Article

Model Free Inference on Multivariate Time Series with Conditional Correlations

Dimitrios Thomakos [1], Johannes Klepsch [2] and Dimitris N. Politis [3],*

[1] Department of Economics, University of Peloponnese, 22100 Tripolis, Greece; thomakos@uop.gr or dimitrios.thomakos@gmail.com
[2] Department of Mathematical Statistics, Technische Universität München, 85748 Munich, Germany; j.klepsch@tum.de
[3] Department of Mathematics and Halicioglu Data Science Institute, University of California, San Diego, CA 92093, USA
* Correspondence: dpolitis@ucsd.edu

Abstract: New results on volatility modeling and forecasting are presented based on the NoVaS transformation approach. 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

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. [1] and Palandri [2], 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 (NOrmalizing and VAriance Stabilizing) transformation approach introduced by Politis [3–6] and significantly extended by Politis and Thomakos [7,8]. 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 [8], 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 advantages 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. Furthermore, we apply NoVaS to portfolio analysis. A referee has suggested that we consider non-linear types of correlations, asymmetric distributions for the returns and other types of portfolios. These are well suited for future research, especially the extensions for non-linear types of association, but this particular extension is beyond the scope of the current paper. The assumption of asymmetric distributions for the returns is innocuous in the current context: the main idea here is a transformation to normality no matter what the original return distribution might be.

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 [9], (the VEC model), Bollerslev [10] (the constant conditional correlation, CCC model), Bollerslev et al. and Engle et al. [11] (the BEKK) model), Pelletier and Silvennoinen et al. [12–15] (extensions of the CCC-GARCH). For an alternative Bayesian treatment of GARCH models see Vrontos et al. [16]

Engle [17] introduced the popular dynamic conditional correlation (DCC) model, which was extended and generalized by various authors: see, among others, Sheppard [18] and Hafner and Frances [19]. Independently Tse and Tsui [20] developed the VC-GARCH. For a review of the class of multivariate GARCH-type models see Bauwens et al. [21] and Silvennoinen and Teräsvirta [14] and for a review of volatility and correlation forecast evaluation see Patton and Sheppard [22]. A recent paper linking BEKK and DCC models is Caporin and McAleer [23].

Part of the literature treats the problem in a semi-parametric or non-parametric manner, such as in Long and Ullah [24] and Hafner et al. [25] Ledoit et al. [1] and Palandri [2] 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 [26] and Patton [27], and see Andersen et al. [28] for the realized correlation measures.

Finally, although time-series correlation modeling originated in the econometrics literature for financial applications, it caught on recently on the neuroscience literature after Lindquist et al. [29] presented DCC to the neuroimaging community. Some very recent and related papers on this strand of the literature, some of which propose robust methods, are those of Warnick et al. [30], John et al. [31], Lee and Kim [32,33], Behboundi and Farnoosh [34] and Faghiri et al. [35]

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 [8].

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' (Departures from the assumption of these) 'stylized facts' have been discussed in [7,8] of financial returns:

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

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 $\mathcal{F}_{t|t-p} \stackrel{\text{def}}{=} \sigma(X_t, X_{t-1}, \dots, X_{t-p})$

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

where α is a scalar control parameter, $\boldsymbol{a} \stackrel{\text{def}}{=} (a_0, a_1, \dots, a_p)^\top$ is a $(p+1) \times 1$ vector of control parameters and $G(\cdot; \alpha, \boldsymbol{a})$ is to be specified. 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:

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

with the implied restrictions (to maintain positivity for γ_t) that $\alpha \geq 0$, $a_j \geq 0$, $g(\cdot) > 0$ 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. The necessity and advantages of including the current value is elaborated upon by Politis [3–6,36–38] 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 [3,4,36]. The use of absolute returns, i.e., $g(z) = |z|$ has also been advocated for volatility modeling; see, e.g., Ghysels and Forsberg [39] 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 [7,8].

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

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

where $\phi(z)$ is the time-localized standard deviation that is defined relative to our choice of $g(z)$, for example $\phi(z) = \sqrt{z}$ if $g(z) = z^2$ or $\phi(z) = z$ if $g(z) = |z|$. The aim now is to choose the NoVaS parameters in such a way as to make W_t follow as closely as possible a chosen target distribution that is easier to work with. The natural choice for such a distribution is the normal—hence the 'normalization' in the NoVaS acronym; other choices (such as the uniform) are also possible in applications, although perhaps not as intuitive—see, e.g., Politis and Thomakos [7,8]. 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 λ; see, e.g., Politis [3,4,36]. 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 [36,37] for the univariate and for the multivariate case respectively. For example, solving for X_t in (3), and using the fact that γ_t depends on X_t, it follows that:

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

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

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

and

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

If one postulates that the U_t are i.i.d. according to some desired distribution, then (4) becomes a bona fide model. In particular, when $g(z) = z^2$, then (4) is tantamount to an $ARCH(p)$ model. For example, if the distribution of U_t is the one implied by (4) with W_t having a (truncated) normal distribution, then (4) is the model that is 'associated' with NoVaS. 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) are available on request.

Remark 2. *Equation (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:*

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

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 $(\alpha, \boldsymbol{a})$. The parameters α and $\boldsymbol{a}$ are constrained to be nonnegative to ensure the same for the variance. In addition, motivated by unbiasedness considerations, Politis [3,4,36] 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 [7,8] 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 [3,4,36]. An alternative allowing for greater weight to be placed on earlier lags is to consider exponential weights of the form:

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

where b is the rate; these are the 'exponential' NoVaS weights proposed in Politis [3,4,36]. 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 $\boldsymbol{\theta} \stackrel{\text{def}}{=} (p, b) \mapsto (\alpha, \boldsymbol{a})$, 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 [3,4,36], or complete distributional matching via any goodness-of-fit statistic like the Kolmogorov–Smirnov statistic, the quantile-quantile correlation coefficient (Shapiro–Wilk 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:

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

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 κ^* denotes the theoretical kurtosis coefficient of the target distribution. For the normal distribution $\kappa^* = 3$, whereas for the Uniform $\kappa^* = 1$.

The objective function for this case can be taken to be the absolute value, i.e., $D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} |\mathcal{K}_n(\boldsymbol{\theta})|$, and one would adjust the values of $\boldsymbol{\theta}$ so as to minimize $D_n(\boldsymbol{\theta})$. As noted by Politis [3,4,36] 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. Politis [3,4,36] 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–Wilk test for normality, when the target distribution is the standard normal. We now have that:

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

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

$$D_n(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \sup_x \sqrt{n} |\widehat{F}_W(x) - F(x)| \tag{11}$$

where $\widehat{F}_W(x)$ and $F(x)$ are two cumulative distribution functions; $\widehat{F}_W(x)$ is the empirical distribution of the W_t data, and $F(x)$ is the target distribution.

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:

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

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

Remark 3. *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 [40], Wang and Serfling [41]) or a multivariate normality statistic (e.g., Royston [42], Villasenor-Alva and Gonzalez-Estrada [43])? 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. Our approach is similar. The first step is to use the univariate NoVaS transformation to obtain 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 $Z_t(i, j) \equiv Z_t \overset{def}{=} W_{t,i}^* W_{t,j}^*$ denote their product.*

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

2. *$\rho_{t|t-s} \overset{def}{=} \mathsf{E}[Z_t|\mathcal{F}_{t-s}] = \mathsf{E}\left[W_{t,i}^* W_{t,j}^* | \mathcal{F}_{t-s}\right]$, for $s = 0, 1$, is the conditional correlation coefficient between the returns. For the case that $s = 0$ the expectation operator is formally redundant but see Equation (13) and the discussion around it.*

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 Equation (14) which can compactly represented as the following autoregressive model:

$$\rho_{t|t-s} \overset{def}{=} \lambda \rho_{t-1|t-1-s} + (1 - \lambda) Z_{t-s} \tag{13}$$

and can therefore be estimated by:

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

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:

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

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

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

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:

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

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:

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

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:

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

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

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

What is left to do is to estimate λ. In the following, we propose two 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.

3.1. Maximum Likelihood Estimation

We first summarize some interesting properties concerning the product of two studentized series 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 ρ, 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 $(-\infty, 0)$. 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 ρ, 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 4. *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 λ, as in Equation (13).

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

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

Remark 5. *Even though we do not need the explicit joint distribution of the studentized series, we still need to know the distribution of the product. Furthermore, using an MLE approach makes the multivariate NoVaS "semi-parametric". We therefore introduce a second method based on cross validation which does not require knowledge of the distribution, and hence keeps NoVaS model-free.*

3.2. Cross Validation (CV)

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

We still use Equation (13) but without knowing the density of Z_t. We only rely on the data. In the following, we define an objective function $Q(\lambda)$, which describes how well the λ is globally suited to describe the conditional correlation. $Q(\lambda)$ is then minimized with respect to λ, in order to find the optimal λ in Equation (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:

$$Q_1(\lambda) \stackrel{def}{=} \sum_{t=1}^{n} \left(\hat{\rho}_{t|t-1} - Z_t \right)^2 \tag{21}$$

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

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

and transform them individually with univariate NoVaS to get:

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

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:

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

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

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

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

$$Q_2(\lambda) \overset{def}{=} \sum_{t=T+1}^{n} \left(\hat{W}_{i,t} - W_{i,t} \right)^2 \tag{23}$$

In this context, T should be chosen large enough, in order to guarantee that the estimate of the conditional correlation in Equation (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 can add to the term in Equation (23) the symmetric term:

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

with

$$\hat{W}_{j,t} \overset{def}{=} \hat{\rho}_{t|t-1} W_{i,t}, \qquad \text{for } t = T+1, \ldots, n$$

to get to the objective function:

$$Q_3(\lambda) \overset{def}{=} \sum_{t=T+1}^{n} \left(\hat{W}_{i,t} - W_{i,t} \right)^2 + \sum_{t=T+1}^{n} \left(\hat{W}_{j,t} - W_{j,t} \right)^2 \tag{24}$$

- **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 Equation (22). One can then use the inverse transformation discussed in Equation (7), namely:

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

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

$$Q_4(\lambda) \overset{def}{=} \sum_{t=T+1}^{n} \left(\hat{X}_{i,t} - X_{i,t} \right)^2 \tag{26}$$

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

$$Q_5(\lambda) \overset{def}{=} \sum_{t=T+1}^{n} \left(\hat{X}_{i,t} - X_{i,t} \right)^2 + \sum_{t=T+1}^{n} \left(\hat{X}_{j,t} - X_{j,t} \right)^2 \tag{27}$$

- **CV 6** With the motivation of capturing the correct sign of the correlation, one can define an objective function that gets larger if the sign of the correlation at time point t is not predicted correctly. More formally, we define the loss function L:

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

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

$$Q_6(\lambda) \overset{def}{=} \sum_{t=T+1}^{n} L(t) \tag{28}$$

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

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

Using this estimate in Equation (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 6. *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:*

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

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

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

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

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

Since analytic computation of the above term is difficult, it is more sensible to use the iid structure of the $(W_{t,i}, W_{t,j})$. Assuming a normal target distribution of the NoVaS transformation, 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 Equation (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 Equation (14).

Interestingly, practical application show 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.

3.3. Going from the Bivariate Paradigm to a Fully Multivariate Setting

All the discussion in Section 3 has focused on estimating the conditional correlation coefficient of a pair (i, j) of studentized returns $W_{t,i}^*$ and $W_{t,j}^*$; the notation $\hat{\rho}_{t|t-s}$ and $\rho_{t|t-s}$ was used for the estimator and estimand respectively.

Of course, the realistic challenge is to construct an estimate of the conditional correlation coefficient matrix—denoted $R_{t|t-s}$—that is associated with a d–dimensional multivariate series, i.e., a vector series that at time t equals $(W_{t,1}^*, \ldots, W_{t,d}^*)'$.

Our proposal is to construct a matrix estimator $\hat{R}_{t|t-s}$ whose (i, j)-th element is given by the aforementioned conditional correlation coefficient of returns $W_{t,i}^*$ and $W_{t,j}^*$. Hence, for each pair (i, j), an optimal estimator of the (i, j) entry of $R_{t|t-s}$ that involves its own optimal smoothing λ parameter. For example, suppose that estimating the correlation between series 1 and 2 requires a small λ whereas for the correlation between series 3 and 4 it would be better to use a large λ. Using the same smoothing λ parameter for all the entries of the matrix could be grossly suboptimal.

Our procedure constructs a matrix estimator that is optimized entry-by-entry; the disadvantage is that the estimator $\hat{R}_{t|t-s}$ will not necessarily be a nonnegative definite matrix. Here, we will adopt the philosophy of Politis [2011], i.e., start with a matrix estimator that is optimized entry-by-entry, and adjust its eigenvalues to make it nonnegative definite.

Although the eigenvalue correction can be performed on the correlation matrix estimator $\hat{R}_{t|t-s}$, it is not guaranteed that the result will be a correlation matrix despite being nonnegative definite. Hence, it is more expedient to do the eigenvalue correction on an estimate of the covariance matrix instead, and then turn the corrected covariance matrix estimator into a new correlation matrix estimator.

The complete algorithm is outlined in what follows.

1. Construct $\hat{R}_{t|t-s}$ which is the estimator of the $d \times d$ correlation matrix $R_{t|t-s}$. The construction is carried out using entry-by-entry optimized estimators each of which is constructed according to one of the methods described in the previous subsections of Section 3.
2. Turn the correlation matrix estimator $\hat{R}_{t|t-s}$ into a covariance matrix estimator (denoted $\hat{C}_{t|t-s}$) by multiplying each correlation entry by the two respective volatilities, i.e., the square root of local variance estimates. In other words, the i, j entry of $\hat{C}_{t|t-s}$ consists of the i, j entry of $\hat{R}_{t|t-s}$ multiplied by the product of the volatility estimate of series i and the volatility estimate of series j.
3. Calculate the spectral decomposition $\hat{C}_{t|t-s} = TDT'$ where T is orthogonal and D diagonal containing the eigenvalues of $\hat{C}_{t|t-s}$.
4. Let D^* denote a version of D where all negative diagonal elements are replaced by zeros (or by a very small positive number—Politis [2011]—if the goal is strict positive definiteness).
5. Define the new covariance matrix estimator $\hat{C}_{t|t-s}^* = TD^*T'$ which is by construction nonnegative definite (or even strictly positive definite as mentioned above).
6. Turn $\hat{C}_{t|t-s}^*$ into a new correlation matrix estimator $\hat{R}_{t|t-s}^*$ by dividing each entry by the appropriate volatilities. In other words, the i, j entry of $\hat{R}_{t|t-s}^*$ consists of the i, j entry of $\hat{C}_{t|t-s}^*$ divided by the square root of the product of the i, i and the j, j entry of $\hat{C}_{t|t-s}^*$.

Note that Step 6 of the above algorithm implies that the square root of the i, i entry of $\hat{C}_{t|t-s}^*$ is to be treated as a new volatility estimate of series i for the purposes of multivariate modeling. However, this volatility estimate differs little from the standard one obtained from univariate modeling, e.g., the one used in Step 6 of the above algorithm. The reason is that the eigenvalue adjustment towards nonnegative definiteness is just a small-sample correction. The estimators $\hat{R}_{t|t-s}$ and $\hat{R}_{t|t-s}^*$ perform identically in large samples; see Politis [2011] for a discussion in a related context.

4. Using NoVaS in Applications

4.1. Model Selection

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. Details and empirical analysis is given in Politis and Thomakos (2008a,b). 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 (7×1) vector, say $\boldsymbol{D}_m(\nu, \tau)$, where $m =$ Method 1, ..., Method 6, MLE Method, $\nu =$ squared, absolute returns and $\tau =$ normal, uniform target distribution. Then, repeat the optimizations with respect to all seven methods for all combinations of (ν, τ). The "optimal" combination is then defined across all possible combinations (m, ν, τ) as follows:

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

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 Equation (13) can be based on the chosen length on the individual NoVaS transformations (i.e., on p from Equation (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).

4.2. Application to Portfolio Analysis

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 μ_1 and μ_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 β_t units of asset 1 and $(1 - \beta_t)$ units of asset 2. The portfolio return is therefore given by

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

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. β is indexed by $t-1$, because the choice of the composition of the portfolio has to 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

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

The goal of portfolio analysis in this context is to choose β_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 η, a general form of the utility function is:

$$U\left(\mathbb{E}\left[r_{p,t}|\mathcal{F}_{t-1}\right], \sigma_{p,t}^2\right) = \mathbb{E}\left[r_{p,t}|\mathcal{F}_{t-1}\right] - \eta\sigma_{p,t}^2 \tag{33}$$

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

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

$$\frac{\partial}{\partial \beta_{t-1}} U\left(\mathbb{E}\left[r_{p,t}|\mathcal{F}_{t-1}\right], \sigma_{p,t}^2\right) \stackrel{!}{=} 0 \Leftrightarrow \beta_{t-1} = \frac{0.5\eta^{-1}(\mu_1 - \mu_2) - (\sigma_{12,t} - \sigma_{2,t}^2)}{\sigma_{1,t}^2 + \sigma_{2,t}^2 - 2\sigma_{12,t}} \tag{34}$$

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

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

Recall that we need to impose that $0 \leq \beta_{t-1} \leq 1$ under the assumption of no short sales. 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 simulation study. We use two types of simulated data. First, we use a simple bivariate model as a data generating process (DGP), as in [22], 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 4.1 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 $\boldsymbol{R}_t \stackrel{\text{def}}{=} [X_t, Y_t]^\top$ denote the (2×1) vector of returns, the DGP-PS is given as follows:

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

where $\bar{\boldsymbol{\Sigma}}$ is a (2×2) matrix with unit diagonals and off-diagonals entries of 0.3. We let $t = 1, \ldots, 1000$. This is a scalar BEKK-GARCH model of order 1.

We use the model selection approach of the previous section. We compute $\boldsymbol{D}_m(\nu, \tau)$ of Equation (30) for all m and (ν, τ) 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 1.

We use the specified NoVaS transformation, where the kurtosis as in Equation (9) was used when fitting to a normal target, and the QQ-correlation as in Equation (11) was used when fitting to a uniform target. Furthermore we set $s = 0$ and used exponential weights.

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.

	Normal Target Squared Returns	Normal Target Absolute Returns	Uniform Target Squared Returns	Uniform Target Absolute Returns
MLE	2.09×10^{-2}	$\mathbf{2.08 \times 10^{-2}}$	3.18×10^{-2}	3.27×10^{-2}
CV 1	2.44×10^{-2}	2.44×10^{-2}	2.32×10^{-2}	2.37×10^{-2}
CV 2	2.24×10^{-2}	2.25×10^{-2}	2.29×10^{-2}	2.34×10^{-2}
CV 3	3.63×10^{-2}	3.89×10^{-2}	3.13×10^{-2}	2.58×10^{-2}
CV 4	2.22×10^{-2}	2.22×10^{-2}	2.34×10^{-2}	2.28×10^{-2}
CV 5	3.23×10^{-2}	3.39×10^{-2}	2.98×10^{-2}	5.78×10^{-2}
CV 6	8.10×10^{-2}	8.18×10^{-2}	8.80×10^{-2}	1.02×10^{-1}

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 might 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:

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

where

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

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

For both multivariate processes, we again compute the $\boldsymbol{D}_m(\nu, \tau)$ (as in Equation (30)). We repeat the computations 1000 times in order to get a robust idea of which method works best. Table 2 shows

the mean squared error between the real deterministic correlation and the estimates using the 28 different NoVaS methods.

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.

	Normal Target Squared Returns	Normal Target Absolute Returns	Uniform Target Squared Returns	Uniform Target Absolute Returns
		Sinusoidal correlation		
MLE	5.07×10^{-2}	5.01×10^{-2}	$\mathbf{2.38 \times 10^{-2}}$	2.42×10^{-2}
CV 1	3.47×10^{-2}	3.46×10^{-2}	2.63×10^{-2}	2.60×10^{-2}
CV 2	3.11×10^{-2}	3.10×10^{-2}	2.51×10^{-2}	2.45×10^{-2}
CV 3	1.07×10^{-1}	1.01×10^{-1}	9.36×10^{-2}	5.93×10^{-2}
CV 4	3.08×10^{-2}	3.07×10^{-2}	2.45×10^{-2}	2.51×10^{-2}
CV 5	1.38×10^{-1}	1.22×10^{-1}	1.04×10^{-1}	1.09×10^{-1}
CV 6	7.72×10^{-2}	7.71×10^{-2}	6.36×10^{-2}	6.40×10^{-2}
		Linear correlation		
MLE	6.00×10^{-2}	5.93×10^{-2}	4.88×10^{-2}	4.76×10^{-2}
CV 1	5.78×10^{-2}	5.73×10^{-2}	5.02×10^{-2}	4.89×10^{-2}
CV 2	5.47×10^{-2}	5.45×10^{-2}	4.89×10^{-2}	4.72×10^{-2}
CV 3	8.71×10^{-2}	8.47×10^{-2}	9.29×10^{-2}	7.12×10^{-2}
CV 4	5.44×10^{-2}	5.42×10^{-2}	$\mathbf{4.71 \times 10^{-2}}$	4.88×10^{-2}
CV 5	9.10×10^{-2}	8.90×10^{-2}	9.33×10^{-2}	1.07×10^{-1}
CV 6	9.88×10^{-2}	9.78×10^{-2}	9.18×10^{-2}	9.02×10^{-2}

As we can see in Table 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.

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 β_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, BEKK and CCC.

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 1 and 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).

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.

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 1 and 2. For the BEKK, we fit a BEKK-GARCH(1,1,1) defined as in Engle and Kroner (1995).

	DGP-PS		MVN-Sinusoidal		MVN-Linear	
	MSE	**Cor**	**MSE**	**Cor**	**MSE**	**Cor**
NoVaS	6.83×10^{-2}	6.50×10^{-1}	5.75×10^{-2}	7.16×10^{-1}	7.05×10^{-2}	7.16×10^{-1}
BEKK	1.91×10^{-2}	5.46×10^{-1}	3.28×10^{-2}	7.55×10^{-1}	4.94×10^{-2}	6.24×10^{-1}
DCC	1.39×10^{-2}	7.27×10^{-1}	2.07×10^{-2}	8.68×10^{-1}	4.26×10^{-2}	7.08×10^{-1}
CCC	3.44×10^{-2}	-	7.91×10^{-2}	-	8.16×10^{-2}	-

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 Equation (34). More precisely, the following algorithm is applied:

Algorithm 1: Computation of portfolio weights and associated returns.

1. Assume we observe $X_{t,1}$ and $X_{t,2}$ for $t = 1, \ldots, N$. Fix an window size T_0, for instance $T_0 = 1/2N$.
2. For every $T_0 \le k \le N - 1$, estimate the correlation $\rho_{12,k}$ based on $X_{t,1}$ and $X_{t,2}$, $t = (k + 1 - T_0), \ldots, k$, using the methods introduced above.
3. Use Equation (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. The simulations are repeated 100 times, and we show average results.

Table 4. 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.

	DGP-PS			MVN-Sinusoidal			MVN-Linear		
	Mean	**St.Dev.**	**Sh.R.**	**Mean**	**St.Dev.**	**Sh.R.**	**Mean**	**St.Dev.**	**Sh.R.**
NoVaS	-6.43×10^{-3}	2.78	1.79×10^{-4}	-5.79×10^{-2}	**3.03**	-1.34×10^{-2}	3.62×10^{-2}	3.12	8.77×10^{-3}
CCC	-3.74×10^{-3}	2.80	7.41×10^{-4}	-6.66×10^{-2}	3.05	-1.56×10^{-2}	4.47×10^{-2}	3.13	1.19×10^{-2}
BEKK	3.89×10^{-4}	**2.77**	2.28×10^{-3}	-6.40×10^{-2}	3.04	-1.46×10^{-2}	4.52×10^{-2}	3.12	1.19×10^{-2}
DCC	-3.85×10^{-3}	**2.77**	8.44×10^{-4}	-6.42×10^{-2}	**3.03**	-1.51×10^{-2}	4.52×10^{-2}	**3.11**	1.17×10^{-2}
LM	-2.01×10^{-3}	**2.77**	1.33×10^{-3}	-6.54×10^{-2}	3.06	-1.55×10^{-2}	4.28×10^{-2}	3.15	1.08×10^{-2}
Naive	1.91×10^{-3}	2.81	2.20×10^{-3}	-4.94×10^{-2}	3.33	$\mathbf{-1.11 \times 10^{-2}}$	3.70×10^{-2}	3.43	8.88×10^{-3}
Asset 1	-9.72×10^{-2}	3.48	-2.72×10^{-2}	-5.00×10^{-2}	3.36	-8.12×10^{-3}	$\mathbf{6.13 \times 10^{-2}}$	3.31	$\mathbf{1.55 \times 10^{-2}}$
Asset 2	$\mathbf{1.01 \times 10^{-1}}$	3.47	$\mathbf{3.21 \times 10^{-2}}$	$\mathbf{-4.87 \times 10^{-2}}$	4.42	-1.17×10^{-2}	1.28×10^{-2}	4.49	3.40×10^{-3}

Table 4 confirms previous results. NoVaS is on structured simulated data not able to outperform known methods based on multivariate GARCH. However, especially when looking at the standard deviation of the returns, NoVaS seems to be on par. In Figure 1 we show plots of the captured correlation by the different methods of exemplary simulated data.

However, 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.
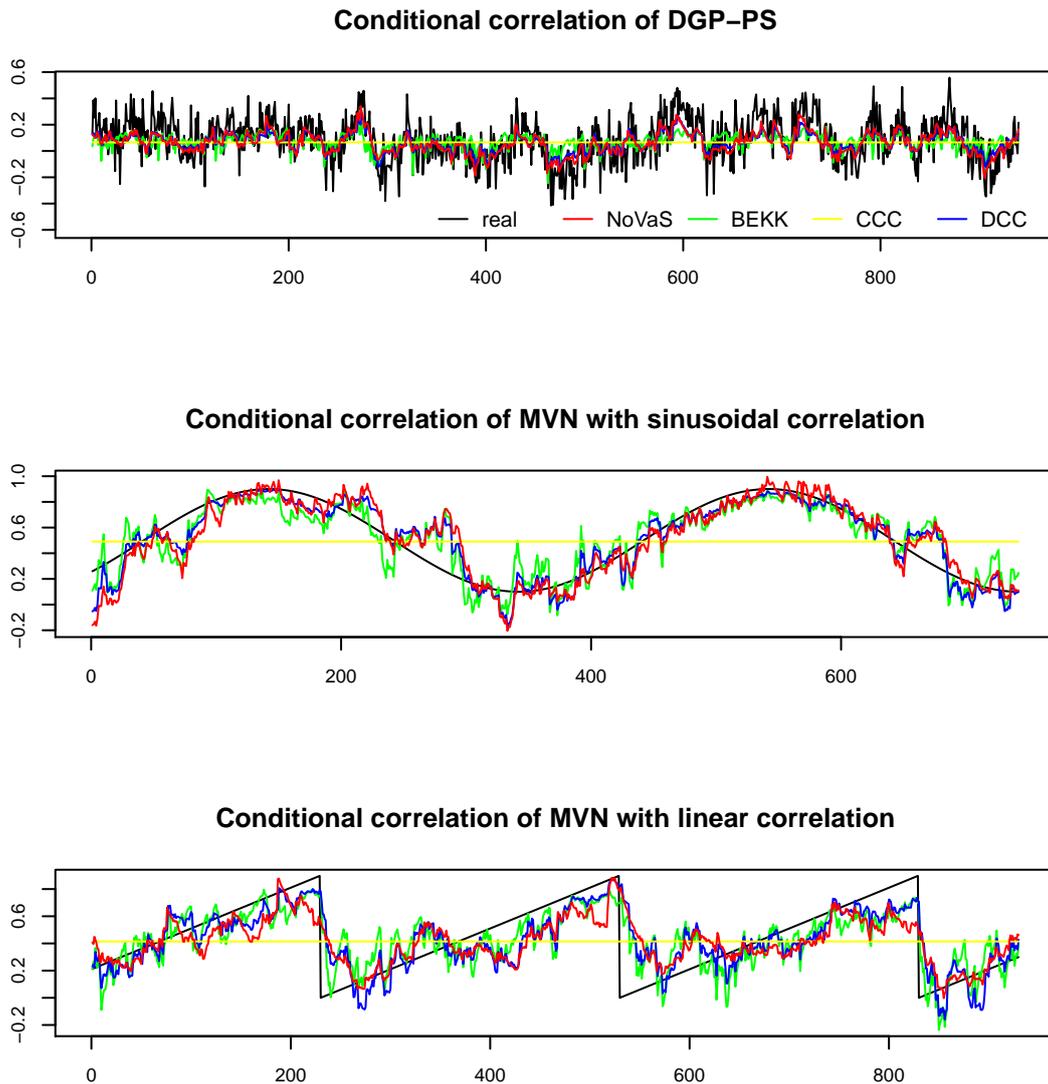
Conditional correlation of DGP–PS

Conditional correlation of MVN with sinusoidal correlation

Conditional correlation of MVN with linear correlation

Figure 1. Comparison of the different methods to capture correlation.

6. Empirical Illustration

In this section we offer a brief empirical illustration of the NoVaS-based correlation estimation using two data sets. First we consider data from the following three series: the S&P500, the 10-year bond (10-year Treasury constant maturity rate series) and the USD/Japanese Yen exchange rate. Then, to assess the performance on smaller samples, we focus on returns of the SPY and the TLT Exchange Traded Funds (ETFs). 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 $n_2 = 135$.

Figure 2 plots the monthly realized and fitted correlations and Tables 5 and 6 summarizes some descriptive statistics. From Tables 5 and 6 we can see that all the 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, the USD/JPY and the SPY series.

Table 5. Descriptive Statistics for monthly data, sample size is $n_1 = 469$ months from 01/1970 to 02/2010; SW is short for Shapiro–Wilk. The first 3 columns refer to returns, the middle 3 columns to realized volatilities, and the last 3 columns to correlations.

	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

Table 6. Descriptive Statistics for monthly data, sample size is $n_2 = 135$ months from 08/2002 to 02/2016, SW short for Shapiro–Wilk.

	SPY ret.	TLT ret.	SPY vol.	TLT vol.	SPY TLT corr.
Mean	0.006	0.006	0.003	0.002	−0.342
Median	0.011	0.001	0.001	0.001	−0.420
Std.Dev.	0.042	0.038	0.007	0.014	0.350
Skewness	−0.871	0.132	6.812	2.984	0.700
Kurtosis	2.130	2.019	57.286	14.002	−0.303
SW−test	0.960	0.967	0.362	0.735	0.944

6.1. Capturing the Correlation

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 7.

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. In the case of the shorter SPY and TLT index return series, using a normal target distribution and absolute returns, together with the MLE method provides the best result.

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 8 summarizes the results and Figure 2 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 four datasets the mean-squared error when using 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 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 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.

Table 7. Multivariate NoVaS model selection: MSE between $\rho_{t|t-1}$ and $\hat{\rho}_{t|t-1}$.

	Normal Target Squared Returns	Normal Target Absolute Returns	Uniform Target Squared Returns	Uniform Target Absolute Returns
		S&P500-Bonds		
MLE	1.32×10^{-1}	1.16×10^{-1}	9.13×10^{-2}	$\mathbf{9.05 \times 10^{-2}}$
CV 1	1.41×10^{-1}	9.48×10^{-2}	9.79×10^{-2}	1.01×10^{-1}
CV 2	1.59×10^{-1}	9.50×10^{-2}	9.77×10^{-2}	1.01×10^{-1}
CV 3	1.28×10^{-1}	1.23×10^{-1}	1.26×10^{-1}	1.26×10^{-1}
CV 4	1.55×10^{-1}	9.49×10^{-2}	9.78×10^{-2}	1.01×10^{-1}
CV 5	1.28×10^{-1}	1.23×10^{-1}	1.26×10^{-1}	1.26×10^{-1}
CV 6	1.81×10^{-1}	1.69×10^{-1}	1.69×10^{-1}	1.67×10^{-1}
		S&P500-USD/JPY		
MLE	7.53×10^{-2}	7.69×10^{-2}	9.68×10^{-2}	1.03×10^{-1}
CV 1	7.46×10^{-2}	7.57×10^{-2}	7.44×10^{-2}	7.69×10^{-2}
CV 2	7.39×10^{-2}	7.56×10^{-2}	7.55×10^{-2}	7.73×10^{-2}
CV 3	7.58×10^{-2}	7.86×10^{-2}	7.70×10^{-2}	7.93×10^{-2}
CV 4	$\mathbf{7.37 \times 10^{-2}}$	7.55×10^{-2}	7.54×10^{-2}	7.73×10^{-2}
CV 5	7.58×10^{-2}	7.86×10^{-2}	7.70×10^{-2}	7.93×10^{-2}
CV 6	2.34×10^{-1}	2.10×10^{-1}	3034×10^{-1}	3.51×10^{-1}
		Bonds-USD/JPY		
MLE	1.13×10^{-1}	1.08×10^{-1}	1.05×10^{-1}	1.11×10^{-1}
CV 1	9.95×10^{-2}	1.02×10^{-1}	1.02×10^{-1}	1.00×10^{-1}
CV 2	$\mathbf{9.92 \times 10^{-2}}$	1.00×10^{-1}	1.02×10^{-1}	9.96×10^{-2}
CV 3	1.09×10^{-1}	1.10×10^{-1}	1.06×10^{-1}	1.05×10^{-1}
CV 4	1.00×10^{-1}	1.02×10^{-1}	1.02×10^{-1}	9.97×10^{-2}
CV 5	1.09×10^{-1}	1.10×10^{-1}	1.06×10^{-1}	1.05×10^{-1}
CV 6	2.75×10^{-1}	3.15×10^{-1}	3.51×10^{-1}	3.40×10^{-1}
		SPY-TLT		
MLE	9.06×10^{-2}	$\mathbf{8.81 \times 10^{-2}}$	1.24×10^{-1}	1.28×10^{-1}
CV 1	9.06×10^{-2}	9.76×10^{-2}	9.20×10^{-2}	9.75×10^{-2}
CV 2	9.27×10^{-2}	9.34×10^{-2}	1.02×10^{-1}	1.10×10^{-1}
CV 3	9.27×10^{-2}	9.81×10^{-2}	9.66×10^{-2}	1.03×10^{-1}
CV 4	9.27×10^{-2}	9.81×10^{-2}	9.60×10^{-2}	1.02×10^{-1}
CV 5	9.27×10^{-2}	9.81×10^{-2}	9.66×10^{-2}	1.03×10^{-1}
CV 6	1.15×10^{-1}	1.28×10^{-1}	1.56×10^{-1}	2.67×10^{-1}

Table 8. 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 7.

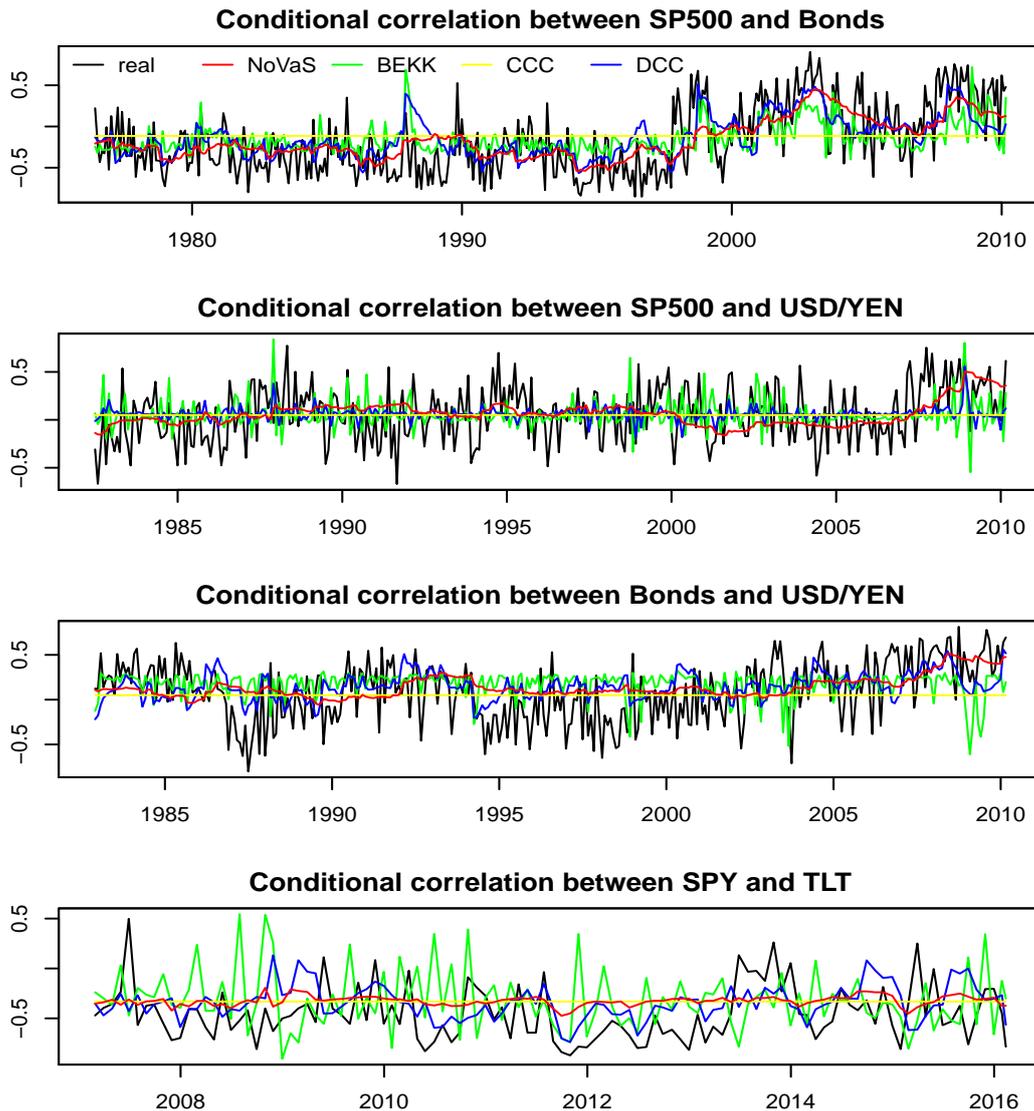|	S&P500-Bonds		S&P500-USD/YEN		Bonds-USD/YEN		SPY-TLT	
	MSE	**Corr**	**MSE**	**Corr**	**MSE**	**Corr**	**MSE**	**Corr**
NoVaS	$\mathbf{9.05 \times 10^{-2}}$	$\mathbf{6.61 \times 10^{-1}}$	7.39×10^{-2}	$\mathbf{3.06 \times 10^{-1}}$	$\mathbf{9.92 \times 10^{-2}}$	$\mathbf{3.64 \times 10^{-1}}$	$\mathbf{8.81 \times 10^{-2}}$	$\mathbf{2.81 \times 10^{-1}}$
DCC	1.05×10^{-1}	5.51×10^{-1}	8.34×10^{-2}	9.69×10^{-2}	1.03×10^{-1}	2.98×10^{-1}	9.82×10^{-2}	1.64×10^{-1}
BEKK	1.16×10^{-1}	4.68×10^{-1}	9.83×10^{-2}	7.34×10^{-2}	1.21×10^{-1}	1.05×10^{-1}	2.00×10^{-1}	1.21×10^{-1}
CCC	1.50×10^{-1}	-	8.22×10^{-2}	-	1.10×10^{-1}	-	8.82×10^{-2}	-

Figure 2. Monthly realized correlations and fitted values.

6.2. Application to Portfolio Analysis

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 calculate the optimal hedge ratio, as in Equation (34). We then compare mean, standard deviation and Sharpe ratio of the portfolio returns. We use Algorithm 1, and set $T_0 = 1/2N$ and $\eta = 3$. Table 9 summarizes the results.

The results of Table 9 show that NoVaS again performs best or on par when comparing to the other methods. In three out of four of our examples, the NoVaS portfolio achieves the highest Sharpe ratio. Only when constructing a portfolio out of Bonds and USD/YEN does the CCC method give a higher Sharpe Ratio, but NoVaS achieves the smallest standard deviation of the returns. In further analysis, not reported here, we compared results based on different choices of the risk aversion coefficient η in Equation (34), and of T_0. The performance of NoVaS somehow did not vary.

Table 9. Mean (M), standard deviation (SD) and Sharpe ratio (SR) of portfolio returns, where the portfolio weights are chosen as in Equation (34), with the correlation estimated with the methods in the left column. The NoVaS transformation used is the one that performed best in Table 7. Naive assumes an equally weighted portfolio. Linear model estimates the conditional correlation based on linear regression.

	S&P500-Bonds			S&P500-USD/YEN		
	M	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	3.91×10^{-2}	1.28×10^{-1}	$\mathbf{3.07 \times 10^{-1}}$	$\mathbf{6.66 \times 10^{-2}}$	1.10×10^{-1}	$\mathbf{6.04 \times 10^{-1}}$
DCC	3.82×10^{-2}	$\mathbf{1.27 \times 10^{-1}}$	3.00×10^{-1}	6.60×10^{-2}	1.12×10^{-1}	5.88×10^{-1}
BEKK	3.48×10^{-2}	1.29×10^{-1}	2.70×10^{-1}	6.45×10^{-2}	1.12×10^{-1}	5.75×10^{-1}
CCC	3.82×10^{-2}	$\mathbf{1.27 \times 10^{-1}}$	3.01×10^{-1}	5.70×10^{-2}	1.10×10^{-1}	5.02×10^{-1}
LM	3.66×10^{-2}	$\mathbf{1.27 \times 10^{-1}}$	2.88×10^{-1}	6.44×10^{-2}	1.12×10^{-1}	5.74×10^{-1}
Naive	7.87×10^{-5}	1.29×10^{-1}	6.09×10^{-4}	8.15×10^{-3}	$\mathbf{9.84 \times 10^{-2}}$	8.29×10^{-2}
Asset 1	$\mathbf{4.48 \times 10^{-2}}$	1.55×10^{-1}	3.12×10^{-1}	4.84×10^{-2}	1.55×10^{-1}	3.12×10^{-1}
Asset 2	-4.83×10^{-2}	1.99×10^{-1}	-2.43×10^{-1}	-3.21×10^{-2}	1.12×10^{-1}	-2.88×10^{-1}
	Bonds-USD/YEN			**SPY-TLT**		
	M	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	-4.78×10^{-2}	$\mathbf{1.02 \times 10^{-1}}$	-4.67×10^{-1}	$\mathbf{1.27 \times 10^{-1}}$	6.08×10^{-2}	**2.10**
DCC	-4.82×10^{-2}	$\mathbf{1.02 \times 10^{-1}}$	-4.71×10^{-1}	$\mathbf{1.27 \times 10^{-1}}$	$\mathbf{6.04 \times 10^{-2}}$	**2.10**
BEKK	-4.71×10^{-2}	1.03×10^{-1}	-4.56×10^{-1}	1.26×10^{-1}	6.07×10^{-2}	2.08
CCC	$\mathbf{-2.32 \times 10^{-2}}$	1.20×10^{-1}	$\mathbf{-1.93 \times 10^{-1}}$	1.26×10^{-1}	6.09×10^{-2}	2.07
LM	-4.86×10^{-2}	$\mathbf{1.02 \times 10^{-1}}$	-4.76×10^{-1}	$\mathbf{1.27 \times 10^{-1}}$	6.09×10^{-2}	2.08
Naive	-4.02×10^{-2}	1.23×10^{-1}	-3.26×10^{-1}	9.63×10^{-2}	6.23×10^{-2}	1.55
Asset 1	-4.83×10^{-2}	1.99×10^{-1}	-2.43×10^{-1}	1.00×10^{-1}	1.24×10^{-1}	8.10×10^{-1}
Asset 2	-3.21×10^{-2}	1.12×10^{-1}	-2.88×10^{-1}	9.23×10^{-2}	1.37×10^{-1}	6.75×10^{-1}

6.3. Do NoVaS Work Well in Larger Dimensions?

There is but one remaining question that we need to address: is NoVaS the way to go when we increase the dimensionality of a problem? The answer to this question can be a paper of its own so here we provide a short-cut illustrating both the efficacy of the purely multivariate algorithm presented in Section 3 but also the efficacy in a practical application. To this end we expand our universe of assets to include more than a pair in two sets of experiments: in the first experiment we add just one more asset, to have a representation of equities, fixed income/bonds and cash; in the second experiment we take several assets that comprise the industry components of S&P500. For the first experiment we add SHY as our cash-proxy ETF and for the second experiment we take the 9 ETFs that make industry-groups for the S&P500: XLB, XLE, XLF, XLI, XLK, XLP, XLU, XLV, XLY.

We thus have a 3-dimensional, a 4-dimensional, and a 9-dimensional example of application of the purely multivariate approach. To better understand the way NoVaS operates on these examples we consider two different sample splits for our estimation, one-third and two-thirds of our total available observations. Furthermore, to examine whether the frequency of portfolio rebalancing matters we consider a weekly and a monthly rebalancing. All our results are collected in Tables 10–13 that follow.

For the weekly rebalancing we find that across the two sample splits NoVaS has the highest or second highest portfolio return in 7 out of 12 cases examined (63%), but only 4 out of 12 (33%) of the highest or second highest Sharpe ratios. For the monthly rebalancing we find that across the two sample splits NoVaS has the highest or second highest portfolio return in 10 out of 12 cases examined (83%), while it has only 3 out of 12 highest or second highest Sharpe ratios (25%). Although the results presented here are only illustrative, they are suggestive of a further avenue of future research for the application of NoVaS in a purely multivariate setting: (a) clearly the application of NoVaS is a return-generator insofar as portfolio performance is concerned, at the (expected) expense of higher portfolio risk; (b) lower frequency rebalancing improves performance dramatically for returns but leaves Sharpe ratios relatively unchanged; (c) NoVaS is highly competitive in larger dimensions

because it involves simpler/less optimization and is robust across sample sizes—a relatively large sample is not a theoretical prerequisite for achieving good estimation performance.

Table 10. Mean (M), standard deviation (SD) and Sharpe ratio (SR) of portfolio returns with monthly rebalancing and with weights computed on 2/3 of available observations.

	SHY-SPY-TLT			OIH-SHY-SPY			OIH-SHY-TLT		
	M	**SD**	**SR**	**M**	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	3.50×10^{-2}	1.59×10^{-2}	**2.19**	$\mathbf{3.91 \times 10^{-2}}$	1.73×10^{-2}	**2.26**	3.60×10^{-2}	2.42×10^{-2}	1.49
CCC	2.74×10^{-2}	$\mathbf{1.30 \times 10^{-2}}$	2.11	1.29×10^{-2}	9.97×10^{-2}	1.29	$\mathbf{1.09 \times 10^{-1}}$	8.73×10^{-2}	1.25
BEKK	2.61×10^{-2}	1.34×10^{-2}	1.95	2.51×10^{-2}	1.38×10^{-2}	1.82	2.51×10^{-2}	1.38×10^{-2}	1.82
DCC	2.75×10^{-2}	$\mathbf{1.30 \times 10^{-2}}$	2.11	2.78×10^{-2}	$\mathbf{1.29 \times 10^{-2}}$	2.15	2.69×10^{-2}	$\mathbf{1.36 \times 10^{-2}}$	**1.98**
Naive	$\mathbf{5.87 \times 10^{-2}}$	4.47×10^{-2}	1.31	2.52×10^{-2}	1.11×10^{-1}	0.23	6.00×10^{-5}	7.69×10^{-2}	8.20×10^{-4}
	OIH-SPY-TLT			OIH-SHY-SPY-TLT			9 ETFs		
	M	**SD**	**SR**	**M**	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	$\mathbf{1.14 \times 10^{-1}}$	8.25×10^{-2}	1.38	3.76×10^{-2}	2.00×10^{-2}	1.88	5.17×10^{-2}	1.39×10^{-1}	0.37
CCC	8.84×10^{-2}	$\mathbf{6.64 \times 10^{-2}}$	1.33	$\mathbf{1.12 \times 10^{-1}}$	7.79×10^{-2}	1.44	3.46×10^{-2}	1.50×10^{-1}	0.23
BEKK	8.83×10^{-2}	6.79×10^{-2}	1.30	2.83×10^{-2}	1.34×10^{-2}	2.11	-	-	-
DCC	9.35×10^{-2}	6.70×10^{-2}	**1.40**	2.81×10^{-2}	$\mathbf{1.31 \times 10^{-2}}$	2.15	9.91×10^{-3}	1.28×10^{-1}	0.77
Naive	3.94×10^{-2}	9.90×10^{-2}	0.40	3.01×10^{-2}	7.43×10^{-2}	0.42	$\mathbf{1.12 \times 10^{-1}}$	1.14×10^{-1}	**0.98**

Table 11. Mean (M), standard deviation (SD) and Sharpe ratio (SR) of portfolio returns with monthly rebalancing and with weights computed on 1/3 of available observations.

	SHY-SPY-TLT			OIH-SHY-SPY			OIH-SHY-TLT		
	M	**SD**	**SR**	**M**	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	4.05×10^{-2}	2.70×10^{-2}	1.50	3.80×10^{-2}	2.80×10^{-2}	1.36	5.77×10^{-2}	4.76×10^{-2}	1.21
CCC	$\mathbf{9.18 \times 10^{-2}}$	5.88×10^{-2}	1.56	$\mathbf{9.88 \times 10^{-2}}$	1.07×10^{-1}	0.93	$\mathbf{6.42 \times 10^{-2}}$	5.61×10^{-2}	1.14
BEKK	3.32×10^{-2}	1.52×10^{-2}	2.18	3.29×10^{-2}	1.51×10^{-2}	2.18	3.18×10^{-2}	1.57×10^{-2}	2.03
DCC	3.32×10^{-2}	$\mathbf{1.48 \times 10^{-2}}$	**2.24**	3.45×10^{-2}	$\mathbf{1.47 \times 10^{-2}}$	**2.35**	3.13×10^{-2}	$\mathbf{1.53 \times 10^{-2}}$	**2.04**
Naive	5.13×10^{-2}	5.97×10^{-2}	0.86	1.36×10^{-2}	1.55×10^{-1}	8.76×10^{-2}	1.92×10^{-2}	9.87×10^{-2}	0.19
	OIH-SPY-TLT			OIH-SHY-SPY-TLT			9 ETFs		
	M	**SD**	**SR**	**M**	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	7.49×10^{-2}	1.06×10^{-1}	0.71	4.63×10^{-2}	3.99×10^{-2}	1.16	3.68×10^{-2}	1.40×10^{-1}	0.26
CCC	$\mathbf{9.97 \times 10^{-2}}$	1.31×10^{-2}	0.76	$\mathbf{1.12 \times 10^{-1}}$	8.78×10^{-1}	1.27	3.25×10^{-2}	1.63×10^{-1}	0.20
BEKK	8.25×10^{-2}	$\mathbf{8.75 \times 10^{-2}}$	**0.94**	3.27×10^{-2}	1.61×10^{-2}	2.04	-	-	-
DCC	7.62×10^{-2}	9.25×10^{-2}	0.82	3.39×10^{-2}	$\mathbf{1.46 \times 10^{-2}}$	**2.32**	3.62×10^{-2}	1.28×10^{-1}	0.28
Naive	3.24×10^{-2}	1.40×10^{-1}	0.23	2.91×10^{-2}	1.04×10^{-1}	0.28	$\mathbf{7.21 \times 10^{-2}}$	1.49×10^{-1}	**0.48**

Table 12. Mean (M), standard deviation (SD) and Sharpe ratio (SR) of portfolio returns with weekly rebalancing and with weights computed on 2/3 of available observations.

	SHY-SPY-TLT			OIH-SHY-SPY			OIH-SHY-TLT		
	M	**SD**	**SR**	**M**	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	2.50×10^{-2}	2.20×10^{-2}	1.14	1.42×10^{-2}	3.40×10^{-2}	0.42	$\mathbf{2.33 \times 10^{-2}}$	4.28×10^{-2}	0.54
CCC	$\mathbf{6.19 \times 10^{-2}}$	3.93×10^{-2}	**1.57**	$\mathbf{5.15 \times 10^{-2}}$	1.26×10^{-1}	0.41	5.71×10^{-3}	6.45×10^{-2}	8.84×10^{-2}
BEKK	5.91×10^{-3}	$\mathbf{5.05 \times 10^{-3}}$	1.17	1.03×10^{-3}	6.55×10^{-3}	0.20	2.53×10^{-3}	4.31×10^{-3}	0.59
DCC	8.03×10^{-3}	5.19×10^{-3}	1.55	2.65×10^{-3}	$\mathbf{4.81 \times 10^{-3}}$	**0.55**	3.92×10^{-3}	$\mathbf{4.23 \times 10^{-3}}$	**0.93**
Naive	2.33×10^{-2}	5.43×10^{-2}	0.43	1.45×10^{-2}	1.38×10^{-1}	0.11	-5.93×10^{-3}	9.90×10^{-2}	-6.00×10^{-2}
	OIH-SPY-TLT			OIH-SHY-SPY-TLT			9 ETFs		
	M	**SD**	**SR**	**M**	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	1.00×10^{-1}	7.34×10^{-2}	1.36	2.78×10^{-2}	3.20×10^{-2}	**0.87**	1.57×10^{-1}	$\mathbf{9.08 \times 10^{-2}}$	**1.73**
CCC	8.65×10^{-2}	8.12×10^{-2}	1.07	$\mathbf{4.43 \times 10^{-2}}$	7.37×10^{-2}	0.60	$\mathbf{1.79 \times 10^{-1}}$	1.12×10^{-1}	1.59
BEKK	$2.098.74 \times 10^{-2}$	$\mathbf{6.12 \times 10^{-2}}$	1.43	3.06×10^{-3}	$\mathbf{5.07 \times 10^{-3}}$	0.60	-	-	-
DCC	$\mathbf{1.01 \times 10^{-1}}$	6.23×10^{-2}	**1.63**	2.50×10^{-3}	5.17×10^{-3}	0.48	1.35×10^{-1}	9.20×10^{-2}	1.47
Naive	1.30×10^{-2}	1.29×10^{-1}	0.10	1.12×10^{-2}	9.60×10^{-2}	0.12	4.70×10^{-2}	1.30×10^{-1}	0.36

Table 13. Mean (M), standard deviation (SD) and Sharpe ratio (SR) of portfolio returns with weekly rebalancing and with weights computed on 1/3 of available observations.

	SHY-SPY-TLT			OIH-SHY-SPY			OIH-SHY-TLT		
	M	**SD**	**SR**	**M**	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	2.52×10^{-2}	1.84×10^{-2}	1.37	1.75×10^{-2}	1.63×10^{-2}	1.07	1.05×10^{-3}	2.57×10^{-2}	4.07×10^{-2}
CCC	$\mathbf{8.61 \times 10^{-2}}$	4.06×10^{-2}	**2.12**	$\mathbf{4.17 \times 10^{-2}}$	1.17×10^{-1}	0.36	$\mathbf{1.29 \times 10^{-2}}$	6.70×10^{-2}	0.19
BEKK	9.19×10^{-3}	5.99×10^{-3}	1.54	6.02×10^{-3}	5.63×10^{-3}	1.07	5.91×10^{-3}	6.01×10^{-3}	**0.98**
DCC	8.79×10^{-3}	$\mathbf{5.46 \times 10^{-3}}$	1.61	7.69×10^{-3}	$\mathbf{5.50 \times 10^{-3}}$	**1.40**	4.70×10^{-3}	$\mathbf{5.67 \times 10^{-3}}$	0.83
Naive	6.20×10^{-2}	5.00×10^{-2}	1.24	5.98×10^{-3}	1.22×10^{-1}	4.91×10^{-2}	-4.27×10^{-3}	8.51×10^{-2}	-5.02×10^{-2}

	OIH-SPY-TLT			OIH-SHY-SPY-TLT			9 ETFs		
	M	**SD**	**SR**	**M**	**SD**	**SR**	**M**	**SD**	**SR**
NoVaS	7.09×10^{-2}	7.69×10^{-2}	9.22×10^{-2}	1.14×10^{-2}	2.30×10^{-2}	0.50	0.15	0.12	1.24
CCC	9.90×10^{-2}	9.96×10^{-2}	0.99	5.22×10^{-2}	7.28×10^{-2}	0.72	**0.17**	0.13	**1.35**
BEKK	$\mathbf{1.13 \times 10^{-1}}$	6.25×10^{-2}	**1.81**	9.24×10^{-3}	$\mathbf{6.13 \times 10^{-3}}$	**1.51**	-	-	-
DCC	1.10×10^{-1}	$\mathbf{6.14 \times 10^{-2}}$	1.80	$\mathbf{5.93 \times 10^{-2}}$	5.29×10^{-2}	1.12	0.14	**0.11**	1.26
Naive	2.85×10^{-2}	1.13×10^{-1}	0.25	2.31×10^{-2}	8.45×10^{-2}	0.27	9.02×10^{-2}	1.23×10^{-1}	0.73

Of course this summary of our multivariate results does not exhaust the problem of NoVaS multivariate applications. It rather leaves open the question as to how further improvements can be made, but as our results show these improvements are necessarily specific to what we are trying to achieve. Further research can show us where multivariate NoVaS fares better, in estimation of correlations or in problems like portfolio allocation.

7. Concluding Remarks

In this paper we extend the univariate NoVaS methodology for volatility modeling and forecasting, put forth by Politis (2003a,b, 2007, 2015) 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; (c) a further and more elaborate exploration of how NoVaS can be applied in high-dimension problems, and what the relevant implications may be.

References

1. Ledoit, O.; Santa-Clara, P.; Wolf, M. Flexible Multivariate GARCH Modeling with an Application to International Stock Markets. *Rev. Econ. Stat.* **2003**, *8*, 735–747. [CrossRef]
2. Palandri, A. Sequential Conditional Correlations: Inference and Evaluation. *J. Econ.* **2009**, *153*, 122–132. [CrossRef]
3. Politis, D.N. *Model-Free Volatility Prediction*; Department of Economics, UCSD: San Diego, CA, USA, 2003.

4. Politis, D.N. A Normalizing and Variance-Stabilizing Transformation for Financial Time Series. In *Recent Advances and Trends in Nonparametric Statistics*; Akritas, M.G., Politis, D.N., Eds.; Elsevier: Amsterdam, The Netherlands, 2003; pp. 335–347.

5. Politis, D.N. Model-free vs. model-based volatility prediction. *J. Financ. Econ.* **2007**, *5*, 358–389.

6. Politis, D.N. *Model-Free Prediction and Regression: A Transformation-Based Approach to Inference*; Springer: New York, NY, USA, 2015.

7. Politis, D.; Thomakos, D. Financial Time Series and Volatility Prediction using NoVaS Transformations. In *Forecasting in the Presence of Parameter Uncertainty and Structural Breaks*; Rapach, D.E., Wohar, M.E., Eds.; Emerald Group Publishing: Bingley, UK, 2008.

8. Politis, D.; Thomakos, D. *NoVaS Transformations: Flexible Inference for Volatility Forecasting*; Working Paper; Department of Economics, UCSD: San Diego, CA, USA, 2008. Available online: http://escholarship.org/uc/item/982208kx (accessed on 28 July 2020).

9. Bollerslev, T.; Engle, R.; Wooldridge, J. A Capital Asset Pricing Model with Time Varying Covariances. *J. Political Econ.* **1988**, *96*, 116–131. [CrossRef]

10. Bollerslev, T. Modeling the coherence in short-run nominal exchange rates: A multivariate Generalized ARCH Model. *Rev. Econ. Stat.* **1990**, *72*, 498–505. [CrossRef]

11. Engle, R.; Kroner, F. Multivariate Simultaneous Generalized ARCH. *Econ. Theory* **1995**, *11*, 122–150. [CrossRef]

12. Pelletier, D. Regime Switching for Dynamic Correlations. *J. Econ.* **2006**, *131*, 445–473. [CrossRef]

13. Silvennoinen, A.; Teräsvirta, T. Modeling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model. *J. Financ. Econ.* **2009**, *7*, 373–411. [CrossRef]

14. Silvennoinen, A.; Teräsvirta, T. Mutlivariate GARCH Models. In *Handbook of Financial Time Series*; Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2009.

15. Silvennoinen, A.; Teräsvirta, T. Modeling conditional correlations of asset returns: A smooth transition approach. *Econ. Rev.* **2015**, *34*, 174–197. [CrossRef]

16. Vrontos, I.D.; Dellaportas, P.; Politis, D.N. Full Bayesian Inference for GARCH and EGARCH Models. *J. Bus. Econ. Stat.* **2000**, *18*, 187–198.

17. Engle, R. Dynamic Conditional Correlation: A simple class of multivariate GARCH models. *J. Bus. Econ. Stat.* **2002**, *17*, 339–350. [CrossRef]

18. Hafner, C.; Franses, P. A generalized dynamic conditional correlation model: Simulation and application to many assets. *Econ. Rev.* **2009**, *28*, 612–631. [CrossRef]

19. Sheppard, K. *Understanding the Dynamics of Equity Covariance*; Unpublished Paper; UCSD: San Diego, CA, USA, 2002.

20. Tse, Y.; Tsui, A. A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations. *J. Bus. Econ. Stat.* **2002**, *20*, 351–362. [CrossRef]

21. Bauwens, L.; Laurent, S.; Rombouts, J.V.K. Multivariate GARCH Models: A Survey. *J. Appl. Econ.* **2006**, *21*, 79–109. [CrossRef]

22. Patton, A.; Sheppard, K. Evaluating Volatility and Correlation Forecasts. In *Handbook of Financial Time Series*; Andersen, T.G., Davis, R.A., Kreiss, J.-P., Mikosch, T., Eds.; Springer: New York, NY, USA, 2009.

23. Caporin, M.; McAleer, M. Do We Really Need Both BEKK and DCC? A Tale of Two Multivariate GARCH Models. *J. Econ. Surv.* **2012**, *26*, 736–751. [CrossRef]

24. Long, X.; Su, L.; Ullah, A. Estimation and Forecasting of Dynamic Conditional Covariance: A Semiparametric Multivariate Model. *J. Bus. Econ. Stat.* **2011**, *29*, 109–125. [CrossRef]

25. Hafner, C.; Dijk, D.; Franses, P. Semi-parametric Modelling of Correlation Dynamics. *Adv. Econ.* **2006**, *20*, 59–103.

26. Jondeau, E.; Rockinger, M. The Economic Value of Distributional Timing. *Swiss Financ. Inst. Res. Pap. Ser.* **2006**, *35*. [CrossRef]

27. Patton, A. Modeling Asymmetric Exchange Rate Dependence. *Int. Econ. Rev.* **2006**, *47*, 527–556. [CrossRef]

28. Andersen, T.G.; Bollerslev, T.; Christoffersen, P.F.; Diebold, F.X. Volatility and Correlation Forecasting. In *Handbook of Economic Forecasting*; Elliott, G., Granger, C.W.J., Timmermann, A., Eds.; Elsevier: Amsterdam, The Netherlands, 2006; pp. 778–878.

29. Lindquist, M.A.; Xu, Y.; Nebel, M.B.; Caffo, B.S. Evaluating Dynamic Bivariate Correlations in Resting-State fMRI: A Comparison Study and a New Aapproach. *NeuroImage* **2014**, *101*, 531–546. [CrossRef]

30. Warnick, R.; Guindani, M.; Erhardt, E.; Allen, E.; Calhoun, V.; Vannucci, M. A Bayesian Approach for Estimating Dynamic Functional Network Connectivity in fMRI Data. *J. Am. Stat. Assoc.* **2018**, *113*, 134–151. [CrossRef] [PubMed]

31. John, M.; Wu, Y.; Narayan, M.; John, A.; Ikuta, T.; Ferbinteanu, J. Estimation of Dynamic Bivariate Correlation using a Weighted Graph Algorithm. *Entropy* **2020**, *22*, 617. [CrossRef]

32. Lee, N.; Kim, J.M. Dynamic Functional Connectivity Analysis of Functional MRI based on copula Time-Varying Correlation. *J. Neurosci. Methods* **2019**, *323*, 32–47. [CrossRef]

33. Lee, N.; Kim, J.M. Dynamic Functional Connectivity Analysis based on Time-Varying Partial Correlation with a copula-DCC-GARCH mode. *Neurosci. Res.* **2020**. [CrossRef]

34. Behboudi, M.; Farnoosh, R. Modified Models and Simulations for Estimating Dynamic Functional Connectivity in Resting State Functional Magnetic Resonance Imaging. *Stat. Med.* **2020**, *39*, 1781–1800. [CrossRef]

35. Faghiri, A.; Iraji, A.; Damaraju, E.; Belger, A.; Ford, J.; Mathalon, D.; Mcewen, S.; Calhoun, V.D. Weighted Average of Shared Trajectory: A New Estimator for Dynamic Functional Connectivity Efficiently Estimates both Rapid and Slow Changes Over Time. *J. Neurosci. Methods* **2020**, *334*. [CrossRef]

36. Politis, D.N. A heavy-tailed distribution for ARCH residuals with application to volatility prediction. *Ann. Econ. Financ.* **2004**, *5*, 283–298.

37. Politis, D.N. A multivariate heavy-tailed distribution for ARCH/GARCH residuals. *Adv. Econ.* **2006**, *20*, 105–124.

38. Politis, D.N. Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econ. Theory* **2011**, *27*, 703–744. [CrossRef]

39. Ghysels, E.; Forsberg, L. Why Do Absolute Returns Predict Volatility So Well? *J. Financ. Econ.* **2007**, *5*, 31–67.

40. Mardia, K. Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika* **1970**, *57*, 519–530. [CrossRef]

41. Wang, J.; Serfling, R. Nonparametric Multivariate Kurtosis and Tailweight Measures. *J. Nonparamet. Stat.* **2005**, *17*, 441–456. [CrossRef]

42. Royston, J.P. Some Techniques for Assessing Multivariate Normality based on the Shapiro-Wilk W. *Appl. Stat.* **1983**, *32*, 121–133. [CrossRef]

43. Villasenor-Alva, J.; Gonzalez-Estrada, E. A Generalization of Shapiro-Wilks Test for Multivariate Normality. *Commun. Stat. Theory Methods* **2009**, *38*, 1870–1883. [CrossRef]