

# Multivariate NoVaS & Inference on Conditional Correlations\*

Dimitrios D. Thomakos<sup>†</sup> , Johannes Klepsch<sup>‡</sup> and Dimitris N. Politis<sup>§</sup>

February 28, 2016

## 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”, which is why we introduce an alternative using cross validation. 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, and evaluate our methodology in the context of portfolio management.

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

---

\*Preliminary material; please do not quote without permission.

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

<sup>‡</sup>Department of Mathematical Statistics, Technische Universität München, Munich, Germany. Email: j.klepsch@tum.de

<sup>§</sup>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 and (c) 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 priori 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. We furthermore apply NoVaS to portfolio analysis.

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). For an alternative Bayesian treatment of GARCH models see Vrontos et al. (2000).

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 and a possible application of the methodology in portfolio analysis, 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’<sup>1</sup> of

---

<sup>1</sup>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  $h_t^2 \stackrel{\text{def}}{=} \mathbb{E} [X_t^2 | \mathcal{F}_{t-1}]$  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  $\mathcal{F}_{t|t-p} \stackrel{\text{def}}{=} \{X_t, X_{t-1}, \dots, X_{t-p}\}$

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

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. The function  $G(\cdot; \alpha, \mathbf{a})$  can be expressed in a variety of ways, using a parametric or a semi-parametric specification. For parsimony assume that  $G(\cdot; \alpha, \mathbf{a})$  is additive and takes the following form:

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

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  $\sigma^2 = \mathbb{E} [X_1^2]$ , or of the mean absolute deviation of the returns  $\delta = \mathbb{E} |X_1|$ , and a weighted average of the current<sup>2</sup> 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 is 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) .

---

<sup>2</sup>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:

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

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, 2006) for the univariate and for the multivariate case respectively. For example, solving for  $X_t$  in eq. (3), and using the fact that  $\gamma_t$  depends on  $X_t$ , it follows that:

$$X_t = U_t A_{t-1} \quad (4)$$

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

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

and

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

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

<sup>3</sup>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).

**Remark 2.** Eq. (4) can not only be viewed as an implied model, but does also give us a backwards transformation from  $W_t$  back to  $X_t$ . Assuming we have transformed our series  $X_t$  and are now working with  $W_t$ , we can recapture  $X_t$ , for example in the case of  $g(z) = z^2$ , by:

$$\hat{X}_{i,t} = \left( \sqrt{\alpha s_{t-1}^2 + \sum_{k=1}^p a_k X_{i,t-k}^2} \right) \frac{W_{i,t}}{\sqrt{1 - a_0 W_{i,t}^2}} \quad (7)$$

This is going to be interesting in later parts of this work.

## 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  $(\alpha, \mathbf{a})$ . The parameters  $\alpha$  and  $\mathbf{a}$  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  $j = 0, 1, \dots, p$ . 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:

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

where  $b$  is the rate; these are the ‘exponential’ NoVaS weights proposed in Politis (2003a,b, 2007). In the exponential NoVaS,  $p$  is chosen as a big enough number, such that the weights  $a_j$  are negligible for  $j > p$ .

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  $\boldsymbol{\theta} \stackrel{\text{def}}{=} (p, b) \mapsto (\alpha, \mathbf{a})$ , and 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, 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-Wilk’s 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:

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

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 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  $D_n(\boldsymbol{\theta})$ .<sup>4</sup> 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  $W_{(t)}$ ,  $W_{(1)} \leq W_{(2)} \leq \dots \leq W_{(n)}$ , 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

---

<sup>4</sup>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  $W_t = \text{sign}(X_t)$  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.

the pairs  $[Q_{(t)}, W_{(t)}]$  is a measure of distributional goodness of fit and corresponds to the well known Shapiro-Wilk test for normality, when the target distribution is the standard normal. We now have that:

$$D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} 1 - \frac{[\sum_{t=1}^n (W_{(t)} - \bar{W}_n)(Q_{(t)} - \bar{Q}_n)]^2}{[\sum_{t=1}^n (W_{(t)} - \bar{W}_n)^2] \cdot [\sum_{t=1}^n (Q_{(t)} - \bar{Q}_n)^2]} \quad (10)$$

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

$$D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \sup_t \sqrt{n} |F_t - \hat{F}_{W,t}| \quad (11)$$

Note that for any choice of the objective function we have that  $D_n(\boldsymbol{\theta}) \geq 0$  and the optimal values of the parameters are clearly determined by the condition:

$$\boldsymbol{\theta}_n^* \stackrel{\text{def}}{=} \underset{\boldsymbol{\theta}}{\text{argmin}} D_n(\boldsymbol{\theta}) \quad (12)$$

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. We start by giving a definition concerning the product of two series.

**Definition 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.

1.  $\rho \stackrel{\text{def}}{=} \mathbb{E}[Z_t] = \mathbb{E}[W_{t,i}^* W_{t,j}^*]$  is the constant correlation coefficient between the returns and can be consistently estimated by the sample mean of  $Z_t$  as  $\hat{\rho}_n \stackrel{\text{def}}{=} n^{-1} \sum_{t=1}^n Z_t$ .
2.  $\rho_{t|t-s} \stackrel{\text{def}}{=} \mathbb{E}[Z_t | \mathcal{F}_{t-s}] = \mathbb{E}[W_{t,i}^* W_{t,j}^* | \mathcal{F}_{t-s}]$ , for  $s = 0, 1$ , is the conditional correlation coefficient between the returns.<sup>5</sup>

The unconditional correlation can be estimated by the sample mean of the  $Z_t$ . The remaining task is therefore to propose a suitable form for the conditional correlation and to estimate its parameters. To stay in line with the "model-free" spirit of this paper, when choosing a method to estimate the conditional correlation, we opt for parsimony, computational simplicity and compatibility with other models in the related literature. The easiest scheme is exponential smoothing as in (14) which can compactly represented as the following autoregressive model:

$$\rho_{t|t-s} \stackrel{\text{def}}{=} \lambda \rho_{t-1|t-1-s} + (1 - \lambda) Z_{t-s} \quad (13)$$

and can therefore be estimated by:

$$\hat{\rho}_{t|t-s} = (1 - \lambda) \sum_{j=s}^{L-1+s} \lambda^{j-s} Z_{t-j} \quad (14)$$

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:

$$\hat{\rho}_{t|t-s} \stackrel{\text{def}}{=} \sum_{j=s}^{L-1+s} w_j(\lambda) B^j Z_t \equiv w(B; \lambda) Z_t \quad (15)$$

with  $B$  the backshift operator. Choosing exponential weights, as in univariate NoVaS, we have

$$w_j(\lambda) = \frac{e^{-\lambda(j-s)}}{\sum_{i=s}^{L+1-s} e^{-\lambda(i-s)}}.$$

---

<sup>5</sup>For the case that  $s = 0$  the expectation operator is formally redundant but see equation 13 and the discussion around it.

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:

$$\hat{\rho}_{t|t-s} \stackrel{\text{def}}{=} w(B; \lambda)Z_t + [1 - w(1, \lambda)]\hat{\rho}_n \quad (16)$$

Note what exactly is implied by the use of  $s = 0$  in the context of equation (13): the correlation is still conditional but now using data up to and including time  $t$ . Both  $s = 0$  and  $s = 1$  options can be used in applications with little difference in their in-sample performance; their out-of-sample performance needs to be further investigated.

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:

$$\rho_{t|t-s} \stackrel{\text{def}}{=} (\lambda + \gamma d_{t-s})\rho_{t-1|t-1-s} + (1 - \lambda - \gamma d_{t-s})Z_{t-s} \quad (17)$$

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:

$$\psi_{t|t-s} = \frac{1}{2} \log \frac{1 + \rho_{t|t-s}}{1 - \rho_{t|t-s}} \quad (18)$$

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

$$\hat{\rho}_{t|t-s} = \frac{\exp(2\psi_{t|t-s}) - 1}{\exp(2\psi_{t|t-s}) + 1} \quad (19)$$

All that is now left to do is to estimate  $\lambda$ . In the following, we will introduce two different approaches. One involves maximum likelihood estimation and is based on the distribution of the product of the two studentized series. The other is more in line with the “model-free” spirit of the NoVaS approach and uses cross-validation to measure conditional correlation.

### 3.1 Maximum Likelihood Estimation

There are some interesting properties concerning the product of two studentized series which we summarize in the following proposition.

**Proposition 1.** With Definition 1, and under the assumptions of strict stationarity and distributional matching the following holds.

1. 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:

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

and  $I_2(z; \rho)$  is the integral of the same function in the interval  $(-\infty, 0)$ . Note that the result in Graig (1936) is for the normal not truncated normal distribution; however, the truncation involved in NoVaS has a negligible effect in the validity of the result.

- (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  $\beta(\rho) \stackrel{\text{def}}{=} \sqrt{3}(1+\rho)$ . In this case, and in contrast to the previous case with a truncated normal target distribution, the result obtained is exact.

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

**Remark 3.** Proposition 1 allows for 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  $\lambda$ , as in (13).

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:

$$\hat{\lambda}_n = \operatorname{argmax}_{\lambda \in [0,1]} \sum_{t=1}^n \log f_Z(Z_t; \lambda) \quad (20)$$

**Remark 4.** Even though we do not need the explicit joint distribution of the studentized series, we still need to know the distribution of the product. Because of that and since we want to stay in the mindset of the NoVaS setting, we will introduce a second method, that is based on cross validation and does not require distributions.

### 3.2 Cross Validation (CV)

Our aim in this subsection is to find an estimate for  $\lambda$  as in (13), without using a maximum likelihood method, and without using the density of  $Z_t$ , for the reason mentioned in Remark 4. We instead use an error minimization procedure, and start by suggesting different objective functions, which we then compare for suitability.

We therefore still use eq. (13), but ignore the density of  $Z_t$ , and only rely on the data. In the following, we define an objective function  $Q(\lambda)$ , which describes how well the  $\lambda$  is globally suited to describe the conditional correlation.  $Q(\lambda)$  is then minimized with respect to  $\lambda$ , in order to find the best  $\lambda$  in (13) to capture the conditional correlation.

- **CV 1** Since  $\rho_{t|t-1} = \mathbb{E}[Z_t | \mathcal{F}_{t-1}]$ , a first intuitive approach is to define the objective function by:

$$Q(\lambda) \stackrel{\text{def}}{=} \sum_{t=1}^n (\hat{\rho}_{t|t-1} - Z_t)^2 \quad (21)$$

- **CV 2** Assume we observe the series :

$$\begin{aligned} X_{i,1}, X_{i,2}, \dots, X_{i,T} \\ X_{j,1}, X_{j,2}, \dots, X_{j,T}, X_{j,T+1}, \dots, X_{j,n} \end{aligned}$$

and transform them individually with univariate NoVaS to get:

$$\begin{aligned} W_{i,1}, W_{i,2}, \dots, W_{i,T} \\ W_{j,1}, W_{j,2}, \dots, W_{j,T}, W_{j,T+1}, \dots, W_{j,n} \end{aligned}$$

Assuming we used NoVaS with a normal target distribution, due to the properties of the multivariate normal distribution, the best estimator for  $W_{i,T+1}$  given  $W_{j,T+1}$ , is:

$$\hat{W}_{i,T+1} \stackrel{\text{def}}{=} \rho_{T+1|T} W_{j,T+1}. \quad (22)$$

Assuming now that we furthermore observe  $X_{i,T+1}, \dots, X_{i,n}$  and therefore the entire series:

$$\begin{aligned} X_{i,1}, X_{i,2}, \dots, X_{i,T}, X_{i,T+1}, \dots, X_{i,n} \\ X_{j,1}, X_{j,2}, \dots, X_{j,T}, X_{j,T+1}, \dots, X_{j,n} \end{aligned}$$

we can use the estimates  $\hat{W}_{i,k+1}$  with  $k = T, \dots, n-1$  as in 22 to get to the objective function:

$$Q(\lambda) \stackrel{def}{=} \sum_{t=T+1}^n \left( \hat{W}_{i,t} - W_{i,t} \right)^2 \quad (23)$$

In this context,  $T$  should be chosen large enough, in order to guarantee that the estimate of the conditional correlation in (13) has enough data to work with. For practical implementation, we use  $T \approx n/4$ .

- **CV 3** To account for the symmetry of the correlation, one might prefer to add to the term in (23) the symmetric term:

$$\sum_{t=T+1}^n \left( \hat{W}_{j,t} - W_{j,t} \right)^2$$

with

$$\hat{W}_{j,t} \stackrel{def}{=} \hat{\rho}_{t|t-1} W_{i,t}, \quad \text{for } t = T+1, \dots, n$$

to get to the objective function:

$$Q(\lambda) \stackrel{def}{=} \sum_{t=T+1}^n \left( \hat{W}_{i,t} - W_{i,t} \right)^2 + \sum_{t=T+1}^n \left( \hat{W}_{j,t} - W_{j,t} \right)^2 \quad (24)$$

- **CV 4** Remaining in the same state of mind as for Method 2 and 3, one might think that  $\rho_{t|t-1}$  should rather describe the dependency between  $X_{i,t}$  and  $X_{j,t}$  then between  $W_{i,t}$  and  $W_{j,t}$ . One could therefore argue, that it would be more sensible to use  $(\hat{X}_{j,t} - X_{j,t})$  as an error. Still, to get to  $\hat{X}_{j,t}$ , one has to go through  $\hat{W}_{j,t}$ , which we get by applying (22). One can then use the inverse transformation discussed in (7), namely:

$$\hat{X}_{i,t} = \left( \sqrt{\alpha s_{t-1}^2 + \sum_{k=1}^p a_k X_{i,t-k}^2} \right) \frac{\hat{W}_{i,t}}{\sqrt{1 - a_0 \hat{W}_{i,t}^2}} \quad (25)$$

Now, one can once again define the objective error function:

$$Q(\lambda) \stackrel{def}{=} \sum_{t=T+1}^n \left( \hat{X}_{i,t} - X_{i,t} \right)^2 \quad (26)$$

- **CV 5** With the same motivation as in Method 3, thus to account for the symmetry of the correlation, one could think about using:

$$Q(\lambda) \stackrel{def}{=} \sum_{t=T+1}^n \left( \hat{X}_{i,t} - X_{i,t} \right)^2 + \sum_{t=T+1}^n \left( \hat{X}_{j,t} - X_{j,t} \right)^2 \quad (27)$$

- **CV 6** A bit different is the following approach: we would like our correlation to be of the right sign. With that motivation, our objective function gets bigger if the sign of the correlation at time point  $t$  is not predicted correctly. More formally, we define the loss function  $L$ :

$$L(t) \stackrel{def}{=} \begin{cases} 1 & \text{if } \hat{W}_{i,t} W_{i,t} < 0 \\ 0 & \text{if } \hat{W}_{i,t} W_{i,t} > 0 \end{cases}$$

for  $t = T + 1, \dots, n$ , and with  $\hat{W}_{i,t}$  defined as in (22). Our objective error function is then:

$$Q(\lambda) \stackrel{def}{=} \sum_{t=T+1}^n L(t) \quad (28)$$

No matter which of the six methods is used, the goal will in every case be to choose  $\hat{\lambda}$  as in:

$$\hat{\lambda} = \underset{\lambda \in [0,1]}{\operatorname{argmin}} Q(\lambda) \quad (29)$$

Using this estimate in eq. (13) than yields the captured correlation:

$$\hat{\rho}_{t|t-s} = (1 - \hat{\lambda}) \sum_{j=s}^{L-1+s} \hat{\lambda}^{j-s} Z_{t-j}$$

**Remark 5.** Note however that the captured correlation is first of all the correlation between the series  $W_{t,i}$  and  $W_{t,j}$ . We are now interested in the correlation between  $X_{t,i}$  and  $X_{t,j}$ . To be more precise, we have an estimate  $\hat{\rho}_{t|t-s,W}$  for:

$$\rho_{t|t-s,W} \stackrel{def}{=} \mathbf{E} [W_{t,i} W_{t,j} | \mathcal{F}_{t-s}], \quad \text{for } s = 0, 1.$$

What we would like to get is an estimate  $\hat{\rho}_{t|t-s,X}$  for

$$\rho_{t|t-s,X} \stackrel{def}{=} \mathbf{E} [X_{t,i} X_{t,j} | \mathcal{F}_{t-s}], \quad \text{for } s = 0, 1$$

With eq.(7), this is in the case of  $g(z) = z^2$ :

$$\begin{aligned} \hat{\rho}_{t|t-s,X} &= \mathbb{E} \left[ \frac{\sqrt{\alpha_i s_{i,t-1}^2 + \sum_{k=1}^{p_i} a_{i,k} X_{i,t-k}^2} \frac{W_{i,t}}{\sqrt{1 - a_{i,0} W_{i,t}^2}}}{\sqrt{\alpha_j s_{j,t-1}^2 + \sum_{k=1}^{p_j} a_{j,k} X_{j,t-k}^2} \frac{W_{j,t}}{\sqrt{1 - a_{j,0} W_{j,t}^2}} \Big| \mathcal{F}_{t-s}} \right] \\ &= \sqrt{\alpha_i s_{i,t-1}^2 + \sum_{k=1}^{p_i} a_{i,k} X_{i,t-k}^2} \sqrt{\alpha_j s_{j,t-1}^2 + \sum_{k=1}^{p_j} a_{j,k} X_{j,t-k}^2} \\ &\quad \mathbb{E} \left[ \frac{W_{i,t} W_{j,t}}{\sqrt{1 - a_{i,0} W_{i,t}^2} \sqrt{1 - a_{j,0} W_{j,t}^2}} \Big| \mathcal{F}_{t-s} \right] \end{aligned}$$

Since we can not easily analytically compute that, we rather use the iid structure of the  $(W_{t,i}, W_{t,j})$ . If our target distribution is normal, we can sample from the multivariate normal distribution of the  $(W_{t,i}, W_{t,j})$  with covariance matrix:

$$\Sigma_t = \begin{pmatrix} 1 & \hat{\rho}_{t|t-s,W} \\ \hat{\rho}_{t|t-s,W} & 1 \end{pmatrix}$$

We then transform the sampled iid  $(W_{t,i}, W_{t,j})$  back to  $(\hat{X}_{t,i}, \hat{X}_{t,j})$  using the backwards transformation (25). Doing that, we can for every  $t$  construct an empirical distribution of the  $(X_{t,i}, X_{t,j})$  which we then use to compute  $\hat{\rho}_{t|t-s,X}$  using again (14).

Interestingly, practical application shows that the captured correlation  $\hat{\rho}_{t|t-s,W}$  barely differs from  $\hat{\rho}_{t|t-s,X}$ . This might be due to the fact that at least in the case of a normal target, the distribution of  $\frac{W_t}{\sqrt{1 - a_{i,0} W_t^2}}$  is actually bell shaped albeit with heavy tails. We are still investigating why this empirical finding also holds for a uniform target.

## 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, different matching functions (kurtosis, QQ-correlation and KS-statistic) and different cross validation methods to capture the conditional correlation. 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 results. In the univariate case, the best “results” were defined by the closest distributional match. In our multivariate setting we are much rather interested in the NoVaS combination that is most suited to capture the correlation.

The choice as to which matching function should be used depends on the target distribution. Even though the kurtosis for instance does make sense when opting for a normal target distribution, it is not the most intuitive choice for a uniform target. Practical experimentation suggests that using the kurtosis as a matching measure works well for the normal target, whereas the QQ-correlation coefficient is more suitable when trying to match a uniform distribution. Another important point that should be made is that we are choosing the same target distribution for both univariate series, as, since we are trying to capture correlation, differently distributed series are undesirable.

The choice as to which combination should be chosen can be made as follows. Consider fixing the type of normalization used (squared or absolute returns) and the target distribution (normal or uniform) and then calculating the correlation between the transformed series with all seven of the described methods in Sections 3.1 and 3.2. Calculate the mean squared error between this captured correlation and the realized correlation. Record the results in a  $(7 \times 1)$  vector, say  $\mathbf{D}_m(\nu, \tau)$ , where  $m = \text{Method 1}, \dots, \text{Method 6}, \text{MLE Method}$ ,  $\nu = \text{squared, absolute returns}$  and  $\tau = \text{normal, uniform target distribution}$ . Then, repeat the optimizations with respect to all seven methods for all combinations of  $(\nu, \tau)$ . The “optimal” combination is then defined across all possible combinations  $(m, \nu, \tau)$  as follows:

$$d^* \stackrel{\text{def}}{=} \operatorname{argmin}_{(m, (\nu, \tau))} \mathbf{D}_m(\nu, \tau). \quad (30)$$

Since the realized correlation is in general not known in practice, one can alternatively evaluate the quality of the captured correlation between say  $X_t$  and  $Y_t$  by using it to forecast  $X_n$  given  $Y_n$  by  $\hat{X}_n = \hat{\rho}_n Y_n$ . Then the optimal NoVaS transformation is the one that minimizes  $(X_n - \hat{X}_n)^2$ .

The choice of the truncation parameter  $L$  in (13) 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).

In what follows we apply the NoVaS transformation to return series for portfolio analysis. We consider the case of a portfolio consisting of two assets, with prices at time  $t$   $p_{1,t}$  and  $p_{2,t}$  and continuously compounded returns  $r_{1,t} = \log p_{1,t}/p_{1,t-1}$  and  $r_{2,t} = \log p_{2,t}/p_{2,t-1}$ . Denote by  $\mu_1$  and  $\mu_2$  the assumed non time varying mean returns. The variances are  $\sigma_{1,t}^2$  and  $\sigma_{2,t}^2$  and the covariance between the two assets is  $\sigma_{12,t} = \rho_{12,t} \sigma_{1,t} \sigma_{2,t}$ .

Let us further assume, that the portfolio consists of  $\beta_t$  units of asset 1 and  $(1 - \beta_t)$  units of asset 2. The portfolio return is therefore given by

$$r_{p,t} \simeq \beta_{t-1} r_{1,t} - (1 - \beta_{t-1}) r_{2,t}, \quad (31)$$

where we use the linear approximation of the logarithm, because we can expect that returns are going to be small, a setting in which this approximation works well.  $\beta$  is indexed by  $t-1$ , because the choice of the composition of the portfolio has to be made before the return in  $t$  is known. We assume that no short sales are allowed, and therefore impose that  $0 \leq \beta_t \leq 1$  for all  $t$ . The portfolio variance is given by

$$\sigma_{p,t}^2 \simeq \beta_{t-1}^2 \sigma_{1,t}^2 + (1 - \beta_{t-1})^2 \sigma_{2,t}^2 + 2\beta_{t-1}(1 - \beta_{t-1})\sigma_{12,t}. \quad (32)$$

The goal of portfolio analysis in this context is to choose  $\beta_t$ , such that the utility of the investor is maximized.

The utility of the investor is a function of the portfolio return and the portfolio variance. Assuming that the investor is risk-averse with risk aversion parameter  $\eta$ , a general form of the utility function is:

$$\begin{aligned} U(\mathbb{E}[r_{p,t}|\mathcal{F}_{t-1}], \sigma_{p,t}^2) &= \mathbb{E}[r_{p,t}|\mathcal{F}_{t-1}] - \eta \sigma_{p,t}^2 \\ &\approx \beta_{t-1}\mu_1 + (1 - \beta_{t-1})\mu_2 - \eta(\sigma_{1,t}^2 + \beta_{t-1}^2\sigma_{2,t}^2 - 2\beta_{t-1}\sigma_{12,t}) \end{aligned} \quad (33)$$

where the last equality is exact if we assume efficient markets.

A rational investor will try to maximize his utility with respect to  $\beta_{t-1}$ :

$$\begin{aligned} \frac{\partial}{\partial \beta_{t-1}} [\beta_{t-1}\mu_1 + (1 - \beta_{t-1})\mu_2 - \eta(\sigma_{1,t}^2 + \beta_{t-1}^2\sigma_{2,t}^2 - 2\beta_{t-1}\sigma_{12,t})] &\stackrel{!}{=} 0 \\ \Leftrightarrow \beta_{t-1} &= \frac{0.5\eta^{-1}(\mu_1 - \mu_2) - (\sigma_{12,t} - \sigma_{2,t}^2)}{\sigma_{1,t}^2 + \sigma_{2,t}^2 - \sigma_{12,t}} \end{aligned} \quad (34)$$

which simplifies to the minimum variance weight when we assume zero means:

$$\beta_{t-1} = \frac{\sigma_{2,t}^2 - \sigma_{12,t}}{\sigma_{1,t}^2 + \sigma_{2,t}^2 - \sigma_{12,t}}$$

Under the assumption that no short sales are allowed, one furthermore has to impose that  $0 \leq \beta_{t-1} \leq 1$ . As expected, the optimal hedge ratio depends on the correlation and can therefore be time varying.

## 5 Simulation study

In this section we report results from a limited simulation study. We use two types of simulated data: first, we use a simple bivariate model as a data generating process (DGP), as in Patton and Sheppard (2008), which we call DGP-PS, that allows for consistent realized covariances and correlations to be computed. Next, we assume two univariate GARCH models and specify a deterministic time varying correlation between them.

We start by illustrating the approach discussed in Section 4 and continue with comparing the performance of NoVaS with other standard methods from literature. Finally we conclude by applying NoVaS to portfolio analysis.

### 5.1 DGP-PS simulation

Letting  $\mathbf{R}_t \stackrel{\text{def}}{=} [X_t, Y_t]^\top$  denote the  $(2 \times 1)$  vector of returns, the DGP-PS is given as follows:

$$\begin{aligned} \mathbf{R}_t &= \Sigma_t^{1/2} \boldsymbol{\epsilon}_t \\ \boldsymbol{\epsilon}_t &= \boldsymbol{\xi}_{kt} \cdot I_2 \quad \text{with } \boldsymbol{\xi}_t \sim \mathcal{N}(0, 1) \text{ and } I_2 \text{ identity matrix} \\ \Sigma_t &= 0.05\bar{\Sigma} + 0.90\Sigma_{t-1} + 0.05\mathbf{R}_{t-1}\mathbf{R}_{t-1}^\top \end{aligned} \tag{35}$$

where  $\bar{\Sigma}$  is a  $(2 \times 2)$  matrix with unit diagonals and off-diagonals entries of 0.3. We let  $t = 1, \dots, 1000$ .

We use the model selection approach of the previous section. We compute  $\mathbf{D}_m(\nu, \tau)$  of (30) for all  $m$  and  $(\nu, \tau)$  and repeat the calculations 1000 times. We summarize the mean squared error between the realized correlation  $\rho_{t|t-s}$  and our estimated conditional correlation  $\hat{\rho}_{t|t-s}$  in all 28 combinations in Table 5.1. We use the specified NoVaS transformation, where the kurtosis

	Normal Target squared returns	Normal Target absolute returns	Uniform Target squared returns	Uniform Target absolute returns
<b>MLE</b>	2.09E-02	<b>2.08E-02</b>	3.18E-02	3.27E-02
<b>CV 1</b>	2.44E-02	2.44E-02	2.32E-02	2.37E-02
<b>CV 2</b>	2.24E-02	2.25E-02	2.29E-02	2.34E-02
<b>CV 3</b>	3.63E-02	3.89E-02	3.13E-02	2.58E-02
<b>CV 4</b>	2.22E-02	2.22E-02	2.34E-02	2.28E-02
<b>CV 5</b>	3.23E-02	3.39E-02	2.98E-02	5.78E-02
<b>CV 6</b>	8.10E-02	8.18E-02	8.80E-02	1.02E-01

Table 1: Model selection on DGP-PS, 1000 iterations. Table entries are the MSE between  $\rho_{t|t-s}$  and  $\hat{\rho}_{t|t-s}$ . Smallest MSE is presented in bold characters.

as in (9) was used when fitting to a normal target, and the QQ-correlation as in (11) was used when fitting to a uniform target. Furthermore we set  $s = 0$  and used exponential weights.

The NoVaS transformation with normal target and absolute returns, combined with the MLE method to estimate the correlation yields the best result. Using a normal target with squared returns combined with methods CV 2 and CV 4 perform competitively. One can expect that the good performance of the MLE compared to the other methods is due to the Gaussian structure of the data. 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 of the NoVaS approach can be different. We explore this in the Section 6.

We now focus on a different type of simulated data with deterministic and time-varying correlation.

## 5.2 Multivariate normal returns

We now assume that our bivariate return series follows a multivariate normal distribution, where the variances are determined by two volatility processes that follows GARCH dynamics. At the same time we specify a deterministic correlation process between the two return series. More precisely:

$$\begin{aligned}
 \mathbf{R}_t &\stackrel{\text{def}}{=} [X_t, Y_t]^\top \\
 \mathbf{R}_t &\sim \mathcal{N}(0, \mathbf{H}_t) \\
 \mathbf{H}_t &= \begin{pmatrix} \sigma_{1t}^2 & \rho_{i,t}\sigma_{1t}\sigma_{2t} \\ \rho_{i,t}\sigma_{1t}\sigma_{2t} & \sigma_{2t}^2 \end{pmatrix}, \quad i = 1, 2,
 \end{aligned} \tag{36}$$

where

$$\begin{aligned}
 \sigma_{1,t}^2 &= 0.01 + 0.05X_{1,t-1}^2 + 0.94\sigma_{1,t-1}^2 \\
 \sigma_{2,t}^2 &= 0.05 + 0.20X_{2,t-1}^2 + 0.50\sigma_{2,t-1}^2 \quad \text{and} \\
 \rho_{1,t} &= 0.5 + 0.4 \cos(2\pi t/400) \quad \text{or} \\
 \rho_{2,t} &= \text{mod}(t/300).
 \end{aligned}$$

Both examples of  $\rho_t$ , the first implying a sinusoidal correlation, the second a linearly increasing correlation, will be examined.

For both multivariate processes, we again compute the  $\mathbf{D}_m(\nu, \tau)$ . We repeat the computations 1000 times in order to get a robust idea of which method works best. Table 5.2 shows the mean squared error between the real deterministic correlation and the estimates using the 28 different NoVaS methods.

As we can see in Table 5.2, in the case of a sinusoidal correlation structure, using NoVaS with a uniform target distribution and absolute returns seems to be working best when using the MLE method to capture the correlation, but CV 2 and CV 4 and absolute returns perform competitively. In the case of the linearly increasing correlation, one should again use uniform target either with squared returns and CV 4 or absolute returns and CV2. Interestingly, in this case, using a uniform target distribution clearly outperforms the normal target. We show plots of the resulting estimated correlation in Figure 1.

	Normal Target squared returns	Normal Target absolute returns	Uniform Target squared returns	Uniform Target absolute returns
<b>Sinusoidal correlation</b>				
<b>MLE</b>	5.07E-02	5.01E-02	<b>2.38E-02</b>	2.42E-02
<b>CV 1</b>	3.47E-02	3.46E-02	2.63E-02	2.60E-02
<b>CV 2</b>	3.11E-02	3.10E-02	2.51E-02	2.45E-02
<b>CV 3</b>	1.07E-01	1.01E-01	9.36E-02	5.93E-02
<b>CV 4</b>	3.08E-02	3.07E-02	2.45E-02	2.51E-02
<b>CV 5</b>	1.38E-01	1.22E-01	1.04E-01	1.09E-01
<b>CV 6</b>	7.72E-02	7.71E-02	6.36E-02	6.40E-02
<b>Linear correlation</b>				
<b>MLE</b>	6.00E-02	5.93E-02	4.88E-02	4.76E-02
<b>CV 1</b>	5.78E-02	5.73E-02	5.02E-02	4.89E-02
<b>CV 2</b>	5.47E-02	5.45E-02	4.89E-02	4.72E-02
<b>CV 3</b>	8.71E-02	8.47E-02	9.29E-02	7.12E-02
<b>CV 4</b>	5.44E-02	5.42E-02	<b>4.71E-02</b>	4.88E-02
<b>CV 5</b>	9.10E-02	8.90E-02	9.33E-02	1.07E-01
<b>CV 6</b>	9.88E-02	9.78E-02	9.18E-02	9.02E-02

Table 2: Model Selection on multivariate normal returns with specified correlation, 1000 iterations. Table entries are the MSE between  $\rho_{t|t-s}$  and  $\hat{\rho}_{t|t-s}$ . Smallest MSE is presented in bold characters

### 5.3 Comparison of NoVaS to standard methods and portfolio analysis

We now evaluate the performance of NoVaS . To do that we compare the error between the correlation captured by NoVaS to the realized correlation to the error made when capturing the correlation with standard methods from literature. We use baseline methods like a naive approach, where the hedge ratio  $\beta_t$  is defined to be constantly 0.5, and a linear model, where the hedge ratio is defined through linear regression. We furthermore compare NoVaS to GARCH based methods like DCC (Engle (2002)), BEKK (Engle and Kroner (1995)) and CCC (Bollerslev (1990)). For an overview of these methods, one can look at Bos and Gould (2007).

In Table 3, we calculate the mean-squared error between captured correlation and realized correlation of the simulation examples as before. We average over 1000 repetitions of the simulations. NoVaS in Table 3 corresponds to the method that performed best in capturing the correlation according to Tables 5.1 and 5.2 (hence normal target and absolute returns, combined with the MLE method for the DGP-PS data; uniform target with squared returns and the MLE method for sinusoidal correlation; and uniform target with absolute returns and CV 4 for linearly

increasing correlation).

	<b>DGP-PS</b>		<b>MVN-sinusoidal</b>		<b>MVN-linear</b>	
	MSE	Cor	MSE	Cor	MSE	Cor
NoVaS	6.83E-02	6.50E-01	5.75E-02	7.16E-01	7.05E-02	7.16E-01
BEKK	1.91E-02	5.46E-01	3.28E-02	7.55E-01	4.94E-02	6.24E-01
DCC	1.39E-02	7.27E-01	2.07E-02	8.68E-01	4.26E-02	7.08E-01
CCC	3.44E-02	NA	7.91E-02	NA	8.16E-02	NA

Table 3: MSE and correlation between realized correlation and estimated correlation by the respective method averaged over 1000 simulations. NoVaS corresponds to the best method according to Tables 5.1 and 5.2.

Table 3 shows that NoVaS gets outperformed by the classic DCC and BEKK approaches with all three types of simulated data, when considering the MSE between realized and estimated correlation. However, considering the structure of the simulated datasets, NoVaS performs better than expected, especially when considering the correlation between realized and estimated correlation. We expect that NoVaS will perform even better on datasets with heavier tails and less structure.

We now apply NoVaS to portfolio analysis. We use the correlation captured by the different methods above, in order to calculate the optimal hedge ratio defined in (34). More precisely, the following algorithm is applied:

**Algorithm 1**

1. Assume we observe  $X_{t,1}$  and  $X_{t,2}$  for  $t = 1, \dots, N$ . Fix an window size  $T_0$ , for instance  $T_0 = 1/3N$ .
2. For every  $T_0 \leq k \leq N - 1$ , estimate the correlation  $\rho_{12,k}$  based on  $X_{t,1}$  and  $X_{t,2}$ ,  $t = (k + 1 - T_0), \dots, k$ , using the methods introduced above.
3. Use (34) to derive the optimal portfolio weights, and the portfolio returns.

We compute mean, standard deviation and Sharpe ratio of the portfolio returns. The results are shown in Table 4. We again repeat the simulations 1000 times and show average results.

One should not be surprised by the negative mean of the returns, since the portfolio choice is always made with the main aim of variance minimization. So the focus should mainly lie on the standard deviation of the returns. Surprisingly, the DCC method gets outperformed by all other methods in the DGP-PS setting. The naive approach, where no variance minimization took place, yields the worst results in all scenarios. NoVaS performs very competitively with all

	DGP-PS			MVN-sinusoidal			MVN-linear		
	Mean	St.Dev.	Sh.R.	Mean	St.Dev.	Sh.R.	Mean	St.Dev.	Sh.R.
<b>NoVaS</b>	-1.39e-3	0.99	-1.43e-3	-1.30e-3	0.86	-1.28e-3	-2.08e-3	0.85	-2.39e-3
<b>DCC</b>	-6.93e-4	1.20	-5.79e-4	-1.34e-3	0.87	-1.26e-3	-1.41e-3	0.84	-1.43e-3
<b>BEKK</b>	-9.89e-4	0.98	-1.02e-3	-1.25e-3	0.84	-1.28e-3	-1.78e-3	0.83	-1.98e-3
<b>CCC</b>	-9.59e-4	0.99	-9.83e-4	-1.36e-3	0.87	-1.35e-3	-1.69e-3	0.85	-1.83e-3
<b>Linear Model</b>	-9.56e-4	0.99	-9.82e-4	-1.15e-3	0.86	-1.14e-3	-1.30e-3	0.84	-1.49e-3
<b>Naive</b>	-9.40e-4	1.34	-7.54e-4	-6.71e-4	1.21	-4.83e-4	-6.94e-4	1.16	-5.95e-4

Table 4: **TO BE UPDATED! SIMULATIONS ARE STILL RUNNING** Mean, standard deviation and sharpe ratio of portfolio returns, where the hedge ratio is based on the methods in the left column. NoVaS stands for the NoVaS method that for the specific dataset had the most convincing results when capturing the correlation.

others methods. In Figure 1 we show plots of the captured correlation by the different methods of exemplary simulated data. The plots show that whenever the MLE method is chosen to capture the correlation in the NoVaS setting, the curve representing the correlation seems to be smoother, and to underestimate the peaks. This is shown in the datasets of the DGP-PS series and the MVN series with sinusoidal correlation. In all three cases, DCC and BEKK provide similar estimates.

But once again, one should not forget that we are dealing with very structured datasets with normal innovation processes. This is something that should be beneficial to parametric GARCH type methods. In the next section, we will observe how NoVaS performs under conditions where the data has heavier tails and the dynamics of the dataset are unknown.

## 6 Empirical illustration

In this section we offer a brief empirical illustration of the NoVaS -based correlation estimation using two data sets. First data from the following three series: the S&P500, the 10-year bond<sup>6</sup> and the USD/Japanese Yen exchange rate, then, to assess the performance on smaller samples, returns of SPY and TLT. 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  $n_1 = 469$  available observations for the first three series, and from 08/2002 to 02/2016 for the second sample for total of  $n_2 = 135$

<sup>6</sup>10-year Treasury constant maturity rate series

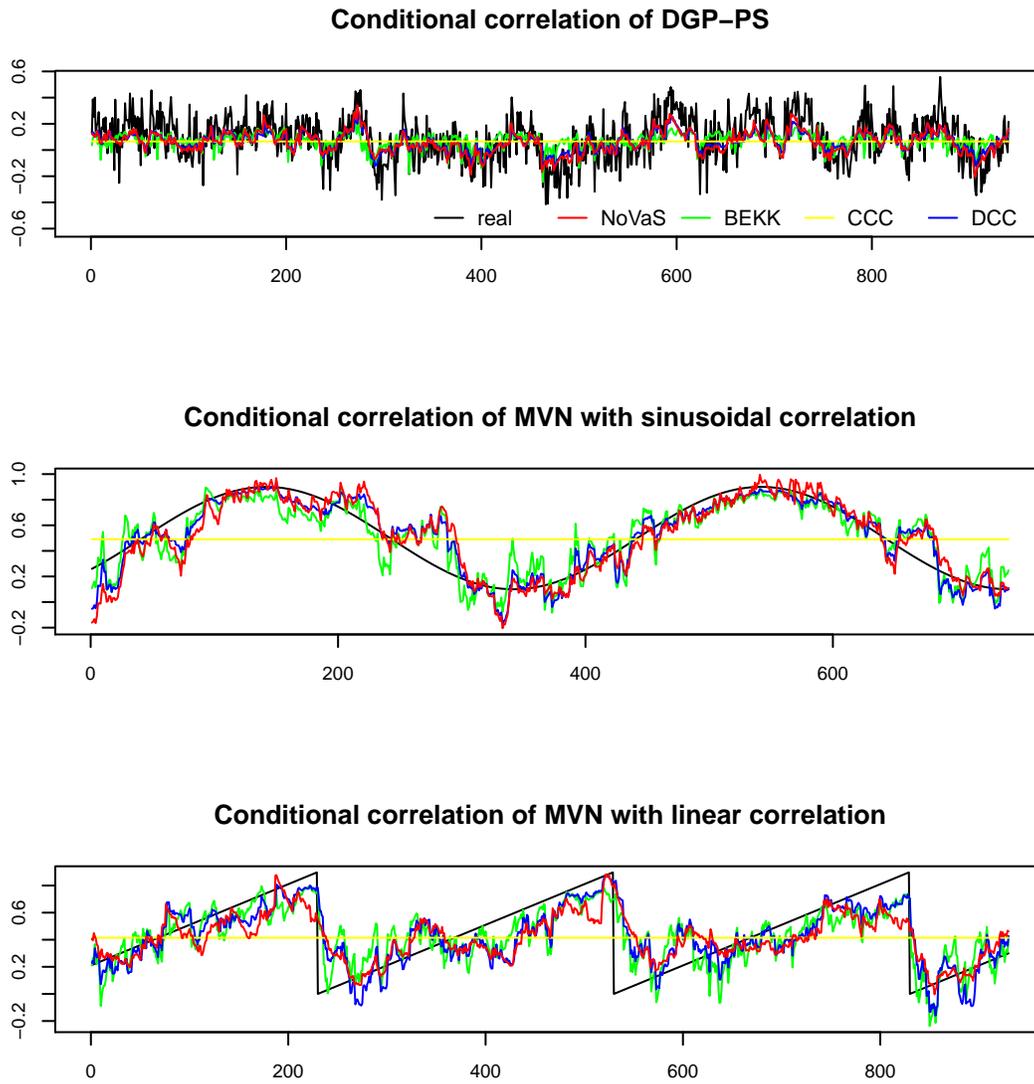


Figure 1: Comparison of the different methods to capture correlation

Figures 3, 4 and 5 plot the monthly returns, realized volatilities and correlations and Table 5 summarizes some descriptive statistics. From Table 5 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.

	S&P500	Bonds	USD/JPY	S&P500	Bonds	USD/JPY	S&P500	S&P500	Bonds
	ret.	ret.	ret.	vol.	vol.	vol.	Bonds	USD/JPY	USD/JPY
							corr.	corr.	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

Table 5: Descriptive Statistics for monthly data, sample size is  $n = 469$  months from 01/1970 to 02/2010, SW short for Shapiro Wilks

After performing univariate NoVaS transformation of the return series individually, we move on to compute the NoVaS -based correlations. We use exponential weights as in equations (15) and (16) with  $s = 0$  and  $L$  set to a multiple of the lags used in the individual NoVaS transformations (results are similar when we use  $s = 1$ ). Applying the model selection approach of Section 4, we look for the optimal combination of target distribution, squared or absolute returns and the method to capture the correlation. The results are summarized in Table 6.

When working with the S&P500 and Bonds dataset, the MSE between realized and estimated correlation is minimized when using the MLE Method, together with uniform target and absolute returns. In the other two cases, normal target and squared returns yield better results. The dataset with S&P500 and USD/YEN works best with CV 4, whereas the Bonds and USD/YEN dataset works better with CV 2.

We now assess the performance of NoVaS using the same measures as in the simulation study and compare the results with the same methods as before. Table 7 summarizes the results and Figure 6 plots the realized correlations along with the fitted values from different benchmark methods.

The table entries show that the NoVaS approach provides better estimates of the conditional correlation than the other models. For all three datasets the mean-squared error when using

	Normal Target squared returns	Normal Target absolute returns	Uniform Target squared returns	Uniform Target absolute returns
<b>S&amp;P500-Bonds</b>				
<b>MLE</b>	1.32E-01	1.16E-01	9.13E-02	<b>9.05E-02</b>
<b>CV 1</b>	1.41E-01	9.48E-02	9.79E-02	1.01E-01
<b>CV 2</b>	1.59E-01	9.50E-02	9.77E-02	1.01E-01
<b>CV 3</b>	1.28E-01	1.23E-01	1.26E-01	1.26E-01
<b>CV 4</b>	1.55E-01	9.49E-02	9.78E-02	1.01E-01
<b>CV 5</b>	1.28E-01	1.23E-01	1.26E-01	1.26E-01
<b>CV 6</b>	1.81E-01	1.69E-01	1.69E-01	1.67E-01
<b>S&amp;P500-USD/JPY</b>				
<b>MLE</b>	7.53E-02	7.69E-02	9.68E-02	1.03E-01
<b>CV 1</b>	7.46E-02	7.57E-02	7.44E-02	7.69E-02
<b>CV 2</b>	7.39E-02	7.56E-02	7.55E-02	7.73E-02
<b>CV 3</b>	7.58E-02	7.86E-02	7.70E-02	7.93E-02
<b>CV 4</b>	<b>7.37E-02</b>	7.55E-02	7.54E-02	7.73E-02
<b>CV 5</b>	7.58E-02	7.86E-02	7.70E-02	7.93E-02
<b>CV 6</b>	2.34E-01	2.10E-01	3.34E-01	3.51E-01
<b>Bonds-USD/JPY</b>				
<b>MLE</b>	1.13E-01	1.08E-01	1.05E-01	1.11E-01
<b>CV 1</b>	9.95E-02	1.02E-01	1.02E-01	1.00E-01
<b>CV 2</b>	<b>9.92E-02</b>	1.00E-01	1.02E-01	9.96E-02
<b>CV 3</b>	1.09E-01	1.10E-01	1.06E-01	1.05E-01
<b>CV 4</b>	1.00E-01	1.02E-01	1.02E-01	9.97E-02
<b>CV 5</b>	1.09E-01	1.10E-01	1.06E-01	1.05E-01
<b>CV 6</b>	2.75E-01	3.15E-01	3.51E-01	3.40E-01
<b>SPY-TLT</b>				
<b>MLE</b>	9.06E-02	8.81E-02	1.24E-01	1.28E-01
<b>CV 1</b>	9.06E-02	9.76E-02	9.20E-02	9.75E-02
<b>CV 2</b>	9.27E-02	9.34E-02	1.02E-01	1.10E-01
<b>CV 3</b>	9.27E-02	9.81E-02	9.66E-02	1.03E-01
<b>CV 4</b>	9.27E-02	9.81E-02	9.60E-02	1.02E-01
<b>CV 5</b>	9.27E-02	9.81E-02	9.66E-02	1.03E-01
<b>CV 6</b>	1.15E-01	1.28E-01	1.56E-01	2.67E-01

Table 6: Multivariate NoVaS model selection: MSE between  $\rho_{t|t-1}$  and  $\hat{\rho}_{t|t-1}$

	<b>S&amp;P500-Bonds</b>		<b>S&amp;P500-USD/YEN</b>		<b>Bonds-USD/YEN</b>	
	MSE	Corr	MSE	Corr	MSE	Corr
<b>NoVaS</b>	9.05e-2	6.61e-1	7.39e-2	3.06e-1	9.92e-2	3.64e-1
<b>DCC</b>	1.05e-1	5.51e-1	8.34e-2	9.69e-2	1.03e-1	2.98e-1
<b>BEKK</b>	1.16e-1	4.68e-1	9.83e-2	7.34e-2	1.21e-1	1.05e-1
<b>CCC</b>	1.50e-1	-	8.22e-2	-	1.10e-1	-

Table 7: Mean-Squared-Error and correlation between realized correlation and captured correlation by the different methods. The NoVaS transformation used is the one that performed best in Table 6

NoVaS is either better or on par with the other methods, and the correlation between realized and estimated correlation is larger.

Our results are, of course, conditional on the data at hand and the benchmark models. 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 than 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 for 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.

We finally want to give the results of the proposed application in portfolio analysis. As in the previous section with simulated data, we use the captured correlation in order to calculate the optimal hedge ratio, as in (34). We then compare mean, standard deviation and Sharpe ratio of the portfolio returns. We use Algorithm 1, and set  $T_0 = 1/3N$  and  $\eta = 3$ . Table 8 summarizes the results.

The results of Table 8 show that NoVaS is again the most successful in minimizing the variance, as the standard deviation of the portfolio returns is smallest when using the NoVaS hedge ratio, at least in the first two datasets. Furthermore, the Sharpe ratio is highest when using NoVaS correlation in the first two datasets. The plots in Figure 6 might explain why the results on portfolio returns are not convincing when looking at Bonds and the USD/YEN exchange rate: the general level of correlation is quite low, which makes it more difficult to construct an effective minimum variance portfolio. Other than that, as seen before, the NoVaS correlation seems to be smoother than other methods.

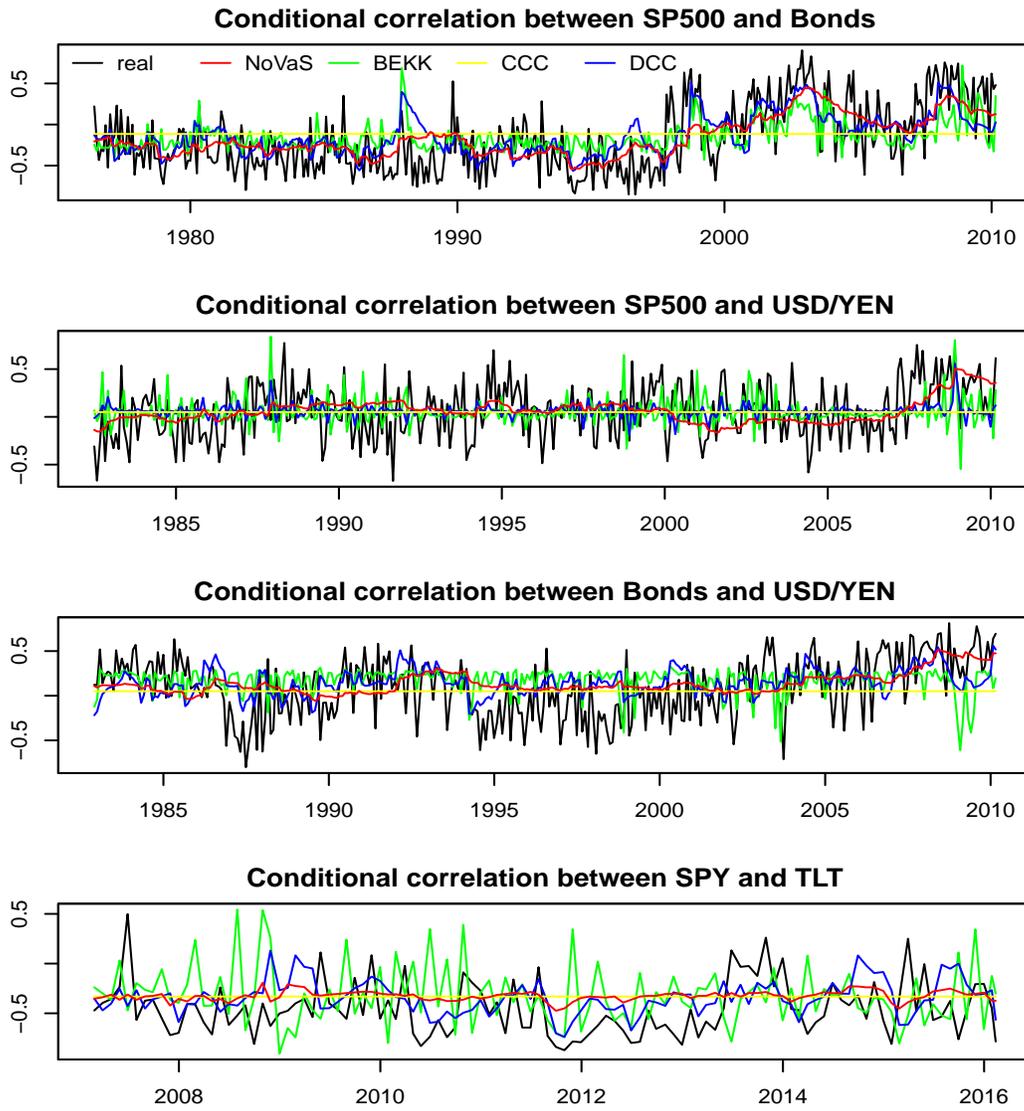


Figure 2: Monthly realized correlations and fitted values

	S&P500-Bonds			S&P500-USD/YEN			Bonds-USD/YEN		
	Mean	St.Dev.	Sh.R.	Mean	St.Dev.	Sh.R.	Mean	St.Dev.	Sh.R
<b>NoVaS</b>	5.71E-02	1.26E-01	<b>4.53E-01</b>	7.70E-02	1.13E-01	6.82E-01	-5.31E-02	1.07E-01	-4.97E-01
<b>DCC</b>	5.37E-02	1.27E-01	4.23E-01	7.75E-02	1.14E-01	6.77E-01	-5.03E-02	1.06E-01	-4.76E-01
<b>BEKK</b>	5.28E-02	1.28E-01	4.14E-01	<b>7.88E-02</b>	1.14E-01	<b>6.89E-01</b>	-4.98E-02	1.06E-01	-4.69E-01
<b>CCC</b>	5.62E-02	1.29E-01	4.37E-01	4.27E-02	1.18E-01	3.61E-01	-5.15E-02	1.12E-01	-4.60E-01
<b>LM</b>	5.26E-02	1.26E-01	4.17E-01	7.61E-02	1.14E-01	6.67E-01	-5.00E-02	<b>1.05E-01</b>	-4.74E-01
<b>Naive</b>	4.26E-03	<b>1.23E-01</b>	3.47E-02	1.13E-02	<b>1.02E-01</b>	1.11E-01	-4.68E-02	1.17E-01	-3.98E-01
<b>Asset 1</b>	<b>6.23E-02</b>	1.60E-01	3.90E-01	6.23E-02	1.60E-01	3.90E-01	-5.38E-02	1.88E-01	-2.87E-01
<b>Asset 2</b>	-5.38E-02	1.88E-01	-2.87E-01	-3.97E-02	1.13E-01	-3.51E-01	<b>-3.97E-02</b>	1.13E-01	-3.51E-01

Table 8: Mean, standard deviation and Sharpe ratio of portfolio returns, where the portfolio weights are chosen as in (34), with the correlation estimated with the methods in the left column. The NoVaS transformation used is the one that performed best in Table 6. Naive assumes an equally weighted portfolio. Linear model estimates the conditional correlation based on linear regression

## 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. 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. We furthermore evaluate the NoVaS based correlations in the context of portfolio management. 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) additional comparisons of NoVaS -based correlations with other benchmark 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] Bos, C.S., Gould, P., 2007. “Dynamic Correlations and Optimal Hedge Ratios”, *Tinbergen Institute Discussion Paper, Department of Econometrics and O.R., Vrije Universiteit Amsterdam*,
- [6] Caporin, M. and McAleer, M., 2010. “Do We Really Need Both BEKK and DCC? A Tale of Two Multivariate GARCH Models”, working paper.
- [7] Craig, C., 1936. “On the Frequency Function of  $xy$ ”, *The Annals of Mathematical Statistics*, Volume 7, Number 1 (1936), 1-15.
- [8] Engle, R., 2002. “Dynamic Conditional Correlation: A simple class of multivariate GARCH models”, *Journal of Business and Economic Statistics*, vol. 17.
- [9] Engle, R., and Kroner, F., 1995. “Multivariate Simultaneous Generalized ARCH”, *Econometric Theory*, vol. 11, pp. 122-150.
- [10] Ghysels, E. and L. Forsberg, 2007. “Why Do Absolute Returns Predict Volatility So Well?”, *Journal of Financial Econometrics*, vol. 5, pp. 31-67.
- [11] Hafner, C. and P. Franses, 2009. “A Generalized Dynamic Conditional Correlation model with Applications to Many Asset Returns”, working paper.
- [12] Hafner, C., Dijk, D., and P. Franses, 2004. “Semi-parametric Modelling of Correlation Dynamics”, working paper.
- [13] 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.

- [14] Long, X. and A. Ullah, 2005. “Nonparametric and Semiparametric Multivariate GARCH Model”, working paper.
- [15] Mardia, K., 1970. “Measures of Multivariate Skewness and Kurtosis with Applications”, *Biometrika*, vol.57, pp. 519-530.
- [16] Palandri, A. 2009. “Sequential Conditional Correlations: Inference and Evaluation”, *Journal of Econometrics*, vol 153, pp. 122-132.
- [17] Patton, A., 2006. “Modeling Asymmetric Exchange Rate Dependence”, *International Economic Review*, vol. 47, pp. 527-556.
- [18] Patton, A. and K. Sheppard, 2008. “Evaluating Volatility and Correlation Forecasts” OFRC Working Papers Series 2008fe22.
- [19] Pelletier, D., 2006. “Regime Switching for Dynamic Correlations”, *Journal of Econometrics*, vol. 131, pp. 445-473.
- [20] Politis, D.N., 2003a. “Model-Free Volatility Prediction”, UCSD Dept. of Economics Discussion Paper 2003-16.
- [21] 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.
- [22] 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.
- [23] Politis, D.N., 2006. “A multivariate heavy-tailed distribution for ARCH/GARCH residuals”, *Advances in Econometrics*, vol. 20, part A, pp. 105-124.
- [24] Politis, D.N., 2007. “Model-free vs. model-based volatility prediction”, *J. Financial Econometrics*, vol. 5, pp. 358-389.
- [25] 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.
- [26] 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>

- [27] Rohatgi, V.K., 1976. *An Introduction to Probability Theory & Mathematical Statistics*. Wiley, New York.
- [28] Royston, J.P., 1983. "Some Techniques for Assessing Multivariate Normality based on the Shapiro-Wilk W", *Applied Statistics*, vol. 32, pp. 121-133.
- [29] Silvennoinen, A. and T. Terasvirta, 2005. "Multivariate Autoregressive Conditional Heteroskedasticity with Smooth Transitions in Conditional Correlations", working paper.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] Wang, J. and R. Serfling, 2005. "Nonparametric Multivariate Kurtosis and Tailweight Measures", *Journal of Nonparametric Statistics*, vol. 17, pp. 441-456.
- [34] Vrontos, I.D., Dellaportas, P. and D.N. Politis, 2000. "Full Bayesian Inference for GARCH and EGARCH Models", *Journal of Business and Economic Statistics*, vol. 18, pp. 187-198.

## Appendix

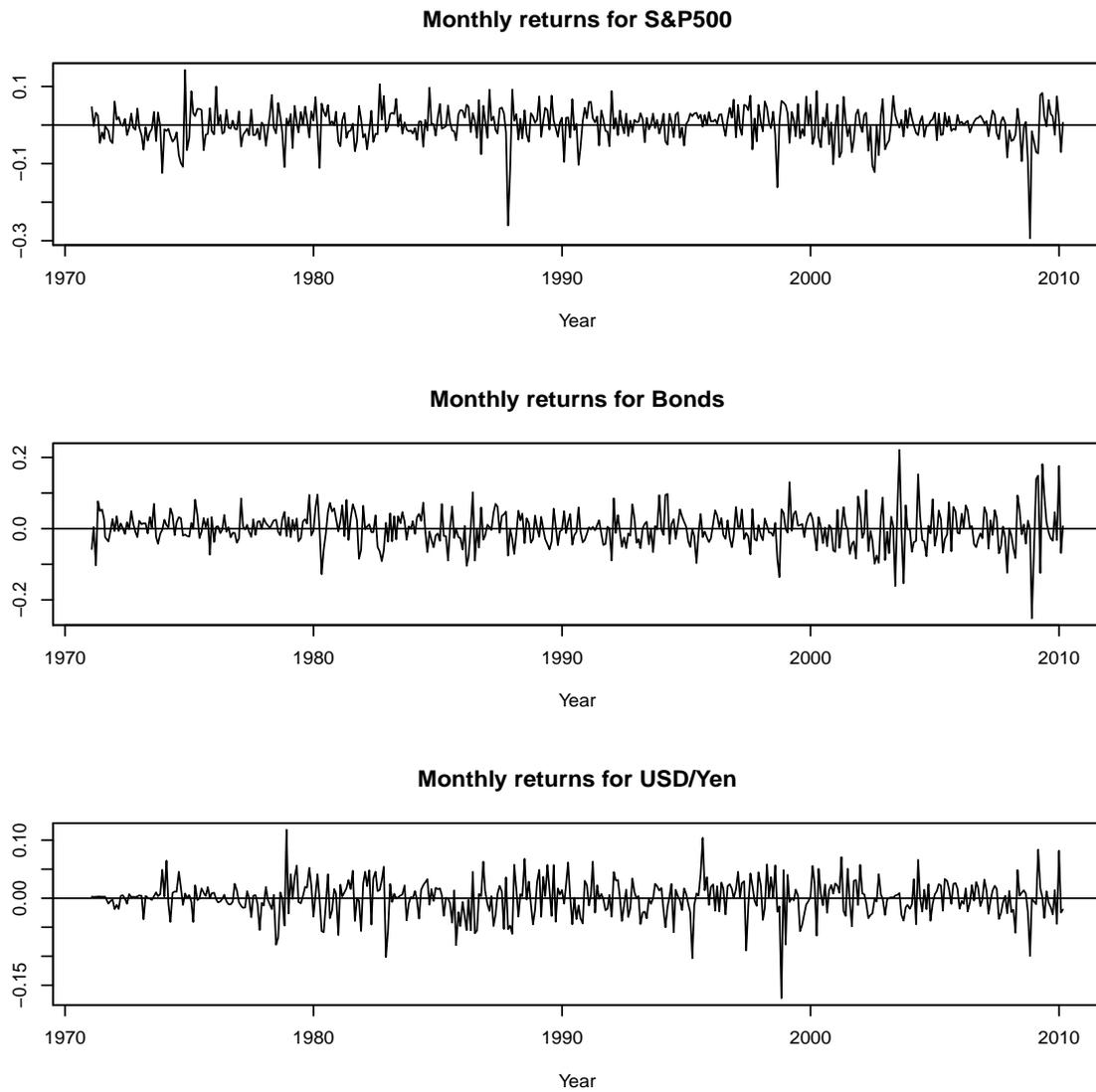


Figure 3: Monthly returns

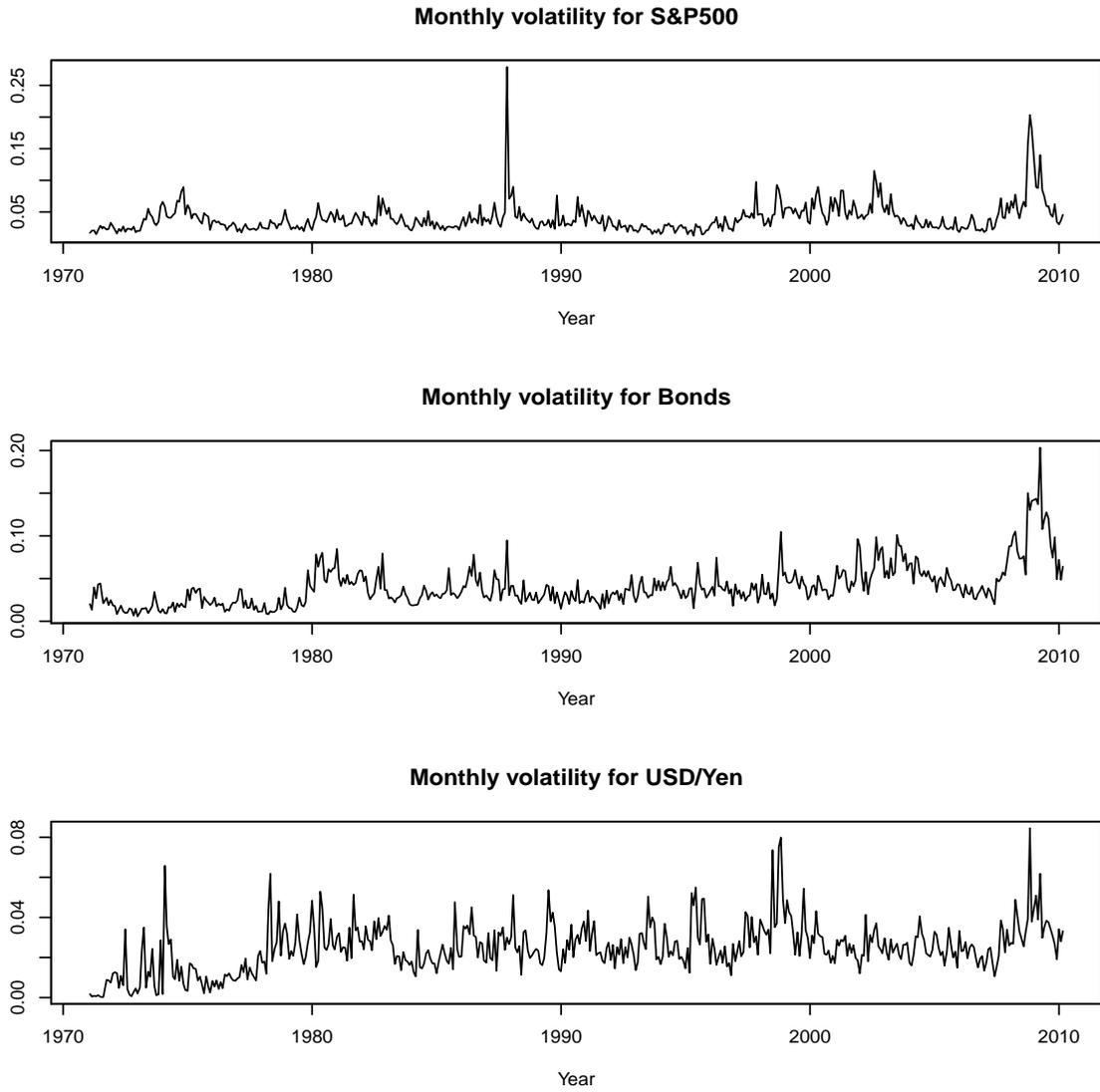


Figure 4: Monthly realized volatilities

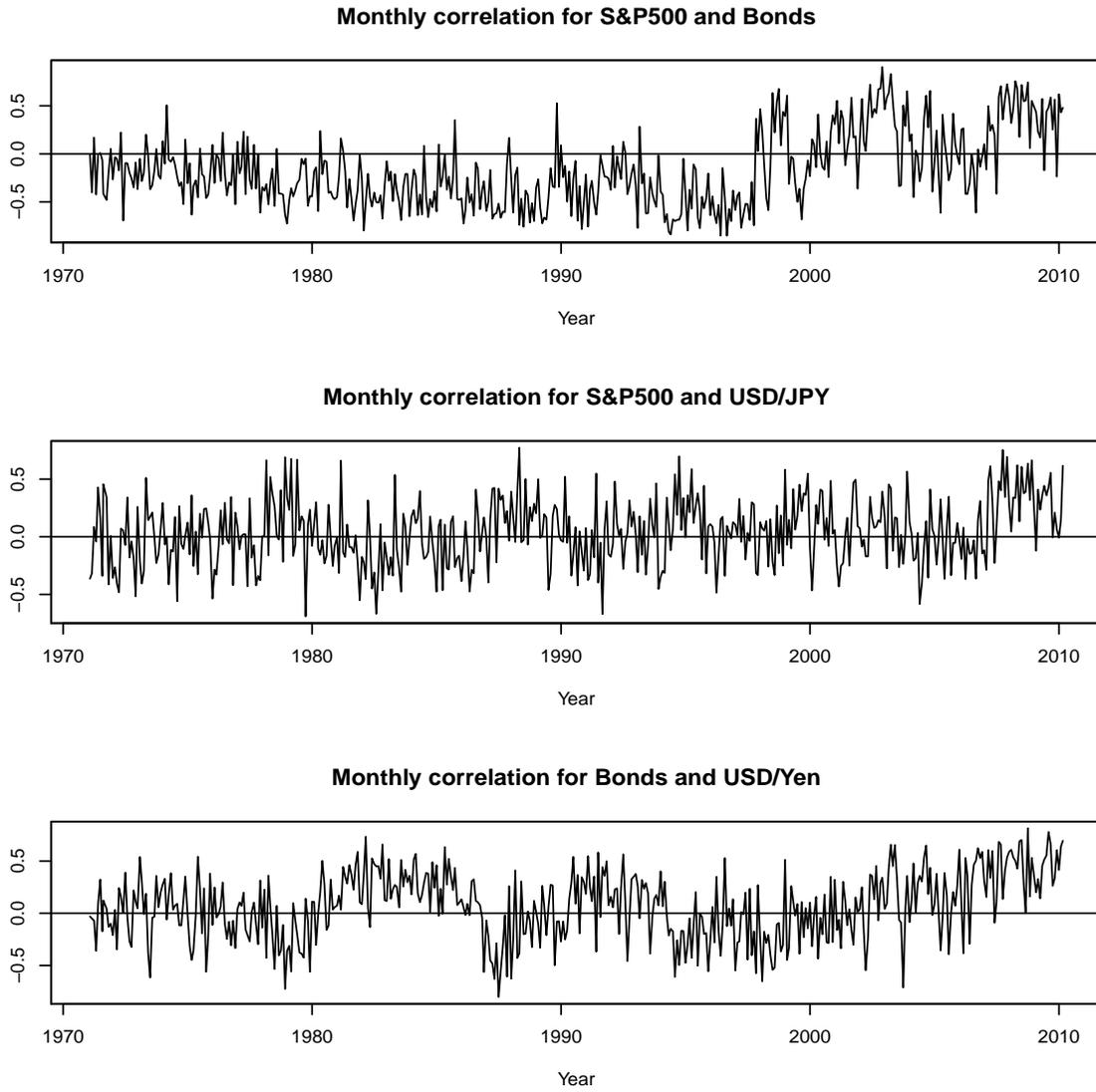


Figure 5: Monthly realized correlations