ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis Quantitative Finance

A cellwise estimator for the DCC-GARCH model

Name Student: Ndubuisi Igwala

Student ID number: 452107

Supervisor: dr M. Zhelonkin
Second assessor: dr D.J.C. van Dijk

Date final version: 23/04/2024

Abstract

This paper presents a robust estimator for the DCC-GARCH model, which was originally proposed by Engle (2002). Unlike previous robust methods that only handle rowwise outliers, our proposed estimator is specifically designed to handle cellwise outliers. To assess its performance, we compare it with the maximum likelihood estimation (MLE) method and an M-estimator introduced by Muler and Yohai (2008). Through simulation experiments, we demonstrate that our cellwise estimator outperforms the other estimators. It successfully estimates the parameters of the DCC-GARCH model even in the presence of outliers. In addition, an empirical study reveals the efficacy of the cellwise estimator for analyzing real-life data. This research is significant as it introduces a robust estimator for the DCC-GARCH model without requiring any modifications to the model itself.

Keywords— DCC-GARCH, M-estimator, cellwise estimator

1 Introduction

In the field of statistics, outliers are an important issue. Outliers in the data may lead to misleading specifications of models. In turn, this may lead to incorrect conclusions. Young (2022) notes that many economic papers rely their findings upon a few outliers. This finding makes it so that the conclusions drawn by these papers carry less weight than initially thought.

Different models aim to explain the behavior of outliers in data. The Tukey-Huber contamination model assumes that a substantial proportion of cases follow a certain clean distribution. Meanwhile, a significant portion of the data is contaminated and follows a different distribution. The model assumes that outliers are row-wise. This implies that an entire observation is considered an outlier. The problem of rowwise outliers in classical econometric models has been extensively covered. However, cellwise outliers have not been investigated to a great extent in the world of quantitative finance. Cellwise outliers are outliers within a multivariate observation. In other words, a certain component of an observation is treated as an outlier. An example of a paper that investigated cellwise outliers is Hubert, J., and Van den Bossche (2019). The authors investigate the impact of cellwise outliers on principal component analysis (PCA).

In the field of portfolio management, understanding the moments of assets is a crucial topic. Additionally, the correlation between these assets holds significance as it is often required for optimal portfolio allocation methods. Multivariate models, such as the Baba-Engle-Kraft-Kroner generalized autoregressive conditional heteroskedasticity (BEKK-GARCH) model introduced by Engle and Kroner (1995), are commonly used for calculating the dynamic behavior of asset correlations. However, the dynamic conditional correlation GARCH (DCC-GARCH) model is predominantly preferred in practice due to its ability to handle a large number of assets without compromising accuracy. Two papers have introduced their own versions of the DCC model, namely the DCC model by Engle (2002) and the DCC model by Tse and Tsui (2002). It is important to note that the estimation results of these models may be unreliable if they fail to handle substantial positive or negative return shocks. This study aims to investigate the robustness of widely used estimation methods for

DCC-GARCH models, as proposed by Engle and Sheppard (2001).

Considering the significance of estimating the correlation matrix for portfolio allocation, it would be worthwhile to explore the impact of return outliers on the conclusions drawn from the DCC-GARCH model. It is possible that outliers may significantly alter the optimal allocation of portfolios by influencing the mean, variance, and covariances of the assets under consideration. In cases where the DCC model's results are adversely affected by outliers, it would be advantageous to develop a robust estimator that addresses this issue.

This paper's relevance lies in the evaluation of common estimation methods for the DCC-GARCH model. If these methods are found to be non-robust, it necessitates a reevaluation of numerous findings derived from these models. Furthermore, non-robustness of the traditional estimation techniques warrants research on robust estimation methods for the DCC-GARCH model. Additionally, this paper may stimulate the search for robust estimation methods that can handle cellwise outliers in other areas of finance. Furthermore, if the known estimation methods for the DCC-GARCH model are unable to handle cellwise outliers, portfolio managers have reason to discontinue their use for correlation estimation purposes. This, in turn, prompts them to explore alternative models that yield better results in the presence of return outliers. Ultimately, this benefits taxpayers since pension funds can make more informed decisions. The risk associated with investing in a potentially risky portfolio perceived as safe decreases when appropriate conclusions are drawn from the most suitable multivariate models.

The objective of this paper is to examine whether our newly introduced estimator robust to cellwise outliers for the the DCC-GARCH model by Engle (2002) is able to produce unbiased estimates in the presence of outliers. Our analysis is twofold. First, we employ simulations of multivariate GARCH returns with significant return outliers. By comparing the estimation results with the underlying parameters, we aim to determine the robustness of the conventional estimation method. We propose a cellwise robust estimator for the DCC-GARCH model that is capable of handling large outliers. The performance of this robust estimator is assessed by comparing it with other estimation methods. Second, we conduct a study on real-life S&P500 data to invigorate our simulation results.

Our findings indicate that the cellwise outlier estimator outperforms both the MLE and

M-estimator in accurately estimating the true parameters in the presence of outliers. The cellwise estimator exhibits unbiasedness in the presence of infrequent large outliers. Moreover, the cellwise estimator is able to handle outliers in real-life data. The DCC-GARCH model making use of this estimation technique is able to produce stable estimates of the variance-covariance structure, whereas the model using traditional MLE estimation fails to do this.

Firstly, we provide a general overview of relevant literature. This is followed by a detailed explanation of all methods in the methodology section. In section 4, we present the results of our simulations. Section 5 analyzes the S&P500 data, followed by a comparison of our estimator with the two-step MLE estimator using this real-life data. Lastly, we discuss the obtained results and draw conclusions.

2 Literature Review

Since the inception of DCC-GARCH models, various estimation approaches have emerged. The MLE estimation method is commonly used to estimate the DCC-GARCH model discussed in this paper. Both the models proposed by Engle (2002) and Tse and Tsui (2002) share the same likelihood function, but differ in the computation of optimal parameters. Engle and Sheppard (2001) propose a consistent two-step estimator for Engle's DCC-GARCH model, where the likelihood function is partitioned into mean and volatility parts and a correlation part. Initially, the individual GARCH models are estimated, followed by the estimation of the DCC-GARCH model based on the optimal GARCH parameters obtained in the first step. However, this two-step estimation is inefficient for estimating the correlation parameters. To address this, a one-step iteration of the Newton-Raphson algorithm is proposed. Aielli (2013), however, disputes the claims made by Engle and Sheppard regarding the consistency of Engle's DCC-GARCH estimator. He demonstrates that the estimator can be biased in the presence of high correlation persistence and suggests a corrected version of the model to mitigate this bias.

While robust estimation of univariate GARCH models has been extensively studied, fewer studies have focused on robust estimation of multivariate GARCH models. One such study is conducted by Muler and Yohai (2008), who propose two robust estimators that outperform the quasi-maximum likelihood (QML) estimator in simulated environments. They recommend the inclusion of their estimator for real-life data. Carnero, Peña, and Ruiz (2012) make slight adjustments to the aforementioned robust estimator, as they argue that small biases in parameter estimates can lead to significant biases in volatility estimates. They propose a method to address this issue.

This paper aims to contribute to the literature on robust estimation of the DCC-GARCH model by proposing an estimator robust to cellwise outliers. Although there have been few studies on the effect of cellwise outliers on estimation in various econometric fields, Hubert et al. (2019) develop a robust PCA estimation method. Boudt, Danielsson, and Laurent (2013) also seek to find a robust DCC-GARCH model by modifying the model itself to minimize the impact of robust outliers. In contrast, our paper focuses solely on developing an estimator without altering the underlying model. Furthermore, our focus is specifically on outliers within observations.

The importance of this paper lies in the potential ramifications of neglecting cellwise outliers in the robust estimation of the DCC-GARCH model. Drawing parallels to a different field, Young (2022) finds that in instrumental variable estimation, the conclusions of many studies depend on a few outliers. This highlights the significance of robust estimation and the necessity for thorough research in this area to prevent erroneous conclusions in the future.

A potential limitation of our approach is the empirical departure from normality in returns. The use of Mahalanobis distances to detect outliers is most effective for elliptical distributions, particularly the multivariate Gaussian distribution. The assumption of multivariate Gaussianity facilitates the calculation of a cutoff value for outlier detection. Prykhodko, Prykhodko, Makarova, and Pukhalevych (2018) propose a solution by suggesting the Yeo-Johnson transformation by Yeo and Johnson (2000) to normalize non-normal data. However, this approach may not be effective when dealing with Student's t-distributed data, which already closely resembles the normal distribution. Another challenge with using Mahalanobis distances to identify outlying returns is heteroskedasticity. In larger samples, heteroskedasticity may lead to false identification of potential outliers, as the variance at that particular time point may be higher or lower than usual.

Similar to the approach taken by Rousseeuw and Van Den Bossche (2018) in detecting deviating data cells, we develop a method to identify outlying data cells. There are notable differences between our method and theirs. Our approach remains effective even when the correlation between assets is very low, although this is rarely the case in real-life data. Additionally, our estimator initially detects outlying rows before focusing on outlying cells directly.

3 Methodology

In this section, we provide a detailed explanation of the methodology employed in our analysis. Firstly, we provide a comprehensive explanation of the DCC-GARCH model proposed by Engle (2002). We utilize this model to simulate DCC-GARCH returns, to which we introduce outliers in order to compare the performance of the two-step estimation method and our robust methods. Subsequently, we introduce our estimator that is robust to cellwise contamination. Furthermore, we delve into the details of an M-estimator that we employ to assess the performance of our cellwise estimator, a method introduced by Muler and Yohai (2008).

3.1 DCC-GARCH

To begin with, we elaborate on the specifications of the DCC-GARCH model considered in this paper. Furthermore, we explain the two-step MLE estimation method with which we compare our estimator. Lastly, we provide a brief overview of how we simulate DCC-GARCH returns.

3.1.1 Model specification

The generalized autoregressive conditional heteroskedasticity (GARCH) model is introduced by Bollerslev (1986). The GARCH(1,1) model is defined as

$$h_t = \omega + \alpha v_{t-1}^2 + \beta h_{t-1}, \tag{1}$$

where h is the conditional variance, ω a constant and v is an error term. This error term is defined by

$$r_t = \mu + v_t. \tag{2}$$

Here, μ is the mean return. Hence, v_t can be considered as the demeaned return at time t. To ensure that the conditional variance is always positive ω, α and $\beta > 0$. Furthermore, to ensure a stationary process we restrict α and β such that $\alpha + \beta < 1$. For this process, weak stationarity means that the mean and variance are constant throughout the sample time. This model and its extensions are widely used in practice. The models are intuitive and are able to produce tractable forecasts. We do not consider different specifications of a GARCH(p,q) setup. Hansen and Lunde (2005) show that no GARCH specification that does not incorporate the leverage effect is significantly better at modeling volatility than GARCH(1,1). Since incorporating extensions, such as the leverage effect is outside the scope of this paper, we choose an ordinary GARCH(1,1) model as the foundation of the DCC-GARCH model.

Following the success of the univariate GARCH model, multivariate extensions were a logical way forward. Multivariate models are not only able to model univariate series simultaneously, they allow the investigation of the dynamics between the univariate return series. The constant conditional correlation GARCH (CCC-GARCH) model by Bollerslev (1990) is an example of such a multivariate model. This model assumes there is a constant correlation between the evaluated time series. This results in a constant correlation matrix. The covariance matrix, however, is time-varying. This must be the case since the univariate conditional variances are also dependent on t. The time-varying covariance matrix $\mathbf{H}_t$ is given by

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t, \tag{3}$$

where D_t is a matrix with the conditional standard deviations on its diagonal. This matrix looks like

$$\mathbf{D}_t = \begin{pmatrix} \sqrt{h_{1,t}} & 0 & \cdots & 0 \\ 0 & \sqrt{h_{2,t}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{h_{k,t}} \end{pmatrix}.$$

Furthermore, R is the constant correlation matrix. To complete the model, the demeaned returns are generated by

$$v_t = \mathbf{H}_t^{1/2} z_t, \tag{4}$$

where $z_t \sim N(0, I_k)$ is a vector with length k. The CCC-GARCH model is a great attempt at modeling multivariate return series. The main assumption of a constant correlation matrix, however, is not applicable in practice. Correlations between assets are time-varying, hence the CCC-GARCH model is generalized by Engle (2002) to allow for a time-varying correlation structure. They alter the equation for the covariance matrix such that

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t. \tag{5}$$

Here, $\mathbf{R}_t$ has the shape

$$\mathbf{R}_t = \begin{pmatrix} 1 & \rho_{12,t} & \cdots & \rho_{1k,t} \\ \rho_{21,t} & 1 & \cdots & \rho_{2k,t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k1,t} & \rho_{k2,t} & \cdots & 1 \end{pmatrix}.$$

This setup allows for a dynamic correlation structure. $\mathbf{R}_t$ is also the conditional covariance matrix of the standardized residuals vector ϵ_t. This vector is defined as

$$\epsilon_t = r_t / \sqrt{h_t}. \tag{6}$$

The introduction of time-varying correlation causes the introduction of a set of equations that explains the conditional correlation matrix. The conditional correlation matrix is defined as

$$\mathbf{R}_t = \mathbf{Q}_t^{*-1}\mathbf{Q}_t\mathbf{Q}_t^{*-1}. \tag{7}$$

Here, $\mathbf{Q}_t$ is defined as

$$\mathbf{Q}_t = (1-a-b)\mathbf{S} + a(\epsilon_{t-1}\epsilon_{t-1}^{'}) + b\mathbf{Q}_{t-1}. \tag{8}$$

This specification ensures that $\mathbf{R}_t$ is positive definite. This is a necessary requirement, since we need to ensure that $\mathbf{H}_t$ is positive definite. This is a necessity since $\mathbf{H}_t$ represents a covariance matrix. None of the values in this matrix can fall below zero. Another requirement that is ensured through this specification is that none of the values in the conditional correlation matrix are above one. $\mathbf{Q}_t^*$ is a diagonal matrix with the diagonal entries of $\mathbf{Q}_t$ on its diagonal. In (8), a and b are the DCC parameters. Note that the specification of $\mathbf{Q}_t$ is similar to the autoregressive pattern in (1). The coefficient a is multiplied with the realized correlation matrix of last period, whereas the coefficient b is multiplied with the Q matrix of last period. This resembles autoregressive nature of the unconditional variance in a GARCH(1,1) equation. $\mathbf{S}$ is the unconditional covariance matrix of the standardized residuals, it is defined as

$$\mathbf{S} = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t\epsilon_t^{'}. \tag{9}$$

There are two constraints on the parameters a and b. Namely, $a,\ b > 0$ and $a+b < 1$. These parameter restrictions strongly resemble the GARCH parameter restrictions. They ensure the stationarity of the model.

3.1.2 MLE Estimation

We compare the performance of our cellwise estimator with the two-step estimator introduced by Engle and Sheppard (2001). This estimator is widely used in practice to estimate the DCC-GARCH model. This estimator is not robust to outliers in the data and likely to fall apart in the presence of outliers. The two-step estimator allows us to separate the likelihood function into a mean and correlation part. This way, parameter stability is improved since fewer parameters are estimated simultaneously. Since $r_t \sim N(0, \mathbf{H}_t)$, the total log-likelihood function is as follows

$$\begin{aligned} L &= -\frac{1}{2}\sum_{t=1}^{n}(k\log(2\pi) + \log(|\mathbf{H}_t|) + r_t^{'}\mathbf{H}_t^{-1}r_t) \\ &= -\frac{1}{2}\sum_{t=1}^{n}(k\log(2\pi) + \log(|\mathbf{D}_t|)\mathbf{R}_t\mathbf{D}_t| + r_t^{'}\mathbf{D}_t^{-1}\mathbf{R}_t^{-1}\mathbf{D}_t^{-1}r_t) \\ &= -\frac{1}{2}\sum_{t=1}^{n}(k\log(2\pi) + 2\log(|\mathbf{D}_t|) + \log(|\mathbf{R}_t|) + \epsilon_t^{'}\mathbf{R}_t^{-1}\epsilon_t). \end{aligned} \tag{10}$$

In the first step, the univariate GARCH parameters are estimated. Subsequently, in the second step, the DCC parameters are estimated. We separate the parameters of the DCC-GARCH model into two parameter sets, namely (θ, ϕ). Here the parameters of θ correspond to the univariate GARCH parameters, such that $\theta_i = (\omega_i, \alpha_i, \beta_i)$. Furthermore, $\phi = (a, b)$. When we replace R_t with the identity matrix I_k, we get the log-likelihood function for the first step as

$$\begin{aligned} L_1(\theta|r_t) &= -\frac{1}{2}\sum_{t=1}^{n}(k\log(2\pi) + 2\log(|\mathbf{D}_t|) + \epsilon_t^{'}\epsilon_t) \\ &= -\frac{1}{2}\sum_{t=1}^{n}\left(k\log(2\pi) + \sum_{i=1}^{k}\left(\log(h_t) + \frac{r_t^2}{h_t}\right)\right) \\ &= -\frac{1}{2}\sum_{i=1}^{k}\left(T\log(2\pi) + \sum_{t=1}^{n}\left(\log(h_t) + \frac{r_t^2}{h_t}\right)\right). \end{aligned} \tag{11}$$

The resulting log-likelihood is the sum of the univariate GARCH log-likelihoods. Hence, we can estimate the parameters of each GARCH model individually. Now in the second step, we estimate the DCC parameters given the estimated univariate GARCH parameters through the following log-likelihood function

$$\begin{aligned} L_2(\phi|\hat{\theta}, r_t) &= -\frac{1}{2}\sum_{t=1}^{n}(k\log(2\pi) + 2\log(|\mathbf{D}_t|) + \log(|\mathbf{R}_t|) + \epsilon_t^{'}\mathbf{R}_t^{-1}\epsilon_t) \\ &= -\frac{1}{2}\sum_{t=1}^{n}(\log(|\mathbf{R}_t|) + \epsilon_t^{'}\mathbf{R}_t^{-1}\epsilon_t). \end{aligned} \tag{12}$$

In the last step, we omit the constant terms since they do not affect the estimation of the optimal parameters.

3.1.3 Returns Algorithm

To simulate DCC-GARCH returns, we propose an algorithm that can iteratively generate returns. We simulate demeaned returns instead of actual returns. This makes the algorithm easier without losing the characteristics of a DCC-GARCH model. Ultimately, a level shift of the returns has no influence on the covariance structure. The return series which result from this algorithm adhere to the individual GARCH specifications and the multivariate DCC specification. The algorithm is given as

Algorithm 1: DCC-GARCH return algorithm

Result: DCC-GARCH returns
Define DCC and GARCH parameters a, b, ω_i, α_i and β_i
Initialize $\mathbf{Q}_0, \mathbf{S}_0, h_0$ and v_0
for $s \in (1, ..., t)$ **do**
  for $i \in (1, ..., k)$ **do**
    $h_{s,i} = \omega_i + \alpha_i v_{s,i}^2 + \beta_i h_{s,i}$
  end
  $\mathbf{D}_s = diag(\sqrt{h_s})$
  $\mathbf{S}_s = \frac{1}{s}\sum_{m=1}^{s} \epsilon_s \epsilon_s^{'}$
  $\mathbf{Q}_s = (1 - a - b)\mathbf{S} + a\epsilon_{s-1}\epsilon_{s-1}^{'} + b\mathbf{Q}_{s-1}$
  $\mathbf{Q}_s^* = diag(\sqrt{\mathbf{Q}_s})$
  $\mathbf{R}_s = \mathbf{Q}_s^{*-1}\mathbf{Q}_s\mathbf{Q}_s^{*-1}$
  $\mathbf{H}_s = \mathbf{D}_s\mathbf{R}_s\mathbf{D}_s$
  $z_s \sim N(0, 1)$
  $v_s = \mathbf{H}_s^{\frac{1}{2}} z_s$
  $\epsilon_s = \mathbf{D}_s^{-1} v_s$
end

The algorithm starts with initializing the DCC and GARCH parameters. Furthermore, the matrices **Q**, **S** and the conditional variances and returns are also initialized. Thereafter, the loop starts by calculating the conditional variances for $s = 1$. The conditional variances are used to obtain the conditional standard deviation matrix **D**. Subsequently, the conditional correlation matrix **R** is calculated according to the steps explained earlier. Lastly, the conditional covariance matrix and the returns are computed. Now, another iteration can commence. The loop ends when there have been t iterations. This algorithm ensures that the univariate return series follow the chosen GARCH parameters. Moreover, the entire system follows the DCC parameters a and b.

3.2 Cellwise outliers

To examine our estimator in a simulation setting, we artificially add outliers to DCC-GARCH simulated returns. We create additive outliers as follows. We introduce two new variables, namely the fraction of outliers $0 < \delta < 1$, and the magnitude of the addition λ. After the returns are simulated, we randomly add λ to a fraction δ of returns. Note that if the selected returns are small, they will not be large outliers after addition. We investigate the effect of these outliers on the estimation of the DCC-GARCH model for differing values of δ and λ. First, we concatenate the return series. Thereafter, we add the outliers at random. Hence, it is almost a certainty that the amount of outlying observations is not equal among the univariate return series. In order to make a valid comparison between our findings and the conclusions drawn by Muler and Yohai (2008) and Boudt et al. (2013), we merely add positive outliers. Hence, we create the concatenated return series v_l^* through

$$v_l^* = \begin{cases} v_l + \lambda, & \text{if } l = l_i, 1 \leq i \leq u = \lambda pn, \\ v_l, & \text{otherwise,} \end{cases}$$

where $l_1, ..., l_u$ are the instances where the outliers are added.

3.3 Cellwise Estimator

3.3.1 Estimator

Now, we present and elaborate on our estimator that is specifically designed to handle cellwise contamination. This estimator enables robust estimation of the DCC-GARCH model. What differentiates our estimator from existing robust estimators is its ability to handle not only rowwise outliers but also cellwise outliers. Consequently, valuable data is not discarded if an observation is flagged as an outlier. Instead, we identify the outlying return or returns within the observation, thereby preserving non-outlying returns in an observation flagged as outlying. This approach prevents the unnecessary removal or shrinkage of data, which would potentially lead to a loss of estimation efficiency in determining the values of the DCC-GARCH parameters. Additionally, the identification of outlying returns provides insight into the return series responsible for the outlying observation.

The purpose of this estimator is to detect and deal with large outliers caused by, for instance, a period of financial instability or a big shock. This period of large outliers heavily influences the estimation of the parameters of the DCC-GARCH model, such that the model is not correct when just looking at times when the financial market is more stable. Since these periods with large returns occur sparsely, we deem it safe to label them as outlying. Another way of looking at this estimator is by correcting for volatility. Hence, devolatizing the returns before looking for the outlying values. This way, returns through the whole sample period would be flagged and labeled as outlying returns. We did not choose for this approach, since we deem it more useful for our estimation method to deliver robust estimates in the event of a small period of large outlying returns. This choice means that our estimation method holds less value in the absence of financial turmoil. In this absence, an approach with devolatized returns would be more sensible.

The estimator consists of three key steps. First, we employ Mahalanobis distances, as introduced by Mahalanobis (1936), to detect outlying observations. An observation is flagged as outlying when the squared Mahalanobis distance exceeds a predetermined threshold. Next, we introduce a method that utilizes partial Mahalanobis distances to detect and eliminate the cellwise outliers. Finally, we employ the Expectation-Maximization (EM) algorithm to impute the values of the missing data. This iterative algorithm fills in the most likely values and allows for the completion of the data.

Mahalanobis distances are commonly used in practice to identify outlying cases in large datasets. However, a prerequisite for its application is that the multidimensional distribution is elliptical. Since return data generally follows a Student's t distribution, our approach is applicable. The Mahalanobis distance of an observation can be viewed as the multivariate equivalent of a z-score. A large distance indicates an outlying observation, which could be due to either outlying cells or a fully outlying observation. We assume that observations with a sufficiently low Mahalanobis distance do not contain cellwise outliers. To calculate the squared Mahalanobis distance for each observation x, we employ the following formula

$$D^2(x) = (x - \mu)' \mathbf{\Sigma}^{-1} (x - \mu), \tag{13}$$

where μ is the location vector of the data and Σ the scatter matrix. However, there is a major problem with this approach. According to Rousseeuw and Van Zomeren (1990), using the location vector and scatter matrix of the contaminated data set will result in contaminated and biased Mahalanobis distances. Ideally, we would like to use the mean vector and covariance matrix of the underlying clean distribution. If we knew the underlying distribution, we could easily locate and remove the contaminated rows. Therefore, we need to find robust estimates of μ and Σ given the contaminated data set.

We find these robust estimates through the Minimum Covariance Determinant (MCD) estimator as introduced by Rousseeuw (1985). The MCD estimator estimates the mean vector and covariance matrix with the d observations that minimize the determinant of the sample covariance matrix, where $d \leq n$, the total number of observations. This procedure yields the raw estimates $\hat{\mu}_0$ and $\hat{\boldsymbol{\Sigma}}_0$. Since only a subset of points is used, the covariance matrix of the subset does not represent the covariance structure of the whole sample. The covariances and variances are underestimated in a sense. After all, points that result in a large determinant of the covariance matrix are not taken into account. Therefore, the covariance matrix needs to be multiplied by a constant, which is called the Fisher consistency correction. Fisher consistency entails that an estimator is the true parameter when the population is used as a sample. To make the estimated covariance matrix Fisher consistent at the normal distribution, a consistency factor c_0 is needed as explained in Butler, Davies, and Jhun (1993). Hence, the Fisher consistent covariance matrix is defined as $c_0\hat{\boldsymbol{\Sigma}}_0$, where c_0 is defined as

$$c_0 = \gamma / F_{\Gamma^2_{\frac{k}{2}+1,1}} \left(\frac{\chi^2_\gamma}{2} \right), \tag{14}$$

where $\gamma = d/n$ and k is the number of return series. In other words, k is the number of dimensions in the DCC-GARCH model.

There is still a shortcoming to the obtained parameter estimates. A trade-off exists in the choice of γ. A lower γ results in a higher robustness of the estimates, but simultaneously makes the MCD estimator less efficient. This is investigated in Croux and Haesbroeck (1999). We choose γ to be 0.5 for our analysis, since we value the robustness of the MCD over its efficiency. Moreover, we use a large horizon in our simulations, which we will elaborate on later in the paper. Ideally, we would want to maintain high robustness while increasing the efficiency of the estimator. A paper by Lopuhaa and Rousseeuw (1991) proposes to apply a weighting step to the MCD estimators. The squared Mahalanobis distances $D^2(x_i, \hat{\mu}_{MCD}, \hat{\boldsymbol{\Sigma}}_{MCD})$ are computed with the previously obtained estimates. Then the reweighted estimates are defined as:

$$\hat{\mu}_{MCD} = \frac{\sum_{i=1}^{n} w(D_i^2) x_i}{\sum_{i=1}^{n} w(D_i^2)}, \tag{15}$$

$$\hat{\boldsymbol{\Sigma}}_{MCD} = c_1 \frac{\sum_{i=1}^{n} w(D_i^2)(x_i - \hat{\mu}_{MCD})(x_i - \hat{\mu}_{MCD})'}{\sum_{i=1}^{n} w_i}. \tag{16}$$

The weights w_i are defined as

$$w_i = \begin{cases} 1, & \text{if } l = l_i, 1 \leq i \leq u = \lambda kn, \\ 0, & \text{otherwise.} \end{cases}$$

In essence, observations with a higher squared Mahalanobis distance than the cutoff value are disregarded for the estimation of the final location and covariance matrix. The Mahalanobis distances are calculated with the location and covariance matrix obtained in the previous step. Once again, we multiply the covariance matrix with a constant c_1 to obtain a Fisher consistent estimator. This time the constant is defined as

$$c_1 = (1 - \delta)/F_{\Gamma(\frac{k}{2}+1,1)}(\chi^2_{1-\delta}/2), \tag{17}$$

where δ is 0.025. Let F denote the F-distribution. With the reweighted mean and covariance matrix, we compute the squared Mahalanobis distances for each observation. If the squared Mahalanobis distance exceeds the cutoff value of $\chi^2_{k,0.975}$, we flag this observation as an outlier.

The subsequent stage in the estimator involves identifying cellwise outliers by examining the flagged observations. Danilov (2010) introduced the procedure for identifying cellwise outliers in the flagged observations, known as the Partial Mahalanobis distance approach (P-approach). This approach compares the partial Mahalanobis distances to the full Mahalanobis distances calculated in the initial step of the algorithm in order to identify outliers. The P-approach employs an iterative procedure where one observation is omitted at a time to calculate the Mahalanobis distance for the remaining values. If the resulting Mahalanobis distance falls below the cutoff value, the remaining observations are considered normal or non-outlying, implying that the omitted observation is an outlier. The partial squared Mahalanobis distance is denoted as MD_j^2, where j represents the index of the omitted observation. The cutoff value for the partial Mahalanobis distance, denoted as C_{n-1}, is determined by the number of return observations, or the number of investigated equity prices, indicated by n. It should be noted that this cutoff value is smaller than the value used to identify outliers, as the degrees of freedom have decreased. When examining the flagged observations, three possible outcomes can occur.

The first outcome can be expressed as $\#(j|MD_j^2 \leq C_{n-1}) = 1$. This signifies that excluding the jth return value results in the remaining observation no longer being classified as an outlier. Eliminating any other data point will not cause the observation to fall below the cutoff value. The outlying observation is consequently removed from the dataset.

The second outcome is $\#(j|MD_j^2 \leq C_{n-1}) = u \geq 2$. This outcome indicates that eliminating any of the u data cells renders the remaining observation non-outlying. This circumstance is misleading because it suggests that if a supposedly non-outlying observation is removed, it suddenly becomes an outlier, and vice versa. To resolve this issue, we adopt a strategy where we remove the data cell that triggers the largest decrease in the squared Mahalanobis distance, denoted as $MD^2 - MD_j^2$. By doing so, we eliminate the observation that is deemed most likely to be an outlier from the dataset.

The third outcome is $\#(j|MD_j^2 \leq C_{n-1}) = 0$. In this case, there is not a single instance where the squared partial Mahalanobis distance falls below the predefined threshold when data cells are omitted. This leads us to suspect that multiple observations are outliers, rather than just one. If the number of observations n is equal to or less than 3, we classify all cells as outliers. However, if n is greater than 3, we proceed to engage in a more thorough detection process by gradually removing pairs of observations. The same logic applies here: if eliminating a pair of observations yields a normal observation, both observations in the pair are flagged as outliers. If, at this level, no outliers are found, we delve even deeper and create groups of three cells. This process continues until a group of outliers is identified. The identified outlying observations are then eliminated from the dataset.

In the third and final stage, the missing values caused by the removal of outliers in the previous step are imputed in the dataset. To accomplish this, we utilize the Expectation Maximization (EM) algorithm, introduced by Dempster, Laird, and Rubin, 1977. We opt for EM imputation over alternative methods such as mean imputation, as Donders, Van Der Heijden, Stijnen, and Moons (2006) demonstrate that mean imputation introduces undesirable biases to the dataset. The resulting clean dataset is subsequently used to compute the parameters of the DCC-GARCH model. The EM algorithm is predicated on the assumption of multivariate normal data, as it simplifies the utilization of conditional normal distributions. The entire data matrix is denoted as $\mathbf{Y}$, where each ith observation constitutes a k-dimensional vector, with k representing the number of distinct returns. Given our assumption of a normal data matrix, y_s follows a multivariate normal distribution $N(\mu, \mathbf{\Sigma})$. Nonetheless, some of these y_s vectors will contain missing components due to the removal of outlying values during the previous step. We denote these missing components as z_s, whereas the observed components are denoted as x_s. We further assume that each subvector follows a normal distribution. Therefore, the following holds true:

$$\begin{pmatrix} x_s \\ z_s \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{s,x} \\ \mu_{s,z} \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{s,xx} & \mathbf{\Sigma}_{s,xz} \\ \mathbf{\Sigma}_{s,zx} & \mathbf{\Sigma}_{s,zz} \end{pmatrix}\right).$$

We are interested in the expected mean and variance of the missing values given the observed values and the parameters. In other words, we are interested in m_s and V_s in

$$p(z_s | x_s, \mu, \mathbf{\Sigma}) \sim N(m_s, \mathbf{V}_s). \tag{18}$$

Following a property of joint normal distributions, we can compute the conditional mean and variance as

$$m_s = \mu_{s,z} + \mathbf{\Sigma}_{s,zx}\mathbf{\Sigma}_{s,xx}^{-1}(x_i - \mu_{s,x}), \tag{19}$$

$$\mathbf{V}_s = \mathbf{\Sigma}_{s,zz} - \mathbf{\Sigma}_{s,zx}\mathbf{\Sigma}_{s,xx}^{-1}\mathbf{\Sigma}_{s,xz}. \tag{20}$$

The previous equations form the basis of the EM-algorithm used to impute the missing values after deletion of the outlying cells. The algorithm is as follows

Algorithm 2: EM missing values imputation

Result: Imputed return matrix

Define the return matrix with removed cells as $\mathbf{Y}$. Furthermore, define the small stopping value $\nu = 0.001$. Define $\mathbf{P}$ as a copy of the initial missing data matrix $\mathbf{Y}$. Furthermore, initialize μ_0 and $\mathbf{\Sigma}_0$ with the observed data.

while $||(\mu_s - \mu_{s-1})||_F > \nu$ *or* $||(\mathbf{\Sigma}_s - \mathbf{\Sigma}_{s-1})||_F > \nu$ **do**

  for $s \in (1, ..., n)$ **do**

    if $\mathbf{P}_s$ *contains missing values* **then**

      E-step:

      $m_s = \mu_{s,z} + \mathbf{\Sigma}_{s,zx}\mathbf{\Sigma}_{s,xx}^{-1}(x_s - \mu_{s,x})$

      $\mathbf{V}_s = \mathbf{\Sigma}_{s,zz} - \mathbf{\Sigma}_{s,zx}\mathbf{\Sigma}_{s,xx}^{-1}\mathbf{\Sigma}_{s,xz}$

      $\mathbf{Y}_{s,z} = m_s$

    end

  end

  M-step:

  $\mu_{w+1} = \frac{1}{n}\sum_{i=1}^{n} y_i$

  $\mathbf{\Sigma}_{w+1} = \frac{1}{n}\sum_{i=1}^{n} y_i y_i^{'} - \mu_{w+1}\mu_{w+1}^{'} + \mathbf{V}_i$

end

where w is the iteration of the algorithm. Furthermore, $||...||_F$ is the Frobenius norm. The Frobenius is another name for the Euclidean norm in the case of a vector. The Frobenius norm of a matrix is defined as

$$||\mathbf{A}||_F = \sqrt{Tr(\mathbf{A}\mathbf{A}^T)}. \tag{21}$$

This algorithm runs as long as there is a significant improvement in the estimation of the mean vector or the covariance matrix. In the Expectation-step, all the missing values are imputed. Subsequently, the mean and covariance structure are updated. Note that $\mathbf{V}_i$ is a zero matrix when there are no missing values in row i of the data set.

3.3.2 Cutoff values

There is an issue when using Mahalanobis distances to determine outliers in a data set of returns. The general Mahalanobis cutoff values are χ^2 distributed since the cutoff is based on a multidimensional normal distribution. However, return data is empirically not normally distributed. Return data is almost always characterised by a multidimensional Student's t distribution. Hence, it would not be appropriate to use the χ^2 cutoff values. The usage of these values would result marking regular observations as outlying. Finding a closed-form solution for cutoff values based on a Student's t distribution is a cumbersome endeavour. Therefore, we take a different approach. We simulate multiple multidimensional Student's t distributions. Thereafter, the Mahalanobis distances for every simulated data set are computed. We define κ as the significance level. If we want to find the cutoff value for $k = 3$ and $\kappa = 0.05$, we select the Mahalanobis distance such that 5 percent of the Mahalanobis distances is bigger than this particular distance. The same reasoning holds for different values of κ. Below, we present a table with Mahalanobis cutoff values for different k and κ. The degrees of freedom for the Student's t distributions is set to 4.

Table 1

Mahalanobis distance cutoff values

		k		
		2	3	4
	0.05	13.89	19.72	25.54
κ	0.025	21.28	29.74	38.36
	0.01	35.98	49.50	64.05

Note. This table displays the Mahalanobis distance cutoff values for different combinations of κ and k. The values are generated through simulating 50,000 samples from a multidimensional Student's t distribution

3.4 M-estimator

To invigorate the effectiveness of our Mahalanobis estimator, we compare the estimating accuracy with the M-estimator used in Muler and Yohai (2008). Note that the estimators used in their paper merely focus on the estimation of the univariate GARCH models. The estimation of the multivariate parameters a and b are thus not compared. M-estimators are a generalization of MLE estimation. They allow for a different specification of the loss function used to estimate the parameters. This allows for putting less weight on outliers while estimating, making these estimators more robust to large outliers.

We now elaborate on the M-estimator suggested by Muler and Yohai. They show that maximizing the log-likelihood function in (11) is the same as minimizing

$$L = \frac{1}{n-k} \sum_{t=k+1}^{n} \psi_0(y_t - \log{(h_t)}), \tag{22}$$

where $\psi_0 = -\log(g_0)$. Here, g_0 is defined as

$$g_0(w) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}(e^w - w)}. \tag{23}$$

This QML estimate of the GARCH parameters is not robust since ψ_0 is unbounded. In other words, one single outlier is able to introduce a large bias to the estimated parameters. The estimation of the log-likelihood can be made more robust by using alternative bounded ψ-functions. Nevertheless, these estimates will still be influenced by outliers since the conditional variances are computed with (1). An outlying conditional variance value at time t influences the conditional variances after t.

Muler and Yohai suggest to use an M-estimate defined as $\psi_1 = m(\rho_0)$, where m(x) is defined as

$$m_x = \begin{cases} x, & \text{if } x \leq 4.02, \\ c_4x^4 + c_3x^3 + c_2x^2 + c_1x + c_0, & \text{if } 4.02 < x \leq 4.30, \\ 4.16, & \text{otherwise,} \end{cases}$$

where $c_0 = 6777, c_1 = -6536.2, c_2 = 2362.3, c_3 = -379.0087$ and $c_4 = 22.777$. Note that this is a smoothed function of a function that attains the value 4.02 after x reaches 4.02. Propagating ψ_0 into this smoothing function results in bounding the conventional ψ-function used in general QMLE of univariate GARCH models. This way, the influence of large outliers diminishes greatly.

4 Simulation

In this section, we conduct a simulation study of the DCC-GARCH model in the presence of additive outliers. The parameters of the model are estimated by two-step maximum likelihood estimation (MLE), M-estimation, and the newly proposed estimator that uses Mahalanobis distances to filter out cellwise outliers. We compare the performance of the three estimation methods in terms of bias and variance.

We simulate a three-dimensional DCC-GARCH model with 5000 data points. For all results, the number of iterations per simulation is 100. The Mahalanobis cutoff values are chosen for $\kappa = 0.01$. The multivariate parameters a and b are set to the values of 0.1 and 0.8 respectively. This specification adheres to the parameter restrictions of $a, \ b < 1$ and $a \ + \ b < 1$. Moreover, these values are realistic to be obtained for real-life time series. The GARCH parameters are $\omega_1 = 0.1$. $\alpha_1 = 0.3, \ \beta_1 = 0.6, \ \omega_2 = 0.1, \ \alpha_2 = 0.2, \ \beta_2 = 0.7, \ \omega_3 = 0.1, \ \alpha_3 = 0.1, \ \beta_3 = 0.8$. Similarly to the multivariate parameters, these parameter values also adhere to the parameter restrictions $\omega, \ \alpha, \ \beta < 1$ and $\omega + \alpha + \beta < 1$.

Table 2

Performance estimators with no outliers.

	Bias			MSE			Variance		
	MLE	M-estimator	Cellwise estimator	MLE	M-estimator	Cellwise estimator	MLE	M-estimator	Cellwise estimator
ω_1	0.000	0.000	-0.006	0.0001	0.0002	0.0002	0.00012	0.00023	0.00012
α_1	0.001	0.004	-0.025	0.0005	0.0009	0.0012	0.00054	0.00085	0.00058
β_1	0.000	-0.002	0.026	0.0007	0.0012	0.0015	0.00068	0.00121	0.00086
ω_2	0.001	0.001	-0.001	0.0002	0.0003	0.0002	0.00018	0.00031	0.00018
α_2	-0.002	-0.002	-0.010	0.0005	0.0007	0.0005	0.00051	0.00066	0.00041
β_2	-0.001	0.000	0.008	0.0006	0.0009	0.0007	0.00062	0.00091	0.00062
ω_3	0.004	0.008	0.003	0.0004	0.0007	0.0004	0.00035	0.00067	0.00037
α_3	0.002	0.005	-0.002	0.0002	0.0003	0.0002	0.00021	0.00030	0.00016
β_3	-0.006	-0.012	-0.002	0.0009	0.0015	0.0009	0.00083	0.00135	0.00085
a	0.004		-0.014	0.0001		0.0005	0.00067		0.00034
b	-0.005		0.007	0.0003		0.0015	0.00027		0.00145

Note. This table displays the bias, mean squared error (MSE) and the variance of the parameter estimates of the three-dimensional DCC-GARCH model. For this simulation, no outliers are added. The three estimators which are compared in this table are the maximum likelihood estimation (MLE) estimator, the Muler-Yohai M-estimator and the cellwise estimator.

We first present the estimation results for particular choices of δ and λ. Table 2 shows the results for a simulation without any outliers. This simulation is carried out to gauge the efficacy of the robust estimation methods when no outliers are present. The biases of the robust estimators are slightly higher than the bias of the MLE method. This is to be expected, since large values are unrightfully marked as outliers. However, since the biases are still small, we can conclude that the amount of wrongly flagged outliers is not large. The highest biases are found for the cellwise estimator parameters estimates of α_1 and β_1. Across the board the variances are low with no significant differences. Hence, we can conclude that in the case of no outliers, the cellwise estimator possibly has a small bias in the estimation of univariate GARCH parameters.

Table 3 shows the results for a simulation where $\delta = 0.01$ and $\lambda = 5$. We observe a clear difference in the performance of the estimators. As expected, the MLE estimator introduces significant bias to the estimates. The positive bias is particularly large for the estimation of the constants ω. This is because the addition of outliers increases the variance of the time series significantly. The α-parameters are affected the least out of all the GARCH parameters. Moreover, there is a large negative bias of the multivariate parameter b. The M-estimator results in less bias than the MLE estimator, especially for the constants ω. Our cellwise estimator is the best in terms of bias. There is still some bias in the estimation of α, but it is smaller than the bias of the MLE and M-estimator.

When examining the variance of the estimator, a distinct observation emerges. In general, all estimators demonstrate low variance for the majority of parameters. However, the variance of the Maximum Likelihood Estimation (MLE) estimate for parameter b deviates notably from this pattern with a value of 0.073. This is significantly higher than the variances of the other parameters.

Table 3

Performance estimators for $\lambda = 5$.

	Bias			MSE			Variance		
	MLE	M-estimator	Cellwise estimator	MLE	M-estimator	Cellwise estimator	MLE	M-estimator	Cellwise estimator
ω_1	0.166	0.052	0.002	0.0335	0.0034	0.0003	0.00585	0.00067	0.00031
α_1	-0.038	-0.014	-0.027	0.0037	0.0016	0.0012	0.00232	0.00140	0.00045
β_1	-0.059	-0.087	0.021	0.0115	0.0100	0.0011	0.00810	0.00238	0.00059
ω_2	0.132	0.057	0.021	0.0223	0.0045	0.0030	0.00493	0.00127	0.00257
α_2	-0.037	-0.031	-0.012	0.0032	0.0019	0.0005	0.00183	0.00095	0.00034
β_2	-0.039	-0.062	-0.005	0.0080	0.0072	0.0016	0.00646	0.00342	0.00159
ω_3	0.108	0.051	0.005	0.0431	0.0070	0.0009	0.03134	0.00433	0.00083
α_3	-0.039	-0.045	-0.013	0.0024	0.0025	0.0005	0.00090	0.00046	0.00033
β_3	-0.025	-0.023	0.011	0.0245	0.0073	0.0009	0.02388	0.00676	0.00077
a	-0.013		-0.020	0.0015		0.0007	0.00129		0.00030
b	-0.283		-0.019	0.1527		0.0025	0.07272		0.00211

Note. This table displays the bias, mean squared error (MSE) and the variance of the parameter estimates of the three-dimensional DCC-GARCH model. For this simulation, $\delta = 0.01$ and $\lambda = 5$. The three estimators which are compared in this table are the maximum likelihood estimation (MLE) estimator, the Muler-Yohai M-estimator and the cellwise estimator.

For the subsequent simulation, the value of λ is modified to 10, resulting in an error magnitude that is twice as large as the previous simulation. The fraction of outliers, however, remains constant at 0.01. The results of this simulation are presented in Table 4. Upon initial examination, it becomes evident that the performance of MLE estimation is severely compromised. The biases for nearly all parameters are substantially large. The biases of the estimated constants exceed 0.3, while the underlying ω parameters all have values of 0.1. Consequently, the estimated constants are at least three times greater than their true values. The sole parameter with relatively low bias are α_3 and a. It is evident that MLE fails to provide accurate estimates when confronted with outliers of this magnitude. As a result, the estimates should be approached with caution, as they are of limited utility. On the other hand, the Muler-Yohai M-estimator exhibits improved performance compared to MLE, with significantly smaller biases for almost all parameters. The cellwise estimator estimations display minimal bias for all parameters. Similar patterns emerge when evaluating the variances. The MLE estimates demonstrate large variances, particularly for the ω parameters. In contrast, the variances of the a estimates are low, indicating that MLE consistently estimates a wrong. The cellwise estimator exhibits minimal variance in its estimations of the DCC-GARCH parameters.

In comparison to the simulation with a fixed value of $\lambda = 5$, several notable differences can be observed in the outcomes. Specifically, when examining the bias, it is evident that the MLE estimates perform significantly worse as the magnitude of the added outliers increases. This outcome was to be expected, as the presence of outliers with larger magnitudes inherently leads to a deteriorated bias in the estimates. Similarly, the estimates obtained through the M-estimator also exhibit a similar trend.

Table 4

Performance estimators for $\lambda = 10$.

	Bias			MSE			Variance		
	MLE	M-estimator	Cellwise estimator	MLE	M-estimator	Cellwise estimator	MLE	M-estimator	Cellwise estimator
ω_1	0.897	0.203	0.001	1.007	0.046	0.0001	0.20261	0.00492	0.00012
α_1	-0.042	0.068	-0.024	0.031	0.006	0.0010	0.02871	0.00162	0.00041
β_1	-0.259	-0.359	0.015	0.140	0.139	0.0010	0.07258	0.01016	0.00074
ω_2	0.756	0.320	-0.002	0.938	0.114	0.0002	0.36579	0.01160	0.00017
α_2	-0.075	0.033	-0.008	0.020	0.005	0.0003	0.01438	0.00342	0.00024
β_2	-0.224	-0.439	0.008	0.175	0.221	0.0008	0.12483	0.02811	0.00069
ω_3	0.335	0.319	-0.001	0.572	0.232	0.0004	0.45926	0.13030	0.00036
α_3	-0.090	-0.095	-0.007	0.009	0.009	0.0002	0.00076	0.00020	0.00016
β_3	-0.025	-0.255	0.006	0.119	0.210	0.0008	0.11859	0.14540	0.00081
a	-0.032		-0.020	0.004		0.0008	0.00296		0.00041
b	-0.450		-0.007	0.363		0.0023	0.16063		0.00224

Note. This table displays the bias, mean squared error (MSE) and the variance of the parameter estimates of the three-dimensional DCC-GARCH model. For this simulation, $\delta = 0.01$ and $\lambda = 10$. The three estimators which are compared in this table are the maximum likelihood estimation (MLE) estimator, the Muler-Yohai M-estimator and the cellwise estimator.

Table 5

Relative variances cellwise and MLE estimator

	$\lambda = 0$	$\lambda = 5$	$\lambda = 10$
α_1	1.014	0.053	0.001
β_1	1.067	0.193	0.014
ω_1	1.264	0.073	0.010
α_2	1.005	0.052	0.001
β_2	0.795	0.185	0.017
ω_2	0.999	0.246	0.006
α_3	1.061	0.026	0.001
β_3	0.767	0.363	0.215
ω_3	1.023	0.033	0.007
a	5.093	0.233	0.140
b	5.474	0.029	0.014

Note. This table displays the relative variances as a ratio of the cellwise estimator variance to the MLE variance. δ is set to 0.01 and the three values for λ's are 0, 5 and 10.

Conversely, in contrast to the aforementioned results, the cellwise estimator demonstrates superior performance in the presence of larger outliers. The biases and variances of the estimates are both reduced, likely due to the increased magnitude of the outliers enabling the estimator to better identify and address the outlying cells. This improvement can be attributed to the clearer distinction between additive outliers and regular returns, thus leading to a greater probability of detecting and imputing additive outliers.

Table 5 shows the ratios of the cellwise estimator and MLE estimation variances. We see that there is evidence for a tradeoff in the preferred estimation method. On top of the higher bias of the parameter estimates, the variance of the cellwise estimator estimates is higher for almost all parameters. Especially the multivariate parameters a and b are difficult to estimate consistently for the cellwise estimator. However, when outliers are present in the data, the cellwise estimator produces significantly lower variance than the MLE estimator. These findings indicate that in the case of no outliers, the MLE estimator performs better. However, its performance falls off quickly when outliers are present.

Now, attention is focused on the estimation bias of specific parameters, namely ω, a and b. Figure 1 illustrates the estimates of ω_1, as well as the DCC parameters a and b, as λ is varied. The frequency of added outliers remains fixed at 0.01 for the purpose of this analysis, and the estimates derived from both MLE and the cellwise estimator are juxtaposed. The scales of the left and right column are different, such that a pattern is discernible in the column with cellwise estimates. Notably, the bias associated with MLE estimation for the constant ω_1 rapidly escalates as λ increases. This observation aligns with expectations, as the increasingly large size of the outliers contributes to an overestimation of the univariate conditional variance within the initial simulated return series. In contrast, the bias exhibited by the cellwise estimator is significantly smaller than that of the MLE estimate across the entire range of λ. Strikingly, the bias initially exhibits an upward trend before subsequently decreasing after a λ value of 4. This peculiar pattern may be explained by the fact that the estimator struggles to detect the presence of smaller outliers. As the magnitude of the outliers grows, however, the estimator effectively identifies and appropriately estimates their underlying parameter value. Similar patterns emerge when inspecting the estimates of the multivariate parameters. Specifically, the estimates of a and b become increasingly biased as λ rises, with biases demonstrating steady growth that accelerates upon exceeding $\lambda = 8$. Notably, the cellwise estimates also exhibit slight bias when the outliers are of a smaller scale. In line with the findings for ω_1, the bias recedes as the outliers grow larger. These findings suggest that the cellwise estimator outperforms MLE estimation, particularly in cases where outliers are considerable in size.

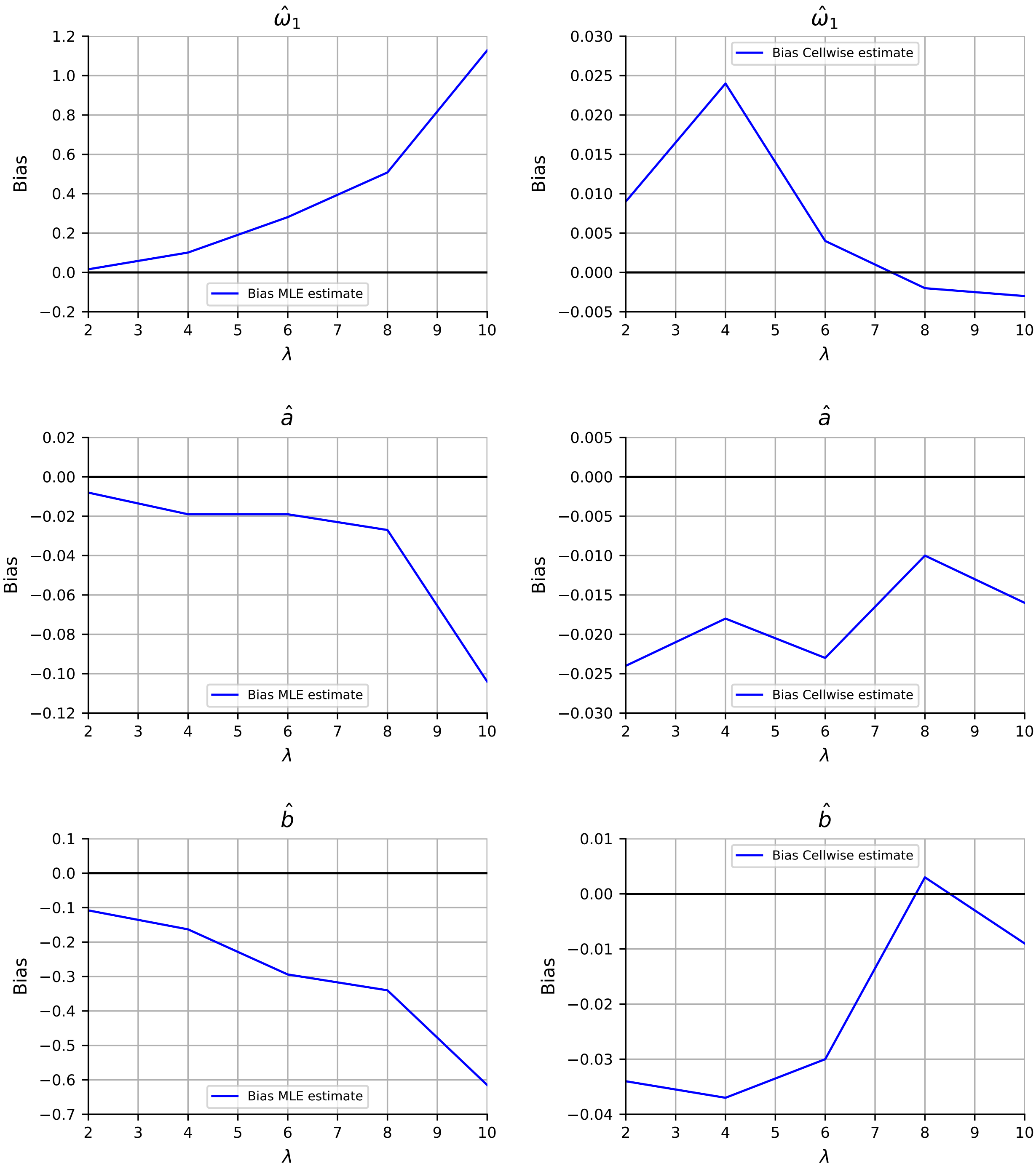

Figure 1. Bias MLE and Cellwise estimates for differing λ.

In Figure 2 we again look at the bias of the estimates of the same parameters. However, now we keep λ fixed at 10 and alter the value of δ, the frequency of outliers. the first column displays the bias for the MLE estimates, whereas the second column displays the bias for the cellwise estimator estimates. The magnitude of δ does not seem to have an effect on the bias of the cellwise estimator estimate of ω_1. The bias is low throughout the range of δ with only small differences. We see that MLE is not able to estimate a accurately, regardless of the fraction of outliers. The cellwise estimator performs slightly better, but the bias increases for increasing δ. The cellwise estimator does a better job at estimating b. However, we see that for increasing δ, the slightly negative bias turns into a slightly positive bias. Regular MLE estimation is not able to estimate b accurately for any value of δ. It must be said that the Mahalanobis cutoff values used for this simulation remain the same, thus $\alpha = 0.01$. These cutoff values do not make sense anymore, since δ increases. This results in the cellwise estimator not being able to accurately estimate a. This result shows that the assumed fraction of outliers is important when using the cellwise estimator. When the fraction of outliers is underestimated, biases can occur.

The simulation results indicate that the cellwise estimator outperforms the Muler-Yohai estimator for the univariate GARCH parameters and the MLE for both the univariate and the multivariate parameters. Furthermore, the cellwise estimator is able to robustly estimate DCC-GARCH parameters in the presence of large outliers. However, the cellwise estimator does exhibit some performance degradation when the proportion of outliers is very large and the cutoff values are not adjusted appropriately. Nevertheless, it still outperforms the other two estimators by a substantial margin. The decreasing performance of the cellwise estimator when the proportion of outliers grows large is not a cause for concern, since these outlier proportions are not typically observed in practice.

We find that our results are consistent with those reported in Muler and Yohai (2008). The cellwise estimator is able to estimate the GARCH parameters with negligible or small bias. Muler and Yohai adjust the model so that their BM-estimator robustly estimates the parameters. We are able to produce similar results even though the GARCH(1,1) models are not modified to account for outliers in the data. This finding is consistent with our comparison to the BIP-GARCH model introduced in Boudt et al. (2013). They also find a method that robustly estimates the DCC-GARCH model, but certain modifications are made to the model to make it robust to outliers. We conclude that the cellwise estimator produces similar results as the estimators in the two aforementioned papers without further

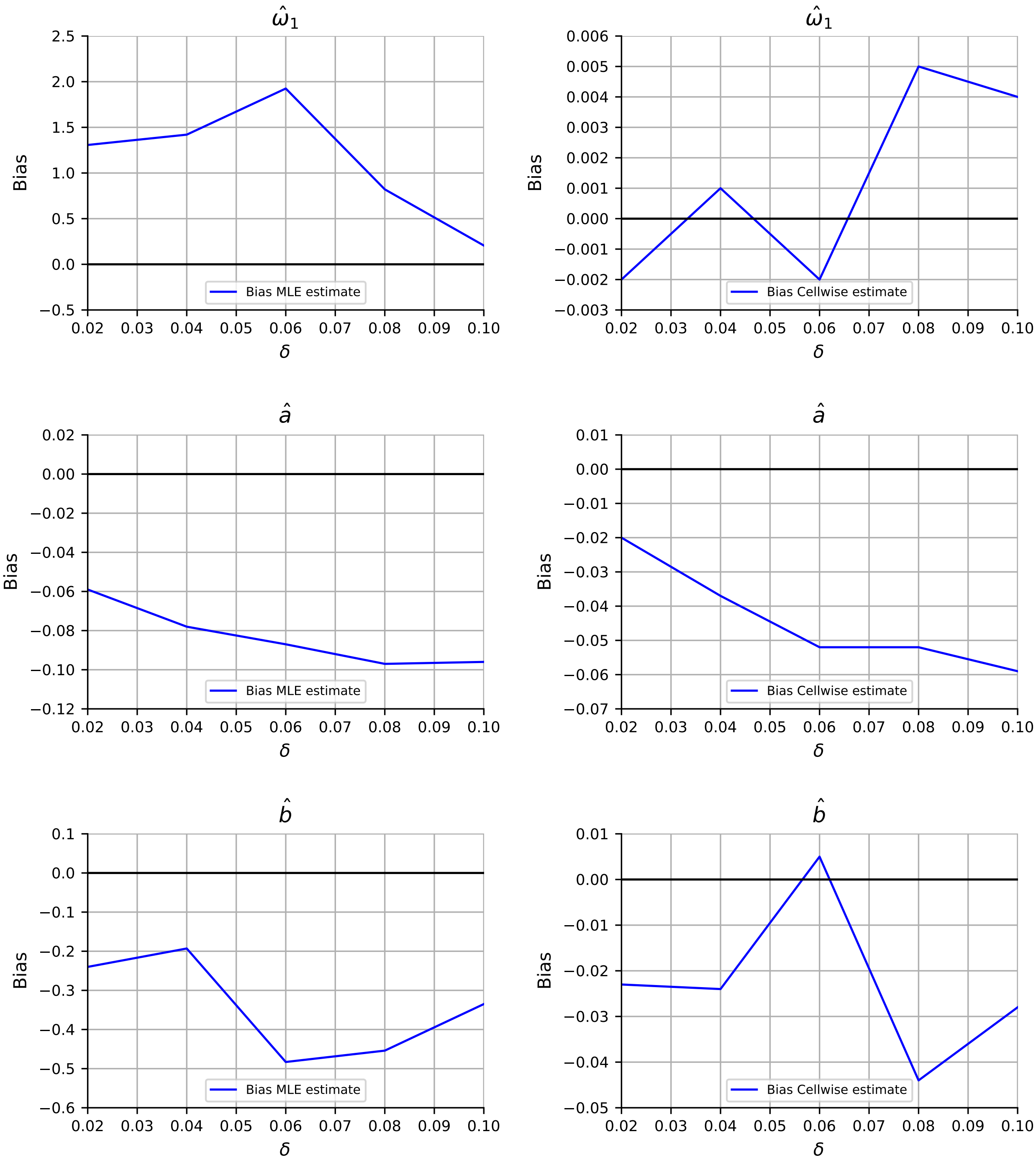

Figure 2. Bias MLE and Cellwise estimates for differing δ.

complicating the model. The cellwise estimator is therefore a good alternative for users who are concerned with ease of use when employing robust estimators for the DCC-GARCH model.

5 Data

In this section, we give information on the data used to carry out the empirical part of our research.

The return data of Apple, Amazon and Johnson & Johnson is gathered from the Yahoo Finance database. The considered time frame is from January 2011 until December 2022. This selection results in 3019 data points. Moreover, Covid-19 and the connected financial turmoil fall into this period. The returns are calculated through taking the percentage increase or decrease of the adjusted close price. So, $r_t = (\frac{p_{close,t}}{p_{close,t-1}} - 1) * 100\%$, where r is the return and p_{close} is the adjusted close price of the stock.

Table 6 displays the descriptive statistics of the return data. The returns of Johnson & Johnson stand out since the standard deviation is substantially lower than every other asset. Furthermore, the maximum and minimum are the highest and lowest of all the stocks. The Pearson correlation coefficients of all returns are displayed in Table 7.

Table 6

Descriptive statistics

	Mean	Standard deviation	Maximum	Minimum
Apple	0.101	1.81	12.0	-12.9
Amazon	0.095	2.08	15.7	-14.0
Johnson & Johnson	0.052	1.08	8.00	-10.0

Note. This table displays the mean, standard deviation, maximum and minimum for the returns of Apple, Amazon and Johnson & Johnson over the period of 2011 through 2022.

Table 7

Pearson correlation coefficients

Stock			
Apple	-		
Amazon	0.487	-	
Johnson & Johnson	0.357	0.289	-

Note. This table displays the Pearson correlation coefficient for the returns of Apple, Amazon and Johnson & Johnson over the period of 2011 through 2022.

The GARCH(1,1) volatility of the Apple returns is displayed below in Figure 3. The volatility does not display a clear trend. It is rather constant with a few large peaks. Most noticeable is the volatility spike during the start of the Covid-19 crisis.

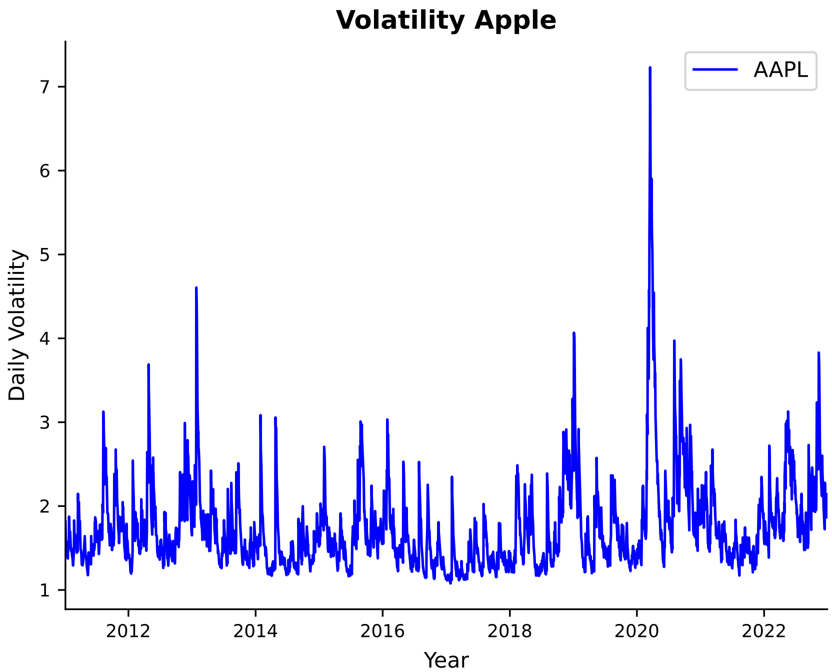

Figure 3. GARCH(1,1) daily volatility of Apple stock.

6 Empirical Results

In this section we study the data discussed in the previous section. The cellwise estimator, M-estimator and MLE estimator are used to estimate the DCC-GARCH model for the data. The parameter estimates are subsequently compared. Thereafter, we look at the covariance structures of the three models. We compare the estimated variances and covariances in times of crisis. Ultimately, this is where users of the DCC-GARCH model are interested in. This comparison leads to conclusions on the effectiveness of estimators for real-life data.

Table 8 displays the estimated DCC-GARCH parameters for a multidimensional model including the three stocks discussed previously. For the cellwise estimator, we chose κ to be 0.05, assuming 5 percent of the data to be outlying. The estimated constants ω of the MLE estimation are significantly larger than the constants estimated by the robust methods. This signifies that base level conditional volatility is estimated to be larger with MLE estimation. This makes sense since these parameter estimates are heavily influenced by outliers. Another interesting observation is that the cellwise estimator consistently estimates β to be larger and α to be smaller than the MLE estimator and the M-estimator estimates. This finding indicates that the autoregressive property of the conditional variance is perceived to be bigger when outliers are filtered out of the dataset. In other words, the estimated conditional volatility is more stable and less spiky. We must note that the numbers in the cellwise estimator column in Table 8 are different when the value of κ is different. However, the overall conclusions do not change.

Table 8

Parameter estimates DCC-GARCH model for S&P 500 data

	MLE	M-estimator	Cellwise estimator
ω_1	0.176	0.086	0.037
α_1	0.116	0.066	0.056
β_1	0.833	0.878	0.929
ω_2	0.390	0.047	0.042
α_2	0.173	0.045	0.052
β_2	0.756	0.917	0.933
ω_3	0.052	0.029	0.029
α_3	0.094	0.077	0.062
β_3	0.858	0.859	0.901
a	0.008		0.008
b	0.990		0.991

Note. This table displays the parameter estimates of the MLE estimator, M-estimator and the cellwise estimator for the DCC-GARCH model. The data ranges from January 2011 until December 2022. The data consists of daily returns from Apple, Amazon and Johnson & Johnson. The Mahalanobis cutoff values are chosen for $\kappa = 0.05$.

This result is invigorated by Table 9. The table displays the DCC-GARCH covariance matrices during financial stability and turmoil. The 16th of March 2020 falls in the height of the high variance return situation induced by the appearance of the Covid-19 pandemic. The 10th of January 2018 is a random date chosen during a time with low volatility on the market. The variances and covariances estimated by MLE explode during the highly volatile market. Moreover, the variances and covariances are also large in 2018. This is due to the overestimation of the covariance matrix elements. This table is another example of MLE estimation not being able to handle outliers in the data. On the other hand, when looking at the covariance matrices produced by the cellwise estimator, we see little difference in the values. This indicates that the outlying period of the Covid-19 pandemic does not heavily influence the variance-covariance structure of the three stocks. This result is also notable in Figure 4. This figure displays the estimated covariance between Amazon and Johnson & Johnson returns. The covariance estimated by the MLE method is not stable and is influenced by periods of financial instability. The covariance estimated by the DCC-GARCH model making use of the cellwise estimator is much more stable and only slightly increases when markets are unstable.

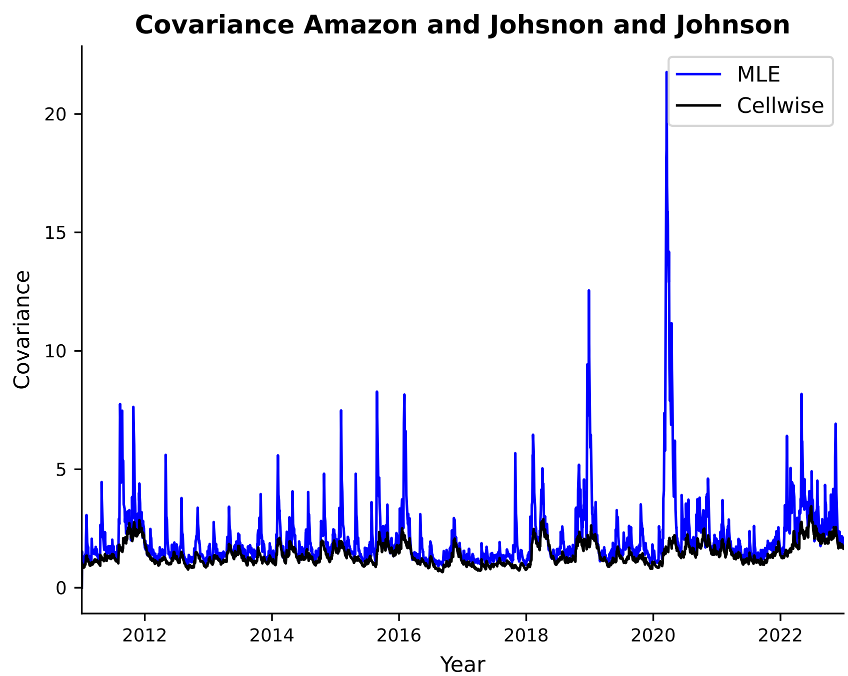

Figure 4. DCC-GARCH covariances of Amazon and Johnson & Johnson.

The findings in Table 9 have great implications for, for instance, portfolio managers. We use the portfolio selection theory of Markowitz (1952) as an example. This theory is one of the most well-known papers in the field of portfolio management. This theory aims to maximize profits while keeping the variance of a portfolio at a minimum. It defines the variance of a portfolio consisting of two assets as follows

$$V_{portfolio} = w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_1\sigma_2\rho_{12}. \tag{24}$$

Here, w_i corresponds to the weight given to asset i. Furthermore, σ_{ii} is the variance of asset i, whereas σ_i is the volatility of asset i. Lastly, ρ_{12} is the correlation between the two

Table 9

Covariance tables during financial stability and turmoil

	Apple	Amazon	J&J
Apple	4.50	3.86	1.79
Amazon	3.86	3.20	1.53
J&J	1.79	1.54	0.71

a *MLE estimator covariance matrix 10/01/2018*

	Apple	Amazon	J&J
Apple	1.28	1.38	0.89
Amazon	1.38	1.47	0.95
J&J	0.89	0.95	0.61

b *Cellwise estimator covariance matrix 10/01/2018*

	Apple	Amazon	J&J
Apple	39.61	30.24	23.77
Amazon	30.24	23.09	18.15
J&J	23.77	18.15	14.27

c *MLE estimator covariance matrix 16/03/2020*

	Apple	Amazon	J&J
Apple	1.59	1.99	1.10
Amazon	1.99	2.49	1.37
J&J	1.10	1.38	0.76

d *Cellwise estimator covariance matrix 16/03/2020*

Note. This table displays the covariance matrices on two specific dates produced by the DCC-GARCH models when the MLE and cellwise estimator are used. For the cellwise estimator, α is chosen to be 0.05.

assets. From the above formula, we can deduce that a wrong estimate of the variances and/or covariance of the assets has a significant impact on the estimation of the portfolio variance. Biased estimates could thus lead to underestimated or overestimated variance portfolios. This in turn could lead to sub-optimal portfolio allocations.

Another benefit of the cellwise estimator over the MLE estimator is the stability of the estimates. Table 9 shows that the covariance matrix is more stable over prolonged periods of time. This means that the optimal portfolio allocation using the cellwise estimator stays more similar compared to an allocation using the MLE estimator. The improved stability thus leads to a lower turnover. This in beneficial since total transaction costs are reduced, Hence, the profitability of a dynamic portfolio allocation increases.

Altogether, we conclude that there are stark differences in the estimated DCC-GARCH models estimated by different estimation methods. The cellwise estimator is able to filter out outlying data cells, resulting in more robust estimates of the parameter values and the variance-covariance structure.

7 Discussion and Conclusion

This study introduces a novel estimator for the DCC-GARCH model. The estimator presented in this paper is robust against cells that contain outliers and utilizes the Mahalanobis distance to detect such outliers. Our findings demonstrate that this robust estimator outperforms both the MLE and the M-estimator proposed by Muler and Yohai (2008). Through simulated experiments, we observe that the cellwise estimator yields smaller biases and variances in parameter estimation compared to the other two estimators. Furthermore, the performance of the cellwise estimator improves as the magnitude of the outliers increases. However, it slightly deteriorates when the number of outliers is high and the cutoff values are not adjusted accordingly.

The empirical part of our research shows that using the cellwise estimator instead of MLE for the DCC-GARCH model brings about significantly different results. The constants in the univariate GARCH equations are estimated to be lower and the autoregressive property of the conditional variance appears to be more present. Moreover, it is demonstrated that the covariance structure modeled by the DCC-GARCH model is more stable when the cellwise estimator is used. Financial turmoil has a smaller effect on the covariance structure for the cellwise estimator than for MLE. MLE-estimated variances and covariances tend to explode when the market is unstable. Hence, it can be concluded that our estimator robust to cellwise outliers is an improvement over the naive MLE estimation method.

Considering these results, it is recommended to use both the MLE estimation method and the cellwise estimator simultaneously when employing the DCC-GARCH model in the presence of financial instability. If the results differ, it is likely that the MLE estimation is biased due to the presence of outliers.

However, there are limitations to consider in this research. Firstly, the simulation study only investigates the performance of the cellwise estimator in the presence of additive outliers. Hence, the behavior of this estimator remains untested in the presence of innovative outliers, which have a lasting impact even after their introduction. Considering the distinct characteristics of innovative outliers, it is expected that the cellwise estimator may not perform as effectively in such scenarios. Secondly, the empirical research of this paper is focused solely on three stocks in the S&P 500 within the DCC-GARCH framework. Therefore, the conclusions cannot be generalized to stocks from different sectors or countries, let alone other types of assets such as foreign currencies. Finally, the results obtained with the cellwise estimator depend on several user-defined settings. For example, the choice of the cutoff value used to determine outlying observations lacks a universally optimal setting. Determining the optimal parameter settings for the estimator requires a case-by-case evaluation by the user.

This paper successfully introduces a cellwise estimator that robustly estimates the parameters of the DCC-GARCH model. Building upon this success, future research could explore the possibility of developing similar estimators for different multivariate GARCH models. For instance, the widely used DCC-GARCH model presented by Tse and Tsui (2002) can be investigated before examining models such as the BEKK-GARCH model. The expansion of robust estimators for multivariate volatility models would not only facilitate performance comparisons but also provide a broader array of options for robust implementation of volatility models. Another avenue for future research is to further refine the imputation step employed in the cellwise estimator. In this study, we utilize an EM-algorithm for imputing values in the flagged cells. However, more advanced imputation techniques, such as multiple imputation, may enhance the performance of the cellwise estimator.

References

Aielli, G. P. (2013). Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics*, *31*(3), 282–299.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics*, 498–505.

Boudt, K., Danielsson, J., & Laurent, S. (2013). Robust forecasting of dynamic conditional correlation garch models. *International Journal of Forecasting*, *29*(2), 244–257.

Butler, R. W., Davies, P. L., & Jhun, M. (1993). Asymptotics for the minimum covariance determinant estimator. *The Annals of Statistics*, 1385–1400.

Carnero, M. A., Peña, D., & Ruiz, E. (2012). Estimating garch volatility in the presence of outliers. *Economics Letters*, *114*(1), 86–90.

Croux, C., & Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis*, *71*(2), 161–190.

Danilov, M. (2010). *Robust estimation of multivariate scatter in non-affine equivariant scenarios* (Doctoral dissertation, University of British Columbia).

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Donders, A. R. T., Van Der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*(10), 1087–1091.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, *20*(3), 339–350.

Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric theory*, *11*(1), 122–150.

Engle, R. F., & Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. *National Bureau of Economic Research Working Paper Series*, *8854*.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a garch (1, 1)? *Journal of Applied Econometrics*, *20*(7), 873–889.

Hubert, M., J., R. P., & Van den Bossche, W. V. D. (2019). Macropca: An all-in-one pca method allowing for missing values as well as cellwise and rowwise outliers. *Technometrics*, *61*(4), 459–473.

Lopuhaa, H. P., & Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, 229–248.

Mahalanobis, P. C. (1936). Mahalanobis distance. In *Proceedings national institute of science of india* (Vol. 49, pp. 234–256).

Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 71–91.

Muler, N., & Yohai, V. J. (2008). Robust estimates for garch models. *Journal of Statistical Planning and Inference*, *138*(10), 2918–2940.

Prykhodko, S., Prykhodko, N., Makarova, L., & Pukhalevych, A. (2018). Application of the squared mahalanobis distance for detecting outliers in multivariate non-gaussian data. In *2018 14th international conference on advanced trends in radioelecrtronics, telecommunications and computer engineering (tcset)* (pp. 962–965). IEEE.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, *8*(283-297), 37.

Rousseeuw, P. J., & Van Den Bossche, W. V. D. (2018). Detecting deviating data cells. *Technometrics*, *60*(2), 135–145.

Rousseeuw, P. J., & Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, *85*(411), 633–639.

Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, *20*(3), 351–362.

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959.

Young, A. (2022). Consistency without inference: Instrumental variables in practical application. *European Economic Review*, *147*, 104112.