LABARR, ARIC DAVID. Multivariate Robust Estimation of DCC-GARCH Volatility Model. (Under the direction of Dr Peter Bloomfield.)

Volatility estimation plays an important role in the fields of statistics and finance. Many different techniques address the problem of estimating volatilities of financial assets. Autoregressive conditional heteroscedasticity (ARCH) models and the related generalized ARCH models are popular models for volatilities. Multivariate approaches to GARCH models, such as Engle's Dynamic Conditional Correlation GARCH (DCC-GARCH), allow for estimation of multiple financial asset volatilities and covariances. However, the parameters of the DCC-GARCH model are typically estimated with Maximum Likelihood Estimation (MLE), which is greatly affected by outliers. Outliers in a DCC-GARCH model affect subsequent estimation of volatilities by the design of the model. These outliers may also affect volatility estimates of other financial assets within the same set of assets due to the correlated nature of the financial asset estimation.

This thesis reviews ARCH / GARCH modeling and robust estimation and proposes a robust estimation method for the DCC-GARCH model based on bounded deviance function estimation. This robust method of the DCC-GARCH model better estimates the volatilities of a set of financial assets in the presence of outliers. The thesis presents a study of the consistency of the robust method of the DCC-GARCH model along with simulation results to explore the characteristics of the robust method of the DCC-GARCH model estimation. For a better evaluation of the robust method, the thesis also examines the distribution structure of foreign exchange rate data. The thesis also discusses possible future topics and research in this field of study.
Multivariate Robust Estimation of DCC-GARCH Volatility Model

by
Aric David LaBarr

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2010

APPROVED BY:

Dr David Dickey

Dr Howard Bondell

Dr Denis Pelletier

Dr Peter Bloomfield
Chair of Advisory Committee
DEDICATION

I want to dedicate this dissertation to all of my friends and family who helped support me through all of my endeavors in getting my PhD. I especially want to thank my Lord and Savior Jesus Christ for giving me the strength and knowledge needed to complete such a large task.

The LORD is my shepherd; I shall not want.
He maketh me to lie down in green pastures: he leadeth me beside the still waters.
He restoreth my soul: he leadeth me in the paths of righteousness for his name’s sake.
Yea, though I walk through the valley of the shadow of death, I will fear no evil: for thou art with me;
thy rod and thy staff they comfort me.
Thou preparest a table before me in the presence of mine enemies: thou anointest my head with oil;
my cup runneth over.
Surely goodness and mercy shall follow me all the days of my life: and I will dwell in the house of the LORD for ever.
BIOGRAPHY

Aric D. LaBarr was born April 19, 1983 in Rochester, NY during a four-foot snowstorm. He is the first of two sons born to David and Julie LaBarr. During his early years, the family moved around a bit and his younger brother, Todd, was not born until 1990. However, Aric always held a special place in his heart for his younger brother, and despite the age difference, the two are very close.

Academically, Aric has always done well in school due to his mother’s influence and watchful eye. It was not until his high school math teacher, Ms. Kellog got a hold of him and really pushed Aric to excel did he start to realize his full potential. During his senior year at Southeast Raleigh High School, Aric was accepted to North Carolina State University. He entered the First-Year College program and from there decided to enter the Statistics department thanks to the influence of Dr. Bill Swallow. He double majored in Statistics and Economics and graduated with high honors in 2005. He continued at NCSU for his master’s degree in Statistics and eventually his Ph.D. As a graduate student, Aric spent much of his time as a teaching assistant. His advisor and many students noticed that he has a natural gift for teaching. This gift allowed him to explore an opportunity with the newly founded Institute of Advanced Analytics on NCSU’s Centennial Campus as a teaching assistant.

No story of Aric’s college career would be complete without mentioning how he failed the master’s qualifier exam after his first year in graduate school, but managed to pass the Ph.D. qualifier the very next day. Baffling his instructors and advisors, Aric was successful on the second test because he was more relaxed; he knew he had done poorly the day before and would have to take it again the following semester. He was successful on his second attempt on the master’s qualifier. It is because of this incident that the Statistics Department reorganized how the master’s/Ph.D. qualifier tests would be administered.

During his time at N.C. State, Aric met his wife Ashley, a North Carolina native and NCSU alumna. After meeting, the pair found out they actually had been so close to meeting several times throughout their lives. They had attended the same elementary and middle schools, in addition to having some of the same friends in high school. They married in May of 2007 and moved to Clayton, North Carolina.

The most critical part of Aric’s life is his faith. Never really brought up in a highly religious family, Aric was always curious about religion. He found his faith with a little help from Ashley. She invited him to attend church with her while they were dating and in their eyes, God took care of the rest. Aric has been involved in many of the youth ministries at their church including Upward basketball and teaching 5th grade Sunday School.

Aric looks to start a career in teaching at the Institute for Advanced Analytics following his time as a student at North Carolina State University. Wherever life takes him, he will always be a wolf at heart. GO PACK!
ACKNOWLEDGEMENTS

I would like to thank my advisor Dr Peter Bloomfield for his help. I also want to thank my advising committee of Dr David Dickey, Dr Howard Bondell, and Dr Denis Pelletier for all of their combined help as well.

I would like to thank the Statistics Department for my 9 years of education here and all of the great relationships I have developed over the years.

I would like to thank both Laura Ladrie and John Jerrigan at the Institute for Advanced Analytics for helping me set up the computing power needed to complete the simulation studies done in this thesis. I also want to thank Anthony Franklin for loaning me some of his computers as well to run simulations.

I want to thank my wife for being extremely understanding as I was sitting and typing for many long nights as she picked up my slack around the house. I also want to thank my family and friends for all of the emotional support given to me during this long and stressful process.
# TABLE OF CONTENTS

List of Tables ........................................................................................................... vii
List of Figures ............................................................................................................... viii

## Chapter 1 Introduction ......................................................................................... 1
  1.1 ARCH/GARCH Volatility Modeling ................................................................. 2
  1.2 Outliers and Robust Estimation ................................................................... 5
  1.3 Summary ......................................................................................................... 10

## Chapter 2 Literature Review ............................................................................... 11
  2.1 ARCH / GARCH Models .............................................................................. 11
    2.1.1 Extensions to Univariate ARCH / GARCH Models ............................ 13
    2.1.2 Multivariate ARCH / GARCH Models .............................................. 17
    2.1.3 Conditional Correlation Approach ....................................................... 19
  2.2 Generalized Robust Estimation ................................................................. 21
    2.2.1 Break-down Point Approach ................................................................. 23
    2.2.2 Deviance Robust Estimation ................................................................. 25
    2.2.3 Robust Estimation of Time Series Data .............................................. 27
    2.2.4 Multivariate Robust Estimation ........................................................... 34
  2.3 Robust Estimation of ARCH / GARCH ...................................................... 41
  2.4 Symmetry of Elliptical Distributions .......................................................... 48

## Chapter 3 Robust Estimation of the DCC-GARCH Model ............................. 53
  3.1 Outlier in General DCC-GARCH ................................................................. 53
  3.2 Robust Estimation of DCC-GARCH ............................................................ 57
    3.2.1 Proposed Robust Method ..................................................................... 57
    3.2.2 Consistency .......................................................................................... 59
  3.3 Initial Tests of Robust Method ..................................................................... 60

## Chapter 4 Data Driven Evaluation ................................................................. 70
  4.1 Multivariate Characterization of Data .......................................................... 70
  4.2 Data Driven Exploration ............................................................................. 76
    4.2.1 Data Driven Simulation and Results .................................................... 76
    4.2.2 Application to Foreign Exchange Rates ............................................ 80

## Chapter 5 Conclusion ....................................................................................... 88
  5.1 Summary ....................................................................................................... 88
  5.2 Conclusion .................................................................................................... 89
  5.3 Suggestions for Future Research ................................................................. 90

References ............................................................................................................... 91
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>$R_0$ Results with First Initial Correlation Matrix</td>
<td>55</td>
</tr>
<tr>
<td>3.2</td>
<td>Results with First Initial Correlation Matrix</td>
<td>57</td>
</tr>
<tr>
<td>3.3</td>
<td>$R_0$ Results with Second Initial Correlation Matrix</td>
<td>57</td>
</tr>
<tr>
<td>3.4</td>
<td>Results with Second Initial Correlation Matrix</td>
<td>58</td>
</tr>
<tr>
<td>3.5</td>
<td>Comparing Two Robust Methods with First Correlation Structure</td>
<td>61</td>
</tr>
<tr>
<td>3.6</td>
<td>Comparing Two Robust Methods with Second Correlation Structure</td>
<td>62</td>
</tr>
<tr>
<td>3.7</td>
<td>Results Across Generating Processes with First Set</td>
<td>65</td>
</tr>
<tr>
<td>3.8</td>
<td>Results Across Generating Processes with Second Set</td>
<td>69</td>
</tr>
<tr>
<td>4.1</td>
<td>Exchange Rate Volatility Parameter Estimation</td>
<td>71</td>
</tr>
<tr>
<td>4.2</td>
<td>Comparing $R^2$ Values of Q-Q Plots</td>
<td>74</td>
</tr>
<tr>
<td>4.3</td>
<td>$R^2$ Combinations for Contaminated Normal</td>
<td>74</td>
</tr>
<tr>
<td>4.4</td>
<td>Set 1 Results for Data Driven Comparison of Methods</td>
<td>78</td>
</tr>
<tr>
<td>4.5</td>
<td>Set 2 Results for Data Driven Comparison of Methods</td>
<td>79</td>
</tr>
<tr>
<td>4.6</td>
<td>Volatility Parameters Between Robust and Traditional Method</td>
<td>80</td>
</tr>
<tr>
<td>4.7</td>
<td>Correlation Parameters Between Robust and Traditional Method</td>
<td>81</td>
</tr>
<tr>
<td>A.1</td>
<td>Complete MSE Results Across Generating Processes with First Portfolio</td>
<td>97</td>
</tr>
<tr>
<td>A.2</td>
<td>Complete Bias Results Across Generating Processes with First Portfolio</td>
<td>98</td>
</tr>
<tr>
<td>A.3</td>
<td>Complete Variance Results Across Generating Processes with First Portfolio</td>
<td>99</td>
</tr>
<tr>
<td>A.4</td>
<td>Complete Results Across Generating Processes with Second Portfolio</td>
<td>100</td>
</tr>
<tr>
<td>A.5</td>
<td>Complete Bias Results Across Generating Processes with Second Portfolio</td>
<td>101</td>
</tr>
<tr>
<td>A.6</td>
<td>Complete Variance Results Across Generating Processes with Second Portfolio</td>
<td>102</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Likelihoods of $\alpha$ (a) and $\beta$ (b) without Split Likelihood Estimation</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Likelihoods of $\alpha$ (a) and $\beta$ (b) with Split Likelihood Estimation</td>
<td>6</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>QQ-plot of Spherical Symmetry</td>
<td>72</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Q-Q Plots of Squared Radii with $\chi^2_{1.36}$ (a) and $F_{2.75,3.48}$ (b)</td>
<td>75</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Overlaid Estimated Volatility Time Plots</td>
<td>82</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Estimated Volatility Time Plots with [a] MLE and [b] Robust Estimation</td>
<td>83</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Overlaid Estimated Volatility Time Plots</td>
<td>84</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Time Plots of Determinant of Estimated Correlation Matrices</td>
<td>86</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Time Plots of Average Correlation Between Assets</td>
<td>87</td>
</tr>
<tr>
<td>Figure B.1</td>
<td>Density Plots of Radii (a) and Squared Radii (b)</td>
<td>104</td>
</tr>
<tr>
<td>Figure B.2</td>
<td>Estimated Volatility Time Plots of Euro (MLE [a], Robust [b])</td>
<td>105</td>
</tr>
<tr>
<td>Figure B.3</td>
<td>Estimated Volatility Time Plots of Pound (MLE [a], Robust [b])</td>
<td>106</td>
</tr>
<tr>
<td>Figure B.4</td>
<td>Estimated Volatility Time Plots of Franc (MLE [a], Robust [b])</td>
<td>107</td>
</tr>
<tr>
<td>Figure B.5</td>
<td>Estimated Volatility Time Plots of Yen (MLE [a], Robust [b])</td>
<td>108</td>
</tr>
<tr>
<td>Figure B.6</td>
<td>Estimated Volatility Time Plots of Dollar (MLE [a], Robust [b])</td>
<td>109</td>
</tr>
<tr>
<td>Figure B.7</td>
<td>Estimated Correlation Time Plots</td>
<td>110</td>
</tr>
<tr>
<td>Figure B.8</td>
<td>Estimated Correlation Time Plots</td>
<td>111</td>
</tr>
<tr>
<td>Figure B.9</td>
<td>Estimated Correlation Time Plots</td>
<td>112</td>
</tr>
<tr>
<td>Figure B.10</td>
<td>Estimated Correlation Time Plots</td>
<td>113</td>
</tr>
<tr>
<td>Figure B.11</td>
<td>Estimated Correlation Time Plots</td>
<td>114</td>
</tr>
</tbody>
</table>
Volatility modeling plays a critical role in mathematical finance and statistical applications. The ability to estimate and forecast volatilities for different assets and groups of assets leads to a better understanding of current and future financial risk. Many different methods of volatility estimation have been developed over the past few decades. Understanding the characteristics of financial assets helps develop estimation methods for volatilities. Studying volatilities of financial assets reveals that volatility seems to vary over time instead of remaining constant. The volatilities also exhibit some persistence, or dependence over time, with a clustering effect of small (or large) number of returns being followed by small (or large) number of returns of either sign.

Let $r_t$ be the daily return of a financial asset, modeled by $r_t = \sqrt{h_t} \varepsilon_t$ and $\varepsilon_t$ as the random error with the variance of 1. These daily returns are defined as the logarithm of the relative prices $\log(P_t/P_{t-1})$ with $P_t$ as the current time period price in dollars. It is reasonable to assume daily returns have a conditional mean of approximately zero. This assumption is reasonable because an extremely high annual return of 25% translates to a daily return of only 0.09%. Time weighted estimates are a reasonable initial guess of volatility estimation to account for persistence:

$$h_t = \sum_{i=1}^{n} \omega_i r_{t-i}^2$$

with $\sum_{i=1}^{n} \omega_i = 1$ and the more recent observations more heavily weighted. A problem with this approach is that many different weights $\omega_i$ need defining. An exponential weight scheme where $\omega_i = \lambda \omega_{i-1}$, with $\lambda$ between 0 and 1, potentially solves the problem. Other estimates have also been proposed.
1.1 ARCH/GARCH Volatility Modeling

Instead of a weighting scheme, Engle (1982) uses an autoregressive time series approach to account for persistence in volatility estimation. He also assumes the conditional variance, or volatility, of returns varies over time. Engle’s autoregressive conditional heteroscedasticity (ARCH) model defines the volatility as

$$h_t = a_0 + a_1 r_{t-1}^2,$$  (1.1)

with $a_0 > 0$ and $a_1 \geq 0$ for the volatility to remain positive. Engle assumes the returns, $r_t$ given $\psi_{t-1}$, follow a normal distribution: $r_t | \psi_{t-1} \sim N(0, h_t)$, where $\psi_{t-1}$ represents all information up to time $t - 1$. This model easily extends to $p$ lags of returns with the ARCH(p) model

$$h_t = a_0 + \sum_{i=1}^{p} a_i r_{t-i}^2,$$  (1.2)

with $a_0 > 0$ and $a_i \geq 0$ for all $i$. However, in practice, we often need large values of $p$ to accurately model real world data.

Bollerslev (1986) avoids the problem of large values of $p$ in Engle’s ARCH model by generalizing the ARCH(p) model into the generalized autoregressive conditional heteroscedasticity (GARCH) model, in much the same way as an autoregressive (AR) model extends to the autoregressive moving average (ARMA) model. The GARCH model allows a longer memory process with more flexibility. The GARCH(p,q) model still assumes normality with $r_t | \psi_{t-1} \sim N(0, h_t)$, but instead of equation 1.2, $h_t$ is defined as

$$h_t = a_0 + \sum_{i=1}^{p} a_i r_{t-i}^2 + \sum_{i=1}^{q} b_i h_{t-i},$$  (1.3)

with $a_0 > 0, a_i \geq 0$, and $b_i \geq 0$. In many cases, $p = q = 1$ is found to give an adequate fit. The univariate ARCH/GARCH framework of models has been adapted into many different forms, which are detailed in Section 2.1.1.

The ARCH/GARCH framework of models also extends into a multivariate context, to model the underlying volatilities and correlations between different market assets. The general multivariate extension to the GARCH model has a vector of assets as a stochastic process $r_t$ of $k \times 1$ dimension defined as

$$r_t = H_t^{1/2} \varepsilon_t,$$  (1.4)

where $H_t^{1/2}$ is a factor of the conditional variance-covariance matrix of size $k \times k$, and with $\text{Var}(\varepsilon_t) = I_k$. Bollerslev et al. (1988) model $H_t$ as

$$\text{vech}(H_t) = c + \text{Avech}(\varepsilon_t \varepsilon_t') + \text{Bvech}(H_t),$$  (1.5)
where vech(·) is the operator that is a column-wise vectorization of the of the lower triangular portion of a matrix, and the matrices A and B are parameter matrices. This specification of the $H_t$ matrix is referred to as the VEC model. The number of parameters in this model grows very quickly as the number of assets in the model grows. To make parameter estimation feasible, Bollerslev, Engle, and Wooldridge proposed to restrict A and B to diagonal matrices. Models with other specifications of $H_t$, such as the BEKK(1,1,K) and Factor GARCH, are described in Section 2.1.2. Most of these approaches involve many parameters to be estimated, which leads to computational burdens for large portfolios of assets.

A less computationally burdensome approach to multivariate GARCH estimation is a combination of univariate estimation of GARCH models and estimation of multivariate correlation matrices. This greatly reduces the number of parameters by separately defining individual conditional variance structures and an overall correlation structure. Bollerslev (1990) designed one of these approaches with the constant conditional correlation GARCH (CCC-GARCH). The CCC-GARCH defines the conditional covariance matrix of returns as

$$H_t = D_t R D_t, \quad D_t = \text{diag}(\sqrt{h_{i,t}}),$$

(1.6)

where $R$ is a correlation matrix containing conditional correlations, and $h_{i,t}$ follows the univariate GARCH model, defined as

$$h_{i,t} = a_{i,0} + \sum_{p=1}^{P_i} a_{i,p} r_{i,t-p}^2 + \sum_{q=1}^{Q_i} b_{i,q} h_{i,t-q}.$$

(1.7)

The conventional sample correlation matrix is a reasonable estimate of $R$. However, in practice the assumption that correlations of assets remain constant over time seems unrealistic. In particular, the constant correlations assumption understates risk if correlations increase in turbulent markets.

Engle (2002) relaxes the assumption of constant correlation in the dynamic conditional correlation GARCH (DCC-GARCH) by allowing the correlation matrix to change over time. This model is widely used for its combination of computational ease as well as the evolving correlation matrix defined by

$$H_t = D_t R_t D_t,$$

(1.8)

Engle mentions two different estimates for the $R_t$ matrix. The first specification involves exponential smoothing with

$$Q_t = (1 - \lambda) \left( \varepsilon_{t-1} \varepsilon_{t-1}^\prime \right) + \lambda Q_{t-1},$$

(1.9)

where $Q_t$ is the positive definite covariance matrix and $R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}$. Another
method uses a GARCH(1,1) model as a specification with

\[ Q_t = R_0(1 - \alpha - \beta) + \alpha(\epsilon_{t-1}\epsilon'_{t-1}) + \beta Q_{t-1}, \]  

(1.10)

with \( R_0 \) as the unconditional correlation matrix and \( \alpha + \beta < 1 \).

This leads to the following specification of the DCC-GARCH model:

\[ r_t | \psi_{t-1} \sim N(0, D_t R_t D_t), \]
\[ D_t^2 = \text{diag}(a_{0,i}) + \text{diag}(a_{1,i}) \odot r_{t-1} r'_{t-1} + \text{diag}(b_{1,i}) \odot D_{t-1}^2, \]
\[ \epsilon_t = D_t^{-1} r_t, \]
\[ Q_t = R_0(1 - \alpha - \beta) + \alpha(\epsilon_{t-1}\epsilon'_{t-1}) + \beta Q_{t-1}, \]
\[ R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}, \]

(1.11)

where \( \odot \) represents the elementwise product of the matrices. The log likelihood we would maximize to estimate the parameters of the model is

\[ L = -\frac{1}{2} \sum_{t=1}^{T} \left[ n \log(2\pi) + 2\log|D_t| + r'_t D_t^{-1} D_t^{-1} r_t - \epsilon'_t \epsilon_t + \log|R_t| + \epsilon'_t R_t^{-1} \epsilon_t \right], \]  

(1.12)

as shown in detail in Appendix A. Maximizing this function over the parameters leads to the maximum likelihood estimates (MLE) of the parameters. Engle suggests splitting the likelihood into the sum of two parts to improve efficiency in calculating the model. The two components are the volatility component, which only depends on the individual GARCH parameters, and the correlation component, which depends on both the correlation parameters and the individual GARCH parameters. Let \( \theta \) denote the volatility parameters in the \( D \) matrix and \( \phi \) denote the correlation parameters in the \( R \) matrix. The split is written

\[ L(\theta, \phi) = L_V(\theta) + L_C(\theta, \phi), \]  

(1.13)

with the volatility part as

\[ L_V(\theta) = -\frac{1}{2} \sum_{t=1}^{T} \left[ n \log(2\pi) + 2\log|D_t| + r'_t D_t^{-1} D_t^{-1} r_t \right], \]  

(1.14)

and the correlation part as

\[ L_C(\theta, \phi) = -\frac{1}{2} \sum_{t=1}^{T} \left[ \log|R_t| + \epsilon'_t R_t^{-1} \epsilon_t - \epsilon'_t \epsilon_t \right]. \]  

(1.15)

Engle first estimates the volatility parameters with ML estimation. He then places the estimates into the correlation portion of the likelihood to estimate the correlation parameters with ML estimation.
Figure 1.1: Likelihoods of $\alpha$ (a) and $\beta$ (b) without Split Likelihood Estimation

After comparing estimates with both whole ML estimation and ML estimation across parts, the lack of efficiency mentioned above arises from problems in the estimation of the correlation parameters. A set of three assets is evaluated with a DCC-GARCH model with correlation parameters $\alpha = 0.24$ and $\beta = 0.7$ over 500 periods in time with an initial correlation matrix $R_0$ defined as

$$
R_0 = \begin{pmatrix}
1 & 0.85 & 0.85 \\
0.85 & 1 & 0.85 \\
0.85 & 0.85 & 1
\end{pmatrix}.
$$

A likelihood function of both the $\alpha$ and $\beta$ parameters shows instability as shown in Figure 1.1. The graph of the likelihood functions above show the value of the likelihood function for changing values of a single parameter as the other parameters in the model are held constant.

The instability in the estimation of these parameters poses a problem in trying to derive conclusions about the model. The likelihoods for the correlation parameters show more stability when imposing the technique of breaking the likelihood into two pieces, as seen in Figure 1.2. The immense improvement in the stability of the estimation of these parameters allows for better estimation of the model parameters. With the split estimation technique, we can better examine the effects of outliers in the maximum likelihood estimation of the DCC-GARCH.

1.2 Outliers and Robust Estimation

Observations that deviate from the general pattern of the data, called outliers, affect the accuracy of standard techniques of analysis and estimation. Outliers reduce the ability of classic techniques
such as sample mean, sample variance, sample correlation, and regression modeling to estimate parameters in the data. Many different robust estimation techniques have been used to estimate models from data when outliers are or are not present. Good robust estimates provide accurate estimation of parameters in the presence or absence of outliers.

Outliers also affect maximum likelihood estimation (MLE). This is a common method of model parameter estimation used in estimating the parameters in the DCC-GARCH model. Define $L(\beta | y_i, \ldots, y_n)$ as the log of the likelihood function for the random variable $y_i$. We derive the MLE by solving either

$$\hat{\beta}_{\text{MLE}} = \arg \max_{\beta} L(\beta | y_i, \ldots, y_n),$$

or taking the derivative of the log of the likelihood function and solving

$$\sum_{i=1}^{n} \ell(\beta | y_i) = 0, \quad (1.16)$$

where $\ell(\cdot)$ is the derivative of the log of the likelihood function. Under certain regularity conditions, $\hat{\beta}_{\text{MLE}}$ is both consistent and asymptotically normally distributed. These maximum likelihood estimators are a specific subset of a general class of estimators called M-estimators developed by Huber (1964). These estimates, $\hat{\beta}_\psi$, are solutions to

$$\sum_{i=1}^{n} \psi(y_i, \beta) = 0, \quad (1.17)$$

instead of equation 1.16. Assume the function is unbiased — defined as $E_{\beta} [\psi(y_i, \beta)] = 0$. Then, under
certain regularity conditions defined by Huber (1973), \( \hat{\beta} \) is consistent and asymptotically normal. Different choices of the \( \psi \) function lead to robust estimation of the parameters. These choices are discussed further in Section 2.2.

It is beneficial to understand outliers in time dependent data because the DCC-GARCH model uses an autoregressive structure in estimating the volatilities and correlations. Time series data, such as financial data, contain greater potential for outliers hindering the estimation process, because of the underlying dependence between observations in the data. In time series data, outliers may occur in patches throughout the data, or in isolation throughout the data. In some cases, entire shifts of the process may occur. A couple of common outliers that may occur in time series data are additive outliers (AO) and innovation outliers (IO).

Maronna et al. (2006) describes additive outliers as outliers where instead of the expected observation \( y_t \), it is replaced by \( y_t + \nu_t \) where \( \nu_t \sim (1 - \epsilon)\delta_0 + \epsilon N(\mu_\nu, \sigma^2_\nu) \). In this definition, \( \delta_0 \) is a point mass distribution at zero and the \( \sigma^2_\nu \) is significantly greater than the variance of \( y_t \). This creates an outlier with probability \( \epsilon \), and \( n \) consecutive outliers with probability \( \epsilon^n \). They defined innovation outliers as outliers that affect both the current and subsequent observations. These outliers are especially relevant in autoregressive (AR) and autoregressive moving average (ARMA) models.

Innovation outliers occur in the error term of the model. In this case, the observation \( y_t \) is actually affected, as shown with a simple AR(1) model:

\[
y_t = \phi y_{t-1} + \epsilon_t.
\]

If \( \epsilon_t \) comes from either a distribution with larger tails than a normal distribution or a mixture of two normal distributions, then \( y_t \) becomes an outlier that directly affects \( y_{t+1} \) by

\[
y_{t+1} = \phi \hat{y}_t + \epsilon_t,
\]

where \( \hat{y}_t \) is the value of \( y_t \) where \( \epsilon_{t-1} \) is not from a normal distribution. Both additive and innovation outliers potentially bias the estimate along with changing the estimate’s variability.

Many different methods have been proposed to handle outliers in time series data, such as robust filter estimation for AR and ARMA models or M-estimation techniques both of which are described in Maronna et al. (2006). These are just some of the different possible approaches, and more are discussed in detail in Section 2.2.3. A robust filter approach replaces prediction residuals \( \hat{\epsilon}_t \) with robust prediction residuals \( \tilde{\epsilon}_t \) by replacing outliers by robust filtered values. Instead of the typical residual definition in an AR(p) model

\[
\hat{\epsilon}_t = (y_t - \mu) - \phi_1(y_{t-1} - \mu) - \ldots - \phi_p(y_{t-p} - \mu),
\]
the new robust residuals are defined by

\[ \tilde{\varepsilon}_t = (y_t - \mu) - \phi_1(\tilde{y}_{t-1} - \mu) - \ldots - \phi_p(\tilde{y}_{t-p} - \mu), \]

where the \( \tilde{y}_t \) are a filtered prediction of the observation, which are approximations to the expected value of that observation. The value \( \tilde{y}_t \) is equal to \( y_t \) if the value does not fall outside a predetermined range from the expected value at that point. If \( \tilde{y}_t \) falls outside the range, then the estimate is equal to an approximation of the expected value at that point, given previous information. This robust filtering approach extends into the ARMA class of models as well. Although this estimation works well for AO, the process does not work as well in the presence of IO.

M-estimation techniques for ARMA models minimize

\[ \sum_{t=p+1}^{T} \rho \left( \frac{\hat{\varepsilon}_t(\beta)}{\hat{\sigma}_\varepsilon} \right), \]

where the \( \psi \) appears in equation 1.17 may be the derivative of the \( \rho \) function appearing here, \( \hat{\varepsilon}_t \) are the model residuals, and \( \hat{\sigma} \) is a robust estimate of the standard deviation of the residuals. As mentioned before, the M-estimation may be implemented using various \( \psi \) functions, such as Yohai (1987) MM-estimate, to help limit effects of outliers. Yohai uses three steps for MM-estimation, where he first computes an initial estimate of \( \hat{\beta} \). From this estimate, he computes a robust scale estimate, \( \hat{\sigma} \) for the residuals. He then uses an iterative process to continue the previous two steps until convergence. These estimates are relatively robust for contamination with AO, but lose their effectiveness as the order \( p \) of the AR(\( p \)) process increases. However, asymptotic theory for M-estimates is based on the assumption that the errors in the model are homoscedastic. The DCC-GARCH model is heteroscedastic by construction. Although some M-estimates do not depend on homoscedastic errors, they have lower efficiency than those that account for the lack of homoscedasticity.

An improved M-estimator without homoscedasticity takes into account the other covariates and possible parameters making the errors heteroscedastic by

\[ y_t = \beta'x_t + g(\xi, \beta'x_t)\varepsilon_t, \]

where \( \xi \) is a parameter vector limited to the error variance. To obtain robust estimates of both sets of parameters, Maronna et al. (2006) suggest computing an initial estimate of the parameter vector \( \beta \) by the proposed above MM-estimate. The residuals of this model are then calculated and used in the computation of an estimate of the parameter vector \( \xi \). From here, robust transformations of the original \( y_t \) and \( x_t \) are calculated by dividing through by the estimated \( g(\cdot) \) function, to produce a more accurate estimate of \( \beta \). The process continues in iterations until reasonable estimates are obtained.

The heteroscedastic errors affect not only univariate estimation, as in the case of the GARCH
parameters in the DCC-GARCH, but also the multivariate estimation of the correlation structure between variables. The DCC-GARCH model requires the estimation of a covariance matrix to describe the relationship between the multiple assets in the portfolio. White (1980) noted that heteroscedasticity not only hinders linear model parameter estimation, but also hinders covariance matrix estimation. He proposes an estimate of the covariance matrix that is not unduly affected by the presence of heteroscedasticity and does not require a specific model of the heteroscedasticity. He assumes that the errors in the model have heteroscedasticity of the form $E(\epsilon_t^2|x_t) = g(x_t)$. Under some basic moment assumptions of the errors in the model, White develops the estimator

$$\hat{\epsilon}_t = 1/n \sum_{t=1}^{n} \hat{\epsilon}_t^{2, MLE} x_t' x_t,$$

where $\hat{\epsilon}_t^{2, MLE}$ are the residuals evaluated with the parameters at the MLE values. Using the previous estimator, the heteroscedasticity-robust covariance matrix is

$$\hat{\Sigma}_R = \left(\frac{x_t' x_t}{n}\right)^{-1} \hat{V}_n \left(\frac{x_t' x_t}{n}\right)^{-1}. \quad (1.18)$$

Outliers also affect covariance matrix estimation. Some proposed robust multivariate estimates of the covariance matrix are computationally burdensome in high dimensional data, such as some financial data. Robust estimation of location and scale using Mahalanobis distances computed from M-estimators are computationally difficult, according to Peña and Prieto (2001). They state that the minimum covariance determinant (MCD) by Rousseeuw (1984) is also computationally intensive. The purpose of the MCD method is to take $h$ observations from the total that have the lowest determinant of the covariance matrix. The MCD estimate of the covariance matrix is just a multiple of these points’ covariance matrix. For this process to work, many iterations of resampling must take place, which lead Rousseeuw and Van Driessen (1999) to create the FAST-MCD algorithm, explained in full detail in Hubert et al. (2008).

Peña and Prieto again suggest that even the FAST-MCD algorithm requires too much resampling and reduces heavy computation time with needed approximations. They suggest that outliers in multivariate data created by a symmetric contamination tend to increase the kurtosis coefficient. The directions of the projected observations, based on kurtosis coefficients, lead to a better idea of which directions contain outliers. They create an algorithm based on these projected kurtosis directions. The details of the algorithm are contained in Peña and Prieto (2001). Other multivariate estimators with outliers are discussed further in Section 2.2.4.
1.3 Summary

The above conclusions about the effects of outliers in autoregressive models, models with heteroscedasticity, and covariance matrix estimation, show the DCC-GARCH model is inherently hindered by outliers. This thesis proposes a robust estimation method for the DCC-GARCH that accounts for outliers present in the data. The second chapter is a literature review of the explored papers and topics in ARCH/GARCH modeling in both the univariate and multivariate context, robust estimation in univariate, multivariate, and time series data, previous attempts of ARCH/GARCH robust estimations, and tests of symmetry for elliptical distributions. Chapter 3 proposes the robust estimation method for the DCC-GARCH method and shows an example of outliers hindering the DCC-GARCH model while discussing the creation and asymptotics of the new robust estimation method. Chapter 4 discusses the attempts to identify real world data distributions for a data driven evaluation of the newly proposed model. Chapter 4 also summarizes the results of simulation studies comparing the maximum likelihood fitting of the DCC-GARCH model with the newly proposed robust method, and displays results of fitting the robust method to foreign exchange rate data. Chapter 5 concludes with a summary of the results along with possible areas of future research in the field.
CHAPTER 2

Literature Review

This chapter reviews the past and current literature on univariate and multivariate ARCH/GARCH modeling, robust estimation techniques, and tests of elliptical symmetry.

2.1 ARCH / GARCH Models

The following section will briefly revisit Engle’s ARCH model as mentioned in Section 1.1. Consider a random variable $r_t$ drawn from a conditional distribution of $f(r_t | \psi_{t-1})$, where $\psi_{t-1}$ is all information up until time $t - 1$. The current period forecast of $r_t$, after some basic assumptions, is the conditional expected value given the previous period’s information, $E(r_t | \psi_{t-1})$. Similarly, the variance of the current period forecast is the conditional variance, $\text{Var}(r_t | \psi_{t-1})$. However, traditional econometric models did not take the previous period’s response of $r_{t-1}$ into the calculation, by assuming constant conditional variance. Engle (1982) proposed a model called the autoregressive conditional heteroscedasticity (ARCH) model, allowing the underlying forecast variability to change over time.

Modeling of heteroscedastic variances allows variances to change and evolve over time. Standard heteroscedasticity corrections to predicting variances introduce an exogenous variable $x_t$ to the calculation as

$$r_t = \epsilon_t x_{t-1}$$

with $E(\epsilon_t) = 0$ and $\text{Var}(\epsilon_t) = \sigma^2$. This leads to $\text{Var}(r_t) = \sigma^2 x_{t-1}^2$. Although this variance changes over time, the variance depends on the changes of the exogenous variable instead of the possible evolution of the response variable's conditional variance on previous periods. Replacing the exogenous variable with the previous period response variable, $r_{t-1}$, leads to the general bilinear model of Granger and
Andersen (1978). This model allows the evolution of the conditional variance based on changes to the response variable, but leads to an unconditional variance of either zero or infinity.

Engle’s ARCH model replaces the bilinear model with the following form

\[ r_t = \varepsilon_t h_t^{1/2} \]

\[ h_t = \alpha_0 + \alpha_1 r_{t-1}^2, \quad (2.1) \]

with \( \text{Var}(\varepsilon_t) = 1 \). Engle assumes the normality of \( r_t \) given all of the information at the previous time period, \( \psi_{t-1} \), with

\[ r_t | \psi_{t-1} \sim N(0, h_t). \]

The ARCH model extends to an order of \( p \), with the ARCH(p) model only differing from the ARCH model through the function

\[ h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2, \quad (2.2) \]

with the \( \alpha_i \)’s restricted to positive values as defined in equation 1.2. Engle (1982) describes details of the distribution of the ARCH process and importantly notes that the unconditional distribution of the error possesses fatter tails than the normal distribution. Also, the ARCH(p) model possesses an arbitrarily long lag structure when it is applied to real life data.

As mentioned in Section 1.1, Bollerslev (1986) corrects the problem of the arbitrarily long lag structure of the conditional variance of the ARCH(p) model by generalizing the ARCH(p) model into the general autoregressive conditional heteroscedasticity (GARCH) model. Bollerslev saw the long lag structure of the ARCH(p) model as potentially burdensome. The GARCH model still assumes normality with

\[ r_t | \psi_{t-1} \sim N(0, h_t), \]

but instead of \( h_t \) defined as in equation 2.2, \( h_t \) is defined as

\[ h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}, \quad (2.3) \]

with \( \alpha_0 > 0, \alpha_i \geq 0, \) and \( \beta_i \geq 0 \). The GARCH model with \( q = 0 \) simplifies to an ARCH(p) model.

Bollerslev focuses on the simplest GARCH model, the GARCH(1,1) model. Bollerslev calculates the moments of the GARCH(1,1) model to find information about the distribution of the process. The detailed calculations are found in Bollerslev’s paper, but the important finding is that the second and fourth order moments exist and are given by

\[ \text{E}(r_t^2) = \alpha_0 (1 - \alpha_1 - \beta_1)^{-1} \]
and

\[ E(r_t^4) = \frac{3\alpha_0^2 (1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)} \]

respectively, with the assumption of \(3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1\) for the fourth moment to be finite. With these two moments, the kurtosis of the distribution of the GARCH(1,1) process is

\[ \kappa = \frac{E(r_t^4) - 3E(r_t^2)^2}{E(r_t^2)^2} \]

\[ = \frac{6\alpha_1^2 (1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)^{-1}}. \]

Coupled with the assumption \(3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1\), the excess kurtosis of the distribution is strictly positive, which leads to a heavy tailed distribution. This is similar to the findings of Engle with the ARCH process. Many extensions of the ARCH/GARCH framework have been proposed since their creation.

### 2.1.1 Extensions to Univariate ARCH / GARCH Models

Both ARCH and GARCH models may be used to account for the presence of volatility clustering in time series such as financial data, where periods of high (or low) volatility are typically followed by further periods of high (or low) volatility. Another typical aspect of financial data is that the unconditional distribution of the returns tends to have fatter tails than the normal distribution. Although ARCH and GARCH models with conditional normal errors possess unconditional error distributions with fatter tails than the normal distribution, the residuals in these models still often exhibit leptokurtosis. Some of the first extensions of the ARCH/GARCH modeling account for the leptokurtosis in the residuals of these models.

Bollerslev (1987) proposes one of the first extensions of the ARCH/GARCH modeling system by making an adjustment to the conditional distribution of the error term. Bollerslev notes the usefulness of the ARCH and GARCH models in portraying the clustering of volatilities in financial data. However, he also notes that financial data are conditionally leptokurtic. Therefore, Bollerslev proposes to switch the conditional error distribution to a \(t\)-distribution. The new \(t\)-distributed GARCH(p,q) he proposes is given by

\[ r_t | \psi_{t-1} \sim f_v(e_t | \psi_{t-1}) \]

\[ = \Gamma\left(\frac{\nu + 1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)^{-1}((\nu - 2)h_t)^{-1/2} \]

\[ \times (1 + e_t^2 h_t^{-1}(\nu - 2)^{-1})^{-(\nu + 1)/2}, \]

with \(\nu > 2\) and \(h_t\) defined in equation 2.3. Bollerslev estimates the degrees of freedom from the \(t\)-distribution, \(\nu\), along with the other parameters in the model. The GARCH(1,1)-\(t\) has both the

13
cluster volatility aspects of the GARCH model and a higher leptokurtosis than the GARCH(1,1) with conditional normally distributed errors.

Nelson (1991) notices three different problems with the GARCH model which Bollerslev proposes for asset pricing applications. Nelson notices that the GARCH model accounts for only the magnitude of the volatility of the previous period and not whether the shift in volatility is up or down. This goes against research that shows a negative correlation between present return and future return volatilities. He also notices the GARCH model is potentially too restrictive on the parameter estimates. Lastly, Nelson shows that interpretation of volatility persistence in GARCH models is difficult. Nelson proposes the Exponential GARCH (EGARCH) model, where the log of the volatilities is an asymmetric function of past returns given by

\[ r_t = h_t^{1/2} \epsilon_t \]

\[ \log(h_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i (\phi \epsilon_{t-i} + \gamma |\epsilon_{t-i}| - E(|\epsilon_{t-i}|)) + \sum_{i=1}^{p} \beta_i \log(h_{t-i}). \] (2.4)

Nelson considers a more general family of distributions for the error term instead of the normal distribution. Nelson uses the Generalized Error Distribution (GED), which is normalized to have a mean of zero and variance of one, for the error distribution. This normalized GED is given by

\[ f(\epsilon) = \frac{\nu e^{-(1/2)|\epsilon/\nu|^\nu}}{\lambda 2^{1+1/\nu} \Gamma(1/\nu)} \]

\[ \lambda \equiv \left[ 2^{-2/\nu} \frac{\Gamma(1/\nu)}{\Gamma(3/\nu)} \right]^{1/2}, \]

with \( \nu \) a factor that determines the thickness of the tails of the distribution. If \( \nu = 2 \), the error follows a normal distribution. For values of \( \nu < 2 \), the distribution has thicker tails than the normal. For values of \( \nu > 2 \), the distribution has thinner tails than the normal distribution.

The EGARCH model possesses fewer restrictions on the parameters of \( \alpha_i \) and \( \beta_i \). Therefore, if \( \alpha_i \phi < 0 \), the model allows for asymmetry in the variances so the volatility tends to rise (or fall) when \( \epsilon_{t-i} \) is negative (or positive). Nelson notes that the EGARCH outperforms the traditional GARCH model in asset pricing applications. Nelson also portrays the results of the EGARCH as easier to interpret than the traditional GARCH.

Jorion (1988) (see also Bollerslev et al. 1992) focuses his research on foreign exchange markets and the examination of discontinuities in the data. Jorion believes that discontinuities in financial data lead to leptokurtosis in their unconditional distribution. Hopes of accounting for discontinuities lead Jorion to combine jump-diffusion processes and ARCH models. He defines \( r_t \) as the logarithm of relative prices \( \log(P_t/P_{t-1}) \) with \( P_t \) as the current time period price in dollars of the foreign currency.
He proposes that prices follow a diffusion process given by

$$\frac{dP_t}{P_t} = \alpha dt + \sigma d\epsilon_t,$$  \hspace{1cm} (2.5)

which leads to a discrete time representation defined by

$$r_t = \mu + \sigma \epsilon_t.$$  

This model is equivalent to Engle’s structure of $r_t$ with an included mean term $\mu$.

However, this model does not account for discontinuities in the data. Therefore, Jorion alters equation 2.5 into a mixed jump-diffusion model given by

$$\frac{dP_t}{P_t} = \alpha dt + \sigma d\epsilon_t + dq_t,$$  \hspace{1cm} (2.6)

where $dq_t$ is a Poisson process characterized with a mean number of jumps with a jump size of $Y$ per unit of time. He assumes the size of the jumps has a lognormal distribution. With this alteration to the process, the new function for returns is

$$r_t = \mu + \sigma \epsilon_t + \sum_{i=1}^{n_t} \log(Y_i),$$  \hspace{1cm} (2.7)

where $n_t$ is the actual number of jumps occurring in the interval.

Jorion combines equation 2.7 with Engle’s ARCH(1) model defined in equation 2.1 to get the following model

$$r_t = \mu + h_t^{1/2} \epsilon_t + \sum_{i=1}^{n_t} \log(Y_i)$$

$$h_t = \alpha_0 + \alpha_1 (r_{t-1} - \mu)^2.$$  \hspace{1cm} (2.8)

Jorion’s model accounts for discontinuities along with the clustering of volatilities present in financial data. In comparison with the ARCH(1) model and the diffusion process, the jump-diffusion model provides a lower Schwarz Criterion defined in Schwarz (1978). This points to the fact that the jump-diffusion process better represents the data compared to the other two models.

Hsieh (1989) agrees that the EGARCH specification can be better interpreted than the traditional GARCH model. However, Hsieh compares the distributional extensions of Bollerslev, Nelson, and Jorion with his own extension in five different foreign exchange rate markets. Hsieh uses the EGARCH model with four different conditional distributions. He compares Bollerslev’s traditional normal distribution approach, Bollerslev’s t-distribution approach, Nelson’s GED approach, Jorion’s normal-Poisson approach, and his own normal log-normal mixture distributional approach. Hsieh uses the
following conditional distribution for the error term:

\[ f(\varepsilon) = \int_{-\infty}^{\infty} e^{1/4\xi^2 - 1/2\xi u} \phi \left( \varepsilon e^{1/4\xi^2 - 1/2\xi u} \right) \phi(u) \, du \]

Hsieh analyzes goodness of fit test statistics for each of the distributions. The goodness of fit test statistics reject the normal distribution for all five of the foreign exchange rates analyzed. However, the goodness of fit test statistics do not reject only the normal log-normal mixture Hsieh proposes in all of the models. The t-distribution and the normal-Poisson mixture distribution are not rejected in four of the currencies, while the GED is not rejected in only three of the currencies.

Engle et al. (1987) note that the conditional mean may also depend on the previous variances of the data. Holding risky assets requires compensation that directly corresponds to the amount of risk in the assets. Engle, Lilien, and Robins develop the ARCH-in-Mean, or ARCH-M, model where the conditional mean is a function of previous variances and possibly other covariates. The function of the returns is defined

\[ r_t = g(x_{t-1}, h_t; \beta) + h_t^{1/2} \varepsilon_t \]  
(2.9)

\[ h_t = a_0 + \sum_{i=1}^{p} \alpha_i (r_{t-i} - g(x_{t-i}, h_t; \beta))^2, \]

with \( g(\cdot) \) commonly a linear or logarithmic function of \( h_t \) and \( x_{t-1} \) a vector of covariates. This model allows a change in the variance of an asset to directly affect the price of the asset either positively or negatively. The ARCH-M model accounts for financial theory that directly relates the trade-off of risk and return of assets. However, Bollerslev et al. (1992) note that consistency of the ARCH-M model requires the correct specification of the model.

Engle and Bollerslev (1986) restructure the GARCH model specified in equation 2.3 to that of a stationary ARMA time series process. They rearrange the GARCH model to

\[ r_t^2 = \sum_{i=1}^{p} (\alpha_i + \beta_i)r_{t-i}^2 - \sum_{j=1}^{q} \beta_j v_{t-j} + v_t, \]  
(2.10)

where \( v_t = r_t^2 - h_t \) is a sequence of uncorrelated random variables. This model has the same correlation structure as an ARMA(p,q) process with AR parameters \( (\alpha_i + \beta_i) \) and the MA parameters \(-\beta_j\).

With the assumption of \( p \geq q \), the above model is stationary if \( \sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i \leq 1 \). However, Engle and Bollerslev note the possibility of a unit root in the GARCH process when \( \sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i = 1 \). If the unit root is present, then the model becomes the Integrated GARCH (IGARCH) model. The IGARCH(1,1) model is defined

\[ h_t = a_0 + \alpha_1 r_{t-1}^2 + (1 - \alpha_1)h_{t-1}, \]
which closely resembles a random walk model with drift because
\[ E(h_{t+1}) = sa_0 + h_t. \]

Engle and Bollerslev mention the difficulties in testing for the presence of persistence in the IGARCH model.

### 2.1.2 Multivariate ARCH / GARCH Models

A multivariate framework leads to more applicable models compared to the univariate approach when studying the relationships between volatilities in multiple assets at the same time. With a multivariate approach, the multivariate distribution directly computes the implied distribution of portfolios compared to single assets. The specification of the multivariate GARCH model has a vector stochastic process \( r_t \) of \( k \times 1 \) dimension. Bauwens et al. (2006) define the process

\[ r_t = \mu_t(\theta) + H_t^{1/2} \epsilon_t, \tag{2.11} \]

where \( H_t^{1/2} \) is the factor of the \( k \times k \) positive definite matrix \( H_t \), \( \epsilon_t \) as a white noise process, with the mean of the error term equaling zero and the variance equaling \( I_k \). Many different specifications for the conditional variance matrix \( H_t \) are defined in this section.

Bollerslev et al. (1988) originally propose a formulation of \( H_t \) where each element of the covariance matrix is a linear function of errors and lagged values of the elements of \( H_t \) defined as

\[ \text{vech}(H_t) = c + \text{Avech}(\epsilon_t' \epsilon_{t-1}) + \text{Bvech}(H_{t-1}). \tag{2.12} \]

They call this the VEC(1,1) model. However, this model is highly parameterized in high dimensional systems, making the VEC(1,1) model hard to estimate. Also, the VEC model cannot guarantee \( H_t \) is positive definite. To help reduce the number of parameters, Bollerslev alters the VEC model into the diagonal VEC (DVEC) model, which limits \( A \) and \( B \) to diagonal matrices. This adaptation is still hard to estimate in high dimensional systems.

Also, Bollerslev et al. need additional conditions to ensure the conditional variance matrices are positive definite. To ensure this, they assume that the matrices in equation 2.12 are \( c = \text{vech}(C^o) \), \( A = \text{diag}(\text{vech}(A^o)) \), and \( B = \text{diag}(\text{vech}(B^o)) \). They assume the matrices \( C^o \), \( A^o \), and \( B^o \) are positive definite.

Engle and Kroner (1995) propose a solution to the positivity issues of the VEC and DVEC models with their BEKK(1,1,K) model. The BEKK model is a special case of the VEC class of models that ensures that the covariance matrix is positive definite. The BEKK(1,1,K) model defines the covariance
matrix as follows:

\[ H_t = C^* C^* + \sum_{k=1}^{K} A_k^r \epsilon_{t-1} \epsilon_{t-1}^r A_k^* + \sum_{k=1}^{K} B_k^r H_{t-1} B_k^* \]  \hspace{1cm} (2.13)

The matrix \( C^* \) is limited to an upper triangular matrix. The BEKK model helps solve the positivity issues of the VEC model, but still contains the difficulty in high dimensional parameter estimation. The BEKK model reduces the parameterizations of the VEC model only slightly. Therefore, the VEC and BEKK models are not widely used for high dimensional estimation problems.

Kawakatsu (2003) details another approach to ensuring the positivity of \( H_t \) in the VEC model without all of the parameter restrictions that the BEKK(1,1,K) model imposes. Kawakatsu proposes the Cholesky factor GARCH that specifies a functional form in terms of the Cholesky factorization of the conditional covariance instead of \( H_t \). The advantage of the Cholesky factor GARCH is the assurance that the conditional covariance is positive definite without imposing restrictions on the parameters that do not identify the model.

The Cholesky factor GARCH specifies \( L_t \) from the decomposition of \( H_t^{-1/2} = L_t L_t' \) as

\[ \text{vech}(L_t) = c + \sum_{i=1}^{p} A_i h_{t-i} + \sum_{j=1}^{q} B_j \epsilon_{t-j}, \]  \hspace{1cm} (2.14)

where \( c, A_i, B_j \) are parameter arrays and the \( \text{vech}(\cdot) \) is previously defined in Section 1.1. Since the Cholesky factor \( L_t \) of a positive definite matrix is not uniquely defined, we assume all the diagonal elements of \( L_t \) are positive. This specification restricts the diagonal elements of \( L_t \) to depend on past values of the diagonal elements and not past values of the innovation vector. Kawakatsu proposes another specification of \( L_t \) as

\[ \text{vech}(L_t) = c + \sum_{i=1}^{p} A_i h_{t-i} + \sum_{j=1}^{q} B_j |\epsilon_{t-j}|. \]

The disadvantage of both these identification restrictions and the model in general is the parameters in the model become very hard to interpret.

Engle et al. (1990b) propose that a small number of underlying factors drive the common persistence in the conditional variances of the assets. The factor GARCH (F-GARCH) model by Engle et al. (1990b) develop is a special case of the BEKK model defined in equation 2.13. Lin (1992) defines the F-GARCH(p,q,K) model by

\[ H_t = \Omega + \sum_{j=1}^{p} \sum_{k=1}^{K} A_{kj} \epsilon_{t-1} \epsilon_{t-1}^r A_{kj} + \sum_{j=1}^{q} \sum_{k=1}^{K} B_{jk}^r H_{t-1} B_{jk} \]  \hspace{1cm} (2.15)

where \( A_{kj} \) and \( B_{kj} \) have rank one and the same left and right eigenvectors \( f_k \) and \( g_k \) with \( A_{kj} = \)
$a_{kj} f_k g_k' \quad \text{and} \quad B_{kj} = \beta_{kj} f_k g_k'$. With this specification, $H_t$ is defined by

$$H_t = \Omega + \sum_{k=1}^{K} g_k g_k' \left( \sum_{j=1}^{p} \alpha_{kj}^2 f_k' \epsilon_{t-j} f_k + \sum_{j=1}^{q} \beta_{kj}^2 f_k' H_{t-j} f_k \right).$$

There are also many different variants to the factor GARCH model in the literature.

### 2.1.3 Conditional Correlation Approach

Separately specified combinations of univariate GARCH model estimation and multivariate correlation matrix estimation are a less computationally burdensome approach to estimating multivariate GARCH models. This nonlinear combination approach greatly reduces the number of estimated parameters in the model.

Bollerslev (1990) proposes a model of this form, where the conditional correlation matrix remains constant. As in equation 1.6, the constant conditional correlation GARCH (CCC-GARCH) model is defined as

$$H_t = D_t R D_t, \quad D_t = \text{diag}(\sqrt{h_{i,t}}),$$

where $R$ is a correlation matrix with conditional correlations and $h_{i,t}$ defined as any univariate GARCH model. The most basic GARCH representation is

$$h_{i,t} = a_{i,0} + \sum_{p=1}^{P} a_{i,p} r_{t-p}^2 + \sum_{q=1}^{Q} b_{i,q} h_{i,t-q}.$$

The matrix $H_t$ is positive definite if all the conditional variances are positive and $R$ is positive definite.

The assumption that correlations of assets remain constant over time seems unreasonable in real world applications. Engle (2002) instead assumes a dynamic conditional correlation GARCH (DCC-GARCH) model where the conditional correlation matrix changes over time. Section 1.1 describes Engle’s DCC-GARCH model in detail. The full specification of the DCC-GARCH model is

$$r_t | \psi_{t-1} \sim N(0, D_t R_t D_t),$$

$$D_t^2 = \text{diag}(a_{0,i}) + \text{diag}(a_{1,i}) \circ r_{t-1} r_{t-1}' + \text{diag}(b_{1,i}) \circ D_{t-1}^2,$$

$$\varepsilon_t = D_t^{-1} r_t,$$

$$Q_t = R_0 (1 - \alpha - \beta) + \alpha \epsilon_{t-1} \epsilon_{t-1}' + \beta Q_{t-1},$$

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

Tse and Tsui (2002) define another specification for the DCC-GARCH model that is less popular than Engle’s DCC-GARCH. The only difference in Tse and Tsui’s DCC-GARCH is the specification of...
the matrix $R_t$ as

$$R_t = (1 - \theta_1 - \theta_2)R_0 + \theta_1 R_{t-1} + \theta_2 \Psi_{t-1},$$

where the elements of $\Psi_t$ are defined as

$$\psi_{i,j,t-1} = \frac{\sum_{h=1}^{M} \epsilon_{i,t-h} \epsilon_{j,t-h}}{\sqrt{\left(\sum_{h=1}^{M} \epsilon_{i,t-h}^2\right) \left(\sum_{h=1}^{M} \epsilon_{j,t-h}^2\right)}}.$$

Therefore, $\psi_{t-1}$ is the sample correlation matrix of an M length rolling window of previous time points. To guarantee $\Psi_t$ is positive definite, place the restriction $M \geq K$.

Audrino and Barone-Adesi (2006) notice that Engle’s DCC-GARCH model constrains the correlation dynamics to be equal across all of the assets. Audrino and Barone-Adesi relax this assumption in the creation of the average conditional correlation GARCH (ACC-GARCH) model. They propose their model as another approach to extend Bollerslev’s CCC-GARCH. The ACC-GARCH has the same functional form as the other conditional correlation approaches with the volatility and correlation structures defined separately. The univariate volatility portion of the model is similarly defined by any univariate GARCH model; the authors use the GARCH(1,1) structure defined by

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1}.$$

Again, the main difference between the ACC-GARCH and other conditional correlation models is the form of the correlation matrix $R_t$. Audrino and Barone-Adesi define the correlation matrix $R_t$ as

$$R_t = (1 - \lambda) \overline{Q}_{t-p}^{-1} + \lambda \overline{R}_t, \quad \lambda \in [0, 1],$$

where $\overline{Q}_{t-p}^{-1}$ is defined as the unconditional correlation matrix of the returns over the past $p$ days. The matrix $\overline{R}_t$ is a matrix with ones on the diagonal and all other elements equal to

$$\overline{r}_t = \frac{1}{k-1} \left( \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{\sigma_{t,i} \sigma_{t,j}}{\left(\sum_{d=1}^{k} \sigma_{t,d}\right)^2} - 1 \right),$$

where $k$ is the number of assets in the portfolio. When the parameter $\lambda$ is zero, the model becomes very similar to the CCC-GARCH with a rolling window correlation estimate. The ACC-GARCH model is estimated similarly to the two stage estimation of Engle’s DCC-GARCH with the details not included here. The detail of the nonparametric procedure is left to Audrino and Barone-Adesi (2006).

Pelletier (2006) develops a regime switching model as a balance between Bollerslev’s CCC-GARCH and dynamic correlation models such as the DCC-GARCH of Engle or the DCC-GARCH of Tse and Tsui. In Pelletier’s regime switching dynamic correlation (RSDC) model, the correlation matrix remains
constant within a regime and changes across different regimes. The RSDC model uses a Markov chain to switch between the regimes. Pelletier still takes the foundational approach of Bollerslev in equation 1.6 with \( r_t = H_t^{1/2} \epsilon_t \), but defines the matrix \( H_t \) as

\[
H_t = S_t \Gamma_t S_t
\]

where \( S_t \) is a diagonal matrix composed of standard deviations and \( \Gamma_t \) is a correlation matrix. The correlation matrix \( \Gamma_t \) is defined as

\[
\Gamma_t = \sum_{n=1}^{N} \mathbb{1}_{\Delta_t = n} \Xi_n
\]

with \( \mathbb{1} \) representing the indicator function, \( \Delta_t \) an unobserved Markov chain process that is independent of \( \epsilon_t \), \( \Xi_n \) are correlation matrices, and \( N \) is the total number of regimes. The Markov chain process \( \Delta_t \) can take integer values from 1 to \( N \). Pelletier imposes constraints on the matrices \( \Xi_n \) to ensure that \( \Gamma_t \) is a correlation matrix. He works with the Cholesky factorization \( \Xi_n = P_n P_n^T \) and imposes constraints on \( P_n \) to make \( \Xi_n \) positive definite.

A benefit of Pelletier’s RSDC model over the DCC-GARCH models of Engle and Tse and Tsui is that the RSDC model allows for computation of multi-step ahead conditional expectations of the variance matrix. This is due to the linearity of the correlation model from the Markov chain. DCC-GARCH models use square roots of variances that input nonlinearities in the model. Furthermore, Pelletier defines the volatility standard deviations with the ARMACH model of Taylor (1986) and Schwert (1989) to perform these calculations for the entire variance matrix. The ARMACH model defines the standard deviations in the volatility portion of the model as

\[
s_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i |r_{t-i}| + \sum_{i=1}^{q} \beta_i s_{t-i}
\]

This model may include a more robust approach to estimating volatility by the use of absolute deviations of returns instead of squared returns. The ARMACH model is not required by the RSDC model but does allow for the computation of multi-step ahead conditional expectations of the entire variance matrix. Pelletier (2006) gives further details of the estimation of the parameters in the RSDC model.

### 2.2 Generalized Robust Estimation

Generalized linear models (GLM) defined by Nelder and Wedderburn (1972) play a prominent role in the field of statistics. Generalized linear models have the following joint density:

\[
f^*(Y; X; \beta) = f(Y; X^T \beta) u(X).
\]
The conditional density of the response variable vector $Y$ given the explanatory variable vector $X = x$ is $f^*(Y; X^T \beta)$, which depends on the unknown parameter vector $\beta$. The marginal density of $X$ is $u(X)$ in the above equation.

Let $(y_i, x_i)$ from $i = 1, \ldots, n$ be independent observations from $f^*(Y, X; \beta)$. Under certain regularity conditions, the maximum likelihood estimator (MLE) $\hat{\beta}_{MLE}$ is the solution to

$$
\sum_{i=1}^{n} \ell(y_i, x_i, \beta) = 0,
$$

with $\ell(\cdot)$ being the derivative of the log of the likelihood function. Under further regularity conditions, $\hat{\beta}_{MLE}$ is both consistent and asymptotically normally distributed.

A few anomalous observations can strongly affect the MLE. These anomalous observations, or outliers, can come in two common forms - leverage points and residuals. A leverage point occurs when a point $x_i$ is an outlier in the covariate space. For example, if $x_i$ is an outlier in the covariate space, then $z_i = (y_i, x_i)$ is a leverage point. These leverage points are either harmful or not, depending on whether the error of $z_i$ is large or small, respectively. Another type of outlier is a vertical outlier that occurs when a point $z_i$ is not a leverage point but still has a large residual, as described in Croux and Haesbroeck (2003). These different types of outliers lead to different problems with the generalized linear model.

Anomalous observations should not affect robust estimators to the extent that they affect the MLE. These robust estimators should be approximately equal to the parameters even in the presence of outliers. The addition or deletion of a few observations should not greatly affect the parameter estimates or analysis in robust estimation.

Multiple approaches to robust estimation of parameters in generalized linear models have been developed over the past few decades. Some of the first approaches to solving the problems with robust estimation in GLM involve a notion of the sensitivity of an estimator. Robust estimators should not be as sensitive to changes in small numbers of observations. This would ensure that one observation being an outlier would not greatly affect the estimated parameter. Creating a specific measure of sensitivity permits the comparison of different estimators. Bounding this measure of sensitivity would then ensure that an estimator could not have an infinitely large sensitivity. Call the function $\Omega$ the influence function, where $\Omega(y_i, x_i)$ represents the effect of a single observation $(y_i, x_i)$ on the estimation. Bounding this influence function $\Omega$ is a form of bounding the sensitivity of the estimator as mentioned in Hampel (1974). The focus here turns to the influence of the class of estimators called M-estimators defined in equation 1.17. From Stefanski et al. (1986), the influence function of an M-estimator is

$$
\Omega(y_i, x_i) = \frac{\psi(y_i, x_i; \beta)}{-\left(\frac{\partial}{\partial \beta} E(\psi(y_i, x_i; \beta))\right)},
$$

where $\psi$ is a measurable function. Minimizing the asymptotic variance of the M-estimator and
bounding the influence function would not only ensure relative efficiency, but also ensure robustness of the estimator because any observation would have a limited effect on the estimator.

2.2.1 Break-down Point Approach

Instead of trying to measure the sensitivity of an estimator with influence functions, another approach is to measure the amount of outliers in the sample it would take to ruin the estimate. Since every sample is of different size, an exact number of outliers to ruin the estimator would not be helpful, but a percentage of observations that are potential outliers is understandable for any sample size. The break-down point is the maximum percentage of outliers in a sample before the estimator becomes completely inaccurate.

Yohai (1987) notes that most robust estimators of the time had extremely low break-down points. Therefore, even if the estimators are robust by definition, it would only take a small percentage of outliers in the sample to ruin even the robust estimator. Yohai points out that Huber’s M-estimation even had a break-down point of zero. Rousseeuw and Yohai (1984) propose a method that tries to keep the flexibility and asymptotic properties of the M-estimators, but has a higher breakdown point. Rousseeuw and Yohai focus their attention on estimating the scale of the residuals to derive their parameter estimates in the regression model

\[ y_i = x_i^T \beta + \xi_i. \]

They define a symmetric and continuously differentiable function \( \rho \) where there exists a positive constant \( c \) such that \( \rho \) is strictly increasing on \([0, c]\). They defined the scale estimates, \( s \), as the solution to the equation

\[
\frac{1}{n} \sum_{i=1}^{n} \rho \left( \frac{\hat{\xi}_i}{s(\beta)} \right) = E \left[ \rho(\hat{\xi}) \right],
\]

where \( \hat{\xi}_i = y_i - x_i^T \beta \). The S-estimator, \( \hat{\beta}_S \), is the solution of

\[
\arg \min_{\hat{\beta}} s(\hat{\xi}_1, \ldots, \hat{\xi}_n).
\]

Rousseeuw and Yohai propose a possible function \( \psi = \rho' \) defined as Tukey’s biweight function

\[
\psi(k) = \begin{cases} 
  k \left[ 1 - \left( \frac{k}{c} \right)^2 \right]^2 & \text{if } |k| \leq c \\
  0 & \text{if } |k| > c
\end{cases},
\]

for a constant \( c \). This function bounds the possible outlier effects in the regression model leading to more robust parameter estimates.
Rousseeuw and Yohai (1984) prove consistency and asymptotic normality for their S-estimators, but mention that the calculation of the estimators may be very difficult and time consuming. The breakdown point of their S-estimators is shown to be 

$$\delta^* = \left\lfloor \frac{n}{2} \right\rfloor + 1.$$ 

As the sample size increases, the value of the breakdown point approaches $$\delta^* = 0.5$$. This is far better than the M-estimators previously defined.

However, Yohai (1987) shows that early attempts at high break-down point estimators, such as the S-estimators of Rousseeuw and Yohai (1984), all became highly inefficient with the regression model with normal errors. Yohai proposes an estimate called the MM-estimate that had both high break-down point and high efficiency under the regression model with normal errors

$$y_i = x_i^T \beta + \xi_i, \quad \xi_i \sim \text{N}(0, \sigma^2).$$

First, Yohai defines a scale M-estimate that is similar to the solution of equation 2.20. He starts with a continuous, monotone increasing function called $$\rho$$. Then he defines the scale estimate $$s$$ as the solution to

$$\frac{1}{n} \sum_{i=1}^{n} \rho \left( \frac{\xi_i}{s} \right) = E\phi(\rho(\xi_i)), \quad \phi \sim \text{N}(0, 1). \quad (2.22)$$

From this scale M-estimate, Yohai takes three steps to derive the MM-estimator. First, Yohai takes one of the previously found high break-down point estimates that had a break-down point as high as 0.5. He then computes the residuals of the model, $$\hat{\xi}_i$$, from this estimate to create a scale M-estimate of these residuals, $$s(\hat{\xi}_i)$$, using a function $$\rho_0$$. Lastly, Yohai takes another function, $$\rho_1$$, completely dominated by $$\rho_0$$ but has the same finite supremum. The MM-estimate, $$\hat{\beta}_{\text{MM}}$$, is the solution to

$$\sum_{i=1}^{n} \rho_1^t \left( \frac{\xi_i(\beta)}{s(\xi_i(\beta))} \right) x_i = 0. \quad (2.23)$$

This solution however may not be unique because local minima might exist along with the global minimum. These other possible solutions still are MM-estimators with high break-down points and high efficiency. Yohai proves the consistency of $$\hat{\beta}_{\text{MM}}$$ and that the estimator also has an asymptotic normal distribution. Yohai (1987) gives the details of these proofs.

Yohai and Zamar (1988) later improve the robust MM-estimate with the robust $$\tau$$-estimate. The process for creating the $$\tau$$-estimate is quite similar in nature to that of creating the MM-estimate with a different scale of the residuals $$\hat{\xi}_i$$. The only advantage of the $$\tau$$-estimate is that it provides a high break-down and highly efficient estimate of the scaled error simultaneously with the estimates of the regression coefficients. To derive the $$\tau$$-estimate, Yohai and Zamar begin with a robust M-scale

24
estimate \( s \) similar to the solution of equation 2.22. They define the scale estimate \( \tau \) as

\[
\tau(\xi)^2 = s^2 \frac{1}{n} \sum_{i=1}^{n} \rho \left( \frac{\xi_i}{s} \right),
\]  

(2.24)

where \( \rho \) is similarly defined as equation 2.22. The robust \( \tau \)-estimates of the parameters are calculated from

\[
\hat{\beta}_\tau = \arg \min_{\beta} \tau(\xi(\beta)).
\]

There are criticisms to the break down point approach to robust estimation. Some of the criticisms question the break-down point approach with break-down point close to 0.5. A break-down point this high could lead to the exclusion of half of the observations in a sample. This could allow incorrect analysis in patterns of data where other possible models might better represent the data, or situations where different halves of the data have different trends. Both of these situations might have half of the observations excluded even though the excluded observations are not outliers.

One can argue that these situations are good examples of why the break-down point approach to robust estimation is beneficial. Break-down point estimation reveals these situations after plotting the fitted model with the data. This plot reveals any existing patterns in excluded observations. In addition, a simple plot of the data itself, or the robust residuals may reveal these situations as well. Blindly applying break-down point estimation, like any statistical method, has potential downfalls without careful analysis of the original data.

### 2.2.2 Deviance Robust Estimation

Other approaches include working with deviances instead of likelihoods and developing a measure of observation influence and trying to bound these measures. Pregibon (1982) approaches the problem of sensitivity by switching between likelihood functions and deviance functions. The log of the likelihood function transforms to a similar deviance function

\[
d(g(x^T \beta); y_i) = -2 \left[ L(y_i, x_i; \beta) - L(y_i, x_i; \hat{\beta}_{MLE}) \right],
\]

(2.25)

where \( g(x^T \beta) \) is the link function of the generalized linear model. The link function provides the relationship between the linear predictor and the mean of the distribution function. Converting the log of the likelihood to the deviance does not alter the results because maximizing the log of the likelihood function is the same as minimizing the deviance function.

This makes the deviance function as sensitive to outliers as the likelihood function. Pregibon
proposes a new less sensitive estimator that minimizes
\[ \sum_{i=1}^{n} \lambda \{ d \{ g(x^T \beta) ; y_i \} \}, \] (2.26)

where \( \lambda \) is a strictly increasing Huber loss function defined in Huber (1973) and the estimate \( \hat{\beta}_\lambda \) exists and is unique. Pregibon’s \( \lambda \) function is:
\[ \lambda_P(t) = \begin{cases} 
  t, & t \leq c \\
  2\sqrt{tc} - c, & t > c 
\end{cases} \] (2.27)

with \( c \) an adjustable tuning constant. However, Pregibon suggests that this is not the only option for the \( \lambda \) function.

Although Pregibon’s goal of resistance to poorly fitted observations is achieved with this method, high leverage points still greatly affect the results. For a solution to this issue, a weighting scheme depending on only the covariate \( x_i \) solves this problem. Pregibon’s estimator also has a bounded influence function, under an adjusted weighting scheme, which is a desired quality of a robust estimator.

Bianco and Yohai (1996) propose their own adaptation of the Pregibon robust estimator. Bianco and Yohai adapt the Pregibon estimate by altering equation 2.27. They propose an M-estimate of the \( \lambda \)-type just as Pregibon, but Bianco and Yohai desire to make their estimator Fisher consistent and asymptotically normally distributed. A Fisher consistent estimator must hold to the following:
\[ E[\psi(y_i, x_i; \beta)] = 0, \]

where \( \psi \) is a measurable function similar to equation 1.17. Bianco and Yohai use a different \( \lambda \) function denoted \( \lambda_{BY} \) where
\[ \lambda_{BY}(t) = \begin{cases} 
  t - \frac{t^2}{2c}, & t \leq c \\
  \frac{c}{2}, & t > c 
\end{cases} \] (2.28)

with \( c \) a positive number. This function bounds the effects of outliers with its truncated function design. Since this \( \lambda \) function converges to the identity function as \( c \to \infty \), \( \hat{\beta}_{BY} \to \hat{\beta}_{MLE} \) as \( c \to \infty \). These properties along with a bias correction factor defined in Section 3.2.2 lead to the proof of \( \hat{\beta}_{BY} \) converging in distribution to a multivariate normal distribution with mean zero and covariance matrix \( V \) defined in Bianco and Yohai (1996). However, \( \hat{\beta}_{BY} \) later proves undesirable in certain situations of robust estimation.

Croux and Haesbroeck (2003) began to critique the Bianco and Yohai estimate in the logistic regression situation. Again, Croux and Haesbroeck show the robustness and inconsistency of the estimator created by Pregibon (1982) and note the benefits of the Bianco and Yohai (1996) estimator.
Although the Bianco and Yohai estimator is both robust and consistent, Croux and Haesbroeck notice that the Bianco and Yohai estimator may not always exist when working with loss functions, even if the model does not contain any outliers. This leads Croux and Haesbroeck to improve the Bianco and Yohai estimator with a newer $\lambda$ function.

The Bianco and Yohai estimator uses the $\lambda_{BY}$ function defined in equation 2.28, while Croux and Haesbroeck suggest using the $\lambda_{CH}$ function defined as

$$
\lambda_{CH}(t) = \begin{cases} 
  t e^{-\sqrt{c}}, & t \leq c \\
  -2e^{-\sqrt{c}(1+\sqrt{t})} + e^{-\sqrt{c}(2+2\sqrt{c}+c)}, & t > c 
\end{cases}
$$

for a given constant $c$, with $c$ determining the balance between efficiency and robustness. Croux and Haesbroeck defined their estimator, $\hat{\beta}_{CH}$, following the same methods as Bianco and Yohai with the only change being $\lambda_{CH}$. The improved estimation method does not downweight the outliers as severely as the Bianco and Yohai method. The new $\hat{\beta}_{CH}$ estimator always exists with a simple condition - the MLE exists. The proof of this point may be found in Croux and Haesbroeck (2003).

### 2.2.3 Robust Estimation of Time Series Data

The structure of the DCC-GARCH model has a time series element to the modeling. The structure of time series data makes the data more susceptible to the effects of outliers because current observations depend on previous observations. Whether outliers occur in patches or are isolated in the data, the parameter estimates of an autoregressive moving average (ARMA) model are affected. The ARMA(p,q) model is defined

$$
y_t = \omega + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t,
$$

for $L$ is the lag operator that takes the previous value of the variable, $\epsilon_t$ is a white noise process, and all the roots of the polynomials $\phi(L)$ and $\theta(L)$ are outside the unit circle. A white noise process is a process that has a zero mean, constant and finite variance, and is serially uncorrelated.

Using this representation, Chen and Liu (1993) describe the effect of additive or innovation outliers...
at the time point $t = t^*$ on the series $y_t$ with

$$
\hat{y}_t = y_t + \delta \zeta(L) \mathbb{1}_{t=t^*},
$$

(2.31)

with $y_t$ following the ARMA process described in equation 2.30, $\delta$ denoting the magnitude of the outlier effect, and $\zeta(L)$ denoting the dynamic pattern of the outlier effect. The dynamic pattern definitions are:

$$
\text{AO: } \zeta_{\text{AO}}(L) = 1 \\
\text{IO: } \zeta_{\text{IO}}(L) = \frac{\theta(L)}{\phi(L)}.
$$

(2.32)

The effects of each outlier are estimated as

$$
\text{AO: } \hat{\delta}_{\text{AO}}(t^*) = \frac{\sum_{t=t^*}^{T} \hat{\epsilon}_t x_{2,t}}{\sum_{t=t^*}^{T} x_{2,t}^2} \\
\text{IO: } \hat{\delta}_{\text{IO}}(t^*) = \hat{\epsilon}_t,
$$

(2.33)

where $x_{i,t} = 0$ for $i = 1, 2$ and $t < t^*$, $x_{1,t^*} = 1$ for $i = 1, 2$, $x_{1,(t^*+k)} = 0$ and $x_{2,(t^*+k)} = -\pi_k$ for all $k \geq 0$.

The values of $\pi_k$ are the coefficients of the Taylor series polynomial

$$
\pi(L) = \frac{\theta(L)}{\phi(L)} = 1 - \pi_1 L - \pi_2 L^2 - \ldots.
$$

(2.34)

Unlike innovation outliers, the effects of additive outliers do not depend on the structure of the model. Innovation outliers occurring at time $t = t^*$ have a dynamic effect on the model for time points $t + k$, with $k > 0$, that depends directly on the structure of the ARMA model.

To evaluate the direct effects of additive and innovation outliers, consider the simple AR(1) model without an intercept defined as

$$
\hat{y}_t = \phi \hat{y}_{t-1} + \epsilon_t.
$$

(2.35)

Obtain the least squares estimate of $\phi$ by minimizing

$$
\sum_{t=2}^{T} (y_t - \phi y_{t-1})^2.
$$

Differentiating the above equation with respect to $\phi$ and solving for $\phi$ leads to the closed form solution to the least squares estimate as

$$
\hat{\phi} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=1}^{T-1} y_t^2}.
$$

(2.36)

In an AR(1) process, the value of $\phi$ also equals the first autocorrelation, $\rho(1)$, of the process because

$$
\rho(1) = \frac{\text{Cov}(y_t, y_{t-1})}{\text{Var}(y_t)} = \frac{\text{Cov}(\phi y_{t-1} + \epsilon_t, y_{t-1})}{\text{Var}(y_t)} = \frac{\phi \text{Var}(y_t) + 0}{\text{Var}(y_t)} = \phi.
$$
Therefore, under certain regularity conditions, \( \hat{\phi} \overset{p}{\to} \rho(1) \) as \( T \to \infty \).

When additive outliers are present, the actual value of \( y_t \) is not observed. Maronna et al. (2006) describe additive outliers as outliers where instead of the expected observation \( y_t \), it is replaced by \( y_t + \upsilon_t \) where \( \upsilon_t \sim (1 - \epsilon)\delta_0 + \epsilon N(\mu_\upsilon, \sigma^2_\upsilon) \). If this is the case, then the calculation of \( \rho(1) \) becomes

\[
\rho_{AO}(1) = \frac{\text{Cov}(y_t + \upsilon_t, y_{t-1} + \upsilon_{t-1})}{\text{Var}(y_t + \upsilon_t)}
= \frac{\text{Cov}(y_t, y_{t-1}) + \text{Cov}(y_t, \upsilon_t) + \text{Cov}(\upsilon_t, y_{t-1}) + \text{Cov}(\upsilon_t, \upsilon_{t-1})}{\text{Var}(y_t) + \text{Var}(\upsilon_t)}
= \frac{\phi \text{Var}(y_t) + \rho_\upsilon(1) \text{Var}(\upsilon_t)}{\text{Var}(y_t) + \text{Var}(\upsilon_t)} \neq \phi.
\]

Only when \( \text{Var}(\upsilon_t) = 0 \) does \( \rho_{AO}(1) = \rho(1) = \phi \).

The effects of innovation outliers occur completely in the error term of the model in equation 2.30. For this reason, the ARMA process \( y_t \) remains unchanged and the estimate of \( \phi \) is consistent if \( \mu_\upsilon = 0 \). However, according to Maronna et al. (2006), in least squares estimation innovation outliers do increase the variance of the estimate of the mean of an ARMA process compared to models estimated without innovation outliers. Therefore, least squares estimation with innovation outliers becomes less efficient.

One approach to account for additive outliers in the data is robust filtering of AR models described in Maronna et al. (2006). A robust filter approach replaces the predicted residuals \( \hat{\epsilon} \) with a filtered predicted residual \( \tilde{\epsilon} \). For an AR(p) model, these two sets of residuals are defined as

\[
\hat{\epsilon}_t = (y_t - \mu) - \phi_1(y_{t-1} - \mu) - \ldots - \phi_p(y_{t-p} - \mu) \quad (2.37)
\]
\[
\tilde{\epsilon}_t = (y_t - \mu) - \phi_1(\tilde{y}_{t-1} - \mu) - \ldots - \phi_p(\tilde{y}_{t-p} - \mu), \quad (2.38)
\]

where \( \tilde{y}_{t-i|t-1} \) are filtered robust estimates of \( y_{t-i} \) given all information up to time \( t - 1 \). The filtered estimates are calculated recursively with the equation

\[
\tilde{y}_{t|t} = \tilde{y}_{t|t-1} + s_t \psi \left( \frac{\hat{\epsilon}_t}{s_t} \right),
\]

where \( s_t \) is an estimate of the scale of \( \hat{\epsilon}_t \), \( \tilde{y}_{t|t-1} \) are filtered robust estimates of \( y_t \) given all information up to time \( t - 1 \), and the function \( \psi \) is defined by

\[
\psi(u) = \begin{cases} u & \text{if } |u| \leq a \\ 0 & \text{if } |u| > b \end{cases},
\]

with a polynomial defined further in Maronna et al. (2006) as the value if \( a < |u| \leq b \) to ensure
smoothness of the function. This means that the estimate $\hat{y}_{t|t}$ is given by

$$\hat{y}_{t|t} = \begin{cases} 
\hat{y}_{t|t-1} & \text{if } |\tilde{\epsilon}_t| > bs_t \\
y_t & \text{if } |\tilde{\epsilon}_t| \leq as_t 
\end{cases},$$

for constants $a$ and $b$. The new robust parameter estimates of the model are calculated from these filtered residuals. The same process is extended to ARMA models in Maronna et al. (2006) with the basic approach remaining the same.

The $\tau$-estimates of Yohai and Zamar (1988) can be applied to the estimation of ARMA models as well. Maronna et al. alter the typical calculation of the maximum likelihood estimates of the parameters, which under the assumption of the normality of $\epsilon_t$, are defined by the minimization of

$$\sum_{t=1}^{T} \log \left( \frac{\sigma_t^2}{\sigma_{\epsilon}^2} \right) + T \log \left( \frac{1}{T} \sum_{t=1}^{T} \frac{\hat{\epsilon}_t^2}{\sigma_t^2/\sigma_{\epsilon}^2} \right),$$

(2.39)

where $\sigma_{\epsilon}^2$ is the variance of the error term. With the same $\tau$-estimate defined in equation 2.24, the robust $\tau$-estimate of the parameters in the ARMA model are estimated from the minimization of the new quantity

$$\sum_{t=1}^{T} \log \left( \frac{\sigma_t^2}{\sigma_{\epsilon}^2} \right) + T \log \left( \tau^2 \left( \frac{\hat{\epsilon}_1}{\sigma_1^2/\sigma_{\epsilon}^2}, \ldots, \frac{\hat{\epsilon}_T}{\sigma_T^2/\sigma_{\epsilon}^2} \right) \right).$$

(2.40)

These estimates become filtered $\tau$-estimates when combined with the filtered residual estimates mentioned in the previous paragraph. The quantity to be minimized now becomes

$$\sum_{t=1}^{T} \log \left( \frac{\sigma_t^2}{\sigma_{\epsilon}^2} \right) + T \log \left( \tau^2 \left( \frac{\hat{\epsilon}_1}{\sigma_1^2/\sigma_{\epsilon}^2}, \ldots, \frac{\hat{\epsilon}_T}{\sigma_T^2/\sigma_{\epsilon}^2} \right) \right).$$

(2.41)

Chen and Liu (1993) design an approach to not only detect outliers in an ARMA process, but to robustly estimate the parameters in the model as well. Equations 2.32 and 2.33 detail the dynamic patterns and effects of both additive and innovation outliers. Chen and Liu create a set of standardized statistics of outlier effects defined as

AO: $\hat{y}_{AO}(t^*) = \frac{\hat{\delta}_{AO}(t^*)}{\sigma_{\epsilon}^2} \left( \sum_{t=t^*}^{T} x_{2,t}^2 \right)^{1/2}$

IO: $\hat{y}_{IO}(t^*) = \frac{\hat{\delta}_{IO}(t^*)}{\sigma_{\epsilon}^2},$

(2.42)

where $\hat{\delta}_{AO}$ and $\hat{\delta}_{IO}$ are defined in equation 2.33.

Chen and Liu use the following iteration process to both detect outliers and robustly estimate the parameters of the ARMA model. First, they estimate the maximum likelihood estimates of the
ARMA parameters based on the original data. Next, they compute \( \hat{\gamma}_{\text{AO}}(t) \) and \( \hat{\gamma}_{\text{IO}}(t) \) for all values of \( t \). If \( \max_t(\hat{\gamma}_{\text{AO}}(t), \hat{\gamma}_{\text{IO}}(t)) \geq C \) for some predetermined constant \( C \), there is a possible outlier at that point. Chen and Liu remove the effects of this outlier using the estimated effects and dynamic patterns defined in equations 2.32 and 2.33. After correcting all \( m \) outlier effects in all possible outlier observations where \( m \geq 0 \), Chen and Liu compute the maximum likelihood estimates of the model. After the iteration process is complete and no more new outliers are detected, they jointly estimate the effects, \( \hat{\delta} \), of all \( m \) previously determined outliers with the regression model

\[
\hat{\varepsilon} = \sum_{j=1}^{m} \delta \pi(L) \zeta_j(L) \mathbb{1}_{t=t^*} + \epsilon_t,
\]

where \( \zeta(L) \) and \( \pi(L) \) are previously defined in equations 2.32 and 2.34. They estimate a new set of \( \gamma \) statistics defined as \( \hat{\gamma}_j = \hat{\delta}_j / \text{SD}(\hat{\delta}_j) \). Using the same predetermined constant \( C \) defined above, they check if the new values give a \( \max_t(\hat{\gamma}_{\text{AO}}(t), \hat{\gamma}_{\text{IO}}(t)) \geq C \). If the value is below \( C \), then disregard the point as an outlier. However, if the value is still greater than \( C \), remove the effect of the outlier using the updated values of \( \hat{\delta} \). After this process is repeated over the entire set of previously detected outliers, Chen and Liu recalculate parameter estimates with maximum likelihood estimation.

The general class of M-estimators defined by Huber (1964) in equation 1.17 is similarly applied to the ARMA structure of models by minimizing

\[
\sum_{t=p+1}^{T} \rho \left( \frac{\hat{\varepsilon}_t(\phi, \theta)}{\hat{\sigma}_\varepsilon} \right), \tag{2.43}
\]

where \( \hat{\sigma}_\varepsilon \) is a robust scale estimate of the residuals. Obtain another solution by setting the derivative of the above quantity equal to zero and solving for the parameters with

\[
\sum_{t=p+1}^{T} y_t \psi \left( \frac{\hat{\varepsilon}_t(\phi, \theta, y_t)}{\hat{\sigma}_\varepsilon} \right) = 0, \tag{2.44}
\]

where \( \psi = \rho' \) and \( y_t \) is the time series process. Assume that the \( \psi \) function is consistent such that

\[
E_{\phi, \theta} \left( y_t \psi \left( \frac{\varepsilon_t}{\sigma_\varepsilon} \right) \right) = 0.
\]

Under certain regularity conditions and the invertibility and stationarity of the ARMA process, the M-estimate of the parameters is asymptotically normal. The asymptotic distribution of the parameter vector \( \lambda = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) \) is given by

\[
\sqrt{T}(\hat{\lambda}_M - \lambda) \xrightarrow{d} N_{p+q+1}(0, V_M), \tag{2.45}
\]

31
with the detailed definition of the variance matrix $V_M$ left in Maronna et al. (2006). Although M-estimators are more robust than the estimates from the least squares approach with additive outliers, the efficiency of the robustness decreases as the order $p$ of the AR portion of the model increases. This leads to bias in the estimates.

Denby and Martin (1979) attempt to correct the problem of bias created by additive outliers in M-estimation with a generalized M-estimation (GM) procedure. Instead of minimizing $2.43$, generalized M-estimation minimizes

$$
\sum_{t=p+1}^{T} W(y_t) \rho \left( \frac{e_t(\phi, \theta)}{\sigma_e} \right),
$$

(2.46)

where $W(\cdot)$ is a nonnegative weight function that down weights certain observations containing additive outliers to limit their effect. Denby and Martin define their weight function as $W(k) = g(k)/k$ with the bounded function $g(\cdot)$ and $k g(k) \geq 0$. Under this representation, obtain GM-estimates by solving for the parameters in the equation

$$
\sum_{t=p+1}^{T} g(y_t) \psi \left( \frac{e_t(\phi, \theta)}{\sigma_e} \right) = 0,
$$

(2.47)

instead of equation 2.44 for M-estimation.

Denby and Martin define $g(\cdot)$ and $\psi(\cdot)$ as functions depending on estimates of the scale of $y$, $s_y$, and the scale of the residuals, $s_r$. These functions are

$$
g(k) = C_y s_y g_0 \left( \frac{k}{C_y s_y} \right),
$$

$$
\psi(k) = C_r s_r \psi_0 \left( \frac{k}{C_r s_r} \right),
$$

where $C_y$ and $C_r$ are predetermined constants chosen to optimize the efficiency of the procedure. Denby and Martin choose $g(\cdot) = \psi(\cdot)$ for their research, but suggest that this is not a requirement. They use two forms of the $\psi$-function. They define them as a monotone Huber influence function and Tukey’s bisquare influence function. These functions are respectively defined

$$
\psi_H(k) = \begin{cases} 
  k & \text{if } |k| \leq 1 \\
  \text{sign}(k) & \text{if } |k| > 1
\end{cases},
$$

(2.48)

$$
\psi_B(k) = \begin{cases} 
  k(k - k^2)^2 & \text{if } |k| \leq 1 \\
  0 & \text{if } |k| > 1
\end{cases}
$$

(2.49)

The choice of $\psi_B$ is more robust to additive outliers than $\psi_H$.

Denby and Martin show in detail not covered here that the GM-estimates are both consistent and asymptotically normal in the presence of innovation outliers. For the additive outlier model,
the GM-estimate is asymptotically biased, which is similar to least square estimation and traditional M-estimation. They mention that with proper selection of \( g(\cdot) \) and \( \psi(\cdot) \) functions, GM-estimated parameters will have less bias than with traditional M-estimation techniques.

Bustos and Yohai (1986) propose a robust estimate of ARMA models based on robust autocovariances (RA) and compare them with the performance of least squares, M, and GM estimation with additive outliers. They alter the equations for the least squares estimation of the parameters in the ARMA\((p,q)\) model. Another method of finding the least squares estimates for the ARMA model is differentiating the sum of squares quantity for every parameter leading to the system of equations

\[
\begin{align*}
\sum_{t=p+1}^{T} \epsilon_t (\partial \epsilon_t / \partial \phi_j) &= 0, \quad 1 \leq j \leq p \\
\sum_{t=p+1}^{T} \epsilon_t (\partial \epsilon_t / \partial \theta_j) &= 0, \quad 1 \leq j \leq q \\
\sum_{t=p+1}^{T} \epsilon_t (\partial \epsilon_t / \partial \mu) &= 0.
\end{align*}
\]

(2.50)

Assuming stationarity and invertibility leads to the above equations simplifying into

\[
\begin{align*}
\sum_{t=p+1}^{T} \epsilon_t \phi^{-1}(L)\epsilon_{t-j} &= 0, \quad 1 \leq j \leq p \\
\sum_{t=p+1}^{T} \epsilon_t \theta^{-1}(L)\epsilon_{t-j} &= 0, \quad 1 \leq j \leq q \\
\sum_{t=p+1}^{T} \epsilon_t &= 0,
\end{align*}
\]

(2.51)

where \( L \) is the lag operator and all the roots of the polynomials \( \phi(L) \) and \( \theta(L) \) are outside the unit circle. By letting \( p_j \) represent the coefficients in \( \phi^{-1}(L) \) and \( t_j \) represent the coefficients in \( \theta^{-1}(L) \), equations 2.51 become

\[
\begin{align*}
\sum_{h=0}^{T-j-p-1} p_h \gamma_{h+j}(\phi, \theta) &= 0, \quad 1 \leq j \leq p \\
\sum_{h=0}^{T-j-p-1} t_h \gamma_{h+j}(\phi, \theta) &= 0, \quad 1 \leq j \leq q \\
\sum_{t=p+1}^{T} \epsilon_t &= 0,
\end{align*}
\]

(2.52)
where \( \gamma_i(\phi, \theta) = \sum_{t=p+1}^{T-i} \hat{\epsilon}_{t+i} \hat{\epsilon}_t \).

Bustos and Yohai (1986) create their robust residual autocovariance (RA) estimates of the ARMA model by adjusting equations 2.52 into

\[
\sum_{n=0}^{T-j-p-1} p_h \tilde{\gamma}_{h+j}(\phi, \theta) = 0, \quad 1 \leq j \leq p
\]

\[
\sum_{h=0}^{T-j-p-1} t_h \tilde{\gamma}_{h+j}(\phi, \theta) = 0, \quad 1 \leq j \leq q
\]

\[
\sum_{t=p+1}^{T} \psi(\hat{\epsilon}_t / \hat{\sigma}) = 0, \quad (2.53)
\]

where \( \hat{\sigma} \) is a robust estimate of the scale of the residuals and \( \tilde{\gamma} = \sum_{t=p+1}^{T} \eta(\hat{\epsilon}_t / \hat{\sigma}, \hat{\epsilon}_{t-i} / \hat{\sigma}) \). Both \( \psi(\cdot) \) and \( \eta(\cdot) \) are bounded continuous functions with \( \eta(u, v) = \psi(u)\psi(v) \) or \( \eta(u, v) = \psi(uv) \). In fact, if \( \eta(u, v) = \psi(u)u \), then the RA estimates are equivalent to the M-estimates of equation 2.43. Also, if \( \psi(u) = u \) in the previous definition of \( \eta(\cdot) \), then the estimates simplify further into the least squares estimate.

Bustos and Yohai mention that although the RA estimates are robust for AR processes, they lose robustness if any MA component exists in the model. The RA estimates with MA models are still more robust than least squares or M-estimates, but gradually lose their effectiveness. They correct this issue with their truncated residual autocovariance (TRA) estimation method. Truncated residuals depend on only the previous \( k \) observations instead of all of the previous observations. These truncated residuals are defined by

\[
\hat{\epsilon}_{t,k} = \sum_{i=0}^{k} t_i \phi(L)(y_t - \mu)
\]

(2.54)

where \( t_i \) is defined in equations 2.53. The only difference with equation 2.53 is that for TRA estimates, replace all residuals, \( \hat{\epsilon} \), with truncated residuals, \( \hat{\epsilon}_{t,k} \). These TRA estimates are more robust than plain RA estimates with MA components in an ARMA model with additive outliers.

2.2.4 Multivariate Robust Estimation

Since the DCC-GARCH model describes multivariate covariance structures in data, we must consider the effects of outliers in multivariate data estimation. Instead of robustly estimating univariate variances, in multivariate data, robust estimates of entire covariance matrices are needed. Define the multivariate parameters of scatter as the matrix \( \Sigma \) with the multivariate estimation of location as \( \mu \). Some different proposals of robust multivariate estimation include M-estimation similar to previous sections, S-estimation, minimum covariance determinants (MCD), and minimum volume ellipsoids (MVE).
Maronna (1976) extends the M-estimators of equation 1.17 into the multivariate context. Maronna is interested in estimating the scatter matrix $\Sigma$ of a set of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of dimension $m$ with density function

$$f(\mathbf{x}) = |\Sigma|^{-1/2} h \left[ (\mathbf{x} - \mu)^\prime \Sigma^{-1} (\mathbf{x} - \mu) \right],$$

where the function $|\cdot|$ denotes the determinant of a matrix, and the function $h(\cdot)$ is a density. For this set of data, Maronna defines the multivariate M-estimators as the solutions to the set of equations

$$\frac{1}{n} \sum_{i=1}^{n} \rho_1 \left[ \left( \mathbf{x}_i - \mu \right)^\prime \Sigma^{-1} \left( \mathbf{x}_i - \mu \right) \right] = 0,$$

$$\frac{1}{n} \sum_{i=1}^{n} \rho_2 \left[ \left( \mathbf{x}_i - \mu \right)^\prime \Sigma^{-1} \left( \mathbf{x}_i - \mu \right) \right] \left( \mathbf{x}_i - \mu \right)^\prime = \Sigma,$$

where $\rho_1$ and $\rho_2$ are defined

$$\rho_1(s) = -\frac{1}{s} \frac{d \log(h(s))}{ds},$$

$$\rho_2(s^2) = \rho_1(s),$$

where $s$ is positive or positive semidefinite in the case of matrices. The functions $\rho_1$ and $\rho_2$ in the above equations relate to the functions $\psi_1$ and $\psi_2$ where $\psi_i(s) = s \rho_i(s)$ for $i = 1, 2$.

Properly selecting these $\psi$-functions makes the M-estimators more robust to outliers than the maximum likelihood estimates. Maronna proposes two sets of $\psi$-functions. The first set of $\psi$-functions are derived from Huber (1964) with

$$\psi_1(s) = \max(-c_1, \min(s, c_1)),$$

$$\psi_2(s^2) = \frac{\max(-c_2, \min(s, c_2^2))}{\max(-c_2^2, \min(||\mathbf{x}||^2, c_2^2))},$$

where $c_1$ and $c_2$ are positive constants and the function $||\cdot||$ is the Euclidean norm. The second set of $\psi$-functions is derived from the maximum likelihood estimation of the $t$-distribution with $\nu$ degrees of freedom

$$\psi_1(s) = \frac{s(m+\nu)}{\nu + s^2},$$

$$\psi_2(s^2) = \frac{s(m+\nu)}{(p+s)},$$

where $m$ is the length of the data vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Maronna (1976) proves these M-estimators are both consistent and asymptotically normal. However, Maronna does show that the breakdown point.
of the M-estimators is \((m + 1)^{-1}\). This implies that M-estimators of this form are not very robust for high dimensional data sets.

Lopuhaä (1989) extends the S-estimator approach of Rousseeuw and Yohai (1984) in Section 2.2.4 to the multivariate framework and compares it to the M-estimator approach developed by Maronna (1976) in equation 2.57. He mentions that although the M-estimator has some appealing asymptotic properties, the S-estimators keep these qualities and also possess a higher breakdown point than the M-estimators. Lopuhaä uses the relationship between S-estimators and M-estimators to develop and compare the influence functions and asymptotic properties. Lopuhaä starts with the same density function as Maronna in equation 2.55. He defines different M-estimators as the solutions to the set of equations

\[
0 = \frac{1}{n} \sum_{i=1}^{n} \rho_1 \left[ (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \right]^{1/2} (x_i - \mu), \quad (2.60)
\]

\[
0 = \frac{1}{n} \sum_{i=1}^{n} \rho_2 \left[ (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \right] (x_i - \mu)(x_i - \mu)' - \frac{1}{n} \sum_{i=1}^{n} \rho_3 \left[ (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \right]^{1/2} \Sigma, \quad (2.61)
\]

where \(\rho_1, \rho_2, \text{ and } \rho_3\) are real valued functions on the nonnegative portion of the real line. Although M-estimators have a breakdown point that decreases as the dimensions of the data increase, they still have a bounded influence function with properly selected \(\rho_1, \rho_2, \text{ and } \rho_3\).

Rousseeuw and Yohai (1984) introduce the S-estimator of equation 2.21 in a univariate regression context. Lopuhaä develops the S-estimator for the multivariate location and scatter problem as the solution when minimizing \(|\Sigma|\) subject to the equation

\[
\frac{1}{n} \sum_{i=1}^{n} \rho \left[ (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \right]^{-1/2} = b, \quad (2.62)
\]

where the constant \(0 < b < \sup \rho\).

Lopuhaä applies the method of Lagrange multipliers to transform the above multivariate S-estimators to the minimization of

\[
\log |\Sigma| - \lambda \left[ \frac{1}{n} \sum_{i=1}^{n} \rho \left[ (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \right]^{-1/2} - b \right],
\]
with $\lambda$ defined as the corresponding Lagrange multiplier. This implies that the parameter estimates $\hat{\theta}$ and $\hat{\lambda}$ satisfy the set of equations

$$0 = \frac{\lambda}{n} \sum_{i=1}^{n} u \left[ \left( (x_i - \mu)'\Sigma^{-1}(x_i - \mu) \right)^{1/2} \right] \Sigma^{-1}(x_i - \mu),$$

$$0 = 2\Sigma^{-1} - \text{diag}(\Sigma^{-1}) + \frac{\lambda}{2n} \sum_{i=1}^{n} u \left[ \left( (x_i - \mu)'\Sigma^{-1}(x_i - \mu) \right)^{1/2} \right] (2V_i - \text{diag}(V_i)),$$

where $u(k) = \psi(k)/k$, with $\psi(\cdot)$ as the derivative of $\rho(\cdot)$, and $V_i = \Sigma^{-1} (x_i - \mu)(x_i - \mu)'\Sigma^{-1}$. Lopuhaä solves for $\lambda$ and substitutes it back into these equations to find that $\theta$ is the solution to the system of equations

$$0 = \frac{1}{n} \sum_{i=1}^{n} u \left[ \left( (x_i - \mu)'\Sigma^{-1}(x_i - \mu) \right)^{1/2} \right] (x_i - \mu),$$

$$0 = \frac{1}{n} \sum_{i=1}^{n} m u \left[ (x_i - \mu)'\Sigma^{-1}(x_i - \mu) \right] (x_i - \mu)(x_i - \mu) - \frac{1}{n} \sum_{i=1}^{n} u \left[ \left( (x_i - \mu)'\Sigma^{-1}(x_i - \mu) \right)^{1/2} \right] \Sigma,$$

where $u(k) = \rho'(k)/k - \rho(k) + b$ and $m$ is the dimension of the data.

These equations have the same structure as M-estimators. The above equations can be rewritten in the form

$$\frac{1}{n} \sum_{i=1}^{n} \Psi(x_i, \theta) = 0,$$

where $\Psi = (\Psi_1, \Psi_2)$ is the function

$$\Psi_1 = u \left[ \left( (x_i - \mu)'\Sigma^{-1}(x_i - \mu) \right)^{1/2} \right] (x_i - \mu),$$

$$\Psi_2 = m u \left[ (x_i - \mu)'\Sigma^{-1}(x_i - \mu) \right] (x_i - \mu)(x_i - \mu)' - u \left[ \left( (x_i - \mu)'\Sigma^{-1}(x_i - \mu) \right)^{1/2} \right] \Sigma.$$
an elliptical distribution minimize the function

\[ \frac{1}{n} \sum_{i=1}^{n} \left[ \rho \left( \| \Sigma^{-1/2}(x_i - \mu) \| \right) + \frac{1}{2} \log |\Sigma| \right], \]

or are solutions to the equation

\[ \frac{1}{n} \sum_{i=1}^{n} w(||\Sigma^{-1/2}(x_i - \mu)||) \left( ||\Sigma^{-1/2}(x_i - \mu)||^{-1} \Sigma^{-1}(x_i - \mu) \right) \left( ||\Sigma^{-1/2}(x_i - \mu)||^{-1} \Sigma^{-1}(x_i - \mu) \right)' = I_m, \]

where \( w(k) = \rho'(|k|) k \), and \( m \) is the dimension of the data vectors. Without loss of generality, Sirkiä et al. define \( \mu = 0 \). They mention that Maronna’s M-estimator defined in equation 2.57 takes the form

\[ \frac{1}{n} \sum_{i=1}^{n} \left( w_1(||\Sigma^{-1/2}x_i||) \left( ||\Sigma^{-1/2}x_i||^{-1} \Sigma^{-1}x_i \right) \right) \left( ||\Sigma^{-1/2}x_i||^{-1} \Sigma^{-1}x_i \right)' = \frac{1}{n} \sum_{i=1}^{n} w_2(||\Sigma^{-1/2}x_i||)I_m. \]

The symmetrized M-estimators focus on the pairwise differences of the observations. By focusing their attention on pairwise differences, Sirkiä et al. create a situation where it is not necessary to impose any location parameter on the model because all pairwise differences in the multivariate setting make univariate projections symmetric with a location of 0. Therefore, the symmetrized M-estimator of scatter, \( \hat{\Sigma}_S \), from a sample of data \( x_1, \ldots, x_n \) is the solution to

\[ \min_{\Sigma} \left( \frac{n}{2} \right)^{-1} \sum_{i<j} \left[ \rho \left( ||\Sigma^{-1/2}d_{i,j}|| \right) + \frac{1}{2} \log |\Sigma| \right], \tag{2.64} \]

where \( d_{i,j} = x_i - x_j \). The symmetrized M-estimator of scatter \( \hat{\Sigma}_S \) is also the solution of either one of the equations

\[ 0 = \left( \frac{n}{2} \right)^{-1} \sum_{i<j} \left[ w(||\Sigma^{-1/2}d_{i,j}||) \left( ||\Sigma^{-1/2}d_{i,j}||^{-1} \Sigma^{-1}d_{i,j} \right) \left( ||\Sigma^{-1/2}d_{i,j}||^{-1} \Sigma^{-1}d_{i,j} \right)' - I_m \right], \]

\[ 0 = \left( \frac{n}{2} \right)^{-1} \sum_{i<j} w_1(||\Sigma^{-1/2}d_{i,j}||) \left( ||\Sigma^{-1/2}d_{i,j}||^{-1} \Sigma^{-1}d_{i,j} \right) \left( ||\Sigma^{-1/2}d_{i,j}||^{-1} \Sigma^{-1}d_{i,j} \right)' - \left( \frac{n}{2} \right)^{-1} \sum_{i<j} w_2(||\Sigma^{-1/2}d_{i,j}||)I_m. \]

Sirkiä mentions that the influence functions for M-estimators and symmetrized M-estimators are not equal to each other. However, both influence functions are bounded.

In the same fashion that Lopuhaä (1989) compares multivariate S- and M-estimators, Roelant et al. (2009) create a symmetrized version of S-estimators they call Generalized S-estimators (GS-estimators)
and compare them to the symmetrized M-estimator of Sirkiai et al. (2007). By focusing on the pairwise differences of the residuals in the multivariate regression model, Roelant et al. eliminate the need of estimating the location parameter of multivariate data and only estimate the scatter matrix. Since S-estimators have the flexibility and asymptotic properties of M-estimators with a higher breakdown point for robust estimation, Roelant et al. create a similar symmetrized version of the estimator to improve the symmetrized M-estimator. They define the GS-estimator as a solution to the equation

$$
\left( \frac{n}{2} \right)^{-1} \sum_{i<j} \rho \left( \left[ (\hat{\xi}_i - \hat{\xi}_j)^\top \Sigma^{-1} (\hat{\xi}_i - \hat{\xi}_j) \right]^{1/2} \right) = b,
$$

where $\hat{\xi}_i = y_i - x_i^\top \beta$.

Similarly Lopuhaä (1989) and Roelant et al. (2009) note that these multivariate GS-estimators satisfy the equations

$$
0 = \sum_{i<j} u \left[ \left[ (\hat{\xi}_i - \hat{\xi}_j)^\top \Sigma^{-1} (\hat{\xi}_i - \hat{\xi}_j) \right]^{1/2} \right] (x_i - x_j)(\hat{\xi}_i - \hat{\xi}_j),
$$

$$
0 = \sum_{i<j} m u \left[ \left[ (\hat{\xi}_i - \hat{\xi}_j)^\top \Sigma^{-1} (\hat{\xi}_i - \hat{\xi}_j) \right]^{1/2} \right] (\hat{\xi}_i - \hat{\xi}_j)^\top (\hat{\xi}_i - \hat{\xi}_j) - \sum_{i<j} u \left[ \left[ (\hat{\xi}_i - \hat{\xi}_j)^\top \Sigma^{-1} (\hat{\xi}_i - \hat{\xi}_j) \right]^{1/2} \right] \Sigma,
$$

where $u(k) = \rho'(k)/k - \rho(k) + b$ and $m$ is the dimension. This implies that the GS-estimators of scatter have the same form as the symmetrized M-estimators of scatter. Since this is true, GS-estimators have the same influence function and asymptotic properties as the symmetrized M-estimators. However, the GS-estimator of scatter still has a breakdown point of $\delta^* = 0.5$, which makes it more robust than symmetrized M-estimators, whose breakdown point decreases as the dimension of the data increases. Roelant et al. mention that the disadvantage of the GS-estimators in relation to the multivariate S-estimators is that they take a tremendous amount of time to calculate.

Rousseeuw (1984) takes a different approach to robustly estimating covariance matrices. Two of the approaches suggested by Rousseeuw are the minimum volume ellipsoid (MVE) and the minimum covariance determinant (MCD). The MVE method looks for the ellipsoid with the smallest volume that covers $h$ observations with $n/2 < h < n$. This approach has a breakdown point of approximately $\delta^* = (n - h)/n$. This implies that the breakdown point is at most $\delta^* = 0.5$. There are many proposals of algorithms to solve the MVE, yet most are extremely time consuming.

Rousseeuw also proposes the minimum covariance determinant (MCD) method that finds the $h$ observations with the lowest value for the determinant of the covariance matrix. The sample covariance matrix of these $h$ points is the MCD estimator of the covariance. The MCD estimator has the same breakdown point as the MVE estimator, but unlike the MVE estimator, the MCD estimator is
asymptotically normal. Rousseeuw and Van Driessen (1999) note that the MCD method is also more efficient than the MVE method because the MVE is only $n^{1/3}$ consistent instead of $n^{1/2}$ like the MCD. Again, many algorithms are proposed to compute the MCD including the FAST-MCD algorithm of Rousseeuw and Van Driessen (1999).

Peña and Prieto (2001) state that the above approaches to robustly estimating multivariate covariance matrices are too time consuming. They mention that most outlier detection techniques rely on the fact that in multivariate contaminated samples, each outlier must be an extreme point along the direction from the mean of the uncontaminated data to the outlier. This requires large numbers of randomly generated directions that lead to heavy computation in the estimation. They propose a method to try and optimize these procedures by limiting the number of appropriate directions to identify outliers. They choose these directions based on kurtosis coefficients of the projected observations. Outliers generated by a usual symmetric contaminated model increase the kurtosis coefficient. Small groups of outliers in asymmetric contaminated models also increase kurtosis coefficients. However, large groups of outliers in asymmetric models could actually decrease the kurtosis coefficients to the minimized value. Therefore, Peña and Prieto suggest searching for outliers by projecting data onto the directions that maximize or minimize the kurtosis coefficient.

Let $X$ be a mixture distribution of the uncontaminated distribution $F$ and the contamination distribution $G$ such that $X = (1 - \epsilon)F + \epsilon G$, where $\epsilon$ is the fraction of outliers in the generated data. Define $\gamma_X$, $\gamma_F$, and $\gamma_G$ as the kurtosis coefficients of the distributions $X$, $F$, and $G$ respectively. Under reasonable assumptions, Peña and Prieto derive the kurtosis coefficient of $X$, $\gamma_X$, in explicit detail found in Peña and Prieto (2001). If both distributions $F$ and $G$ have the same mean, then the kurtosis coefficient of $X$ is

$$\gamma_X = \frac{\gamma_F \left(1 + \epsilon \left(\theta \frac{m_G m_G(2)^2}{m_F m_F(2)^2} - 1\right)\right)}{\left(1 + \epsilon \left(m_G(2) m_F(2)^2 - 1\right)\right)^2},$$

where $m_i(j)$ is the $j^{th}$ centered moment of the distribution $i$. However, if the distributions $F$ and $G$ have different means, then the kurtosis coefficient of $X$ approaches

$$\gamma_X \to \frac{\epsilon^3 + (1 - \epsilon)^3}{\epsilon(1 - \epsilon)}.$$

Peña and Prieto (2001) describe the advanced algorithm to analyze different projections of the data to find the directions for kurtosis analysis. The details are left for the reader to research in Peña and Prieto (2001). Once the directions for possible outliers are calculated, these projections are studied to find possible outliers points. To calculate a robust covariance matrix estimate, Peña and Prieto calculate a sample covariance estimate of the data points that are not labeled outliers after the previous process is complete. Since there are fewer directions to make calculations than either the MVE or MCD, the kurtosis coefficient calculation is faster.
2.3 Robust Estimation of ARCH / GARCH

The ARCH/GARCH model estimation is inherently hindered by outliers because of the autoregressive design and maximum likelihood estimation of the models. Many robust methods for general linear models, time series models, and multivariate estimation are previously displayed in this chapter. Over the past decade, there are many attempts to apply these and other techniques to creating more robust estimation of ARCH/GARCH models.

Franses and Ghijsels (1999) note the usefulness of GARCH models to capture the features of financial data such as excess kurtosis. They mention that Bollerslev’s GARCH-t model mentioned in Section 2.1.1 does a reasonable job at measuring the excess kurtosis in financial data. However, some residuals still exhibit excess kurtosis even with distributional accommodations. They propose that certain observations are additive outliers (AO) not accounted for in standard GARCH modeling approaches. These additive outliers lead to incorrect parameter estimation and incorrect forecasts of volatility. Additive outliers were first introduced in Section 1.2 and described further in Section 2.2.3.

Franses and Ghijsels rewrote the GARCH model into the form of an ARMA model and took the same approach as Chen and Liu (1993) (see Section 2.2.3) to handle additive outliers. Franses and Ghijsels alter Bollerslev’s GARCH model in equation 2.3 by assuming the asset returns follow a normal distribution with a possibility of additive outliers. They transformed the GARCH(1,1) structure into its ARMA formulation describing the squared returns

\[ r_t^2 = \alpha_0 + (\alpha_1 + \beta_1) r_{t-1}^2 + \nu_t - \beta_1 \nu_{t-1}, \]

similar to Engle and Bollerslev in equation 2.10 where \( \nu_t = r_t^2 - h_t \) is a sequence of uncorrelated random variables. They notice that the process \( \nu_t \) is heterogeneous as well.

The process of robustly estimating the parameters is similar to Chen and Liu’s approach in their ARMA additive outlier model. The first step estimates the parameters in the model with maximum likelihood to obtain \( \hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1, \) and \( \hat{h}_t \). The estimate \( \hat{\nu}_t = r_t^2 - \hat{h}_t \) comes from the previous parameter estimates. The next step in the process calculates the estimated polynomial

\[ \pi(L) = \frac{1 - (\hat{\alpha}_1 + \hat{\beta}_1)L}{1 - \hat{\beta}_1 L} = (1 - \pi_1 L - \pi_2 L^2 - \ldots), \]

where \( L \) is the lag operator. This polynomial estimation leads to the estimation of the weight \( \omega \) with

\[ \hat{\omega}(\tau) = \frac{\sum_{t=\tau}^{n} \hat{\nu}_t x_t}{\sum_{t=\tau}^{n} x_t^2}, \]  

(2.67)
with \( x_t \) defined the same as equation 2.33. From the weight estimate \( \hat{\omega} \), the estimate of \( \hat{\tau} \) is calculated

\[
\hat{\tau} = \frac{\hat{\omega}(\tau)}{\hat{\sigma}_v} \left( \sum_{t=\tau}^{n} x_{1,t}^2 \right)^{1/2},
\]

where \( \hat{\sigma}_v \) is the omit-one variance. The observation at \( t = \tau \) with the largest value of \( \hat{\tau} \), which should exceed the value of 4 according to Chen and Liu, is replaced by \( \hat{\tau}_t^* = \hat{\tau}_t - \hat{\omega} \). Calculate the new robust estimate \( r_t^* \) with the estimates \( \hat{\tau}_1^* \) and \( \hat{h}_t \). The robust estimate of the return becomes

\[
r_t^* = \text{sign}(r_t)(r_t^{*2})^{1/2},
\]

repeat the above process until no value of the statistic \( \hat{\tau} \) exceeds the value of 4. With the final robust series of returns \( r_t^* \), calculate the new robust parameter estimates for the GARCH model with maximum likelihood estimation.

The process above by Francses and Ghijsels accounts only for additive outliers. Charles and Darne (2005) extend the Francses and Ghijsels additive outlier approach to robust estimation to include innovation outliers. They specifically deal with the GARCH(1,1) process as do Francses and Ghijsels. They state that when an additive or innovation outlier occurs, \( \tilde{r}_t \) is observed instead of the true value \( r_t \). This new observation is defined

\[
\tilde{r}_t^2 = r_t^2 + \omega_i \xi_i(L) \mathbb{I}_i(\tau),
\]

with \( i = 1, 2 \) for additive and innovation outliers respectively, \( \mathbb{I}_i(\tau) \) as the indicator function that takes a value of one at the point an outlier occurs and zero otherwise, and \( \xi_i(L) \) as the magnitude and pattern of the outlier. The equation for \( \xi_i(L) \) is defined

\[
\xi_1(B) = \frac{1 - \hat{\beta}_1 L}{1 - (\hat{\alpha}_1 + \hat{\beta}_1)L} = \hat{\pi}(L)^{-1},
\]

\[
\xi_2(B) = \frac{1 - \hat{\beta}_1 L}{1 - (\hat{\alpha}_1 + \hat{\beta}_1)L} = \hat{\pi}(L)^{-1},
\]

where \( \hat{\pi}(L) \) is defined the same as Francses and Ghijsels above. Instead of only using a single weight calculation as Francses and Ghijsels, Charles and Darne have a weight for additive and innovation outliers. They defined

\[
\hat{\tau}_1 = \frac{\hat{\omega}_1(\tau)}{\hat{\sigma}_v} \left( \sum_{t=\tau}^{n} x_{1,t}^2 \right)^{1/2},
\]

\[
\hat{\tau}_2 = \frac{\hat{\omega}_2(\tau)}{\hat{\sigma}_v},
\]

with \( \hat{\tau}_1 \) and \( \omega_i \) the same as Francses and Ghijsels in equation 2.67 with \( x_{i,t} = 0 \) for \( i = 1, 2 \) and \( t < \tau, x_{i,t} = 1 \) for \( i = 1, 2 \) and \( t = \tau, x_{1,t+k} = -\pi_k \) and \( x_{2,t+k} = 0 \) for \( t > \tau \) and \( k > 0 \).

Charles and Darne use the same iteration approach as Francses and Ghijsels to recalculate robust estimates of the returns to get better estimates of the GARCH process. One difference between the
two iteration processes is that Frances and Ghijsels only calculate one value of \( \hat{\tau} \), while Charles and Darne calculate \( \hat{\tau}_1 \) and \( \hat{\tau}_2 \) for each point and take the maximum value of these two estimates. If the maximum of these estimates exceed a certain prespecified critical value \( C \), then the return at that point is replaced. For additive outliers the squared robust returns are replaced with \( r_t^2 - \hat{\omega}_1 \), while innovation outliers have the squared returns replaced with \( r_{t+j}^2 - \hat{\omega}_2 \hat{\tau}_j(L)^{-1} \). Repeat this iteration process until the maximum of the two estimates of \( \tau \) are below the critical value and estimate the new robust GARCH parameters.

Park (2002) also references Frances and Ghijsels’ claims of traditional estimation techniques of the GARCH model not accounting for possible additive outliers in the data. He notes that this leads to limitations in the volatility forecasting ability of the GARCH model under traditional estimation techniques. Park proposes the use of a least absolute deviation (LAD) approach of estimating the GARCH model as an alternate solution that accounts for additive outliers in the data as well as a lack of normality.

Instead of modeling the conditional variance of the returns, as in the definition of the univariate GARCH model from equation 2.3, Park models the conditional standard deviation with a GARCH structure as follows:

\[
\sqrt{h_t} = a_0 + \sum_{i=1}^{p} a_i |r_{t-i}| + \sum_{i=1}^{q} \beta_i \sqrt{h_{t-i}}. \quad (2.71)
\]

This is the ARMACH model mentioned earlier in equation 2.17. This model works with the absolute value of the returns instead of the squared returns in the GARCH model.

Define \( \theta = (a_0, a_1, \ldots, a_p, \beta_1, \ldots, \beta_q)' \) as the parameter vector of this model. Park uses a Taylor series approach explained in detail in Park (2002) to estimate the parameters as

\[
\hat{\theta}_{Park} = \arg \min_{\theta} \sum_{t=1}^{T} \left| r_t - a_0 - \sum_{i=1}^{p} a_i |r_{t-i}| - \sum_{i=1}^{q} \beta_i \sqrt{h_{t-i}} \right|. \quad (2.72)
\]

Park mentions that this objective function is not differentiable, which implies the need of minimization algorithms not based on differentiation.

Park compares his estimation technique to that of a GARCH model and a random walk model. The LAD approach to estimating the GARCH model out-performs the other two models in simulation studies with respect to mean square error (MSE) and mean absolute error (MAE) as the number of outliers increases. The LAD approach performs comparably when outliers do not exist.

Peng and Yao (2003) also take an absolute value estimation approach to robustly estimating the parameters in a GARCH model. They focus on a quasi-maximum likelihood point of view where a Gaussian likelihood assumption in quasi-maximum likelihood is sensitive to heavy tails. An LAD approach in quasi-maximum likelihood estimation provides a more robust estimate. Under the assumption of conditional normality of \( r_{t+1} \) with \( \psi \leq t \leq \nu \sim N(0, h_t) \) with \( \nu \geq \max(p, q) \), the conditional
density function of \( r_{\nu+1}, \ldots, r_T \) is proportional to
\[
\left( \prod_{t=\nu+1}^{T} h_t \right)^{-1/2} \exp \left( -\frac{1}{2} \sum_{t=\nu+1}^{T} \frac{r_t^2}{h_t} \right). \tag{2.73}
\]
Maximizing this quantity leads to the quasi-maximum likelihood estimates of the GARCH model with
\[
\hat{\theta}_{\text{MLE}} = \arg \min_{\theta} \sum_{t=\nu+1}^{T} \left( \frac{r_t^2}{\tilde{h}_t} + \log \tilde{h}_t \right),
\]
where \( \tilde{h}_t \) is an \( r_1, \ldots, r_t \) based approximation of \( h_t \) defined in Peng and Yao (2003). Peng and Yao reparameterize the returns so that the squared returns now have a median equal to 1 instead of the variance of the returns equaling 1. They also adjusted the structure of the returns from \( r_t = \sqrt{h_t} \epsilon_t \) to
\[
\log(r_t^2) = \log(h_t) + \log(\epsilon_t^2). \tag{2.74}
\]
This leads to the creation of their quasi-maximum likelihood based least absolute deviations estimator
\[
\hat{\theta}_{\text{PY}} = \arg \min_{\theta} \sum_{t=\nu+1}^{T} \left| \log(r_t^2) - \log(\tilde{h}_t) \right|, \tag{2.75}
\]
since \( \text{median} (\log(\epsilon_t^2)) = 0 \). They show that this estimator is both unbiased and asymptotically normal under certain conditions. Their estimation technique out-performs the maximum likelihood estimates when the normality assumption is broken, but performs worse than the MLE under normality with no outliers.

Muler and Yohai (2002) adapt the robust \( \tau \)-estimates of Yohai and Zamar (1988) mentioned in Section 2.2.1 to the ARCH model. Muler and Yohai consider only a model that corrects isolated additive outliers in the data. The robust scale approach of \( \tau \)-estimates obtains both high breakdown point and high efficiency under a normality assumption. Muler and Yohai reparameterize the traditional ARCH model
\[
h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \delta_{t-i}^2,
\]
by defining \( \eta = \sum_{i=0}^{p} \alpha_i \) and \( \delta_i = \alpha_i / \eta \) with \( \delta_0 = 1 - \sum_{i=1}^{p} \delta_i \). Obtain the maximum likelihood estimates of \( \delta_1, \ldots, \delta_p \) by solving
\[
\arg \min \left[ \log \left( \frac{1}{n-p} \sum_{t=p+1}^{n} \delta_{0} + \sum_{i=1}^{p} \delta_i \delta_{t-i}^2 \right) + \frac{1}{n-p} \sum_{t=p+1}^{n} \log \left( \delta_{0} + \sum_{i=1}^{p} \delta_i \delta_{t-i}^2 \right) \right],
\]
as shown in detail in Muler and Yohai (2002). Calculate the estimates of \( \alpha_0, \ldots, \alpha_p \) from the estimates.
of $\delta_0, \ldots, \delta_p$.

Muler and Yohai use the same $\tau$-estimate approach of Yohai and Zamar (1988) where they define the scale estimate $s$ as the solution to

$$\frac{1}{n} \sum_{i=1}^{n} \rho \left( \frac{\xi_i}{s} \right) = E\phi(\rho(\xi_i)), \quad \phi \sim N(0, 1).$$

Define the $\tau$-estimate

$$\tau_n(\xi) = s_n \left( \frac{1}{n} \sum_{i=1}^{n} \rho \left( \frac{\xi_i}{s_n(\xi)} \right) \right)^{1/2},$$

(2.76)

where $s_n$ is the scale estimate previously described, and $\xi = (\xi_1, \ldots, \xi_n)$. Robust estimates of the parameters $\delta_0, \ldots, \delta_p$ are obtained by replacing

$$\frac{1}{n-p} \sum_{t=p+1}^{n} \frac{r_t^2}{\delta_0 + \sum_{i=1}^{p} \delta_i r_{t-i}^2}$$

in the above likelihood equation with

$$\tau_n^2 \left( \frac{r_{p+1}}{\delta_0 + \sum_{i=1}^{p} \delta_i r_{p+1-i}^2} \right)^{1/2}, \ldots, \left( \frac{r_n}{\delta_0 + \sum_{i=1}^{p} \delta_i r_{n-i}^2} \right)^{1/2}.$$
modeling the log of the squared returns, similar to Peng and Yao (2003). Muler and Yohai generalize the minimizing equation 2.75 of Peng and Yao to an M-estimation structure of

\[ M_T = \frac{1}{T - p} \sum_{t=p+1}^{T} \rho(\log(r_t^2) - \log(h_t)). \]  

(2.77)

Minimizing this quantity leads to the M-estimates, \( \hat{\gamma}_1 \) for the GARCH model where \( \gamma \) is a vector of the parameters \( (\alpha_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^T \). When the function \( \rho \) is unbounded, but the derivative of \( \rho \) is bounded, the estimates are robust in the presence of a heavy tailed residual distribution, but still affected by outliers. When both the function \( \rho \) and its derivative are bounded, the estimates are robust to any form of outlier.

However, even when both \( \rho \) and its derivative are bounded, outliers may still affect the quasi-maximum likelihood estimation, because outliers at time \( t \) may affect estimates of the conditional variance at time \( t + i \) with \( i > 0 \). Muler and Yohai correct this problem with the use of robust filtering similar to their approach in Muler and Yohai (2002). Muler and Yohai replace \( h_t \) in the calculation of \( M_T \) with a filtered version defined by

\[ h_t^* = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{t-i}^* g_k \left( \frac{r_{t-i}^2}{h_{t-i}^*} \right) + \sum_{i=1}^{q} \beta_i h_{t-i}^*, \]

where the function \( g_k \) is defined

\[ g_k(u) = \begin{cases} 
  u & \text{if } u \leq k, \\
  k & \text{if } u > k.
\end{cases} \]

Minimizing the new filtered \( M_T^* \) leads to the robust estimate \( \hat{\gamma}_2 \). The final BM-estimate of Muler and Yohai combines \( \hat{\gamma}_1 \) and \( \hat{\gamma}_2 \) as

\[ \hat{\gamma}^B_T = \begin{cases} 
  \hat{\gamma}_{1,T} & \text{if } M_T(\hat{\gamma}_{1,T}) \leq M_T^*(\hat{\gamma}_{2,T}), \\
  \hat{\gamma}_{2,T} & \text{if } M_T(\hat{\gamma}_{1,T}) > M_T^*(\hat{\gamma}_{2,T}).
\end{cases} \]

Muler and Yohai (2008) prove this estimator is highly efficient and asymptotically normal. They also compare their estimator to Peng and Yao’s estimator, the traditional QMLE, and the maximum likelihood estimator with a \( t \)-distribution with 3 degrees of freedom. Their model outperforms each of these estimates with regards to MSE.

Boudt and Croux (2009) use Muler and Yohai’s filter structure with quasi-maximum likelihoods in their recent approach to creating a robust multivariate GARCH model. They use the multivariate structure mentioned in equation 2.11 where

\[ r_t = H_t^{1/2} \epsilon_t, \]
with \( \text{Var}(\varepsilon_t) = I_k \). Under this specification, the quasi-maximum likelihood estimator of the parameter vector \( \theta \) is defined as

\[
\hat{\theta}_{\text{QMLE}} = \arg\max_{\theta} \frac{1}{T} \sum_{t=1}^{T} \left( -\log|H_t| + 2 \log g(r_t' H_t^{-1} r_t) \right),
\]

(2.78)

where \( g(\cdot) \) is an assumed density function. This quasi-maximum likelihood function generalizes to the broader class of M-estimators defined as

\[
M = \frac{1}{T} \sum_{t=1}^{T} \left( \log|H_t| + \sigma \rho(r_t' H_t^{-1} r_t) \right),
\]

(2.79)

where the function \( M \) is minimized, with the continuous, positive and nondecreasing scalar loss function \( \rho(\cdot) \).

Boudt and Croux combine two approaches to make a more outlier robust estimation method for a multivariate GARCH model. The first approach is to use a loss function \( \rho(\cdot) \) that downweights the observations with large Mahalanobis Distances where \( MD = r_t' H_t^{-1} r_t \). The \( t \)-distribution leads to a value of \( \rho'(z) = \frac{N-\nu}{\nu-2+z} \), which approaches zero as the value of \( z \) approaches infinity. Boudt and Croux use a \( t \)-distribution with four degrees of freedom in their research.

The second approach of Boudt and Croux bounds the effect of outliers at a point \( t \) on the volatility at point \( t+i \) with \( i > 0 \). They use a similar filtering approach as Muler and Yohai (2002) and Muler and Yohai (2008). They replace the matrix \( H_t \) with a filtered version calculated as

\[
\tilde{H}_t = H_t(\tilde{r}_1, \ldots, \tilde{r}_{t-1}), \quad \tilde{r}_t = r_t \sqrt{w(r_t' \tilde{H}_t^{-1} r_t)},
\]

(2.80)

where \( w(\cdot) \) is a weight function that bounds the effect of \( r_t \) on \( \tilde{H}_t \). Boudt and Croux combine these two methods and the weight function

\[
w(z) = \begin{cases} 
1 & \text{if } z \leq c_1 \\
1 - (1 - c_1/z)^3 & \text{if } c_1 < z \leq c_2 \\
(c_2/z)(1 - (1 - c_1/c_2)^3) & \text{else.}
\end{cases}
\]

in creating a robust BEKK model, where the BEKK model is defined in equation 2.13. They conclude that their robust model – called BIP-BEKK – produces parameter estimates relatively close in efficiency to the QMLE in the presence of no outliers. The BIP-BEKK has better efficiency in the presence of outliers. However, they conclude this is due more to the \( t \)-distribution assumption than the filtering approach because a \( t \)-distributed BEKK without the filtered approach produces similar results.
2.4 Symmetry of Elliptical Distributions

Various issues arise in statistics where observations exist in more than one dimension. Financial data provides opportunities to examine volatilities of assets in multiple dimensions. With univariate data, the normal distribution along with certain other distributions can be generalized to a symmetric location distribution to provide for flexibility in data modeling. In the multivariate setting, the multivariate normal distribution may be generalized to a symmetrical elliptical distribution for the same flexibility. Beran (1979) mentions that observations with this distribution have the following density:

\[ |\Sigma|^{-1} h(\Sigma^{-1/2}(x - \mu)) \]  

where \( \mu \) is a \( p \times 1 \) vector, \( \Sigma \) is a \( p \times p \) nonsingular matrix, and the function \( h(\cdot) \) is a spherical symmetric density on the space \( \mathbb{R}^p \) centered at the origin. A vector \( x \) has a spherically symmetric distribution if for every orthogonal matrix \( O \), \( Ox \) has the same distribution as \( x \).

Since elliptical distributions have desirable properties for estimation, a test of elliptical symmetry is valuable. Beran proposes a test of elliptical symmetry for \( p \)-dimensional data. Define the sample of \( p \)-dimensional vectors \( x_1, \ldots, x_n \) from the space \( \mathbb{R}^p \). Under the hypothesis of elliptical symmetry, each observation \( x_i \) has the density in the form of equation 2.81. He defined \( R_i(\hat{\mu}, \hat{\Sigma}) \) as the ranks of 

\[ \frac{|\hat{\Sigma}^{-1/2}(x_i - \hat{\mu})|}{n + 1}, \]

where \( \hat{\mu} \) and \( \hat{\Sigma} \) are estimates of the parameters \( \mu \) and \( \Sigma \) respectively, such as maximum likelihood estimates. Transform the estimated direction vector 

\[ \frac{\Sigma^{-1/2}(x_i - \hat{\mu})}{|\Sigma^{-1/2}(x_i - \hat{\mu})|}, \]

into the polar coordinate representation \( \Theta_i(\hat{\mu}, \hat{\Sigma}) \). Use these two quantities to calculate the test statistic

\[ S(\hat{\mu}, \hat{\Sigma}) = \sum_{k=1}^{K} \sum_{m=1}^{M} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_k(R_k)b_m(\Theta_i) \right)^2, \]

where \( a_k \) and \( b_m \) are families of orthonormal functions with respect to the Lebesgue measure on the unit interval and the uniform measure on \([0, \pi]^{p-2} \times [0, 2\pi]\) respectively. Beran mentions that large values of this test statistic imply that the data are not from an elliptical distribution. The statistic \( S \) follows an asymptotically normal distribution if the true parameters \( \mu \) and \( \Sigma \) are known.

Li et al. (1997) develop a visual test for elliptical symmetry of multivariate distributions instead. They adapt the popular Q-Q plot technique to test for elliptical and spherical symmetry. A \( p \)-dimensional random vector \( x_i \) is a symmetric elliptical distribution with parameters \( \mu \) and \( \Sigma \) if its
characteristic function \( \psi(t) \) has the structure

\[
\psi(t) = e^{i t^\top \mu} \phi(t^\top \Sigma t),
\]

for a scalar function \( \phi(\cdot) \). This form represents a symmetric spherical distribution if \( \mu = (0) \) and \( \Sigma = I_p \).

They first develop a test of spherical symmetry called the \( t \)-plot. They calculate

\[
t(x) = \frac{\sqrt{p} \bar{X}}{s}, \tag{2.83}
\]

where \( x = (X_1, \ldots, X_p) \), \( \bar{X} = \frac{1}{p} \sum_{j=1}^{p} X_j \), and \( s = \sqrt{\frac{1}{p-1}(\sum_{j=1}^{p} X_j - \bar{X})^2} \). The test statistic \( t(x) \) follows a \( t_{p-1} \) distribution. Therefore, the test statistics

\[
Z_i = t(x_i) = \frac{\sqrt{p} \bar{X}_i}{s_i}, \tag{2.84}
\]

are independent and identically distributed observations according to the \( t_{p-1} \) distribution. Li et al. (1997) tested elliptical symmetry by plotting the order statistics \( Z_{(i)} \) against the quantiles of the \( t_{p-1} \) distribution with the same approach of a Q-Q plot. If the distribution is spherical, the points on the Q-Q plot will be relatively linear across the 45\(^o\) diagonal.

Li et al. (1997) generalize the test of spherical symmetry to test for elliptical symmetry. If \( x \) follows an elliptical distribution instead of a spherical distribution, then they transform the elliptically distributed variable into a spherically symmetric variable by

\[
y = \Sigma^{-1/2}(x - \mu), \tag{2.85}
\]

where \( \mu \) and \( \Sigma \) may be estimated. Once these vectors are spherically symmetric, the same technique mentioned above is applied to the vector \( y \) with the Q-Q plot now testing for elliptical symmetry instead of spherical symmetry.

Another approach to handling observations in more than one dimension uses the surface of a unit sphere as a sample space to describe the directions in space of the observations. A test of uniformity on the unit sphere would test if the directions of the observations are uniformly positioned in space like a spherically symmetric distribution. The theory of spherical inference is an extension to inference of circular data in \( p \) dimensions, according to Mardia and Jupp (2000). Directions in \( p \) dimensions on the unit sphere are represented with unit vectors, \( x \), centered at zero with length one on the \( (p - 1) \) dimensional sphere. Define the sample mean vector in \( \mathbb{R}^p \) as

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{2.86}
\]
To express this mean in polar coordinates with direction and distance, define the mean direction of the sample as

$$\bar{x}_0 = ||\bar{x}||^{-1} \bar{x}$$

and the mean resultant length as

$$\bar{R} = ||\bar{x}||$$

Rayleigh develops one of the most basic tests of uniformity on a sphere. This is just a generalization of Rayleigh’s test of uniformity on a circle as defined in Mardia and Jupp (2000). Rayleigh assumes it is reasonable to reject uniformity on a sphere when the vector sample mean $\bar{x}$ is far from the vector $0$ because when $x$ follows the uniform distribution, $E[x] = 0$. When $\bar{x}$ is far from the vector $0$, the mean resultant length $\bar{R}$ is large. Therefore, the test statistic for the Rayleigh test of uniformity is

$$S = pn\bar{R}^2,$$  \hspace{1cm}(2.87)

which under the null hypothesis of uniformity has the following asymptotic distribution:

$$pn\bar{R}^2 \sim \chi^2_p.$$  \hspace{1cm}(2.88)

Jupp later modifies Rayleigh’s test of uniformity to reduce the error in the approximation from $O(n^{-1/2})$ to $O(n^{-1})$. The modified Rayleigh statistic $S^*$ is given by

$$S^* = \left(1 - \frac{1}{2n}\right)S + \frac{1}{2n(p+2)}S^2,$$  \hspace{1cm}(2.89)

which still has a $\chi^2_p$ distribution in the limit. However, the obvious problem with Rayleigh’s test is that an alternative distribution with zero mean resultant length that is not uniform leads to low power.

Gine’s $F_n$ test of uniformity in the $p = 3$ dimension is consistent against all alternatives. The $F_n$ test of uniformity rejects uniformity for large values of the following quantity:

$$F_n = \frac{3n}{2} - \frac{4}{n\pi} \sum_{i<j} (\Psi_{ij} + \sin \Psi_{ij}),$$  \hspace{1cm}(2.90)

where $\Psi_{ij} = \cos^{-1}(x_i^T x_j)$. The 10%, 5%, and 1% quantiles for the limiting distribution are 2.355, 2.748, and 3.633 respectively, as listed in Mardia and Jupp (2000).

The $A_n$ statistic for testing uniformity on a circle developed by Ajne (in Mardia and Jupp (2000)) generalizes to the $(p - 1)$ dimensional sphere. The test statistic is

$$A_n = \frac{n}{4} - \frac{1}{n\pi} \sum_{i<j} \Psi_{ij},$$  \hspace{1cm}(2.91)
where $\Psi_{ij}$ is defined above. The 10%, 5%, and 1% quantiles for the limiting distribution are 1.816, 2.207, and 3.090 respectively for the $p = 3$ dimension as listed in Mardia and Jupp (2000).

Pycke (2007) later develops the U-statistic to test uniformity on the sphere. Pycke’s U-statistic is based on geometric means of distances between points on the surface of a sphere. This test is consistent against all alternatives. Pycke focuses on the $p = 3$ dimensional unit sphere. A unit vector $\mathbf{x}$ is characterized by Cartesian coordinates given by $x = \sin \theta$, $y = \sin \theta \cos \phi$, and $z = \cos \theta$, where $\theta$ and $\phi$ correspond to spherical coordinates of colatitude and longitude respectively. Pycke tests the null hypothesis that $\xi_1, \ldots, \xi_n$ are $n$ independent observations from the uniform distribution where $\xi_i$ is a point specified by spherical coordinates:

$$H_0 : f(\xi) = f_0(\xi) \text{ vs. } H_a : f(\xi) \neq f_0(\xi).$$

Pycke uses a similar kernel to the $U_n$ statistic for testing uniformity of a circle by Watson (1961). The kernel is defined

$$\Gamma(\xi_1, \xi_2) = -\frac{1}{4\pi} \log \frac{e}{2} (1 - \mathbf{x}_1 \cdot \mathbf{x}_2). \quad (2.92)$$

Pycke proves that the U-statistic

$$U_{\Gamma,n}(\xi_1, \ldots, \xi_n) = \frac{2}{n-1} \sum_{1 \leq i < j \leq n} \Gamma(\xi_i, \xi_j) \quad (2.93)$$

converges as $n \to \infty$ to a random variable with the distribution

$$\sum_{\ell=1}^{\infty} \frac{C_{\ell}(2\ell + 1)}{\ell(\ell + 1)} \quad (2.94)$$

where $C_{\ell}(k) = \chi^2_{\ell} - \ell$.

Manzotti et al. (2002) develop a test for the symmetry of an elliptical distribution based on points projected on the unit sphere. They transform the vector $\mathbf{x}$ in the same way Li, Fang, and Zhu in equation 2.85 to get spherically symmetric data vectors $\mathbf{y}$. Next, they project these values onto the unit sphere as $w_i = y_i / \|y_i\|$. If the original vector $\mathbf{x}$ is elliptically symmetric, then the $w_i$’s are approximately uniformly distributed on the unit sphere. They test the uniformity on the unit sphere with averaging the spherical harmonics of the transformed data.

A degree $j$ spherical harmonic is a restriction to the unit sphere of a degree $j$ homogeneous polynomial $p(\mathbf{x})$ on the space $\mathbb{R}^p$ such that $\sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2} (p) \equiv 0$. Denote $\mathcal{S}_j$ as the set of spherical harmonics of degree $j$ and the union from $j$ to $l$ of these sets as $\mathcal{S}_{j,l}$. They fix a positive value $\epsilon$ and let $p_n$ denote the $\epsilon$-quantile for the variables $\|y_1\|, \ldots, \|y_n\|$. Define the function

$$Q(h) = \frac{1}{n} \sum_{i \leq n} h(w_i) I_{\|y_i\| > p_n}.$$
where $h(\cdot)$ is a function defined on the unit sphere. They create the following statistic to test the uniformity on the unit sphere:

$$Z^2 = n \sum_{h \in \mathcal{H}_j} Q^2(h).$$  \hfill (2.95)

The statistic $Z^2$ has a $\chi^2$ limiting distribution of

$$Z^2 \sim (1 - \epsilon) \chi^2_j,$$

regardless of the underlying distribution. The details of this calculation are left for Manzotti et al. (2002). They mention that the test does not test against all possible alternatives to elliptical symmetry because they do not consider spherical harmonics of degree smaller than three. However, they assert that the test detects most nonsymmetrical elliptical distributions.
As shown in Section 1.2, a few anomalous observations in the data affect the accuracy of estimation of the DCC-GARCH model of Engle (1982). The autoregressive (AR) structure, maximum likelihood estimation method, and conditional correlation structure of the DCC-GARCH model makes the parameter estimation inherently susceptible to the effects of outliers in the data. This is due to observations having a limitless effect on the likelihood function of the DCC-GARCH model. A single observation outside the pattern of the data greatly affects the estimates of both the univariate GARCH parameters and the parameters in the correlation structure. This chapter uses simulations to demonstrate the effects of outliers in the current DCC-GARCH and proposes a new robust estimation method of the DCC-GARCH.

### 3.1 Outlier in General DCC-GARCH

To show the effects of outliers in the DCC-GARCH model, five different types of data generating processes are analyzed using the DCC-GARCH model. An innovation outlier approach is considered instead of an additive outlier approach. Good initial tests of the robust estimation of the DCC-GARCH model would include a model without outliers, models with different levels of random outliers, and distributions that simply have heavy-tails. Return data are generated using an underlying multivariate normal distribution as defined by the DCC-GARCH in equation 1.11, a contaminated normal distribution with 5% contamination, a contaminated normal distribution with 25% contamination, a multivariate standardized student $t$-distribution with 6 degrees of freedom, and a multivariate standardized student $t$-distribution with 3 degrees of freedom. The contaminated normal distributions
are defined with the following mixture of multivariate distributions

$$\varepsilon_t \equiv (1 - \eta)\varepsilon_{1,t} + \eta\varepsilon_{2,t}, \quad \varepsilon_{1,t} \sim \mathcal{N}(0, 1), \quad \varepsilon_{2,t} \sim \mathcal{N}(0, \sigma_{\varepsilon_{2,t}}^2),$$

(3.1)

where $\eta$ is a Bernoulli random variable with a probability equal to the percentage of contamination and $\sigma_{\varepsilon_{2,t}}^2 >> 1$. The simulation studies use $\sigma_{\varepsilon_{2,t}}^2 = 9$ to make the standard deviation three times larger. These contaminated normal distributions are standardized to have a variance equal to 1. These situations will test the effectiveness of the robust estimation method when no outliers are present, random outliers appear in normal data, and there exists an underlying distribution with heavier tails than the normal distribution. These $t$-distribution situations are relevant for testing the robustness of the proposed method when outliers are so prevalent in the data that the data no longer belongs to a normal distribution.

A data set with three assets has volatility parameters defined by a GARCH(1,1) as

$$a_0 = \begin{pmatrix} 0.12 \\ 0.60 \\ 0.32 \end{pmatrix}, \quad a_1 = \begin{pmatrix} 0.23 \\ 0.10 \\ 0.15 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 0.59 \\ 0.82 \\ 0.74 \end{pmatrix},$$

(3.2)

and correlation parameters as

$$\alpha = 0.24, \quad \beta = 0.70,$$

(3.3)

is generated from each of the above mentioned five distributions for over 500 time periods to create five sets of generated daily returns.

Two initial unconditional correlation matrices, $R_0$, are used for each distribution. The first initial correlation matrix is

$$\begin{pmatrix} 1 & 0.85 & 0.85 \\ 0.85 & 1 & 0.85 \\ 0.85 & 0.85 & 1 \end{pmatrix},$$

(3.4)

which describes a set of assets where all assets are closely related to each other. This set of assets is selected because possible outliers in each asset may impact all of the assets in the set. The second initial correlation matrix represents a more diversified set of assets with one of the three assets less correlated to the other two as follows:

$$\begin{pmatrix} 1 & 0.85 & 0.10 \\ 0.85 & 1 & 0.10 \\ 0.10 & 0.10 & 1 \end{pmatrix}. $$

(3.5)

This set is considered to determine if diversifying a set of assets leads to better volatility and correlation.
A Monte Carlo simulation study with 250 repetitions and 500 time points is computed for each distribution. The first initial unconditional correlation matrix in equation 3.4 is estimated with sample correlations for each distribution. Table 3.1 shows the estimates for the unconditional correlations in the first set of assets with standard errors in parentheses. For the volatility parameters $a_0$, $a_1$, and $b_1$, the average Monte Carlo mean squared errors (MSE) of the estimates across the three assets is calculated with the average standard error of the MSE values in parentheses. For the correlation parameters $\alpha$ and $\beta$, the estimates’ MSE values are shown with the standard error of the estimated MSE values in parentheses. The results from the correlation matrix defined in equation 3.4 are shown in Table 3.2. Appendix A contains the full detailed results for each asset and its respective parameters.

The estimates of the unconditional correlations barely change between the underlying distributions in the DCC-GARCH simulations. All of the estimates are within two standard deviations of each other. Heavy tailed distributions or outlier contaminated distributions do not appear to unduly affect the sample correlation matrix in the DCC-GARCH returns when all assets are highly correlated.

Using an underlying multivariate standardized student $t$-distribution with 6 degrees of freedom instead of an underlying multivariate normal distribution begins to affect the accuracy of the maximum likelihood estimation in the DCC-GARCH. First, the volatility parameters are examined. The estimate of the intercept parameter $a_0$ in an underlying $t_6$-distribution is comparable with the underlying normal distribution with its average MSE value slightly under two standard errors apart. However, both the estimates of the $a_1$ and $b_1$ parameters with a $t_6$-distribution have an average MSE over twice that of their respective counterparts in the normal distribution. The correlation parameter $\alpha$ has an estimate with an MSE value almost twice as large in the $t_6$-distribution compared to the normal distribution. The correlation estimate of $\beta$ is over ten times larger than the Normal distribution’s MSE for $\beta$.

The results of the multivariate standardized student $t$-distribution with 3 degrees of freedom compared to the Normal distribution are predictably more extreme than the previous $t_6$-distribution results. The estimate of the volatility intercept parameter $a_0$ has an average MSE six times larger than the respective MSE value in the $t_6$-distribution and eight times larger than its respective MSE value.
in the Normal distribution. This is similar to the estimates of the volatility parameters of $a_1$ and $b_1$. Their MSE values are three times larger than the MSE values of the $t_6$-distribution and five times larger than MSE values of the Normal distribution. The estimate of the correlation parameter $\alpha$ has a MSE value 5 times larger than the $t_6$-distribution's respective value and ten times larger than the Normal distribution's respective value. The MSE of the estimated value of $\beta$, which has a larger value for the $t_6$-distribution already, is now over 30 times larger than the respective MSE of the Normal distribution. Both the $t_6$- and the $t_3$-distributions affect the estimation of the parameters in the DCC-GARCH model.

The results with the contaminated Normal distributions are similar to the $t$-distributions in that both the volatility and correlation parameter estimations in the DCC-GARCH is inaccurate. The only difference is the magnitude of the inaccuracy. In the 5% contaminated Normal (CN) distribution, the average MSE values of the volatility parameter estimates are larger than their respective estimates in the Normal distribution. The estimate of $a_0$ actually has an average MSE that is approximately three times larger in the 5% CN distribution compared to the Normal distribution estimate. Both the $a_1$ and $b_1$ parameter estimates average MSE is twice as large as the respective Normal distribution parameter estimate's MSE. The correlation parameters are more extensively affected by the contaminated Normal distribution. The MSE of the estimates of $\alpha$ and $\beta$ are three and fourteen times greater than the Normal distribution's estimates’ MSE values respectively.

The results of the 25% contaminated Normal distribution are even more extreme in the effects of the estimation of the DCC-GARCH parameters. The volatility intercept estimate of $a_0$ has an average MSE value only twice as large as the MSE value of the estimates from the Normal distribution for the same parameter. The average MSE values for the estimates of $a_1$ and $b_1$ are higher than their respective estimates in the 5% CN distribution. Similarly to the 5% CN distribution, the 25% CN distribution has correlation parameter estimates’ MSE values greater than the respective Normal distribution MSE values with the only difference being the magnitude. The MSE values of the estimates of $\alpha$ and $\beta$ are approximately five and seventy times larger than their respective estimates from the Normal distribution.

A Monte Carlo simulation study with 250 repetitions and 500 time points is computed for each distribution for the second initial correlation matrix. The results from the correlation matrix mentioned in equation 3.5 are shown in Table 3.4. The second initial unconditional correlation matrix in equation 3.5 is estimated with sample correlations for each distribution. Table 3.3 shows the estimates for the unconditional correlations in the second set of assets with standard errors in parentheses. Again, the estimates of the unconditional correlations barely change between the underlying distributions in the DCC-GARCH simulations. Heavy tailed distributions or outlier contaminated distributions do not appear to have too much of an effect on the calculation of the sample correlation matrix in the DCC-GARCH returns when the set of assets is more diversified in correlation structure.

Even in a more diversified set of assets, the results are similar for each of the underlying distribu-
Table 3.2: Results with First Initial Correlation Matrix

Average MSE of Parameter Estimates

<table>
<thead>
<tr>
<th>Para.</th>
<th>Normal</th>
<th>$t_6$</th>
<th>$t_3$</th>
<th>CN 5%</th>
<th>CN 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0.0471</td>
<td>0.0629 (0.0092)</td>
<td>0.3919 (0.0730)</td>
<td>0.1246 (0.0157)</td>
<td>0.0871 (0.0094)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0069</td>
<td>0.0147 (0.0017)</td>
<td>0.0466 (0.0062)</td>
<td>0.0159 (0.0015)</td>
<td>0.0248 (0.0026)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0088</td>
<td>0.0167 (0.0018)</td>
<td>0.0486 (0.0056)</td>
<td>0.0199 (0.0018)</td>
<td>0.0218 (0.0021)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0026</td>
<td>0.0043 (0.0003)</td>
<td>0.0225 (0.0009)</td>
<td>0.0084 (0.0004)</td>
<td>0.0101 (0.0003)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0012</td>
<td>0.0124 (0.0011)</td>
<td>0.0318 (0.0016)</td>
<td>0.0169 (0.0006)</td>
<td>0.0201 (0.0005)</td>
</tr>
</tbody>
</table>

Table 3.3: $R_0$ Results with Second Initial Correlation Matrix

Parameter Estimates of Unconditional Correlations

<table>
<thead>
<tr>
<th>Para.</th>
<th>Normal</th>
<th>$t_6$</th>
<th>$t_3$</th>
<th>CN 5%</th>
<th>CN 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{1,2}$</td>
<td>0.7952</td>
<td>0.7834 (0.0062)</td>
<td>0.7868 (0.0089)</td>
<td>0.7838 (0.0076)</td>
<td>0.7867 (0.0066)</td>
</tr>
<tr>
<td>$\rho_{1,3}$</td>
<td>0.0920</td>
<td>0.0631 (0.0129)</td>
<td>0.0737 (0.0172)</td>
<td>0.0692 (0.0132)</td>
<td>0.1112 (0.0140)</td>
</tr>
<tr>
<td>$\rho_{2,3}$</td>
<td>0.0860</td>
<td>0.0665 (0.0120)</td>
<td>0.0538 (0.0167)</td>
<td>0.0861 (0.0129)</td>
<td>0.1117 (0.0141)</td>
</tr>
</tbody>
</table>

Much of the same results occurred for this second initial correlation matrix (again Appendix A contains the full detailed estimates). A majority of the average MSE values for the volatility parameter estimates in the second, more diversified set are within one standard error of their respective average MSE values in the first initial correlation structure situation. For the correlation parameters in the second set of assets, all of the $\beta$ estimates’ MSE values are within one standard error of their respective first set’s counterparts. The respective $\alpha$ correlation estimates between the two sets are not equal, but the effects of the heavy tailed $t$-distributions and contaminated Normal distributions still remain in both sets. In both sets of assets, anomalous observations from either heavy tailed distributions or mixed distribution contamination greatly affect the estimation capability of the DCC-GARCH model. The next section introduces a new robust estimation method for the DCC-GARCH to account for these estimation issues.

### 3.2 Robust Estimation of DCC-GARCH

#### 3.2.1 Proposed Robust Method

The proposed robust method for the DCC-GARCH takes the bounded deviance function approach mentioned in Section 2.2.2. This approach will limit the observational effects of unexpectedly large magnitude in the data, but not affect observations that are not unexpectedly large in magnitude. The
DCC-GARCH estimation involves the split estimation of two components of a linear equation, as mentioned in Section 1.1. Since the volatility and correlation components are a linear combination and the $\lambda$-function adjustment is nonlinear, the $\lambda$-function is not applied to each individual component. Instead it is applied to the sum of both components. However, it is shown earlier in Figure 1.1 that the estimation of both components simultaneously leads to unstable likelihood functions.

For this reason, a multiple iteration approach is needed for the estimation of the proposed robust method for the DCC-GARCH. Before the iteration process, estimate the unconditional correlation matrix for an initial value of $R_0$. As previously seen, the sample correlation matrix performs similarly well in any of the tested distributions; so the sample correlation matrix is used.

The first iteration involves estimation of only the volatility parameters by assuming the parameters in the correlation component are constant, using the MLE values

$$
\sum_{t=1}^{T} \lambda \left\{ d(\theta, \phi) \right\} = \sum_{t=1}^{T} \lambda \left\{ -2 \left[ L_V(\theta) + L_C(\hat{\theta}_{MLE}, \hat{\phi}_{MLE}) - L_V + L_C(\hat{\theta}_{MLE}, \hat{\phi}_{MLE}) \right] \right\}.
$$

From this modified deviance function, the robust estimates of the volatility parameters are calculated by

$$
\hat{\theta}_R = \arg \min_{\theta} \sum_{t=1}^{T} \lambda(d(\theta, \phi_{MLE})).
$$

(3.6)

The second iteration involves estimation of only the correlation parameters by assuming the volatility parameters are constant, using the previously calculated robust values:

$$
\sum_{t=1}^{T} \lambda \left\{ d(\theta, \phi) \right\} = \sum_{t=1}^{T} \lambda \left\{ -2 \left[ L_V(\hat{\theta}_R) + L_C(\hat{\theta}_R, \phi) - L_V + L_C(\hat{\theta}_{MLE}, \hat{\phi}_{MLE}) \right] \right\}
$$

### Table 3.4: Results with Second Initial Correlation Matrix

<table>
<thead>
<tr>
<th>Para.</th>
<th>Normal</th>
<th>$t_6$</th>
<th>$t_3$</th>
<th>CN 5%</th>
<th>CN 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0.0581 (0.0239)</td>
<td>0.0792 (0.0175)</td>
<td>0.2453 (0.0599)</td>
<td>0.1666 (0.0150)</td>
<td>0.0895 (0.0106)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0073 (0.0013)</td>
<td>0.0147 (0.0021)</td>
<td>0.0474 (0.0068)</td>
<td>0.0184 (0.0014)</td>
<td>0.0237 (0.0025)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0092 (0.0015)</td>
<td>0.0154 (0.0021)</td>
<td>0.0469 (0.0057)</td>
<td>0.0218 (0.0015)</td>
<td>0.0225 (0.0023)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0041 (0.0002)</td>
<td>0.0035 (0.0004)</td>
<td>0.0213 (0.0008)</td>
<td>0.0067 (0.0003)</td>
<td>0.0081 (0.0002)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0013 (0.0001)</td>
<td>0.0101 (0.0006)</td>
<td>0.0341 (0.0011)</td>
<td>0.0156 (0.0004)</td>
<td>0.0195 (0.0004)</td>
</tr>
</tbody>
</table>
The robust correlation parameters are estimated by

\[ \hat{\phi}_R = \arg \min_{\phi} \sum_{t=1}^T \lambda(d(\hat{\theta}_R, \phi)). \]  

(3.7)

The iteration process does not need to continue past this point because the estimation of the volatility component does not depend on the correlation parameters.

As mentioned in Section 2.2.2, the modified deviance approach is defined for many different \( \lambda \) functions. The \( \lambda \) function selected is the bounded function described by Bianco and Yohai (1996) as:

\[ \lambda_{BY}(t) = \begin{cases} 
\frac{t^2}{2c^2}, & t \leq c \\
\frac{c}{2}, & t > c 
\end{cases} \]

The Bianco and Yohai function is chosen because it has a strict flat bound after the bounding threshold \( c \) is reached by the deviances. Pregibon’s \( \lambda \) function did not have an upper bound on the deviances.

### 3.2.2 Consistency

It is beneficial to study the asymptotic behaviors of the robust estimators defined in equations 3.6 and 3.7 with regards to Fisher-consistency. Bianco and Yohai (1996) adapt the deviance function estimation previously mentioned by Pregibon in equation 2.26 to the minimization of

\[ \sum_{i=1}^n \left\{ \lambda \left[ d(g(x^T \beta); y_i) \right] + C(x^T \beta) \right\}, \]  

(3.8)

where \( C(x^T \beta) \) is a bias correction term to make the estimator conditionally Fisher-consistent. The bias correction function is calculated from the score function of the estimation equation. If the expected value of the score function equals zero under the true value of the parameters, the estimators are conditionally Fisher-consistent. This makes the bias correction term equal to zero.

Due to the complexity of the multivariate deviance estimating equation, the score functions for the robust estimation equations 3.6 and 3.7 are difficult to calculate. Estimating the score functions through simulations eliminates the difficulty of obtaining the true score functions. Finite differences estimate the derivative at a point in a function. After calculating an estimator \( \hat{\beta} \), the deviance equation is evaluated at \( \hat{\beta} + h \) and \( \hat{\beta} - h \). The value of the slope of the chord with \( h \) sufficiently small is an approximation of the derivative of \( \hat{\beta} \). Resampling similar data sets in a bootstrap method for different samples of the same distribution and taking an average value of these slopes approximately estimates the expected score function of the deviance estimation.

Using this estimation technique, the bias correction term may be approximated by a linear function of the parameters with the coefficients equalling the estimated score functions for that
This makes the robust parameter estimates for the volatility parameters the solution to

$$
\hat{\theta}_R = \arg \min_{\theta} \left\{ \sum_{t=1}^{T} \lambda(d(\theta, \hat{\phi}_{MLE})) - \hat{\gamma}^T \theta \right\},
$$

(3.9)

where $\hat{\gamma}_\theta$ is a vector of estimated coefficients from the score estimation process. The robust parameter estimates for the correlation parameters are

$$
\hat{\phi}_R = \min_{\phi} \left\{ \sum_{t=1}^{T} \lambda(d(\hat{\theta}_R, \phi)) - \hat{\gamma}^T \phi \right\},
$$

(3.10)

where $\hat{\gamma}_\phi$ is a vector of estimated coefficients from the score estimation process.

### 3.3 Initial Tests of Robust Method

The previous section introduces the robust estimation for the DCC-GARCH model. As seen in the beginning of this chapter, a multivariate $t$-distribution and contaminated Normal distribution greatly affect the maximum likelihood estimates of the DCC-GARCH. This demonstrates how outliers cause bias and increases in the sampling variance of the values of the estimates. For an initial test, the robust method is tested with the same return data generated to initially test the DCC-GARCH in Section 3.1. The five sets of returns used are generated by a Normal distribution, $t_6$-distribution, $t_3$-distribution, 5% contaminated Normal distribution, and a 25% contaminated Normal distribution. The mean squared error (MSE) of each of the estimates of the parameters is estimated with a Monte Carlo simulation similar to the process of testing the maximum likelihood estimators mentioned in Section 3.1. The value of the constraint for the robust estimation method for the DCC-GARCH in the $\lambda$ function is $c = 80$ to make the highest value of the bounded deviances equal to 40. Testing revealed that the deviances under the Normal distribution have approximately 95% of their values below 40.

A good robust model follows two general principles:

1. Keeps estimates from dramatically changing in the presence of outliers.
2. Does not dramatically change estimates without the presence of outliers.

It is important to remember the second principle, because this principle allows the use of a robust model in analysis when it is unknown whether outliers are present and they might not be easily detected.

The MSE of the parameter estimates is recorded for both the robust estimation method without the bias correction in equations 3.6 and 3.7 and the robust estimation with the bias correction in equations 3.9 and 3.10. Both robust estimation results are recorded to compare the MSE values between the two estimation methods. Although the bias correction term might improve the bias of
Table 3.5: Comparing Two Robust Methods with First Correlation Structure

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>$t_6$</th>
<th>$t_3$</th>
<th>CN 5%</th>
<th>CN 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0.2213</td>
<td>0.5220</td>
<td>0.7001</td>
<td>0.5125</td>
<td>1.1778</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.2459</td>
<td>0.5630</td>
<td>0.7231</td>
<td>0.6186</td>
<td>1.1022</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.2188</td>
<td>0.1948</td>
<td>0.2254</td>
<td>0.1641</td>
<td>0.2683</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.6486</td>
<td>3.8681</td>
<td>5.1612</td>
<td>3.0057</td>
<td>6.8042</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0260</td>
<td>0.1588</td>
<td>0.2540</td>
<td>0.1488</td>
<td>0.3659</td>
</tr>
</tbody>
</table>

the parameter estimates, the estimation process may introduce unneeded variation and lead to a worse MSE in comparison. The average MSE Ratio between the two robust estimates for each of the parameters is shown in Tables 3.5 and 3.6 for both of the initial correlation structures mentioned in equations 3.4 and 3.5 respectively.

The first set’s initial correlation structure is not diversified in the sense that the three assets are all highly correlated with each other. The bias correction negatively affects the estimation more than it helps the estimation for all of the volatility parameters across all distributions tested except for the 25% contaminated Normal distribution. Upon closer inspection though, only four of the nine volatility parameter estimates have a lower MSE for the bias correction estimation in the 25% contaminated Normal distribution. Three of these four estimates are the intercept parameter estimates. Although the parameter estimates might be less biased with the correction term, the variability introduced to the estimation out-weighs the benefit of the correction of bias. The $\alpha$ correlation parameter estimate has a lower MSE with the correction bias compared to without bias correction for every distribution except the Normal distribution. However, as fully detailed in the results below, the MLE technique outperforms both robust estimation techniques with regards to the MSE of the $\alpha$ estimate in half of the distributions. The other correlation parameter $\beta$ is similar to the volatility parameters in that the bias correction hinders the estimation more than it helps for any distribution.

The second set’s initial correlation structure is more diversified in the sense that only two of the assets in the set are highly correlated, while the third asset is not strongly correlated with either of the previous two. The bias correction affects the estimation more than it helps the estimation for all of the volatility parameters across all tested distributions. None of the average MSE values in the bias-corrected robust estimation method for the DCC-GARCH are larger than in the robust method without the bias correction. Unlike the previous analysis with the first initial correlation structure, the correlation parameter $\alpha$ estimation has higher MSE with the correction bias compared to without bias correction for every distribution. The other correlation parameter $\beta$ is similar to the volatility parameters in that the bias correction negatively affects the estimation more than it helps for any distribution.
Table 3.6: Comparing Two Robust Methods with Second Correlation Structure

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>$t_6$</th>
<th>$t_3$</th>
<th>CN 5%</th>
<th>CN 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0.1489</td>
<td>0.2222</td>
<td>0.4592</td>
<td>0.2481</td>
<td>0.3394</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.4504</td>
<td>0.3174</td>
<td>0.5699</td>
<td>0.4006</td>
<td>0.5217</td>
</tr>
<tr>
<td>$b_2$</td>
<td>0.1883</td>
<td>0.1924</td>
<td>0.2804</td>
<td>0.2118</td>
<td>0.4923</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.6929</td>
<td>0.5262</td>
<td>0.2686</td>
<td>0.8127</td>
<td>0.1439</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.3593</td>
<td>0.0119</td>
<td>0.0104</td>
<td>0.0299</td>
<td>0.0044</td>
</tr>
</tbody>
</table>

distribution. This analysis implies that a more diversified set of assets is better estimated without the bias correction term for nearly every situation and parameter. Perhaps the diversified structure of the returns makes the robust version of the DCC-GARCH less biased in its estimation of both the volatility and correlation parameters in the model.

Table 3.7 summarizes the results of analyzing each distribution with the robust estimation for the DCC-GARCH without the bias correction term. The average parameter estimate MSE over all of the assets is displayed for the set of assets with the first correlation structure. Complete details of all the estimates for each individual asset in the set is located in the Appendix in Tables A.1, A.2, and A.3.

When the model has an underlying Normal distribution, the robust method is outperformed by the maximum likelihood estimation of the traditional DCC-GARCH model. Most of the volatility parameters are better estimated with the DCC-GARCH maximum likelihood method. The volatility intercept parameter $a_0$ in the robust method has an estimate with an MSE value double that of the DCC-GARCH model in each of the three assets. The MSE of the estimate of the volatility parameter $a_1$ has the reverse results when averaged across the three assets. The robust estimation of the DCC-GARCH provides an MSE that is under half the value of the MLE’s MSE. While the robust method of the DCC-GARCH $a_1$ estimate MSE is lower for all three assets, the first asset’s estimate has a robust MSE that is under 30% of the traditional model estimate’s MSE, where the other two assets’ estimates have a robust MSE that is approximately 75% of the MLE’s MSE. The volatility parameter $b_1$ is similar to $a_0$. On average and individually, the robust estimates’ MSE is over twice the value of the maximum likelihood value estimates’ MSE. The comparison of the correlation parameter estimation is not even close. For both the estimates of $\alpha$ and $\beta$, the robust method MSE is almost four times larger than the MSE from the traditional DCC-GARCH method. When no outliers are present in the model, maximum likelihood is the best estimation method, as expected.

When the underlying distribution of the returns follows a heavier tailed distribution than the Normal distribution, the results for the robust method improve in comparison to the traditional DCC-GARCH. When the underlying model follows a $t_6$-distribution, the estimation of the volatility
parameters improves on average. The average robust model has MSE values for estimates of both the volatility parameters \(a_0\) and \(a_1\) that are approximately equal to the maximum likelihood estimation's MSE values. The estimates are within one to two standard errors of each other. However, upon closer inspection of the individual assets shown in Table A.1, only one estimate from the \(a_0\) parameter and one \(a_1\) parameter estimate have robust model MSE values lower than the traditional MSE values. The other two assets' parameters are not as close in their estimates' MSE comparison as the average MSE value makes it appear. The last volatility parameter \(b_1\) is not as closely estimated with the robust method as the other volatility parameters. The average robust model MSE for the estimates of \(b_1\) is over 1.67 times larger than the MSE from the DCC-GARCH. Unfortunately, the MSE value for the correlation parameter estimates continues the same trend with the underlying \(t_6\)-distribution as the Normal distribution. Neither parameter is estimated as well with the robust estimation method for the DCC-GARCH compared to the traditional DCC-GARCH.

As the tails of the underlying distribution continue to get heavier, the robust method begins to greatly out-perform the traditional DCC-GARCH. When the underlying distribution is a \(t_5\)-distribution, the trend continues to improve for the robust estimation method. On average, each of the volatility parameter estimates from the robust model have a lower MSE than their respective maximum likelihood estimates' MSE values. The intercept parameter \(a_0\) has the most striking improvement with its estimates' average MSE value using the robust method at one sixth the average MSE value using the DCC-GARCH model. The \(a_1\) and \(b_1\) estimates from the robust model have average MSE values at 55% and 80% of their respective MLE's MSE values. However, upon closer inspection of the \(b_1\) parameter estimates for each asset in Table A.1, only two of the three assets' robust estimate's MSE values out-perform the MSE values using maximum likelihood estimation. Although the estimates of the correlation parameters have better MSE ratios than the Normal distribution or \(t_6\)-distribution, the maximum likelihood estimation still out-performs the robust estimation. This distribution is the only distribution in which the bias correction robust estimation MSE of the \(\alpha\) parameter out-performs the MLE value of the MSE. As the tails of the underlying distribution get heavier, the robust estimation method for the DCC-GARCH begins to out-perform the traditional DCC-GARCH in the estimation of the volatility parameters in the model. However, the correlation parameters' maximum likelihood estimates are still the most accurate in comparison to the robust estimation method.

Similar to heavy tailed distributions, as distributions become contaminated with outliers in a mixed distribution sense, the robust method performs better than the DCC-GARCH. With a 5% contaminated Normal distribution with a contaminated standard deviation three times larger than the uncontaminated portion, the volatility parameter estimation with the robust method is a slight improvement over the traditional DCC-GARCH. On average and individually, the assets' estimate of the intercept volatility parameter \(a_0\) and parameter \(a_1\) have a robust model MSE value lower than the maximum likelihood estimation's MSE value. However, one of the three assets' estimates of \(a_0\) has a MSE value within one standard error of each other and can be seen as comparable. None of the assets'
estimates have a better robust model MSE value for the $b_1$ parameter. The trend of poor correlation parameter estimation with the robust method compared to the DCC-GARCH continues for the 5% contaminated normal distribution with robust model MSE values approximately four times larger for both the estimates of $\alpha$ and $\beta$ parameters.

As the contamination increased to 25%, the effectiveness of the robust estimation method for the DCC-GARCH decreases. The only significant benefit of the robust method in the 25% contaminated Normal distribution is the estimation of the volatility parameter $a_0$. The average MSE from the robust model's estimates are smaller than the MSE value under maximum likelihood estimation. The estimates of five of the six remaining volatility parameters have a significantly higher average MSE value with the robust method compared to the DCC-GARCH. A possible explanation of this is the volatility parameters $a_1$ and $b_1$ are bounded by the constraint that $a_1 + b_1 < 1$. This extreme outlier model might push the estimation of the parameters to the edge of the parameter space. Even though the estimates are more biased, the variability in these estimates is small enough to make the MSE value low. Similar to all the previous distributions, the correlation parameter estimation of $\alpha$ and $\beta$ is more efficient with maximum likelihood estimation compared to the robust estimation.

The robust estimation method for the DCC-GARCH did not perform as well in the outlier contaminated distributions as compared to the heavy tailed distributions using the first set of assets. The robust estimation of the model does a reasonable job at correcting the estimation concerns with the volatility modeling in these predetermined distributional cases. The estimation of the correlation parameters never improves with the robust model compared to the traditional DCC-GARCH. It appears that the robust method does not accurately estimate the correlation structure when the assets in the set are highly correlated. The next part of the simulation study will determine if a more diversified set of assets helps in the robust method parameter estimation.
Table 3.7: Results Across Generating Processes with First Set

**Average MSE for Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
<td>Robust</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0471 (0.0061)</td>
<td>0.0869 (0.0094)</td>
<td>0.0629 (0.0092)</td>
<td>0.0750 (0.0065)</td>
<td>0.3919 (0.0730)</td>
<td>0.0664 (0.0032)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0069 (0.0008)</td>
<td>0.0030 (0.0002)</td>
<td>0.0147 (0.0017)</td>
<td>0.0144 (0.0008)</td>
<td>0.0466 (0.0062)</td>
<td>0.0262 (0.0020)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0088 (0.0009)</td>
<td>0.0199 (0.0020)</td>
<td>0.0167 (0.0018)</td>
<td>0.0279 (0.0032)</td>
<td>0.0486 (0.0056)</td>
<td>0.0410 (0.0044)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0026 (0.0002)</td>
<td>0.0095 (0.0002)</td>
<td>0.0043 (0.0003)</td>
<td>0.0348 (0.0008)</td>
<td>0.0225 (0.0009)</td>
<td>0.0345 (0.0007)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0012 (0.0001)</td>
<td>0.0047 (0.0003)</td>
<td>0.0124 (0.0011)</td>
<td>0.0717 (0.0040)</td>
<td>0.0318 (0.0016)</td>
<td>0.1234 (0.0063)</td>
</tr>
</tbody>
</table>

**Average MSE for Parameter Estimates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CN 5%</th>
<th></th>
<th>CN 25%</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
<td>Robust</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.1246 (0.0133)</td>
<td>0.0703 (0.0045)</td>
<td>0.0871 (0.0084)</td>
<td>0.0836 (0.0027)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0159 (0.0012)</td>
<td>0.0141 (0.0004)</td>
<td>0.0248 (0.0023)</td>
<td>0.0440 (0.0031)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0199 (0.0015)</td>
<td>0.0260 (0.0024)</td>
<td>0.0218 (0.0019)</td>
<td>0.0430 (0.0038)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0084 (0.0003)</td>
<td>0.0363 (0.0006)</td>
<td>0.0101 (0.0002)</td>
<td>0.0418 (0.0007)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0169 (0.0005)</td>
<td>0.0729 (0.0036)</td>
<td>0.0201 (0.0005)</td>
<td>0.1771 (0.0069)</td>
</tr>
</tbody>
</table>
The second set of assets, which has the initial correlation matrix defined in equation 3.5, is a more diversified set. In this set, only two of the assets are highly correlated, with one asset only slightly correlated to the other two. Table 3.8 summarizes the results of analyzing each distribution with the robust estimation method for the DCC-GARCH without the bias correction term, which hampers the estimation more than it benefits it. The average MSE over all of the assets in each parameter is displayed for the set of assets with the second correlation structure defined in equation 3.5. Complete details of all the estimates for each individual asset in the set is located in the Appendix in Tables A.4, A.5, and A.6.

When the model has an underlying Normal distribution, the effectiveness of the robust estimation of the DCC-GARCH with the diversified set of assets is similar to the results with the previous set, which lacked diversification. The average MSE value with the robust method for the estimate of the volatility intercept parameter \( a_0 \) is comparable to the MSE value with the DCC-GARCH with the estimates being within one standard error of each other. However, the remaining volatility parameters are not as accurately estimated. The average robust method estimates have a higher MSE for both the remaining volatility parameters \( a_1 \) and \( b_1 \). This is true not just on average, but each asset in the set has this same quality with regards to the parameter estimation of \( a_1 \) and \( b_1 \). The same trend of a lack of effectiveness in the robust model's ability to estimate the correlation parameters continues here. Both of the robust method's MSE values for the estimates of the parameters \( \alpha \) and \( \beta \) are higher than their respective maximum likelihood estimate MSE values. Despite the change in the initial correlation structure, the estimation of the parameters with an underlying Normal distribution remains relatively weak with the robust method compared to the DCC-GARCH.

Similarly to the previous initial correlation structure, as the tails in the underlying distribution get heavier in nature, the effectiveness of the robust method increases. With an underlying \( t_6 \)-distribution, the effectiveness in estimating the parameters increases with the robust method. As seen in closer detail in Table A.4 in the Appendix, estimates of two of the three assets' volatility intercept parameters have lower MSE values when estimated with the robust estimation of the DCC-GARCH. The estimate of the volatility parameter \( a_1 \) has a lower average robust method MSE value compared to the traditional maximum likelihood MSE value. This also holds true for two of the three assets in the set. However, the robust method is not more effective in estimating the final volatility parameter \( b_1 \) on average. When it comes to the correlation parameters, the trend breaks with the underlying \( t_6 \)-distribution with the more diversified set of assets. Although the correlation parameter \( \alpha \) is still better estimated with the MLE values from the DCC-GARCH model, the \( \beta \) parameter is better estimated from the robust method when comparing MSE values of the estimates. The diversification in the set of assets appears to allow better estimation of the correlation structure with the robust method compared to a set that lacks diversification.

As the tails in the underlying distribution get even heavier with the \( t_3 \)-distribution, the robust estimation method for the DCC-GARCH becomes more effective in estimating the model parameters.
for both the volatility and correlation portions. As seen in further detail in Table A.4, the robust method has a lower MSE for every volatility parameter estimate except for one $b_1$ parameter estimate. The most surprising element of these results is that both the correlation parameters $\alpha$ and $\beta$ are better estimated with the robust estimation method. The diversification of the set seems to allow the robust method to better estimate the correlation structure of the data compared to a set that is not diversified.

The same results appear in underlying distributions with the outlier contamination instead of heavy tails. When the underlying distribution is a 5% contaminated Normal distribution, the robust method still performs extremely well in the estimation of the volatility parameters. With regard to the estimate of the intercept parameter $a_0$, the robust method has a lower MSE value in two of the three assets; this still leads to an average robust method MSE value which is lower when compared to the maximum likelihood MSE value. On average and individually across assets, the pattern of effective estimation of the $a_1$ parameter continues with this underlying distribution. The estimates from the robust method have a lower MSE value for every asset’s estimation of the $a_1$ parameter. Two of the three assets have a lower robust method MSE value for the $b_1$ parameter estimate compared to the DCC-GARCH model estimate’s MSE value. The asset that does not have a lower $b_1$ estimate MSE has an MSE value comparable to the MSE value of the maximum likelihood estimation. Although the estimation of the correlation parameter $\alpha$ returns to being more effective in maximum likelihood estimation, the correlation parameter $\beta$ is still being better estimated with the robust version of the model. Again, the diversification of the set of assets seems to help in the robust model’s estimation of the parameters.

When the underlying distribution of the returns is a 25% contaminated Normal distribution, the robust estimation method for the DCC-GARCH begins to lose its effectiveness in estimating the volatility parameters. This trend continues to be true when the returns are not as diversified as they are with this set of simulations. However, unlike the previous set of simulations, the more diversified set does allow more volatility parameters to be better estimated with the robust method. The average robust method MSE value is lower for both the estimates of the parameters $a_0$ and $a_1$ compared to their respective maximum likelihood MSE values. The $b_1$ parameter however is better estimated with the DCC-GARCH compared to the robust method in every asset. However, both of the correlation parameters are better estimated with the robust method. Even though the effectiveness of the robust method seems to drop slightly in this scenario, it still performs better at estimating the parameters with a more diversified set.

The diversified set of assets is introduced in hopes of lessening the effects of the outliers in the maximum likelihood estimation. Although this did not hold true, the more diversified set did aid in the estimation of the returns with the robust method. In every situation, except for the underlying distribution not having heavy tails or outlier contamination, the robust estimation method effectiveness improved in comparison with the previous set of simulations which has a less diversified
set of assets. It seems that a set of assets should be diversified to not only better protect the holder of the set of assets from risk, but to also help in the estimation of the volatility of that set.

Although the robust method seems less affected by outliers compared to the maximum likelihood estimate in certain situations, real world data may not follow a multivariate $t$-distribution or contaminated Normal distribution. If the distribution of some real world data — such as a set of foreign exchange rates — follows a multivariate $t$-distribution with degrees of freedom in the same range as those used in these tests or a similar contaminated Normal distribution used in these tests, then the effectiveness of the robust estimation of the DCC-GARCH model could be estimated.
Table 3.8: Results Across Generating Processes with Second Set

Average MSE for Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>$t_6$</th>
<th>$t_3$</th>
<th>Normal</th>
<th>$t_6$</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
<td>Robust</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0581 (0.0239)</td>
<td>0.0501 (0.0015)</td>
<td>0.0792 (0.0175)</td>
<td>0.0549 (0.0046)</td>
<td>0.2453 (0.0599)</td>
<td>0.0650 (0.0044)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0073 (0.0013)</td>
<td>0.0083 (0.0003)</td>
<td>0.0147 (0.0021)</td>
<td>0.0043 (0.0004)</td>
<td>0.0474 (0.0068)</td>
<td>0.0096 (0.0003)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0092 (0.0015)</td>
<td>0.0204 (0.0009)</td>
<td>0.0154 (0.0021)</td>
<td>0.0198 (0.0020)</td>
<td>0.0469 (0.0057)</td>
<td>0.0378 (0.0030)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0041 (0.0002)</td>
<td>0.0172 (0.0002)</td>
<td>0.0035 (0.0004)</td>
<td>0.0089 (0.0003)</td>
<td>0.0213 (0.0008)</td>
<td>0.0051 (0.0002)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0013 (0.0001)</td>
<td>0.0643 (0.0003)</td>
<td>0.0101 (0.0006)</td>
<td>0.0044 (0.0005)</td>
<td>0.0341 (0.0011)</td>
<td>0.0046 (0.0003)</td>
</tr>
</tbody>
</table>

Average MSE for Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CN 5%</th>
<th>CN 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>Robust</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.1666 (0.0150)</td>
<td>0.0496 (0.0030)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0184 (0.0014)</td>
<td>0.0055 (0.0002)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0218 (0.0015)</td>
<td>0.0166 (0.0013)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0067 (0.0003)</td>
<td>0.0151 (0.0003)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0156 (0.0004)</td>
<td>0.0126 (0.0007)</td>
</tr>
</tbody>
</table>
In this chapter, we examine the underlying distribution in a set of daily returns for five United States foreign exchange rates. The daily foreign exchange rates studied are the Euro, British Pound, Swiss Franc, Japanese Yen, and Canadian Dollar, from the beginning of the Euro entering the exchange market on January 4, 1999 until October 16, 2009. After removing the mean and variance structure from the set of daily returns of foreign exchange rates, the residuals reveal the remaining aspects of the set’s distribution. By examining the distribution of the squared radii of the multivariate observations, a close approximation to the distribution of the residuals is estimated. Since a multivariate $t$-distribution or contaminated Normal distribution may not represent real world data, the information from the residual distribution allows for a better data generating process to be tested with the robust estimation method for the DCC-GARCH model. If the robust method outperforms the DCC-GARCH in this newly revealed distribution, then the robust method provides a better estimate of the foreign exchange rate data compared to the MLE of the DCC-GARCH.

### 4.1 Multivariate Characterization of Data

The underlying distribution of a set of returns of United States exchange rates comes from the conditional errors of the market data. To get to the conditional errors in the model, the mean, variance, and correlation structures must be removed. Vector autoregressive (VAR) models estimate the mean structure of multivariate data. The general VAR(p) model is

$$r_t = c + A_1 r_{t-1} + \ldots + A_p r_{t-p} + e_t$$

(4.1)
Table 4.1: Exchange Rate Volatility Parameter Estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Euro</th>
<th>Pound</th>
<th>Franc</th>
<th>Yen</th>
<th>Dollar</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0.29e-4</td>
<td>0.21e-4</td>
<td>0.28e-4</td>
<td>0.30e-4</td>
<td>0.91e-5</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.183</td>
<td>0.204</td>
<td>0.171</td>
<td>0.146</td>
<td>0.198</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.127</td>
<td>0.270</td>
<td>0.152</td>
<td>0.127</td>
<td>0.442</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.181</td>
<td>0.134</td>
<td>0.191</td>
<td>0.373</td>
<td>0.024</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.170</td>
<td>0.120</td>
<td>0.170</td>
<td>0.121</td>
<td>0.113</td>
</tr>
</tbody>
</table>

where each $r_t$ is a $k \times 1$ vector of daily returns of foreign exchange rates and each $A_i$ is a $k \times k$ matrix of parameters.

The VAR(1) model defined as

$$ r_t = c + A_1 r_{t-1} + e_t $$

is used in estimating the set of returns for the five foreign exchange rates of the United States Dollar with the Euro (EU), British Pound (BP), Swiss Franc (SF), Japanese Yen (JY), and Canadian Dollar (CD). The results for the VAR(1) model are

$$
\begin{bmatrix}
\hat{r}_{EU,t} \\
\hat{r}_{BP,t} \\
\hat{r}_{SF,t} \\
\hat{r}_{JY,t} \\
\hat{r}_{CD,t}
\end{bmatrix} =
\begin{bmatrix}
0.000 \\
0.097 \\
0.143 \\
0.127 \\
0.126
\end{bmatrix} +
\begin{bmatrix}
0.160 & -0.055 & -0.110 & -0.040 & 0.000 \\
0.000 & 0.000 & -0.095 & -0.033 & 0.053 \\
0.143 & 0.000 & -0.123 & 0.000 & 0.000 \\
0.127 & 0.000 & -0.091 & 0.000 & 0.000 \\
0.126 & 0.000 & -0.134 & -0.036 & 0.000
\end{bmatrix} \begin{bmatrix}
r_{EU,t-1} \\
r_{BP,t-1} \\
r_{SF,t-1} \\
r_{JY,t-1} \\
r_{CD,t-1}
\end{bmatrix}.
$$

From this estimate of the vector $\hat{r}_t$, the residuals of the data without the estimated mean structure are calculated by $r_t - \hat{r}_t$. The set of residuals without mean structure alone still has a correlation and variance structure. This correlation and variance structure influences the results of the tests of elliptical symmetry for the set of residuals and must be removed from the data.

A DCC-GARCH(p,q) model described in section 2.1.2 estimates the variance and correlation structure of the set of the five exchange rates. The GARCH(3,1) structure is selected as the best model for the data. The correlation parameters $\alpha$ and $\beta$ are estimated as

$$ \hat{\alpha} = 0.212, \quad \hat{\beta} = 0.776. $$

After estimating the parameters, the estimated covariance matrix $\hat{H}_t$ is calculated for each time point along with the Cholesky factorization of $\hat{H}_t$. The residuals from the above calculations are approximations to the conditional errors in the set of foreign exchange rates. Estimating the density
of these conditional residuals in the set of foreign exchange rates allows for the creation of a data generating process derived from real world data. This better tests the effectiveness of the robust method in comparison with the DCC-GARCH model. If the mean, variance, and correlation structure are completely removed from the foreign exchange rate data, the remaining residuals would resemble spherical symmetry, if originally the residuals are elliptically symmetric. Li et al. (1997) proposes a simple test of spherical symmetry with their Q-Q plot test mentioned in section 2.4. They recommend plotting their ordered test statistic mentioned in equation 2.84 with the quantiles from the $t_4$ distribution. The $t$-distribution has four degrees of freedom because the data are five dimensional.

Figure 4.1 displays the above-mentioned Q-Q plot. The plot is not a straight line and its S-shape suggests that the residuals are not spherically symmetric. This implies that the residuals from our data set do not have the full mean, variance, or correlation structure removed or that the residuals are not originally elliptically symmetric. Trying a different estimation process for the covariance structure is possible, but the DCC-GARCH is designed to achieve the relative accuracy of previous models with faster speed and less computing power. Since the DCC-GARCH is suggested as an improvement to previous models in both time and inference, it seems counter productive to suggest estimating real world data with a different process in order to aid in the evaluation of the robust estimation method for the DCC-GARCH. This discussion may open future research proposals, as suggested in section 5.3.
Even though the residuals are not spherically symmetric, an approximation to a spherically symmetric distribution is estimated from the squared radii of the points from the origin. The five dimensional residual cloud should be centered around the origin after subtracting off the estimate of the mean structure. The radius of each point from the origin is calculated using the Euclidean norm

$$\|\epsilon_t\| = \sqrt{\epsilon_{EU,t}^2 + \epsilon_{BR,t}^2 + \epsilon_{SF,t}^2 + \epsilon_{Y,t}^2 + \epsilon_{CD,t}^2}$$  \hspace{1cm} (4.4)$$

with $\epsilon_t$ is defined as the residual vector at time point $t$. The sample density plots for both the radii and squared radii are displayed in Figure B.1 in the Appendix. The squared radii provide valuable information when estimating the distribution of the residuals. If the residuals are spherically symmetric with a multivariate Normal distribution, the squared radii of the residuals follow a $\chi^2_5$ distribution. The degrees of freedom equals five because the data is five dimensional. If the residuals are spherically symmetrical with a multivariate $t_\nu$-distribution, the squared radii of the multivariate residuals follows an $F_5,\nu$ distribution. The numerator degrees of freedom equals five because the data is five dimensional. The denominator degrees of freedom $\nu$ equals the degrees of freedom of the multivariate $t_\nu$-distribution.

We know from the Q-Q plots developed by the Li et al. (1997) test in Figure 4.1 that the residuals are not spherically symmetric. However, fitting the $\chi^2$ and $F$ distributions to the data set of squared radii and estimating the respective degrees of freedom provides an idea of the closest spherical approximation to the distribution of our residuals. Calculating the maximum likelihood estimate of the degrees of freedom of the $\chi^2$ distribution with the squared radii provides an estimated 1.36 degrees of freedom. Figure 4.2(a) is a Q-Q plot of a $\chi^2_{1.36}$ distribution with the squared radii and reveals that the data do not follow the $\chi^2_{1.36}$ distribution very well.

Calculating the maximum likelihood estimates for both the numerator and denominator degrees of freedom in the likelihood of the $F$ distribution with the squared radii provides an estimated 2.75 numerator degrees of freedom and 3.48 denominator degrees of freedom. Figure 4.2(b) is a Q-Q plot of an $F_{2.75,3.48}$ distribution with the squared radii and reveals that the data follows the $F_{2.75,3.48}$ distribution better than the $\chi^2$ distribution.

One way to compare Q-Q plots is the $R^2$ value of the line through the origin in each of plots. Table 4.2 compares the $R^2$ value of the line through the origin of the $\chi^2_{1.36}$ plot and the $F_{2.75,3.48}$ plot. Since the last two points in each plot appear to fall outside the pattern of the rest of the data, the $R^2$ values are calculated with and without these last two points. The $R^2$ value for the $F_{2.75,3.48}$ distribution is extremely close to one, especially compared to the $\chi^2_{1.36}$ distribution.

Unfortunately, it is harder to estimate the quantiles for contaminated Normal distributions to get Q-Q plots. Optimizing contaminated Normal distribution likelihoods as previously done with the $\chi^2$ and $F$ distributions is also a difficult task. To get around both of these difficulties, squared radii are calculated from a randomly sampled five dimensional multivariate contaminated Normal
Table 4.2: Comparing $R^2$ Values of Q-Q Plots

<table>
<thead>
<tr>
<th>$R^2$ Values</th>
<th>Distribution</th>
<th>With Outliers</th>
<th>Without Outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2_{1.36}$</td>
<td>0.662</td>
<td>0.773</td>
<td></td>
</tr>
<tr>
<td>$F_{2.75,3.48}$</td>
<td>0.953</td>
<td>0.994</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: $R^2$ Combinations for Contaminated Normal Monte Carlo $R^2$ Values of Q-Q plots

<table>
<thead>
<tr>
<th>Contamination Std. Dev. $\sigma$</th>
<th>Contamination Probability $p$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.698 (0.003)</td>
<td>0.711 (0.002)</td>
<td>0.695 (0.002)</td>
<td>0.666 (0.002)</td>
<td>0.648 (0.002)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.856 (0.002)</td>
<td>0.787 (0.002)</td>
<td>0.746 (0.003)</td>
<td>0.708 (0.003)</td>
<td>0.678 (0.002)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.857 (0.003)</td>
<td>0.784 (0.003)</td>
<td>0.732 (0.002)</td>
<td>0.705 (0.002)</td>
<td>0.674 (0.002)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.829 (0.002)</td>
<td>0.772 (0.003)</td>
<td>0.728 (0.002)</td>
<td>0.701 (0.002)</td>
<td>0.666 (0.002)</td>
<td></td>
</tr>
</tbody>
</table>

distribution defined as

$$p \times N_5(\mathbf{0}, \mathbf{I}_5) + (1 - p) \times \sigma \times N_5(\mathbf{0}, \mathbf{I}_5),$$

with $p$ as the contamination probability and the contamination standard deviation $\sigma >> 1$. An $R^2$ value of a Q-Q plot between the sampled data and the squared radii is calculated. This process is repeated 100 times to get a Monte Carlo estimate of the $R^2$ value. This Monte Carlo simulation is repeated for the twenty combinations of $p = (0.05, 0.10, 0.15, 0.20, 0.25)$ and $\sigma = (2, 3, 4, 5)$. Table 4.3 summarizes these results. The combination of $p$ and $\sigma$ that produced the highest average $R^2$ value is $p = 0.05$ and either $\sigma = 3$ or $\sigma = 4$. Another pair of Monte Carlo simulations is calculated for these two combinations without the two possible outlier points in the data set. Their average $R^2$ values were 0.932 (0.01) for $\sigma = 3$ and 0.912 (0.01) for $\sigma = 4$. None of these values are close to the $R^2$ value using the optimized $F$ distribution.

Therefore, of all the tested distributions, an $F_{2.75,3.48}$ distribution appears to give the closest approximation to the distribution of the squared radii. The next section discusses the calculation of a spherically symmetric approximation to the distribution of the actual residuals using the estimated distribution of the squared radii.
Figure 4.2: Q-Q Plots of Squared Radii with $\chi^2_{1.36}$ (a) and $F_{2.75, 3.48}$ (b)
4.2 Data Driven Exploration

As shown in the previous section, the multivariate residuals of the foreign exchange rate data are not spherically symmetric. If the closest spherically symmetric approximation of the residuals is either a multivariate $t$-distribution or a contaminated Normal distribution, the results from section 3.3 compare the effectiveness of the new robust method with the traditional DCC-GARCH. The robust estimation of the DCC-GARCH is more effective in a couple of the tested situations from the previous chapter when the initial correlation matrix has highly correlated assets. When the assets are more diversified, the robust method performs much better in its estimation.

However, after studying the squared radii of the residuals from the origin, the closest spherically symmetric distribution is not either a multivariate $t$-distribution or a contaminated Normal distribution. Instead, the squared radii follow an $F_{2.75,3.48}$ distribution, which leads the residuals to follow a different distribution than the previously tested ones. Now a new data-driven data generating process (DDDGP) that more closely resembles real world data is created. A more accurate demonstration of the possible effectiveness of the robust estimation method for the DCC-GARCH in more realistic settings comes from this new data-driven data generating process. If the robust method outperforms the traditional DCC-GARCH in this controlled setting, then it is reasonable to assume that the analysis from the robust method on the actual foreign exchange rate data is more reliable than the traditional DCC-GARCH estimation results.

4.2.1 Data Driven Simulation and Results

A spherically symmetric distribution is created from the estimated $F_{2.75,3.48}$ distribution to compare the effectiveness of the DCC-GARCH and robust method in a controlled setting that closely resembles real world data. The same three asset parameters as in section 3.1 that are used in the previous comparisons of the two models are used for this controlled situation. This makes the results comparable to the previously tested underlying distributions. Again, the GARCH(1,1) volatility parameters of the three assets are

$$
a_0 = \begin{pmatrix} 0.12 \\
0.60 \\
0.32 \end{pmatrix}, \quad a_1 = \begin{pmatrix} 0.23 \\
0.10 \\
0.15 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 0.59 \\
0.82 \\
0.74 \end{pmatrix},
$$

(4.6)

and the correlation parameters are

$$\alpha = 0.24, \quad \beta = 0.70.
$$

(4.7)

The simulation is done with both the highly correlated initial correlation structure of equation 3.4 and the more diversified initial correlation structure of equation 3.5.

First, 500 points are uniformly generated across the three dimensional unit sphere. Random samples from a three dimensional multivariate Normal distribution with a mean of zero and identity
covariance matrix is projected onto the unit sphere by
\[
\frac{x_t}{\sqrt{\sum_{i=1}^{5} x^2_{i,t}}}, \quad x_t = (x_{1,t}, \ldots, x_{5,t}).
\]
In addition, 500 random observations are drawn from an \( F_{2.75,3.48} \) distribution, because this is the estimated distribution of the squared radii. After taking the square roots of these observations, each observation is randomly multiplied to one of the 500 points on the unit sphere. The covariance matrix estimate of 500 of these distributions is taken to estimate the scale to make \( \text{Var}(\epsilon_t) = I_5 \). This provides a random sample of a spherically symmetric distribution with similar squared radii to the estimated foreign exchange rate residuals. Using the same DCC-GARCH structure mentioned in equation 1.11 and the above mentioned parameter values, estimates of the covariance structure matrix \( H_t \) are calculated for each time point. The Cholesky factorization of this covariance matrix is multiplied by the vector of generated residuals at each time point as
\[
r_{DD} = H_t^{-1/2} \epsilon_{t,DD},
\]
where \( r_{DD} \) and \( \epsilon_{t,DD} \) are the estimated returns and residuals from the data-driven data generating process.

The calculated data-driven returns, \( r_{DD} \), are then fitted with both the traditional DCC-GARCH model and the proposed robust estimation method for the DCC-GARCH model, with and without the bias correction term. This process is repeated 250 times to get a Monte Carlo simulation similar to the previous tests of these models. The first analysis has the initial correlation matrix equal to the first initial correlation matrix in equation 3.4. The results of this data-driven data generating process are in Table 4.4. The Robust \( BC \) method is the robust method with bias correction. The second analysis has the second initial correlation matrix from equation 3.5. These results are located in Table 4.5.

With regards to all of the volatility parameters, the robust method MSE values are lower than their respective MLE counterparts in most assets when analyzing the first correlation structure. When examining each asset individually, the estimated MSE value with the robust estimation of the DCC-GARCH model is lower than the MSE value with the DCC-GARCH for every volatility parameter estimate in the model except the first \( a_1 \) and \( b_1 \). The third asset’s estimated \( b_1 \) parameter has a comparable MSE value with the traditional DCC-GARCH estimate because the estimates are within one standard error of each other. The bias corrected robust method produces lower MSE values for the volatility estimates of \( a_0 \) compared with the same MSE values produced by ML estimation. However, the MSE values for the estimates of \( a_1 \) and \( b_1 \) are higher for the bias corrected robust method compared to the DCC-GARCH model in every asset. In fact, the MSE values for the robust method without the bias correction are lower than the MSE values with the bias correction for a majority of the estimates of \( a_1 \) and \( b_1 \). The robust method with bias correction better estimates the volatility
Table 4.4: Set 1 Results for Data Driven Comparison of Methods

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimation Method</th>
<th>MLE</th>
<th>Robust</th>
<th>Robust(_{BC})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.0129 (0.0020)</td>
<td>0.0049 (0.0002)</td>
<td>0.0052 (0.0003)</td>
</tr>
<tr>
<td>(a_{0,1})</td>
<td></td>
<td>0.7012 (0.1050)</td>
<td>0.1939 (0.0086)</td>
<td>0.1436 (0.0093)</td>
</tr>
<tr>
<td>(a_{0,2})</td>
<td></td>
<td>0.0875 (0.0127)</td>
<td>0.0498 (0.0016)</td>
<td>0.0354 (0.0022)</td>
</tr>
<tr>
<td>(a_{0,3})</td>
<td></td>
<td>0.0697 (0.0071)</td>
<td>0.1027 (0.0105)</td>
<td>0.0841 (0.0073)</td>
</tr>
<tr>
<td>(a_{1,1})</td>
<td></td>
<td>0.0201 (0.0028)</td>
<td>0.0069 (0.0001)</td>
<td>0.0308 (0.0048)</td>
</tr>
<tr>
<td>(a_{1,2})</td>
<td></td>
<td>0.0317 (0.0048)</td>
<td>0.0151 (0.0003)</td>
<td>0.0320 (0.0035)</td>
</tr>
<tr>
<td>(a_{1,3})</td>
<td></td>
<td>0.0528 (0.0043)</td>
<td>0.1023 (0.0083)</td>
<td>0.1898 (0.0084)</td>
</tr>
<tr>
<td>(b_{1,1})</td>
<td></td>
<td>0.0361 (0.0047)</td>
<td>0.0257 (0.0040)</td>
<td>0.3669 (0.0184)</td>
</tr>
<tr>
<td>(b_{1,2})</td>
<td></td>
<td>0.0320 (0.0040)</td>
<td>0.0362 (0.0046)</td>
<td>0.2678 (0.0143)</td>
</tr>
<tr>
<td>(\alpha)</td>
<td></td>
<td>0.0180 (0.0010)</td>
<td>0.0326 (0.0005)</td>
<td>0.0037 (0.0002)</td>
</tr>
<tr>
<td>(\beta)</td>
<td></td>
<td>0.2872 (0.0128)</td>
<td>0.4842 (0.0009)</td>
<td>0.4884 (0.0003)</td>
</tr>
</tbody>
</table>

The intercept parameter \(a_0\) compared to the robust method without the bias correction. However, both are better estimates than the DCC-GARCH when comparing MSE values of the estimates.

The correlation parameters continue to have the same pattern with the data-driven data generating process as they did with the other underlying distributions with the first initial correlation structure. Both of the correlation parameter estimates of \(\alpha\) and \(\beta\) have lower estimated MSE values for the DCC-GARCH compared to the robust estimation method for the DCC-GARCH. The estimates of the correlation parameters from the robust method without bias correction have never been better than the DCC-GARCH estimates in any distribution tried in this study with the first initial correlation structure. The robust method with a bias correction does produce a significantly lower estimated MSE value for the \(\alpha\) parameter estimate in comparison to the value produced by the DCC-GARCH. However, the correlation parameter \(\beta\) is still better estimated with the traditional DCC-GARCH model.

The same analysis is performed with the more diversified initial correlation structure. In the previous test distributions the diversified set of assets increased the effectiveness of the robust estimation. That trend slightly continues with the data-driven data generating process. On average and individually across assets, the intercept parameter \(a_0\) is better estimated with the robust method, with and without the bias correction, in comparison with the DCC-GARCH model. The bias corrected robust method does outperform the robust method in these parameter estimates as well. The volatility parameter estimates of \(a_1\) have a lower estimated robust MSE compared to the maximum likelihood MSE value in two of the three assets. The bias corrected robust method only has one asset with a lower MSE value for the estimated parameters. However, the bias corrected robust method loses its effectiveness in the \(b_1\) volatility parameter compared to either of the other two methods. Two of the
Table 4.5: Set 2 Results for Data Driven Comparison of Methods

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>Robust</th>
<th>Robust&lt;sub&gt;BC&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_{0,1})</td>
<td>0.0154 (0.0025)</td>
<td>0.0025 (0.0002)</td>
<td>0.0048 (0.0003)</td>
</tr>
<tr>
<td>(a_{0,2})</td>
<td>0.5871 (0.0838)</td>
<td>0.1822 (0.0067)</td>
<td>0.1375 (0.0086)</td>
</tr>
<tr>
<td>(a_{0,3})</td>
<td>0.0970 (0.0127)</td>
<td>0.0395 (0.0016)</td>
<td>0.0294 (0.0023)</td>
</tr>
<tr>
<td>(a_{1,1})</td>
<td>0.0572 (0.0058)</td>
<td>0.1454 (0.0131)</td>
<td>0.1019 (0.0099)</td>
</tr>
<tr>
<td>(a_{1,2})</td>
<td>0.0237 (0.0031)</td>
<td>0.0070 (0.0001)</td>
<td>0.0070 (0.0001)</td>
</tr>
<tr>
<td>(a_{1,3})</td>
<td>0.0460 (0.0053)</td>
<td>0.0148 (0.0010)</td>
<td>0.0344 (0.0046)</td>
</tr>
<tr>
<td>(b_{1,1})</td>
<td>0.0471 (0.0037)</td>
<td>0.2347 (0.0098)</td>
<td>0.1975 (0.0091)</td>
</tr>
<tr>
<td>(b_{1,2})</td>
<td>0.0350 (0.0046)</td>
<td>0.0299 (0.0046)</td>
<td>0.3764 (0.0182)</td>
</tr>
<tr>
<td>(b_{1,3})</td>
<td>0.0456 (0.0048)</td>
<td>0.0427 (0.0054)</td>
<td>0.3337 (0.0134)</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.0155 (0.0009)</td>
<td>0.0698 (0.0011)</td>
<td>0.0091 (0.0005)</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.2390 (0.0132)</td>
<td>0.4706 (0.0022)</td>
<td>0.4895 (0.0001)</td>
</tr>
</tbody>
</table>

three assets’ estimates have lower MSE values for the robust method compared to the DCC-GARCH. Overall, the robust method outperforms the DCC-GARCH in volatility parameter estimation in the more diversified set of assets.

Unlike the previously tested distributions, the robust method does not outperform the DCC-GARCH on either of the correlation parameters under this second initial correlation structure. The correlation parameter estimates of \(\alpha\) and \(\beta\) have MSE values that are four times and two times higher in value compared to their respective maximum likelihood estimate MSE values. Although the bias corrected robust method best predicts the \(\alpha\) parameter, that same method is also the worst model for predicting the \(\beta\) parameter. The data-driven data generating process is the only tested distribution where the more diversified set does not aid the robust method in estimating the correlation parameters in the model. This analysis is similar to the first set with the volatility parameters best estimated with the robust method while the correlation parameters are best estimated with the DCC-GARCH model on average.

From these results, the estimation of the volatility parameters by the robust estimation method for the DCC-GARCH without bias correction mostly outperforms the DCC-GARCH estimation in a realistic test of the two methods using a data-driven data generating process in either type of initial correlation structure. However, the traditional DCC-GARCH still estimates the correlation parameters of the data better than the robust method without bias correction. Even with the bias correction term, only the estimation of the correlation parameter \(\alpha\) benefits. Since the correlation parameters have a mutual bound of \(\alpha + \beta < 1\), it is not recommended to use different estimation techniques to estimate the \(\alpha\) and \(\beta\) parameters separately. Either the maximum likelihood estimation or the bias correction is preferred.

79
### Table 4.6: Volatility Parameters Between Robust and Traditional Method

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Euro MLE</th>
<th>Euro Robust</th>
<th>Pound MLE</th>
<th>Pound Robust</th>
<th>Franc MLE</th>
<th>Franc Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0.29e-4</td>
<td>0.92e-5</td>
<td>0.21e-4</td>
<td>0.11e-4</td>
<td>0.28e-4</td>
<td>0.15e-4</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.183</td>
<td>0.103</td>
<td>0.204</td>
<td>0.196</td>
<td>0.171</td>
<td>0.133</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.127</td>
<td>0.047</td>
<td>0.270</td>
<td>0.251</td>
<td>0.152</td>
<td>0.125</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.181</td>
<td>0.146</td>
<td>0.134</td>
<td>0.168</td>
<td>0.191</td>
<td>0.201</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.170</td>
<td>0.444</td>
<td>0.120</td>
<td>0.134</td>
<td>0.170</td>
<td>0.233</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Yen MLE</th>
<th>Yen Robust</th>
<th>Dollar MLE</th>
<th>Dollar Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_0$</td>
<td>0.30e-4</td>
<td>0.16e-4</td>
<td>0.91e-5</td>
<td>0.98e-5</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.146</td>
<td>0.131</td>
<td>0.198</td>
<td>0.186</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.127</td>
<td>0.199</td>
<td>0.442</td>
<td>0.312</td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.373</td>
<td>0.356</td>
<td>0.024</td>
<td>0.055</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.121</td>
<td>0.137</td>
<td>0.113</td>
<td>0.184</td>
</tr>
</tbody>
</table>

The corrected robust method should be used in the estimation of both parameters.

#### 4.2.2 Application to Foreign Exchange Rates

The previous section discusses the effectiveness of the robust estimation method for the DCC-GARCH model relative to the DCC-GARCH model with a distribution estimated from foreign exchange rate data. On average, the volatility parameters in the model are better estimated with the proposed robust method instead of maximum likelihood estimation. The correlation parameters however are still better modeled with maximum likelihood estimation. The same foreign exchange rate data from section 4.1 are now analyzed with the robust estimation of the DCC-GARCH model.

The set of daily returns consists of the five United States foreign exchange rates with the Euro, British Pound, Swiss Franc, Japanese Yen, and Canadian Dollar. The daily returns of these exchange rates are collected from the beginning of the Euro entering the exchange market on January 4, 1999 until October 16, 2009. The mean structure of the set of returns is modeled with a vector autoregressive model of order one (VAR(1)) in section 4.1. The residuals of this VAR(1) model are analyzed originally with the DCC-GARCH(3,1) model to achieve the parameter estimates in Table 4.1.

The parameter estimates with the robust method are different. Table 4.6 compares the values of the volatility parameter estimates in the two different models while Table 4.7 compares the values of the correlation parameter estimates. The time structure of both the volatility and correlation
Table 4.7: Correlation Parameters Between Robust and Traditional Method

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimation</th>
<th>MLE</th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td></td>
<td>0.212</td>
<td>0.299</td>
</tr>
<tr>
<td>$\beta$</td>
<td></td>
<td>0.776</td>
<td>0.695</td>
</tr>
</tbody>
</table>

parameters change.

In the maximum likelihood estimation of the DCC-GARCH model, the volatility parameters have a mean sum value of the terms $a_1 + a_2 + a_3 + b_1$ of 0.723. This implies a time dependency as the volatilities change between periods. In the robust estimation method for the DCC-GARCH these parameters have an average sum estimate of 0.748, with four of the five assets having a larger estimated value under the robust method. This implies a higher dependency on the previous period’s volatility estimate than originally estimated with the DCC-GARCH. The intercept parameters slightly shrink in the robust model for four of the five assets. From the previous simulation study, the robust method estimates are believed to be more accurate than the DCC-GARCH.

Figures 4.3, 4.4(a), and 4.4(b) compare the average estimated univariate variances across the assets in the set of returns with both the maximum likelihood estimation (black line) and robust estimation (red line). The new robustly predicted variances appear to have smaller shifts than the traditionally predicted volatilities. Also, these robust variance estimates are slightly smaller in value due to the estimated intercept term decreasing in the robust method. The individual foreign exchange rate comparison plots between the robust estimated volatility and the maximum likelihood estimated volatility are in Appendix B in Figures B.2, B.3, B.4, B.5, and B.6.
Figure 4.3: Overlayed Estimated Volatility Time Plots
Figure 4.4: Estimated Volatility Time Plots with [a] MLE and [b] Robust Estimation
In the estimation of the correlation parameters, the two models are similar in their interpretation of the set of returns. The parameter estimation from both models is displayed in Table 4.7. The maximum likelihood estimation of the correlation parameter puts a higher weight on the parameter estimate of $\beta$ compared to the robust estimation technique. In contrast, the robust technique puts a higher weight on the parameter estimate of $\alpha$ compared to the maximum likelihood estimation. The robust technique implies that more weight should be placed on the incoming data compared to the maximum likelihood estimation. In the same regards, the robust technique implies less weight should be placed on the previous values of the correlation matrix compared to the maximum likelihood estimation.

Figure 4.5 displays the time plot of the determinant of the correlation matrix under both the robust estimation and ML estimation. The black line on the plot represents the determinant of the estimated correlation matrix across time for the ML estimation, while the red line represents the robust method estimation. This figure displays the differences in how the entire correlation matrices change over time between the two models. Figure 4.6(a) and (b) displays the time plot of the determinant of the correlation matrix under these two methods separately. The maximum likelihood estimation is believed to be more reliable because of the results of the previous simulation study, where the DCC-GARCH better estimates the correlation parameters compared to the robust estimation method for the DCC-GARCH.

Figure 4.7(a) and (b) displays the time plot of the average value of the correlations between assets under both the ML estimation and robust estimation respectively. The black line on the plot represents
the average value of the estimated correlations across time for the ML estimation, while the red line represents the robust method estimation. The underlying pattern of the correlations does seem about equal, but the magnitudes of the shifts are different. Under maximum likelihood estimation, the average correlation between the assets does not have shifts that are as large as the shifts in correlation for the robust method. The comparisons of each of the correlation combinations between the assets in located in Appendix B in Figures B.7, B.8, B.9, B.10, and B.11. The maximum likelihood estimation is believed to be more reliable because of the results of the previous simulation study, where the DCC-GARCH better estimates the correlation parameters compared to the robust method.
Figure 4.6: Time Plots of Determinant of Estimated Correlation Matrices

[a] MLE Estimated Determinant

[b] Robust Estimated Determinant

86
Figure 4.7: Time Plots of Average Correlation Between Assets
5.1 Summary

Volatility modeling plays a crucial role in the study of financial mathematics, economics, and statistics. Estimating and forecasting volatilities of financial assets is important in making financial decisions through the understanding of potential risk. Inaccurately modeling volatilities leads to potentially higher risk for a set of assets. Higher financial risk leads to higher potential loss.

One of the popular multivariate models for the volatilities and correlations of a set of assets is the Dynamic Conditional Correlation Generalized Autoregressive Heteroscedasticity (DCC-GARCH) model developed by Engle (2002). This is one of the many models in the long line of adaptations to the original ARCH and GARCH models developed by Engle (1982) and Bollerslev (1986) respectively. However, the structure of the DCC-GARCH model makes the model susceptible to estimation concerns in the presence of anomalous observations, called outliers, or heavy tailed distributions.

The effects of outliers have been studied for many years. It has been shown that outliers affect the estimation of models with an autoregressive structure. Both additive and innovation outliers lead to estimation problems in these types of models. Outliers also affect models that have heteroscedasticity. It has also been shown that outliers affect covariance estimation. The DCC-GARCH model structure is constructed with each of these components as assumptions or parts of the model. This makes the DCC-GARCH model inherently affected by outliers. Through simulation studies, the DCC-GARCH model estimation is shown to deteriorate in the presence of heavy tailed distributions or outlier-contaminated distributions. For these reasons, a robust estimation method for the DCC-GARCH model using the technique of bounded deviance estimation is proposed to alleviate these estimation concerns.
concerns.

5.2 Conclusion

Two robust estimations of the DCC-GARCH model are proposed to solve the estimation problems of the DCC-GARCH model in the presence of outliers or heavy tailed distributions. The first robust method does not contain a bias correction term to adjust for potential bias in the estimation technique. The second robust method contains this bias correction term. Although the bias correction term adjusts for some potential bias in the model, it adds unneeded variability in the estimation process. For this reason, the robust method without bias correction has better estimates of the volatility parameters in the model with regards to MSE, while the bias-corrected robust method better estimates the correlation parameters. Two situations are also studied, with one set of assets being more diversified than the other.

First, the robust estimation method for the DCC-GARCH model is tested with a set of assets that is not diversified. Even though the robust method without bias correction better estimated the volatility portion of the model compared to the bias corrected version, the robust method did not consistently outperform the maximum likelihood estimates in models with only slight outlier contamination or slightly heavier tails. The robust method does not possess one of the desired properties, of accurate estimation in models with a lack of outliers. The robust estimation method for the DCC-GARCH outperforms the DCC-GARCH in estimation with an underlying distribution with heavy tails. This shows that the robust method still possesses the property of better estimation in the presence of extreme tailed distributions. The robust method is never able to compete with the DCC-GARCH in the estimation of the correlation parameters in the model. The bias-corrected robust method does a better job of predicting one of the correlation parameters, but not both.

Next, the robust estimation method for the DCC-GARCH model is tested with a set of assets that is more diversified. This results in better effectiveness of the robust method compared to the non-diversified set of assets. The robust method began to consistently beat the DCC-GARCH in the effectiveness of modeling the volatility parameters when either heavy tailed distributions or outlier contaminated distributions were present. The robust method also performs much better in estimating the correlation parameters with this more diversified set. In four out of the five distributions tested, the robust method outperforms the DCC-GARCH in estimating at least one of the two correlation parameters. However, the robust method still does not possess one of the desired properties of accurate estimation in models with a lack of outliers.

The robust estimation method for the DCC-GARCH does not consistently outperform the DCC-GARCH in every situation, so a study of the distribution of real world data is needed to determine the true effectiveness of the robust method. A set of foreign exchange rates is studied in detail to generate a data-driven data generating process, to compare the effectiveness of these two models.
in a controlled setting. With highly correlated assets and diversified assets, the robust method significantly outperformed the DCC-GARCH for every volatility parameter estimation. However, in neither situation does the robust estimation of the DCC-GARCH outperform the DCC-GARCH in correlation parameter estimation. Therefore, the use of the robust method would benefit the estimation of the volatility parameters of real world data, while the maximum likelihood estimates of the correlation parameters should be used. The robust estimation method for the DCC-GARCH model is a split estimation based model; this means that the volatility parameters can be estimated with the robust method without affecting the ML estimates of the correlation parameters.

5.3 Suggestions for Future Research

The idea of applying robust estimation techniques to multivariate ARCH/GARCH modeling is still a young one. Only in the last decade has the idea of robust estimation of univariate ARCH/GARCH models become popular. There are many different robust estimation approaches to correct for the effect of outliers or heavy tailed distributions. Bounded deviance estimation is only one of the many different techniques available to the many different multivariate ARCH/GARCH models. The M-estimation and scaled residuals approaches of robustly estimating the BEKK model by Boudt and Croux (2009) is one possible combination. There are many different directions to which the field of robust estimation of multivariate GARCH models may deviate.

Another concern for future research is adapting another possible method to the robust approach to improve the estimation of the correlation parameters in the model. Although the robust method performs very well in estimating the volatility parameters, there exists a need for a robust estimation process that accurately describes the correlation parameters. Since the robust method is a multiple step robust procedure, a possible adaptation of the model is using different bounding functions for each portion of the estimation. Perhaps a different bounding function $\lambda$ could be developed to improve the correlation parameter estimation.

Another area for future research involves the data-driven data generating process. In the estimation of the distribution of the foreign exchange rate data, the DCC-GARCH model is used. However, the DCC-GARCH is already shown to be inaccurate in the presence of outliers or heavy tailed distributions. A robust estimate of this distribution would be beneficial, but the robust method would have been inappropriate as it had not yet been proven to outperform the DCC-GARCH. A more parameter intensive model, such as the BEKK model, is a possibility. The intensive models were not used here because it was felt inappropriate to suggest a computationally intensive model to initially estimate data before estimating with a robust approach focused on a less computationally intensive model.
REFERENCES


A.1 DCC-GARCH Likelihood Function

The DCC-GARCH model developed by Engle (2002), assumes the assets’ returns follows a multivariate Normal distribution as $r_t | \psi_{t-1} \sim N(0, H_t)$. This leads to the likelihood function mentioned in equation 1.12. The detailed results of how Engle calculated this log likelihood equation is as follows:

$$L = -\frac{1}{2} \sum_{t=1}^{T} \left( n \log 2\pi + \log |H_t| + r_t' H_t^{-1} r_t \right), \quad H_t = D_t R_t D_t$$

$$= -\frac{1}{2} \sum_{t=1}^{T} \left( n \log 2\pi + \log |D_t R_t D_t| + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t \right), \quad \epsilon_t = D_t^{-1} r_t$$

$$= -\frac{1}{2} \sum_{t=1}^{T} \left( n \log 2\pi + 2 \log |D_t| + \log |R_t| + \epsilon_t' R_t^{-1} \epsilon_t \right)$$

$$= -\frac{1}{2} \sum_{t=1}^{T} \left( n \log 2\pi + 2 \log |D_t| + \log |R_t| + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t - \epsilon_t' \epsilon_t + \epsilon_t' R_t^{-1} \epsilon_t \right)$$

A.2 Complete Initial DCC-GARCH Simulations

The robust estimation method for the DCC-GARCH is compared to the DCC-GARCH model in the estimation of return data with five underlying distributions. Tables A.1, A.2, and A.3 use the initial correlation matrix in equation 3.4. Tables A.4, A.5, and A.6 use the initial correlation matrix in equation 3.5.
Table A.1: Complete MSE Results Across Generating Processes with First Portfolio

MSE for Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal MLE</th>
<th>Normal Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{0,1}$</td>
<td>0.0026 (0.0002)</td>
<td>0.0047 (0.0005)</td>
</tr>
<tr>
<td>$a_{0,2}$</td>
<td>0.1066 (0.0136)</td>
<td>0.1883 (0.0188)</td>
</tr>
<tr>
<td>$a_{0,3}$</td>
<td>0.0320 (0.0044)</td>
<td>0.0795 (0.0085)</td>
</tr>
<tr>
<td>$a_{1,1}$</td>
<td>0.0141 (0.0018)</td>
<td>0.0042 (0.0003)</td>
</tr>
<tr>
<td>$a_{1,2}$</td>
<td>0.0026 (0.0003)</td>
<td>0.0019 (0.0001)</td>
</tr>
<tr>
<td>$a_{1,3}$</td>
<td>0.0039 (0.0004)</td>
<td>0.0030 (0.0026)</td>
</tr>
<tr>
<td>$b_{1,1}$</td>
<td>0.0141 (0.0015)</td>
<td>0.0294 (0.0009)</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>0.0040 (0.0004)</td>
<td>0.0099 (0.0021)</td>
</tr>
<tr>
<td>$b_{1,3}$</td>
<td>0.0083 (0.0010)</td>
<td>0.0209 (0.0002)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0026 (0.0002)</td>
<td>0.0095 (0.0002)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0012 (0.0001)</td>
<td>0.0047 (0.0002)</td>
</tr>
</tbody>
</table>

MSE for Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$t_6$ MLE</th>
<th>$t_6$ Robust</th>
<th>$t_3$ MLE</th>
<th>$t_3$ Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{0,1}$</td>
<td>0.0032 (0.0002)</td>
<td>0.0045 (0.0002)</td>
<td>0.0092 (0.0012)</td>
<td>0.0047 (0.0002)</td>
</tr>
<tr>
<td>$a_{0,2}$</td>
<td>0.1463 (0.0230)</td>
<td>0.1832 (0.0167)</td>
<td>0.8985 (0.1680)</td>
<td>0.1533 (0.0074)</td>
</tr>
<tr>
<td>$a_{0,3}$</td>
<td>0.0392 (0.0045)</td>
<td>0.0373 (0.0025)</td>
<td>0.2679 (0.0495)</td>
<td>0.0413 (0.0021)</td>
</tr>
<tr>
<td>$a_{1,1}$</td>
<td>0.0270 (0.0027)</td>
<td>0.0278 (0.0021)</td>
<td>0.0619 (0.0066)</td>
<td>0.0599 (0.0057)</td>
</tr>
<tr>
<td>$a_{1,2}$</td>
<td>0.0081 (0.0014)</td>
<td>0.0051 (0.0001)</td>
<td>0.0307 (0.0053)</td>
<td>0.0059 (0.0001)</td>
</tr>
<tr>
<td>$a_{1,3}$</td>
<td>0.0089 (0.0009)</td>
<td>0.0102 (0.0003)</td>
<td>0.0471 (0.0066)</td>
<td>0.0127 (0.0003)</td>
</tr>
<tr>
<td>$b_{1,1}$</td>
<td>0.0245 (0.0020)</td>
<td>0.0424 (0.0044)</td>
<td>0.0484 (0.0043)</td>
<td>0.0672 (0.0060)</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>0.0098 (0.0015)</td>
<td>0.0165 (0.0022)</td>
<td>0.0421 (0.0063)</td>
<td>0.0267 (0.0037)</td>
</tr>
<tr>
<td>$b_{1,3}$</td>
<td>0.0157 (0.0018)</td>
<td>0.0247 (0.0030)</td>
<td>0.0553 (0.0063)</td>
<td>0.0293 (0.0034)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0043 (0.0003)</td>
<td>0.0348 (0.0008)</td>
<td>0.0225 (0.0009)</td>
<td>0.0345 (0.0007)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0124 (0.0011)</td>
<td>0.0717 (0.0040)</td>
<td>0.0318 (0.0016)</td>
<td>0.1234 (0.0063)</td>
</tr>
</tbody>
</table>

MSE for Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CN 5% MLE</th>
<th>CN 5% Robust</th>
<th>CN 25% MLE</th>
<th>CN 25% Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{0,1}$</td>
<td>0.0033 (0.0002)</td>
<td>0.0033 (0.0001)</td>
<td>0.0043 (0.0004)</td>
<td>0.0052 (0.0002)</td>
</tr>
<tr>
<td>$a_{0,2}$</td>
<td>0.3114 (0.0342)</td>
<td>0.1718 (0.0115)</td>
<td>0.2181 (0.0213)</td>
<td>0.1990 (0.0069)</td>
</tr>
<tr>
<td>$a_{0,3}$</td>
<td>0.0592 (0.0055)</td>
<td>0.0360 (0.0018)</td>
<td>0.0389 (0.0035)</td>
<td>0.0466 (0.0010)</td>
</tr>
<tr>
<td>$a_{1,1}$</td>
<td>0.0264 (0.0019)</td>
<td>0.0262 (0.0009)</td>
<td>0.0517 (0.0046)</td>
<td>0.1107 (0.0090)</td>
</tr>
<tr>
<td>$a_{1,2}$</td>
<td>0.0073 (0.0007)</td>
<td>0.0052 (0.0001)</td>
<td>0.0107 (0.0013)</td>
<td>0.0068 (0.0001)</td>
</tr>
<tr>
<td>$a_{1,3}$</td>
<td>0.0140 (0.0012)</td>
<td>0.0109 (0.0002)</td>
<td>0.0121 (0.0011)</td>
<td>0.0147 (0.0002)</td>
</tr>
<tr>
<td>$b_{1,1}$</td>
<td>0.0243 (0.0013)</td>
<td>0.0320 (0.0023)</td>
<td>0.0357 (0.0028)</td>
<td>0.0909 (0.0069)</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>0.0154 (0.0016)</td>
<td>0.0192 (0.0022)</td>
<td>0.0135 (0.0014)</td>
<td>0.0193 (0.0026)</td>
</tr>
<tr>
<td>$b_{1,3}$</td>
<td>0.0200 (0.0016)</td>
<td>0.0267 (0.0026)</td>
<td>0.0162 (0.0014)</td>
<td>0.0188 (0.0019)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0084 (0.0003)</td>
<td>0.0363 (0.0006)</td>
<td>0.0101 (0.0002)</td>
<td>0.0418 (0.0007)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0169 (0.0005)</td>
<td>0.0729 (0.0036)</td>
<td>0.0201 (0.0005)</td>
<td>0.1771 (0.0069)</td>
</tr>
<tr>
<td>Parameter</td>
<td>Normal</td>
<td>$t_6$</td>
<td>$t_3$</td>
<td>CN 5%</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
<td>Robust</td>
</tr>
<tr>
<td>$a_{0,1}$</td>
<td>-0.0296</td>
<td>0.0117</td>
<td>-0.0232</td>
<td>-0.0393</td>
</tr>
<tr>
<td>$a_{0,2}$</td>
<td>-0.0570</td>
<td>0.0661</td>
<td>-0.0080</td>
<td>-0.1679</td>
</tr>
<tr>
<td>$a_{0,3}$</td>
<td>-0.0142</td>
<td>0.0559</td>
<td>0.0012</td>
<td>-0.0912</td>
</tr>
<tr>
<td>$a_{1,1}$</td>
<td>0.0545</td>
<td>-0.294</td>
<td>0.0651</td>
<td>-0.1292</td>
</tr>
<tr>
<td>$a_{1,2}$</td>
<td>0.0209</td>
<td>-0.0141</td>
<td>0.0304</td>
<td>-0.0678</td>
</tr>
<tr>
<td>$a_{1,3}$</td>
<td>0.0218</td>
<td>-0.0240</td>
<td>0.0343</td>
<td>-0.0978</td>
</tr>
<tr>
<td>$b_{1,1}$</td>
<td>0.0337</td>
<td>-0.0594</td>
<td>0.0269</td>
<td>-0.0571</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>-0.0011</td>
<td>-0.0241</td>
<td>-0.0156</td>
<td>-0.0145</td>
</tr>
<tr>
<td>$b_{1,3}$</td>
<td>-0.0032</td>
<td>-0.0405</td>
<td>-0.0147</td>
<td>-0.0276</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0439</td>
<td>0.0949</td>
<td>-0.0510</td>
<td>0.1864</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.0029</td>
<td>-0.0597</td>
<td>0.0874</td>
<td>-0.2488</td>
</tr>
<tr>
<td>Parameter</td>
<td>Normal MLE</td>
<td>Robust MLE</td>
<td>Normal Robust MLE</td>
<td>Robust MLE</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>------------</td>
<td>-------------------</td>
<td>------------</td>
</tr>
<tr>
<td>$a_{0,1}$</td>
<td>0.0017</td>
<td>0.0048</td>
<td>0.0027</td>
<td>0.0030</td>
</tr>
<tr>
<td>$a_{0,2}$</td>
<td>0.1034</td>
<td>0.1754</td>
<td>0.1463</td>
<td>0.1550</td>
</tr>
<tr>
<td>$a_{0,3}$</td>
<td>0.0318</td>
<td>0.0730</td>
<td>0.0392</td>
<td>0.0290</td>
</tr>
<tr>
<td>$a_{1,1}$</td>
<td>0.0112</td>
<td>0.0033</td>
<td>0.0227</td>
<td>0.0111</td>
</tr>
<tr>
<td>$a_{1,2}$</td>
<td>0.0022</td>
<td>0.0017</td>
<td>0.0072</td>
<td>0.0005</td>
</tr>
<tr>
<td>$a_{1,3}$</td>
<td>0.0035</td>
<td>0.0023</td>
<td>0.0078</td>
<td>0.0007</td>
</tr>
<tr>
<td>$b_{1,1}$</td>
<td>0.0129</td>
<td>0.0261</td>
<td>0.0238</td>
<td>0.0391</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>0.0040</td>
<td>0.0091</td>
<td>0.0096</td>
<td>0.0163</td>
</tr>
<tr>
<td>$b_{1,3}$</td>
<td>0.0083</td>
<td>0.0188</td>
<td>0.0155</td>
<td>0.0240</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0006</td>
<td>0.0005</td>
<td>0.0017</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0048</td>
<td>0.0098</td>
</tr>
</tbody>
</table>
Table A.4: Complete Results Across Generating Processes with Second Portfolio

### MSE for Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td></td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0027 (0.0002)</td>
<td>0.0038 (0.0002)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.1455 (0.0679)</td>
<td>0.1140 (0.0042)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0263 (0.0037)</td>
<td>0.0325 (0.0016)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0142 (0.0026)</td>
<td>0.0185 (0.0010)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0027 (0.0003)</td>
<td>0.0026 (0.0001)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0051 (0.0008)</td>
<td>0.0039 (0.0001)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0144 (0.0019)</td>
<td>0.0298 (0.0016)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0052 (0.0017)</td>
<td>0.0134 (0.0009)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0081 (0.0010)</td>
<td>0.0179 (0.0011)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0041 (0.0002)</td>
<td>0.0172 (0.0002)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0013 (0.0001)</td>
<td>0.0643 (0.0004)</td>
</tr>
</tbody>
</table>

### MSE for Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$t_6$</th>
<th>Robust</th>
<th>$t_3$</th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td></td>
<td>MLE</td>
<td></td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0032 (0.0004)</td>
<td>0.0040 (0.0004)</td>
<td>0.0165 (0.0081)</td>
<td>0.0038 (0.0001)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.1866 (0.0422)</td>
<td>0.1413 (0.0138)</td>
<td>0.5892 (0.1395)</td>
<td>0.1584 (0.0113)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0476 (0.0098)</td>
<td>0.0195 (0.0014)</td>
<td>0.1300 (0.0320)</td>
<td>0.0328 (0.0017)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0024 (0.0026)</td>
<td>0.0072 (0.0011)</td>
<td>0.0605 (0.0068)</td>
<td>0.0171 (0.0005)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0095 (0.0012)</td>
<td>0.0021 (0.0001)</td>
<td>0.0278 (0.0058)</td>
<td>0.0040 (0.0001)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0101 (0.0018)</td>
<td>0.0036 (0.0002)</td>
<td>0.0538 (0.0079)</td>
<td>0.0078 (0.0002)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0201 (0.0018)</td>
<td>0.0364 (0.0035)</td>
<td>0.0514 (0.0043)</td>
<td>0.0721 (0.0046)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0121 (0.0028)</td>
<td>0.0117 (0.0016)</td>
<td>0.0334 (0.0057)</td>
<td>0.0208 (0.0024)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0140 (0.0018)</td>
<td>0.0113 (0.0015)</td>
<td>0.0558 (0.0072)</td>
<td>0.0203 (0.0021)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0035 (0.0004)</td>
<td>0.0089 (0.0003)</td>
<td>0.0213 (0.0008)</td>
<td>0.0051 (0.0002)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0101 (0.0006)</td>
<td>0.0044 (0.0005)</td>
<td>0.0341 (0.0011)</td>
<td>0.0046 (0.0003)</td>
</tr>
</tbody>
</table>

### MSE for Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CN 5%</th>
<th>Robust</th>
<th>CN 25%</th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td></td>
<td>MLE</td>
<td></td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0028 (0.0001)</td>
<td>0.0035 (0.0002)</td>
<td>0.0054 (0.0006)</td>
<td>0.0034 (0.0002)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.4527 (0.0415)</td>
<td>0.1246 (0.0074)</td>
<td>0.2231 (0.0272)</td>
<td>0.1744 (0.0190)</td>
</tr>
<tr>
<td>$a_0$</td>
<td>0.0444 (0.0034)</td>
<td>0.0207 (0.0012)</td>
<td>0.0402 (0.0042)</td>
<td>0.0347 (0.0024)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0307 (0.0020)</td>
<td>0.0100 (0.0004)</td>
<td>0.0484 (0.0047)</td>
<td>0.0161 (0.0006)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0084 (0.0007)</td>
<td>0.0027 (0.0001)</td>
<td>0.0663 (0.0007)</td>
<td>0.0038 (0.0002)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.0160 (0.0014)</td>
<td>0.0039 (0.0002)</td>
<td>0.0163 (0.0021)</td>
<td>0.0070 (0.0003)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0252 (0.0013)</td>
<td>0.0278 (0.0021)</td>
<td>0.0356 (0.0032)</td>
<td>0.0363 (0.0041)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0206 (0.0017)</td>
<td>0.0094 (0.0007)</td>
<td>0.0122 (0.0015)</td>
<td>0.0170 (0.0023)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.0195 (0.0013)</td>
<td>0.0126 (0.0010)</td>
<td>0.0198 (0.0023)</td>
<td>0.0260 (0.0025)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0067 (0.0003)</td>
<td>0.0151 (0.0003)</td>
<td>0.0081 (0.0002)</td>
<td>0.0035 (0.0002)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0156 (0.0004)</td>
<td>0.0126 (0.0007)</td>
<td>0.0195 (0.0004)</td>
<td>0.0021 (0.0002)</td>
</tr>
</tbody>
</table>
Table A.5: Complete Bias Results Across Generating Processes with Second Portfolio

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Normal</th>
<th>$t_6$</th>
<th>$t_3$</th>
<th>CN 5%</th>
<th>CN 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
</tr>
<tr>
<td>$a_{0,1}$</td>
<td>-0.0279</td>
<td>-0.0009</td>
<td>-0.0163</td>
<td>-0.0074</td>
<td>-0.0002</td>
</tr>
<tr>
<td>$a_{0,2}$</td>
<td>-0.0606</td>
<td>-0.0380</td>
<td>0.0174</td>
<td>-0.0174</td>
<td>0.1501</td>
</tr>
<tr>
<td>$a_{0,3}$</td>
<td>-0.0131</td>
<td>0.0075</td>
<td>0.0079</td>
<td>-0.0354</td>
<td>0.0639</td>
</tr>
<tr>
<td>$a_{1,1}$</td>
<td>0.0570</td>
<td>-0.0322</td>
<td>0.0716</td>
<td>-0.0609</td>
<td>0.0930</td>
</tr>
<tr>
<td>$a_{1,2}$</td>
<td>0.0204</td>
<td>-0.0273</td>
<td>0.0348</td>
<td>-0.0360</td>
<td>0.0574</td>
</tr>
<tr>
<td>$a_{1,3}$</td>
<td>0.0193</td>
<td>-0.0331</td>
<td>0.0376</td>
<td>-0.0449</td>
<td>0.0941</td>
</tr>
<tr>
<td>$b_{1,1}$</td>
<td>0.0300</td>
<td>-0.0518</td>
<td>0.0099</td>
<td>-0.0747</td>
<td>-0.0164</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>-0.0013</td>
<td>-0.0136</td>
<td>-0.0225</td>
<td>-0.0319</td>
<td>-0.0578</td>
</tr>
<tr>
<td>$b_{1,3}$</td>
<td>-0.0032</td>
<td>-0.0335</td>
<td>-0.0155</td>
<td>-0.0255</td>
<td>-0.0756</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0582</td>
<td>0.1277</td>
<td>-0.0408</td>
<td>0.0948</td>
<td>-0.1386</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.0130</td>
<td>-0.1557</td>
<td>0.0904</td>
<td>-0.0924</td>
<td>0.1779</td>
</tr>
<tr>
<td>Parameter</td>
<td>Normal</td>
<td>$t_0$</td>
<td>$t_3$</td>
<td>CN 5%</td>
<td>CN 25%</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
<td>Robust</td>
<td>MLE</td>
</tr>
<tr>
<td>$a_{0,1}$</td>
<td>0.0019</td>
<td>0.0038</td>
<td>0.0030</td>
<td>0.0040</td>
<td>0.0165</td>
</tr>
<tr>
<td>$a_{0,2}$</td>
<td>0.1418</td>
<td>0.1125</td>
<td>0.1863</td>
<td>0.1410</td>
<td>0.5667</td>
</tr>
<tr>
<td>$a_{0,3}$</td>
<td>0.0261</td>
<td>0.0325</td>
<td>0.0475</td>
<td>0.0182</td>
<td>0.1259</td>
</tr>
<tr>
<td>$a_{1,1}$</td>
<td>0.0109</td>
<td>0.0174</td>
<td>0.0192</td>
<td>0.0035</td>
<td>0.0518</td>
</tr>
<tr>
<td>$a_{1,2}$</td>
<td>0.0023</td>
<td>0.0018</td>
<td>0.0083</td>
<td>0.0008</td>
<td>0.0245</td>
</tr>
<tr>
<td>$a_{1,3}$</td>
<td>0.0048</td>
<td>0.0028</td>
<td>0.0087</td>
<td>0.0016</td>
<td>0.0450</td>
</tr>
<tr>
<td>$b_{1,1}$</td>
<td>0.0135</td>
<td>0.0271</td>
<td>0.0200</td>
<td>0.0308</td>
<td>0.0511</td>
</tr>
<tr>
<td>$b_{1,2}$</td>
<td>0.0052</td>
<td>0.0132</td>
<td>0.0116</td>
<td>0.0106</td>
<td>0.0301</td>
</tr>
<tr>
<td>$b_{1,3}$</td>
<td>0.0081</td>
<td>0.0168</td>
<td>0.0137</td>
<td>0.0106</td>
<td>0.0501</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0018</td>
<td>0.0001</td>
<td>0.0021</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.0011</td>
<td>0.0401</td>
<td>0.0019</td>
<td>0.0041</td>
<td>0.0024</td>
</tr>
</tbody>
</table>
B.1 Densities of Squared Radii

The distribution of the squared radii of the residual points helps to determine the multivariate distribution of the residuals. Figure B.1(a) is the sample estimated distribution function of the radii. Figure B.1(b) is the sample estimated distribution function of the squared radii.

B.2 Volatility Estimation of Foreign Exchange Rates

Figures B.2, B.3, B.4, B.5, and B.6 compare the average estimated univariate volatilities across the assets in the set with both the maximum likelihood estimation and robust estimation.

B.3 Correlation Estimation of Foreign Exchange Rates

Figures B.7, B.8, B.9, B.10, and B.11 compare the estimated correlations between each pair of assets in the set with both the maximum likelihood estimation and robust estimation.
Figure B.1: Density Plots of Radii (a) and Squared Radii (b)
Figure B.2: Estimated Volatility Time Plots of Euro (MLE [a], Robust [b])
Figure B.3: Estimated Volatility Time Plots of Pound (MLE [a], Robust [b])
Figure B.4: Estimated Volatility Time Plots of Franc (MLE [a], Robust [b])
Figure B.5: Estimated Volatility Time Plots of Yen (MLE [a], Robust [b])
Figure B.6: Estimated Volatility Time Plots of Dollar (MLE [a], Robust [b])
Figure B.7: Estimated Correlation Time Plots
Figure B.8: Estimated Correlation Time Plots
Figure B.9: Estimated Correlation Time Plots
Figure B.10: Estimated Correlation Time Plots
Figure B.11: Estimated Correlation Time Plots