Conditional Correlation in Financial Returns

A Model-Free Approach

Masterarbeit

von

Johannes Klepsch

Themensteller: Prof. Claudia Klüppelberg

Abgabetermin: 8. November 2013
Abstract

In the modeling of financial time series, mathematicians have always incurred limits when trying to fit models to datasets. It seems like the evolution of market phenomena is too complex to be captured by an easy parametric model. In 2003, Politis introduced a model-free data-based approach that he named NoVaS (Normalizing and Variance Stabilizing) method. As no particular model is invoked, the loss of information whilst fitting a model to the data is avoided. At the same time, due to its parsimony, NoVaS is a very intuitive and numerically feasible approach.

In this thesis, we give an overview of the existing NoVaS theory in an univariate setting and extend it to the case of multivariate time series. We then introduce possible applications of the methodology, including the construction of minimum variance portfolio and intuitive estimation of parameters of a GARCH(1, 1) process without invoking likelihood methods. In the meantime, we lead an empirical analysis, in which we compare the performance of NoVaS to standard methods in the literature in a multivariate setting.
Hiermit erkläre ich, dass ich die Masterarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Hong Kong, den 08. November 2013
Acknowledgements

First of all I would like to thank my supervising Professor Claudia Klüppelberg, for the motivating words before starting the thesis, and for constant supervision, despite a very full schedule. I owe special thanks to Professor Dimitris Politis from the UC San Diego, who during his stay at the TU Munich in the context of the TUM-IAS visiting fellowship, was very patient, and, due to his undisputed expertise in the domain and as ”the father” of NoVaS, was of great help. Furthermore, I would like to thank Professor Dimitrios Thomakos, for granting access to the code that was used in previous papers on NoVaS. Having the code at my disposal made my life much easier.
Further acknowledgements go to my family. Without their moral and financial support, I would have never made it until here.
Contents

1 Introduction 2

2 Univariate NoVaS 4
   2.1 Motivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
   2.2 Setting . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
   2.3 Transformation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
      2.3.1 Choice of $G(\cdot; \alpha, \lambda)$ . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
      2.3.2 Choice of Weights . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
      2.3.3 Distributional matching . . . . . . . . . . . . . . . . . . . . . . . . . 13
   2.4 Empirical Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
      2.4.1 NoVaS on DGP-process . . . . . . . . . . . . . . . . . . . . . . . . . 16
      2.4.2 NoVaS on Real Data . . . . . . . . . . . . . . . . . . . . . . . . . . . 20

3 Multivariate NoVaS 29
   3.1 Idea . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
   3.2 Form of Correlation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30
      3.2.1 Approach via MLE . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
      3.2.2 Approach via LSE . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
   3.3 Empirical Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
      3.3.1 Realized Correlation . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
      3.3.2 Multivariate NoVaS on simulated Data . . . . . . . . . . . . . . . . . . 37
      3.3.3 Multivariate NoVaS on Real Data . . . . . . . . . . . . . . . . . . . . 46
Chapter 1

Introduction

Time series analysis is playing a very important role in today’s society, a role that has been gaining importance over the last years due to an ongoing quick growth in the amount of stored data. The goals of time series analysis are very broad and range from descriptive analysis, such as determining trends (increase, decrease), patterns (cyclic patterns, seasonal effects), to spectral analysis (analysis in the frequency domain), through forecasting, intervention analysis (“Is there a change in the time series before and after a certain event?”) and explanatory analysis (“What relationship is there between two time series?”). Most of these goals require first of all, to find a relationship between the data and a certain model. We will give an overview of the existing literature of time series modeling in the beginning of Chapter 2. After the model is chosen, the model is supposed to be fitted to the data. In these two steps, a lot of the information that is contained in the data is lost. The amount of loss depends on the complexity of the model: a more complex model allows a smaller loss of information. Unfortunately, the more complex the model, the more difficult it is to handle numerically. Therefore seemingly only a certain degree of complexity in the data can be captured by a model.

With that in mind, Politis introduced a new method in 2003, that he called Normalizing and Variance Stabilizing (NoVaS) method. With the aim of not invoking any particular model, Politis presented a method that was completely data-based, and relied on no specific model. As we will further explain in following chapters, Politis basically introduced an invertible transformation from the given data to a new dataset with known distribution, that one can easily handle.

In his papers Politis (2003a), Politis (2003b), Politis (2004), Politis (2007), Politis, Thomakos (2008a) and Politis, Thomakos (2008b), Politis focusses primarily on univariate time series, and his main application of the new method is the in financial mathematics omnipresent volatility forecasting. He concludes in numerical studies, that
his NoVaS volatility predictor is at least competitive with GARCH based predictors, which is surprising considering the simplicity and parsimony of the method.

Even though univariate series have a very high practical utility, in many fields it is most of all the dependence structure between different univariate time series that is relevant. In portfolio analysis for instance, with increasing number of assets in the portfolio, the volatility of the individual asset looses importance, and the correlation between the assets becomes the predominant factor. Literature on conditional correlation modeling is very broad. We will give a brief summary in the beginning of Chapter 3.

In multivariate time series, the problems of univariate models remain, and even gain importance: due to the higher complexity, the number of parameters of a model necessary to capture the information increases as well. As we will see later in this thesis, the amount of parameters quickly gets out of hand, which makes numerical evaluation difficult.

This does again call for a model free method. In their paper Politis and Thomakos (2011), Politis and Thomakos introduce an extension of the univariate NoVaS theory to multivariate time series.

We will in this thesis start by giving an overview of univariate NoVaS in Chapter 2, extend the theory to multivariate time series in Chapter 3, where we will present the ideas of Politis and Thomakos (2011), and refine them, trying to improve the numerical results. We continue by introducing possible applications of NoVaS. In Chapter 4 we will give a possible application in the setting of portfolio analysis: using NoVaS, we try to calculate the optimal hedge ratio of a portfolio. In the meantime, we will lead an empirical study in order to compare multivariate NoVaS to standard methods used to model multivariate time series. We will conclude with Chapter 5 by introducing a possible method to estimate the parameters of an GARCH model with the help of NoVaS, and again compare the performance of the method to standard likelihood based methods.
Chapter 2

Univariate NoVaS

In the following, we will present the NoVaS transformation, as first introduced by Politis (2003). We will start by giving a quick overview over the history of time series modelling, with the goal of motivating the model-free approach NoVaS. We will then move on to describe the setting and the actual NoVaS transformation, and will finish by presenting some empirical results of the univariate NoVaS transformation.

2.1 Motivation

The task of modeling financial returns has occupied financial mathematicians since the beginning of the last century. Pioneering work has been done by Bachelier (1900), who started by using a random walk model for the logarithm of stock market prices. Bachelier argued that a return series $X_t$ can be modelled as independent, identically distributed (i.i.d.) random variables with Gaussian $\mathcal{N}(0, \sigma^2)$ distribution.

The argument of Gaussianity was refuted in the 60s when the heavy tails of the distribution of the returns were noticed. For instance, the paper of Fama (1965) presents results which show, that the tails of a normal distribution are less fat than the ones of the distribution of returns. A solution seemed to be to choose a non-normal distribution with heavier tails for the returns.

In Mandelbrot (1963), the author pointed out that it was not only the heavy tails that were causing problems: the phenomenon of ‘volatility clustering’, i.e. the fact that high (and low) volatility days occur in clusters (see Figure 2.1), made the original model of Bachelier look too simple. Actually ‘volatility clustering’ refuted the assumption of independent returns and suggested positive correlation of squared returns.
2.1 Motivation

It was only in 1982 that Engle presented his famous ARCH (Autoregressive Conditional Heteroscedasticity) model (Engle (1982)). ARCH models are designed to capture the effect of 'volatility clustering' by assuming a particular structure of dependence for the time series of squared returns \( \{X_t^2\} \). A typical ARCH\( (p) \) model would be described by the following equation:

\[
X_t = \sqrt{a + \sum_{i=1}^{p} a_i X_{t-i}^2} Z_t
\]  

(2.1)

where \( Z_t \) is assumed to be i.i.d. \( \mathcal{N}(0, 1) \) and \( p \) an integer indicating the order of the model. ARCH models did indeed allow for a marginal distribution of the returns to have heavier tails than the normal. However, the degree of heavy tails empirically found in the distribution of returns was still not matched by the ARCH model. Several deviations from the ARCH (GARCH, EGARCH, ..., see Bollerslev (1992) for a review) had the same properties. This can nicely be seen in the example of the market crash of October 1987 where Nelson (1991) shows that empirical distribution is 6 to 7 standard deviations away from the best ARCH model.

The obvious consequence was to again use a distribution with heavier tails as innovation in the ARCH model. A popular choice for a distribution of \( Z_t \) is the t-distribution, where the degrees of freedom are empirically chosen such that the degree of heavy tails is matched.
2.2 Setting

This choice of a $t$-distribution seems somehow to be rather arbitrary. It feels like we are in the same situation as in the 1960s trying to model the excess kurtosis by an arbitrary chosen heavy-tailed distribution.

It is probable that phenomena happening in the real world like the evolution of market returns are just too complex to be captured by a neat parametric model such as a GARCH model. In the next sections we will present an alternative to model-based representation of returns: we will instead introduce a data-based, non parametric approach. We will discuss a normalizing and variance-stabilizing (NoVaS) transformation, introduced by Politis (2003); (2007); (2008), show some of its properties and present empirical results on simulated as well as real data.

2.2 Setting

Let us consider a zero mean, strictly stationary time series $\{X_t\}_{t \in \mathbb{Z}}$ corresponding to the returns of a financial asset. We furthermore assume:

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 $h^2_t \overset{def}{=} \mathbb{E}[X^2_t | \mathcal{F}_{t-1}]$ that exhibits strong dependence, with $\mathcal{F}_{t-1} \overset{def}{=} \sigma(X_{t-1}, X_{t-2}, \ldots)$.

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

The usual procedure would now be to assume a model and fit it to the data. For instance if a GARCH($p,q$) model seems suitable, one uses pseudo maximum likelihood methods to choose the parameters. But as mentioned earlier, this may not be satisfactory, as the loss of information might be too high. We will proceed in a different way.

Our goal is to find a transformation $T$ that maps our data $\{X_t\}_{1 \leq t \leq n}$ onto a sequence $W_t = \{W_1, \ldots, W_n\}$, where the $W_i$, $i = 1, \ldots, n$ are i.i.d. with identical distribution $F$.

The first step to finding such a transformation is to account for the time varying conditional variance of the returns. To do so, we construct an empirical measure $\gamma_t$ of the time-localized variance of $X_t$ based on the information set $\mathcal{F}_{t(t-p)} \overset{def}{=} \{X_t, X_{t-1}, \ldots, X_{t-p}\}$:

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

(2.2)

where $\alpha$ is a scalar control parameter, $a \overset{def}{=} (a_0, a_1, \ldots, a_p)^T$ is a $(p+1)$-dim vector of control parameters and $G(\cdot; \alpha, a)$ is to be specified. We will discuss the choice of the function $G$.
and of the parameters $a$, $p$ and $\alpha$ later.

In a second step, $\gamma_t$ is used to construct a studentized version of the returns. We define:

$$W_t = W_t(\alpha, a) = \frac{X_t}{\Phi(\gamma_t)}$$

(2.3)

where $\Phi(\cdot)$ is defined relative to our choice of $G(\cdot; \alpha, a)$. We will also come back to this later.

As already indicated in the acronym NoVaS, the next step is to choose the parameters $\alpha$ and $a$ in order to "normalize" and "stabilize the variance".

**Remark 2.2.1** As one can tell by looking at the acronym NoVaS, when Politis first thought about his transformation, the only targeted distribution $F$ for the $W_i$ was the normal distribution. The normal distribution is indeed in practise the most intuitive one and works well as we will show empirically. Nevertheless, it was shown since then that other typical distribution function can also be used. We will later on also work with the uniform distribution, which will allow us to achieve good results.

**Remark 2.2.2** When choosing the parameters in the distribution matching procedure, the goal would be to achieve for instance joint normality of $W_i$. This could be achieved by attempting to normalize linear combinations of the form $W_i + \lambda W_{i-k}$ for different values of $k$ and the weights $\lambda$. We will for now focus on the distributional matching of the first marginal distribution. As it appears, it is sufficient for practical application.

We would like to emphasize that until now, no model has been defined and no structural assumptions were required. Actually nothing other than basic properties of the time series, corresponding to the ‘stylized facts’ of financial returns, has been assumed.

### 2.3 Transformation

After introducing the general setting, let us now present possible intuitive choices for the function $G(\cdot; \alpha, a)$, and the control parameters $\alpha$ and $a$.

#### 2.3.1 Choice of $G(\cdot; \alpha, a)$

The function $G(\cdot; \alpha, a)$ can be expressed in a variety of ways, using a parametric or a semi-parametric specification. When motivating the choice of $G(\cdot; \alpha, a)$ made by Politis in e.g. [Politis (2003b)](Politis), one should first reconsider our goal: we want to construct a measure
of the time-localized variance of $X_t$, without invoking any particular model. We want the measure of the variance to be intuitive and parsimonious and we want to be able to finally use the measure to transform our data.

To motivate the choice made by Politis, we should start by observing the already mentioned ARCH($p$) Model [2.1]. When solving for the innovation term, we find:

$$X_t = \sqrt{a + \sum_{i=1}^{p} a_i X_{t-i}^2} Z_t$$

$$\Leftrightarrow Z_t = \frac{X_t}{\sqrt{a + \sum_{i=1}^{p} a_i X_{t-i}^2}}$$

where $Z_t$ is thought of being perfectly normalized and variance stabilized as it is assumed to be $\mathcal{N}(0,1)$. Therefore, [2.4] describes a possible transformation from an arbitrary dataset of financial returns \{X_t\}, to a series \{Z_t\}, which consists out of normal i.i.d. random variables. Indeed, dividing $X_t$ by $\sqrt{a + \sum_{i=1}^{p} a_i X_{t-i}^2}$ yields a standard normal random variable, if we believe in ARCH models. Even if we do not believe in these models, we can assume, that $Z_t$ might be close to a standard normal random variable, and can take this approximation as a starting point.

Furthermore, there seems to be no reason - other than coming up with a neat model - to exclude the value of $X_t$ - the "current" value of the series - from an empirical, causal estimate of the time-localized variance. Hence, we could define our $\gamma_{t}$ by:

$$\gamma_t \overset{\text{def}}{=} G(\mathcal{F}_t|t-p; \alpha, a) = \alpha s_{t-1}^2 + \sum_{i=0}^{p} a_i g(X_{t-i}) \quad \text{for } t = p + 1, p + 2, \ldots, n$$

(2.5)

where in the above, $s_{t-1}^2$ is an estimator of the unconditional variance of $X_t$ $\sigma_X^2 = Var(X_1)$, based on the data up to time $t$. Assuming a zero-mean series $X_t$, the natural estimator is

$$s_{t-1}^2 = \frac{1}{t-1} \sum_{k=1}^{t-1} g(X_k)$$

Looking closely at (2.5) and Model (2.1), the only differences are that (2.5) has an extra term, $a_0 g(X_t)$, and that there is more flexibility because of the possible choices of $g$.

By the definition, we have restrictions on the parameters, mostly to maintain positivity of $\gamma_t$: we require that $\alpha \geq 0$, $a \geq 0$, $g(\cdot) \geq 0$ and $a_p \neq 0$ for identifiability. The natural choices for $g(z)$, especially if we again compare (2.5) to Model (2.1), are $g(z) = z^2$ and $g(z) = |z|$.

Our measure of time-localized variance is therefore a combination of the unconditional variance of the returns, and a weighted average of the current and last $p$ values of the
squared or absolute returns.

Our studentized returns therefore look like:

\[
W_t = W_t(\alpha, a) = \frac{X_t}{\Phi(\gamma_t)} = \frac{X_t}{\Phi(\alpha s_{t-1}^2 + \sum_{i=0}^{p} a_i g(X_{t-i}))}
\]  

(2.6)

where \(\Phi(z) = \sqrt{z}\) if \(g(z) = z^2\) and \(\Phi(z) = z\) if \(g(z) = |z|\).

Even though we did explicitly not want \(X_t\) to follow model, one should mention the notion of implied model. In fact, rearranging (2.6), with \(g(z) = z^2\), yields:

\[
W_t^2 = \frac{X_t^2}{\alpha s_{t-1}^2 + \sum_{i=0}^{p} a_i X_{t-i}^2} 
\]

\[\iff \quad X_t^2 = \left(\alpha s_{t-1}^2 + \sum_{i=0}^{p} a_i X_{t-i}^2\right) W_t^2 \]

\[\iff \quad X_t^2 - a_0 X_t^2 W_t^2 = \left(\alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i X_{t-i}^2\right) W_t^2 \]

\[\iff \quad X_t = \left(\sqrt{\alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i X_{t-i}^2}\right) \frac{W_t}{\sqrt{1 - a_0 W_t^2}} \]  

(2.7)

Analogously, with \(g(z) = |z|\):

\[
X_t = \left(\alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i |X_{t-i}|\right) \frac{W_t}{1 - a_0 |W_t|}
\]

Therefore we obtain, in both cases, a model of the sort:

\[
X_t = U_t A_{t-1} \quad (2.8)
\]

where \(U_t\) and \(A_t\) are given by:

\[
U_t \overset{\text{def}}{=} \begin{cases} 
W_t / \sqrt{1 - a_0 W_t^2} & : \text{if } g(z) = z^2 \\
W_t / (1 - a_0 |W_t|) & : \text{if } g(z) = |z|
\end{cases}
\]

and

\[
A_{t-1} \overset{\text{def}}{=} \begin{cases} 
\sqrt{\alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i X_{t-i}^2} & : \text{if } g(z) = z^2 \\
\alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i |X_{t-i}| & : \text{if } g(z) = |z|
\end{cases}
\]

Observing (2.7), one immediately sees the resemblance to an ARCH(\(p\)) model. Actually, the only differences lie in the constant term \(\alpha s_{t-1}^2\) and especially in the distribution of the innovation \(\frac{W_t}{\sqrt{1 - a_0 W_t^2}}\). These are two important points, which deserve their own remarks.
Remark 2.3.1 In Politis (2004), the distribution of \( U_t \) is discussed in the context of heavy-tailed distribution of ARCH-Residuals. He motivates it as a more 'natural' and less ad hoc distribution of ARCH/GARCH residuals. Let us quickly discuss some of the properties of this distribution, in the case \( g(z) = z^2 \):

To be able to calculate the density of \( U_t \), we will first of all need the distribution of \( W_t \).

We will in a next step try to choose the open parameters \( \alpha \) and \( a \) such that the distance between the distribution of \( W_t \) and a specified, feasible distribution \( F \) is minimized. Let us therefore assume that \( W_t \) approximately follows a normal distribution, as this is the most intuitive choice for \( F \). Looking closely at (2.7) one sees:

\[
\frac{1}{W_t^2} = \frac{\alpha s^2_t + a_0 \sum_{i=1}^{p} a_i X^2_{t-i}}{X^2_t} \geq a_0
\]

and thus:

\[
|W_t| \leq \frac{1}{\sqrt{a_0}} \quad (2.9)
\]

which means that exact normality may not hold for \( W_t \). A natural way to handle such a situation would be to assume a truncated normal distribution, i.e., to assume that the \( W_t \) are i.i.d. with density:

\[
\frac{\Phi(x)1\{|x| \leq C_0\}}{\int_{-C_0}^{C_0} \Phi(y)dy} \quad \text{for all } x \in \mathbb{R}
\]

where \( \Phi \) denotes the standard normal density, and \( C_0 = 1/\sqrt{a_0} \). Actually, when \( a_0 \) is chosen small enough, the boundedness is effectively not noticeable: Recalling that 99.7% of the mass of the \( \mathcal{N}(0,1) \) distribution lies in the range \( \pm 3 \), having \( a_0 \leq 1/9 \) seems to be a good reference.

Assuming thus, that \( W_t \) follows a truncated normal distribution, we find:

\[
P \left[ U_t \leq z \right] = P \left[ \frac{W_t}{\sqrt{1 - a_0 W_t^2}} \leq z \right] = P \left[ W_t \leq z \sqrt{1 - a_0 W_t^2} \right] = P \left[ W_t^2 \leq z^2 \left( 1 - a_0 W_t^2 \right) \right] = P \left[ W_t^2 \left( 1 + a_0 z^2 \right) \leq z^2 \right] = P \left[ W_t \leq \frac{z}{\sqrt{1 + a_0 z^2}} \right]
\]
2.3 Transformation

and hence

\[
f_{U_t}(x) = f_{W_t}\left(\frac{x}{\sqrt{1 + a_0x^2}}\right) \cdot \left(\frac{x}{\sqrt{1 + a_0x^2}}\right)'
\]

\[
= \phi\left(\frac{x}{\sqrt{1 + a_0x^2}}\right) \mathbf{1}\{\frac{x}{\sqrt{1 + a_0x^2}} \leq C_0\} \cdot (1 + a_0x^2)^{-3/2} \int_{-C_0}^{C_0} \phi(y)dy
\]

\[
= (1 + a_0x^2)^{-3/2} \exp\left(-\frac{x^2}{2(1+a_0x^2)}\right) \sqrt{2\pi} \left(\Phi(1/\sqrt{a_0}) - \Phi(-1/\sqrt{a_0})\right)
\]

Similarly, if \(g(z) = |z|\),

\[
\mathbb{P}[U_t \leq z] = \mathbb{P}\left[W_t \leq \frac{z}{1 + a_0|z|}\right]
\]

and therefore:

\[
f_{U_t}(x) = \frac{(1 + a_0|x|)^{-2} \exp\left(-\frac{x^2}{2(1+a_0|x|)^2}\right)}{\sqrt{2\pi} \left(\Phi(1/|a_0|) - \Phi(-1/|a_0|)\right)}
\]

In a numerical study, Politis compares the distribution of \(U_t\) to other heavy tailed distributions. He realizes that for \(g(z) = z^2\) the distribution is close to a \(t_2\) distribution, as the rate by which the density of \(U_t\) tends to 0 in the tails is the same as in the \(t_2\) case. Due to the different constants associated with the rate of convergence, Politis explains that the tails of \(U_t\) are still lighter than those of the \(t_2\) distribution, and concludes that the density of \(U_t\) "achieves its degree if heavy tails in a subtler way". \(U_t\) seems therefore to be a good choice for the innovations in a (G)ARCH process. For more details and empirical proof, check Politis (2004).

Remark 2.3.2 The second difference between ARCH and the implied NoVaS equation if a normal target distribution \(g(z) = z^2\) is used, is the difference in the constant term. But from a practical point of view, replacing the term \(a\) in (2.1) with \(\alpha s_t^{-1}\) in (2.7) is only natural: \(a\) is not scale invariant whereas \(\alpha s_t^{-1}\) is, because it has by necessity units of variance. Thus it is much easier to compare the \(\alpha\), than to compare the \(a\) between two models.

Even though (2.7) looks neat, one should not view it as a model: our focus will lie on (2.6).
2.3 Transformation

2.3.2 Choice of Weights

Remember our transformation:

\[ W_t = \frac{X_t}{\Phi(\alpha s_{t-1}^2 + \sum_{i=0}^{p} a_i g(X_{t-i}))} \]

We will now have to decide on how to choose the open parameters \( p \) and \((\alpha, a)\). What seems obvious, is that \( \alpha \) and \( a \) should be chosen non-negative. Next, to ensure unit of variance and unbiasedness Politis in Politis (2003b) suggests:

\[ \alpha + \sum_{i=0}^{p} a_i = 1 \]

A further possible restriction is to impose \( a_i \geq a_j \) for \( i < j \) because the coefficients can be thought of as smoothing weights. We should not lose sight of the fact that we want to achieve a degree of conditional homoscedasticity. We therefore would like for \( p \) and \( \alpha \) to be small enough, so that a local and not a global estimator of scale is obtained.

Bearing in mind, that we opt for simplicity and parsimony, our objective is to achieve a distributional matching procedure with as few parameters as possible.

Under consideration of these requirements, Politis suggests two choices:

- **Simple NoVaS**: \( \alpha = 0 \) and \( a_j = 1/(p + 1) \) for all \( j = 0, 1, ..., p \). This results in an equal weighting of the last \( p \) returns. It requires the calibration of only one parameter namely \( p \). The simple NoVaS is a very intuitive approach as it basically corresponds to the popular time series method of obtaining a local average.

- **Exponential NoVaS**: \( \alpha = 0 \) and

\[
a_j = \begin{cases} 
1/\sum_{i=0}^{p} \exp(-ci) & j = 0 \\
 a_0 \exp(-cj) & j = 1, 2, ..., p 
\end{cases} \quad (2.10)
\]

with \( c > 0 \). This results in greater weight placed in earlier lags. It requires the calibration of 2 parameters: our rate \( c \), and the lag \( p \), even though because of the decreasing nature of the weights, \( p \) is of secondary importance, as we will further see in Chapter 5.

While the simple NoVaS is equivalent to forming a local average, the exponential NoVaS can be compared to the time series method of exponential smoothing.

The second choice, namely the exponential NoVaS allows for greater flexibility, and will be our preferred method.
2.3 Transformation

The case of $\alpha \neq 0$ will be treated later on, but will not cause more difficulties. When choosing the weights, we should also be alert of the boundedness of $W_t$, as seen in (2.9). To ensure that the truncation of the normal distribution is at a practically acceptable level, one should remember the rule that for $a_0 \leq 1/9$, in the case $g(z) = z^2$, the values lie within $\pm 3$, resulting in 99.7% of the mass of a normal distribution being covered. We should therefore make sure, that $a_0 \leq 1/9$, for this will yield acceptable truncation.

Remark 2.3.3 We have, by choosing either simple or exponential NoVaS, achieved to reduce the number of parameters from $p+2$ to 1 (simple NoVaS) or 2 (exponential NoVaS), which is in our interest, as we are looking for a parsimonious and simple approach.

For the sake of simple notation, let us call the open parameters $\theta$, and therefore: $\theta \mapsto (\alpha, a)$. Let us denote $W_t \equiv W_t(\theta)$.

2.3.3 Distributional matching

In the next step we will calibrate the parameters in either exponential or simple NoVaS with the goal of distributional matching. Hence we will try to minimize the ‘distance’ of the distribution of our $W_t$ to the target distribution. To measure the distance, many different functions may be used.

In his papers of 2003 and 2007, Politis proposed two main methods:

- **Moment Based Matching:**
  With our assumptions on the data (zero mean, symmetric with excess kurtosis), the first moment worth matching is the fourth moment. Let us therefore define the sample excess kurtosis of the studentized returns as:
  \[
  \kappa_n(\theta) \overset{df}{=} \frac{\sum_{t=1}^n (W_t - \bar{W}_n)^4}{n s_n^4} - \kappa^*
  \]  
  \[(2.11)\]
  where $\bar{W}_n \overset{df}{=} \frac{1}{n} \sum_{t=1}^n W_t$ is the sample mean, $s_n^2 \overset{df}{=} \frac{1}{n} \sum_{t=1}^n (W_t - \bar{W}_n)^2$ is the sample variance and $\kappa^*$ denotes the theoretical kurtosis of the target distribution. The goal would then be to minimize the absolute excess kurtosis. In other words, we would want to find $\theta$, such that the objective function $D_n(\theta) \overset{df}{=} |\kappa_n(\theta)|$ is minimized. An appropriate algorithm will be discussed in the next section.

- **Distributional matching via goodness-of-fit statistics:**
  The standard statistics that are used to check for the similarity of two distributions are Shapiro-Wilks type of statistics like the quantile-quantile correlation coefficient,
2.4 Empirical Analysis

or the Kolmogorov statistic.
The construction of the first one works as follows: For any given value of \( \theta \), one computes the order statistic \( W_t \), that is \( W_1 \leq W_2 \leq W_3 \leq \cdots \leq W_n \), and the quantiles of the target distribution \( Q(t) \). To measure the distributional goodness of fit, one can calculate the squared correlation coefficient in the simple regression on the pairs \([Q(t), W(t)]\). If the target distribution is the normal, this corresponds to the Shapiro-Wilks test for normality, and we find that:

\[
D_n(\theta) \overset{de}{=} 1 - \frac{\left| \sum_{t=1}^{n} (W_t - \bar{W}_n) (Q(t) - \bar{Q}(t)) \right|^2}{\left[ \sum_{t=1}^{n} (W_t - \bar{W}_n)^2 \right] \left[ \sum_{t=1}^{n} (Q(t) - \bar{Q}(t))^2 \right]}
\]

Alternatively, on objective function based on the Kolmogorov-Smirnov statistic would be:

\[
D_n(\theta) \overset{de}{=} \sup_t \sqrt{n} \left| F_t - \hat{F}_{W,t} \right|
\]

No matter what we choose as an objective function, our optimized \( \theta^*_n \), hence the \( \theta \) that allows for the highest degree of distributional matching, is determined by:

\[
\theta^*_n \overset{de}{=} \arg \min_{\theta} D_n(\theta)
\]

This results in our studentized series:

\[
W^*_t \equiv W_t(\theta^*_n)
\]

2.4 Empirical Analysis

In this section, we will look at the performance of the NoVaS empirically. We will first shortly discuss a possible implementation of the algorithm, then see how NoVaS performs with randomly generated data, and finally observe the performance of NoVaS with real financial returns.

As we have seen earlier, there are many ways to perform a NoVaS transformation: one can choose the target distribution (normal or uniform), the \( g(z) \) (absolute or squared) and the objective function (kurtosis-based, Kolmogorov-Smirnov statistics, QQ-correlation). The first question is therefore how to choose the appropriate combination.

Our favourite combination will be the one that minimizes the objective function \( D_n \). We will therefore go through all combinations, and pick the one that results in the smallest \( D_n(\theta) \): For instance, we can fix the type of normalization (squared or absolute returns) and the target distribution, and then perform the matching procedure using all three
2.4 Empirical Analysis

Objective functions, resulting in a $(3 \times 1)$ vector, which we call $D_m(\nu, \tau)$, with $m = \text{kurtosis, QQ-correlation, KS-statistics, } \nu = \text{squared, absolute returns and } \tau = \text{normal, uniform target distribution.}$ Then, the optimization procedure has to be repeated for all combinations of $(\nu, \tau)$. Finally one can choose the optimal combination across all possible combinations $(m, \nu, \tau)$ by:

$$d_m^* \overset{\text{def}}{=} \arg\min_{(\nu, \tau)} D_m(\nu, \tau)$$

$$d^* \overset{\text{def}}{=} \arg\min_m D_m(\nu, \tau)$$

We can specify an algorithm for the NoVaS transformation, both for the simple and exponential NoVaS.

- **Algorithm for Simple NoVaS**

1. Let $\alpha = 0$ and $a_i = 1/(p + 1)$ for all $0 \leq i \leq p$
2. Pick $p$ such that $D_n(\theta)$ is minimized

**Remark 2.4.1** One might wonder, if there exists a convincingly minimizing $p$. Let us have a look at the moment matching procedure. We can note, that for $p = 0$, we have $a_0 = 1$, $W_t = \text{sign}(X_t)$, and $K(W_t) = 1$. If we let, on the other hand, $p$ be large, one can expect, that due to the large kurtosis of $X_t$, $K(W_t)$ will become large (for instance if the target distribution is the normal, larger than 3). Actually, the law of large numbers implies that for increasing values of $p$, $K(W_t)$ will tend to the real kurtosis of $X_t$, which is by assumption large. Therefore, if we see $K(W_t)$ as a smooth function of $p$, the intermediate value theorem suggests that there exists a $p$ for which the desired value of $K(W_t)$ (for instance 3, if the target distribution is the normal distribution) is approximately attained. This is what happens in practice.

Even though the algorithm seems to work quite well in practice, there is one remaining problem: in the case of a normal target distribution, we need to make sure, that our truncation is not too noticeable. More precisely, we wanted, according to the Remark 2.3.1, that $a_0 = 1/(p + 1) \leq 1/9$. If this condition is not satisfied, the following step can be added to the algorithm:

3. If $p$ is such that $a_0 > 1/9$, increase $p$ until it is the smallest integer such that $1/(p + 1) \leq 1/9$

- **Algorithm for exponential NoVaS**

1. Let $p$ take a very high starting value, e.g. $p = n/5$, where here, $n$ is the length of our sample.
2.4 Empirical Analysis

2. Let $\alpha = 0$ and $a_i = c' \exp(-ci)$, for all $0 \leq i \leq p$, where $c' = 1/\sum_{i=0}^{p} \exp(-ci)$.

3. Pick $c$ such that $D_n(\theta)$ is minimized.
   The same problem as above with the truncation parameter might occur, in which case one adds:

4. If $c$ as found is such that $a_0 > 1/9$, decrease $c$ until the condition is satisfied $a_0 < 1/9$.

5. Finally, the value of $p$ has to be trimmed: One can simply discard all coefficients $a_i$ that fall below a certain threshold. A threshold of 0.01 seems to be reasonable, and works well in practice. Thus: if $a_i < 0.01$, for all $i > i_0$, let $p = i_0$ and renormalize the $a_i$ so that $\sum_{i=0}^{p} a_i = 1$.

2.4.1 NoVaS on DGP-process

Let us now have a look at the result of a NoVaS transformation of simulated data. To this end, we look at a GARCH(1,1) process with t-innovations, with three degrees of freedom and $n = 300$:

$$X_t = \sqrt{2.29 \cdot 10^{-6} + 0.062X_{t-1}^2 + 0.898\sigma_{t-1}^2} \ Z_t$$

where $Z_t$ follows a $t$-distribution with three degrees of freedom. In Figure 2.2 you can see the series, its volatility process, the histogram of the distribution of the return and the QQ-plot of the empirical distribution of the returns against a standard normal distribution. We have omitted the QQ-plot to check whether the empirical distribution follows a uniform distribution, as one can clearly exclude this assumption by looking at the histogram.

One can again see that the assumptions made in Section 2.2 are valid. The non-normality is easily visible when looking at the QQ-plot. The time varying volatility with possible volatility clustering phenomena can be seen when looking at the volatility process. That the distribution is centred and symmetric can be seen when looking at the histogram.
In the following Table 2.1, the results of the NoVaS transformation with the different methods are listed. The entries correspond to the value of the objective function. For the kurtosis, the QQ-correlation and the KS-statistic, the smaller the value, the better the matching. These results are the basis for the model selection for the univariate NoVaS transformation. The method (Normal target squared returns, Normal target absolute returns, Uniform target squared returns or Uniform target absolute returns) with the best resulting value of the objective function (Exc. Kurtosis, QQ-Correlation and KS-p-value) is going to be the one chosen for performing the NoVaS transformation.

Figure 2.2: Simulated GARCH(1,1)-process, it’s volatility, Empirical distribution, QQ-Plot with respect to normal distribution
2.4 Empirical Analysis

We can see that a uniform target distribution seems to yield better results than the normal target distribution. Furthermore, the statistics Excess Kurtosis and Kolmogorov-Smirnov-p-value agree on $g(z) = z^2$ being a better choice than $g(z) = |z|$. This is therefore the method that we are going to use to perform our transformation. In the following Figure 2.3, one can see the resulting return series with its volatility process, the QQ-Plot and the histogram of the transformed series.
2.4 Empirical Analysis

We can see that except a couple of unexpected outliers at around the index 100, the volatility process is stabilized. The distribution of the transformed series is now very close to a uniform distribution. Therefore, we can say that we have achieved our goals.

Until now, we are not able to quantitatively judge the performance of the transformation: We can not really compare it to anything. Politis used the NoVaS transformation for volatility forecasting, and seemed to achieve satisfying results, compared to benchmark methods (see Politis (2004)). What we know, is that we are able to achieve to match the distribution that we wanted, as one can see from the QQ-plot and Table 2.1. We will only be able to judge the method when talking about multivariate NoVaS, because we will then be able to measure the ability to capture the correlation of two return series (especially in Chapter 4, where we are going to compare standard correlation capturing methods).

Figure 2.3: Simulated GARCH(1,1)-process, it’s volatility, Empirical distribution, QQ-Plot with respect to normal distribution
2.4 Empirical Analysis

2.4.2 NoVaS on Real Data

We will now look at the performance of NoVaS when trying to transform data from real return series. To this end we are going to use three series: The S&P500, the 10-year bond (10-year maturity constant maturity rate series) and the USD/Japanese Yen exchange rate. From daily data from the beginning of the series, we trim, align and finally compute monthly returns and realized volatilities and correlations. The final data sample is from 01/1971 to 02/2010 for total of n=469 observations. Figure 2.5 shows the return series, with their respective volatility.

First of all, let us check the assumptions made on the data in Section 2.2: from the QQ-plots in Figure 2.6, one can see that the non-Gaussianity assumption is verified. The returns are obviously not distributed uniformly either. The Table 2.2 summarizing some statistical information on the series, shows that all three series have excess kurtosis. In Figure 2.4, one can verify the assumption of low autocorrelation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.004</td>
<td>-0.002</td>
<td>-0.003</td>
<td>0.040</td>
<td>0.040</td>
<td>0.025</td>
</tr>
<tr>
<td>Median</td>
<td>0.007</td>
<td>-0.003</td>
<td>-0.001</td>
<td>0.035</td>
<td>0.035</td>
<td>0.024</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.045</td>
<td>0.049</td>
<td>0.031</td>
<td>0.023</td>
<td>0.024</td>
<td>0.012</td>
</tr>
<tr>
<td>Skewness</td>
<td>-1.213</td>
<td>0.054</td>
<td>-0.449</td>
<td>4.247</td>
<td>2.070</td>
<td>0.878</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>9.149</td>
<td>5.799</td>
<td>5.730</td>
<td>33.938</td>
<td>10.134</td>
<td>5.550</td>
</tr>
</tbody>
</table>

Table 2.2: Statistics three financial return series
Figure 2.4: The autocorrelation plots of the three financial return series
2.4 Empirical Analysis

Figure 2.5: Returns and volatility of our three financial return series
2.4 Empirical Analysis

Figure 2.6: QQ-Plot to check for normal or uniform distribution of our three financial return series

Now we will again want to choose the right combination for our NoVaS transformation. We therefore compute the value of our objective functions after optimization for all three time series. The results are shown in the following Table 2.3.
### Table 2.3: Model Selection Univariate NoVaS for Real Data

<table>
<thead>
<tr>
<th></th>
<th>Normal target squared returns</th>
<th>Normal target absolute returns</th>
<th>Uniform target squared returns</th>
<th>Uniform target absolute returns</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S&amp;P 500</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exc. Kurtosis</td>
<td>0.000115</td>
<td>0.000241</td>
<td>0.000003</td>
<td>0.000006</td>
</tr>
<tr>
<td>QQ-Correlation</td>
<td>0.008344</td>
<td>0.011550</td>
<td>0.002774</td>
<td>0.006820</td>
</tr>
<tr>
<td>KS-pvalue</td>
<td>0.335374</td>
<td>0.110251</td>
<td>0.0321913</td>
<td>0.018121</td>
</tr>
<tr>
<td><strong>Bonds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exc. Kurtosis</td>
<td>0.000150</td>
<td>0.000156</td>
<td>0.000007</td>
<td>0.000018</td>
</tr>
<tr>
<td>QQ-Correlation</td>
<td>0.002276</td>
<td>0.003295</td>
<td>0.002322</td>
<td>0.005776</td>
</tr>
<tr>
<td>KS-pvalue</td>
<td>0.696003</td>
<td>0.076299</td>
<td>0.0764131</td>
<td>0.036183</td>
</tr>
<tr>
<td><strong>USD/YEN</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exc. Kurtosis</td>
<td>0.000174</td>
<td>0.000020</td>
<td>0.000007</td>
<td>0.000009</td>
</tr>
<tr>
<td>QQ-Correlation</td>
<td>0.006730</td>
<td>0.007462</td>
<td>0.004309</td>
<td>0.007168</td>
</tr>
<tr>
<td>KS-pvalue</td>
<td>0.260712</td>
<td>0.193019</td>
<td>0.0229988</td>
<td>0.028211</td>
</tr>
</tbody>
</table>

According to Table 2.3, for all three time series, it seems again to be the transformation with uniform target distribution and squared returns, which delivers the best results. Especially the Excess Kurtosis and the QQ-Correlation make this observation possible. This is therefore the combination that we are going to use for the transformation, resulting in the following series:
Figure 2.7: Transformed S&P500 return series
2.4 Empirical Analysis

Figure 2.8: Transformed Bonds return series
Comparing Figure 2.6 to Figures 2.7, 2.8 and 2.9, one can see that the NoVaS transformation has clearly managed to transform the distribution of the individual series from an unknown distribution to the desired uniform distribution. Furthermore, the volatility of the transformed is obviously far from being time invariant, but has stabilized in the sense that the outliers from the volatility series have been eliminated.

The transformation can therefore be seen as successful. Now that we know the distribution of our transformed dataset, it is going to be easy to work with the data. Forecasting
volatility is for example a far easier task, on a dataset which has a know, easy distribution, and where the individual entries of the series are independent.

For more information on forecasting with NoVaS, Politis, Thomakos (2008a) is of interest.
Chapter 3

Multivariate NoVaS

Now that we are able to handle univariate series, and to transform them into a studentized i.i.d. version, let us look at multivariate data. One could argue that the approach of Chapter 3 is theoretically suitable for multivariate series as well: one could use a multivariate version of the kurtosis statistic, or a multivariate normality statistic to construct the objective function used for optimization. Unfortunately, one quickly gets into situations where numerical applications are not feasible. Multivariate numerical optimization, for instance in a unit hyperplane, is unattractive as soon as the dataset gets larger. Therefore the distribution matching approach is not suitable in multivariate context for larger scale problems.

We will in this chapter present an alternative approach, again inspired by the work of Politis and Thomakos (Politis and Thomakos (2011)). We will in the first section present the idea of NoVaS in a multivariate setting. In a second step, we will introduce possible realizations and implementations of this idea, and will finally choose the most suited idea by looking at the performance of the approach in capturing the conditional correlation by comparing it to the known realized correlation. It is only in Chapter 4 that we will compare the performance of multivariate NoVaS against other benchmarking methods, in the context of a motivating application.

3.1 Idea

As mentioned earlier, we need to distance ourselves from the idea of being able to reproduce the theory used in univariate NoVaS with multivariate series. Instead, we are going to follow an approach similar to that of many other correlation models in the literature. If one has determined a model suiting the data, one would normally use it to obtain the fitted volatility of the series, which one would then use to standardize the returns. Finally,
3.2 Form of Correlation

one would use those standardized returns to build a model for the correlations.
Our approach is going to be rather similar. Let us assume we are in a bivariate setting.
We can now use the method of Chapter 2 to transform both univariate series, one by
one. We therefore have for a pair of returns \((i, j)\) the studentized series, which we call
\(W_{t,i}^*\) and \(W_{t,j}^*\). One should keep in mind, that unlike most methods in the literature, we
have constructed our studentized returns with information up to and including time \(t\). In
a parametric context, the standardization would have been done not with data, but with
the underlying assumed model. Our advantage is therefore that we are allowed to use the
information available at time \(t\) when computing the correlation.
Furthermore, since we know the distribution of both \(W_{t,i}^*\) and \(W_{t,j}^*\), we should be able to
gain information about the properties of the distribution of the product \(Z_t(i, j) = W_{t,i}^*W_{t,j}^*\).
With help of these studentized series, we can easily get an estimator for the constant,
unconditional correlation between the returns \(\rho = \mathbb{E}[Z_t(i, j)]\), namely the sample mean of
\(Z_t(i, j)\):
\[
\hat{\rho}_n \overset{\text{def}}{=} \frac{1}{n} \sum_{t=1}^{n} Z_t(i, j) \quad (3.1)
\]
But we are more interested in the time-varying conditional correlation \(\mathbb{E}[Z_t|\mathcal{F}_{t-s}] \overset{\text{def}}{=} \rho_t|t-s\). To that end, we will in the following present two methods: the first one was presented
by Politis and Thomakos in Politis and Thomakos (2011). They proposed a parametric
model for the conditional correlation, and then, using the density of \(Z_t(i, j) = W_{t,i}^*W_{t,j}^*\),
fit the parameters of the model the conditional correlation using maximum-likelihood
methods. The second method we want to introduce uses the same form as in Politis and
Thomakos (2011), but fits the parameters in a different way, namely via some least squared
methods, which is more in accordance with the model-free setting in which NoVaS is. After
having introduced both techniques, we compare them empirically and show how the final
estimated conditional correlation behaves in comparison to the realized correlation.

3.2 Form of Correlation

To stay in mindset of the NoVaS setting, when choosing a form for our conditional cor-
relation, we opt for a simple, parsimonious and intuitive approach. We therefore use an
approach similar to an exponential smoothing method. For the conditional correlation at
time point \(t|t-s\), we use the product \(Z_{t-s}\) of the studentized returns and the conditional
correlation at time point $t - 1|t - 1 - s$, namely:

$$
E[Z_t|\mathcal{F}_{t-s}] \overset{\text{def}}{=} \rho_{t|t-s}
$$

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

(3.2)

By iteration, this yields:

$$
\rho_{t|t-s} = \lambda (\lambda \rho_{t-2|t-s-2} + (1 - \lambda) Z_{t-s-1}) + (1 - \lambda) Z_{t-s}
$$

$$
= \lambda^2 (\lambda \rho_{t-3|t-s-3} + (1 - \lambda) Z_{t-s-2}) + (1 - \lambda) (Z_{t-s} + \lambda Z_{t-s-1})
$$

$$
\vdots
$$

$$
\approx (1 - \lambda) \sum_{j=s}^{L+s-1} \lambda^{j-s} Z_{t-j}
$$

(3.3)

because $\lambda^n \to 0$ for $n \to \infty$. In (3.3) $L$ is a truncation parameter.

A further requirement for the estimator of conditional correlation could be unbiasedness.

To achieve that, we would require that the expectation of conditional correlation equals the unconditional correlation $E[Z_t] \overset{\text{def}}{=} \rho$, and therefore:

$$
E[\rho_{t|t-s}] = \rho_n = \frac{1}{n} \sum_{i=1}^{n} Z_i
$$

but since with the choice in (3.3), we would obtain:

$$
E[\rho_{t|t-s}] = E \left[ (1 - \lambda) \sum_{j=s}^{L+s-1} \lambda^{j-s} Z_{t-j} \right]
$$

$$
= (1 - \lambda) \sum_{j=s}^{L+s-1} \lambda^{j-s} E[Z_{t-j}]
$$

$$
= (1 - \lambda) \sum_{j=s}^{L+s-1} \lambda^{j-s} \rho
$$

$$
\neq \rho
$$

a more appropriate choice for $\rho_{t|t-s}$ would be:

$$
\rho_{t|t-s} = (1 - \lambda) \left( \sum_{j=s}^{L+s-1} \lambda^{j-s} Z_{t-j} + \frac{\rho}{1 - \lambda} - \sum_{j=s}^{L+s-1} \lambda^{j-s} \rho \right)
$$

(3.4)

because that way one immediately finds:

$$
E[\rho_{t|t-s}] = \rho
$$
3.2 Form of Correlation

One can of course choose different weights for the averaging procedure - a more general form of (3.3) would then be:

$$\rho_{t|t-s} = \sum_{j=s}^{L+s-1} w_j(\lambda)B^jZ_t = w(B; \lambda)Z_t$$

with $B$ the backshift operator. (3.4) is then:

$$\rho_{t|t-s} = w(B; \lambda) + (1 - w(1; \lambda))\rho$$

All that is left to do now to get the conditional correlation is to determine the parameter $\lambda$, which we are going to do in two different ways, presented in the next two subsections.

3.2.1 Approach via MLE

In this approach, formulated in Politis and Thomakos (2011), one uses the knowledge of the distribution of the studentized returns. In fact we use the following result of Rohatgi (1976):

**Theorem 3.2.1** Under the assumption that both series $W_{t,i}^*$ and $W_{t,j}^*$ were transformed with use of the same target distribution, the density function of the product of the returns $Z_t$ has the form:

$$f_Z(z) \overset{def}{=} \int_D f_{W_i,W_j}(w_i, z/w_i) \frac{1}{|w_i|} dw_i$$

where $f_{W_i,W_j}(x_i, x_j)$ is the joint density of the studentized series.

In particular, a result from Craig (Craig (1936)) gives:

**Corollary 3.2.2** Under the assumption that both series $W_{t,i}^*$ and $W_{t,j}^*$ were transformed with use of the normal distribution, the density function of the product of the returns $Z_t$ has the form $f_Z(z; \rho) = I_1(z; \rho) - I_2(z; \rho)$, with:

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

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

And using the Karhunen-Loeve transform, one can derive:
Corollary 3.2.3 Under the assumption that both series $W_{t,i}^*$ and $W_{t,j}^*$ were transformed with use of the uniform distribution, the density function of the product of the returns $Z_t$ has the form:

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

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

We can therefore, with the help of the above theorems determine the distribution of $Z_t(i,j) = W_{t,i}^* W_{t,j}^*$, under the assumption that $W_{t,i}^*$ and $W_{t,j}^*$ are both normal (Corollary 3.2.2) or both uniform (Corollary 3.2.3).

Remark 3.2.4 • A problem occurs if different target distributions are used for the two transformations of the series, because the above results only hold for the case where the target distribution is the same. In the scenario that the distributions used in the univariate modeling do not coincide, a decision has to be made as to which product distribution should be used. Again, according to Politis and Thomakos (2011), the better choice is the product distribution based on uniform marginals, because there seems to be more ‘robustness’ in practice.

• As to the choice of $L$, it can be based on the length of the NoVaS transformation $(p)$, or selected with the AIC criterion, as a likelihood function is available.

• Even though, the above functions in the corollaries use the unconditional correlation $\rho$, and not as desired the conditional correlation $\rho_{|t-s}$, according to Politis and Thomakos (2011), similar results hold for the conditional case, and in practice, the above results can be used for the conditional case.

Having agreed on a density of $Z_t$, one can now use it to get a maximum likelihood estimator of $\lambda$, the parameter in (3.2), by:

$$\hat{\lambda}_n = \arg\max_{\lambda \in [0,1]} \sum_{t=1}^{n} \log F_{Z}(Z_t; \lambda)$$

(3.5)

Up to this point, except for the maximum likelihood method, we have achieved to work only on the data, without using any model or parametric method. Therefore, and because of the possible problems mentioned in Remark 3.2.4, this approach is not satisfactory, even though it might be useful in practice. Hence, we are going to give an alternative method in the next subsection.
3.2 Form of Correlation

3.2.2 Approach via LSE

Our aim in this subsection is to find an estimate for $\lambda$ as in (3.2), without using an maximum likelihood method, and without using the density of $Z_t$, for the reason mentioned in Remark 3.2.4. We are instead going to use an error minimization procedure, and are to this end starting by suggesting different objective functions, which we are finally going to compare for suitability.

We are therefore sill using (3.2), but we are now forgetting about the density of $Z_t$, and are again only relying on the data. In the following, we are to this end defining 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 (3.2) to capture the conditional correlation.

- **Method 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:

  $$Q(\lambda) \overset{def}{=} \sum_{t=1}^{n} (\rho_{t|t-1} - Z_t)^2 \quad (3.6)$$

- **Method 2**
  Assume we know the series:

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

  and transform them individually with univariate NoVaS to get:

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

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

  $$\hat{W}_{i,T+1} \overset{def}{=} \rho_{T+1|T} W_{j,T+1} \quad (3.7)$$

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

  $$X_{i,1}, X_{i,2}, \ldots, X_{i,T}, X_{i,T+1}, \ldots X_{i,n}$$
  $$X_{j,1}, X_{j,2}, \ldots, X_{j,T}, X_{j,T+1}, \ldots X_{j,n}$$
we can use the estimates $\hat{W}_{i,k+1}$ with $k = T, \ldots, n - 1$ to get to the objective function:

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

(3.8)

In this context, $T$ should be chosen large enough, in order to guarantee that the estimate of the conditional correlation in (3.2) has enough data to work with. For practical implementation, we used $T \approx n/4$.

- **Method 3**

To account for the symmetry of the correlation, one might prefer to add to the term in (3.8) the symmetric term:

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

with

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

to get to the objective function:

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

(3.9)

- **Method 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 (3.7). One can then use the inverse transformation discussed in (2.7), namely:

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

(3.10)

Now, with help of (3.10), one can once again define the objective error function:

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

(3.11)

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

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

(3.12)
• **Method 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 predicted in the wrong way. More formally, we define the loss function $L$:

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

for $t = T + 1, \ldots, n$, and with $\hat{W}_{i,t}$ defined as in (3.7). Our objective error function is then:

$$Q(\lambda) \overset{\text{def}}{=} \sum_{t=T+1}^{n} L(t)$$  \hspace{1cm} (3.13)

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

$$\hat{\lambda} = \arg\min_{\lambda \in [0,1]} Q(\lambda)$$  \hspace{1cm} (3.14)

and therefore our conditional correlation process would be given by:

$$\rho_{t|t-s} = \hat{\lambda}\rho_{t-1|t-s-1} + (1 - \hat{\lambda})Z_{t-s}.$$  \hspace{1cm} (3.15)

Let us now see which of these methods yields the best results. We will to this end again conduct an empirical study.

## 3.3 Empirical Analysis

In the following, we will first introduce a method that will allow us to compare our estimates with the true correlation of the series. We will then present some results that will help us to compare the different methods to find $\lambda$ in (3.2). We will once again first use simulated data, to finally work on real data.

### 3.3.1 Realized Correlation

Multivariate NoVaS delivers an estimate for the correlation. To be able to judge whether these estimates are good, we first have to introduce a method that gives us the true correlation of the data. This is necessary because unlike the situation where we work with simulated data, where the 'true' underlying correlation is known, in applied work matters are more complicated. The underlying correlation is not known, and as a matter of fact,
3.3 Empirical Analysis

neither is the volatility. We are going to use a method introduced by Andersen, Bollerslev (1998), namely realized correlation. We use high-frequency datasets for the construction of correlation and volatility measures.

Let $H_{rea,i}$ denote the realized correlation for month $i$, and let us assume that for month $i$, we have $K$ demeaned returns $r_{j,i}$, $j = 1, \ldots, K$. In our case, $K$ stands for the number of daily returns available for month $i$, and $r_{j,i}$ is a two dimensional vector consisting of the returns of two assets at day $j$ in month $i$. Then:

$$H_{rea,i} = \sum_{j=1}^{K} r_{j,i} r_{j,i}^\top$$  \hspace{1cm} (3.16)

The diagonal elements of $H_{rea,i}$ are an estimate of the volatility of the assets in month $i$, whereas the off-diagonal elements can be used to get estimates for the correlation.

In theory, when the observation frequency increases, the estimate should converge to the real values, as was again pointed out by Andersen, Bollerslev (1998). Obviously, due to a lack of data and properties of the market like bid-ask-spreads, nonsynchronous trading effects, this convergence can not be achieved. Still, Andersen, Bollerslev (1998) show that the measures perform well if the frequency is high enough.

It is in this context quite interesting to observe, that a long calendar span of data is not required. As long as the returns are sampled frequently enough, volatility can be estimated as well as required from an arbitrarily short span of data. The situation is completely different when one wants to estimate the drift, where a long calendar span is required, regardless of the sampling frequency.

In our setting, we have daily returns for the three series S&P500, Bonds and the exchange rate USD/YEN. As mentioned earlier, from these daily returns, we compute monthly returns, as well as monthly realized correlations and volatilities. These will in the next sections be used to compare the different methods to capture conditional correlation.

First we will first focus on simulated data.

3.3.2 Multivariate NoVaS on simulated Data

We will in the following consider two types of simulated data, in order to see the functionality of our multivariate NoVaS transformation.

The first one is going to be a simple bivariate model, that we call data generating process (DGP), as introduced by Patton, Sheppard (2008). Next, we will assume two univariate GARCH models, and specify a deterministic time varying correlation between them.
Let $R_t = (X_{1,t}, X_{2,t})^T$ be our vector of returns. The DGP is then defined as:

$$R_t = \Sigma_t^{1/2} \varepsilon_t$$

$$\varepsilon_t = \sum_{k=1}^{78} \xi_{kt} \quad \text{where} \quad \xi_{kt} \sim \mathcal{N}(0, 78^{-1}) \quad \text{(3.17)}$$

$$\Sigma_t = 0.05\bar{\Sigma} + 0.90\Sigma_{t-1} + 0.05R_{t-1}R_{t-1}^T$$

where $\bar{\Sigma}$ denotes the unconditional covariance matrix.

We simulate our return process with $n = 1000$ and you can see the plotted return series and realized correlation process in Figure 3.1.
3.3 Empirical Analysis

According to the model selection in Table 3.1, both the methods uniform target with squared returns and uniform target with absolute returns seem to result in good transformation. Depending on which objective function we look at, one seems to be better than the other, or vice versa. We have to make a choice and choose uniform target with absolute returns. Since both transformations work, this choice is not of big relevance. After transforming each univariate series individually with the chosen transformation, we want to obtain estimates for the conditional correlation. To that end we need to find the best method to choose \( \lambda \). Hence, for every described method, we calculate the lambda and the resulting conditional correlation, and compare it to the realized correlation of simulated data. The results are shown in Table 3.2, where the enumeration of the methods is as in Section 3.2. We calculate the mean squared error and the correlation between realized and estimated correlation, and present the used \( \lambda \).

<table>
<thead>
<tr>
<th></th>
<th>( X_{1,t} )</th>
<th>( X_{2,t} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal target</td>
<td>Normal target</td>
</tr>
<tr>
<td>squared returns</td>
<td>0.235882</td>
<td>5.90130e-02</td>
</tr>
<tr>
<td>absolute returns</td>
<td>1.26160e-03</td>
<td>1.26677e-03</td>
</tr>
<tr>
<td></td>
<td>0.272581</td>
<td>3.18787e-04</td>
</tr>
</tbody>
</table>

Table 3.1: Model Selection Univariate NoVaS

<table>
<thead>
<tr>
<th></th>
<th>( X_{1,t} )</th>
<th>( X_{2,t} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exc. Kurtosis</td>
<td>QQ-Correlation</td>
</tr>
<tr>
<td></td>
<td>0.252690</td>
<td>1.33402e-03</td>
</tr>
<tr>
<td></td>
<td>0.1358397</td>
<td>8.69752e-04</td>
</tr>
<tr>
<td></td>
<td>1.39809e-05</td>
<td>4.55791e-04</td>
</tr>
<tr>
<td></td>
<td>4.76079e-05</td>
<td>3.31393e-03</td>
</tr>
</tbody>
</table>

Table 3.2: Comparison of Methods to calculate lambda

All methods yield comparable results, except of the Method 6, which was the method where the only factor taken into account was the sign of the correlation. Apparently, the loss of information when using this method is too big. Furthermore, the maximum-
likelihood method performs surprisingly badly, and is outperformed by all other methods, except of Method 6. We are using Method 5, as it yields the highest correlation to the realized correlation, as one can see in Table 3.2. The Figure 3.2 shows the plots of the resulting estimated correlations.

![Comparison of different methods to estimate correlation](image)

Figure 3.2: Visualisation of performance of Method 5 compared to MLE method and the realized correlation

Multivariate normal returns

We are now going to 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, we have:

\[
R_t = (X_{1,t}, X_{2,t})' \\
R_t \sim \mathcal{N} (0, H_t) \\
H_t = \begin{pmatrix} \sigma_{1t}^2 & \rho_t \sigma_{1t} \sigma_{2t} \\ \rho_t \sigma_{1t} \sigma_{2t} & \sigma_{2t}^2 \end{pmatrix}
\]  

(3.18)
where

\[
\sigma^2_{1,t} = 0.01 + 0.05X^2_{1,t-1} + 0.94\sigma^2_{1,t-1}
\]
\[
\sigma^2_{2,t} = 0.05 + 0.20X^2_{2,t-1} + 0.50\sigma^2_{2,t-1}
\]

and

\[
\rho_t = 0.5 + 0.4 \cos(2\pi t/400)
\]

or

\[
\rho_t = \text{mod}(t/300)
\]

where we will go through both examples of \(\rho_t\), the first implying a sinusoidal correlation, the second a linearly increasing correlation. After having simulated the series, we can proceed as above, namely selection of the model, comparison of the different objective functions \(Q(\lambda)\) and plot of the resulting captured correlation. We choose a sample size of \(n = 1000\).

- **Sinusoidal correlation:**

  We first use the sinusoidal correlation, hence:

  \[
  \rho_t = 0.5 + 0.4 \cos(2\pi t/400) \quad (3.19)
  \]

![Multivariate Return Series](image_url)

**Figure 3.3:** Return Series and Correlation process of Multivariate Normal Returns with Sinusoidal Correlation
3.3 Empirical Analysis

As earlier, we start by selecting the model:

<table>
<thead>
<tr>
<th>Normal target squared returns</th>
<th>Normal target absolute returns</th>
<th>Uniform target squared returns</th>
<th>Uniform target absolute returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_{1,t} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exc. Kurtosis</td>
<td>1.93516e-02</td>
<td>1.06737e-03</td>
<td>1.97589e-05</td>
</tr>
<tr>
<td>QQ-Correlation</td>
<td>3.988263e-03</td>
<td>4.73686e-03</td>
<td>2.63363e-03</td>
</tr>
<tr>
<td>KS-pvalue</td>
<td>3.81259e-01</td>
<td>7.88708e-01</td>
<td>2.14623e-01</td>
</tr>
</tbody>
</table>

\[ X_{2,t} \]

| Exc. Kurtosis                 | 1.54634e-04                   | 1.25340e-04                    | 2.04218e-05                    | 7.75656e-06                            |
| QQ-Correlation                | 6.70468e-03                   | 6.17981e-03                    | 3.94996e-03                    | 4.48064e-03                            |
| KS-pvalue                     | 5.61847e-01                   | 4.65733e-01                    | 5.73960e-01                    | 1.05326e-01                            |

Table 3.3: Model Selection Univariate NoVaS in the case of sinusoidal correlation

Observing the above Table 3.3 one sees that with the Kurtosis and KS-pvalue methods, the optimal model uses uniform target distribution and absolute returns. With this setting, we can now again compare our different methods to compute the correlation. This time, we do not have to only rely on the measure used above, namely the Realized correlation \( H_{\text{real}} \), but we can use the actual deterministic correlation, for comparison. The calculations are summarized in the following Table 3.4:

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation with real Corr.</th>
<th>MSE to real Corr.</th>
<th>Lambda</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.8716346</td>
<td>0.02115454</td>
<td>0.9501566</td>
</tr>
<tr>
<td>Method 1</td>
<td>0.8107868</td>
<td>0.02797518</td>
<td>0.9643933</td>
</tr>
<tr>
<td>Method 2</td>
<td>0.8358809</td>
<td>0.02531119</td>
<td>0.9545646</td>
</tr>
<tr>
<td>Method 3</td>
<td>0.8054239</td>
<td>0.02936121</td>
<td>0.9693726</td>
</tr>
<tr>
<td>Method 4</td>
<td>0.8388725</td>
<td>0.02510483</td>
<td>0.9520657</td>
</tr>
<tr>
<td>Method 5</td>
<td>0.7991878</td>
<td>0.03086811</td>
<td>0.9701286</td>
</tr>
<tr>
<td>Method 6</td>
<td>0.8016514</td>
<td>0.03646713</td>
<td>0.9070909</td>
</tr>
</tbody>
</table>

Table 3.4: Comparison of Methods to calculate lambda in the case of sinusoidal correlation

In the case of multivariate normal returns with sinusoidal correlation, the maximum-likelihood outperforms the rest of the methods. The only methods that are able to compete are Methods 2 and 4, that are better than the others, but the maximum-likelihood still produces a 5% higher correlation with the realized correlation. The great performance of the maximum-likelihood method seems in this context to be dependent of the data. One can expect, that the other methods, since they are not dependent on the structure of the data (for example of the distribution)
are more stable from one dataset to the other. Figure 3.4 shows the resulting estimated correlation with Method 4, compared to the theoretical correlation, and the correlation, that one gets using the maximum-likelihood method.

Figure 3.4: Visualisation of performance of Method 1 compared to MLE method

- **Linear increasing correlation:**

  Similarly for linear increasing correlation, hence for a multivariate normal returns with a covariance matrix as in (3.18) with the following correlation process:

  \[ \rho_t = \text{mod} \left( \frac{t}{300} \right) \]  

  \[ (3.20) \]

  resulting in the following time series, and correlation process:
3.3 Empirical Analysis

Figure 3.5: Return Series and Correlation process of Multivariate Normal Returns with linearly increasing Correlation

One again, we will first want to choose the model. We therefore compute the objective functions for the different combinations of the NoVaS approach.

<table>
<thead>
<tr>
<th></th>
<th>Normal target squared returns</th>
<th>Normal target absolute returns</th>
<th>Uniform target squared returns</th>
<th>Uniform target absolute returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_{1,t}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exc. Kurtosis</td>
<td>4.06123e-01</td>
<td>2.95837e-01</td>
<td>1.14280e-05</td>
<td>1.43865e-05</td>
</tr>
<tr>
<td>QQ-Correlation</td>
<td>3.12765e-03</td>
<td>2.10812e-03</td>
<td>1.11553e-03</td>
<td>4.28217e-03</td>
</tr>
<tr>
<td>KS-pvalue</td>
<td>1.90068e-01</td>
<td>2.13984e-03</td>
<td>1.43105e-01</td>
<td>3.27692e-05</td>
</tr>
</tbody>
</table>

|               |                               |                               |                               |                               |
| $X_{2,t}$     |                               |                               |                               |                               |
| Exc. Kurtosis | 2.67428e-01                   | 3.87380e-02                   | 2.99730e-06                   | 1.91458e-05                   |
| QQ-Correlation| 2.77971e-03                   | 2.82397e-03                   | 8.37137e-04                   | 3.96208e-03                   |
| KS-pvalue     | 1.90073e-01                   | 2.13320e-03                   | 2.10849e-01                   | 1.24636e-04                   |

Table 3.5: Model Selection Univariate NoVaS in the case of linearly increasing correlation

The excess kurtosis and QQ-correlation statistics agree on the uniform target distribution with squared returns to be the combination with the best results. We
therefore once again compute the resulting conditional correlation, and compare it to the actual realized correlation, to get the results listed in Table 3.6 for the different methods introduced earlier.

<table>
<thead>
<tr>
<th>Method</th>
<th>Correlation with real Corr.</th>
<th>MSE to real Corr.</th>
<th>Lambda</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.7246316</td>
<td>0.0486802</td>
<td>0.9562257</td>
</tr>
<tr>
<td>Method 1</td>
<td>0.7323281</td>
<td>0.0437819</td>
<td>0.9586223</td>
</tr>
<tr>
<td>Method 2</td>
<td>0.7679904</td>
<td>0.0399146</td>
<td>0.9405139</td>
</tr>
<tr>
<td>Method 3</td>
<td>0.7091871</td>
<td>0.0467344</td>
<td>0.9640928</td>
</tr>
<tr>
<td>Method 4</td>
<td>0.7697121</td>
<td>0.0398651</td>
<td>0.9486638</td>
</tr>
<tr>
<td>Method 5</td>
<td>0.6514904</td>
<td>0.1057375</td>
<td>0.7717661</td>
</tr>
<tr>
<td>Method 6</td>
<td>0.6919922</td>
<td>0.0786802</td>
<td>0.9045524</td>
</tr>
</tbody>
</table>

Table 3.6: Comparison of Methods to calculate lambda in the case of linearly increasing correlation

With this time series, the maximum-likelihood method gets outperformed by most of the cross validation methods, namely Method 1, 2 and 4. In terms of the correlation with the realized correlation, Method 4 produces the most convincing result, and is therefore the method that we are going to use in the following Figure 3.6.
3.3 Empirical Analysis

Figure 3.6: Visualisation of performance of Method 4 compared to MLE method

3.3.3 Multivariate NoVaS on Real Data

We use the same series that we used in Chapter 2 hence with a S&P500 and a 10-year bond (10-year maturity constant maturity rate series) in the same time frame and frequency as before.
The model selection is analogue to the one in Section 2.4.2 as the considered time series are the same. Therefore we can use the results and are now interested in the correlation between the two processes. Let us now compare the methods introduced in Section 3.2 to estimate the optimal $\lambda$ in (3.2). The results are presented in the following Table 3.7.

<table>
<thead>
<tr>
<th></th>
<th>Correlation with real Corr.</th>
<th>MSE to real Corr.</th>
<th>Lambda</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.6633962</td>
<td>0.0912843</td>
<td>0.9785523</td>
</tr>
<tr>
<td>Method 1</td>
<td>0.6810836</td>
<td>0.0979265</td>
<td>0.9618249</td>
</tr>
<tr>
<td>Method 2</td>
<td>0.6832314</td>
<td>0.0977426</td>
<td>0.9566795</td>
</tr>
<tr>
<td>Method 3</td>
<td>0.6833019</td>
<td>0.0977446</td>
<td>0.9564425</td>
</tr>
<tr>
<td>Method 4</td>
<td>0.6834949</td>
<td>0.0977552</td>
<td>0.9557501</td>
</tr>
<tr>
<td>Method 5</td>
<td>0.6106173</td>
<td>0.1167147</td>
<td>0.9922018</td>
</tr>
<tr>
<td>Method 6</td>
<td>0.5960527</td>
<td>0.1787241</td>
<td>0.7637949</td>
</tr>
</tbody>
</table>

Table 3.7: Comparison of Methods to calculate lambda in the correlation between S&P500 and a 10-year bond.
### 3.3 Empirical Analysis

According to Table 3.7, Methods 2, 3 and 4 yield the best results, with Method 4 having slight advantages in terms of the correlation with the realized correlation of the dataset. The MLE method performs less convincingly than on the previous datasets, in comparison to the more stable cross validation models. But it is still only 3\% worse than the best performing model.

![Comparison of different methods to estimate the correlation](image.png)

**Figure 3.8**: Visualisation of performance of Method 4 compared to MLE method and realized correlation

To be able to actually judge the performance of NoVaS, we need to compare the results to the ability of benchmarking methods in capturing the conditional correlation. We will in another empirical analysis in Chapter 4, in the context of an application in portfolio analysis, see how NoVaS performs compared to other standard methods used in literature.
Chapter 4

Application to Portfolio Analysis

In this chapter, we will use the introduced method of multivariate NoVaS for portfolio analysis. On the basis of the time varying conditional correlation that we get from the technique above, we will construct an optimal portfolio in the sense of portfolio variance minimization. Basically, the goal is to calculate an optimal hedge ratio based on the captured correlation. Since the volatility of financial returns was found to be time varying, one can assume that correlation between financial returns will be time varying as well. This implies that our optimal hedge ratio will also be dynamic as it is based on the correlation.

To be able to judge the performance of NoVaS in this context, we will compare it to a range of standard methods in dealing with multivariate returns. We will consider baseline approaches like naive hedging, we will base the hedge ratio on linear models, and will finally use multivariate GARCH techniques to capture the correlation, and base the hedge ratio on that captured (time-varying) correlation.

We will lead an empirical study, in which we will first compare the different models in their ability to capture the conditional correlation. We will then calculate the time-varying hedge ratios based on the different methods, and will compare the standard deviation of the portfolio returns, in order to be able to judge the quality of the hedge ratios. Like previously, we will use different data sets, on which the empirical work is led. We will first use several simulated datasets, and will finish by using the previous real financial returns.

4.1 Optimizing the utility of a portfolio

Let us consider the easy 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 $r_{2,t} = \log(p_{2,t}/p_{2,t-1})$. The variances are $\sigma_{1,t}^2$ and $\sigma_{2,t}^2$ and the covariance between the two
4.1 Optimizing the utility of a portfolio

Let us further assume, that the portfolio consists of \( \beta_t \) units of asset 2 per unit of asset 1. The portfolio return is therefore given by \( r_{p,t} = r_{1,t} - \beta_{t-1} r_{2,t} \), where \( \beta \) is indexed by \( t-1 \), because the choice of the composition of the portfolio has to made before the return in \( t \) is known. The negative before the \( \beta \) indicates that the investor takes a short position.

The utility of the investor is a function of the portfolio return and the portfolio variance. If we assume that the investor is risk-averse, with risk aversion parameter \( \eta \), a general form of the utility function is:

\[
U \left( \mathbb{E} [r_{p,t} | \mathcal{F}_{t-1}], \sigma_{p,t}^2 \right) = \mathbb{E} [r_{p,t} | \mathcal{F}_{t-1}] - \eta \sigma_{p,t}^2 \\
\approx 0 - \eta \left( \sigma_{1,t}^2 + \beta_{t-1}^2 \sigma_{2,t}^2 - 2 \beta_{t-1} \sigma_{12,t} \right)
\]

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

If the investor is rational, he will try to maximize his utility with respect to \( \beta \):

\[
\frac{\partial}{\partial \beta_{t-1}} \left[ 0 - \eta \left( \sigma_{1,t}^2 + \beta_{t-1}^2 \sigma_{2,t}^2 - 2 \beta_{t-1} \sigma_{12,t} \right) \right] = \eta \left( -2 \beta_{t-1} \sigma_{12,t}^2 + 2 \sigma_{12,t} \right)
\]

\[
\Rightarrow \beta_{t-1} = \frac{\sigma_{12,t}}{\sigma_{2,t}^2}
\]

Therefore, a rational, risk-averse investor will at time point \( t-1 \) choose the following portfolio composition for time point \( t \): for every unit of asset 1, he will hold \( \frac{\sigma_{12,t}}{\sigma_{2,t}^2} \) units of asset 2, if he wants to maximize his utility, or minimize his variance.

As expected, the optimal hedge ratio depends on the correlation and can therefore be time varying. We are in the following chapter going to present several methods with different complexity, with the common aim to capture correlation. These methods will be used to compare them to the introduced method in Chapter 3 to evaluate the performance of multivariate NoVaS.
4.2 Standard approaches in correlation modeling

There is a lot of literature on multivariate models with constant or dynamic correlation. Starting with simple approaches with constant correlation like in the context of the Capital Asset Pricing Model (CAPM) [Sharpe (1964)], there is the GARCH framework of Bollerslev, Engle, Wooldridge (1988) and the stochastic volatility approach, introduced by Harvey, Ruiz and Sheppard (1994), just to name a few. A nice collection is given in Bos and Gould (2007) which we are partly going to use.

We will start by looking at models which imply a direct linear relationship between the returns of two assets. Later, we will move on to models where the variance process of each process is modelled separately, combined with a possibly dynamic correlation process.

4.2.1 Naive models

Let us first look at simple baseline models.

- **'Naive model’**
  
The simplest case is where we do not consider any information given by the data. We just fix the correlation at one, implying that there is a perfect co-movement between the two series. This would also imply that there is perfect hedge in the portfolio and therefore a constant $\beta_t = 1$ for all $t$.

- **Linear regression**
  
  Involving a bit more information, the linear regression approach implies that there is a time-invariant linear relation between the two assets. This is actually in line with the CAPM approach [Sharpe (1964)], which implies that the theoretical rate of return of an asset should be equal to the risk-free rate of return plus the market return multiplied by a factor $\beta$ which represents the sensitivity of that return to non-diversifiable risk, namely:

  $$\mathbb{E}[R_i] = R_f + \beta_i (\mathbb{E}[R_m] - R_f)$$

  where $\mathbb{E}[R_i]$ is the expected rate of return of a capital asset, $R_f$ is the risk-free rate of interest such as interest arising from government bonds, $\beta_i = \frac{\text{cov}(R_i, R_m)}{\text{var}(R_m)}$ and $\mathbb{E}[R_m]$ is the expected rate of return of the market.

  In our scenario, the equation of the linear regression resembles:

  $$r_{1,t} = \alpha + \beta r_{2,t} + \varepsilon_t$$

  (4.4)
4.2 Standard approaches in correlation modeling

where $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon)$. The $\beta$ in this context can be estimated by standard linear regression methods, which is what we are going to do in later applications.

It is of course unlikely that the above relationship holds for two arbitrary financial returns.

The assumptions made on the returns for the above models to work are obviously very strong and are not in accordance with the standard assumptions in literature, especially the fact that volatilities are assumed to be constant. We will see in the applications whether they are nevertheless able to lead to acceptable hedging decisions.

4.2.2 GARCH-based models

We are in the following assuming that our bivariate return process $r_t$ follows a normal distribution, hence:

$$r_t \sim \mathcal{N}(0, H_t)$$

$$H_t = \begin{pmatrix} \sigma^2_{1t} & \sigma_{12t} \\ \sigma_{12t} & \sigma^2_{2t} \end{pmatrix} = \begin{pmatrix} \rho_t \sigma_{11} \sigma_{2t} & \rho_t \sigma_{12} \sigma_{2t} \\ \rho_t \sigma_{12} \sigma_{2t} & \sigma^2_{2t} \end{pmatrix}$$

$H_t$ is going to follow a multivariate GARCH model, as introduced in Bollerslev, Engle, and Wooldridge (1988).

A natural multivariate extension of the univariate GARCH($p,q$) model is the so called vec-representation by Engle and Kroner (1995):

$$\text{vec}(H_t) = W + \sum_{i=1}^{q} A_i^* \text{vec}(r_{t-1} r_{t-1}') + \sum_{j=1}^{p} B_j^* \text{vec}(H_{t-1})$$

$$= W + A^*(L) \text{vec}(r_t r_t') + B^*(L) \text{vec}(H_t)$$

where

$$A^*(L) = A_1^* L + \cdots + A_q^* L^q$$

$$B^*(L) = B_1^* L + \cdots + B_p^* L^p$$

and $L$ can be viewed as the backward operator, and vec denotes the vector-half operator, which stacks the lower triangular elements of an $N \times N$ matrix as an $[N(N+1)/2] \times 1$ vector. For instance, let $A$ be $(2 \times 2)$, with

$$A = \begin{bmatrix} a_{11} & a_{21} \\ a_{21} & a_{22} \end{bmatrix}$$
4.2 Standard approaches in correlation modeling

Then

\[ \text{vec}(A) = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{22} \end{bmatrix} \]

Therefore we would have \( W \) being an \([N(N+1)/2] \times 1\) vector, and \( A_i^* \) and \( B_j^* \) \([N(N + 1)/2] \times [N(N + 1)/2]\) matrices.

As a simple example, we can give the Vec-GARCH(1,1) model, with \( N = 2 \):

\[
\begin{bmatrix}
  h_{11,t}^* \\
  h_{21,t}^* \\
  h_{22,t}^*
\end{bmatrix} = \begin{bmatrix}
  w_1^* \\
  w_2^* \\
  w_3^*
\end{bmatrix} + \begin{bmatrix}
  a_{11}^* & a_{12}^* & a_{13}^* \\
  a_{21}^* & a_{22}^* & a_{23}^* \\
  a_{31}^* & a_{32}^* & a_{33}^*
\end{bmatrix} \begin{bmatrix}
  r_{1,t-1}^2 \\
  r_{1,t-1}r_{2,t-1} \\
  r_{2,t-1}^2
\end{bmatrix} + \begin{bmatrix}
  b_{11}^* & b_{12}^* & b_{13}^* \\
  b_{21}^* & b_{22}^* & b_{23}^* \\
  b_{31}^* & b_{32}^* & b_{33}^*
\end{bmatrix} \begin{bmatrix}
  h_{11,t-1}^* \\
  h_{21,t-1}^* \\
  h_{22,t-1}^*
\end{bmatrix}
\]

The problems with this representation are obvious: the number of parameters is \([1 + (p+q)(N(N+1)/2)^2]\), which even for small values of \( p \) and \( q \) leads to a very large number of parameters. Let \( p = q = 1 \) and \( N = 5 \): then the vec-representation contains 465 parameters.

Even more disturbing is the fact that there is no assurance for the matrix \( H_t \) to be positive definite, which is requirement for the parametrization to be sensible. That is why literature proposes several extensions of the above model. Bollerslev, Engle, Wooldridge (1988) presented a simplified version of the vec-model, by assuming that \( A_i \) and \( B_j \) in (4.5) are diagonal matrices. Even though that reduces the amount of parameters to \((p+q+1)N(N+1)/2\), the model seems too restrictive, as it doesn’t allow for interaction between the individual conditional variances and correlations. Other extensions introduced in literature are the BEKK (Engle, Kroner (1995)) and the DCC model (Engle (2002)), which we will briefly introduce in the following, and consider in the empirical work later on.

- **BEKK model** The BEKK (Baba-Engle-Kraft-Kroner) model can be viewed as a restriction from the vec-representation, that ensures positive-definite conditional covariance matrices by construction. It is given by:

\[
H_t = CC' + \sum_{k=1}^{K} \sum_{i=1}^{q} A_{ik}r_{t-i}r_{t-i}'A'_{ik} + \sum_{k=1}^{K} \sum_{i=1}^{p} B_{ik}H_{t-i}B'_{ik}
\]

where \( C, A_{ik} \) and \( B_{ik} \) are \((N \times N)\), and \( C \) is a lower triangular matrix, which with no further assumptions implies that \( CC' \) is positive definite.

We are going to use the simple case \( K = 1 \), in which context the GARCH(1,1) model is given by:

\[
H_t = CC' + A_1r_{t-1}r_{t-1}'A_1' + B_1H_{t-1}B_1'
\]

Engle, Kroner (1995) give us the following propositions:
Proposition 4.2.1 Suppose that the diagonal elements are restricted to be positive and that $a_{11}$ and $b_{11}$ are also restricted to be positive. Then if $K = 1$, there exists no other $C$, $A_1$, $B_1$ in the Model 4.6 that will give an equivalent representation.

Proposition 4.2.2 If $H_0$, $H_{-1}$, $\ldots$, $H_{-p+1}$ are all positive definite, then the BEKK parametrization (with $K = 1$) yields a positive definite $H_t$ for all possible values of $r_t$ if $C$ is a full rank matrix are $B_1$ is a full rank matrix.

Furthermore, the model is covariance stationary if and only if the eigenvalues of

$$
\sum_{k=1}^{K} \sum_{i=1}^{q} A_{ki} \otimes A_{ki} + \sum_{k=1}^{K} \sum_{i=1}^{p} B_{ki} \otimes B_{ki},
$$

where $\otimes$ denotes the Kronecker product, are less than one in modulus.

Let us as an example give the MGARCH(1,1)-BEKK representation, in the case $N = 2$:

$$
H_t = CC' + \begin{bmatrix}
  a_{11} & a_{12} \\
  a_{21} & a_{22}
\end{bmatrix}
\begin{bmatrix}
  r_{11,t-1} & r_{12,t-1} \\
  r_{21,t-1} & r_{22,t-1}
\end{bmatrix}
\begin{bmatrix}
  a_{11} & a_{12} \\
  a_{21} & a_{22}
\end{bmatrix}'

+ \begin{bmatrix}
  b_{11} & b_{12} \\
  b_{21} & b_{22}
\end{bmatrix}
\begin{bmatrix}
  h_{11,t-1} & h_{12,t-1} \\
  h_{21,t-1} & h_{22,t-1}
\end{bmatrix}
\begin{bmatrix}
  b_{11} & b_{12} \\
  b_{21} & b_{22}
\end{bmatrix}'.
$$

Estimation of a BEKK model still involves quite heavy computations due to a the still large number of parameters: $(p+q)KN^2 + N(N+1)/2$, which in the case $p = q = K = 1$ and $N = 5$ is still 65 parameters. Furthermore, in the estimation procedure, further described in Engle and Kroner (1995) several matrix inversions are required, which are also quite costly. A further problem with the BEKK representation is the difficulty to interpret of the parameters.

• CCC model

The CCC (constant conditional correlation) model is based on a decomposition of $H_t$ into conditional standard deviations and correlations. It was introduced by Bollerslev in [Bollerslev (1990)]. The conditional correlation matrix is time-invariant, so the conditional covariance matrix can be expressed as follows:

$$
H_t = D_tPD_t
$$

where: $H_t = [h_{ij,t}]$, $D_t = \text{diag}(h_{11,t}^{1/2}, \ldots, h_{NN,t}^{1/2})$ and $P = [\rho_{ij}]$ is positive definite with $\rho_{ii} = 1$, $i = 1, \ldots, N$. This means that the off-diagonal elements are given by:

$$
[h_t]_{ij} = h_{ii,t}^{1/2}h_{jj,t}^{1/2}\rho_{ij}, \quad i \neq j, \quad 1 \leq j, i \leq N
$$
The $h_{ii,t}$ are modelled as in the univariate GARCH model, hence:

$$h_{ii,t} = \omega_i + \kappa_i r_{t-1}^2 + \lambda_i h_{ii,t-1}$$

Covariance stationarity is ensured if $\sum_{i=1}^{N} \kappa_i + \sum_{i=1}^{N} \lambda_i < 1$ (for more details see McNeil, Frey, Embrechts (2010)).

The great advantage of CCC-GARCH model is that it simplifies computational work: one can estimate the parameters in two steps: first the conditional variances are estimated using the marginal likelihoods, then $R$ is estimated using the sample estimator of standardized residuals $\hat{D}_t^{-1} r_t$. Actually, one has to invert the conditional correlation matrix only once per iteration. Furthermore, the number of parameters equals $(N(N-1)/2)$ in addition to those required for the univariate GARCH equations, which is an improvement to the earlier models.

Despite the attractive parametrization, the assumption of constant conditional correlations might be too restrictive. That is why the CCC evolved into the Dynamic Correlation Model (DCC).

**DCC model**

The DCC is basically a generalization of the CCC, introduced by Engle (2002) allowing the conditional correlation matrix to be time-varying. The model is therefore:

$$H_t = D_t P_t D_t$$

where again:

$$h_{ii,t} = \omega_i + \kappa_i r_{t-1}^2 + \lambda_i h_{ii,t-1}, \text{ and}$$

$$D_t = \text{diag}(h_{11,t}^{1/2}, \ldots, h_{NN,t}^{1/2}),$$

but now:

$$u_t = D_t^{-1} r_t$$

$$Q_t = (1 - \delta)(u_{t-1} u_{t-1}') + \delta Q_{t-1}$$

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}$$

As for the CCC, the idea is to first model the conditional variance of each individual time series $h_{ii,t}$. The difference is to then compute the standardized residuals $u_t$ from these models, and compute the conditional correlation matrix $R_t$ by an exponentially weighted moving average procedure applied to the scaled residuals $u_t$, yielding a new parameter $\delta$. The positive definiteness of $R_t$ and the fact that we have ones on the diagonal to form a proper correlation matrix are ensured by (4.7).
4.3 Comparison of the model performance

Compared to the CCC model, the advantage of numerical simplicity is lost, as the matrix inversion has to be done not once every iteration, but for every $t$ at every iteration. But we can still use the fact that the problem of estimation can be split into two lower-dimensional problems: the estimation of the parameters for the volatilities, and the estimation of the parameters for the correlation. Furthermore, the number of parameters is still quite low: we have $N(N - 1)/2 + 1$ parameters in each correlation equation.

4.3 Comparison of the model performance

We will in this section use different data sets and apply the different methods introduced so far to evaluate the correlation. We will then finally be able to compare the NoVaS method to standard literature methods. We will start by using simulated data, where we will use the same data generating processes as we did in Chapter 3. We will continue by using real data processes for the comparison, and finish by applying the theory of Section 4.1.

4.3.1 With Simulated Data

Similarly to Section 3.3.2, we will first generate our data with a data generating process and then use multivariate normal returns with a deterministic correlation process.

DGP

As before, we simulate $X_{1,t}$ and $X_{2,t}$ that follow the same dynamics as in (3.17). Choosing a sample size of 1000, we determine the conditional correlation according to our NoVaS approach, and according to methods introduced above. To be able to compare the determined correlation processes, we will as we did earlier calculate the mean squared error and the correlation of each conditional correlation process to the realized correlation process. To check for stability of the estimates, the procedure is repeated 100 times, followed by calculation of mean and variance of the resulting vectors. The results are displayed in the following Table 4.1.
4.3 Comparison of the model performance

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td>NoVaS with normal target</td>
<td>0.02488459</td>
<td>1.489240e-05</td>
</tr>
<tr>
<td>NoVaS with uniform target</td>
<td>0.02582906</td>
<td>1.721627e-05</td>
</tr>
<tr>
<td>DCC</td>
<td>0.01659117</td>
<td>1.846382e-05</td>
</tr>
<tr>
<td>BEKK</td>
<td>0.02129870</td>
<td>4.696327e-05</td>
</tr>
<tr>
<td>Linear Model</td>
<td>0.03276383</td>
<td>2.025572e-05</td>
</tr>
<tr>
<td>CCC</td>
<td>0.03208375</td>
<td>1.631397e-05</td>
</tr>
<tr>
<td>Naive</td>
<td>0.8412974</td>
<td>0.002662113</td>
</tr>
</tbody>
</table>

Table 4.1: Statistical values of the different multivariate correlation estimation techniques. "MSE" stands for "Mean Squared Error", and the MSE and Correlation are calculated with the realized correlation. The used methods are named as in Section 4.2.

There are no entries in rows CCC and Naive of correlation column. This is simply due to the fact, that since CCC and Naive result in a constant correlation process, there will not be any correlation with the realized correlation process.

At first sight, one can say that the baseline models, like the linear model or the naive model are too simple, to be able to capture the real correlation. Obviously, a constant correlation of 1 does not yield good results.

Observing the results more closely, one can see that NoVaS seems to perform competitively with the other methods in terms of the mean squared error with the realized correlation. Especially the variance of the MSE is lower, which indicates more stable results. On the other hand, both the methods with normal and uniform target distribution get outperformed by the DCC approaches in terms of correlation with the realized correlation. The NoVaS with normal target somehow performs better then the BEKK approach.

One has somehow to consider, that this simulated data set has a very GARCH like form, that one will not encounter in practice. It is therefore only normal that GARCH-based methods are doing well on this dataset. NoVaS being databased, one can assume that it is going to perform equally well on other less structured datasets, whereas GARCH-based are likely to loose performance. We will check this assumption on the next datasets, and especially on real financial returns.

Furthermore, one always has to consider that NoVaS is far less complex, and more parsimonious.

In the Figure 4.1 the plots of the different captured correlations are compared, for one arbitrary simulation of DGP process. As the NoVaS with normal target distribution had a higher correlation with the realized correlation, we choose to plot it instead of the one
4.3 Comparison of the model performance

with uniform target distribution.
4.3 Comparison of the model performance

Figure 4.1: Comparison of different methods to estimate the correlation to the realized correlation in the case of simulated returns with DGP (Note: NoVaS refers to NoVaS with normal target distribution)
Multivariate normal returns

We will now look at another type of generated process, that we already encountered in Section 3.3.2, namely multivariate normal returns. We will, as we did for the DGP simulate two return series, first with sinusoidal correlation, then with linearly increasing correlation, and try to rebuilt the correlation with the returns series, with help of the introduced techniques.

- **Sinusoidal correlation:**
  
  We first use the sinusoidal correlation, hence:

  \[ \rho_t = 0.5 + 0.4 \cos\left(\frac{2\pi t}{400}\right) \]

As before, we display in the following Table 4.2 the mean squared error and the correlation between the estimated correlation and the actual correlation \( \rho_t \). This time however, we do not need the realized correlation from the dataset, because we know the actual deterministic underlying correlation process \( \rho_t \).

We again simulate the multivariate normal returns 100 times, and perform our estimations on every iteration, in order to check for stability of the estimates. The results can be found in Table 4.2.

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td><strong>NoVaS with normal target</strong></td>
<td>0.03400907</td>
<td>7.307077e-05</td>
</tr>
<tr>
<td><strong>NoVaS with uniform target</strong></td>
<td>0.02451641</td>
<td>4.859978e-05</td>
</tr>
<tr>
<td><strong>DCC</strong></td>
<td>0.02125749</td>
<td>2.621581e-05</td>
</tr>
<tr>
<td><strong>BEKK</strong></td>
<td>0.03381179</td>
<td>0.0002743331</td>
</tr>
<tr>
<td><strong>Linear Model</strong></td>
<td>0.1282934</td>
<td>0.0005550216</td>
</tr>
<tr>
<td><strong>CCC</strong></td>
<td>0.0806377</td>
<td>2.13586e-06</td>
</tr>
<tr>
<td><strong>Naive</strong></td>
<td>0.3586047</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.2: Statistical values of the different multivariate correlation estimation techniques. "MSE" stands for "Mean Squared Error", and the MSE and Correlation are calculated with the realized correlation. The used methods are named as in Section 4.2.

As before, the entries in the table concerning correlation between CCC and Naive and the realized correlation are left empty, because the CCC and Naive approaches yield in constant correlation processes. The variance of the mean squared error of
the naive approach is 0 because both the real correlation $\rho_t$ and the assumed correlation 1 are deterministic.

The naive approach understandably yields the worst results. The CCC does not perform well either, which could have been expected as well, since it is a constant process, and is likely to have a high mean squared error to a time-varying process.

The NoVaS, both with normal and uniform target distribution performs well in comparison to the other models. The only model that can keep up with NoVaS with uniform target is the DCC, with comparable results both in mean and variance.

Furthermore, both the normal and uniform version of the NoVaS method outperform the BEKK in terms of the correlation with the realized correlation, with again the normal version yielding the best results. Especially the variance of the results of the NoVaS estimations are lower then the ones for the BEKK, which implies a higher stability and reliability of the method.

These results are more even remarkable bearing the parsimony of NoVaS and the dynamics of the underlying dataset in mind.

In Figure 4.2 we show a plot of the different correlation estimations with an arbitrary simulation of the multivariate normal returns. We again show the plot of the normal version of NoVaS, since it performed better, as one can see in the above table.
Figure 4.2: Comparison of different methods to estimate the correlation to the real correlation in the case of multivariate normal returns with a sinusoidal correlation- NoVaS stands for the NoVaS transformation with normal target distribution
4.3 Comparison of the model performance

- **Linear increasing correlation:**

Let us now apply the same procedure for multivariate normal returns, as in Section 3.3.2, this time for linear increasing correlation, hence with:

\[ \rho_t = \text{mod} \left( \frac{t}{300} \right) \]

Once again, we will use the different introduced methods to capture the correlation of the process, and we will compare the results by comparing with help of the actual correlation process \( \rho_t \). We repeat this procedure 100 times, in order to check for the stability of our estimates. Table 4.3 summarizes the results, giving in the first column the mean squared error between captured correlation and \( \rho_t \), and in the second column the correlation between those processes.

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE Mean</th>
<th>MSE Variance</th>
<th>Correlation Mean</th>
<th>Correlation Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoVaS with normal target</td>
<td>0.0649021</td>
<td>1.000489e-04</td>
<td>0.548453</td>
<td>0.009119217</td>
</tr>
<tr>
<td>NoVaS with uniform target</td>
<td>0.0529214</td>
<td>8.64628e-05</td>
<td>0.654991</td>
<td>0.005014432</td>
</tr>
<tr>
<td>DCC</td>
<td>0.0434841</td>
<td>2.457743e-05</td>
<td>0.707309</td>
<td>0.001254434</td>
</tr>
<tr>
<td>BEKK</td>
<td>0.0515732</td>
<td>2.568918e-04</td>
<td>0.606751</td>
<td>0.03946373</td>
</tr>
<tr>
<td>Linear Model</td>
<td>0.1546325</td>
<td>5.57735e-04</td>
<td>-0.005179</td>
<td>0.01716009</td>
</tr>
<tr>
<td>CCC</td>
<td>0.0837661</td>
<td>1.195194e-06</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Naive</td>
<td>0.3087542</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4.3: Statistical values of the different multivariate correlation estimation techniques. "MSE" stands for "Mean Squared Error", and the MSE and Correlation are calculated with the realized correlation. The used methods are named as in Section 4.2.

As usual, the results of the naive approach are not very convincing. The surprising part is that the good results of the NoVaS on multivariate normal returns with sinusoidal correlation is not recognisable anymore, even though the only thing that changed was the shape of the correlation. A possible explanation might be that NoVaS is not very good at handling small correlation, as the minimum of the sinusoidal correlation was 0.1, whereas in this case of linearly increasing correlation, it is 0. Another reason might be that the model runs into trouble when dealing with jumps in the correlation.

The uniform NoVaS still performs a lot better than the BEKK method when looking at the correlation with the realized correlation. But in this setting, the DCC method clearly outperforms both the normal and the uniform version of NoVaS.
Furthermore, both the BEKK and the DCC seem to yield more stable results, when looking at the variance of the correlation.
Figure 4.3: Comparison of different methods to estimate the correlation to the real correlation in the case of multivariate normal returns with a linearly increasing correlation. NoVaS stands for the NoVaS transformation with uniform target distribution.
4.3 Comparison of the model performance

4.3.2 With Real Data

Similarly with real data, we are once again observing the correlation between S&P500 and 10 year bonds, where the dataset remains the same since the last empirical study, hence monthly financial returns, from 1971 to 2010. Applying the same procedure as with simulated data, we obtain the following results, presented in Table 4.4 and Figure 4.4.

Unlike the previous studies, we are not able to check for the stability of the methods by repeating the operation several times, as we are this time not working on simulated data, but on a real dataset, which we can not gradually change.

<table>
<thead>
<tr>
<th>Method</th>
<th>MSE</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoVaS with normal target</td>
<td>0.1021672</td>
<td>0.6718035</td>
</tr>
<tr>
<td>NoVaS with uniform target</td>
<td>0.0977426</td>
<td>0.6832314</td>
</tr>
<tr>
<td>DCC</td>
<td>0.1170555</td>
<td>0.5772323</td>
</tr>
<tr>
<td>BEKK</td>
<td>0.1287656</td>
<td>0.5082495</td>
</tr>
<tr>
<td>Linear Model</td>
<td>0.1772967</td>
<td>0.1440715</td>
</tr>
<tr>
<td>CCC</td>
<td>0.1736798</td>
<td>-</td>
</tr>
<tr>
<td>Naive</td>
<td>1.521307</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 4.4: Statistical values of the different multivariate correlation estimation techniques in the case of Correlation between SP&500 and Bonds returns. "MSE" stands for "Mean Squared Error", and the methods are named as in Section 4.2.

As we have already seen in Section 3.3, the NoVaS performs rather well on this dataset. We saw especially, that the uniform NoVaS performs better than the normal NoVaS, a result, that we get proof of again here.

The naive approach is as usual not very convincing. A rather surprising result is that the DCC and the BEKK, which performed well on simulated GARCH-like data, are now outperformed by the NoVaS method by more then 20%.

One can clearly see, that NoVaS, as a model free method, does not depend on the characteristics of the dataset as much as model based methods do. The GARCH based methods like DCC and BEKK performed very well on a simulated dataset but perform far less convincingly once the desired shape of the distribution is lost. The NoVaS seems to be far less dependent on the dataset.

In Figure 4.4 one can once again see a plot of the resulting extracted correlations, where we choose to plot the uniform NoVaS correlation because it yields the higher correlation with the realized correlation.
Figure 4.4: Comparison of different methods to estimate the correlation to the real correlation
4.3 Comparison of the model performance

4.3.3 Optimal hedge ratio

As discussed in Section 4.1, we will in the following try to optimize the utility of a portfolio, which is equivalent, as we saw, to trying to find the optimal hedge ratio. We are considering the simple case of portfolio consisting of only two assets, because we are primarily focussing on the comparison between the different models that we can use to find the optimal hedge ratio.

In accordance with the theory discussed in Section 4.1 we will report the standard deviation of the portfolio returns. The portfolio that we consider is constructed by choosing the time-varying hedge ratio, as in (4.3). As we saw, the lower the standard deviation, the higher the utility for a risk-averse investor, and therefore the better the method in capturing the correlation.

We will in the following first consider simulated data, where the data generating processes are the same as in previous empirical studies. Secondly, we will as usual consider financial returns, where the datasets used consist like previously out of monthly returns from the S&P500, 10-year Bonds and USD/YEN Exchange rate.

We are in the study first using the different methods introduced in Section 4.2 to get the correlations captured by the models. With that, we are calculating the optimal hedge ratio as determined in (4.3):

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

Since

$$\sigma_{12,t} = \rho_{12,t}\sigma_{1,t}\sigma_{2,t},$$

this yields:

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

We are then calculating the portfolio returns, consisting out of 1 unit of Asset 1 and $-\beta_{t-1}$ units of Asset 2 at time point $t$. Finally, we are calculating the standard deviation of these portfolio returns.

Let us look at the results, where as we said, we first apply the above method to simulated multivariate returns, and then to real financial returns.
4.3 Comparison of the model performance

- Simulated Data

In the Table 4.5 we have listed the standard deviations of the portfolio returns, in the different columns, we used different data generating processes, and in the rows, different methods to calculate the optimal hedge ratio. We again run the simulation 100 times, in order to check the stability of the calculations. We therefore show the mean and the variance of the resulting standard deviations.

<table>
<thead>
<tr>
<th></th>
<th>Sinusoidal</th>
<th>Linearly Increasing</th>
<th>DGP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Mean</td>
</tr>
<tr>
<td>NoVaS w. norm.</td>
<td>0.8163172</td>
<td>0.01853028</td>
<td>0.8540315</td>
</tr>
<tr>
<td>NoVaS w. unif.</td>
<td>0.8183481</td>
<td>0.01973081</td>
<td>0.8681584</td>
</tr>
<tr>
<td>DCC</td>
<td>0.8016871</td>
<td>0.01248542</td>
<td>0.8046404</td>
</tr>
<tr>
<td>BEKK</td>
<td>0.8356860</td>
<td>0.01440624</td>
<td>0.8291800</td>
</tr>
<tr>
<td>Linear Model</td>
<td>0.8583696</td>
<td>0.01621564</td>
<td>0.8567994</td>
</tr>
<tr>
<td>CCC</td>
<td>0.8634421</td>
<td>0.01555058</td>
<td>0.8507062</td>
</tr>
<tr>
<td>Naive</td>
<td>1.017659</td>
<td>0.02016047</td>
<td>0.9908466</td>
</tr>
</tbody>
</table>

Table 4.5: Portfolio Standard Deviation of a portfolio consisting of two returns modelled according to the specified method

In Table 4.5 Sinusoidal refers to multivariate normal returns where the correlation between the returns is defined as in (3.19). Linearly Increasing refers to multivariate normal returns where the correlation between the returns is defined as in (3.20). DGP refers to the data generating process defined in (3.17). We are as usual using 1000 iterations of the simulating algorithm, using the first 100 as a burn-in.

One can observe that as expected, a naive hedge of 1 to 1 is not a good alternative, as it results in all cases in a risk from 30 to 50 % higher than in the best case. Furthermore, one can see that NoVaS, especially with a normal target, performs better for sinusoidal then for linearly increasing correlation. This might be an indicator that NoVaS can run into trouble when estimating low correlations, or that it has problems when handling jumps. Whether one uses a normal or a uniform target distribution does not seem to have a big effect on the resulting standard deviations.

One can say that NoVaS is again competitive with the more complex GARCH methods. In two out of three cases, it performs only slightly worse, in the order of 2 %. It is only in the case of normal returns with linearly increasing correlation, that NoVaS gets outperformed by more than 7%. But this is the dataset, on which NoVaS had already previously not shown satisfying results. NoVaS without doubt beats the
baseline methods.
One should again acknowledge that generated data should be favourable to the GARCH approaches, as it is structurally very close to GARCH processes, which amplifies the positive results of NoVaS.

- **Real Data**

We will now apply the same methodology to real financial returns. We will once again use the dataset consisting of monthly returns from the S&P500, 10-year bonds and the exchange rate between USD and YEN, from 1971 to 2010. The monthly returns are gained from daily returns, with help of which we constructed measures for the realized volatility and correlation.

In Table 4.6, we present the standard deviation of a portfolio consisting of the three possible combinations of the three above returns, represented in the columns. In the rows, we have again the different methods to capture the correlation.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P500 and Bonds</th>
<th>S&amp;P500 and USD/YEN</th>
<th>Bonds and USD/YEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoVaS with normal target</td>
<td>0.04399740</td>
<td>0.04544646</td>
<td>0.05124284</td>
</tr>
<tr>
<td>NoVaS with uniform target</td>
<td>0.04635919</td>
<td>0.04532445</td>
<td>0.05196564</td>
</tr>
<tr>
<td>DCC</td>
<td>0.04466355</td>
<td><strong>0.04420814</strong></td>
<td>0.04808003</td>
</tr>
<tr>
<td>BEKK</td>
<td>0.04657302</td>
<td>0.04474114</td>
<td><strong>0.04716842</strong></td>
</tr>
<tr>
<td>Linear Model</td>
<td>0.04528793</td>
<td>0.04718628</td>
<td>0.05437742</td>
</tr>
<tr>
<td>CCC</td>
<td>0.04542493</td>
<td>0.04466922</td>
<td>0.07882086</td>
</tr>
<tr>
<td>Naive</td>
<td>0.07153294</td>
<td>0.06572808</td>
<td>0.06536087</td>
</tr>
</tbody>
</table>

Table 4.6: Portfolio Standard Deviation of a portfolio consisting of two real financial returns presented above

As already observed before, the naive approach, consisting of just taking as many units of ”Asset 1” as of ”Asset 2” gives the worst results. The NoVaS seem to work competitively with the other approaches in the case of S&P500 and Bonds, yielding even the best result with a normal target distribution. In the other two cases, the standard deviation of a portfolio with a hedging ration determined by a NoVaS approach are slightly worse than the other standard methods. This might again be due to the fact, that the correlation between S&P500 and USD/YEN or Bonds and USD/YEN is rather low: we have already seen before, that NoVaS seems to have problems to capture low correlation. In a setting usually used in practice, where one
hedges a financial asset with for example a future or an option on the same asset, this problem of low correlation should not occur. Hedging with a financial asset with low correlation does not seem very common. Once again it seems like the DCC approach is producing the most stable results, even though it is sometimes outperformed by NoVaS.
Chapter 5

Application to estimation of GARCH parameters

5.1 Introduction

In this chapter, we will introduce another possible application of the NoVaS theory. We will in the following use the NoVaS transformation to estimate the parameters of a GARCH(1, 1) process.

In fact, the common methods for estimation of these parameters are based on maximum likelihood methods. Depending on the amount of available information, one uses maximum-likelihood estimators, where the conditional probability density function is fully known. One can also use a semiparametric approach, where the density is estimated with a data-based procedure, or a quasi-maximum-likelihood estimator, where conditional normality is assumed, though this assumption may not be correct.

Unfortunately these estimators have weaknesses listed for example in Francq, Zakoian (2009). For instance, the QMLE estimator is not explicit, it requires the existence of the 4th order moments for the i.i.d. innovation process, and the estimator is in general inefficient. For more information on efficiency comparisons between likelihood-based estimators of GARCH processes, check González-Rivera, Drost (1999). We are furthermore confident that our NoVaS based approach is going to provide a higher numerical stability then the usual likelihood-based approach.

In the following, we will introduce the theory behind our approach, and later implement it and compare it to standard likelihood-based estimators.
5.2 Theory

Remarks on the GARCH(1,1) process

As was already mentioned earlier, the GARCH(1,1) process is defined in the following way:

\[ \begin{align*}
X_t &= \sqrt{\sigma_t^2} \epsilon_t \\
\sigma_t^2 &= C + AX_t^2 - 1 + B\sigma_{t-1}^2
\end{align*} \tag{5.1} \]

where \((\epsilon_t)\) is a sequence of i.i.d. random variables, and \(A\), \(B\) and \(C\) are non-negative constants. By iterating (5.1), one gets:

\[ \begin{align*}
\sigma_t^2 &= C + AX_t^2 - 1 + B\sigma_{t-1}^2 \\
&= C + AX_t^2 - 1 + B(C + AX_{t-1}^2 + B\sigma_{t-2}^2) \\
&= C(1 + B) + A(X_{t-1}^2 + BX_{t-2}^2 + B^2\sigma_{t-2}^2) \\
&= C(1 + B + B^2) + A\left(X_{t-1}^2 + BX_{t-2}^2 + B^2X_{t-3}\right) + B^3\sigma_{t-3}^2 \\
&\vdots \\
&= C \sum_{i=0}^{n-1} B^i + A \sum_{i=1}^{n} B^{i-1}X_{t-i} + B^n\sigma_{t-n}^2 \\
\xrightarrow{n \to \infty} \frac{C}{1 - B} + A \sum_{i=1}^{\infty} B^{i-1}X_{t-i}^2 \tag{5.2}
\end{align*} \]

Looking at the definition of an ARCH(\(q\)) process:

\[ \begin{align*}
X_t &= \sqrt{\sigma_t^2} \epsilon_t \\
\sigma_t^2 &= C + \sum_{i=1}^{q} A_iX_{t-i}^2 \tag{5.3}
\end{align*} \]

with again the coefficients \(A_i\) and \(C\) non-negative constants, one sees that equations (5.2) and (5.3) are the same, with \(q = \infty\) and \(A_i\) and \(C\) chosen appropriately. Hence every GARCH(1, 1) can be written as a ARCH(\(\infty\)) process.

Comparison to NoVaS equation

Let us now have a look at the equation of our NoVaS transformation. Looking back at
5.2 Theory

equation (2.7), hence at:

\[ X_t = \left( \alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i X_{t-i}^2 \right)^{1/2} \frac{W_t}{\sqrt{1 - a_0 W_t^2}} \]

one can see that it is of the form:

\[ X_t = \left( \alpha s_{t-1}^2 + \sum_{i=1}^{p} a_i X_{t-i}^2 \right)^{1/2} Z_t \tag{5.4} \]

where \( Z_t \) has a heavy tailed distribution. This is with the right choice of parameters \( a_i \) (larger than 0) an ARCH(\( p \)) process. As numeric implementation shows, the \( p \), hence the order of the NoVaS transformation, is usually quite large (of the order 20). As the order of the parameters decays exponentially with \( p \) in a normal ARCH processes, we do not loose much information when treating our NoVaS process as an ARCH(\( \infty \)) process. As every ARCH(\( \infty \)) can be interpreted as a GARCH(1,1), we will in the next step try to use the NoVaS transformation to get estimates for the parameters of a GARCH(1,1) process. In other words, we will try to fit a GARCH(1,1) by using our NoVaS theory.

**Remark 5.2.1** Another important point is the following: when defining our NoVaS transformation, we were usually choosing the weights according to the so called exponential NoVaS, as defined in (2.10), hence:

\[
\begin{align*}
    a_j &= \begin{cases}
        1/\sum_{j=0}^{p} \exp(-bj) & j = 0 \\
        a_0 \exp(-bj) & j = 1, 2, \ldots, p
    \end{cases} \\
    \alpha &= 0.
\end{align*}
\]

and \( \alpha = 0 \). This leads to the following model:

\[ X_t = \left( \sum_{i=1}^{p} a_i X_{t-i}^2 \right)^{1/2} Z_t \]

This is in our situation not ideal, because it would lead to a ARCH(\( \infty \)) with no constant term. If the coefficients \( a_i \) are chosen correctly, in the sense that the defined process is weakly stationary, the process would converge to 0. This is something we want to avoid. We therefore need to refine the algorithm, and find a "good" \( \alpha \neq 0 \).

In his paper Politis (2007), Politis introduces an algorithm that optimizes NoVaS for volatility prediction, that he calls General Exponential NoVaS. Basically, the search for optimal parameters is conducted not only in terms of distributional matching, but a new optimisation goal is introduced: we want to find the optimal model for volatility prediction. The search is performed over a grid of possible values of \( \alpha \), namely
\(\alpha_1, \ldots, \alpha_K\). For each \(\alpha_k, 1 \leq k \leq K\), we perform an exponential NoVaS transformation, and name the resulting model with \(\alpha_k W^GE_k\). The algorithm can be written as follows:

**Algorithm for General Exponential NoVaS**

1. For \(k = 1, \ldots, K\), perform the following steps:
   
   (a) Let \(p\) take a very high starting value, e.g. \(p = n/5\).
   
   (b) Let \(\alpha = \alpha_k\) and \(a_i = c' \exp(-ci)\), for all \(0 \leq i \leq p\), where \(c' = (1 - \alpha_k)/(\sum_{i=0}^{p} \exp(-ci))\).
   
   (c) Pick \(c\) such that \(D_{n,k}(\theta)\) is minimized.
   
   The same problem as above with the truncation parameter might occur, in which case one adds:
   
   (d) If \(c\) as found is such that \(a_0 > 1/9\), decrease \(c\) until \(a_0 \leq 1/9\).
   
   (e) Finally, the value of \(p\) has to be trimmed: One can simply discard all coefficients \(a_i\) that fall below a certain threshold. A threshold of 0.01 seems to be reasonable, and works good in practice. Thus: if \(a_i < 0.01\), for all \(i > i_0\), let \(p = i_0\) and renormalize the \(a_i\) so that \(\sum_{i=0}^{p} a_i = 1\).

2. Compare the models \(\{W^GE_k, k = 1, \ldots, K\}\) in terms of their ability to predict volatility, and pick the model with optimal such performance.

In Step 2, we use the following predictor for the volatility, introduced in Politis (2007):

\[
\hat{X}_{n+1}^2 \overset{def}{=} \text{Median} \left[ \left( \sqrt{\alpha s_n^2 + \sum_{i=1}^{p} a_i X_{n+1-i}^2} \frac{W_{n+1}}{\sqrt{1 - a_0 W_{n+1}^2}} \right)^2 \right] \mathcal{F}_n
\]

where \(\mathcal{F}_n\) represents the observed past, \(W_n, a_i\) and \(s_n^2\) are defined as in the NoVaS context. Hence with the algorithm for General Exponential NoVaS, we find an optimal parametrization with \(\alpha \neq 0\). This is the parametrization that we are using in the estimation of the GARCH\((1, 1)\) parameters.

### Equating coefficients

Comparing equations (5.2) and (5.4), we can easily by equating the coefficients get the
5.2 Theory

following estimates:

\[
\frac{C}{1 - B} = \alpha s_{t-1}^2 \\
A = a_1 \\
A \cdot B = a_2
\]

In the standard exponential NoVaS, the parameters \(a_i\) are chosen according to equation (2.10), as discussed earlier. Therefore, our estimates for the parameters of our GARCH(1,1) process are:

\[
\hat{A} = a_0 \exp(-b) = \frac{\exp(-b)}{\sum_{j=0}^{p} \exp(-bj)} \\
\hat{B} = a_2 \frac{a_2}{a_1} = a_0 \exp(-2b) = \exp(-b) \\
\hat{C} = \alpha s_{t-1}^2(1 - \hat{B}) = \alpha s_{t-1}^2(1 - \exp(-b))
\]

where \(s_{t-1}^2\) is the unconditional variance of the data up to time point \(t - 1\).

Without looking at the results, one can already foresee that the performance might not be satisfying: we are using two parameters, namely \(b\) and \(p\) to get estimates for three parameters, \(A\), \(B\) and \(C\). This might easily have consequences on the results. We are therefore going to change the NoVaS algorithm, in order to make it less parsimonious: instead of defining \(a_0\) with \(b\) and \(p\), we are going to choose it independently of them, and therefore introduce one more step in the optimization procedure in the NoVaS transformation. In fact, we are going to choose \(a_0\) between 0 and 1 such that the transformed distribution of the data is closest to our target distribution. The algorithm to choose the other parameters remains untouched. To simplify things, we are going to optimize over a grid of values for \(a_0\), and for every such \(a_0\) calculate the distance of the distribution of our transformed data to the target distribution.

This algorithm can surely be optimized, and will in this form be very inefficient, as it will multiply the operations needed by the number of grid points used in the optimisation of \(a_0\). But we will in a first step not be interested in efficiency, but rather in effectiveness.

**Remark 5.2.2 (Target distribution of the NoVaS transformation)** As we have seen in the empirical studies of the previous chapters, often the uniform distribution turned out to be the better choice as a target distribution. We will somehow in this context focus on a normal target distribution. For the simple reason that the distribution of the innovation term \(Z_t\) in (5.4) of the implicit model:

\[
\frac{W_t}{\sqrt{1 - a_0 W_t^2}}
\]
has nice properties when the target distribution, hence the distribution of \( W_t \), is chosen to be normal. We have already discussed these properties in Remark 2.3.1. So we are taking into account the risk of achieving a worse fit, in order to have an appropriate distribution of the innovation term.

We can therefore for a given data set after a NoVaS transformation fit a GARCH(1, 1) process to that same data set. Let us now look at the results.

### 5.3 Results

To be able to judge the performance of the introduced theory, we will now first of all simulate a GARCH(1, 1) process, with given parameters, and check how well the estimates are. We will in a next step use well-known likelihood based algorithms to be able to compare the results. To be able to capture the distribution of the estimates, we will run the estimation 1000 times. That way, one can judge the variance of the estimates. We will use the NoVaS with two and with three open parameters.

The simulated GARCH(1, 1) process we use is the following:

\[
X_t = \sqrt{\sigma_t^2} \epsilon_t \\
\sigma_t^2 = 2.29 \cdot 10^{-6} + 0.062 \cdot \sigma_{t-1}^2 + 0.898 \cdot X_{t-1}^2
\]

where \( \epsilon \) follows a normal \( \mathcal{N}(0, 1) \) distribution and the parameters are chosen according to a fit of an GARCH(1, 1) process to the Yen/Dollar exchange rate with normal \( \mathcal{N}(0, 1) \) innovations.
5.3 Results

NoVaS with two open parameters

To give an impression on the choice of $\alpha$, we will start by showing the results of the General Exponential NoVaS algorithm, more precisely we will show how the forecasting performance changes with the $\alpha$, as it is done in Figure 5.2.
5.3 Results

As one can see, the graph in Figure 5.2 has a clear minimum at about 0.35. This is the value that we are going to use in our later estimations.

The following Graph 5.3 shows the distribution of the estimators after 1000 iterations of the estimation algorithm, with a new simulation of the GARCH process in every iteration.
5.3 Results

Keeping the real values of the parameters of the GARCH process in mind, namely $C = 2.29 \cdot 10^{-6}$, $A = 0.062$ and $B = 0.898$, we can see that the estimates are not far off. To be able to judge the performance, we will later compute likelihood-based estimates. We will start by checking if we achieve improvement, by changing the NoVaS algorithm and introducing more parameters.
NoVaS with three open parameters

As described above, the estimation of three GARCH parameters might require more than the usual two NoVaS parameters. We have therefore introduced a changed form of the choice of weights, where the $a_0$ is not determined by $p$ and $b$, but is chosen independently of them, as described before. The results of the estimation of the GARCH(1, 1) are given in Figure 5.4.
5.3 Results

As one can see the results are much worse than they were with the NoVaS with two open parameters. After looking into the calculations, the reasons are quickly apparent. When looking at the order of the NoVaS transformation, hence the number of weights used, or more precisely the $p$, one sees that it is much smaller than in the case of NoVaS with two

Figure 5.4: Distribution of the NoVaS with three open parameters estimates of the parameters of a GARCH(1, 1)
open parameters. It is on average at 3, which is not enough to consider it as a ARCH(∞) process. The equating coefficients step of the algorithm does therefore not make sense, and consistently, the results are not good.

The reason for $p$ being so small might simply be that $a_0$ gives a good enough approximation already, since it is chosen independently of the rest of the weights. There is therefore less need than before for a big $p$, because we require less information from past $X_{t-i}$ values.

**Comparison to likelihood-based estimates**

For the same GARCH(1,1) process, with the same parameters, we now calculate the likelihood-based estimates, again 1000 times, in order to be able to visualize the distribution of the estimates. We again plot the histograms, in Figure 5.5.
5.3 Results

Figure 5.5: Distribution of the likelihood-based estimates of the parameters of a \text{GARCH}(1,1)

We will in the following Table 5.1 compare statistical values from the estimated parameters, one the one hand from NoVaS with two open parameters, and on the other the classical likelihood-based estimation.

We will not consider NoVaS with three open parameters, as one can see in the figures,
that it will not be able to compete with the two other methods.

<table>
<thead>
<tr>
<th></th>
<th>NoVaS</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Mean of Est.</td>
<td>7.3816E-01</td>
<td>1.1173E-01</td>
</tr>
<tr>
<td>Dev. f. Mean</td>
<td>1.5984e-01</td>
<td>4.9730e-02</td>
</tr>
<tr>
<td>Variance</td>
<td>3.8875E-04</td>
<td>1.3738E-02</td>
</tr>
<tr>
<td>MAD</td>
<td>1.6295E-02</td>
<td>8.2816E-02</td>
</tr>
<tr>
<td>Coef. of Vari.</td>
<td>0.026711</td>
<td>1.049040</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-7.8230E-01</td>
<td>3.1455E+00</td>
</tr>
</tbody>
</table>

Table 5.1: Comparison of the statistical values between the NoVaS and likelihood estimates. In bold the better value between the two. "Dev. f. Mean" stands for "Deviation from mean" (|mean-real|). "MAD" stands for "Mean Absolute Deviation", "Coef. of Var." for "Coefficient of Variation". All other values have their usual meaning.

As one can see in Table 5.1, the NoVaS parameters are at least competitive with the likelihood estimates. Depending on the parameter, the mean of the estimated values seems to be slightly closer to the real values for the likelihood estimates, even though for the parameter C, this is not the case.

Similarly, the variances of the NoVaS estimates are not far away from the variances of the likelihood variances. In the case of the parameter A, the variance is even lower for NoVaS estimates.

The most interesting results are achieved when looking at the tails of the distribution of the estimates. As anticipated, the number of outliers seems to be higher when using the likelihood estimates, as we can tell by observing the kurtosis, and the mean absolute deviation.

We have to furthermore say, that once again the NoVaS is completely model independent. Obviously both the methods would not make sense in a setting, where the data is not featuring the properties of a GARCH process. But one could expect the likelihood method to work particularly well in a setting where the data is actually generated by a GARCH process, because the assumed distribution by the likelihood method fits the data perfectly.
Chapter 6

Conclusion

In this thesis, we presented an approach for modeling time series and especially conditional correlations based on the NoVaS transformation. The approach differs from related literature because no assumption on the distribution of the returns are necessary, and because there is no need to assume a model for the volatility and correlation. Furthermore, the numerical implementation is very easy and only requires one-dimensional optimization in the unit interval.

We modified the method from Politis and Thomakos (2011) in order to improve the numerical results. Due to the changes, the estimation of the conditional correlation is now in line with model-free mindset in which the NoVaS transformation in a univariate setting is.

We led an empirical analysis to study the performance of our approach. The results were satisfying: on simulated data, as long as the correlation was not too low or had jumps, the NoVaS approach performed competitively with related methods from the literature. On real datasets, NoVaS even managed to outperform the commonly used methods.

We then introduced a possible application of the NoVaS transformation in portfolio analysis: we used the captured correlation for the calculation of minimum variance hedge ratios between pairs of assets. Again comparing the performance of NoVaS with other industry standard methods, the results again led to believe that NoVaS works better on datasets with high correlation. The results are still competitive, even though NoVaS got slightly outperformed on simulated datasets.

We finally introduced a novel approach to estimate a GARCH(1, 1) model with help of the NoVaS transformation. Using the relation between ARCH(∞) and GARCH(1, 1), and the similarity between the implied model of the NoVaS transformation and the ARCH(∞), we were able to achieve satisfying results compared to the standard likelihood-based methods, especially in terms of the stability of the estimates.
## List of Figures

2.1 The daily returns of the S&P500 from 1970 to 2010 ........................................... 5
2.2 Simulated GARCH(1,1)-process, its volatility, Empirical distribution, QQ-Plot with respect to normal distribution ................................................................. 17
2.3 Simulated GARCH(1,1)-process, its volatility, Empirical distribution, QQ-Plot with respect to normal distribution ................................................................. 19
2.4 The autocorrelation plots of the three financial return series ................................. 21
2.5 Returns and volatility of our three financial return series .................................... 22
2.6 QQ-Plot to check for normal or uniform distribution of our three financial return series ................................................................. 23
2.7 Transformed S&P500 return series ..................................................................... 25
2.8 Transformed Bonds return series ...................................................................... 26
2.9 Transformed USD/YEN return series .................................................................. 27
3.1 Return Series of the simulated DGP and the realized correlation ......................... 38
3.2 Visualisation of performance of Method 5 compared to MLE method and the realized correlation ................................................................. 40
3.3 Return Series and Correlation process of Multivariate Normal Returns with Sinusoidal Correlation ................................................................. 41
3.4 Visualisation of performance of Method 1 compared to MLE method ................. 43
3.5 Return Series and Correlation process of Multivariate Normal Returns with linearly increasing Correlation ................................................................. 44
3.6 Visualisation of performance of Method 4 compared to MLE method ................. 46
3.7 Monthly return series and correlation process of S&P500 and Bonds .................. 47
3.8 Visualisation of performance of Method 4 compared to MLE method and realized correlation .................................................. 48

4.1 Comparison of different methods to estimate the correlation to the realized correlation in the case of simulated returns with DGP (Note: NoVaS refers to NoVaS with normal target distribution) .................. 59

4.2 Comparison of different methods to estimate the correlation to the real correlation in the case of multivariate normal returns with a sinusoidal correlation- NoVaS stands for the NoVaS transformation with normal target distribution .................................................. 62

4.3 Comparison of different methods to estimate the correlation to the real correlation in the case of multivariate normal returns with a linearly increasing correlation- NoVaS stands for the NoVaS transformation with uniform target distribution .................................................. 65

4.4 Comparison of different methods to estimate the correlation to the real correlation .................................................. 67

5.1 Returns and Volatility of a GARCH(1,1) with above parameters .................................................. 78

5.2 Visualisation of the Results of Generalized NoVaS Algorithm .................................................. 79

5.3 Distribution of the NoVaS-based estimates of the parameters of a GARCH(1,1) .................................................. 80

5.4 Distribution of the NoVaS with three open parameters estimates of the parameters of a GARCH(1,1) .................................................. 82

5.5 Distribution of the likelihood-based estimates of the parameters of a GARCH(1,1) .................................................. 84
Bibliography

Andersen, T.; Bollerslev, T.; Diebold, F.; Labys, P., 1999:
(Understanding, Optimizing, Using, Forecasting) Realized Volatility and Correlation Risk, March 2000, pp. 105-108

Andersen, T.; Bollerslev, T., 1998
Answering the skeptics: Yes, Standard Volatility Models do provide accurate forecasts

Bachelier, L., 1900:
Theorie de la speculation
Annales Scientifiques de l’Ecole Normale Superieure 3 (17): pp. 21-86

Bollerslev, T., 1990:
Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model

Bollerslev, T., 1992:
ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence
Journal of Econometrics, Vol. 52 (1-2), pp. 5-59

A Capital Asset Pricing Model with Time-Varying Covariances
Journal of Political Economy, Vol. 96, pp. 116-131

Bos, C.S., Gould, P., 2007:
Dynamic Correlations and Optimal Hedge Ratios
Tinbergen Institute Discussion Paper, 2007
Craig, C., 1936:

On the frequency of xy


Engle, R., 1982:

Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation


Engle, R., 2002:

Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models


Engle, R.; Kroner, K., 1995

Multivariate simultaneous generalized ARCH

Econometric Theory, Vol. 11, pp. 122-150

Fama, E., 1965:

The Behavior of Stock Market Prices

Journal of Business, January 1965, pp. 34-105

Francq, C.; Zakoïan, J.-M., 2009:

Handbook of Financial Time Series: A tour in the asymptotic theory of GARCH estimation


González-Rivera, G.; Drost, F.C., 1999:

Efficiency comparisons of maximum-likelihood based estimators in GARCH models


Mandelbrot, B., 1963:

The Variation of Certain Speculative Prices

McNeil, A.J.; Frey, R.; Embrechts, P., 2010:
Quantitative Risk Management: Konzepte, Techniken und Werkzeuge
Princeton University Press, September 2010

Nelson, D.B., 1991:
Conditional Heteroskedasticity in Asset Returns: A New Approach

Patton, A.; Sheppard, K., 2008
Evaluating Volatility and Correlation Forecasts

Politis, D.N., 2003a:
Model-Free Volatility Prediction
UCSD Dept. of Economics Discussion Paper, 2003-16

Politis, D.N., 2003b:
A Normalizing and Variance-Stabilizing Transformation for Financial Time Series
Recent Advances and Trends in Nonparametric Statistics, M.G. Akritas and D.N. Politis, (Eds.), Elsevier: North Holland, pp. 335-347

Politis, D.N., 2004:
A Heavy-Tailed Distribution for ARCH Residuals with Application to Volatility Prediction

Politis, D.N., 2007.
Model-free vs. Model-based Volatility Prediction

Politis, D.N; Tomakos, D., 2008a:
Financial Time Series and Volatility Prediction using NoVaS Transformations
Forecasting in the Presence of Parameter Uncertainty and Structural Breaks, D.E. Rapach and M.E. Wohar (Eds.), Emerald Group Publishing Ltd.

Politis, D.N; Tomakos, D., 2008b:
NoVaS Transformations: Flexible Inference for Volatility Forecasting
Politis, D.N.; Thomakos, D., 2011:
*Conditional Correlation in Financial Returns*

Rohatgi, V.K., 1976:
*An Introduction to Probability Theory & Mathematical Statistics*
Wiley, New York, June 1976

Sharpe, W.F., 1964:
*Capital Asset Prices—A Theory of Market Equilibrium Under Conditions of Risk*

Silvennoinen, A.; Teräsvirta, T., 2008:
*Multivariate GARCH models*