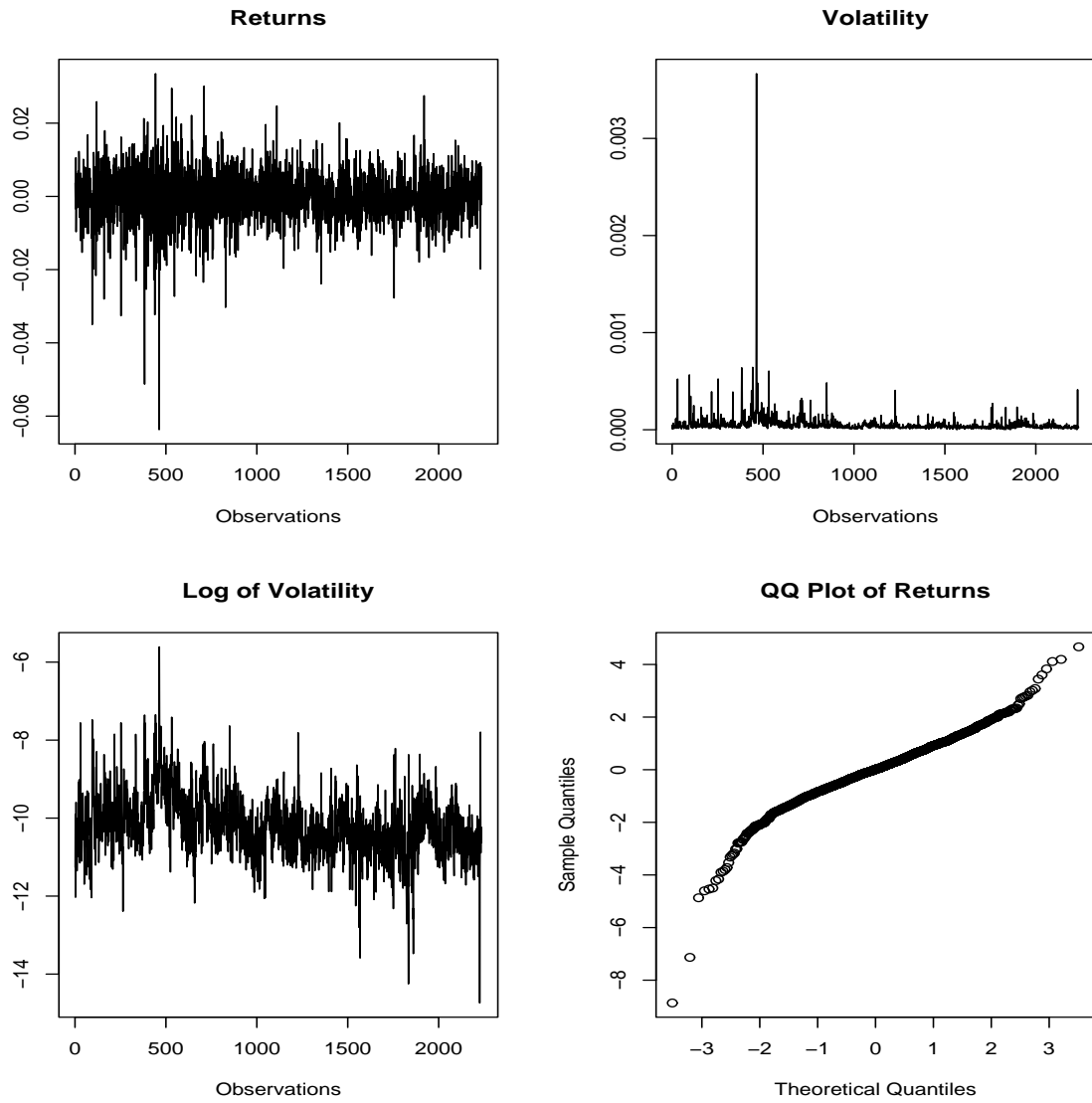Figure 5: Recursive moments for the monthly MSFT series

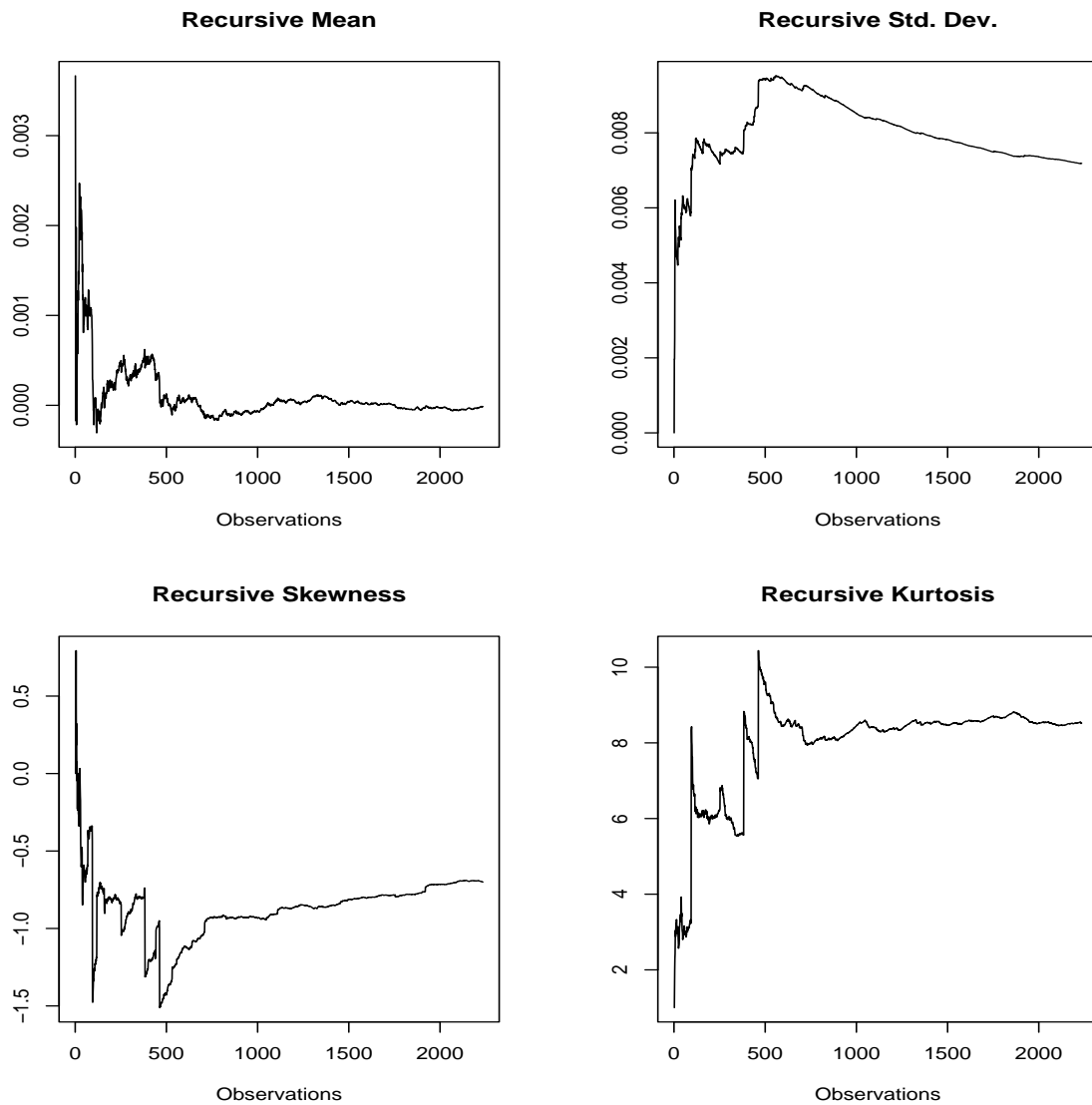Figure 6: Return, volatility and QQ plots for the daily USD/Yen series

Figure 7: Recursive moments for the daily USD/Yen series
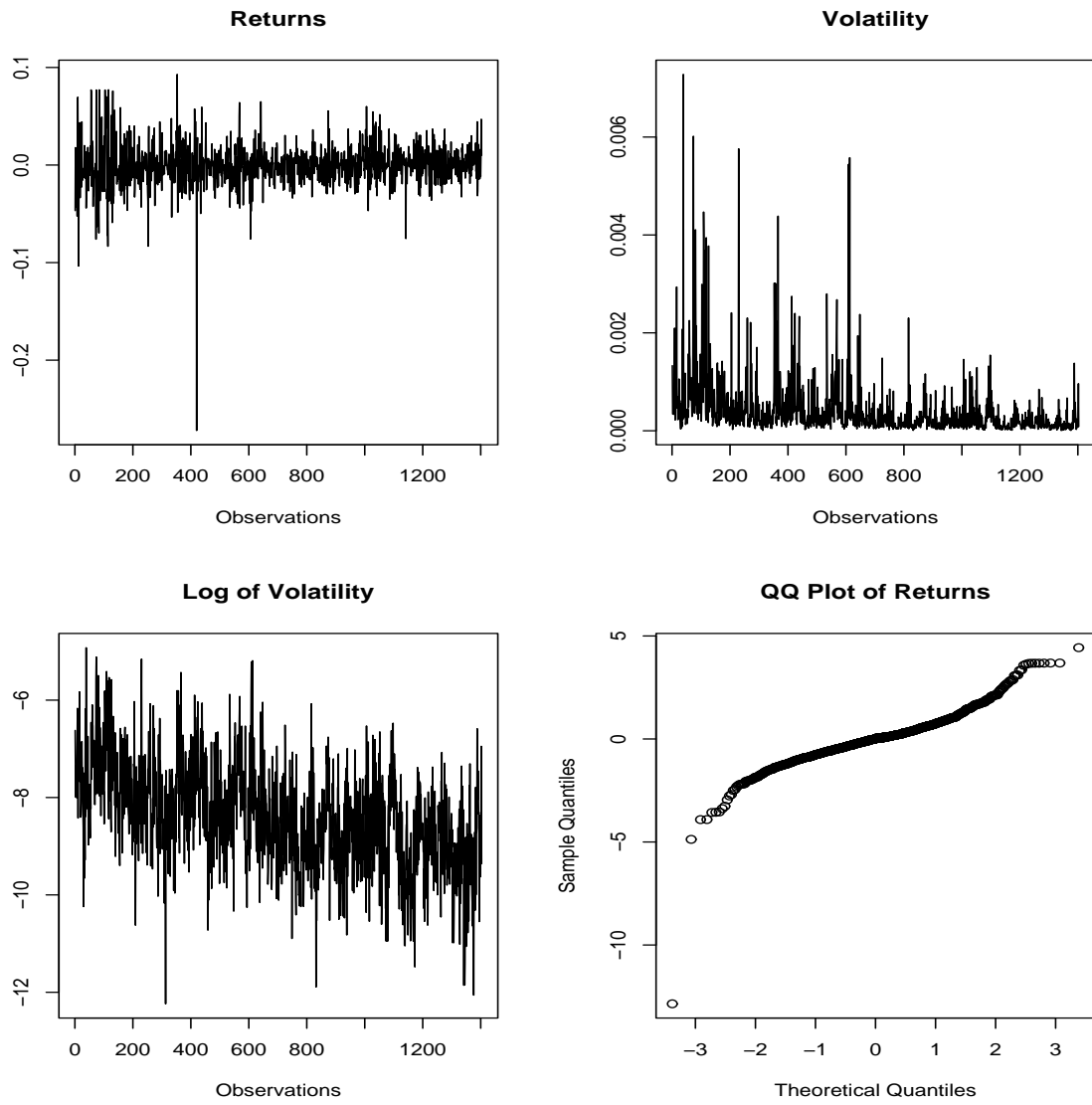
Figure 8: Return, volatility and QQ plots for the daily EFG series
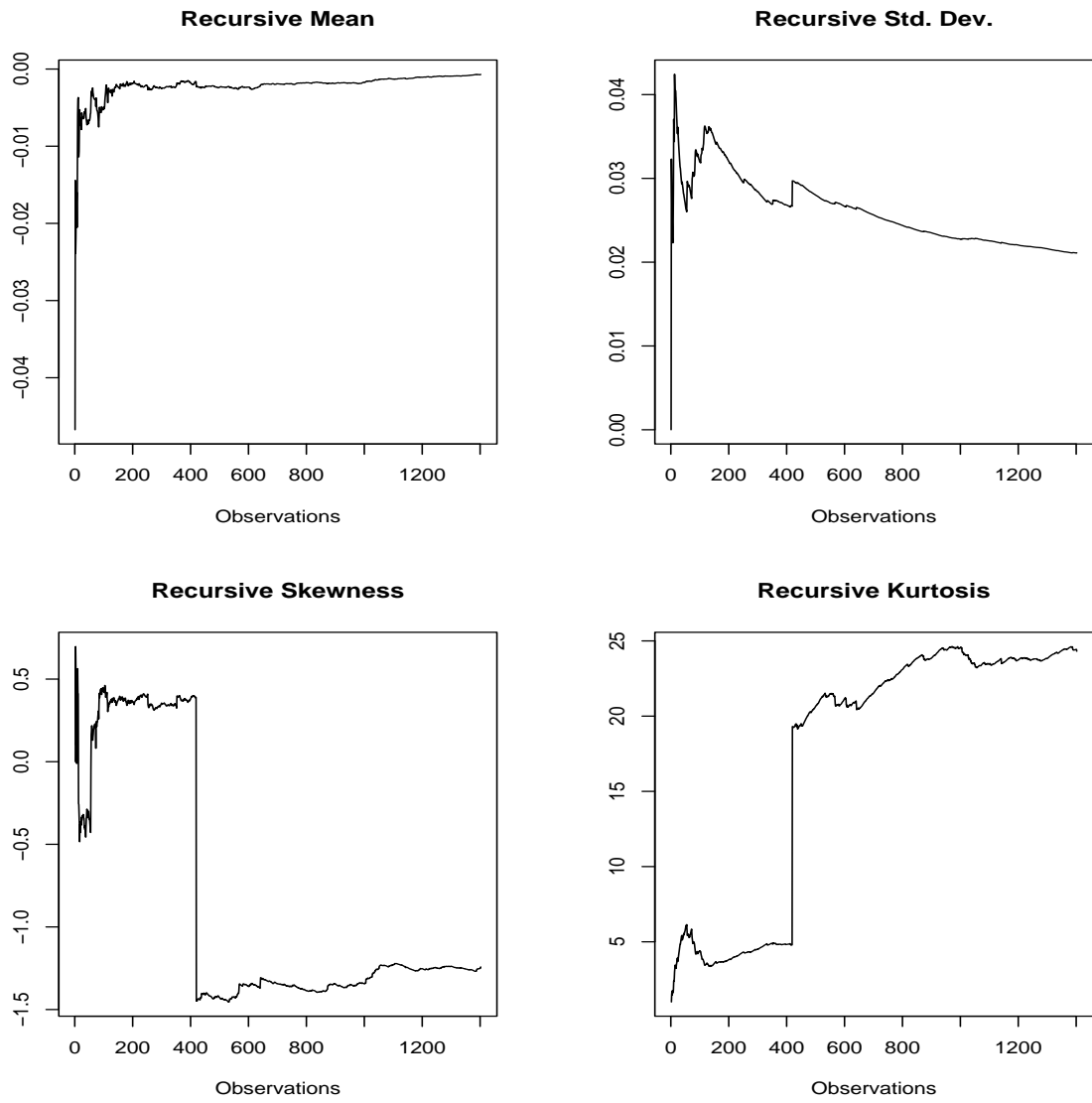
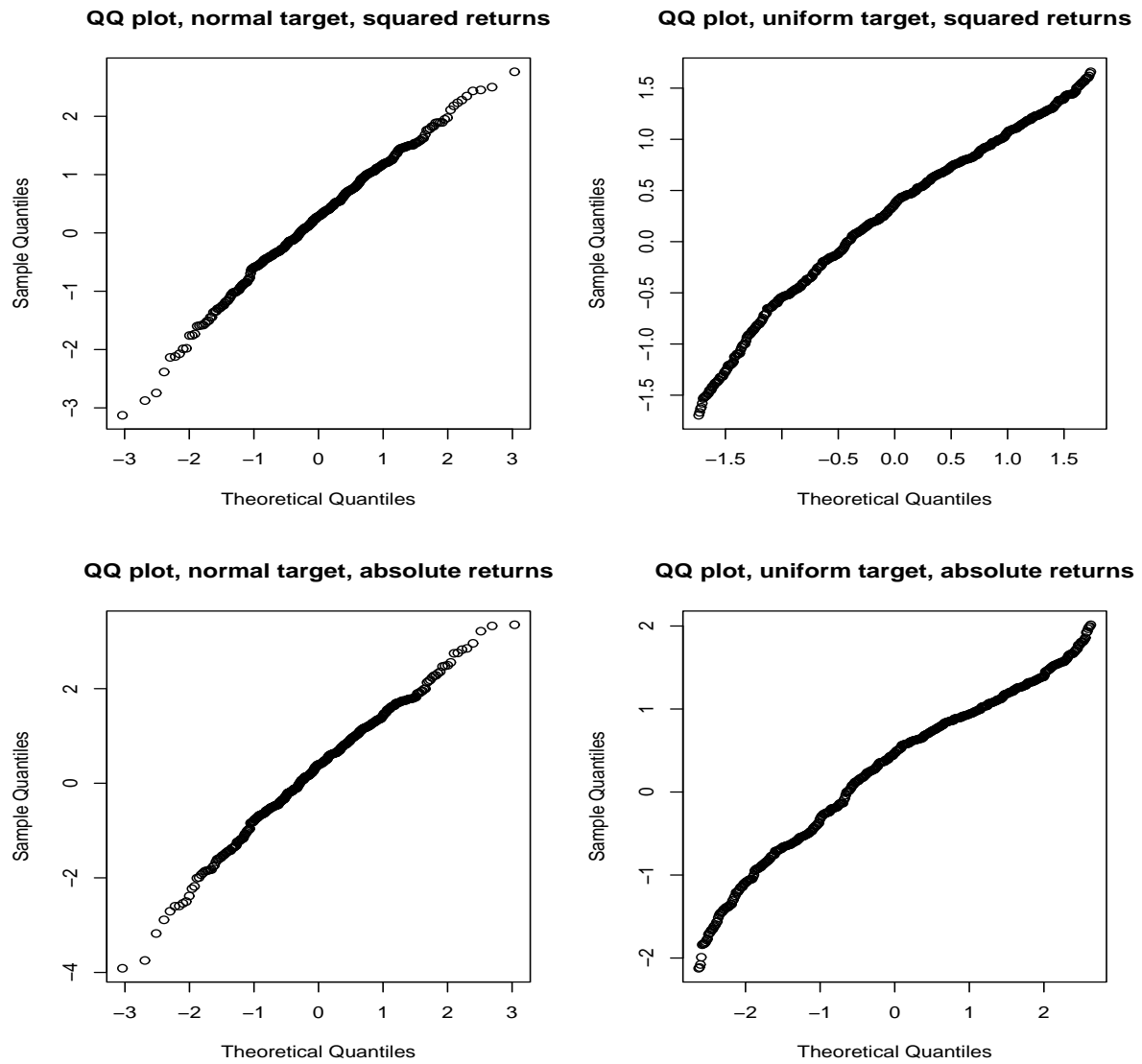Figure 9: Recursive moments for the daily EFG series

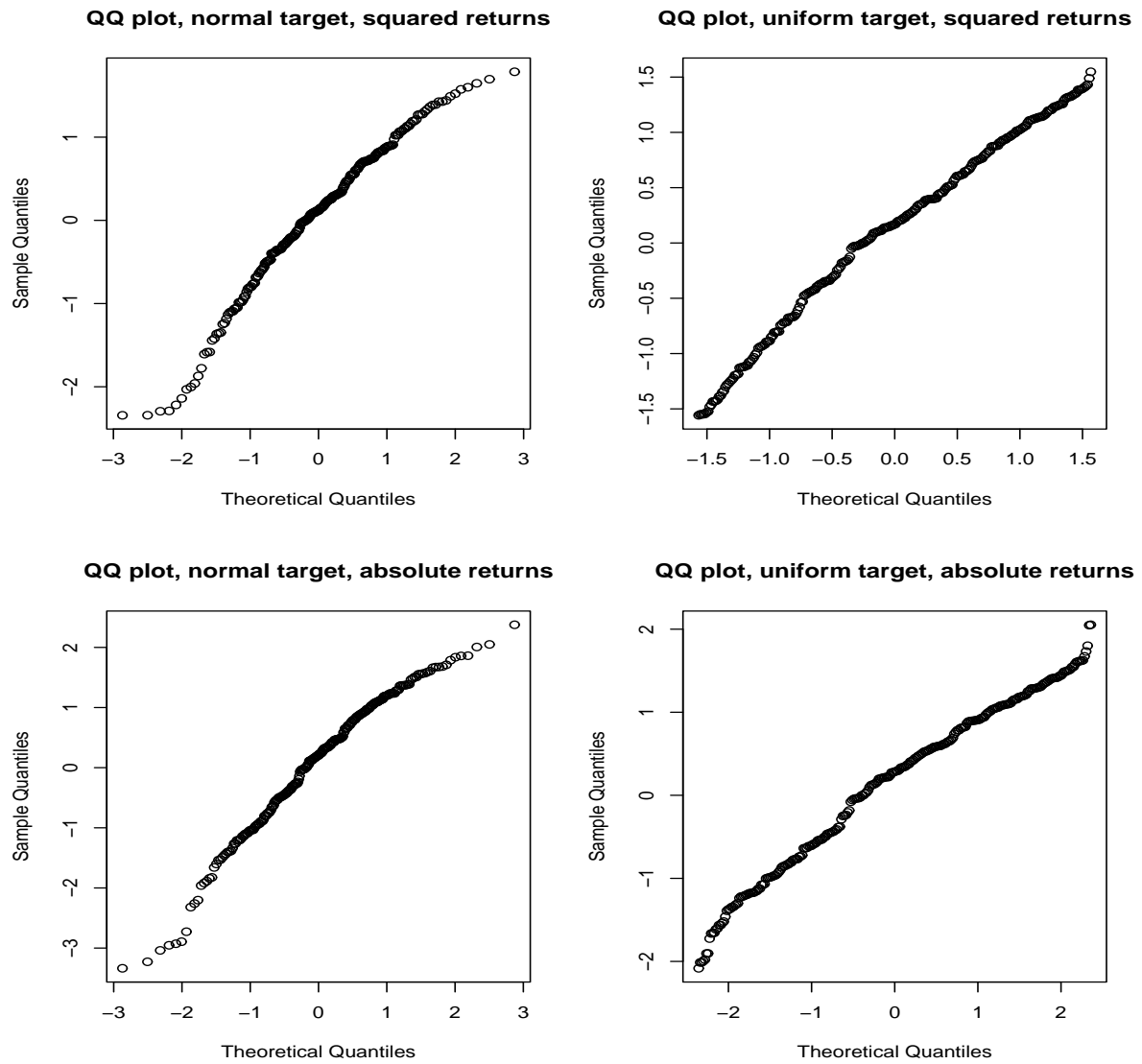Figure 10: QQ plots of the NoVaS -transformed W series for the monthly S&P500 series

Figure 11: QQ plots of the NoVaS -transformed W series for the monthly MSFT series

QQ plot, normal target, squared returns

QQ plot, uniform target, squared returns

QQ plot, normal target, absolute returns

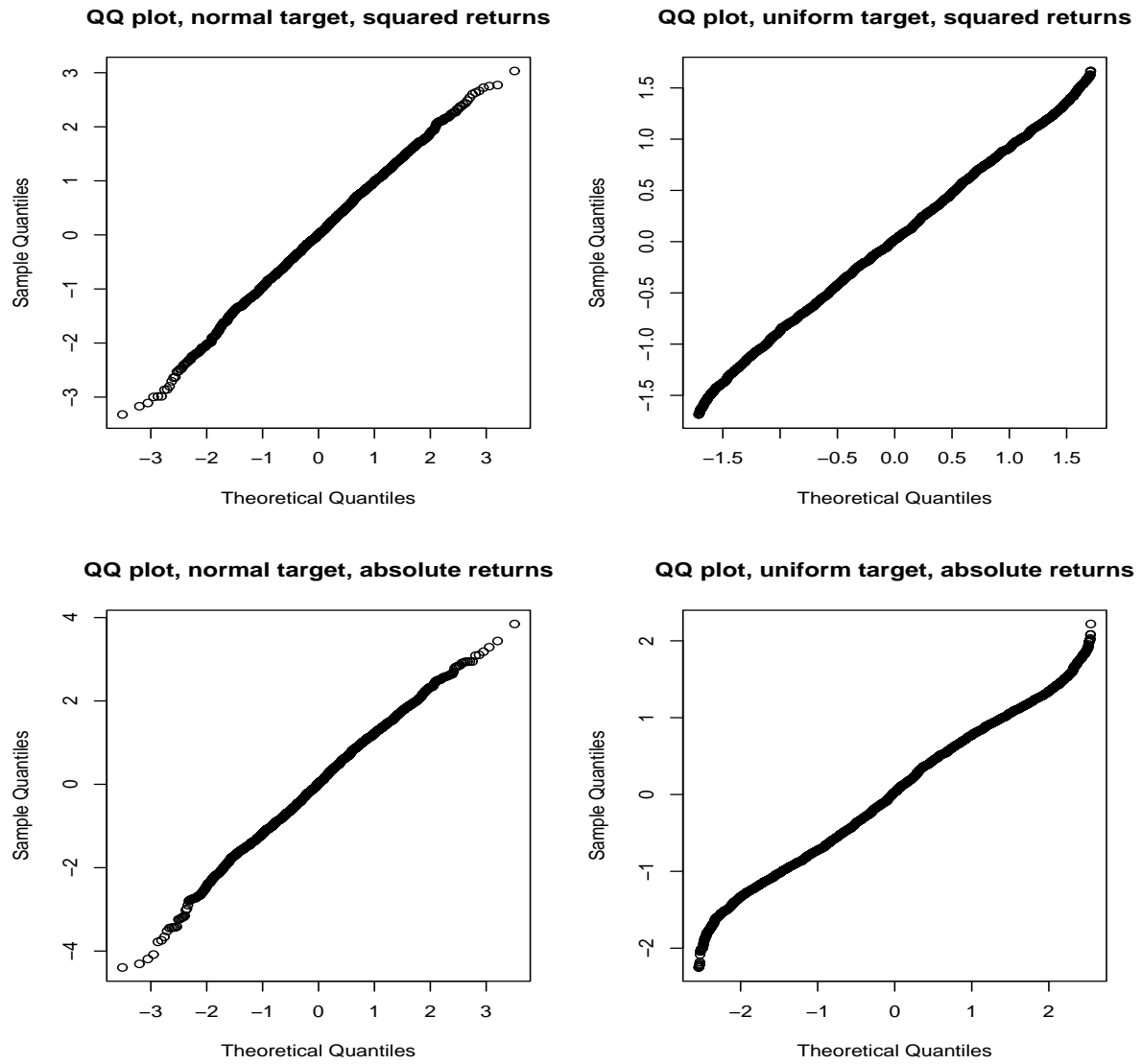**QQ plot, uniform target, absolute returns**

Figure 12: QQ plots of the NoVaS -transformed W series for the daily USD/Yen series
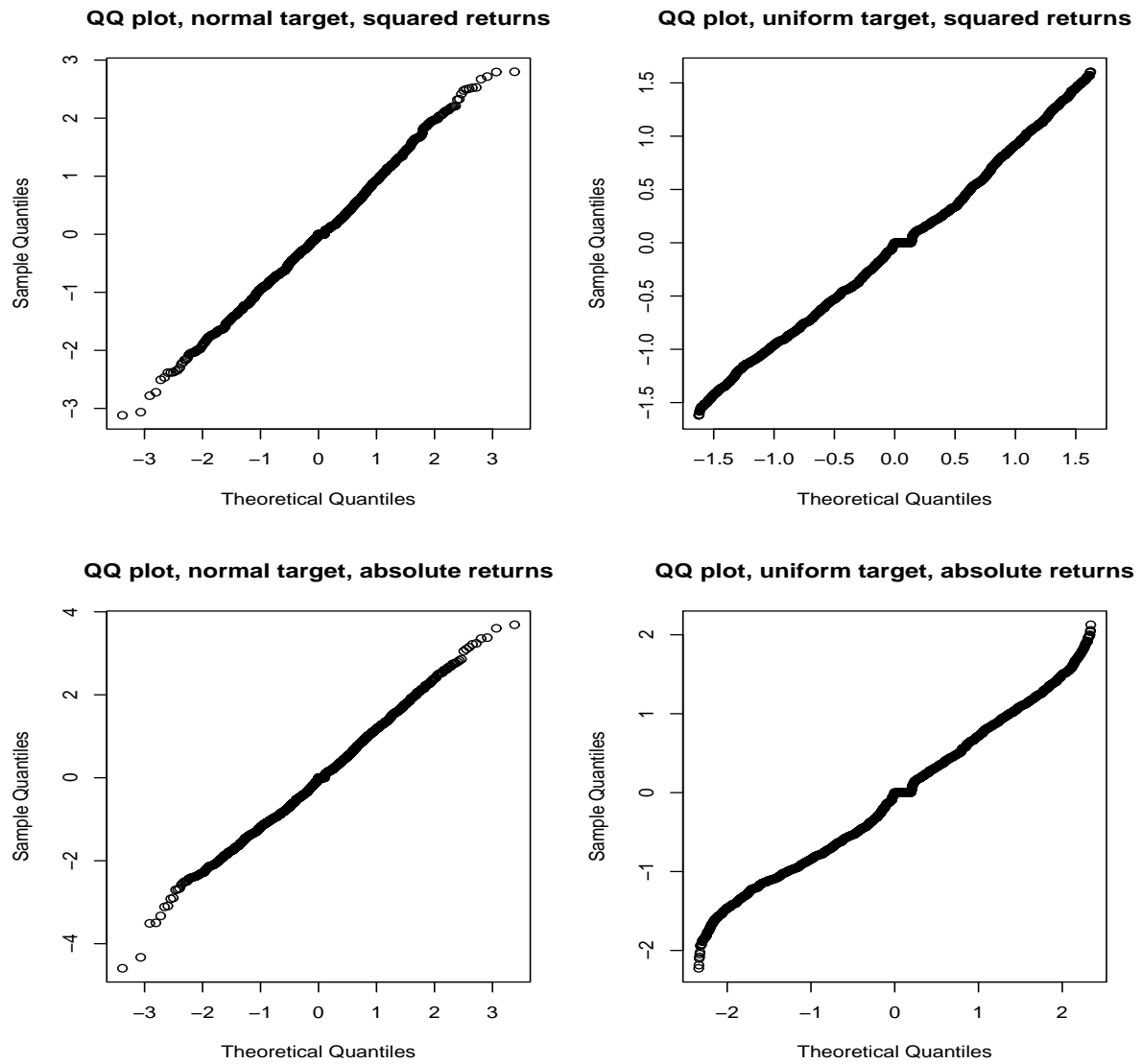
Figure 13: QQ plots of the NoVaS -transformed W series for the daily EFG series