# Multivariate NoVaS & Inference on Conditional Correlations*

Dimitrios D. Thomakos† and Dimitris N. Politis‡

March 11, 2010

**Abstract**

In this paper we present new results on the NoVaS transformation approach for volatility modeling and forecasting, continuing the previous line of research by Politis (2003a,b, 2007) and Politis and Thomakos (2008a, b). Our main contribution is that we extend the NoVaS methodology to modeling and forecasting conditional correlation, thus allowing NoVaS to work in a multivariate setting as well. We present exact results on the use of univariate transformations and on their combination for joint modeling of the conditional correlations: we show how the NoVaS transformed series can be combined and the likelihood function of the product can be expressed explicitly, thus allowing for optimization and correlation modeling. While this keeps the original "model-free" spirit of NoVaS it also makes the new multivariate NoVaS approach for correlations "semi-parametric". We also present a number of auxiliary results regarding the empirical implementation of NoVaS based on different criteria for distributional matching. We illustrate our findings using simulated and real-world data.

*Keywords:* conditional correlation, forecasting, NoVaS transformations, volatility.

*Preliminary material; please do not quote without permission.

†Department of Economics, University of Peloponnese, Greece & Rimini Center for Economic Analysis, Italy. Email: thomakos@uop.gr, dimitrios.thomakos@gmail.com.

‡Department of Mathematics and Department of Economics, University of California, San Diego, USA. Email: dpolitis@ucsd.edu

# 1 Introduction

Joint modeling of the conditional second moments, volatilities and correlations, of a vector of asset returns is considerably more complicated (and with far fewer references) than individual volatility modeling. With the exception of realized correlation measures, based on high-frequency data, the literature on conditional correlation modeling is plagued with the "curse of dimensionality": parametric or semi-parametric correlation models are usually dependent on a large number of parameters (always greater than the number of assets being modeled). Besides the, always lurking, misspecification problems one is faced with the difficult task of multi-parameter numerical optimization under various constraints. Some recent advances, see for example Ledoit et al. (2003) and Palandri (2009), propose simplifications by breaking the modeling and optimization problem into smaller, more manageable, sub-problems but one still has to make ad-hoc assumptions about the way volatilities and correlations are parametrized.

In this paper we present a novel approach for modeling conditional correlations building on the NoVaS transformation approach introduced by Politis (2003a,b, 2007) and significantly extended by Politis and Thomakos (2008a, b). Our work has both similarities and differences with the related literature. The main similarity is that we also begin by modeling the volatilities of the individual series and estimate correlations using the standardized return series. The main differences are that (a) we do not make distributional assumptions for the distribution of the standardized returns, (b) we assume no "model" for the volatilities and the correlations, (c) we use a correctly specified likelihood function for the correlations and (d) calibration-estimation of parameters requires only one-dimensional optimizations in the unit interval and simple numerical integration.

The main advantages of using NoVaS transformations for volatility modeling and forecasting, see Politis and Thomakos (2008b), are that the method is data-adaptable without making any a prior assumptions about the distribution of returns (e.g. their degree of kurtosis) and it can work in a multitude of environments (e.g. global and local stationary models, models with structural breaks etc.) These advantage carry-over to the case of correlation modeling. In addition to our main results on correlations we also present some auxiliary results on the use of different criteria for distributional matching thus allowing for a more "automated" application of the NoVaS methodology.

The related literature on conditional correlation modeling is focused on finding parsimonious, easy to optimize, parametric and semi-parametric representations of volatilities and correlations, and on approaches that can handle the presence of excess kurtosis in asset returns. Early references for parametric multivariate models of volatility and correlation include Bollerslev, Engle

and Woolridge (1988) (the VEC model), Bollerslev (1990) (the constant conditional correlation, CCC model), Bollerslev and Woolridge (1992) and Engle and Kroner (1995) (the BEKK model).

Engle (2002) introduced the popular dynamic conditional correlation DCC model, which was extended and generalized by various authors: see, among others, Tse and Tsui (2002), Sheppard (2002), Pelletier (2006), Silvennoinen and Terasvirta (2005, 2009) and Hanfner and Frances (2009). For a review of the class of multivariate GARCH-type models see Bauwens et al. (2006) and for a review of volatility and correlation forecast evaluation see Patton and Sheppard (2008). A recent paper linking BEKK and DCC models is Caporin and McAleer (2010).

Part of the literature treats the problem in a semi-parametric or non-parametric manner, such as in Long and Ullah (2005) and Hafner et al. (2004). Ledoit et al. (2003) and Palandri (2009) propose simplifications to the modeling process, both on a parametrization and optimization level.

The NoVaS approach we present in this paper also has some similarities with copula-based modeling where the marginal distributions of standardized returns are specified and then joined to form a multivariate distribution; for applications in the current context see Jondeau and Rockinger (2006) and Patton (2006). Finally, see Andersen et al. (2006) for the realized correlation measures.

The rest of the paper is organized as follows: in Section 2 we briefly review the general development of the NoVaS approach; in Section 3 we present the new results on NoVaS -based modeling and forecasting of correlations; in Section 4 we present a proposal for "model" selection in the context of NoVaS ; in Section 5 we present some limited simulation results while in Section 6 we present an illustrative empirical application; section 7 offers some concluding remarks.

## 2 Review of the NoVaS Methodology

In this section we present a brief overview of the univariate NoVaS methodology: the NoVaS transformation, the implied NoVaS distribution and the methods for distributional matching. For brevity we do not review the NoVaS volatility forecasting methodology, which can be found along with additional discussion in Politis and Thomakos (2008b).

### 2.1 NoVaS transformation and implied distribution

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

[1]Departures from the assumption of these 'stylized facts' have been discussed in Politis and Thomakos (2008a, b)

financial returns:

1. $X_t$ has a non-Gaussian, approximately symmetric distribution that exhibits excess kurtosis.

2. $X_t$ has time-varying conditional variance (volatility), denoted by $h_t^2 \stackrel{\text{def}}{=} \mathsf{E}\left[X_t^2|\mathcal{F}_{t-1}\right]$ that exhibits strong dependence, where $\mathcal{F}_{t-1} \stackrel{\text{def}}{=} \sigma(X_{t-1}, X_{t-2}, \dots)$.

3. $X_t$ is dependent although it possibly exhibits low or no autocorrelation which suggests possible nonlinearity.

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

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

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

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

with the implied restrictions (to maintain positivity for $\gamma_t$) that $\alpha \geq 0$, $a_i \geq 0$, $g(\cdot) > 0$ and $a_p \neq 0$ for identifiability. The "natural" choices for $g(z)$ are $g(z) = z^2$ or $g(z) = |z|$. With these designations, our empirical measure of the time-localized variance becomes a combination of an unweighted, recursive estimator $s_{t-1}$ of the unconditional variance of the returns $\sigma^2 = \mathsf{E}\left[X_1^2\right]$, or of the mean absolute deviation of the returns $\delta = \mathsf{E}|X_1|$, and a weighted average of the current[2] 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. $g(z) = |z|$ has also been advocated for volatility modeling; see e.g. Ghysels and Forsberg (2007) and the references therein. Robustness in the presence of outliers in an obvious advantage of absolute vs. squared returns. In addition, note that the mean absolute deviation is *proportional* to the standard deviation for the symmetric distributions that will be of current interest. The practical usefulness of the absolute value measure was demonstrated also in Politis and Thomakos (2008a, b).

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

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:

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

where $\phi(z)$ is the time-localized standard deviation that is defined relative to our choice of $g(z)$, for example $\phi(z) = \sqrt{z}$ if $g(z) = z^2$ or $\phi(z) = z$ if $g(z) = |z|$. The aim now is to choose the NoVaS parameters in such a way as to make $W_t$ follow as closely as possible a chosen target distribution that is easier to work with. The natural choice for such a distribution is the normal—hence the 'normalization' in the NoVaS acronym; other choices (such as the uniform) are also possible in applications, although perhaps not as intuitive–see e.g. Politis and Thomakos (2008a, b). Note, however, that the uniform distribution is far easier to work with in both the univariate and multivariate context.

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

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

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

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

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

and

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

If one postulates that the $U_t$ are i.i.d. according to some desired distribution, then eq. (4) becomes a *bona fide* model.[3] For example, if the distribution of $U_t$ is the one implied by eq. (4)

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

with $W_t$ having a (truncated) normal distribution, then eq. (4) is the model that is 'associated' with NoVaS. The appendix has details on the exact form and probabilistic properties of the resulting *implied* distributions for $U_t$ for all four combinations of target distributions (normal and uniform) and variance estimates (squared and absolute returns).

## 2.2 NoVaS distributional matching

### 2.2.1 Weight selection

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

We discuss the case when $\alpha = 0$; see Politis and Thomakos (2008a, b)for the case of $\alpha \neq 0$. The simplest scheme that satisfies the above conditions is equal weighting, that is $a_j = 1/(p+1)$ for all $j = 0, 1, \ldots, p$. These are the 'simple' NoVaS weights proposed in Politis (2003a,b, 2007). An alternative allowing for greater weight to be placed on earlier lags is to consider exponential weights of the form:

$$a_j = \left\{ \begin{array}{lll} 1/\sum_{j=0}^{p} \exp(-bj) & \text{for} & j = 0 \\ a_0 \exp(-bj) & \text{for} & j = 1, 2, \ldots, p \end{array} \right\} \tag{7}$$

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

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

### 2.2.2 Objective functions for optimization

To evaluate whether the distributional matching to the target distribution has been achieved, many different objective functions could be used. For example, one could use moment-based matching (e.g. kurtosis matching as originally proposed by Politis [2003a,b, 2007]), or complete

distributional matching via any goodness-of-fit statistic like the Kolmogorov-Smirnov statistic, the quantile-quantile correlation coefficient (Shapiro-Wilks type of statistic) and others. All these measures are essentially distance-based and the optimization will attempt to minimize the distance between empirical (sample) and target values.

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

$$\mathcal{K}_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\sum_{t=1}^{n}(W_t - \bar{W}_n)^4}{n s_n^4} - \kappa^* \tag{8}$$

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 $\kappa^*$ denotes the theoretical kurtosis coefficient of the target distribution. For the normal distribution $\kappa^* = 3$.

The objective function for this case can be taken to be the absolute value, i.e., $D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} |\mathcal{K}_n(\boldsymbol{\theta})|$, and one would adjust the values of $\boldsymbol{\theta}$ so as to minimize $D_n(\boldsymbol{\theta})$.[4] Politis [2003a, 2007] describes a suitable algorithm that can be used to optimize $D_n(\boldsymbol{\theta})$.

Alternative specifications for the objective function that we have successfully used in previous applied work include the QQ-correlation coefficient and the Kolmogorov-Smirnov statistic. The first is easily constructed as follows. For any given values of $\boldsymbol{\theta}$ compute the order statistics $W_{(t)}$, $W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(n)}$, 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 $\left[Q_{(t)}, W_{(t)}\right]$ 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:

$$D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} 1 - \frac{\left[\sum_{t=1}^{n}(W_{(t)} - \bar{W}_n)(Q_{(t)} - \bar{Q}_n)\right]^2}{\left[\sum_{t=1}^{n}(W_{(t)} - \bar{W}_n)^2\right] \cdot \left[\sum_{t=1}^{n}(Q_{(t)} - \bar{Q}_n)^2\right]} \tag{9}$$

In a similar fashion one can construct an objective function that is based on the Kolmogorov-Smirnov statistic as:

$$D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \sup_t \sqrt{n}|F_t - \widehat{F}_{W,t}| \tag{10}$$

Note that for any choice of the objective function we have that $D_n(\boldsymbol{\theta}) \geq 0$ and the optimal values

[4] As noted by Politis (2003a,b, 2007) such an optimization procedure will always have a solution in view of the intermediate value theorem. To see this, note that when $p = 0$, $a_0$ must equal 1, and thus $W_t = sign(X_t)$ that corresponds to $\mathcal{K}_n(\boldsymbol{\theta}) < 0$ for any choice of the target distribution. On the other hand, for large values of $p$ we expect that $\mathcal{K}_n(\boldsymbol{\theta}) > 0$, since it is assumed that the data have large excess kurtosis. Therefore, there must be a value of $\boldsymbol{\theta}$ that will make the sample excess kurtosis approximately equal to zero.

of the parameters are clearly determined by the condition:

$$\boldsymbol{\theta}_n^* \stackrel{\text{def}}{=} \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, D_n(\boldsymbol{\theta}) \tag{11}$$

with the final studentized series given by $W_t^* \equiv W_t(\boldsymbol{\theta}_n^*)$.

**Remark 2.** While the above approach is theoretically and empirically suitable for achieving distribution matching in a univariate context the question about its suitability in a multivariate context naturally arises. For example, why not use a multivariate version of a kurtosis statistic (e.g. Mardia [1970], Wang and Serfling [2005]) or a multivariate normality statistic (e.g. Royston [1982], Villasenor-Alva and Gonzalez-Estrada [2009])? This is certainly possible, and follows along the same arguments as above. However, it also means that multivariate numerical optimization (in a unit hyperplane) would need to be used thus making the multivariate approach unattractive for large scale problems. Our preferred method is to perform univariate distributional matching for the individual series and then model their correlations, as we show in the next section.

## 3 Multivariate NoVaS & Correlations

We now turn to multivariate NoVaS modeling. Our starting point is similar to that of many other correlation modeling approaches in the literature. In a parametric context one first builds univariate models for the volatilities and then uses the fitted volatility values to standardize the returns and use those for building a model for the correlations. We can do the same here after having obtained the (properly aligned) studentized series $W_{t,i}^*$ and $W_{t,j}^*$, for a pair of returns $(i, j)$. There are two main advantages with the use of NoVaS in the present context: (a) the individual volatility series are potentially more accurate since there is no problem of parametric misspecification and (b) there is only one univariate optimization per pair of returns analyzed. To fix ideas first remember that the studentized return series use information up to *and including* time $t$. Note that this is different from the standardization used in the rest of the literature where the standardization is made from the model not from the data, i.e. from $X_t/A_{t-1}$ in the present notation. This allows us to use the time $t$ information when computing the correlation measure. Second, there are some interesting properties concerning the product of two studentized series which we summarize in the following proposition.

**Proposition 1.** Consider a pair $(i, j)$ of studentized returns $W_{t,i}^*$ and $W_{t,j}^*$, which have been scaled to zero mean and unit variance, and let $Z_t(i, j) \equiv Z_t \stackrel{\text{def}}{=} W_{t,i}^* W_{t,j}^*$ denote their product. Under the assumptions of strict stationarity and distributional matching we can show that the

following hold.

1. $\rho \stackrel{\text{def}}{=} \mathsf{E}[Z_t] = \mathsf{E}\left[W^*_{t,i}W^*_{t,j}\right]$ is the constant correlation coefficient between the returns and can be consistently estimated by the sample mean of $Z_t$ as $\widehat{\rho}_n \stackrel{\text{def}}{=} n^{-1}\sum_{t=1}^{n} Z_t$.

2. $\rho_{t|t-s} \stackrel{\text{def}}{=} \mathsf{E}[Z_t|\mathcal{F}_{t-s}] = \mathsf{E}\left[W^*_{t,i}W^*_{t,j}|\mathcal{F}_{t-s}\right]$, for $s = 0, 1$, is the conditional correlation coefficient between the returns.

3. Assuming that both studentized series were obtained using the same target distribution then the (conditional or unconditional) density function of $Z_t$ can be obtained from the result of Rohatgi (1976) and has the generic form of:

$$f_Z(z) \stackrel{\text{def}}{=} \int_{\mathcal{D}} f_{W_i,W_j}(w_i, z/w_i)\frac{1}{|w_i|}dw_i$$

   where $f_{W_i,W_j}(w_i, w_j)$ is the joint density of the studentized series. In particular:

   (a) If the target distribution is normal, and using the unconditional correlation $\rho$, the density function of $Z_t$ is given by Craig (1936) and has the following form $f_Z(z;\rho) = I_1(z;\rho) - I_2(z;\rho)$ where:

$$I_1(z;\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}}\int_0^\infty \exp\left\{-\frac{1}{2\sqrt{1-\rho^2}}\left[w_i^2 - 2\rho z + (z/w_i)^2\right]\right\}\frac{dw_i}{w_i}$$

   and $I_2(z;\rho)$ is the integral of the same function in the interval $(-\infty, 0)$.

   (b) If the target distribution is uniform, and again using the unconditional correlation $\rho$, the density function of $Z_t$ can be derived using the Karhunen-Loeve transform and is given (apart from a constant) as:

$$f_Z(z;\rho) = \frac{1}{\sqrt{1-\rho^2}}\int_{-\beta(\rho)}^{+\beta(\rho)}\frac{dw_i}{|w_i|}$$

   where $\beta(\rho) \stackrel{\text{def}}{=} \sqrt{3}(1+\rho)$.

4. A similar result as in 3 above holds when we use the conditional correlation $\rho_{t|t-s}$, for $s = 0, 1$.

**Remark 2.** Proposition 1 allows us a straightforward interpretation of unconditional and conditional correlation using NoVaS transformations on individual series. Moreover, note how we can make use of the distributional matching, based on the marginal distributions, to form an explicit likelihood for the product of the studentized series; this is different from the copula-based approach to correlation modeling where from marginal distributions we go to a joint distribution – the joint distribution is just not needed in the NoVaS context. We can now use the

likelihood function of the product $Z_t$ to obtain an estimate of the conditional or unconditional correlation. Since the unconditional correlation can be estimated by the sample mean of $Z_t$, the only remaining task is, therefore, to propose a suitable form for the conditional correlation and calibrate/estimate its parameters using maximum likelihood.

There are many options in setting up an estimable form for the conditional correlation. We opt for parsimony, computational simplicity and compatibility with other models in the related literature. The easiest scheme is one that follows an autoregressive model as in:

$$\begin{aligned} \rho_{t|t-s} &\stackrel{\text{def}}{=} \lambda \rho_{t-1|t-1-s} + (1-\lambda) Z_{t-s} \\ &\approx (1-\lambda) \textstyle\sum_{j=s}^{L-1+s} \lambda^{j-s} Z_{t-j} \end{aligned} \tag{12}$$

for $s = 0, 1$, $\lambda \in (0, 1)$ the smoothing parameter and $L$ a (sufficiently high) truncation parameter. This is of the form of a local average so different weights can be applied. An alternative general formulation could, for example, be as follows:

$$\rho_{t|t-s} \stackrel{\text{def}}{=} \sum_{j=s}^{L-1+s} w_j(\lambda) B^j Z_t \equiv w(B; \lambda) Z_t \tag{13}$$

with $B$ the backshift operator. Choosing exponential weights, as in univariate NoVaS , we can have $w_j(\lambda) \stackrel{\text{def}}{=} e^{-\lambda(j-s)}/(L-1+s)$. For any specification similar to the above, we can impose an "unbiasedness" condition (similar to other models in the literature) where the mean of the conditional correlation matches the unconditional correlation as follows:

$$\rho_{t|t-s} \stackrel{\text{def}}{=} w(B; \lambda) Z_t + [1 - w(1, \lambda)] \, \widehat{\rho}_n \tag{14}$$

Other specifications are, of course, possible but they would entail additional parameters and move us away from the NoVaS smoothing approach. For example, at the expense of one additional parameter we could account for asymmetries in the correlation in a standard fashion such as:

$$\rho_{t|t-s} \stackrel{\text{def}}{=} (\lambda + \gamma d_{t-s}) \rho_{t-1|t-1-s} + (1 - \lambda - \gamma d_{t-s}) Z_{t-s} \tag{15}$$

with $d_{t-s} \stackrel{\text{def}}{=} I(Z_{t-s} < 0)$ the indicator function for negative returns.

Finally, to ensure that the estimated correlations lie within $[-1, 1]$ it is convenient to work with an (optional) scaling condition, such as the Fisher transformation and its inverse. For example, we can model the series:

$$\psi_{t|t-s} = \frac{1}{2} \log \frac{1 + \rho_{t|t-s}}{1 - \rho_{t|t-s}} \tag{16}$$

and then transform and recover the correlations from the inverse transformation:

$$\rho_{t|t-s} = \frac{\exp(2\psi_{t|t-s}) - 1}{\exp(2\psi_{t|t-s}) + 1} \tag{17}$$

Given the form of the conditional correlation function, the truncation parameter $L$ and the above transformation we have that the smoothing parameter $\lambda$ is estimated by maximum likelihood as:

$$\widehat{\lambda}_n = \underset{\lambda \in [0,1]}{\operatorname{argmax}} \sum_{t=1}^{n} \log f_Z(Z_t; \lambda) \tag{18}$$

# 4 Using NoVaS in applications

The NoVaS methodology offers many different combinations for constructing the volatility measures and performing distributional matching. One can mix squared and absolute returns, uniform and normal marginal target distributions and different matching functions (kurtosis, QQ-correlation and KS-statistic). In applications one can either proceed by careful examination of the properties of individual series and then use a particular NoVaS combination or we can think of performing some kind of "model selection" by searching across the different combinations and selecting the one that gives us the best distributional matching.

Consider fixing the type of normalization used (squared or absolute returns) and the target distribution (normal or uniform) and then performing distributional matching using all three measures (kurtosis, QQ-correlation and KS-statistic). Record the results in a $(3 \times 1)$ vector, say $\boldsymbol{D}_m(\nu, \tau)$, where $m = \text{kurtosis}, \text{QQ-correlation}, \text{KS-statistic}$, $\nu = \text{squared}, \text{absolute returns}$ and $\tau = \text{normal}, \text{uniform target distribution}$. Each element of $\boldsymbol{D}_m(\nu, \tau)$ corresponds to one of the $D_n(\boldsymbol{\theta})$ objective functions from equations (8), (9) and (10). Then, repeat the optimizations with respect to all three measures for all combinations of $(\nu, \tau)$. The "optimal" combination is then defined both for each measure $m$ and across all possible combinations $(m, \nu, \tau)$ as follows:

$$\begin{aligned} d^*_m &\overset{\text{def}}{=} \text{argmin}_{(\nu,\tau)}\, \boldsymbol{D}_m(\nu, \tau) \\ d^* &\overset{\text{def}}{=} \text{argmin}_{(m)}\, \boldsymbol{D}_m(\nu, \tau) \end{aligned} \tag{19}$$

Once a decision is made as to which combinations will be used to model volatility, the studentized series $W^*_{t,i}$ are extracted, aligned, centered and scaled. If the selected target distributions are the same then one can use the results from Proposition and equations (12) to (18) to estimate the conditional correlation. With this procedure there is always the potential that the marginal target distributions that were selected in univariate modeling do no coincide, i.e. have one series studentized with the normal and another series studentized with the uniform distribution. In such a case a subjective decision has to be made as to which product distribution from Proposition 1 should be used; from our preliminary experimentation there appears to be more "robustness" in using the product distribution based on the uniform marginals. This is something that needs to be further explored. The choice of the truncation parameter $L$ can be based on the chosen length on the individual NoVaS transformations (i.e. on $p$ from (2)) or to

a multiple of it or it can be selected via the AIC or similar criterion (since there is a likelihood function available)

# 5 Simulation results

In this section we report results from a limited simulation study. We use a simple bivariate model for the data generating process (DGP), as in Patton and Sheppard (2008), that allows for DGP-consistent realized covariances and correlations to be computed. This is useful as in the next section we will be using realized correlation measures to assess the performance of NoVaS -based correlations. Letting $\boldsymbol{R}_t \stackrel{\text{def}}{=} [X_t, Y_t]^\top$ denote the $(2 \times 1)$ vector of returns, the DGP is given as follows:

$$\begin{aligned} \boldsymbol{R}_t &= \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\epsilon}_t \\ \boldsymbol{\epsilon}_t &= \textstyle\sum_{k=1}^{78} \boldsymbol{\xi}_{kt} \quad \text{with} \quad \boldsymbol{\xi}_{kt} \sim N(0, 78^{-1}) \\ \boldsymbol{\Sigma}_t &= 0.05\bar{\boldsymbol{\Sigma}} + 0.90\boldsymbol{\Sigma}_{t-1} + 0.05\boldsymbol{R}_{t-1}\boldsymbol{R}_{t-1}^\top + \gamma I_A \boldsymbol{R}_{t-1}\boldsymbol{R}_{t-1}^\top \end{aligned} \tag{20}$$

with $\bar{\boldsymbol{\Sigma}}$ the unconditional covariance matrix, $\gamma$ the asymmetry coefficient and $I_A \stackrel{\text{def}}{=} (X_t Y_t < 0)$ the indicator function of negative returns.

We consider two versions of the same DGP, one without $(\gamma = 0)$ and one with $(\gamma = -0.03)$ asymmetries. We use the "model selection" approach of the previous section and the original NoVaS approach using squared returns and a normal target distribution, and we compute a total of four (4) NoVaS -based correlations. Equationa (13) and (14) are used along with $s = 0$ and exponential weights. Finally a standard DCC model is used for comparison. A sample size of $n = 200$ in both cases. For each simulation run we compute the correlation coefficient between the realized and fitted volatility and correlation values and the root-mean-squared error between the realized and fitted volatility values. Note that the DGP for the conditional covariance is very similar to that for a DCC model.

The results from these simulations are given in Tables 1 and 2. We can see that there is at least one combination of NoVaS -based volatilities and correlations that are competitive to the DCC ones, both in terms of fit and in terms of root-mean-squared error. In this context, and given the nature of the DGP, it would be hard for a non-parametric and "model-free" method to beat a parametric one, especially when using a normal distribution for constructing the model's innovations. In practice, when the DGP is unknown and the data have much more kurtosis, the results between the NoVaS approach and the DCC can be more different. We explore this in the next section.

# 6 Empirical illustration

In this section we offer a brief empirical illustration of the NoVaS -based correlation estimation using data from three series: the S&P500, the 10-year bond[5] and the USD/Japanese Yen exchange rate. Daily data are obtained from the beginning of the series and then trimmed and aligned. Daily log-returns are computed and from them we compute monthly returns, realized volatilities and realized correlations. The final data sample is from 01/1971 to 02/2010 for a total of $n = 469$ available observations.

Figures 2, 3 and 4 plot the monthly returns, realized volatilities and correlations and Table 3 has some descriptive statistics. From the table we can see that all three series have excess kurtosis and appear to be non-normally distributed (bootstrapped p-values from the Shapiro-Wilk normality test – not reported – reject the hypothesis of normality). In addition, there is negative skewness for the S&P500 and the USD/JPY series. In Figure 5 we present normal QQ-plots for all series, along with a chi-square plot for their joint normality. The figure provides us with visual confirmation on the rejection of the hypothesis of normality and we next move on to perform individual NoVaS transformations, using the model selection approach of section 4.

The results from the individual NoVaS model selection are given in Table 4. All possible combinations of target distributions and squared and absolute returns are considered for each of the three objective functions used for distributional matching (kurtosis, QQ-correlation and KS-statistic). The results in the table suggests that all series can be NoVaS -transformed using the uniform target distribution and squared returns. Applying the individual NoVaS transformation to each of the series we obtained the studentized returns. Although the results of Table 4 suggest that the transformations are successful we can see this by repeating the QQ plots from Figure 5; in Figure 6 we present the QQ-plots for the studentized series, using the uniform distribution. We can see that there is a very close match between the order statistics and the theoretical quantiles, with almost all of the values falling on a straight line.

Using now the studentized series we move on to compute the NoVaS -based correlations. We use exponential weights as in equations (13) and (14) with $s = 0$ and $L$ set to a multiple of the lags used in the individual NoVaS transformations (results are similar when we use $s = 1$). We assess the performance of our approach using the same measures as in the case of our simulations, and compare our results with the DCC model as a benchmark. Table 5 holds the results and Figure 7 plots the realized correlations along with the fitted values from the NoVaS -based correlations and the DCC-based correlations. Both the table entries and the figure show that the NoVaS approach provides us with far better results, for both the volatilities and the correlations, than

[5] 10-year Treasury constant maturity rate series

the DCC model. The fit is better in all cases and the root-mean-squared error is either better or on par with the DCC. All in all, the NoVaS approach appears to be able to outperform the popular DCC benchmark.

Our results are, of course, conditional to both the data being used and the single benchmarking model. However, we should note that one of the advantages of NoVaS is data-adaptability and parameter parsimony. There are different, more complicated, types of correlation models that include asymmetries, regime switching, factors etc. All these models operate on distributional assumptions, parametric assumptions and are far less parsimonious that the NoVaS approach suggested in this paper. In addition, they are computationally very hard to handle even when the number of series used is small (this is true even fro the sequential DCC model of Palandri [2009]). NoVaS -based correlations do not suffer from these potential problems and they can be very competitive in applications.

# 7 Concluding remarks

In this paper we extend the univariate NoVaS methodology for volatility modeling and forecasting, put forth by Politis (2003a,b, 2007) and Politis and Thomakos (2008a, b), to a multivariate context. Our main is that we show how the individual NoVaS -transformed series can be combined and derive explicit expressions for likelihood function of their product. Using a simple, parsimonious parametrization and smoothing arguments similar to the univariate case, we show how the conditional correlation can be estimated and predicted. A limited simulation study and an empirical application using real data show that the NoVaS approach to correlation modeling can be very competitive, possibly outperform, a popular benchmark as the DCC model. An important advantage of the whole NoVaS approach is data-adaptability and lack of distributional or parametric assumptions. This is particularly important in a multivariate context where most of the competitive models are parametric and much more difficult to handle in applications, especially when the number of assets is large.

There are, of course, open issues that we do not address in this paper but are important both in terms of further assessing the NoVaS approach to correlation modeling and in terms of practical usefulness. Some of them are: (a) evaluation of the forecasting performance of NoVaS -based correlations; (b) evaluation of NoVaS -based correlations in the context of portfolio and risk management; (c) additional comparisons of NoVaS -based correlations with other benchmarking models. We are currently pursuing these issues in ongoing research.

## 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] Bauwens, L., Laurent, S., and Rombouts, J.V.K., 2006. "Multivariate GARCH Models: A Survey", *Journal of Applied Econometrics*, vol. 21, pp. 79-109.

[3] Bollerslev, T., 1990. "Modeling the coherence in short-run nominal exchange rates: A multivariate Generalized ARCH Model", *Review of Economics and Statistics*, vol. 72, pp. 498-595.

[4] Bollerslev, T., Engle, R., and Wooldridge, J., 1988. "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy*, vol. 96, pp. 116-131.

[5] Caporin, M. and McAleer, M., 2010. "Do We Really Need Both BEKK and DCC? A Tale of Two Multivariate GARCH Models", working paper.

[6] Craing, C., 1936. "On the Frequency Function of xy", *The Annals of Mathematical. Statistics*, vol. VII.

[7] Engle, R., 2002. "Dynamic Conditional Correlation: A simple class of multivariate GARCH models", *Journal of Business and Economic Statistics*, vol. 17.

[8] Engle, R., and Kroner, F., 1995. "Multivariate Simultaneous Generalized ARCH", *Econometric Theory*, vol. 11, pp. 122-150.

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

[10] Hafner, C. and P. Franses, 2009. "A Generalized Dynamic Conditional Correlation model with Applications to Many Asset Returns", working paper.

[11] Hafner, C., Dijk, D., and P. Franses, 2004. "Semi-parametric Modelling of Correlation Dynamics", working paper.

[12] Ledoit, O., Santa-Clara, P., Wolf, M., 2003. "Flexible Multivariate GARCH Modeling with an Application to International Stock Markets", *The Review of Economics and Statistics*, vol. 8, pp. 735-747.

[13] Long, X. and A. Ullah, 2005. "Nonparametric and Semiparametric Multivariate GARCH Model", working paper.

[14] Mardia, K., 1970. "Measures of Multivariate Skewness and Kurtosis with Applications", *Biometrika*, vol.57, pp. 519-530.

[15] Palandri, A. 2009. "Sequential Conditional Correlations: Inference and Evaluation", *Journal of Econometrics*, vol 153, pp. 122-132.

[16] Patton, A., 2006. "Modeling Asymmetric Exchange Rate Dependence", *International Economic Review*, vol. 47, pp. 527-556.

[17] Patton, A. and K. Sheppard, 2008. "Evaluating Volatility and Correlation Forecasts" OFRC Working Papers Series 2008fe22.

[18] Pelletier, D., 2006. "Regime Switching for Dynamic Correlations", *Journal of Econometrics*, vol. 131, pp. 445-473.

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

[20] 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.

[21] 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.

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

[23] Politis, D. and D. Thomakos, 2008a. "Financial Time Series and Volatility Prediction using NoVaS Transformations", in *Forecasting in the Presence of Parameter Uncertainty and Structural Breaks*, D. E. Rapach and M. E. Wohar (Eds.), Emerald Group Publishing Ltd.

[24] Politis, D. and D. Thomakos, 2008b. "NoVaS Transformations: Flexible Inference for Volatility Forecasting", working paper, Department of Economics, UCSD, http://escholarship.org/uc/item/982208kx

[25] Rohatgi, V.K., 1976. *An Introduction to Probability Theory & Mathematical Statistics*. Wiley, New York.

[26] Royston, J.P., 1983. "Some Techniques for Assessing Multivariate Normality based on the Shapiro-Wilk W", *Applied Statistics*, vol. 32, pp. 121-133.

[27] Silvennoinen, A. and T. Terasvirta, 2005. "Multivariate Autoregressive Conditional Heteroskedasticity with Smooth Transitions in Conditional Correlations", working paper.

[28] Silvennoinen, A. and T. Terasvirta, 2009. "Modeling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model", *Journal of Financial Econometrics*, vol. 7, pp. 373-411.

[29] Tse, Y.and A. Tsui, 2002. "A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations", *Journal of Business & Economic Statistics*, vol. 20, pp. 351-62.

[30] Villasenor-Alva, J. and Gonzalez-Estrada, E., 2009. "A Generalization of Shapiro-Wilks Test for Multivariate Normality", *Communications in Statistics: Theory and Methods*, vol. 38, pp. 1870-1883.

[31] Wang, J. and R. Serfling, 2005. "Nonparametric Multivariate Kurtosis and Tailweight Measures", *Journal of Nonparametric Statistics*, vol. 17, pp. 441-456.

## Appendix

To understand the implied distribution of $U_t$ first note that the range of $W_t$ is bounded. Using equation (3) 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 $\pm 3$ 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 $\left[-1/\sqrt{a_0}, +1/\sqrt{a_0}\right]$, 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 $\gamma_t$ 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 $U_t$:

$$\begin{aligned} f_1(u, a_0) &= c_1(a_0) \times (1 + a_0 u^2)^{-1.5} \exp\left[-0.5u^2/(1 + a_0 u^2)\right] \quad \text{when } g(z) = z^2 \\ f_2(u, a_0) &= c_2(a_0) \times (1 + a_0|u|)^{-2} \exp\left[-0.5u^2/(1 + a_0|u|)^2\right] \quad \text{when } g(z) = |z| \end{aligned} \tag{21}$$

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

$$\begin{aligned} f_3(u, a_0) &= c_3(a_0) \times (1 + a_0 u^2)^{-1.5} \quad \text{when } g(z) = z^2 \\ f_4(u, a_0) &= c_4(a_0) \times (1 + a_0|u|)^{-2} \quad \text{when } g(z) = |z| \end{aligned} \tag{22}$$

The densities from the use of the uniform target distribution are new, in this and related contexts.

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.[6] 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 $f_2(u, a_0)$.

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

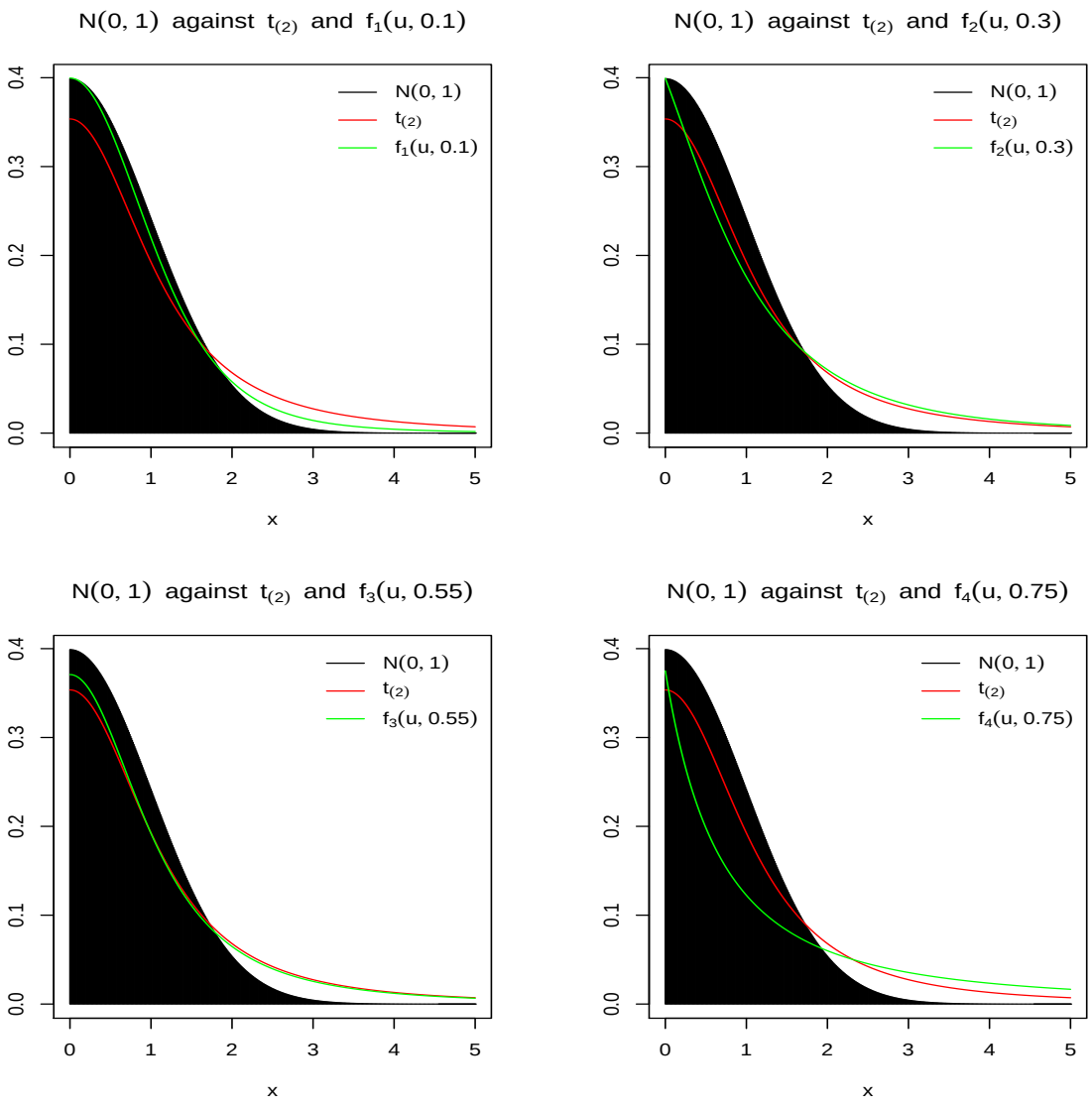


Figure 1: Implied NoVaS distributions compared to the $\mathcal{N}(0,1)$ and the $t_{(2)}$ distributions

We graphically illustrate the differences among the implied densities in equations (21) and (22) 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 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.

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

$\mathsf{E}_j|u|^a \approx \int_{-100}^{100} |u|^a f_j(u, a_0)du$ 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 (21) and (22)

To see how the $f_i(u, a_0)$ distributions compare with the standard normal and the $t_{(2)}$ distributions, we report in Table A 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 A 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.

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

**Table 1. Simulation results on the performance of NoVaS -based correlations**

DGP is given in equation (20), no asymmetries used ($\gamma = 0$)

| Model | $\mathsf{Corr}(h_x, \widehat{h}_x)$ | $\mathsf{Corr}(h_y, \widehat{h}_y)$ | $\mathsf{Corr}(\rho_{x,y}, \widehat{\rho}_{x,y})$ | $RMSE(\widehat{\rho}_{x,y})$ |
|---|---|---|---|---|
| NoVaS #1 | 0.48 | 0.47 | 0.52 | 0.04 |
| NoVaS #2 | 0.48 | 0.48 | 0.51 | 0.04 |
| NoVaS #3 | 0.54 | 0.54 | 0.36 | 0.03 |
| NoVaS #4 | 0.52 | 0.51 | 0.35 | 0.02 |
| DCC | 0.62 | 0.63 | 0.57 | 0.02 |

– $\mathsf{Corr}(h_x, \widehat{h}_x)$ denotes the sample correlation coefficient between the realized and fitted values for the volatility of the $X_t$ series; $\mathsf{Corr}(\rho_{x,y}, \widehat{\rho}_{x,y})$ denotes the sample correlation coefficient between the realized and fitted values for the correlation between $X_t$ and $Y_t$; $RMSE(\widehat{\rho}_{x,y})$ denotes the root-mean-squared error between the realized and fitted values for the correlation between $X_t$ and $Y_t$.

– NoVaS #1 denotes results from NoVaS -based correlations when individual NoVaS transformations are performed using the kurtosis for distributional matching; NoVaS #2 denotes results from NoVaS -based correlations when individual NoVaS transformations are performed using the QQ-correlation for distributional matching; NoVaS #3 denotes results from NoVaS -based correlations when individual NoVaS transformations are performed using the KS-statistic for distributional matching; NoVaS #4 denotes results from NoVaS -based correlations when individual NoVaS transformations are performed using the kurtosis for distributional matching with squared returns and a normal target distribution.

– Sample size is $n = 200$.

**Table 2. Simulation results on the performance of NoVaS -based correlations**

DGP is given in equation (20), no asymmetries used ($\gamma = -0.03$)

| Model | $\mathsf{Corr}(h_x, \widehat{h}_x)$ | $\mathsf{Corr}(h_y, \widehat{h}_y)$ | $\mathsf{Corr}(\rho_{x,y}, \widehat{\rho}_{x,y})$ | $RMSE(\widehat{\rho}_{x,y})$ |
|---|---|---|---|---|
| NoVaS #1 | 0.40 | 0.41 | 0.45 | 0.04 |
| NoVaS #2 | 0.40 | 0.42 | 0.46 | 0.04 |
| NoVaS #3 | 0.45 | 0.46 | 0.32 | 0.03 |
| NoVaS #4 | 0.42 | 0.42 | 0.28 | 0.02 |
| DCC | 0.49 | 0.51 | 0.52 | 0.02 |

– See notes in Table 1.

**Table 3. Descriptive Statistics for monthly data**

sample size is $n = 469$ months from 01/1970 to 02/2010

| | S&P500 ret. | Bonds ret. | USD/JPY ret. | S&P500 vol. | Bonds vol. | USD/JPY vol. | S&P500 Bonds corr. | S&P500 USD/JPY corr. | Bonds USD/JPY corr. |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 0.004 | -0.002 | -0.003 | 0.040 | 0.040 | 0.025 | -0.172 | 0.036 | 0.068 |
| Median | 0.007 | -0.003 | -0.001 | 0.035 | 0.035 | 0.024 | -0.238 | 0.027 | 0.076 |
| Std.Dev. | 0.045 | 0.049 | 0.031 | 0.023 | 0.024 | 0.012 | 0.383 | 0.286 | 0.322 |
| Skewness | -1.213 | 0.054 | -0.449 | 4.247 | 2.070 | 0.878 | 0.597 | 0.115 | -0.160 |
| Kurtosis | 9.149 | 5.799 | 5.730 | 33.938 | 10.134 | 5.550 | 2.631 | 2.588 | 2.436 |
| SW-test | 0.936 | 0.971 | 0.969 | 0.683 | 0.844 | 0.956 | 0.959 | 0.994 | 0.991 |

Note: SW denotes the Shapiro-Wilks test for normality; all other statistics have their standard meanings

**Table 4. Univariate NoVaS Model Selection**

| | **S&P500** | | | |
|---|---|---|---|---|
| | Normal Target squared returns | Normal Target absolute returns | Uniform Target squared returns | Uniform Target absolute returns |
| Kurtosis | 1e-4 | 2e-4 | 3e-6 | 6e-6 |
| QQ-correlation | 8e-3 | 1e-2 | 3e-3 | 7e-3 |
| KS-statistic | 0.33 | 0.11 | 0.32 | 0.02 |
| | **Bonds** | | | |
| | Normal Target squared returns | Normal Target absolute returns | Uniform Target squared returns | Uniform Target absolute returns |
| Kurtosis | 1e-4 | 1e-4 | 7e-6 | 2e-5 |
| QQ-correlation | 2e-3 | 3e-2 | 2e-3 | 6e-3 |
| KS-statistic | 0.69 | 0.08 | 0.76 | 0.04 |
| | **USD/JPY** | | | |
| | Normal Target squared returns | Normal Target absolute returns | Uniform Target squared returns | Uniform Target absolute returns |
| Kurtosis | 2e-4 | 2e-5 | 9e-6 | 8e-5 |
| QQ-correlation | 7e-3 | 8e-3 | 4e-3 | 7e-3 |
| KS-statistic | 0.26 | 0.19 | 0.23 | 0.03 |

– Table entries correspond to the value of the objective function used for distributional matching

– Column correspond to combinations of different target distributions with squared or absolute returns

– Rows correspond to the type of objective function used for distributional matching.

– For the kurtosis and QQ-correlation measures smaller values indicate better matching.

– For the KS-statistic higher values indicate better matching.

**Table 5. Correlation Modeling Results**

NoVaS -based correlations estimated using parameters from univariate optimization

| $(X_t, Y_t)$ pair | $\mathsf{Corr}(h_x, \widehat{h}_x)$ | $\mathsf{Corr}(h_y, \widehat{h}_y)$ | $\mathsf{Corr}(\rho_{x,y}, \widehat{\rho}_{x,y})$ | $RMSE(\widehat{\rho}_{x,y})$ |
|---|---|---|---|---|
| S&P500-Bonds | 0.749 | 0.770 | 0.664 | 0.091 |
| DCC | 0.494 | 0.665 | 0.568 | 0.101 |
| S&P500-USD/JPY | 0.749 | 0.629 | 0.272 | 0.105 |
| DCC | 0.494 | 0.448 | 0.205 | 0.081 |
| Bonds-USD/JPY | 0.770 | 0.629 | 0.379 | 0.104 |
| DCC | 0.665 | 0.448 | 0.285 | 0.104 |

– $\mathsf{Corr}(h_x, \widehat{h}_x)$ denotes the sample correlation coefficient between the realized and fitted values for the volatility of the $X_t$ series; $\mathsf{Corr}(\rho_{x,y}, \widehat{\rho}_{x,y})$ denotes the sample correlation coefficient between the realized and fitted values for the correlation between $X_t$ and $Y_t$; $RMSE(\widehat{\rho}_{x,y})$ denotes the root-mean-squared error between the realized and fitted values for the correlation between $X_t$ and $Y_t$.

– For each pair of variables the first line indicates the NoVaS -based results and the second line indicates the DCC-based results.

– NoVaS -based correlations computed using exponential weights as in equations (13) and (14) with $s = 0$ and $L$ set to a multiple of the lags used in individual NoVaS transformations.

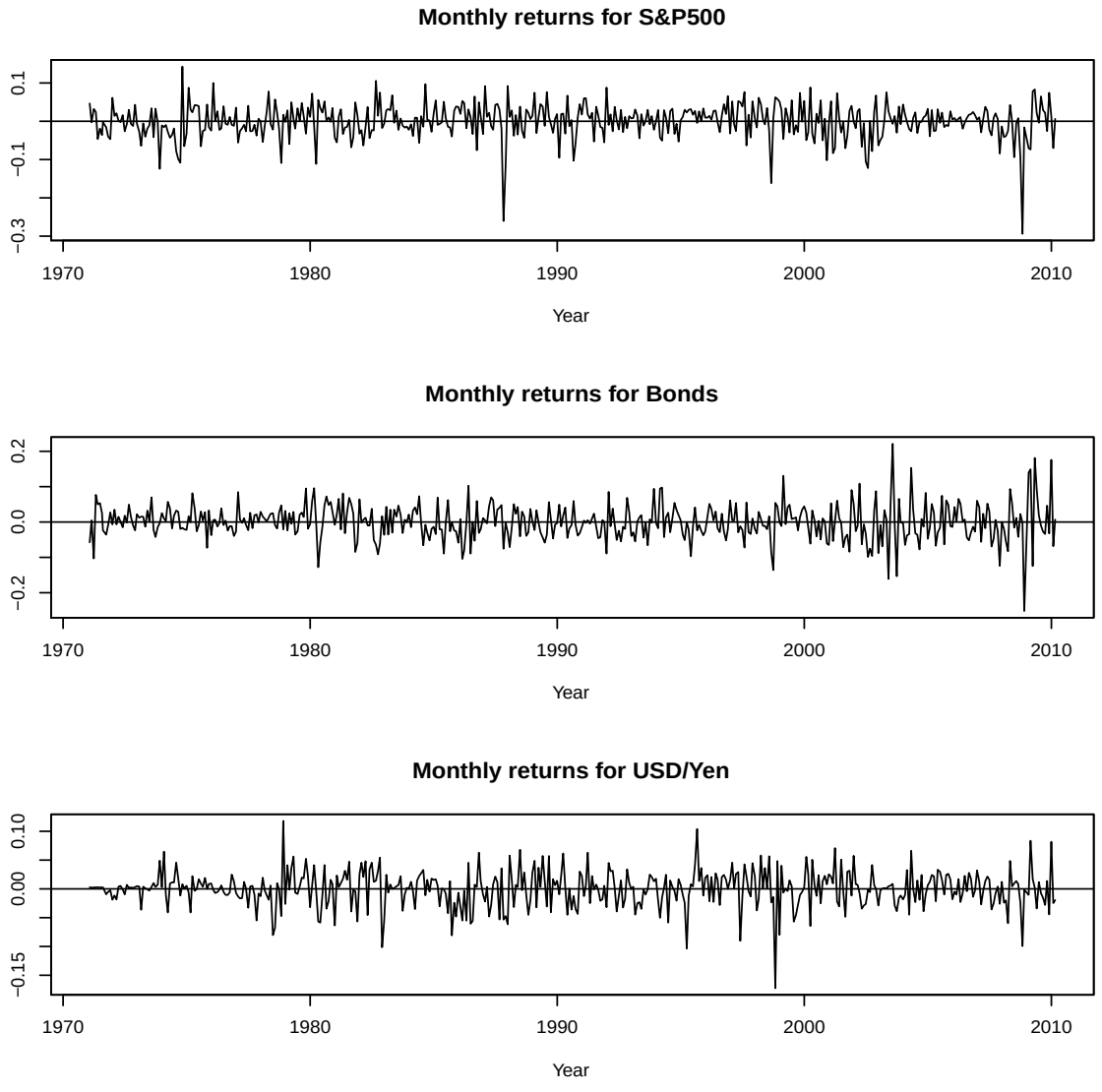


Figure 2: Monthly returns

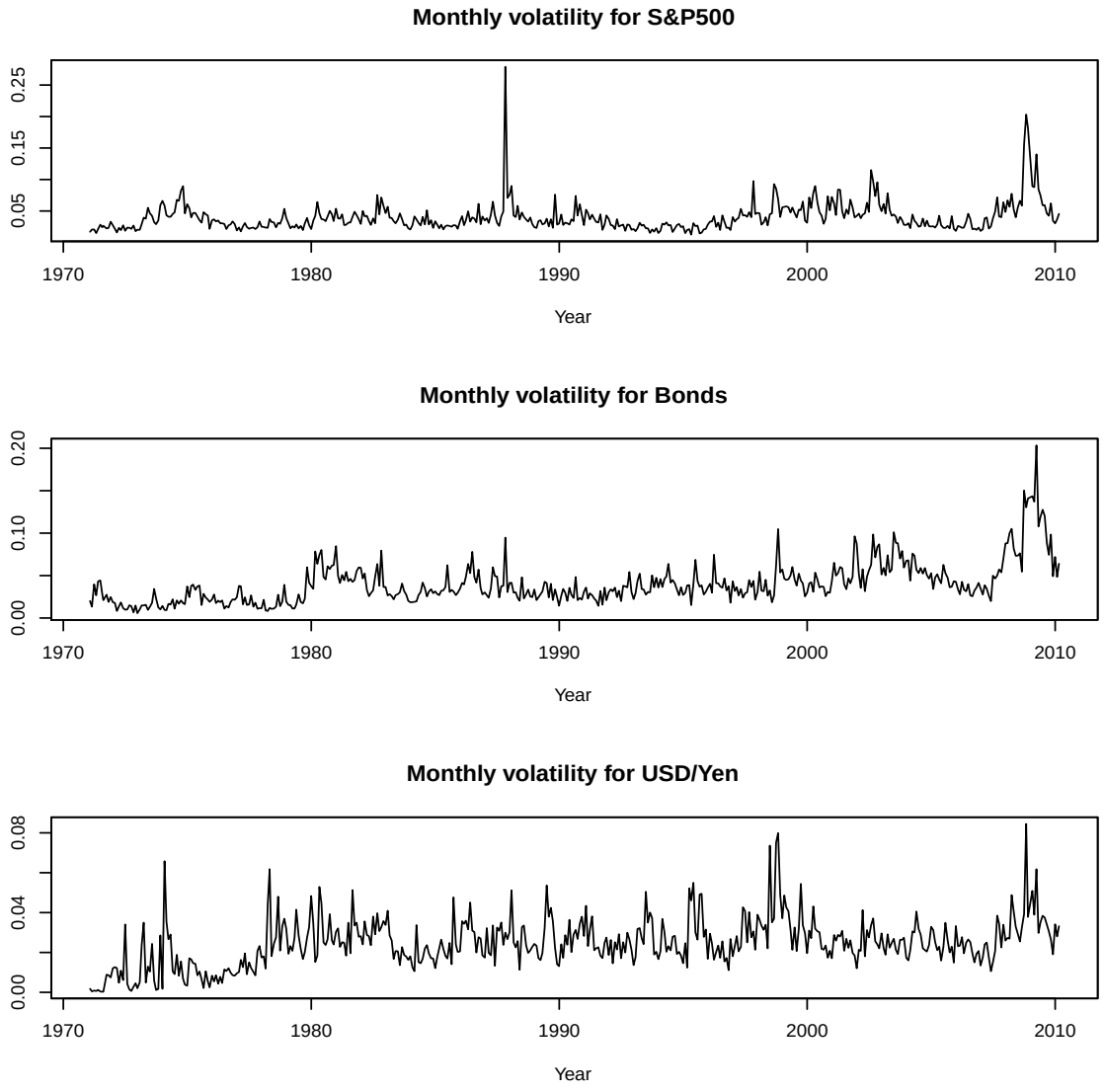


Figure 3: Monthly realized volatilities

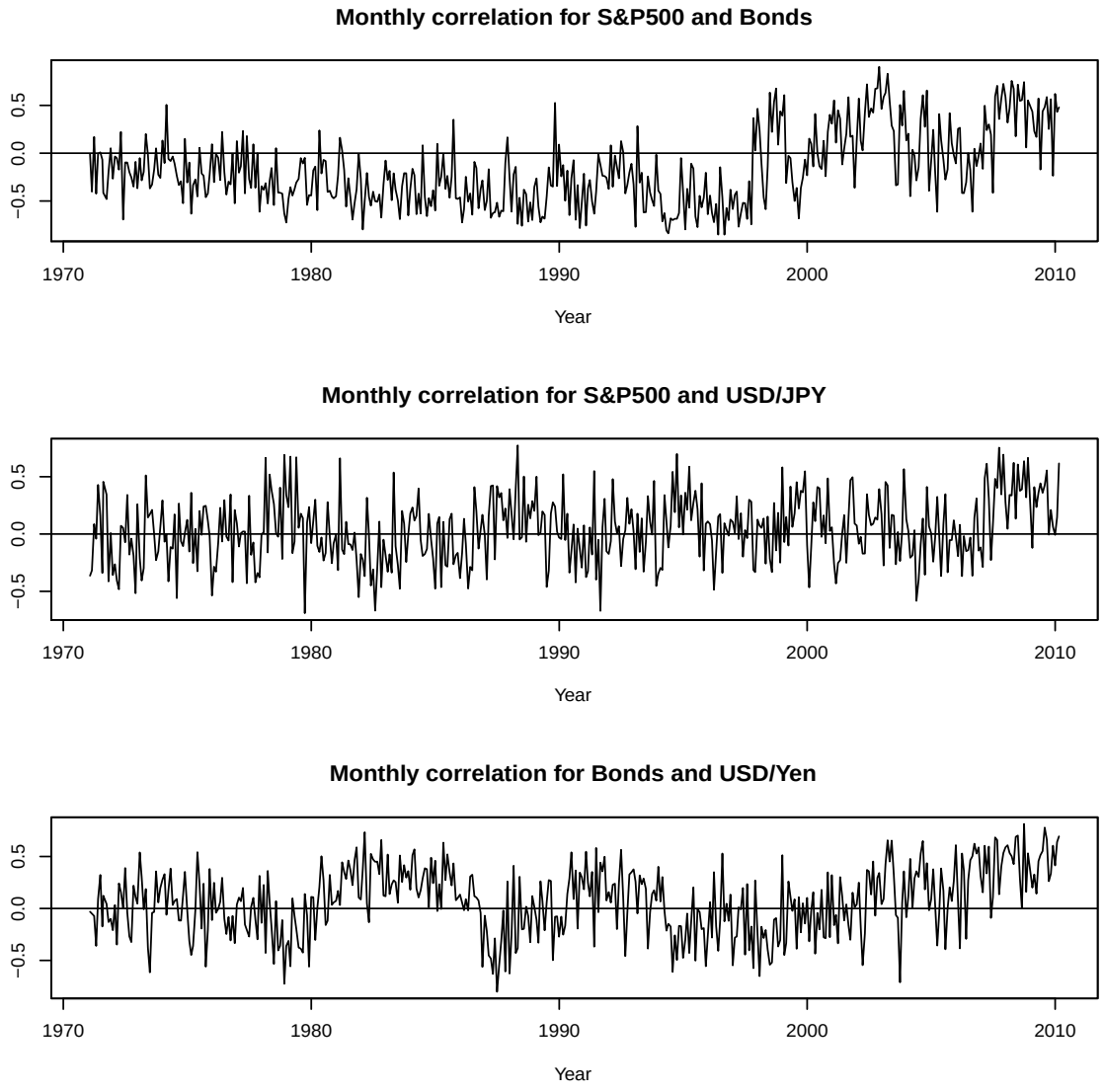


Figure 4: Monthly realized correlations

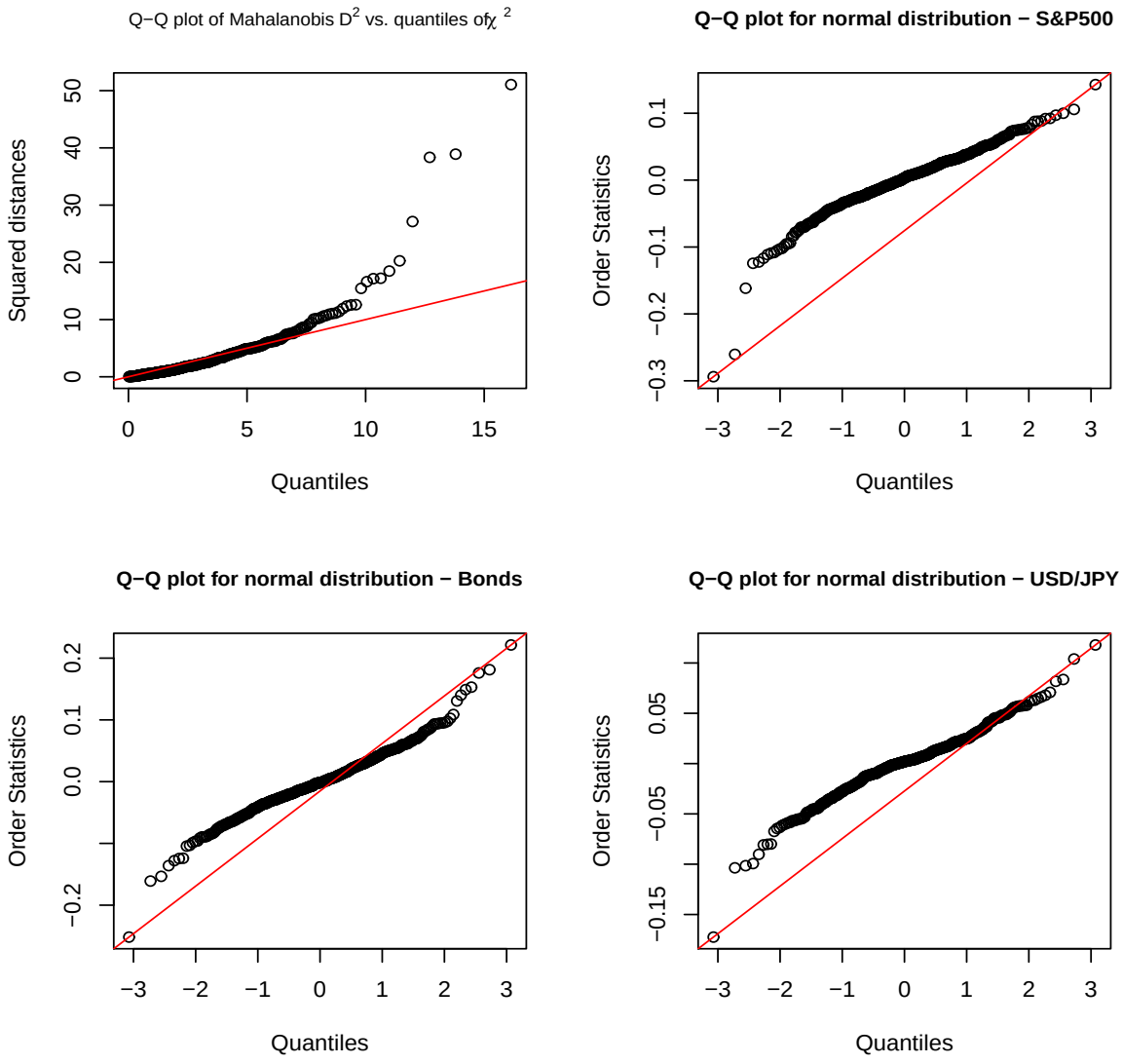


Figure 5: Normal QQ-plots, original data

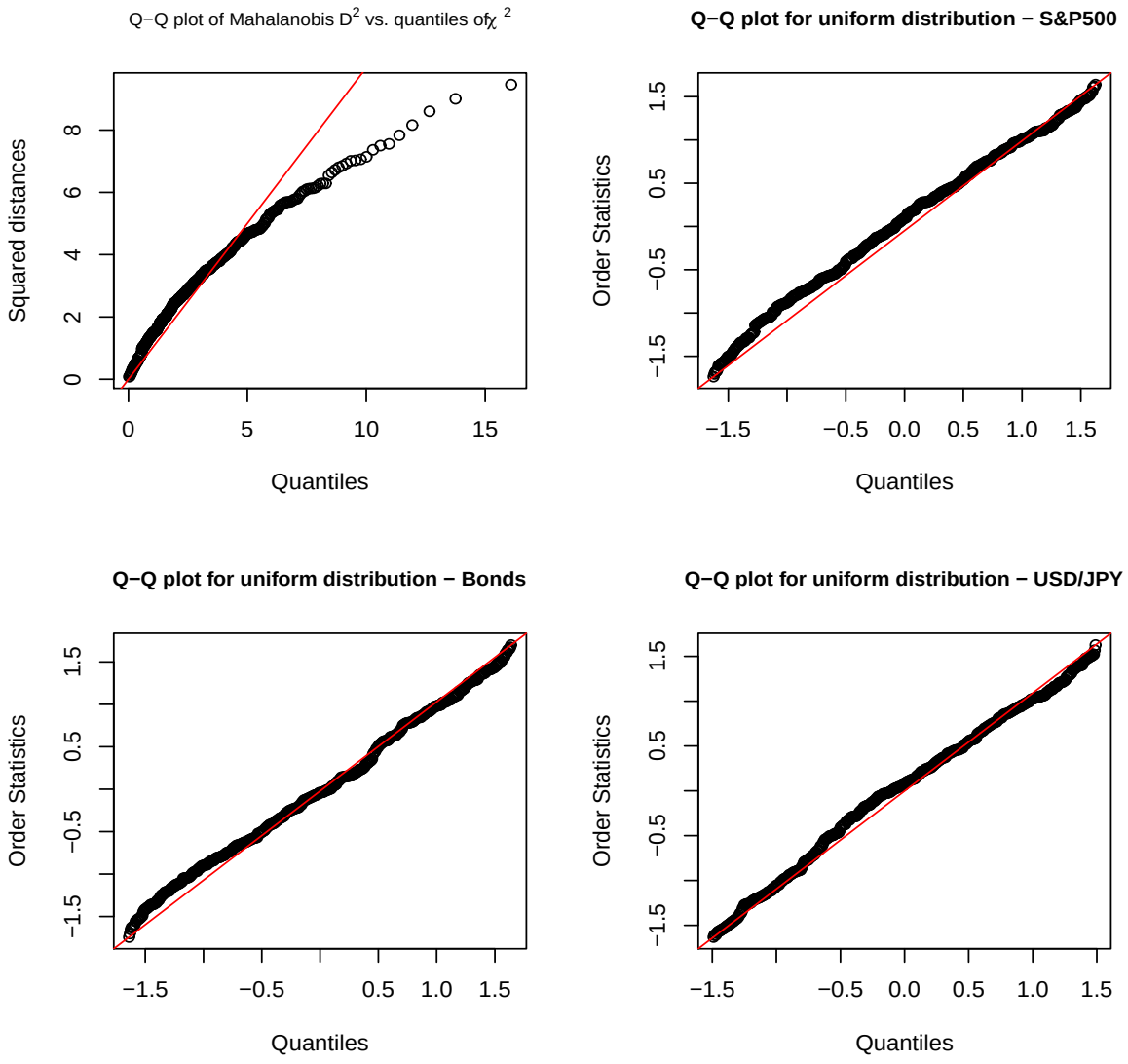


Figure 6: Uniform QQ-plots, transformed data

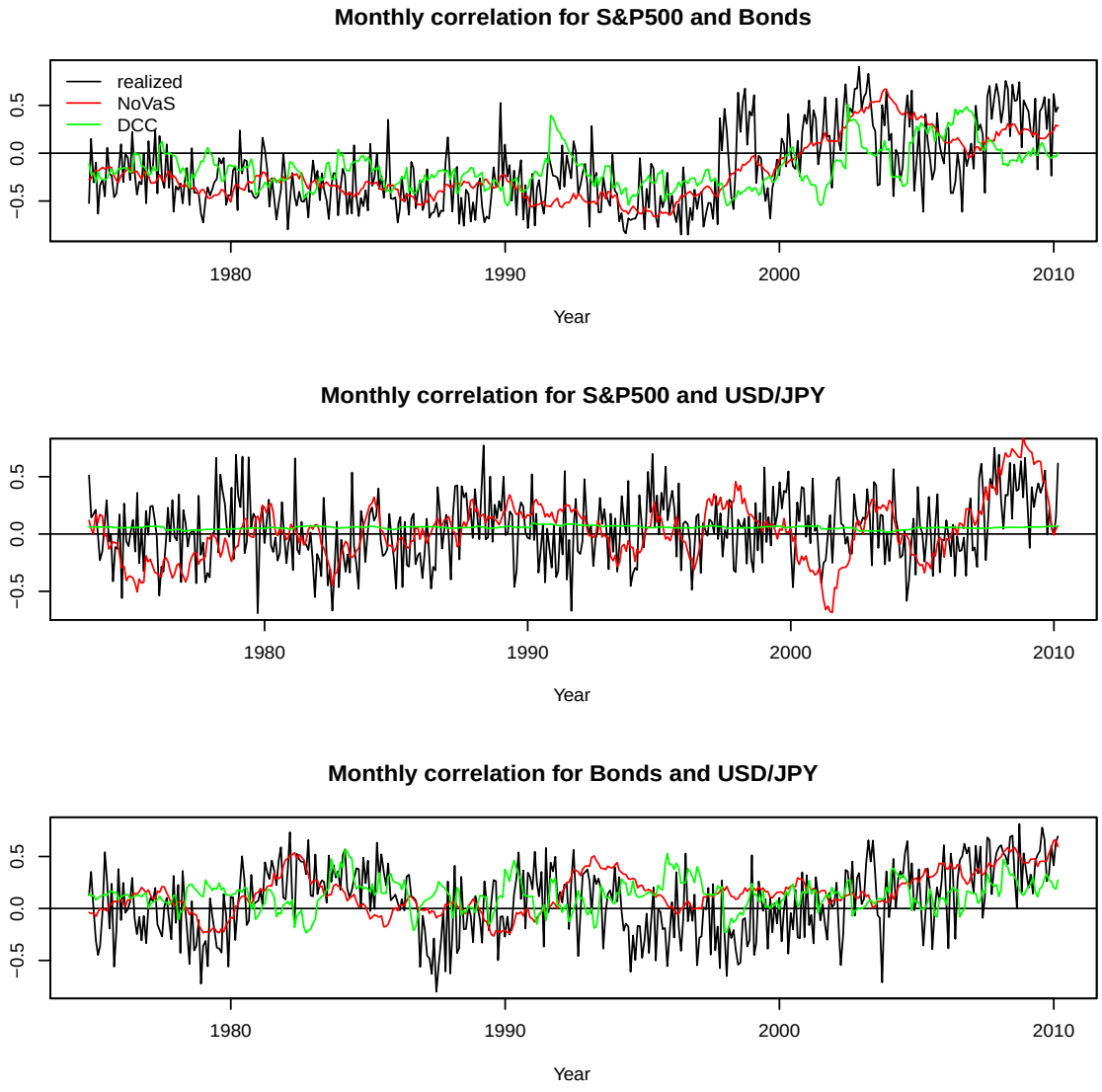


Figure 7: Monthly realized correlations and fitted values