

# Financial Time Series and Volatility Prediction using NoVaS Transformations

Dimitris N. Politis\*   Dimitrios D. Thomakos†

December 21, 2006

## Abstract

We extend earlier work on the NoVaS transformation approach introduced by Politis (2003a,b). The proposed approach is model-free and especially relevant when making forecasts in the context of model uncertainty and structural breaks. We introduce a new implied distribution in the context of NoVaS, a number of additional methods for implementing NoVaS and we examine the relative forecasting performance of NoVaS for making volatility predictions using real and simulated time series. We pay particular attention to data generating processes with varying coefficients and structural breaks. Our results clearly indicate that the NoVaS approach outperforms GARCH-made forecasts models in all cases we examined, except (as expected) when the data generating process was itself a GARCH model.

*Keywords:* volatility, forecasting, financial returns, model uncertainty.

*JEL Classification:* C14, C16, C22, C53, G10.

---

\*Department of Mathematics and Department of Economics, University of California, San Diego, 92093 USA. Email: politis@math.ucsd.edu, tel. +1-858-534-5861, fax: +1-858-534-5273.

†Corresponding author. Department of Economics, University of Peloponnese, Tripolis Campus, 22100 Greece. Email: thomakos@uop.gr, tel. +30-2710-230132, fax: +30-2710-230139.

# 1 Introduction

Making accurate predictions for the volatility of financial returns is an important part of applied financial research. In this paper we extend earlier work on a novel approach for making volatility predictions that was proposed by Politis (2003a,b). The proposed method is based on an exploratory data analysis idea, is model-free and is especially relevant when making forecasts in the context of model uncertainty and structural breaks. The method is called NoVaS , short for normalizing and variance stabilizing transformation. NoVaS is completely data-based, 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) and the conditional variance (volatility) of the series under study. Hence its usefulness under a variety of contexts where we do not know *a priori* which parametric family of models is appropriate for our data. Because of its flexibility, the NoVaS approach can easily handle, in addition to model uncertainty and structural breaks, arbitrary forms of nonlinearity in returns and volatility.

The original development of the NoVaS approach was made in Politis (2003a,b), in the context of volatility prediction and with the popular ARCH model with normal innovations as its starting point. In these papers the problem of prediction in a NoVaS context was addressed using the  $L_1$  prediction for the special case of a single, parametric expression for the dispersion of the returns (a modified ARCH equation) and standardization to normality. In the paper at hand we make a number of additional contributions: we introduce a new implied distribution in the context of NoVaS , we present a number of additional methods for implementing NoVaS and making forecasts, and we examine the relative forecasting performance of NoVaS for making volatility predictions using real and simulated time series.

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 semipara-

metric 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 prediction and the evaluation of volatility forecasts is very large and rapidly expanding. We can only selectively mention certain relatively recent papers (in chronological order) that are related to the problems we address: Mikosch and Starica (2000) for change in structure in time series and GARCH modeling; Peng and Yao (2003) for robust LAD estimation of GARCH models; Poon and Granger (2003) for assessing the forecasting performance of various volatility models; 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. Finally, 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). Of course this list is by no means complete.

The rest of the paper is organized as follows: in section 2 we present the general development of the NoVaS approach, introducing the NoVaS transformation and the implied NoVaS distributions; in section 3 we present the NoVaS distributional matching,

including parametrization and optimization; in section 4 we present NoVaS -based forecasting; in section 5 we present results from a detailed empirical illustration of the NoVaS approach; finally, in section 6 we offer some concluding remarks.

## 2 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. We 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.
2.  $X_t$  has time-varying conditional variance (volatility), denoted by  $\sigma_t^2 \stackrel{\text{def}}{=} \mathbf{E} [X_t^2 | \mathcal{F}_{t-1}]$ ,  $\mathcal{F}_{t-1} \stackrel{\text{def}}{=} \sigma(X_{t-1}, X_{t-2}, \dots)$ , that exhibits strong dependence.
3.  $X_t$  is dependent although it possibly exhibits low or no autocorrelation.

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 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 by non-Gaussianity, i.e. high volatility and dependence: 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 sequence is variance stabilization that takes care of the time-varying conditional variance property of the returns. We construct an empirical measure of the time *time-localized* variance of  $X_t$  based on the information set  $\mathcal{F}_{t|t-p} \stackrel{\text{def}}{=} \{X_t, X_{t-1}, \dots, X_{t-p}\}$

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

where  $\alpha$  is a scalar control parameter,  $\mathbf{a} \stackrel{\text{def}}{=} (a_0, a_1, \dots, a_p)^\top$  is a  $(p+1) \times 1$  vector of control parameters and  $G(\cdot; \alpha, \mathbf{a})$  is to be specified.<sup>1</sup> Note that the first novel element here is the introduction of the current value of  $X_t$  in constructing  $\gamma_t$ ; this is a small but crucial difference in the NoVaS approach which is fully explained below when we talk about the implied distributions of the NoVaS methodology. The function  $G(\cdot; \alpha, \mathbf{a})$  can be expressed in a variety of ways, using a parametric or a semiparametric specification. To keep things simple we assume that  $G(\cdot; \alpha, \mathbf{a})$  is additive and takes the following form:

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

with the implied restrictions (to maintain positivity for  $\gamma_t$ ) 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  $\sigma^2 = \mathbf{E}[X_1^2]$ , 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).

**Remark 1.** The use of absolute returns, in addition to the more common squared returns, in constructing both the recursive estimator  $s_{t-1}$  and the empirical measure  $\gamma_t$

---

<sup>1</sup>See the discussion about the calibration of  $\alpha$  and  $\mathbf{a}$  in the next section.

was not considered in Politis (2003a,b). It can, however, be justified for several reasons. Robustness in the presence of outliers is an obvious one. In addition, the mean absolute deviation is proportional to the standard deviation for the symmetric distributions that will be of current interest, their differences being rather small. For example, for a standard normal distribution the mean absolute deviation is about 0.8 while for a  $t_5$  distribution the standard deviation is about 1.29 while the mean absolute deviation is about 0.95.

**Remark 2.** To account for the possibility of an asymmetric response of return volatility to past returns, another kind of ‘stylized fact’, we can modify equation (1) as follows:

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

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

The second step in the NoVaS transformation is to use  $\gamma_t$  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:

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

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  $W_t$  resemble 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.<sup>2</sup> However, we are not constrained to use only the standard normal as

---

<sup>2</sup>The standard normal has an added advantage that comes useful in prediction, namely that it implies optimal predictors that are linear.

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 part of the ‘distributional goodness of fit’ component of NoVaS .

**Remark 3.** 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 also consider distributional matching for *joint* distributions of  $W_t$ . It is shown in Politis (2003a,b) 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  $\lambda$ .

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  $\gamma_t$  depends on  $X_t$ , we would obtain that:

$$X_t = U_t A_{t-1} \quad (5)$$

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

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

for  $U_t$  and by:

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

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 what the distribution of  $U_t$  is, 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). In addition to that case we consider also the cases where  $g(z) = |z|$  and where the distribution of  $W_t$  is uniform.

To understand the implied distribution of  $U_t$  first note that the range of  $W_t$  is bounded. Using equation (4) it is straightforward 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  $\pm 3$  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  $[-1/\sqrt{a_0}, +1/\sqrt{a_0}]$ , for  $g(z) = z^2$ , and  $[-1/a_0, +1/a_0]$ , for  $g(z) = |z|$ .

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  $\gamma_t$  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  $U_t$ :

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

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

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

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.<sup>3</sup> 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.

### FIGURE 1 AROUND HERE

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 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  $a$  as long as  $a < 2$ , whereas  $f_2(u, a_0)$  and

---

<sup>3</sup>Basically,  $f_1(u, a_0)$  looks like a  $\mathcal{N}(0, 1)$  distribution for small  $u$  but has a  $t_{(2)}$ -type tail.

$f_4(u, a_0)$  have finite moments of order  $a$  as long as  $a < 1$ . 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.

### TABLE 1 AROUND HERE

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 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 4.** 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, we just illustrated that there can be potential return distributions like the  $f_i(u, a_0)$  where fourth moments do not exist! This 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 much preferred choice for comparing the forecasting performance of returns, squared returns and volatility.<sup>4</sup>

---

<sup>4</sup>See also the recent paper by Hansen and Lunde (2006) about the relevance of MSE in evaluating volatility forecasts. These authors make a strong case for using the squared loss function for performing

### 3 NoVaS distributional matching

#### 3.1 Parametrization

We next turn to the issue of parameter selection or calibration.<sup>5</sup> 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  $(\alpha, \mathbf{a})$  or  $(\alpha, \mathbf{a}, \mathbf{b})$  if we want to account for possible asymmetries. The rest of the discussion will be in terms of  $p$ ,  $\alpha$  and  $\mathbf{a}$ . See Remark 5 below for the case where  $\mathbf{b}$  is also present. The parameters  $\alpha$  and  $\mathbf{a}$  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 and 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  $j = 0, 1, \dots, p$ . These are the simple NoVaS weights proposed in Politis (2003a,b). An alternative allowing for greater weight to be placed on earlier lags is to consider exponential weights of the form:

$$a_j = \begin{cases} 1/\sum_{j=0}^p \exp(-bj) & \text{for } j = 0 \\ a_0 \exp(-bj) & \text{for } j = 1, 2, \dots, p \end{cases} \quad (10)$$

where  $b$  is a control parameter. These are the exponential NoVaS weights proposed in Politis (2003a,b). Figure 2 shows the squared frequency response of the NoVaS weights for the equal weighting and exponential weighting schemes for various values of the control parameter  $b$ .

#### FIGURE 2 AROUND HERE

---

and evaluating volatility forecasts.

<sup>5</sup>It is important to stress that there is no formal estimation taking place in NoVaS . The parameters are chosen so as to achieve *distributional matching* and the objective functions discussed in the next section reflect this aim.

The exponential weighting scheme allows for greater flexibility at the cost of one extra calibrating 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  $\boldsymbol{\theta} \stackrel{\text{def}}{=} (p, b) \mapsto (\alpha, \mathbf{a})$  it will be convenient in what follows to denote the studentized series as  $W_t \equiv W_t(\boldsymbol{\theta})$  rather than  $W_t \equiv W_t(\alpha, \mathbf{a})$ . 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 5.** 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  $\mathbf{b}$  we get the following parameter representation:

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

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

### 3.2 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]), 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.<sup>6</sup>

Let us consider the simplest case first, and one easily used in applications, moment-based 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:

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

where  $\bar{W}_n \stackrel{\text{def}}{=} (1/n) \sum_{t=1}^n W_t$  denotes 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  $\kappa^*$  denotes the theoretical kurtosis coefficient of the target 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) such a procedure will work in lieu of the intermediate value theorem. For  $p = 0$  we have that  $a_0 = 1$  and thus  $W_t = \text{sign}(X_t)$  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  $B \stackrel{\text{def}}{=} (b_{(1)}, b_{(2)}, \dots, b_{(M)})$ ,  $M > 0$ . Find the optimal value of  $b$ , say  $b^*$ , that solves  $\min_{b \in B} D_n(\boldsymbol{\theta})$  and compute

---

<sup>6</sup>Although distance-based measures are frequently used for assessing goodness-of-fit they are rarely used for parameter estimation or calibration. However, the idea of using a distance-based objective function for parameter selection is not new and goes back to Wolfowitz (1957).

the optimal parameter vector  $\mathbf{a}^*$  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  $C \stackrel{\text{def}}{=} (c_{(1)}, c_{(2)}, \dots, c_{(M)})$ .

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<sup>7</sup>, 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. However, we describe a couple of other potential objective functions below for completeness.

As an alternative to moment kurtosis matching one can use more robust measures: Kim and White (2004) recommend that one should consider robust measures of (skewness and) kurtosis that are essentially based on quantiles rather than moments. For assessing kurtosis Kim and White recommend four alternative measures to the sample kurtosis coefficient; the simplest such measure is that of Moors (1988) and is based on octiles:

$$\mathcal{K}_n^M(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{(\widehat{E}_7 - \widehat{E}_5) + (\widehat{E}_3 - \widehat{E}_1)}{(\widehat{E}_6 - \widehat{E}_2)} - \kappa^* \quad (13)$$

where  $\widehat{E}_i$  is the  $i$ th sample octile (i.e.  $E_i \stackrel{\text{def}}{=} F^{-1}(i/8)$  for  $i = 1, 2, \dots, 7$ ). The scaling in the denominator ensures that the measure is invariant under a linear transformation. One could substitute  $\mathcal{K}_n^M(\boldsymbol{\theta})$  in place of  $\mathcal{K}_n(\boldsymbol{\theta})$  in the objective function, i.e.  $D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} |\mathcal{K}_n^M(\boldsymbol{\theta})|$ .

---

<sup>7</sup>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 objective function that is based on the QQ-correlation coefficient can also be easily constructed. For any given values of  $\boldsymbol{\theta}$  compute the order statistics  $W_{(t)}$ ,  $W_{(1)} \leq W_{(2)} \leq \dots \leq W_{(n)}$ , and the corresponding quantiles of the target distribution, say  $Q_{(t)}$ , obtained from the inverse cdf. The squared correlation coefficient in the simple regression on the pairs  $[Q_{(t)}, W_{(t)}]$  is a measure of distributional goodness of fit and corresponds to the well known Shapiro-Wilks test for normality, when the target distribution is the standard normal. We now have that:

$$D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} 1 - \frac{[\sum_{t=1}^n (W_{(t)} - \bar{W}_n)(Q_{(t)} - \bar{Q}_n)]^2}{[\sum_{t=1}^n (W_{(t)} - \bar{W}_n)^2] \cdot [\sum_{t=1}^n (Q_{(t)} - \bar{Q}_n)^2]} \quad (14)$$

In a similar fashion one can construct an objective function that is based on the Kolmogorov-Smirnov statistic. 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, the optimal values of the parameters are clearly determined by the condition:

$$\boldsymbol{\theta}_n^* \stackrel{\text{def}}{=} \underset{\boldsymbol{\theta}}{\text{argmin}} D_n(\boldsymbol{\theta}) \quad (15)$$

**Remark 6.** An interesting point raised by the referee was as to whether one can possibly test for the presence of nonlinearities, leverage or breaks within the NoVaS context. While the method is not geared towards formal hypothesis testing (its not an issue in this model-free context), the above discussion indicates how one can make (at least an informal) comparison for the presence of such effects. Consider, for example, the presence of asymmetries. One can perform NoVaS with and without the asymmetries in the time-localized measure of the variance [using equations (2) and (3)] and then compare the values of the objective functions. A potential indication that asymmetries are indeed present would be a smaller value for the objective function when equation (3) is used.

**Remark 7.** The discussion so far was made under the assumption that the parameter  $\alpha$ , 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  $\alpha + \sum_{j=0}^p a_j = 1$ ), choose that value from the grid that optimizes out-of-sample predictive performance. See Politis (2003a,b) for more details.

**Remark 8.** It is important to stress that the specialized form that  $G(\cdot; \alpha, \mathbf{a})$  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  $g(z) = z^2$ :

$$G(X_t, X_{t-1}, \dots, X_{t-p}; 0, a_0) \stackrel{\text{def}}{=} a_0 X_t^2 + \xi_{t-1}(\mathbf{x}_{t-p}^2) \quad (16)$$

where  $\mathbf{x}_{t-p}^2 \stackrel{\text{def}}{=} (X_{t-1}^2, \dots, X_{t-p}^2)^\top$  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:

$$\widehat{\xi}_t(\mathbf{x}_{t-p}^2) \stackrel{\text{def}}{=} \sum_{j=1}^p a_j X_{t-j}^2 \quad (17)$$

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

$$a_j \stackrel{\text{def}}{=} \frac{K_h(X_{t-j}^2 - \mathbf{x}_{t-p}^2)}{\sum_{j=1}^p K_h(X_{t-j}^2 - \mathbf{x}_{t-p}^2)} - \frac{a_0}{p} \quad (18)$$

If we take  $p = p_n \rightarrow \infty$  as  $n \rightarrow \infty$  then, in large samples, the above approach should produce consistent estimates of  $\xi_{t-1}(\cdot)$ . 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.

## 4 NoVaS Forecasting

Once the NoVaS parameters are calibrated one can compute volatility forecasts. In fact, as Politis (2003a,b, 2004) 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 4). 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  $U_n^* \stackrel{\text{def}}{=} \{U_1(\boldsymbol{\theta}_n^*), \dots, U_n(\boldsymbol{\theta}_n^*)\}$ , 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  $\Pi_k [X|F]$  denote the  $k^{\text{th}}$  (regular or absolute) conditional power operator of the argument  $X$  given the argument  $F$ . For example,  $\Pi_1 [XF|F] = XF$ ,  $\Pi_2 [XF|F] = (X^2|F) \cdot F^2$  etc. Applying the power operator in the definition of the *implied* model of equation (5) at time  $n + 1$  we obtain:

$$\Pi_k [X_{n+1}|\mathcal{F}_n] = \Pi_k [U_{n+1}^*|\mathcal{F}_n] \Pi_k [A_n] \quad (19)$$

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:

$$\text{Med} [\Pi_k [X_{n+1}|\mathcal{F}_n]] = \text{Med} [\Pi_k [U_{n+1}^*|\mathcal{F}_n]] \Pi_k [A_n] \quad (20)$$

where  $\text{Med} [x]$  stands for the median of  $x$ . Therefore, what we are after is an estimate of

the conditional median of the series  $\Pi_k [U_{n+1}^* | \mathcal{F}_n]$ .<sup>8</sup>

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  $\text{Med} [\Pi_k [U_{n+1}^* | \mathcal{F}_n]]$  is the unconditional sample median of the appropriate power of the  $U_n^*$  series, namely  $\widehat{\text{Med}} [\Pi_k [U_n^* | \mathcal{F}_n]]$ . 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  $\widehat{\text{Med}} [\Pi_k [U_n^* | \mathcal{F}_n]]$  is still the best estimate of the conditional  $\text{Med} [\Pi_k [U_{n+1}^* | \mathcal{F}_n]]$ .

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  $t = \max(p, q) + 1, \dots, n$  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  $\tilde{W}_t^* \stackrel{\text{def}}{=} e_t + \widehat{W}_{n+1}^*$ . Then, letting  $\widehat{U}_{n+1}^*$  denote the series constructed using these shifted forecast errors, e.g.  $\widehat{U}_{t+1}^* \stackrel{\text{def}}{=} \tilde{W}_t^* / \sqrt{1 - a_0 \tilde{W}_t^*}$  when using squared returns, we have that the best estimate of the conditional median  $\text{Med} [\Pi_k [U_{n+1}^* | \mathcal{F}_n]]$  is the unconditional sample median of the appropriate power of the  $\widehat{U}_{n+1}^*$  series, namely  $\widehat{\text{Med}} [\Pi_k [\widehat{U}_{n+1}^* | \mathcal{F}_n]]$ .

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

---

<sup>8</sup>It should be apparent that, in principle, one can obtain median forecasts for any measurable function of the returns.

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 nonparametric methods, see for example Cai (2002), Gannoun, Sarraco and Yu (2003).

**Remark 9.** We note that we can obtain volatility forecasts  $\hat{\sigma}_t^2$  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  $\hat{\gamma}_{n+1}$ . For example, when using (a) with squared returns the forecast of volatility would be:

$$\hat{\sigma}_t^2 \equiv \hat{X}_{n+1}^2 \stackrel{\text{def}}{=} \widehat{\text{Med}} [\Pi_2 [U_n^* | \mathcal{F}_n]] \Pi_2 [A_n] \quad (21)$$

while when using (c) with squared returns the forecast of volatility would be:

$$\hat{\sigma}_t^2 \equiv \hat{\gamma}_{n+1} \stackrel{\text{def}}{=} \left\{ a_0^* \widehat{\text{Med}} [\Pi_2 [U_n^* | \mathcal{F}_n]] + 1 \right\} \Pi_2 [A_n] \quad (22)$$

Forecasts using absolute returns are constructed in a similar fashion, the only difference being that we will be forecasting directly standard deviations  $\hat{\sigma}_t$  and not variances.

## 5 Empirical Examples

In this section we provide an empirical illustration of the application and potential of the NoVaS approach in practice. We use two real and three simulated time series, which are described in greater detail below. We perform both in-sample and out-of-sample analysis and examine in some detail how NoVaS performs under different data generating processes (DGP) and various different measures of ‘true’ volatility.

### 5.1 Data, DGP and Summary Statistics

Our first dataset consists of daily futures returns and associated realized volatility for the S&P500 index. The data were previously analyzed in Thomakos and Wang (2003).

The sample period is from 1995 to 1999 for a total of  $n = 1260$  observations. The associated realized volatility, used for assessing forecasting performance, was constructed using squared intraday returns as  $\sigma_t^2 \stackrel{\text{def}}{=} \sum_{i=1}^m r_{t,i}^2$ , where  $m$  is the number of intraday returns  $r_{t,i}$ . Our second dataset consists of daily returns and range-based volatility for 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 construct a measure of volatility, for assessing forecasting performance, using daily high and low prices. The corresponding range-based volatility was computed using the Parkinson (1980) estimator as  $\sigma_t^2 \stackrel{\text{def}}{=} [\ln(H_t) - \ln(L_t)]^2 / [4 \ln(2)]$ , where  $H_t$  and  $L_t$  denote the daily high and low prices.

The remaining three series are simulated.<sup>9</sup> The first simulated dataset is a standard  $GARCH(1, 1)$  process with an underlying  $t_{(4)}$  distribution an implied annualized volatility of 20%. The second simulated dataset is a  $GARCH(1, 1)$  process with breaks in the constant term and the GARCH parameter; the break takes place in the second half of the sample and the change in the implied annualized volatility due to the breaks is from 16% to 25%. Finally, the third simulated dataset is a two-regime Markov switching (MS)  $GARCH(1, 1)$  process where the parameter values imply annualized volatility transitions from a low volatility regime of 5% to high volatility regime of 14%. The parameters of the models were tailored after Hillebrand (2005) and Francq and Zakoïan (2005). The true generated volatility  $\sigma_t^2$  is used for assessing forecasting performance. The total number of observations for all three simulated datasets was set to  $n = 1250$ .

Descriptive statistics of the returns for all five of our datasets are given in Table 2. We are mainly interested in the kurtosis of the returns, as we will be using kurtosis-based matching in performing NoVaS .

## TABLE 2 AROUND HERE

All return series have kurtosis in excess of 3: the largest kurtosis is that of the EFG

---

<sup>9</sup>Complete DGP details and the series are available upon request from the authors.

stock (24.3), followed by that of the three simulated return series (about 14) while the S&P500 futures returns exhibit the lowest kurtosis (9.1). A formal test for kurtosis being equal to three rejects the hypothesis for all five series. Similarly, a formal test for zero skewness does not reject the hypothesis of an underlying symmetric distribution for all four series. Figures 3 and 4 show recursive estimates of the standard deviation and kurtosis of the series.

### FIGURES 3 AND 4 AROUND HERE

**Remark 10.** We should stress that according to the most recent literature, see for example Hansen and Lunde (2006), it is important to evaluate volatility forecasts using an appropriate measure for the ‘true’ underlying volatility. The consensus is that some form of ‘realized volatility’ is the best one to use, unlike past standard practice that used squared or absolute returns. For the S&P500 series we do have the corresponding realized volatility. Another recent strand of the literature has advocated range-based estimators as a second-best volatility measure when one does not have access to high frequency returns. We use a range-based estimator for the EFG series. For the three simulated series we have available the true volatility series and use that for evaluating our forecasts.

## 5.2 NoVaS Optimization and Forecasting Specifications

Our NoVaS in-sample analysis is performed for all four possible combinations of target distributions and variance measures, that is squared and absolute returns using a normal distribution and squared and absolute returns using a uniform distribution. We use the exponential NoVaS algorithm as given in the text, 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 (15). The results of our in-sample analysis are given in Tables 3 to 6. In the tables we present the optimal values of the exponential constant

$b^*$ , the first coefficient  $a_0^*$ , the implied lag length  $p^*$ , the value of the objective function  $D_n(\boldsymbol{\theta}^*)$  and two measures of fit. The first is the quantile-quantile correlation coefficient for the  $W_t(\boldsymbol{\theta}^*)$  series, denoted  $r_{QQ}$  in the tables, and the second is the correlation between the fitted NoVaS variance  $\hat{\gamma}_t$ , equations (1) and (2), and the corresponding ‘true’ volatility measure of the datasets.

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 Table 7. All forecasts are based on a rolling sample of  $n_0 = 900$  observations with evaluation samples  $n_1 = 350$  for the three simulated datasets,  $n_1 = 360$  for S&P500 and  $n_1 = 503$  for the EFG stock. All predictions 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. 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  $\hat{\gamma}_{n+1}$  - see remark 9 before.<sup>10</sup> 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), 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 MAD of the forecast errors  $e_t \stackrel{\text{def}}{=} \sigma_t^2 - \hat{\sigma}_t^2$ , given by:

$$MAD(e) \stackrel{\text{def}}{=} \frac{1}{n_1} \sum_{t=n_0+1}^n |e_t| \quad (23)$$

where  $\hat{\sigma}_t^2$  denotes the forecast for any of the methods/models we use. As a naïve benchmark we use the (rolling) sample variance. Our forecasting results are summarized in

---

<sup>10</sup>All NoVaS predictions were made without applying an autoregressive filter as all  $W_t(\boldsymbol{\theta}^*)$  series were uncorrelated.

Table 7. Similar results were obtained when using a recursive sample and are available on request.

### 5.3 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 3 to 6 we see that it is zero to three decimals, except for minor deviations for the S&P500 and EFG series when using squared returns and normal target. 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 0.997 in all cases examined, frequently being practically one). A visual confirmation of the differences in the distribution of returns before and after NoVaS is given in Figures 5 to 7.

#### FIGURES 5 TO 7 AROUND HERE

In these figures we have quantile-quantile plots for the S&P500, EFG and MS-GARCH series respectively using absolute returns and a normal target distribution.<sup>11</sup> It is apparent from these figures that normalization has been achieved. For completeness, we present in Figure 8 the quantile-quantile plot for the S&P500 series when using absolute returns and a uniform target.

#### FIGURE 8 AROUND HERE

#### TABLES 3 TO 6 AROUND HERE

A second noticeable result is the optimal lag length chosen by the different NoVaS specifications. In particular, we see from Tables 3 to 6 that the optimal lag length is

---

<sup>11</sup>Quantile-quantile plots for other NoVaS combinations were practically the same and are available on request.

general much greater when using the normal distribution as a target: it is about three 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-half which implies an approximate interval for the uniform distribution in the range  $\pm 1.5$  to  $\pm 2$ , which can also be seen in Figure 8.

Finally, the correlation between ‘true’ volatility and the NoVaS variance measure  $\gamma_t$  appears to be moderate to high, depending on the data series and the target distribution used. For example, for the S&P500 series the highest correlation is about 69% when using squared returns and a uniform target and 56% when using absolute returns and a normal target distribution. For the EFG series, which has the highest sample kurtosis, the best ‘fit’ of 40% is found when using absolute returns and a normal target distribution. For the simulated series the correlation is higher, as expected, when using the normal rather than the uniform distribution as a target. A visual summary of the in-sample NoVaS computations and fit is given in Figures 9 to 11, for the S&P500, the EFG and the MS-GARCH series. The figures show the original return series, the corresponding volatility measure, the NoVaS transformed series  $W_t(\theta^*)$ , and the NoVaS variance measure (the figures were computed using absolute returns and a normal target distribution).

### FIGURES 9 TO 11 AROUND HERE

We now turn to the out-of-sample results on the forecasting performance of NoVaS , which are summarized in Table 7. Both the NoVaS made forecasts and the GARCH-made forecasts easily outperform the simple MAD benchmark, except for the mean GARCH forecasts for the MS-GARCH series.

### TABLE 7 AROUND HERE

We can also observe that the GARCH forecasts made using the  $L_1$  norm are better than their corresponding  $L_2$  forecasts, except for the case where the DGP is a GARCH (the latter result was expected, as the GARCH models estimated coincided with the DGP). These results were also noted in Politis (2003a,b), that is that the use of  $L_1$ -based forecasts can improve the forecasting performance of GARCH-type models. However, our interest here lies in the performance of the NoVaS -made forecasts and the results of Table 7 are clear: there is at least one NoVaS specification whose forecasts outperform the GARCH-made forecasts. This is true for all series except when the DGP itself is of the GARCH-type. Specifically, NoVaS specifications have the best performance for both real world series and the GARCH with breaks and MS-GARCH series. This is true for both types of forecasts, based on squared or absolute returns and based on the NoVaS variance (except for the case where the DGP is GARCH with breaks; there the NoVaS forecast based on the NoVaS variance is best). In fact, the MAD ratio of the best GARCH forecast to the best NoVaS forecast for these four series is 2, 1.7, 1.1 and 4 respectively. We can also see that the best NoVaS specifications are those using the normal target distribution, although the differences in forecasting performance between the uniform and normal target distributions are not very large - for the MS-GARCH series the results using the uniform target are, in fact, a little better than those using the normal target distribution.

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 to overcome 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. Further work is needed to examine whether NoVaS can outperform the ‘best’ parametric model in a return series.

## 6 Concluding Remarks

In this paper we reviewed and presented some additional results on financial time series and volatility prediction using the NoVaS transformation method introduced by Politis (2003a,b). NoVaS is based on an exploratory data analysis idea, is model-free and is especially relevant when making forecasts in the context of model uncertainty and structural breaks. It allows for very flexible modeling of financial returns as the use of the NoVaS implied distribution allows one to capture any degree of heavy tails in the returns’ distribution without having to resort to a particular *a priori* distribution or parametric model and functional form.

Here we introduced a new implied distribution in the context of NoVaS , we presented a number of additional methods for implementing and forecasting with NoVaS , using squared and absolute returns, and we examined the relative forecasting performance of NoVaS for making volatility predictions using real and simulated time series. Our empirical results are very encouraging about the applicability and prospects of the NoVaS approach as a serious competitor to other methods for forecasting return volatility. Further work is needed in order to fully explore the capabilities and full potential of the method.

**Acknowledgements:** This paper is part of a larger project on the use of the NoVaS transformation approach and time series prediction. A earlier version of the paper was

presented at the Department of Economics, University of Cyprus, and the Department of Economics, University of Crete, Greece. We would like to thank Elena Andreou and seminar participants for useful comments and suggestions. We would also like to thank an anonymous referee for his careful reading and useful comments on our manuscript and co-editor Mark Wohar for his kind invitation to participate in this volume. All errors are ours. All computations performed by the authors. R source code and certain series used are available on request.

## 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, 778-878.
- [2] Cai, Z., 2002. “Regression Quantiles for Time Series”, *Econometric Theory*, 18, 169-192.
- [3] Francq, C. and J-M. Zakoian, 2005. “L2 Structures of Standard and Switching-Regime GARCH Models”, *Stochastic Processes and Their Applications*, 115, 1557-1582.
- [4] Gannoun, A., Saracco, J. and K. Yu, 2003. “Nonparametric Prediction by Conditional Median and Quantiles”, 117, 207–223.
- [5] Ghysels, E. and L. Forsberg, 2004. “Why Do Absolute Returns Predict Volatility So Well?”, mimeo.
- [6] 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.
- [7] Hansen, B., 2006. “Interval Forecasts and Parameter Uncertainty”, *Journal of Econometrics*, forthcoming.
- [8] Hansen, P. R. and A. Lunde, 2006. “Consistent ranking of volatility models”, *Journal of Econometrics*, 131, 97-121.
- [9] Hillebrand, E. 2005. “Neglecting Parameter Changes in GARCH Models”, *Journal of Econometrics*, 129, 121-138.

- [10] Kim, T. H. and H. White, 2004. "On more robust estimation of skewness and kurtosis", *Finance Research Letters*, 1, 56-73.
- [11] 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), 873-889.
- [12] 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.
- [13] Moors, J.J.A., 1988. "A quantile alternative for kurtosis", *The Statistician*, 37, 2532.
- [14] Parkinson, M., 1980. "The Extreme Value Method for Estimating the Variance of the Rate of Return", *Journal of Business*, 53, 6168.
- [15] Peng, L. and Q. Yao, 2003. "Least absolute deviations estimation for ARCH and GARCH models", *Biometrika*, 90, 967-975.
- [16] Politis, D.N., 2003a. "Model-Based vs. Model-Free Volatility Prediction", UCSD Dept. of Economics Discussion Paper 2003-16.
- [17] 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, 335-347.
- [18] Politis, D., 2004. "A Heavy-Tailed Distribution for ARCH Residuals with Application to Volatility Prediction", UCSD Dept. of Economics Discussion Paper 2004-01.
- [19] Poon, S. and C. Granger, 2003. "Forecasting Volatility in Financial Markets: A Review", *Journal of Economic Literature*, 41, 478539.
- [20] Thomakos, D. and T. Wang, 2003. "Realized Volatility in the Futures Markets", *Journal of Empirical Finance*, 10, 321-353.

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

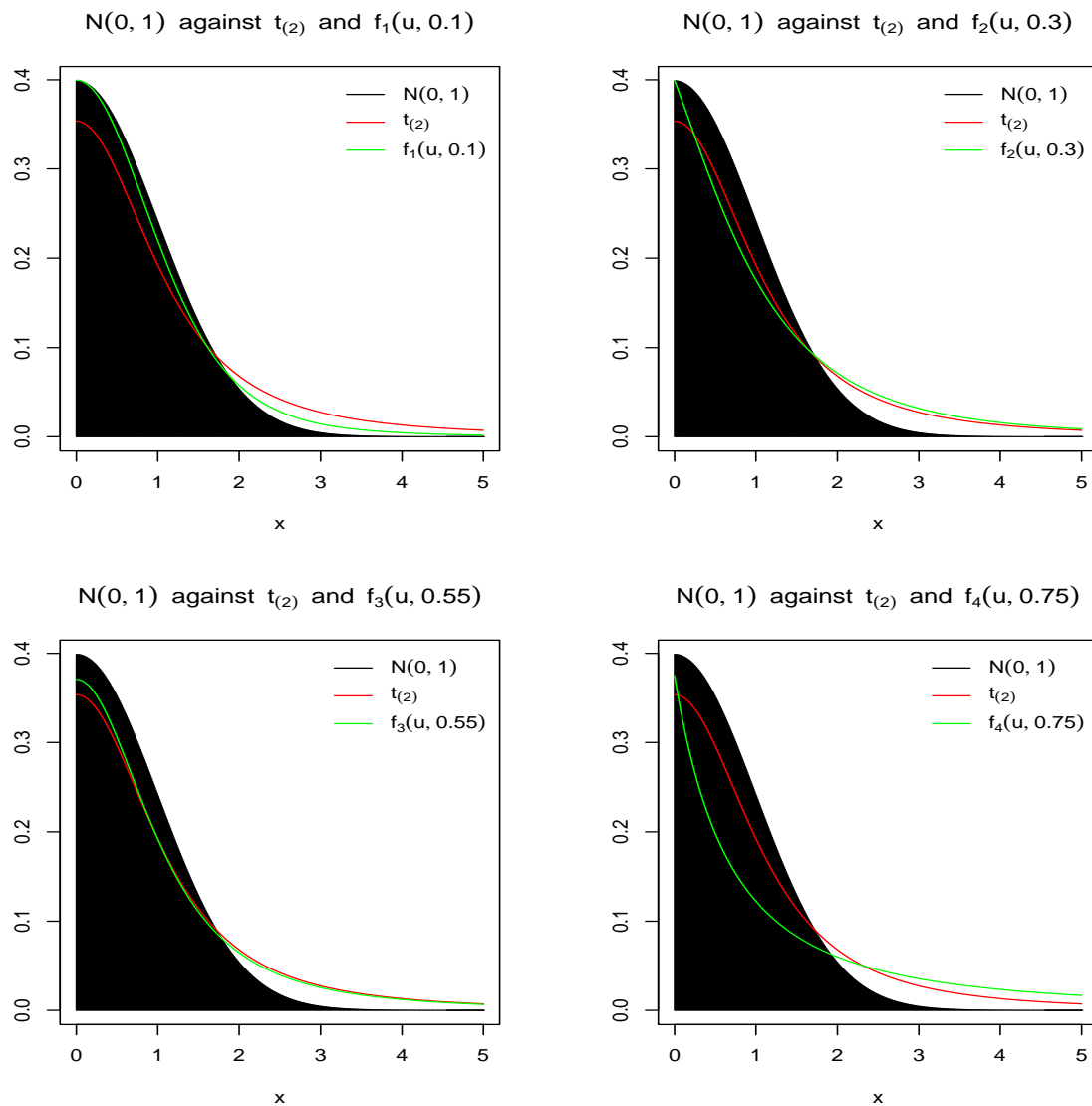


Figure 1: Implied NoVaS distributions compared to the  $\mathcal{N}(0, 1)$  and the  $t_{(2)}$  distributions

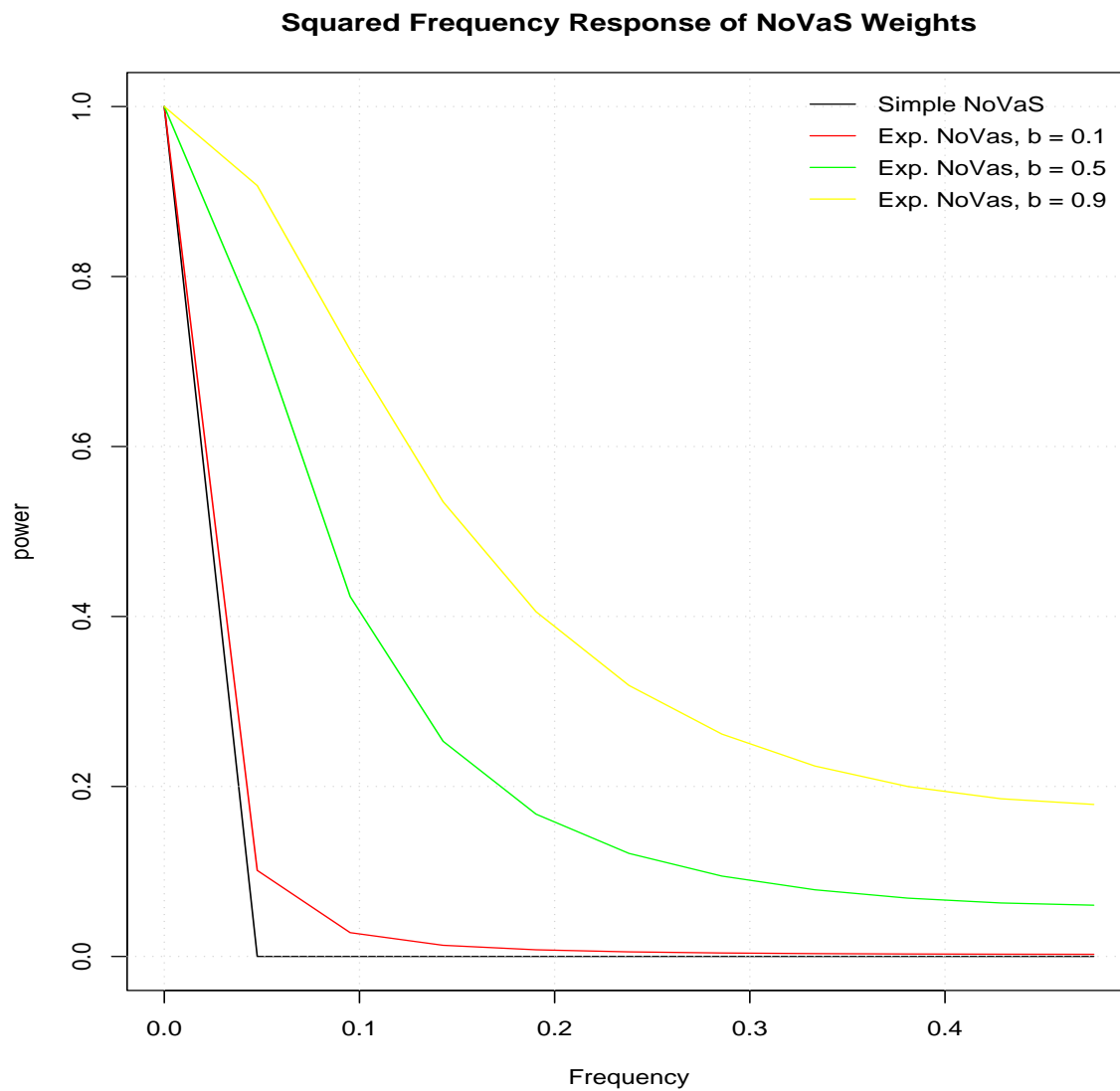


Figure 2: The squared frequency response of the NoVaS weights

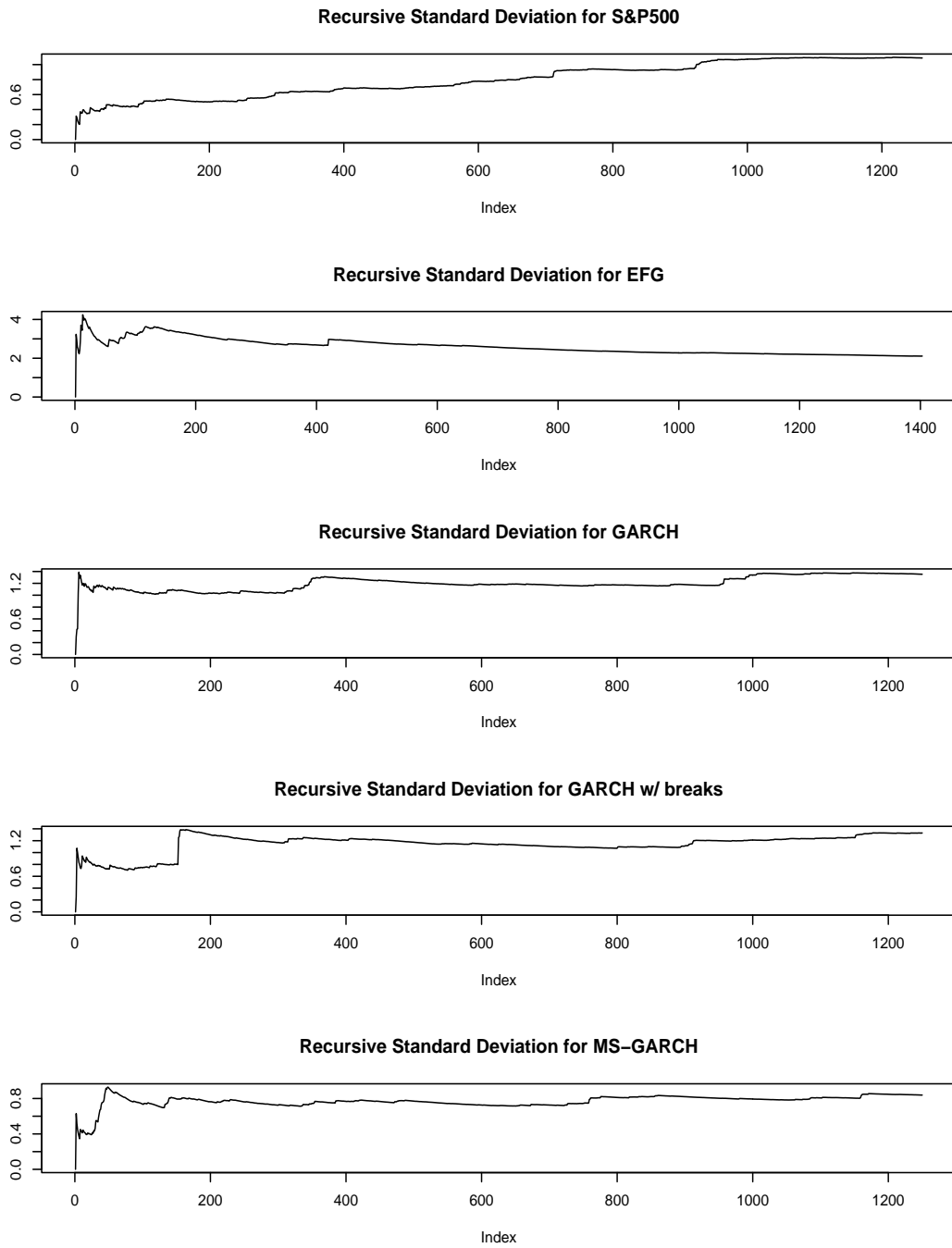


Figure 3: Recursive Standard Deviations

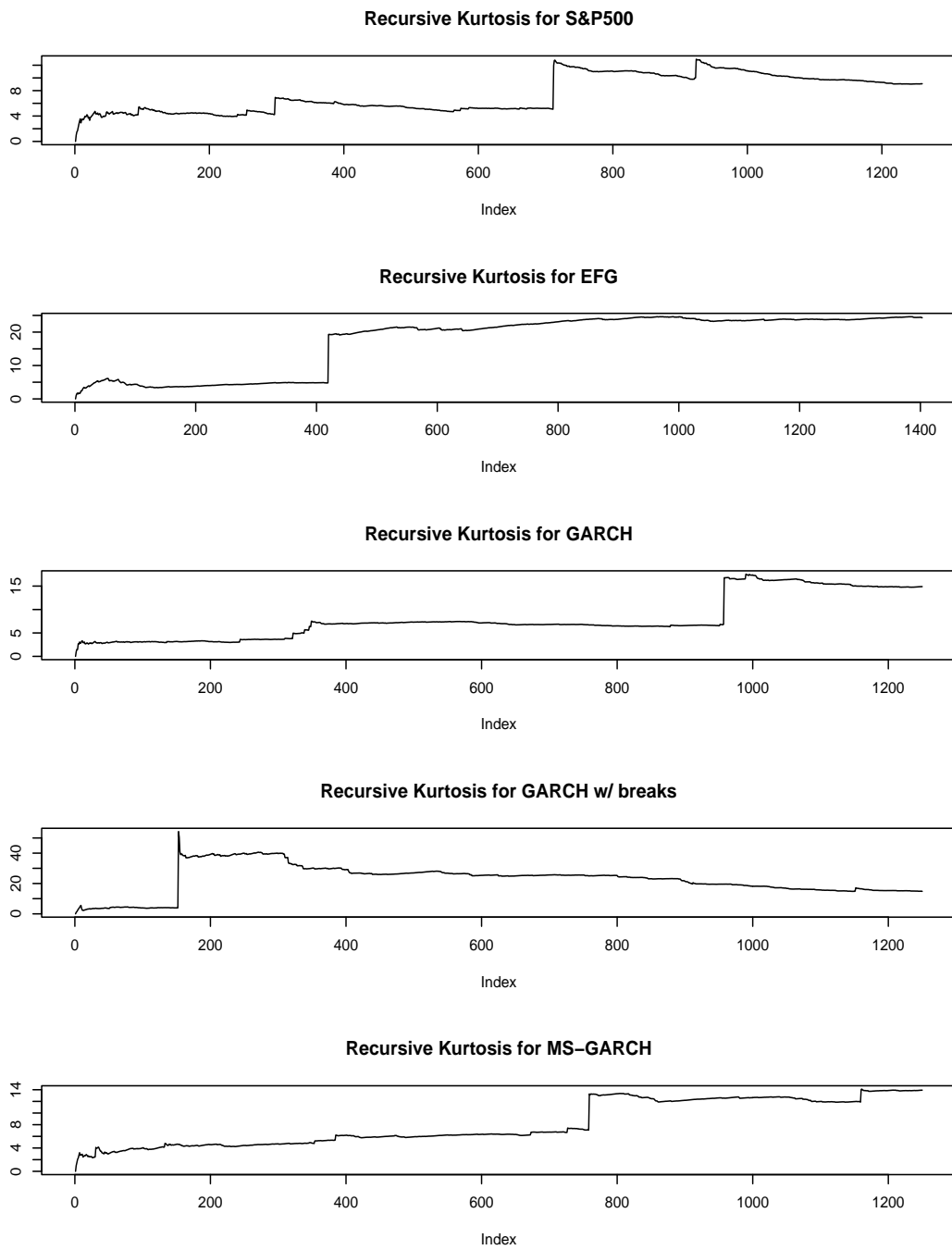


Figure 4: Recursive Kurtosis

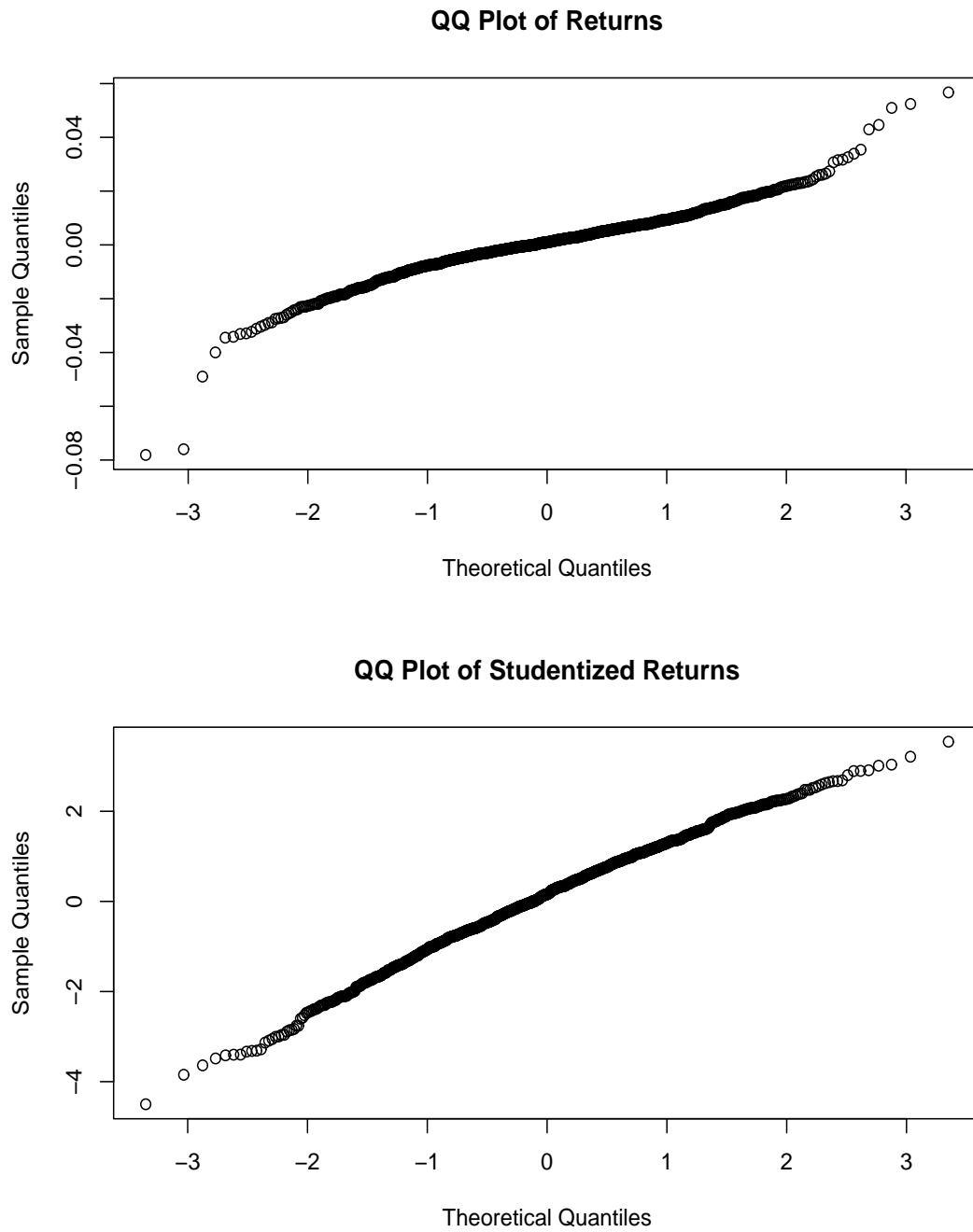


Figure 5: Quantile-Quantile Plots of Returns Before and After NoVaS transformation for the S&P500 series

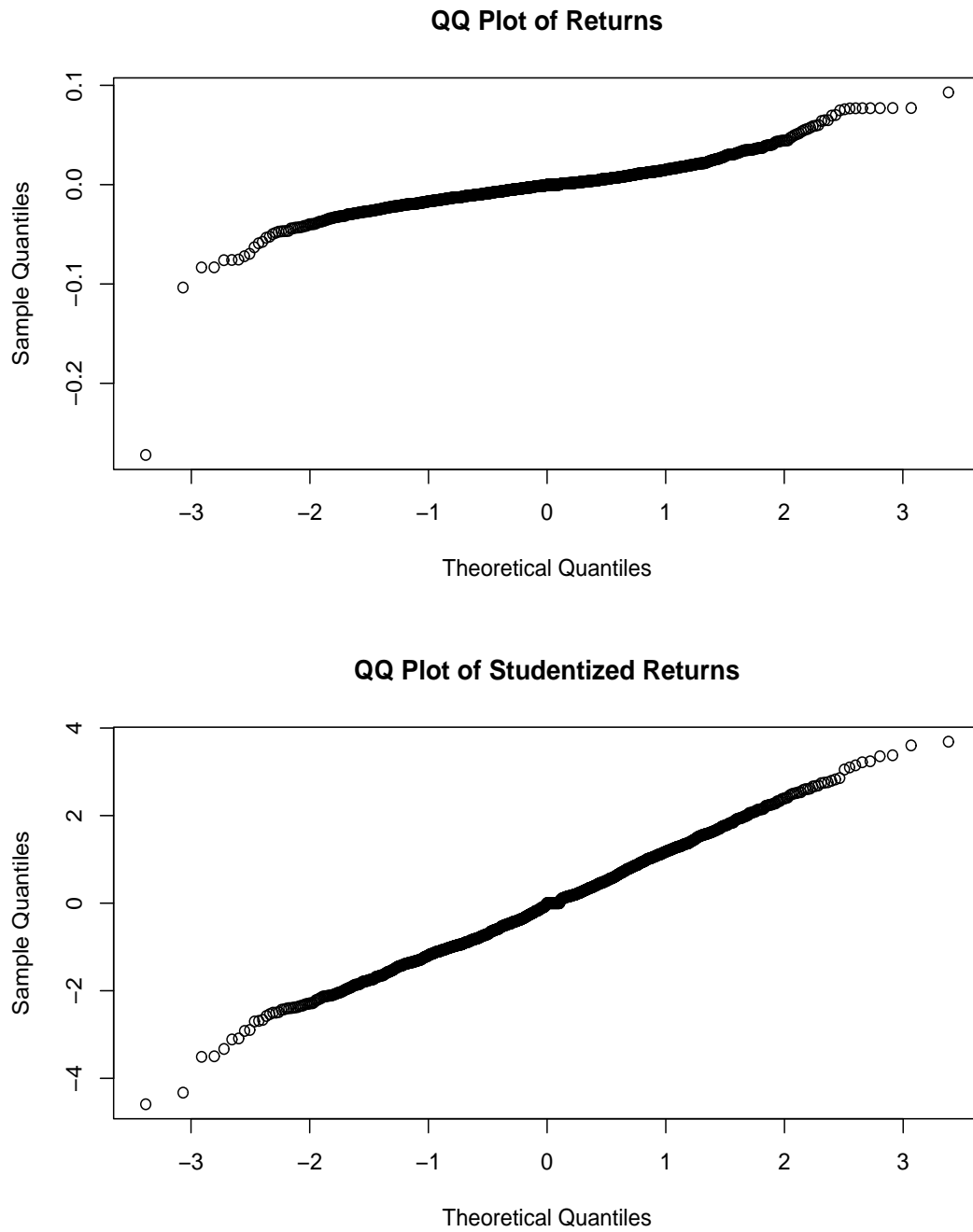


Figure 6: Quantile-Quantile Plots of Returns Before and After NoVaS transformation for the EFG series

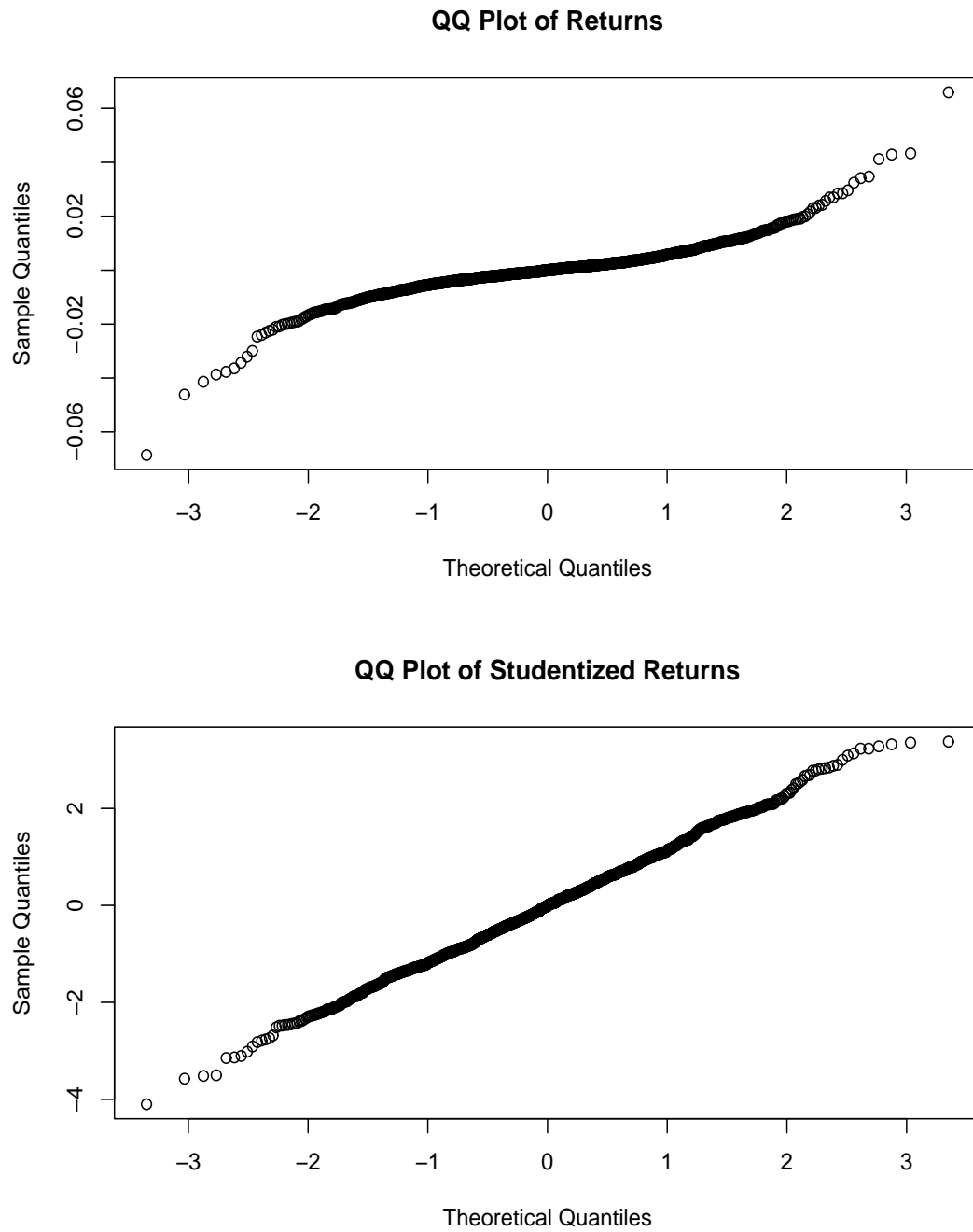


Figure 7: Quantile-Quantile Plots of Returns Before and After NoVaS transformation for the MS-GARCH series

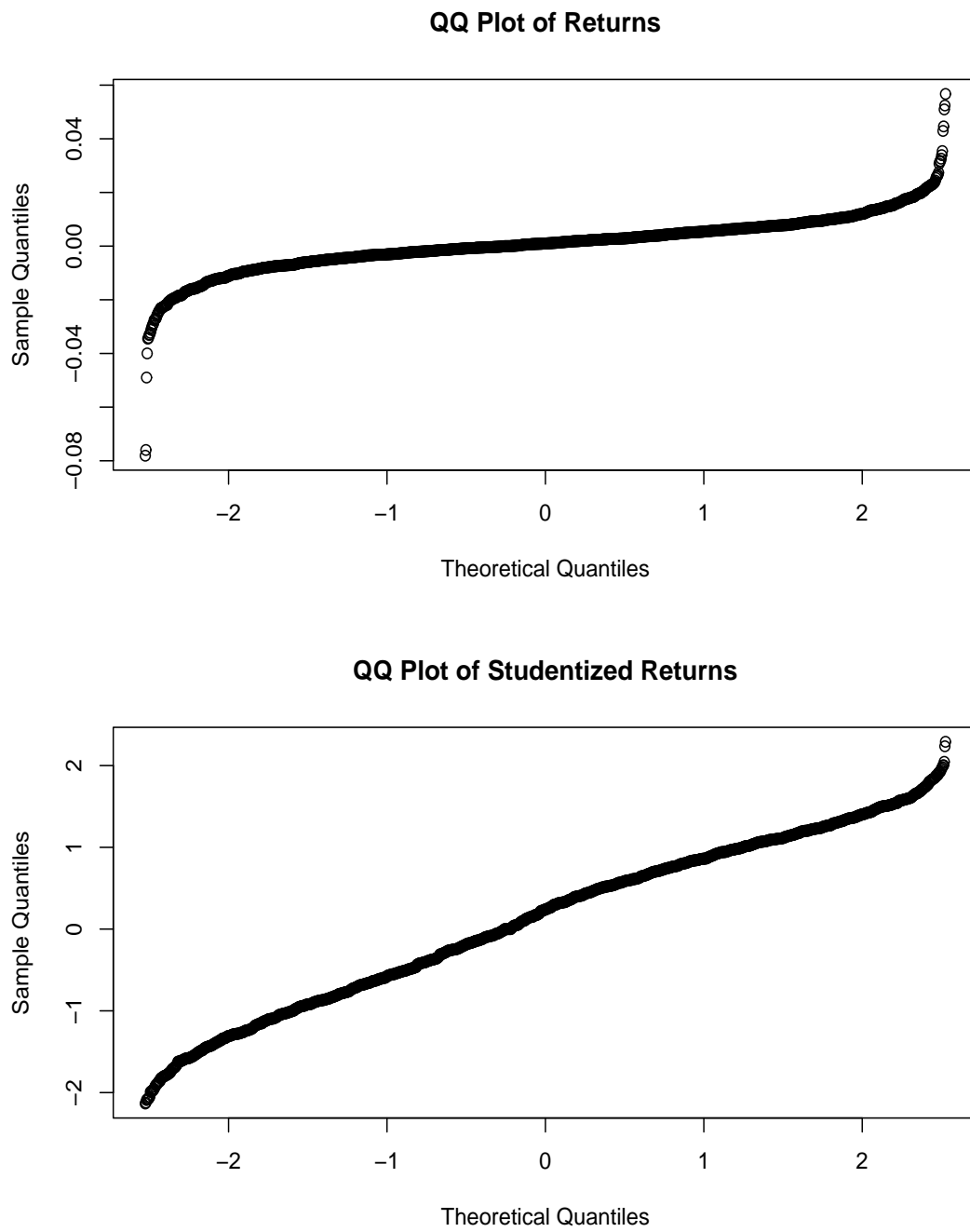


Figure 8: Quantile-Quantile Plots of Returns Before and After NoVaS transformation for the S&P500 series - Uniform Target Distribution

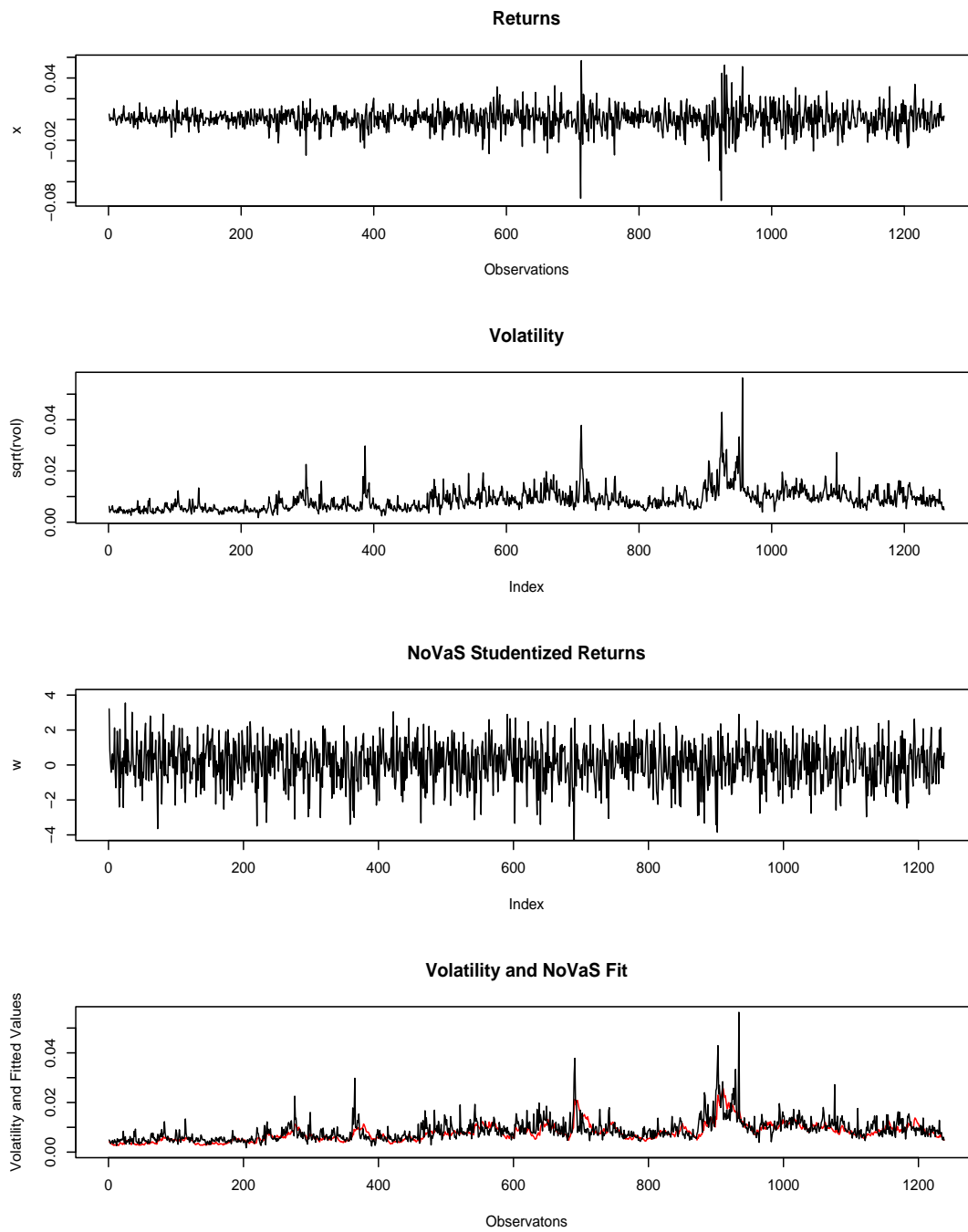


Figure 9: Summary Plots for the S&P500 series

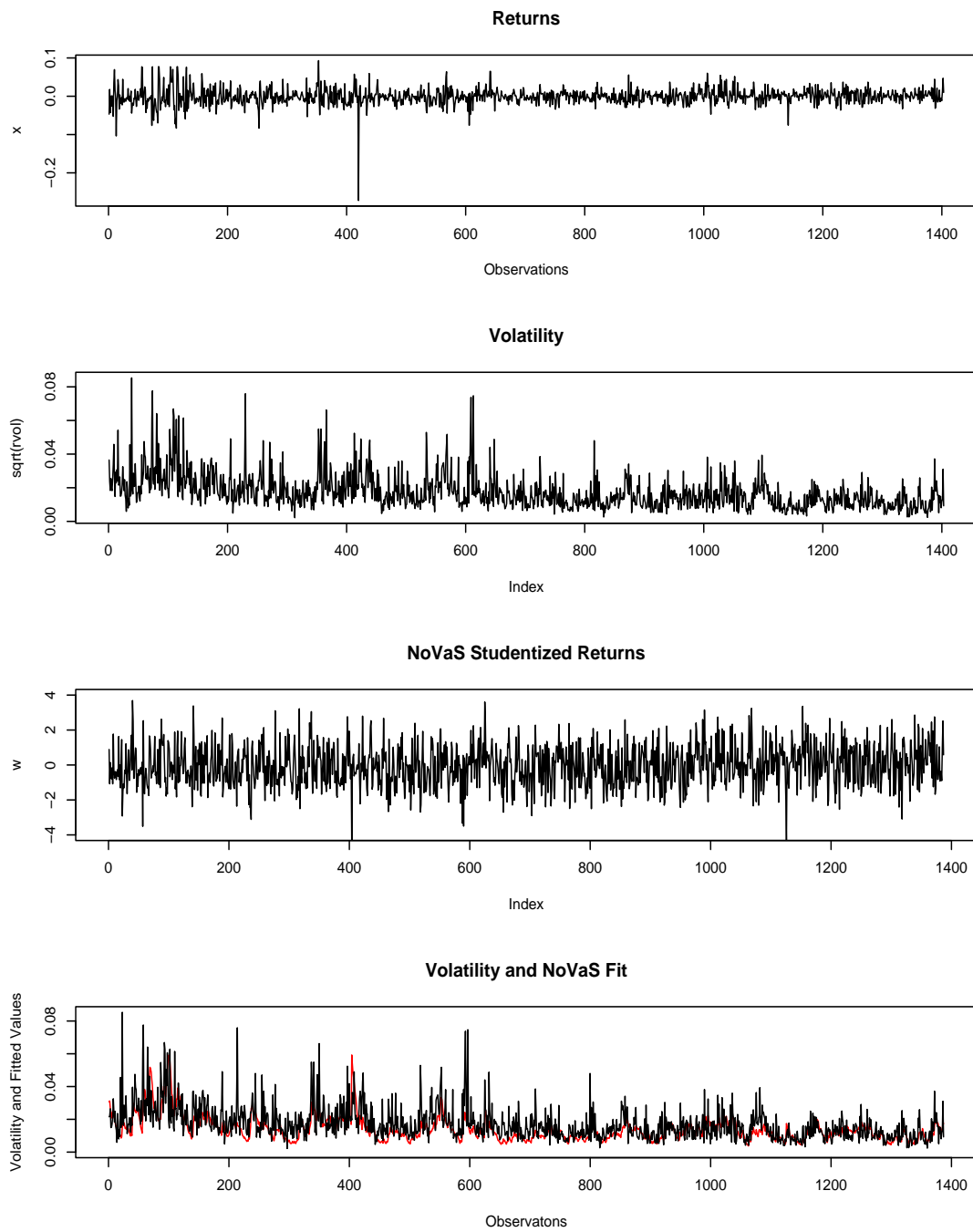


Figure 10: Summary Plots for the EFG series

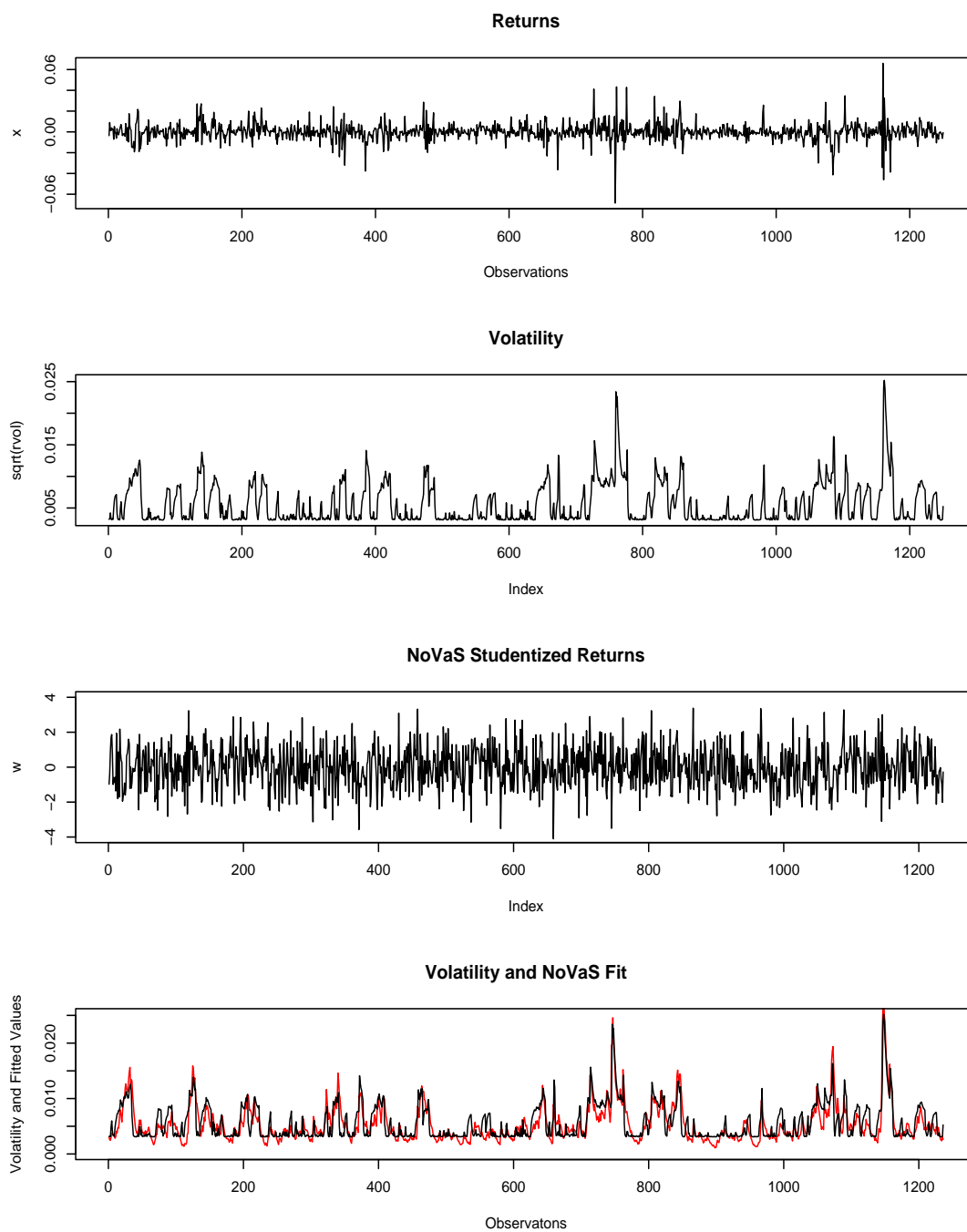


Figure 11: Summary Plots for the MS-GARCH series

**Table 1. Absolute Moments of Implicit NoVaS Distributions**

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

	$a = 1$	$a = 2$	$a = 3$	$a = 4$
$\mathcal{N}(0, 1)$	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. Descriptive Statistics of Returns**

Statistic	S&P500	EFG	GARCH	GARCH	GARCH
				w/ Break	w/ MS
$n$	1260	1403	1250	1250	1250
Mean	0.077	-0.069	0.049	0.001	0.003
RMSE	1.088	2.111	1.353	1.330	0.839
MAD	0.768	1.422	0.912	0.889	0.536
Skewness	-0.510	-1.243	-0.338	0.278	-0.006
Kurtosis	9.082	24.321	14.890	14.826	13.928

Notes: returns are expressed as percentage points  $r_t \stackrel{\text{def}}{=} 100 \cdot \ln(P_t/P_{t-1})$  for computing the descriptive statistics.

**Table 3. In-Sample Results using Squared Returns and Normal Target**

	S&P500	EFG	GARCH	GARCH	GARCH
				w/ Break	w/ MS
$b^*$	0.057	0.089	0.098	0.097	0.140
$a_0^*$	0.068	0.096	0.105	0.103	0.140
$p^*$	29	24	22	22	18
$D_n(\boldsymbol{\theta}^*)$	0.001	0.007	0.000	0.000	0.000
$r_{QQ}$	0.997	0.999	0.999	0.999	0.998
$Fit$	0.543	0.338	0.903	0.887	0.827

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.  $r_{QQ}$  is the quantile-quantile correlation coefficient for the  $W_t(\boldsymbol{\theta}^*)$  series.
4.  $Fit$  is the correlation coefficient between actual and fitted values based on the NoVaS variance  $\gamma_t$ .

**Table 4. In-Sample Results using Absolute Returns and Normal Target**

	S&P500	EFG	GARCH	GARCH	GARCH
				w/ Break	w/ MS
$b^*$	0.102	0.171	0.171	0.175	0.210
$a_0^*$	0.107	0.166	0.166	0.171	0.198
$p^*$	22	16	16	15	14
$D_n(\boldsymbol{\theta}^*)$	0.000	0.000	0.000	0.000	0.000
$r_{QQ}$	0.997	0.999	1.000	0.999	0.999
$Fit$	0.561	0.403	0.861	0.898	0.865

Notes: see table 3.

**Table 5. In-Sample Results using Squared Returns and Uniform Target**

	S&P500	EFG	GARCH	GARCH	GARCH
				w/ Break	w/ MS
$b^*$	0.420	0.460	0.487	0.450	0.515
$a_0^*$	0.351	0.378	0.394	0.373	0.409
$p^*$	8	7	7	7	7
$D_n(\boldsymbol{\theta}^*)$	0.000	0.000	0.000	0.000	0.000
$r_{QQ}$	0.997	0.999	1.000	0.999	1.000
$Fit$	0.689	0.290	0.690	0.718	0.743

Notes: see table 3.

**Table 6. In-Sample Results using Absolute Returns and Uniform Target**

	S&P500	EFG	GARCH	GARCH	GARCH
				w/ Break	w/ MS
$b^*$	0.491	0.540	0.538	0.519	0.551
$a_0^*$	0.396	0.427	0.426	0.411	0.433
$p^*$	7	6	6	7	6
$D_n(\boldsymbol{\theta}^*)$	0.000	0.000	0.000	0.000	0.000
$r_{QQ}$	0.996	0.998	0.999	0.998	0.998
$Fit$	0.680	0.350	0.556	0.710	0.720

Notes: see table 3.

**Table 7. Rolling Out-of-Sample Results on Forecasting Performance**

	S&P500	EFG	GARCH	GARCH	GARCH
				w/ Break	w/ MS
GARCH Mean	0.103	0.188	0.020	0.065	0.072
GARCH Median	0.079	0.135	0.106	0.149	0.041
Normal Target Distribution					
NoVaS SQR	0.045	0.073	0.099	0.145	0.011
NoVaS Var-SQR	0.054	0.107	0.046	0.058	0.015
NoVaS ABR	0.042	0.072	0.100	0.144	0.009
NoVaS Var-ABR	0.036	0.082	0.079	0.116	0.010
Uniform Target Distribution					
NoVaS SQR	0.049	0.087	0.098	0.147	0.009
NoVaS Var-AQR	0.050	0.093	0.079	0.106	0.011
NoVaS ABR	0.053	0.089	0.099	0.145	0.009
NoVaS Var-ABR	0.053	0.095	0.091	0.134	0.009
Benchmark	0.132	0.251	0.189	0.161	0.068
$n_1$	360	503	350	350	350

Notes:

- Table entries are the MAD of the forecasting errors ( $\times 1,000$ ).
- GARCH-Mean denotes GARCH-made forecasts using the  $L_2$  norm.
- GARCH-Median denotes GARCH-made forecasts using the  $L_1$  norm.
- NoVaS SQR and NoVaS ABR denotes NoVaS -made forecasts using either squared or absolute returns, see equation (21).
- NoVaS Var-SQR and NoVaS Var-ABR denotes NoVaS -made forecasts using the NoVaS time-localized variance measure  $\hat{\gamma}_{n+1}$ , see equation (22), using either squared or absolute returns.
- Benchmark denotes the MAD of the naïve forecast based on the (rolling) sample variance.
- $n_1$  is the number of observations in the evaluation period.