A Multivariate GARCH Model for the Non-Normal Behaviour of Financial Assets

Juan Pablo Cajigas

Thesis submitted for the Degree of Doctor of Philosophy

Faculty of Finance
Cass Business School, City University

February 2007

Contents

List of Tables

Chapter 2

Chapter 4

Table 4.2b. List of univariate-GARCH specifications selected for each asset (bonds)

Table 4.3a. P-values for the Value-at-Risk backtesting results

Table 4.3b. P-values for the Value-at-Risk backtesting results (unconditional tests)

Table 4.3c. P-values for the Value-at-Risk backtesting results (independent tests)

Table 4.4. P-values for the expected shortfall (ES) backtesting results (conditional tests)

List of Figures

Chapter 1

Chapter 2

Figure 2.8B. Series of the variances of the portfolios composed of all assets in the sample data for the several DCC models

Figure 2.9. Distribution of the dynamic correlation between the return of UK shares and US shares employing the AGDCC, DCC, GDCC, and ADCC models with AML distributions.

Figure 2.10. Distribution of the dynamic correlation between the return of Japan bonds and UK bonds employing the AGDCC, DCC, GDCC, and ADCC models with AML distributions.

Figure 2.11. Distribution of the dynamic correlation between the return of UK shares and Mexico shares employing the AGDCC, DCC, GDCC, and ADCC models with AML distributions.

Figure 2.12. Distribution of the dynamic correlation between the return of UK bonds and Switzerland shares employing the AGDCC, DCC, GDCC, and ADCC models with AML distributions.

Figure 2.13.Plot of the correlation series between the returns of "Japan shares" and "U.K. shares" estimated with the DCC (1,1), ADCC(1,1,1) and GDCC(1,1) models and with the AML distribution.

Figure 2.14. Plot of the correlation series between the returns of "Switzerland bonds" and "U.K. bonds" estimated with the DCC (1,1), ADCC(1,1,1) and GDCC(1,1) models and with the AML distribution.

Figure 2.15. Plot of the correlation series between the returns of "U.S. shares" and "U.K. shares" estimated with the DCC (1,1), ADCC(1,1,1) and GDCC(1,1) models and with the AML distribution.

Figure 2.16. Plot of the correlation series between the returns of "Mexico shares" and "U.K. shares" estimated with the DCC (1,1), ADCC(1,1,1) and GDCC(1,1) models and with the AML distribution.

Chapter 4

Figure 4.1. Value-at-Risk from two different probability distributions. The confidence level is 0.5%

Figure 4.2. Loss estimates of quantile 99 as a function of exceedences over the threshold

Figure 4.3. Estimates of shape parameter ξ of the Generalised Pareto distribution at different number of exceedences

Figure 4.4. Tail estimate for the filtered data over the 70 exceedences threshold in Generalised Pareto distribution.

Figure 4.5. QQ-plot of residuals from the GPD fit to the filtered data over the selected threshold

Acknowledgements

First and foremost I would like to gratefully acknowledge the enthusiastic and excellent supervision of Giovani Urga during the development of this work. His great dedication, constant support and acute advice made possible the completion of this project. I am also thankful to Lorenzo Trapani and Chiwa Kao for their guidance all along the research process that had as result this thesis.

I received useful comments from participants in the international conference "Common Features in London" (Cass Business School, London, 16-17 December 2004), the "1st Italian Congress of Econometrics and Empirical Economics" (Universita' Ca' Foscari, Venice, 24-25 January 2005), the "Simulation Based and Finite Sample Inference in Finance II" Conference (Québec City, 29-30 April 2005), the ESRC Econometric Study Group Conference (Bristol, 14-16 July, 2005), the CEA@Cass/ESRC seminar on "The Use of Copulas and DCC Models to Measure Dependence in Finance" (Cass Business School, London, 17 March 2006), the Finance Seminar (Tilburg, 20 March 2006), the CEA@Cass/ESRC seminar on "Using Thick Distributions, DCC and Copula Models (Cass Business School, London,19 May 2006). I greatly benefited from stimulating discussions with and comments from Valentina Corradi, Tongshu Ma, Javier Mencia, Andrew Patton, Enrique Sentana. Kevin Sheppard and Lorenzo Cappiello provided the data set used in this dissertation and useful insights to estimate the various models. The usual disclaimer applies.

Last but not least, I express my sincere thanks to my wife Carolina for her support and immense understanding during these years.

Declaration

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement

Abstract

This thesis extends the dynamic conditional correlation (DCC) model proposed in Engle (2002) to the case of conditional returns supposed to follow an asymmetric multivariate Laplace (AML) distribution as presented in Kotz, Kozubowsky and Podgorski (2003). We prove that maximum likelihood estimator provides optimal estimates of the relevant parameters estimated. We show the applicability of our approach in a comprehensive set of risk management implementations where we compute Value-at-Risk and Expected-Shorfall measures for portfolios composed by a large number of assets.

Introduction

Financial econometrics deals with the application of econometric tools to financial data (Bollerslev, 2001; Engle, 2001). We may identify two milestones in the development of financial econometrics: (a) the formulation of time-varying volatility models in the form of ARCH (Engle, 1982, for which Engle got the Nobel Prize in Economics in 2003 and (b) the development of robust methods-of-moments-based estimation procedures. We now understand the salient distributional features of daily and lower frequency speculative market (Forex, stock, interest) returns, and the main empirical findings are that there exist such as (a) strong persistent volatility dependencies, (b) spillover and linkages across different assets and markets; (c) asymmetry and leverage effects in both volatilities and correlations.

Still important contributions are expected from questions related to: (a) models for ultra high-frequency data; (b) longer-run dependencies and large dimensional systems.

This dissertation focuses on the latter issue, how to estimate and test dependence between multiple financial time series.

A strand of the literature has focused on studying interdependence using conditional correlations. These include Kaplanis (1988), Bertero and Mayer (1990), King and Wadhwani (1990), Koch and Koch (1991), Ratner (1992), Longin and Solnik (1995), Ang and Chen (2002), Dumas, Harvey and Ruiz (2003).

With respect to the modelling of contagion, defined in their work as the increase in correlation between two variables during a crisis period, Forbes and Rigobon (2002) propose a test where the correlation between two asset returns during the crisis period is adjusted to account for the fact that correlations

are positive function of volatility. Rigobon (2003) suggests an alternative multivariate test for contagion based on the covariance matrices across two samples. Butler and Joaquin (2002) test for asymmetries in correlation across bull and bear markets. In addition to asymmetries in the level of the shocks, it may be the asymmetric volatility of the shocks that matters. Bekaert, Harvey and Ng (2003) incorporate asymmetric GARCH in the idiosyncratic or country-specific shocks of the latent factor model. All these studies have main limitations. First, the pairwise correlation coefficient is assumed constant over time. Second, the studies fail to propose any forecasting method with predictive power.

Some papers have attempted to model correlations dynamically. Hartmann, Straetmans and de Vries (2001) and Chan-Lau, Mathieson and Yao (2004) focus on the application of "extreme value theory" given that it is in periods of extreme negative returns that the question of cross-market correlation is most relevant. Li (2002) employs an asset-pricing model to demonstrate that the correlation of stock and bond returns can be explained by common exposure to macroeconomic factors. Ball and Torous (2000) assume correlation as a latent variable and use filtering methods to extract stochastic correlation from return data. Campbell, Koedijc and Kofman (2000), using VaR methodology, extract the quantile correlation structure implicit in asset returns. Jacquier and Marcus (2001) model correlations as resulting from the common dependence of returns on a market wide factor (base on CAPM). Changing correlations have also been studied under the framework of ultra-high frequency data (tick by tick). Barndorff-Nielsen and Shephard (2004) provide an asymptotic distribution theory for standard methods including high frequency realized correlations, though in a univariate framework.

There is more progress in the research on time-varying volatility with the introduction of multivariate ARCH-GARCH models: Bollerslev, Engle, and Wooldridge (1988), Bollerslev (1990), Engle, Ng, and Rothschild (1990), Engle and Kroner (1995), Kroner and Ng (1998), Baur (2002). To overcome the complexity of the estimation procedures strong restrictions on the parameters are imposed while the positive definiteness of covariance matrix (measuring financial risk) is not always guaranteed. However, all models continue to assume constant conditional correlations among assets.

A recent development in modelling correlation is the Dynamic Conditional Correlation model (DCC) proposed by Engle (2002), extended by Engle and Sheppard (2002) and Cappiello, Engle and Sheppard (2004). This multivariate model parameterises the conditional correlation directly in a similar way as GARCH models parameterise conditional volatility. It considers time-varying correlations among assets and uses a two step estimation procedure that reduces the number of parameters to calculate. Engle's procedure is a practical tool to model correlations in an empirical framework: first, it allows to estimate high dimensional systems in a short period of time and without demanding excessive computational capability and, secondly, guarantees positive parameter estimates for the variance-covariance matrix in all the estimation process. Correlation is conditionally known and forecasts can be generated.

However, some important issues are at present under investigation. The DCC framework is based on the assumption of normality. Normality is not a satisfactory property for financial time series. But normality-MLE/QMLE provides a feasible and consistent but inefficient DCC coefficients (Bollerslev and Wooldridge, 1992). Current routes of investigation to overcome the limits in the assumption of normality are mainly two:

1. Semi-parametric methods of investigations (Engle and González-Rivera (1991), Drost and Klaassen (1997), González-Rivera (1997), González-Rivera and Drost (1999), Hafner and Rombouts (2004);

2. Use of thick (non normal) distributions to achieve efficiency with implication for the first stage (Bollerslev (1987), Baillie and Bollerslev (1989), Nelson (1991), Fiorentini, Sentana, Calzolari (2003))

Contributions and originality of the dissertation

In this dissertation, by improving the econometric quality of parameters in the DCC model of Engle (2002), we develop a methodology that is able to model the variances and covariances of dozens of time series in a way that captures more closely the real distributional features of the data. The big improvement of our method in relation to the work of Engle (2002) is that the estimation of

parameters is performed without mis-specifying the distribution of the data. By doing this, and as we show in the development of the dissertation, we enhanced the performance of Engle's model when this is used in financial applications such as Value-at-Risk.

The main vehicle used in the enhancement of Enlgle's model is the Asymmetric Multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003) . We employed this distribution, instead of the normal distribution used by Engle (2002), for the estimation of parameters by the method of maximum-likelihood-estimation (MLE). As far as we are concerned this is the first time the AML distribution is employed in multivariate GARCH applications. Another work that is original of this dissertation, and that is related to asymptotic theory as studied in econometrics, is the analysis of the implications of the estimation of the DCC model by MLE under the AML assumption: we evaluate whether the use of MLE with innovations distributed as a AML has implications for consistency of estimates of conditional variances and correlations. Finally, we report a comprehensive empirical exercise using a sample of 21 FTSE All-World stock indices and 13 bond return indices to show the empirical relevance of our model for risk management purposes.

To summarize, this dissertation presents the following four contributions to the financial econometrics literature:

1. A feasible framework able to capture not only the dynamic behaviour of volatilities and covariances, but also possible time-dependent correlations between assets.

2. The proposed framework addresses the problem of large dimension ("dimensionality curse") of the parameter space in multivariate models of conditional second moment.

3. We prove that MLE provides optimal (consistency and efficiency) estimates of dynamic conditional correlations under the AML assumption of innovations.

4. We prove that AML is an appropriate distribution which well characterises the behaviour of financial returns.

The first contribution is important because empirical work (Cappiello *et al*, 2004) has shown that dependence between financial assets is not static but dynamic. Point 2 is also relevant because as we show in Chapter 1, previous specifications (like some multivariate GARCH models for example) although theoretically correct can not be implemented because the "curse of dimensionality" in the presence of reasonable number of assets makes then difficult to apply. Contribution 3 is important because we validate the theoretical use of the estimation methodology under the AML distribution. In our method, we employ Maximum Likelihood Estimation (MLE) despite the relaxation of the normality assumption. This technique provides consistent and efficient estimates under some conditions. Our aim is to develop a model where these conditions are mild and easy to verify. Finally, given that the returns of financial assets are random variables these must be described in terms of their distributional properties. For practical purposes the most relevant of these are the first four moments. Nonetheless, practitioners and academics usually use the paradigm of the normal distribution which can only describe the first two. One of our aims is to produce a model that is able to describe more closely the real distribution of financial returns.

Structure of the dissertation

In what follows, we provide a short description of the individual chapters.

Chapter 1: We provide a brief description of the main ways to measure dependence in finance and we then give a short survey of multivariate heteroskedastic time series models. We introduce Engle's DCC model and explain that although it represents an important advance in solving the "dimensionality curse" in MGARCH models, it is quiet limited as it assumes an unrealistic distribution for standardised residuals.

Chapter 2. We present an extension of the DCC model by allowing for an Asymmetric Multivariate Laplace (AML) distribution for the standardised residuals. We give details regarding the nature of this distribution and explain why it is an appropriate choice. We present MLE estimates of alternative models using the same data set in the seminal paper by Cappiello et (2003) .

Chapter 3: We provide a rigorous analysis to prove that MLE provides consistency of the DCC parameter estimates under AML distribution of innovations. This implies to prove strict stationarity and thus consistency under some regularity conditions.

Chapter 4: We implement our preferred model in a series of risk management applications. Using the data set of 34 financial assets, we compute the conditional VaR and ES using several variants of the DCC model. We also show how it is possible to combine our approach with Extreme-Value-Theory to enhance the quality of risk management measures.

Chapter 1

Dependence in financial time series

1.1 Introduction

The main objective of this chapter is to provide a brief description of the main ways to measure dependence in finance. We then give a short survey of multivariate heteroskedastic time series models and we introduce Engle's DCC model and explain that although it represents an important advance in solving the "dimensionality curse" in MGARCH models, however it is quiet limited as it assumes an unrealistic distribution for standardised residuals. Finally, we report a short description of test statistics used to test for constant correlations and mispecification.

1.2 Main measures of dependence in finance

1.2.1 Linear correlations

Definition 1.1 *Consider the two random variables* $\mathbf{x}_i$ *and* $\mathbf{x}_j$*. The correlation coefficient between the variables* $\mathbf{x}_i$ *and* $\mathbf{x}_j$*, when* $E(\mathbf{x}_i\mathbf{x}_j) < \infty, E(\mathbf{x}_i) < \infty$*, and* $E(\mathbf{x}_j) < \infty$ *is given by*

$$\rho(\mathbf{x}_i, \mathbf{x}_j) = \frac{Cov(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{\sigma^2(\mathbf{x}_i)\sigma^2(\mathbf{x}_j)}}, i \neq j \tag{1.1}$$

where $Cov(\mathbf{x}_i, \mathbf{x}_j)$ is the covariance between $\mathbf{x}_i$ and $\mathbf{x}_j$ defined as $Cov(\mathbf{x}_i, \mathbf{x}_j) = E(\mathbf{x}_i\mathbf{x}_j) - E(\mathbf{x}_i)E(\mathbf{x}_j)$ and $\sigma^2(\mathbf{x}_i) < \infty, \sigma^2(\mathbf{x}_j) < \infty$ denote the variances of x_i and $\mathbf{x}_j$.

Properties

- *Linearity.* Correlation is a measure of linear dependence. The connection between correlation and simple linear regression can be seen in the coefficient of determination

$$R^2 = \rho(\mathbf{x}_i, \mathbf{x}_j)^2 = \frac{\sigma^2(\mathbf{x}_j) - \min_{\alpha,\beta} E[(\mathbf{x}_j - (\alpha + \beta\mathbf{x}_i)]}{\sigma^2(\mathbf{x}_j)} \tag{1.2}$$

- *Invariance under strictly increasing linear transformations.* Correlation is invariant only under positive affine transformations. This property derives from the previous one. Formally

$$\rho(\alpha + \beta\mathbf{x}_i, \delta + \gamma\mathbf{x}_j) = sgn(\beta \cdot \gamma)\rho(\mathbf{x}_i, \mathbf{x}_j) \tag{1.3}$$

- *Functional under Elliptical distributions.* Correlation is a valid measure of dependence only when $\mathbf{X}$ is multivariate elliptical distributed. This property has important implications in the use of correlations as measures of dependence in financial applications.

Definition 1.2 *The N-dimensional random vector $\mathbf{X} = (\mathbf{x}_1, ...\mathbf{x}_N)^T$ has a multivariate elliptical distribution, i.e. $\mathbf{X} \sim \mathbf{E}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, if its characteristic function can be expressed as*

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\left(it^T\mu\right)\psi\left(\frac{1}{2}\mathbf{t}^T\mathbf{H}\mathbf{t}\right) \tag{1.4}$$

for some column-vector $\boldsymbol{\mu}$, and a $N \times N$ positive-definite matrix denoted by $\mathbf{H}$. If the density exists it has the form

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{c_N}{\sqrt{|\mathbf{H}|}} g_N\left[\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T\mathbf{H}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right] \tag{1.5}$$

for some function $g_N(\cdot)$ called the density generator. The normalizing constant c_N can be explicitly expressed as

$$c_N = \frac{\Gamma(N/2)}{(2\pi)^{N/2}} \left[\int_0^\infty x^{N/2-1} g_N(x) dx \right]^{-1} \tag{1.6}$$

Thus, the condition $\int_0^\infty x^{N/2-1} g_N(x) dx < \infty$ guarantees $g_N(\cdot)$ as a density generator.

Some examples of bivariate elliptical distributions are given in Figure 1.1.

[Insert Figure 1.1. here]

1.2.2 Copulas

Instead of measuring the dependence between random variables by the ratio of covariances and variances (second moments), a copula extracts the dependence structure from the entire joint distribution. Schweizer and Sklar (1983), show that a joint distribution of the variables $\mathbf{x}_1, ..., \mathbf{x}_N$ can be written as

$$F(\mathbf{x}_1, ..., \mathbf{x}_N) = C(F_1(\mathbf{x}_1), ..., F_N(\mathbf{x}_N)) \tag{1.7}$$

i.e. as a C function (copula) of the marginal distributions of each one of the random variables in $\mathbf{X}$. This function C, can be seen as a multivariate distribution function with standard uniform marginal distributions. The following definition formalises this concept.

Definition 1.3 *A copula is any function $C : [0,1]^N \longrightarrow [0,1]$ which has the properties:*

1. *$C(\mathbf{x}_1, ..., \mathbf{x}_N)$ is increasing in each component $\mathbf{x}_i$.*
2. *$C(1, ..., 1, \mathbf{x}_i, 1, ..., 1) = \mathbf{x}_i$ for all $i \in \{1, ..., N\}, \mathbf{x}_i \in [0,1]$.*
3. *For all $(a_1, .., a_N), (b_1, .., b_N) \in [0,1]^N$ with $a_i \leq b_i$ we have:*

$$\sum_{i_1=1}^{2} ... \sum_{i_N=1}^{2} (-1)^{i_1+...+i_N} C(x_{1i_1}, ..., x_{Ni_N}) \geq 0 \tag{1.8}$$

where $x_{j1} = a_j$ and $x_{j2} = b_j$ for all $j \in \{1, ..., N\}$.

The main appeal of this kind of dependence measure is its invariance under increasing and continuous transformations of the marginal distributions. This is particularly useful when we are working with non-linear transformations of the individual processes and correlation coefficients can not be used. To see this, assume $y_1 = \log(x_1)$ and $y_2 = \log(x_2)$ and $y_1 \sim N(0,1)$ and $y_2 \sim N(0,4)$. Embrechts *et al* (1999) show that the attainable interval of correlations under lognormal distributions and listed parameters is $[-0.090, 0.666]$. In this case, is not possible to assign a value of 1 when we have perfect positive dependence or -1 when we have perfect negative dependence. In the other hand, if we use for example the Gumbel copula (Joe (1997))

$$C_\beta(u,v) = \exp\left[-\{(-\log u)^{1/\beta} + (-\log v)^{1/\beta}\}^\beta\right], 0 < \beta \leq 1 \tag{1.9}$$

we can obtain a bivariate distribution $F(x_1, x_2) = C_\beta(F_1(x_1), F_2(x_2))$, such that a correlation coefficient of let say 0.70 between $\mathbf{x}_1$and $\mathbf{x}_2$ is attainable.

1.2.3 Rank correlations

Definition 1.4 *Let* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *be random variables with distribution functions* F_1 *and* F_2 *and joint distribution function* F*. Spearman's rank correlation is given by,*

$$\rho_S(\mathbf{x}_1, \mathbf{x}_2) = \rho(F_1(\mathbf{x}_1), F_2(\mathbf{x}_2)) \tag{1.10}$$

where ρ *is the usual linear correlation. Basically,* $\rho_S(\cdot)$ *defines a linear correlation of probability-transformed variables. It can also be seen as the correlation of the copula* C *associated with* $(\mathbf{x}_1, \mathbf{x}_2)^t$*.*

As copulas, rank correlations have the advantage of invariance under transformations of risks, and attainability of the entire interval $[-1, 1]$. Unfortunately, can not be manipulated in the variance-covariance framework as linear correlations because are not moment-based. This precludes their use for many financial applications.

1.2.4 Time variation and breakdown

One area that has received a lot of attention in the last years in the empirical study of the properties of correlations is the one related to "Correlation Breakdown" or "Structural Change in Correlations". The reason comes from the fact that given that correlations are key parameters in the implementation of risk management techniques like VaR, a drastic change in the measure of correlation can alter risk calculations. It is clear that in any analysis of methodologies related to the forecast of correlations the issue of structural breaks has to be fully studied.

The term "correlation breakdown " refers to the situation where the empirical regularity of the measure of co-movement between series change drastically between periods of time. Papers related to the stability of correlation matrices include Kaplanis (1988), Ratner (1992), Koch and Koch (1991), Von Furstenberg and Jeon (1989), King and Wadhwani (1990), Bertero and Mayer (1990), Longin and Solnik (1995), Karolyi and Stulz (1995), Goetzman *et al* (2001) and Forbes and Rigobon (2002) among many others.

Many of the literature cited above, connects correlation breakdowns to extreme changes in volatility. This issue is crucial in portfolio diversification and hedging . If diversification has as goal the reduction of the overall risk vis-a-vis a reduced correlation, the effect might be neutralized at moments of high volatility when is more desirable. A group of recent papers (Ronn (1998), Boyer *et al* (1999), Loretan and English (2000), Forbes and Rigobon (2002)), however, claim that shifts in correlation are in many occasions only the reflex of heteroskedasticity in market returns and not a real change in the data generating process driving the co-movement between assets. To prove this, Boyer *et al* derive the analytical relationship between conditional and unconditional correlations. Suppose x and y are two bivariate normal random variables. Consider the event $x \in F$ where $F \subset \mathbf{R}$ such that $0 < P(F) < 1$, then

$$\rho_F = \frac{\rho}{\sqrt{\rho^2 + (1-\rho^2)Var(x)/Var(x \mid x \in F)}} \tag{1.11}$$

where the unconditional correlation is given by $\rho = \sigma_{xy}/(\sigma_x \sigma_y)$ and ρ_F is

the correlation conditioned on the event $x \in F$. The Theorem holds not only when x and y are bivariate normal but also in any situation when y can be stated as an affine function of x and an independent error term. It is clear from the equation that the conditional correlation is directly proportional to the conditional variance and indirectly proportional to the unconditional variance, i.e. the conditional correlation will augment if given an event ($x \in F$), the conditional variance augments with respect to the unconditional variance. This explains why after an extreme event like for example a financial crisis, the conditional correlation changes while the "real" or unconditional correlation remains constant. Forbes and Rigobon (2002) derive the same relationship but with a quite different statistical framework. In their paper they call this effect "bias in correlation", in the sense that correlation coefficients are inaccurate due to time-varying sampling volatility.

An important theoretical consideration in equation (1.11) refers to the assumptions under which the equality holds. To see this consider the derivation of the equation by Forbes and Rigobon[1]. Assume that the returns x and y are related by the following linear representation,

$$y_t = \alpha + \beta x_t + \varepsilon_t \tag{1.12}$$

where $E(\varepsilon_t^2) = c < \infty$, $E(\varepsilon_t) = 0$, and $E(x_t \varepsilon_t) = 0$. The last two conditions refer to omission of variables and endogeneity between markets. Define β as the beta when $x \notin F$ and β_F as the beta when $x \in F$. Since residuals and regressors are independent, i.e. $E(x_t \varepsilon_t) = 0$, OLS estimates are consistent for both groups and $\beta = \beta_F$. If we consider $\{x \in F\}$ an extreme event that produces major variation, then,

$$\beta = \frac{Cov(x, y)}{Var(x)} = \frac{Cov(x, y \mid x \in F)}{Var(x \mid x \in F)} = \beta_F \tag{1.13}$$

and

[1]In Forbes and Rigobon (2003) equation (3.28) is given by $\rho_F = \rho\sqrt{\frac{1+\delta}{1+\delta\rho^2}}$ where δ is the relative increase in the variance of x, i.e. $\delta \equiv \frac{Var(x|x \in F) - Var(x)}{Var(x)}$

$$Cov(x, y) < Cov(x, y \mid x \in F) \tag{1.14}$$

$$Corr(x, y) < Corr(x, y \mid x \in F) \tag{1.15}$$

We see how the correlation between x and y conditioned to $x \in F$ is greater than the unconditional correlation. The assumptions under which the correlation bias is found ($E(\varepsilon_t) = 0$, and $E(x_t\varepsilon_t) = 0$), are strong. The proof given by Forbes and Rigobon is valid only if there are no exogenous shocks and no feedback between x and y. They show that after correcting a correlation with equation (1.11), if the two assumptions are not fulfilled then the bias will still preserve.

1.3 MGARCH and constant conditional correlation models

1.3.1 BEKK and VECH models

Autoregressive Conditional Heteroskedasticity (ARCH(q)) models are one of the most widely used tools in the forecasting and modelling of volatility of returns in financial markets. In this class, the most popular model by far is the Generalised Autoregressive Conditional Heteroskedasticity (GARCH(p,q)) model. The success of the model resides in its parsimoniousness, in the fact that the variance is conditionally known allowing for relative straight forward estimation, and in that it captures many stylized facts found in the time series of financial assets like excess kurtosis and volatility clustering.

The first generalization to the multivariate case was made by Bollerslev *et al* (1988) (VECH model), where the vector x_t is replaced by the matrix $\mathbf{x}_t/\mathcal{F}_{t-1} \sim N(0, \mathbf{H}_t)$, where $\mathcal{F}_{t-1}$ is the filtration up to time $t-1$. The dynamics

of the conditional variance are replaced by the *vech*[2] specification

$$vech(\mathbf{H}_t) = \zeta + \sum_{j=1}^{q} \mathbf{\Delta}_j vech(x_{t-j}x'_{t-j}) + \sum_{j=1}^{p} \mathbf{B}_j vech(\mathbf{H}_{t-j}) \tag{1.16}$$

where $\mathbf{H}_t$ is the variance-covariance matrix of errors, ζ is a parameter vector, and $\mathbf{\Delta}_j$ and $\mathbf{B}_j$ are square matrices. The model is far from parsimonious, the number of parameters explodes with only a few number of assets and few lags. The reason is that each conditional variance is not only function of its own lags but also of lags of all the conditional variances of other series and all the possible cross products between series (covariances). To make possible a practical implementation, Bollerslev *et al* (1988) present in the same paper a more parsimonious representation called diagonal-vech (DVECH). In this version the matrices $\mathbf{\Delta}_j$ and $\mathbf{B}_j$ are diagonal instead of full matrices, making the model not only more tractable, but also more congruent with the nature of the univariate GARCH models, in the sense that the variances and covariances of each series depend only on own past squared errors and on own past cross-products of errors.

Although the DVECH model reduces the number of parameters to estimate, to secure $\mathbf{H}_t$ as a positive-definite matrix in the estimation process is still difficult. The reason is that given that the model estimates the covariance as a geometrically declining weighted average of past cross products of unexpected returns, without suitable restrictions the rates at which the weights are reduced for older observations might cause that the off-diagonal terms of the estimated covariance matrix became too big relative to the diagonal terms, producing a negative definite matrix. The BEKK model developed by Engle and Kroner (1995) solves this problem by imposing some restrictions directly on the variance parameterisation

$$\mathbf{H}_t = \zeta'\zeta + \sum_{i=1}^{s}\sum_{j=1}^{q} \mathbf{\Delta}'_{ij} x_{t-j}x'_{t-j}\mathbf{\Delta}_{ij} + \sum_{i=1}^{s}\sum_{j=1}^{p} \mathbf{B}'_{ij}\mathbf{H}_{t-j}\mathbf{B}_{ij} \tag{1.17}$$

[2]The *vech* operator stacks the lower triangular portion of a matrix into a vector. If a matrix has AxA dimensions, then the *vech* vector will have $\frac{A^2+A}{2} \times 1$ dimensions.

where $\boldsymbol{\zeta}$ is an upper triangular matrix and s determines the generality of the process. Usually, empirical applications (Bera *et al* (1997), Hafner and Herwartz (1998), and Kearney and Patton (2000)) restrict the value of s to one. The contribution of the model to the literature, resides in that it secures positive-definetness under weak conditions by specifying the constant, autoregressive, and persistent coefficients as products of triangular and square matrices. The parameterization is also very general, Engle and Kroner show that it includes all positive-definite DVECH models and almost all positive-definite VECH models.

1.3.2 Factor models

Another parameterization of MGARCH commonly found in the literature is the FGARCH (Factor-GARCH) model of Engle, Ng, and Rothschild (1990). In this variation it is assumed that a factor or few factors are responsible for the co-movement of all series,

$$\mathbf{H}_t = \boldsymbol{\zeta}'\boldsymbol{\zeta} + \sum_{i=1}^{s} \lambda_i \lambda_i' \left(\sum_{j=1}^{q} \alpha_{ij}^2 w_i' x_{t-j} x_{t-j}' w_i + \sum_{j=1}^{p} \beta_{ij}^2 w_i' \mathbf{H}_{t-j} w_i \right) \tag{1.18}$$

where λ_i and w_i are vectors of dimensions Nx1, and α_{ij} and β_{ij} are scalars. The model can be understood as a special case of the BEKK model where the matrices $\boldsymbol{\Delta}_j$ and $\mathbf{B}_j$ in (1.17) are rank one and equal except for the parameters α_{ij} and β_{ij} which can be considered scale factors,

$$\boldsymbol{\Delta}_{ij}' = \alpha_{ij} w_i \lambda_i' \tag{1.19}$$

$$\mathbf{B}_j = \beta_{ij} w_i \lambda_i' \tag{1.20}$$

This of course alleviates the over-parameterization found in VEC and BEKK models. The main limitation of these models is that it is difficult to generalise to cases where assets have extra risk factors.

Alexander and Chibumba (1996) present the Orthogonal-GARCH (OGA-

RCH) model which can be considered a generalization of the FGARCH model where instead of one factor we have multiple orthogonal factors. The method is very simple in terms of computation. In a first stage the entire system of returns is divided into categories according to the type of asset or industry. The full covariance matrix of each one of the categories is generated with the main orthogonal factors obtained by a Principal Component Analysis (PCA). Next, in a second phase, the factor weights from the PCA are employed to merge together the large covariance matrix of the original structure. More formally, standardise the $\mathbf{x}_{kt}$ matrix (where $k = 1, ...K$ and K defines the number of categories) of returns with $n \leq N$ number of series in to a new matrix $\mathbf{X}_t$ with elements $y_{it} \equiv (x_{it} - \mu_{it})/\sigma_{it}$ where μ_{it} and σ_{it} are respectively the mean and standard deviation of x_{it}. Compute a matrix $\mathbf{W}$ of eigenvectors of $\mathbf{Y}'\mathbf{Y}/T$ and the associated diagonal matrix of eigenvalues $\mathbf{\Lambda}$ prearranged along with the size of each eigenvalue. The matrix of principal components of $\mathbf{y}_{kt}$ is given then by $\mathbf{P} = \mathbf{YW}$. The variance of each one of the principal components p_i, $i = 1, ...m \leq n$ is defined by a univariate GARCH process. The resulting variance-covariance matrix of returns for the category k is,

$$\mathbf{H}_{kt} = \mathbf{W}^*\mathbf{D}\mathbf{W}^{*\prime} \tag{1.21}$$

where $\mathbf{W}^*$ is a matrix of transformed eigenvectors composed of elements $w^*_{ij} = w_{ij}\sigma_i$, and $\mathbf{D}$ is a diagonal matrix composed of m univariate GARCH processes. In the simple case where the original system was subdivided into two categories the variance-covariance matrix of the entire system is,

$$\mathbf{H}_t = \begin{pmatrix} \mathbf{W}^*\mathbf{D}_1\mathbf{W}^{*\prime} & \mathbf{W}^*\mathbf{C}\mathbf{V}^{*\prime} \\ (\mathbf{W}^*\mathbf{C}\mathbf{V}^{*\prime})' & \mathbf{V}^*\mathbf{D}_2\mathbf{V}^{*\prime} \end{pmatrix} \tag{1.22}$$

where $\mathbf{C}$ is the variance matrix of principal components across the two categories and $\mathbf{V}^*$ is the matrix of transform eigenvectors of the second category computed in the same way as $\mathbf{W}^*$.

The key of the procedure resides in the orthogonality of the principal components in each category that implies that the covariance matrix is simply given by the diagonal matrix of variances. This greatly reduces the number

of parameters to estimate.

The matrix $\mathbf{H}_{kt}$ will be positive semi-definite but not strictly positive-definite, unless the number of principal components is the same as the number of series, i.e. if $m = n$. In the case where $m < n$ the variance-covariance matrix has to be run through an eigenvalue check to guarantee strict positive definiteness. Alexander (2001) shows that $\mathbf{H}_t$ is positive semi-definite only if

$$|corr\,(p_i, q_i)| \leq (ml)^{-1/2} \tag{1.23}$$

where q_i is the ith principal component in the $\mathbf{V}$ system and l is the number of principal components in the second category. This condition can be violated when the number of principal components is high.

The main drawback of the OGARCH estimator is that principal components are only unconditionally correlated, meaning that within a single category conditional correlations have to be zero. To solve this, Van der Weide (2002) introduces the Generalized Orthogonal GARCH model (GOGARCH), where instead of estimating univariate GARCH processes for the orthogonal principal components $\mathbf{P} = \mathbf{XW}$, a transformed principal component matrix $\mathbf{P}^* = \mathbf{XWU}_0$ is estimated, where $\mathbf{U}_0$ is an orthogonal matrix containing conditional information. The parameterisation of $\mathbf{U}_0$ defined as $\mathbf{U}$ is given by the product of $\binom{N}{2}$ rotation matrices

$$\prod_{i<j} G_{ij}(\theta_{ij}) \quad -\pi \leq \theta_{ij} \leq \pi \tag{1.24}$$

where θ_{ij} is an Euler angle that determines the degree of rotation $G_{ij}(\cdot)$ in the plane spanned by the ith and j^{th} vectors of the canonical basis of $\Re^N$. The estimation of θ_{ij} is done by maximum likelihood by augmenting the parameter space of the original OGARCH likelihood function by $\binom{N}{2}$ new parameters.

To investigate if the inclusion of conditional correlation effects in the principal components helps to better describe the linear co-movement of financial variables, van der Weide performs a comparison between the OGARCH and GOGARCH using DJIA and NASDAQ data. He computes a Likelihood Radio statistic to test the OGARCH against the GOGARCH, and rejects the null of orthogonal linkage at the 1% level for several lengths of the time series.

1.3.3 Constant conditional correlation (CCC) models

Bollerslev (1990) presents a specification where the variance $\mathbf{H}_t$ is parameterized directly as a function of dynamic univariate variances and a constant correlation matrix

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t \tag{1.25}$$

where $\mathbf{R} = [\rho_{ij}]$, $\rho_{ii} = 1$ for all i, and $\mathbf{D}_t$ is a diagonal matrix composed of variances h_t which are computed from an univariate GARCH process. It is a very parsimonious model with a simple estimation procedure compared to other representations. It eliminates all possibility of modelling the conditional correlation as a dynamic process, but it is a plausible representation when given empirical evidence the correlation can be considered constant (Section 1.4.2 in this Chapter discusses tests for constant correlations).

Kroner and Ng (1998) nest several of the existent models in the General Dynamic Covariance (GDC) Model. The appeal of this representation is that it facilitates model selection; they provide evidence that four MGARCH models (BEKK, VECH, FGARCH, and CCC) produce very different variance and covariance estimates for the same dataset. The variance-covariance matrix is specified by

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t + \mathbf{\Phi} \circ \mathbf{\Theta}_t \tag{1.26}$$

where $\mathbf{R} = [\rho_{ij}]$, $\rho_{ii} = 1$ for all i, $\circ$ denotes the Hadamard product, $\mathbf{\Phi} = [\phi_{ij}]$, $\phi_{ii} = 0$ for al i, $\mathbf{\Theta}_t = [h_{ijt}]$, and $\mathbf{D}_t$ is a diagonal matrix composed of variances h_t which are not computed from a univariate GARCH process like in the CCC model but from a BEKK (1,1,1) specification,

$$h_{ijt} = \zeta_{ij} + \alpha_i' x_{t-1} x_{t-1}' \alpha_j + \beta' \mathbf{H}_{t-1} \beta_j \tag{1.27}$$

The model is clearly an hybrid between the CCC and BEKK(1,1,1) models, and it encompasses the VECH, BEKK, FGARCH, and CCC specifications.

1.4 Dynamic conditional correlation models

1.4.1 Models

Engle (2002) and Engle and Sheppard (2001) developed the Dynamic Conditional Correlation (DCC) model, generalising the CCC model of Bollerslev (1990) by allowing for time varying correlations

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t \tag{1.28}$$

$$\mathbf{R}_t = (diag\mathbf{Q}_t)^{-\frac{1}{2}} \mathbf{Q}_t (diag\mathbf{Q}_t)^{-\frac{1}{2}} \tag{1.29}$$

$$\mathbf{Q}_t = \left(1 - \sum_{l=1}^{L} \alpha_l - \sum_{s=1}^{S} \beta_s\right) \mathbf{Q}^* + \sum_{l=1}^{L} \alpha_l u_{t-l} u'_{t-l} + \sum_{s=1}^{S} \beta_s \mathbf{Q}_{t-s} \tag{1.30}$$

where $u_{it} = x_{it}/\sqrt{h_{iit}}$, h_{ij} is the ij element of the variance matrix $\mathbf{H}_t$, Q^* is the NxN unconditional variance matrix of u_t, and α_l and β_s are scalar parameters satisfying $\sum_{l=1}^{L} \alpha_l + \sum_{s=1}^{S} \beta_s < 1$.

This version is an important step in the quest for a feasible empirical implementation of MGARCH models because it solves two major problems present in previous specifications: non-positive-definiteness, and computational hurdle. The first problem refers to the fact that given the complexity in the MLE process encountered in many other multivariate models (see section 1.3.1.), it is extremely difficult to guaranteed a positive definite variance-covariance matrix along all the estimation process. The second problem refers to the huge number of parameters that need to be estimated, when more than three or four assets are considered.

Another important consideration evaluating the performance of these models relates to the restrictions imposed to the dynamics of conditional variances, conditional covariances, and conditional correlations. In this respect DCC has also advantages. Given the two step estimation procedure implemented for the estimation of the parameters, the model allows for different persistence between variances and correlations. This is a desirable property. For instance,

although the OGARCH (1,1,N) model implies fewer parameters to estimate than the DCC (1,1), the former implies common persistence in all elements. This also applies for the VEC(1,1), BEKK(1,1,1), and FGARCH (1,1,1) models.

Nonetheless, the DCC model is not free from criticisms. Although it allows for different persistence between variances and correlations, it imposes common persistence among all correlations. This of course is unrealistic in situations where a big number of assets is being analyzed and different types, or at least different groups of persistence in correlation are expected to be found[3]. Some solutions to this hurdle are proposed in Hafner and Franses (2004), and Billio *et al* (2004).

Hafner and Franses (2004) propose a generalized DCC model (GDCC) in which equation (1.30) is modified to

$$\begin{aligned} Q_t &= \left(1-\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)^2-\frac{1}{N}\left(\sum_{i=1}^{N}\beta_i\right)^2\right)Q^*+\sum_{l=1}^{L}\boldsymbol{\alpha}_l\boldsymbol{\alpha}_l^{'}\circ u_{t-l}u_{t-l}^{'} \\ &+\sum_{s=1}^{S}\boldsymbol{\beta}_s\boldsymbol{\beta}_s^{'}\circ Q_{t-s} \end{aligned} \tag{1.31}$$

where α_l and β_s are Nx1 parameter vectors and $\circ$ denotes the Hadamard product. The cost of this gain in flexibility is an increase of $2(N-1)$ in the number of parameters to estimate. The authors argue that given that in many empirical applications the elements of the autoregressive vector β_s are very stable, a simplification of the representation can be made by assuming that as in the DCC the autoregressive coefficients are only scalars. In this case the extra-number of parameters reduces to $N-1$.

Billio *et al* (2004) propose a very similar specification (Block DCC), but to assuage the problem of excessive parameters they proposed the grouping of the N variables in to w sets, each one containing elements with common features that allows to model them with an homogeneous structure in correlation. The

[3]The DCC model imposes common persistence among all correlations because the parameters α_l and β_s in equation (4.3) are scalars

extra-number of parameters with respect to the DCC model is $2(w-1)$.

Another inconvenient of the DCC estimator is its non-linear evolution that makes impossible the computation of exact multi-step ahead conditional expectations. Engle and Sheppard (2001) propose two linear approximations. The first one is given by $E_t\left[\varepsilon_{t+i}\varepsilon'_{t+1}\right] \approx \mathbf{Q}_{t+i}$ for $i \in [1,...,r]$ where r is the number of forecasting steps. In this case the forecasting function is given by

$$\begin{aligned} E_t[\mathbf{Q}_{t+r}] &= \sum_{i=0}^{r-2}\left(1-\sum_{l=1}^{L}\alpha_l-\sum_{s=1}^{S}\beta_s\right)\mathbf{Q}^*\left(\sum_{l=1}^{L}\alpha_l+\sum_{s=1}^{S}\beta_s\right)^i \\ &+\left(\sum_{l=1}^{L}\alpha_l+\sum_{s=1}^{S}\beta_s\right)^{r-1}\mathbf{Q}_{t+1} \end{aligned} \tag{1.32}$$

and the forecast of the conditional correlation is given by

$$\mathbf{R}_{t+r} = (diag\mathbf{Q}_{t+r})^{-\frac{1}{2}}\mathbf{Q}_{t+r}(diag\mathbf{Q}_{t+r})^{-\frac{1}{2}} \tag{1.33}$$

.

The second approximation is give by $\mathbf{Q}^* \approx \mathbf{R}^*$ and $E_t\left[\mathbf{Q}_{t+1}\right] \approx E\left[\mathbf{R}_{t+1}\right]$. In this case we can forecast $\mathbf{R}_{t+r}$ directly by

$$\begin{aligned} E_t[\mathbf{R}_{t+r}] &= \sum_{i=0}^{r-2}\left(1-\sum_{l=1}^{L}\alpha_l-\sum_{s=1}^{S}\beta_s\right)\mathbf{R}^*\left(\sum_{l=1}^{L}\alpha_l+\sum_{s=1}^{S}\beta_s\right)^i \\ &+\left(\sum_{l=1}^{L}\alpha_l+\sum_{s=1}^{S}\beta_s\right)^{r-1}\mathbf{R}_{t+1} \end{aligned} \tag{1.34}$$

After some Monte Carlo experiments, Engle and Sheppard find that both approximations produce a very small bias towards 1 or -1, depending if the series has positive or negative unconditional correlation. In general, the method for solving $\mathbf{R}_{t+r}$ forward has better bias properties than solving forward for $\mathbf{Q}_{t+r}$.

Another parameterization under the philosophy of conditional correlation is presented by Audrino and Barone-Adesi (2003). They introduced two vari-

ants: the Rolling Window Average Conditional Correlation (RW-ACC) model and the Rolling Window Tree-Structured Average Conditional Correlation (RW-TACC) model.

In the case of the RW-ACC model the conditional correlation is given by,

$$\mathbf{R}_t = (1-\lambda)\overline{\mathbf{Q}}_{t-p}^{t-1} + \lambda\overline{\mathbf{R}}, \quad \lambda \in [0,1] \tag{1.35}$$

where $\overline{\mathbf{Q}}_{t-p}^{t-1}$ is the unconditional correlation for the last p days, and $\overline{\mathbf{R}}$ is a $(N \times N)$ matrix with ones on the diagonal and remaining elements given by $\overline{r_t} = (N\overline{\rho_t} - 1)/(N-1) \leq 1$. The parameter $\overline{\rho_t}$, called *average conditional correlation,* is the central concept in this specification defining many properties of the RW-ACC and RW-TACC estimators. To see how it is constructed, lets first define Δ_t as the average of portfolio returns constructed from the N individual assets, i.e. $\Delta_t = (1/N)\sum_{i=1}^{N} r_{it}$. The conditional variance of this portfolio is given by,

$$\sigma^2_{\Delta_t t} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sigma_{it}\sigma_{jt}\rho_{ijt} \tag{1.36}$$

where ρ_{ijt} is simply the rolling correlation estimator calculated over the entire sample. In the case where all covariances σ_{ijt} for all $i,j = 1,...,N$ are equal to the product of standard deviations $\sqrt{\sigma^2_{it}\sigma^2_{jt}}$ we have a perfect correlation among all assets, and the portfolio conditional variance is given by

$$\left(\sigma^2_{\Delta_t t}\right)^{\rho_{ij}=1} = \frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma_{it}\right)^2 \tag{1.37}$$

The average conditional correlation is defined as the ratio of the conditional variance of the portfolio when assets are not perfectly correlated and the conditional variance when $\rho_{ijt} = 1$,

$$\overline{\rho_t} = \frac{\sigma^2_{\Delta_t t}}{\left(\sigma^2_{\Delta_t t}\right)^{\rho_{ij}=1}} = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\sigma_{it}\sigma_{jt}}{\left(\sum_{k=1}^{N}\sigma_{kt}\right)^2}\rho_{ijt} \tag{1.38}$$

The Rolling Window Tree-Structured Average Conditional Correlation (RW-TACC) model is given by,

$$\begin{aligned} \mathbf{R}_t &= \left(1-\sum_{k=1}^{P}\lambda_k \mathbf{I}_{[(\overline{\rho}_{t-1},r_{t-1})\in R_k]}\right)\overline{\mathbf{Q}}_{t-p}^{t-1}+\left(\sum_{k=1}^{P}\lambda_k \mathbf{I}_{[(\overline{\rho}_{t-1},r_{t-1})\in R_k]}\right)\mathbf{I}_N, \quad (1.39)\\ \lambda_k &\in [0,1]\forall k, \end{aligned}$$

where $\mathbf{I}_{[(\overline{\rho}_{t-1},r_{t-1})\in R_k]}$ is an indicator function with value one when the average conditional correlation and the return at time $t-1$, belong to the R_k element of the optimal partition $P = \{R_1, ..., R_P\}$ of the predictive space $G = [0,1] \times \Re^N$ of $(\overline{\rho}_{t-1}, r_{t-1})^N$. The optimal partition is obtained following the tree-structured AR-GARCH methodology of Audrino and Trojani (2006).

After an empirical examination Audrino and Barone-Adesi find that the RW-ACC and RW-TACC models outperform the DCC model in terms of out-of-sample prediction capability. Four statistics including negative log-likelihood (NL), multivariate versions of mean absolute errors (MAE), root mean squared errors (RMSE) and mean of absolute empirical correlations (R^2), were used in in-sample and out-of-sample scenarios. They discovered overfitting problems associated with the DCC estimator when after reaching optimal in-sample values, a quite poor out-of-sample performance is produced in comparison to the rolling window estimators performance. The authors suggest that a possible reason for this is that the rolling window methods produce estimates that show a more progressive behaviour with very small scale fluctuations than those of the DCC approach.

In Audrino and Trojani (2004) the authors propose a DCC specification also estimated in two stages. Univariate volatilities are estimated in the first stage according to the tree-structured threshold proposed in Audrino and Trojani (2006), and correlations by a multivariate extension of the tree structured approach. The model is a generalisation of the DCC as proposed by Engle (2002). There is a "local" type of DCC specification for each regime introduced by each multivariate threshold. They present an empirical application where the data consists of nine international equity returns. They found strong GARCH and multivariate threshold effects in conditional volatilities, and threshold but no GARCH effects in conditional correlations.

By means of Monte Carlo simulations they compare the performance of

the model against four other specifications: A CCC-GARCH (1,1) model as in Bollerslev (1990), a tree structured CCC-GARCH (1,1) as in Audrino and Buhlmann (2001), a DCC-GARCH (1,1) as in Engle (2002), and a regime-switching DCC-GARCH (1,1) as in Pelletier (2006). They found that their threshold specification better fits GARCH dynamics as well as threshold structures in conditional volatilities and correlations.

See *inter alia* recent contributions by Palandri (2005), and Engle and Colacito (2006). Interesting is also the decomposition of the conditional density into the product of the marginals and the copula proposed by Granger, Teräsvirta and Patton (2006).

1.4.2 Tests for constant correlation

All the methods and models explained in the previous sections are appropriate only if there is a dynamic specification in the correlation between financial assets. This of course is a question of empirical nature that needs to be tested. In this section I briefly present three tests on conditional correlations: Bera and Kim (2002), Tse (2000), Engle and Shephard (2001), and the test by Goetzman *et al* (2002) on unconditional correlations. Although there are more studies in the literature (Kaplanis (1988), Ratner (1992), King and Wadhwani (1990), and Longin and Solnik (1995), just to cite some of them), I consider only this four as they appear to me as the more representative from a theoretical perspective.

Bera and Kim (2002), adopting the interpretation that the Information Matrix test is a test of parameter variation, test the hypothesis of constant correlation in a bivariate GARCH model. Basically, they derive Rao's score test of the null that the variance of the correlation parameter is zero. Consider the returns of two assets given by

$$x_{1t} = \sqrt{h_{1t}}\varepsilon_{1t}, \quad \varepsilon_{1t} \sim N(0, \sigma_{\varepsilon_1}) \tag{1.40}$$

and

$$x_{2t} = \sqrt{h_{2t}}\varepsilon_{2t}, \quad \varepsilon_{2t} \sim N(0, \sigma_{\varepsilon_2}) \tag{1.41}$$

where h_{1t} and h_{2t} are respectively the variances of x_{1t} and x_{2t}. Define ρ as the constant correlation coefficient between $\mathbf{x}_1$ and $\mathbf{x}_2$, $\widehat{\rho}$ as the MLE of ρ, and $\widehat{\varepsilon}_{1t}$ and $\widehat{\varepsilon}_{2t}$ as the estimated standardize residuals. They develop an efficient score (ES) form of the Information Matrix test as in Orme (1990). The test its given by

$$IM = \frac{\left[\sum_{t-1}^{T}\left(\upsilon_{1t}^2\upsilon_{2t}^2 - 1 - 2\widehat{\rho}^2\right)\right]^2}{4T\left(1 + 4\widehat{\rho}^2 + \widehat{\rho}^4\right)} \tag{1.42}$$

where

$$\upsilon_{1t} = \frac{\widehat{\varepsilon}_{1t} - \widehat{\rho}\widehat{\varepsilon}_{2t}}{\sqrt{1-\widehat{\rho}^2}} \quad \text{and} \quad \upsilon_{2t} = \frac{\widehat{\varepsilon}_{2t} - \widehat{\rho}\widehat{\varepsilon}_{1t}}{\sqrt{1-\widehat{\rho}^2}} \tag{1.43}$$

Under the null of constant correlation the IM statistic is asymptotically distributed as a $\chi^2(1)$.

In this specification, the variance of the indicator of the test for the constancy of ρ_t is derived using moment conditions of bivariate normal distribution. To make the test more robust to non-normalities, Bera and Kim (2002) transform the statistic to a Student-t version of the form,

$$IM_s = \frac{\left(\sum_{t=1}^{T}\eta_t\right)^2}{\sum_{t=1}^{T}(\eta_t - \overline{\eta})^2} \tag{1.44}$$

where $\eta_t = \upsilon_{1t}^2\upsilon_{2t}^2 - 1 - 2\widehat{\rho}^2$ and $\overline{\eta} = \sum_{t=1}^{T}\eta_t/T$. Monte Carlo experiments show that this version of the test has good finite-sample behaviour when data are not normal.

Another specification is the one by Tse (2000), who makes use of the CCC model by Bollerslev (1990) to construct a LM based test. He proposes the following equation to allow for time-varying correlations

$$\rho_{ijt} = \rho_{ij} + \delta_{ij}x_{i,t-1}x_{j,t-1} \tag{1.45}$$

where $\mathbf{x}_t = (x_{1t}, ..., x_{Nt})'$ is assumed to be conditionally multivariate normal with mean zero and covariance matrix $\mathbf{H}_t$. The null hypothesis of constant correlation is given by

$$H_0 : \delta_{ij} = 0, \text{ for } 1 \leq i \leq j \leq N \tag{1.46}$$

In the restricted model there are $N(N-1)/2$ independent parameters while in the unrestricted one there are N^2+2N . The construction of the LM test is natural in this case as we can easily obtain the restricted estimates by QMLE as is done in the CCC model. Lets denote θ_0 as the QMLE of θ under H_0 (under the CCC model). Then

$$LM = [s(\theta_0)]^2 I(\theta_0)^{-1} \sim \chi^2(N(N-1)/2) \tag{1.47}$$

where s is the score vector and $I(\theta)$ is the information matrix. Tse uses instead of the information matrix the sum of the cross products of the first derivatives of the likelihood function

$$LM = s(\theta_0)' \left[S(\theta_0)'S(\theta_0)\right]^{-1} s(\theta_0) \tag{1.48}$$

where $S(\theta)$ is a $T \times (N^2+2N)$ matrix, the rows of which are T partial derivatives of the likelihood function with respect to θ.

Engle and Sheppard (2001) construct a test under their DCC specification. They propose the test,

$$H_0 : R_t = \overline{R} \tag{1.49}$$

against

$$\begin{aligned} H_1 \quad : \quad & vech(R_t) = vech(\overline{R}) + \beta_1 vech(R_{t-1}) + \beta_2 vech(R_{t-2}) \\ & +... + \beta_s vech(R_{t-s}) \end{aligned} \tag{1.50}$$

The idea of the test is to use the standardised residuals from the first estimation stage ($\varepsilon_t = D_t^{-1} r_t$). These residuals have to be standardized again by the symmetric square root decomposition of the constant correlation $\overline{R}$,

$$v_t = \varepsilon_t \overline{R}^{-1/2} \tag{1.51}$$

Let's define $Y_t = vech^u \left[v_t v_t' - I_k\right]$, where $vech^u$ is a vech operator which only selects elements above the diagonal and I_k is the covariance matrix of residuals v_t. Under the null of constant correlation the residuals v_t should be zero mean *iid*, and the constant and all the lagged parameters in the vector

autoregression $Y_t = \alpha + \beta_1 Y_{t-1} + ... + \beta_s Y_{t-s} + \eta_t$ should be zero. The test statistic is given by

$$\frac{\widehat{\delta} X' X \widehat{\delta}'}{\widehat{\sigma}^2} \sim \chi^2(s+1) \tag{1.52}$$

A different approach is the one by Goetzmann *et al* (2002), where the test formulated is not for conditional correlations but for unconditional correlations. It is very flexible as it relaxes assumptions of normality in asset returns[4] and as it can be used to test cross sectional equality in correlations, something that can not be done in the MGARCH based tests. The test is based on the asymptotic distribution of the correlation matrix developed by Browne and Shapiro (1986) and Neudecker and Wesselman (1990). The hypotheses are the following

$$H_0 : \rho_1 = \rho_2 = \rho \text{ and } \Omega_1 = \Omega_2 = \Omega \tag{1.53}$$

$$H_1 : \rho_1 \neq \rho_2 \text{ or } \Omega_1 \neq \Omega_2 \tag{1.54}$$

where Ω is the asymptotic variance matrix of ρ. Under the null hypothesis the correlation matrices are equal element by element. Under H_0 we have

$$vech(\widehat{\rho_1} - \widehat{\rho_2}) \xrightarrow{d} N\left(0, \left[\frac{1}{N_1} + \frac{1}{N_2}\right]\Omega\right) \tag{1.55}$$

and the test statistic (Wald test) is,

$$[vech(\widehat{\rho_1} - \widehat{\rho_2})]' \left(\left[\frac{1}{N_1} + \frac{1}{N_2}\right]\Omega\right)^{-1} [vech(\widehat{\rho_1} - \widehat{\rho_2})] \xrightarrow{d} \chi^2\left(rk(\Omega)\right) \tag{1.56}$$

where $k(\cdot)$ is the number of elements in a *vech* operator matrix. Simulation results show that the test is robust to sample size and non-normality of the data.

[4]However it requires *iid* observation vectors with a multivariate distribution with finite fourth moments

1.4.3 Mis-specification tests

Once a model is selected and estimated, the next obvious step is to verify if it adequately captures the dynamics of the data. In what respects to univariate GARCH models the literature is quite comprehensive[5], but regarding multivariate specifications only two papers, Ding and Engle (2001) and Kroner and Ng (1998), present procedures specifically designed to test for misspecification in multivariate GARCH scenarios.

Ding and Engle (2001) generalizes the diagnostic test proposed in Engle (1982). Assume a specification of returns given by,

$$x_t/\mathcal{F}_{t-1} \sim N(0, \mathbf{H}_t) \tag{1.57}$$

where $\mathbf{H}_t = \mathbf{H}_t(\theta)$, and θ is the vector of parameters defining the variance-covariance matrix. If the model is correctly specified and θ is known then the standardize residuals will be given by,

$$u_t = \mathbf{H}_t^{-1/2}\varepsilon_t/\mathcal{F}_{t-1} \sim N(0, \mathbf{I}_N) \tag{1.58}$$

i.e., the variance matrix of standardize residuals will be equal to the identity matrix and the square standardize residuals will be serially uncorrelated. Three moment conditions are derived,

$$E(u_t u_t^{'}) = \mathbf{I}_N \tag{1.59}$$

$$Cov\left(u_{it}^2, u_{jt}^2\right) = 0, \text{ for all } i = j \tag{1.60}$$

$$Cov\left(u_{it}^2, u_{jt-k}^2\right) = 0, \text{ for } k > 0 \tag{1.61}$$

The second and third conditions have good power to detect non-normalities and misspecification of the conditional covariance equation.

In a more realistic scenario, the vector of parameters θ is unknown and has to be estimated by maximum likelihood. In this case we do not have exact moment conditions but sample estimates. Define m_t as the *vech* of standardize

[5] See Bollerslev et al (1994) and Bera and Higgins (1993).

residuals with dimensions $N(N-1)/2 \times 1$given by,

$$m_t = (u_{it}^2 - 1)(u_{jt}^2 - 1), \quad i \neq j, \; i = 1, ..., N, \; j = 1, ..., N \tag{1.62}$$

In the case where the MGARCH model is well specified ($u_t \sim N(0, \mathbf{I}_N)$) we have that $E(m_t) = 0$ and asymptotically that $m_T(\widehat{\theta}_T) = T^{-1} \sum_{t=1}^{T} m_t(\widehat{\theta}_t) = 0$. Using the m_t vector and the conditional moment test of Newey (1985) and Tauchen (1985), Engle and Ding derive the test statistic,

$$d_T = TR^2 \tag{1.63}$$

where R^2 is calculated from a regression of 1 on $[m_t', s_t']$, where s_t is the gradient of parameters defining the variance matrix, and $d_T \sim \chi^2(N(N+1)/2)$.

In a different approach Kroner and Ng (1998) generalise the "news impact curve" of Engle and Ng (1993) to a multivariate scenario (news impact surfaces) by plotting the conditional variance, covariance, and correlation against shocks from the last period. Combining these surfaces and the robust conditional moment test of Wooldrige (1990), they developed a diagnostic test by measuring the vertical distance between the scatter plot of the cross-product of residuals and the surface. More formally, consider a "generalized residual" given by,

$$g_{ijt} = \varepsilon_{it}\varepsilon_{jt} - h_{ijt} \tag{1.64}$$

If the model is correctly specified then $E_{t-1}(g_{ijt}) = 0$, i.e. the generalized residual is independent to any information available at time $t-1$. To make the test functional, Kroner and Ng select a group of exogenous variables called "misspecification indicators", which are taken as proxies for the information set I_{t-1} to be independent of g_{ijt}. Ten misspecification indicators are formed

$$
\begin{aligned}
x_{1t-1} &= \mathbf{I}(\varepsilon_{it-1} < 0; \varepsilon_{jt-1} < 0) \\
x_{2t-1} &= \mathbf{I}(\varepsilon_{it-1} < 0; \varepsilon_{jt-1} > 0) \\
x_{3t-1} &= \mathbf{I}(\varepsilon_{it-1} > 0; \varepsilon_{jt-1} < 0) \\
x_{4t-1} &= \mathbf{I}(\varepsilon_{it-1} > 0; \varepsilon_{jt-1} > 0) \\
x_{5t-1} &= \mathbf{I}(\varepsilon_{it-1} < 0) \\
x_{6t-1} &= \mathbf{I}(\varepsilon_{jt-1} < 0) \\
x_{7t-1} &= \varepsilon_{it-1}^2 \mathbf{I}(\varepsilon_{it-1} < 0) \\
x_{8t-1} &= \varepsilon_{it-1}^2 \mathbf{I}(\varepsilon_{jt-1} < 0) \\
x_{9t-1} &= \varepsilon_{jt-1}^2 \mathbf{I}(\varepsilon_{it-1} < 0) \\
x_{10t-1} &= \varepsilon_{jt-1}^2 \mathbf{I}(\varepsilon_{jt-1} < 0)
\end{aligned}
\tag{1.65}
$$

where $\mathbf{I}(\cdot)$ is the indicator function. The first four indicators corresponds to the possible sign partition of the $(\varepsilon_{it-1}, \varepsilon_{jt-1})$ space. This partition enables to compare asymmetric properties of different models. The fifth and sixth variables are sign indicators, and the remaining four are indicators controlling the size of the shocks.

The test statistic based on the robust conditional moment test of Wooldrige (1990) is given by

$$
C = \frac{\left[\sum_{t=1}^{T} g_{ijt}\lambda_{mt-1}\right]^2}{T\sum_{t=1}^{T} g_{ijt}^2\lambda_{mt-1}^2} \tag{1.66}
$$

where $\lambda_{mt-1}, m = 1..., 10$ is the residual from the regression

$$
x_{mt-1} = \sum_{k=1}^{K} \alpha_k \frac{\partial h_{ijt}}{\partial \theta_k} + \lambda_{mt-1} \tag{1.67}
$$

where $\boldsymbol{\theta} = (\theta_1, ...\theta_K)$ is the vector of parameters in the MGARCH model. Under general regularity conditions the C test has an asymptotic $\chi^2(1)$ distribution.

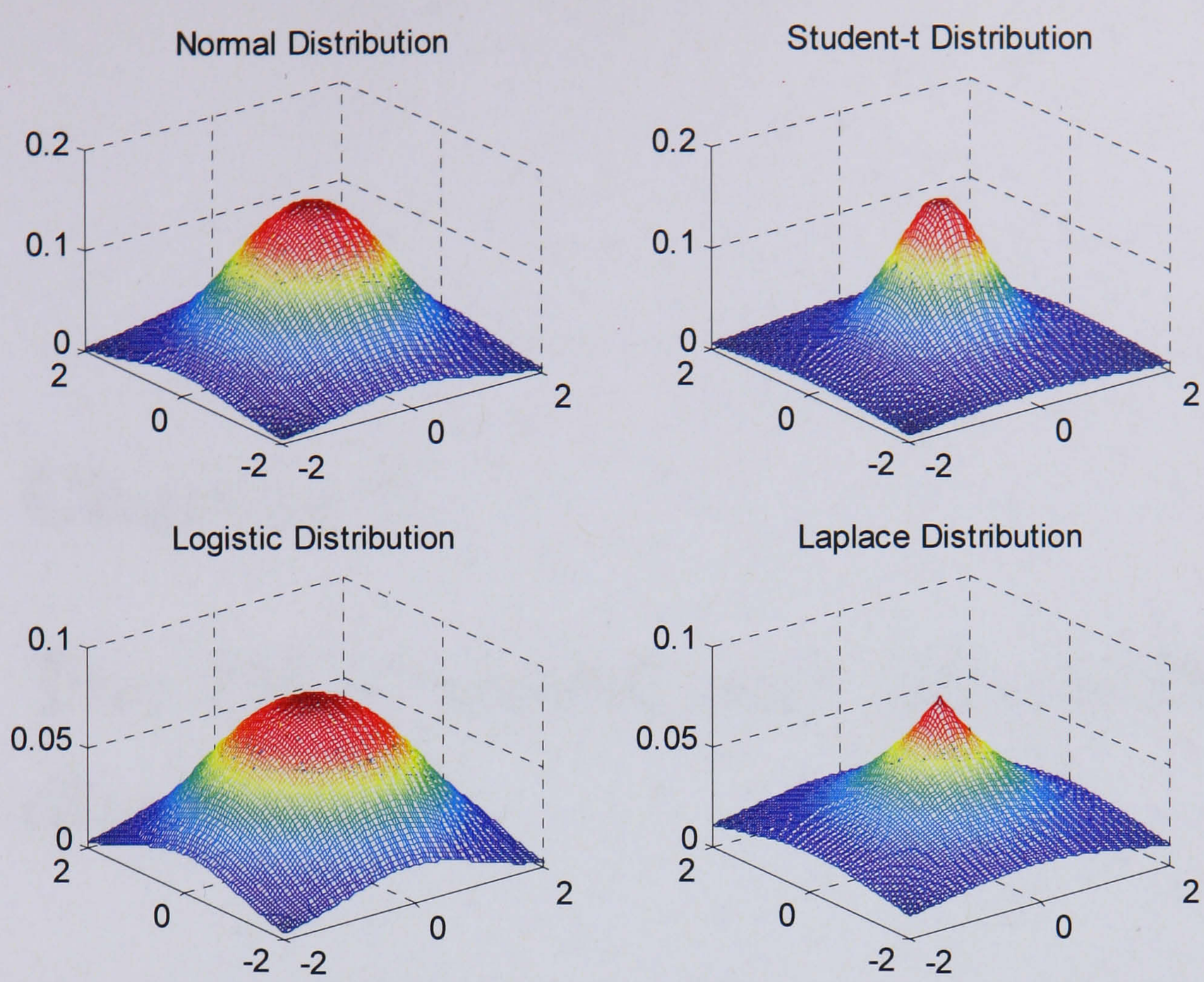


Figure 1.1. Examples of four bi-variate elliptical distributions.

Chapter 2

The DCC model and the AML distribution

2.1 Introduction

A good understanding of the dynamic properties of cross-market correlation (or dependence across markets) is vital for assessing the level of integration between international markets both for investment purposes and for increasing the capacity to produce reliable forecasts. Modelling the dynamics of volatilities of returns from financial assets has been one of the working horses in the development of financial econometrics over the last years (Bollerslev, 2001; Engle, 2001). Nonetheless, most of the advances, especially if we consider the use of the proposed framework for practical purposes, have been seen almost exclusively in univariate cases. The growth in techniques modelling the dynamics of covariances and correlations has lagged considerably behind the growth in modelling time-varying volatility, as evidenced by the shortage of the literature on time-varying correlations compared to that of modelling time-varying volatility. One of the main reasons for this uneven expansion is the "curse of dimensionality", due to the extremely cumbersome problems faced in estimating unrestricted multivariate GARCH (MGARCH) models in highly dimensioned settings. Bauwens *et al* (2006) provide a comprehensive survey on MGARCH models. See also Kroner and Ng (1998).

Among the various MGARCH specifications recently proposed in the liter-

ature, one in particular has proved to be particular suitable to provide a parsimonious, flexibility, and feasible model that significantly reduces the "curse of dimensionality". This is the Dynamic Conditional Correlation (DCC) model proposed in Engle (2002) and Engle and Sheppard (2001). In this model, the dynamic variance-covariance matrix of conditional returns is specified as a function of univariate variances and linear correlations. When the model is estimated by maximum likelihood this framework allows to "break" the log-likelihood function in two parts, one for the parameters determining univariate volatilities and another for the parameters determining the correlations (the so-called DCC two-step estimation technique). By using this technique large systems can be consistently estimated with limited computational costs without imposing too many restrictions like in the case of factor models.

A vital assumption of the DCC model is that standardized residuals are normally distributed. Normality allows (Q)MLE to provide feasible and consistent but inefficient DCC coefficients of conditional correlations (Bollerslev and Wooldridge, 1992). Nevertheless, financial time series do not favour this assumption. Where time-varying volatilities are estimated by assuming a normal-GARCH process for the innovations, it is easy to show that even for correctly specified models, statistically significant levels of leptokurtosis and excess kurtosis can still be found.

Current routes of investigation to overcome the limits in the assumption of normality are mainly two. First, there is strand of literature which aims at using non parametric and semiparametric methods of investigations: Engle and González-Rivera (1991), Drost and Klaassen (1997), González-Rivera (1997), González-Rivera and Drost (1999), Hafner and Rombouts (2004), Long and Ullah (2005). For the few papers studying the multivariate case, estimation and inference is quite difficult and feasible only for a small number of assets. A second strand of contributions use thick (non normal) distributions to achieve efficiency with implication for the first stage. See for instance Bollerslev (1987), Baillie and Bollerslev (1989), Nelson (1991), Verhoeven and McAleer (2003), Fiorentini, Sentana, Calzolari (2003) and Mencia and Sentana (2005).

As already pointed out, returns from financial assets show well defined patterns of leptokurtosis and skewness which cannot be captured by the nor-

mality assumption. There are several multivariate distributions in the literature that present high levels of kurtosis as well as asymmetries and that could be used in a MGARCH framework. Bauwens and Laurent (2004) review multivariate asymmetric densities. However, the majority of these distributions are either too complicated to be estimated for GARCH purposes or present undesirable properties (like an infinite variance) that limit their use for financial applications. One multivariate distribution that parsimoniously captures the main features of financial returns and keeps flexibility is the Asymmetric Multivariate Laplace (AML) distribution, as recently proposed by Kotz, Kozubowski, and Podgorski (2003). In the univariate context, the Laplace or double-exponential distribution has been widely used in financial modelling. Some applications include Madan and Seneta (1990), Madan *et al.* (1988), Linden (2001), Kou and Wang (2001), Hanson and Zhu (2004), Sepp (2004), Heyde and Kou (2004), Komunjer (2005) among many others. The asymmetric multivariate version used in this paper is defined as a subclass of geometric stable distributions (see Section 2.3), a characteristic that in the case of the AML distribution can be used to model linear combinations of random variables with univariate symmetric Laplace distributions. This feature is extremely important as it allows to use this distribution in the computation of the parametric-VaR of portfolios of financial assets, characteristic that was thought exclusive of the Pareto-stable distribution and of its most widely used limiting case such as the normal distribution.

Our work is in the spirit of Mencia and Sentana (2005) who use a generalised hyperbolic distribution in a model where the variance matrix dynamics follow a conditionally heteroskedastic single factor model and the conditional variance of the factor obeys a univariate GQARCH (1,1) process; and Bauwens and Laurent (2004) who use a type of multivariate skewed Student-t distribution to fit a DCC (1,1) model to two sets of three assets data. As far as we are concerned this is the first work where the AML distribution is used to model the returns of financial assets in a MGARCH setting.

The main aim of this chapter is to develop a multivariate time-varying framework for modelling and forecasting cross-market correlations where innovations are assumed to follow an AML distribution. The outline of this chapter is as follows. In Section 2.2, we review the dynamic conditional cor-

relation (DCC) model of Engle (2002) and Engle and Sheppard (2001) and the extensions that allow for asymmetries in the dynamics and asset-specific correlations, as proposed by Cappiello, Engle and Sheppard (2004, CES henceforth). In Section 2.3, we present a framework where the DCC is enriched by the asymmetric multivariate Laplace (AML) distribution, and we discuss the implications of the estimation of the DCC model by maximum likelihood under the AML assumption. In Section 2.6 we report the results from an empirical application using a sample of 21 FTSE All-World stock indices and 13 bond return indices. Section 2.7 concludes.

2.2 DCC models

Consider the $n-$dimensional returns process $r_t \in \mathbb{R}^{T\times n}$, $t = 1, .., T$ generated as,

$$r_t = \mathbf{H}_t^{1/2}(\boldsymbol{\theta})\boldsymbol{\varepsilon}_t \tag{2.1}$$

$$\mathbf{H}_t = Var\left(r_t \mid \Omega_{t-1}\right) \tag{2.2}$$

where Ω_τ is the information set at time τ, and $\boldsymbol{\varepsilon}_t$ is an i.i.d. process. In the DCC setting $\mathbf{H}_t$ is modelled directly as a function of dynamic univariate variances and dynamic linear correlations,

$$\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t \tag{2.3}$$

Where $\mathbf{D}_t \in \mathbb{R}^{T\times n}$ is a diagonal matrix with elements $\sqrt{h_{it}}, i = 1, ..., n$, $t = 1, .., T$, and $\mathbf{R}_t$ is defined as

$$\mathbf{R}_t = \left(\mathbf{Q}_t^*\right)^{-1}\mathbf{Q}_t\left(\mathbf{Q}_t^*\right)^{-1} \tag{2.4}$$

where

$$\mathbf{Q}_t^* = \begin{bmatrix} \sqrt{q_{11}} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{q_{22}} & 0 & \dots & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & \dots & \sqrt{q_{nn}} \end{bmatrix}, \tag{2.5}$$

and

$$\mathbf{Q}_t = \left(1 - \sum_{l=1}^{L} \alpha_l - \sum_{s=1}^{S} \beta_s\right) \overline{\mathbf{Q}} + \sum_{l=1}^{L} \alpha_l \boldsymbol{\varepsilon}_{t-l} \boldsymbol{\varepsilon}_{t-l}' + \sum_{s=1}^{S} \beta_s \mathbf{Q}_{t-s} \tag{2.6}$$

$\overline{\mathbf{Q}} \in \mathbb{R}^{n \times n}$ is the unconditional variance-covariance matrix of $\boldsymbol{\varepsilon}_t$, i.e. $\overline{\mathbf{Q}} = E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t')$, and α_l and β_s are scalar parameters satisfying $\sum_{l=1}^{L} \alpha_l + \sum_{s=1}^{S} \beta_s < 1$. The specification in (2.4) secures that $\mathbf{R}_t$ will be a valid correlation matrix while (2.3) and (2.6), in addition to the condition of stationarity, secure $\mathbf{H}_t$ to be a positive definite matrix.

The dynamics in (2.3) is particularly appealing, because it allows for a two step estimation that makes feasible the estimation of highly dimensioned processes, estimation that for many non-factor models is usually not possible because of the "curse of dimensionality".

To illustrate the two-step estimation technique let us assume first normality for the vector of standardised residuals, i.e. $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{R}_t)$. Denoting $\boldsymbol{\theta}$ as the vector of parameters in the conditional variance-covariance matrix $\mathbf{H}_t$, the log-likelihood $L_T(\boldsymbol{\theta})$ for the T observations of this estimator,

$$L_T(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log f(r_t \mid \boldsymbol{\theta}, \Omega_{t-1}) \tag{2.7}$$

is given by,

$$L_T(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{T} \left\{ n \log(2\pi) + \log |\mathbf{H}_t| + \mathbf{r}_t' \mathbf{H}_t^{-1} \mathbf{r}_t \right\} \tag{2.8}$$

Following (2.3) we have,

$$L_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left\{n\log(2\pi) + \log|\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t| + \mathbf{r}_t^{'}(\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t)^{-1}\mathbf{r}_t\right\} \tag{2.9}$$

and

$$L_T(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left\{n\log(2\pi) + \log\left|\mathbf{D}_t^2\right| + \log|\mathbf{R}_t| + \boldsymbol{\varepsilon}_t'\mathbf{R}_t^{-1}\boldsymbol{\varepsilon}_t\right\} \tag{2.10}$$

with the standardised residuals $\boldsymbol{\varepsilon}_t = \mathbf{D}_t^{-1}\mathbf{r}_t$. Engle (2002) proposes to estimate the first stage by assuming $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{I})$ where $\mathbf{I} \in \mathbb{R}^{n\times n}$ is an identity matrix. By partitioning the vector of parameters in two subsets $\boldsymbol{\theta} = (\boldsymbol{\zeta}, \boldsymbol{\varphi})$, where $\boldsymbol{\zeta}$ contains the parameters of the n univariate volatilities and $\boldsymbol{\varphi}$ contains the parameters of the correlations, the log-likelihood function can be expressed as,

$$L_T(\boldsymbol{\theta}) = L_T(\boldsymbol{\zeta}) + L_T(\boldsymbol{\varphi} \mid \boldsymbol{\zeta}) \tag{2.11}$$

The estimation of the first stage consists in the maximization of the function,

$$L_T(\boldsymbol{\zeta}) = -\frac{1}{2}\sum_{t=1}^{T} n\log(2\pi) + \log\left|\mathbf{D}_t^2\right| + \mathbf{r}_t^{'}\mathbf{D}_t^{-2}\mathbf{r}_t \tag{2.12}$$

Once the vector $\boldsymbol{\zeta}$ is estimated, the vector of standardise residuals $\boldsymbol{\varepsilon}_t = \mathbf{D}_t^{-1}\mathbf{r}_t$ is employed in the second stage, which corresponds to the maximization of the function,

$$L(\boldsymbol{\varphi} \mid \boldsymbol{\zeta}) = -\frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{R}_t| + \boldsymbol{\varepsilon}_t'\mathbf{R}_t^{-1}\boldsymbol{\varepsilon}_t, \tag{2.13}$$

under the assumption $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{R}_t)$.

The novelty of this technique is that estimation is speeded up by employing $n+1$ log-likelihood functions[1] instead of one single but nonetheless extremely

[1] The first stage implies the estimation of n univariate volatility processes and the second stage the estimation of one single correlation process.

flat log-likelihood function.

To show consistency of the two-step DCC estimator Engle (2002) employs the results in Newey and McFadden (1994) for the two-step Generalised Method of Moments (GMM). The result follows from the fact that Maximum Likelihood estimation can be considered a special case of the GMM when the moment conditions are set equal to the scores of the log-likelihoods,

$$\nabla_{\zeta} L_T(\zeta) = \mathbf{0} \tag{2.14}$$

$$\nabla_{\varphi|\zeta} L_T(\varphi \mid \zeta) = \mathbf{0} \tag{2.15}$$

The specification in (2.6) can be enriched by allowing for asymmetries in conditional correlations and covariances as well as for asset-specific correlations. We will refer to this general model as the Asymmetric Generalised Dynamic Conditional Correlation (AGDCC) (L,S,U) model, where L corresponds to the number of autoregressive lags, S corresponds to the number of persistence lags, and U corresponds to the number of asymmetric shock lags[2]. The specification of the matrix $\mathbf{Q}_t$ for the AGDCC (1,1,1) case is given by,

$$\begin{aligned} \mathbf{Q}_t &= \left(\overline{\mathbf{Q}} - \mathbf{A}'\overline{\mathbf{Q}}\mathbf{A} - \mathbf{B}'\overline{\mathbf{Q}}\mathbf{B} - \mathbf{G}'\overline{\mathbf{N}}\mathbf{G}\right) + \mathbf{A}'\varepsilon_{t-1}\varepsilon'_{t-1}\mathbf{A} + \mathbf{B}'\mathbf{Q}_{t-1}\mathbf{B} \\ &\quad + \mathbf{G}'\eta_{t-1}\eta'_{t-1}\mathbf{G} \end{aligned} \tag{2.16}$$

where $\mathbf{A}, \mathbf{B}$, and $\mathbf{G}$ are diagonal parameter matrices $(\mathbf{A}, \mathbf{B}, \mathbf{G} \in \mathbb{R}^{\mathbf{n}\times\mathbf{n}})$ with elements a_{ii}, b_{ii} and g_{ii} respectively, $\eta_\tau = I\left[\varepsilon_\tau < 0\right] \circ \varepsilon_\tau$, " $\circ$ " denotes the Hadamard product and $\overline{\mathbf{N}} = E(\eta_t\eta'_t)$. $\mathbf{Q}_t$ will be positive-definite if

$$\left(\overline{\mathbf{Q}} - \mathbf{A}'\overline{\mathbf{Q}}\mathbf{A} - \mathbf{B}'\overline{\mathbf{Q}}\mathbf{B} - \mathbf{G}'\overline{\mathbf{N}}\mathbf{G}\right)$$

is positive definite[3].

[2]The extension to an Asymmetric Dynamic Conditional Correlation (ADCC) (L,S,U) model is presented in Cappiello et al (2004) . Alternative generalisations that allow for asset-specific correlations and group-specific correlations are presented in Billio et al (2004, "Flexible DCC" model), Hafner and Franses (2003, "Generalised DCC" model), and also in Cappiello et al (2004) under the name "Asymmetric Generalised DCC" model.

[3]Because this cannot always be guaranteed, Hafner and Franses (2003) replace this

The AGDCC (L,S,U) model (2.16) nests several specifications:

- DCC (L,S) model : $\mathbf{G} = [0], \mathbf{A} = \sqrt{a}, \mathbf{B} = \sqrt{b}$
- ADCC (L,S,U) model: $\mathbf{G} = \sqrt{g}, \mathbf{A} = \sqrt{a}, \mathbf{B} = \sqrt{b}$
- GDCC (L,S) model: $\mathbf{G} = [0]$.

2.3 The asymmetric multivariate Laplace distribution

The hypothesis of normality of the returns series makes (Q)MLE feasible, providing consistent though inefficient estimates of the dynamic conditional correlation coefficients (Bollerslev and Wooldridge, 1992). However, normality is not a satisfactory property for financial time series. This has important implications not only for the econometric properties of parameter estimates, but also for the use of the models in applications such as portfolio allocation, VaR and Expected Shortfall analyses.

In this chapter we propose to explore the properties of the Asymmetric Laplace distribution proposed by Kozubowski and Podgorski (2001) as a subclass of geometric stable distributions. In particular we will consider the results in Kotz *et al* (2003), who generalised the laws of an Asymmetric Laplace to the multivariate case.

In the geometric stable model, the return $r_{f(p)}$ is consider to be the sum of smaller returns $r^{(i)}$ over the period of time $f(p)$ which is a stopping time random variable with geometric probability function $P(f(p) = j) = p(1-p)^{j-1}$, $j = 1, 2, \ldots$. The geometric stable distribution can be approximated to a normalised geometric stable model sum when the p parameter of the stopping time function $f(p)$ approaches zero. More formally, the random array $\mathbf{X}$ has a geometric stable distribution in $\mathbb{R}^n$ if and only if,

---

expression by $\left(1 - \bar{a}^2 - \bar{b}^2\right)\overline{\mathbf{Q}}$. The correlation-targeting approach implicit in (2.16) is sacrificed with this substitution but $\mathbf{Q}_t$ will be positive-definite.

$$a(p)\sum_{i=1}^{f(p)}\left(\boldsymbol{\kappa}(p)+\mathbf{r}^{(i)}\right)\xrightarrow{d}\mathbf{X},\qquad \text{as } p\to 0 \tag{2.17}$$

where $\left\{\{\mathbf{r}^{(d)}=\left(r_1^{(d)},...,r_n^{(d)}\right),d\geq 1\right\}$ is a sequence of i.i.d. random vectors in $\mathbb{R}^n$ independent of $f(p)$, $a(p)>0$, $\boldsymbol{\kappa}(p)\in\mathbb{R}^n$, and $\xrightarrow{d}$ denotes convergence in distribution. The AML distribution appears in this context when the distributional limit (2.17) is restricted to have a finite second moment. More precisely, Kozubowski and Podgorski (2001) shows that when each vector in $\mathbf{r}$ has a mean $m_i, i=1,...n$, a variance $\sigma_{ij}, i=1,...n, j=1,...n$, and when we let $a(p)=\sqrt{p}$ and $\boldsymbol{\kappa}(p)=\mathbf{m}\left(\sqrt{p}-1\right)$, the random variable $\mathbf{X}$ defined by the convergence in distribution in (2.17) will have an AML distribution with the characteristic function,

$$\Psi(\mathbf{t})=\frac{1}{1+\frac{1}{2}\mathbf{t}'\mathbf{H}\mathbf{t}-i\mathbf{t}'\mathbf{m}} \tag{2.18}$$

where $\mathbf{t}\in\mathbb{R}^n$, and $\mathbf{H}\in\mathbb{R}^{n\times n}$ is a positive-definite matrix.

The density function of the AML distribution in $\mathbf{H}$ and $\mathbf{r}$ is given by,

$$f(\mathbf{r})=\frac{2\exp\left(\mathbf{r}_t'\mathbf{H}^{-1}\mathbf{m}\right)}{(2\pi)^{n/2}\left|\mathbf{H}\right|^{1/2}}\left(\frac{\mathbf{r}_t'\mathbf{H}^{-1}\mathbf{r}_t}{2+\mathbf{m}'\mathbf{H}^{-1}\mathbf{m}}\right)^{v/2}K_v\left(\sqrt{(2+\mathbf{m}'\mathbf{H}^{-1}\mathbf{m})(\mathbf{r}_t'\mathbf{H}^{-1}\mathbf{r}_t)}\right) \tag{2.19}$$

where $v=(2-n)/2$ and $K_v(u)$ is the modified Bessel function of the third kind defined by $K_v(u)=\frac{(u/2)^v\Gamma(1/2)}{\Gamma(v+1/2)}\int_1^\infty e^{-ut}(t^2-1)^{v-1/2}dt$, $u>0, v\geq -1/2$. The vector $\mathbf{m}$ is the location parameter and the matrix $\mathbf{H}$ is the scale parameter of this distribution. These are related by the relation $\mathbf{m}\equiv\mathbf{H}\mathbf{b}$ where $\mathbf{b}\in\mathbb{R}^n$.

A very important characteristic of the AML distribution is that it is unimodal with the mode equal to zero. Because of this the $\mathbf{m}$ parameter does not only determines the mean of the distribution, but also its level of asymmetry. When $\mathbf{m}=\mathbf{0}$ the distribution is symmetric collapsing as it can clearly be seen in equation (2.18) to the elliptical case (see discussion in Johnson and Kotz, 1972).

Figures 2.1-2.6 present alternative bivariate AML densities with alternative $\mathbf{m}$ vectors and correlation levels.

[Insert Figures 2.1-2.6 here]

2.4 Generalised hyperbolic distributions

As shown in Kotz *et al* (2003), AML distributions can also be obtained as a limiting case of the Generalised Hyperbolic (GH) distribution, introduced by Barndorff-Nielsen (1977). These are location-scale mixtures of normal distributions, i.e. if $\mathbf{X}$ has a GH distribution in $\mathbb{R}^n$ then,

$$\mathbf{X} \stackrel{D}{=} \boldsymbol{\mu} + \mathbf{m}\boldsymbol{\xi} + \boldsymbol{\xi}^{1/2}\mathbf{Z}^4 \tag{2.20}$$

where "$\stackrel{D}{=}$" denotes equality in distribution, $\mathbf{Z} \sim N_n(\mathbf{0}, \mathbf{H})$, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $\boldsymbol{\xi}$ is a generalised inverse Gaussian variable with parameters ν, γ, and δ, i.e. $\boldsymbol{\xi} \sim GIG(\nu, \gamma, \delta)$. AML distributions appear when $\boldsymbol{\mu} = \mathbf{0}$ and when $\boldsymbol{\xi}$ is not $GIG(\nu, \gamma, \delta)$ but standard exponential, i.e. $\boldsymbol{\xi} \sim EXP(1)$. Note that the limiting case $GIG(1, 0, 2)$ is equivalent to $EXP(1)$.

Mencia and Sentana (2005) analyse the GH distribution in multivariate conditionally heteroskedastic dynamic regression models. The dynamics of the conditional covariance matrix $\mathbf{H}_t$ are given by a single factor model with a GQARCH(1,1) specification for the common factor, and a time-invariant diagonal matrix for the idiosyncratic terms. Given that $\boldsymbol{\mu} = \mathbf{0}$ because the mean of the returns has been removed prior to estimation, the only difference with the AML distribution resides in the employed mixing distribution. The generalised inverse Gaussian distribution allows for flexible tail modelling but at the cost of limiting the inclusion of rich dynamics for the conditional variance matrix because of the "course of dimensionality"[5]. For the case of a highly parameterised specification like the AGDCC model the estimation using the GH distribution is extremely difficult.

The representation of the AML distribution as a location-scale mixture of normal distributions is given by,

$$\mathbf{X} \stackrel{D}{=} \mathbf{m}\boldsymbol{\xi} + \boldsymbol{\xi}^{1/2}\mathbf{Z} \tag{2.21}$$

[5] It also limits the application of a two-step estimation

where in this case $\boldsymbol{\xi} \sim EXP(1)$. From this it can easily be seen that $E(\mathbf{X}) = \mathbf{m}$ and $Var(\mathbf{X}) = \mathbf{H} + \mathbf{m}\mathbf{m}'$. This is of particular importance for the estimation of the MGARCH model. Contrary to the Gaussian case, the variance of a random variable with AML distribution does not coincide with the scale parameter of the distribution. Note that $Var(\mathbf{X}) = \mathbf{H}$ only when the distribution is elliptical, i.e. when $\mathbf{m} = \mathbf{0}$).

In contrast with the majority of GH distributions, the AML distribution in the special case $\mathbf{m} = \mathbf{0}$ is stable, just as the normal. This condition implies an important property necessary for the modelling of financial portfolios known as the additivity property, which is basically the concept that a linear combination of independent random variables with stability index α is also stable with the same parameter α (See Khindanova *et al*, 2001).

Pareto stable distributions are stable under random summation. Formally, the random variable $\mathbf{X}$ is said to be Pareto stable if for any $a_i > 0$, $i = 1, ..., d$, there exist a constant $c = d^{1/\alpha}$ and $\mathbf{u}_d \in \mathbb{R}^n$ for any $d \geq 2$ such that,

$$a_1\mathbf{X}^{(1)} + ... + a_d\mathbf{X}^{(d)} \stackrel{D}{=} c\mathbf{X} + \mathbf{u}_d \tag{2.22}$$

where $\mathbf{X}^{(1)}, ..., \mathbf{X}^{(d)}$ are independent copies of $\mathbf{X}$. In an alike way Laplace laws are stable, but under geometric summation instead of random summation. To be able to preserve stability we have to constraint the normalising constants $a(p)$ and $\boldsymbol{\kappa}(p)$ in (2.17) to,

$$a(p) = \sqrt[\alpha]{p}, \quad \boldsymbol{\kappa}(p) = \mathbf{0} \tag{2.23}$$

The first condition implies that for the case of the AML distribution $\alpha = 2$. This is the same alpha value of the normal distribution which is the only Pareto-stable distribution with a finite second moment. The second condition $\boldsymbol{\kappa}(p) = \mathbf{0}$ implies $\mathbf{m} = \mathbf{0}$, restricting the use of the distribution for portfolio-VaR applications to the symmetric case.

2.5 Estimation

We turn now to the estimation of DCC models employing AML distributions. The log-likelihood function $L_T^{AML}(\boldsymbol{\theta})$ assuming a AML distribution for the conditional returns is proportional to,

$$\begin{aligned} L_T^{AML}(\boldsymbol{\theta}) \;=\; & \sum_{t=1}^{T}\Big\{\mathbf{r}_t'\mathbf{H}_t^{-1}\mathbf{m} - \frac{1}{2}\ln|\mathbf{H}_t| + \\ & \frac{v}{2}\left(\ln(\mathbf{r}_t'\mathbf{H}_t^{-1}\mathbf{r}_t) - \ln(2+\mathbf{m}'\mathbf{H}_t^{-1}\mathbf{m})\right) \\ & + \ln\left[K_v\left(\sqrt{(2+\mathbf{m}'\mathbf{H}_t^{-1}\mathbf{m})(\mathbf{r}_t'\mathbf{H}_t^{-1}\mathbf{r}_t)}\right)\right]\Big\} \end{aligned} \tag{2.24}$$

From $H_t \equiv D_t R_t D_t$, equation (2.24) can be written as

$$\begin{aligned} L_T^{AML}(\boldsymbol{\theta}) \;=\; & \sum_{t=1}^{T}\Big\{\mathbf{r}_t'(\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t)^{-1}\mathbf{m} - \frac{1}{2}\ln|(\mathbf{D}_t\mathbf{R}_t\mathbf{D}_t)| \\ & + \frac{v}{2}\left(\ln(\mathbf{r}_t'\mathbf{D}_t^{-1}\mathbf{R}_t^{-1}\mathbf{D}_t^{-1}\mathbf{r}_t) - \ln(2+\mathbf{m}'(\mathbf{D_t}\mathbf{R_t}\mathbf{D_t})^{-1}\mathbf{m})\right) \\ & + \ln\left[K_v\left(\sqrt{(2+\mathbf{m}'(\mathbf{D_t}\mathbf{R_t}\mathbf{D_t})^{-1}\mathbf{m})(\mathbf{r}_t'\mathbf{D_t}^{-1}\mathbf{R_t}^{-1}\mathbf{D_t}^{-1}\mathbf{r}_t)}\right)\right]\Big\} \end{aligned} \tag{2.25}$$

The $\mathbf{m}$ parameter cannot be estimated in the first step because it is a function of the conditional covariances which are estimated only in the second stage. Thus, assume $R_t = I$ and $\mathbf{m} = [\mathbf{0}]$ and let us denote with $\boldsymbol{\zeta}$ the set of parameters in the matrix $\mathbf{D}_t$. The first stage likelihood function is

$$\begin{aligned} L_T^{AML}(\boldsymbol{\zeta}) \;=\; & \sum_{t=1}^{T}\Big\{-\frac{1}{2}\ln\left|\mathbf{D}_t^2\right| + \frac{v}{2}\left(\ln(\mathbf{r}_t'\mathbf{D}_t^{-2}\mathbf{r}_t) - \ln(2)\right) \\ & + \ln\left[K_v\left(\sqrt{2(\mathbf{r}_t'\mathbf{D_t}^{-2}\mathbf{r}_t)}\right)\right]\Big\} \end{aligned} \tag{2.26}$$

Contrary to the normal case, $L_T^{AML}(\boldsymbol{\zeta})$ cannot be expressed as the sum of n-log-likelihood functions, i.e. the parameters in $\boldsymbol{\zeta}$ have to be estimated maximizing one single log-likelihood function. This, however, does allow to

continue to use the two-step estimation technique[6] although it does extend the computing time for estimation.

Defining $\boldsymbol{\varepsilon}_t = \mathbf{r}_t'\mathbf{D}_t^{-1}$ and $\boldsymbol{\varepsilon}_t^* = \mathbf{m}'\mathbf{D}_t^{-1}$, the second-stage log-likelihood is given by,

$$\begin{aligned} L_T^{AML}(\boldsymbol{\varphi}|\,\boldsymbol{\zeta}) \;=\; & \sum_{t=1}^{T}\Big\{\boldsymbol{\varepsilon}_t\mathbf{R}_t^{-1}\left(\boldsymbol{\varepsilon}_t^*\right)' - \frac{1}{2}\ln|\mathbf{R}_t| + \\ & \frac{v}{2}\left(\ln(\boldsymbol{\varepsilon}_t\mathbf{R}_t^{-1}\boldsymbol{\varepsilon}_t') - \ln(2+\boldsymbol{\varepsilon}_t^*\mathbf{R}_t^{-1}\left(\boldsymbol{\varepsilon}_t^*\right)')\right) \\ & + \ln K_v\left(\sqrt{(2+\boldsymbol{\varepsilon}_t^*\mathbf{R}_t^{-1}\left(\boldsymbol{\varepsilon}_t^*\right)')(\boldsymbol{\varepsilon}_t\mathbf{R}_t^{-1}\boldsymbol{\varepsilon}_t')}\right)\Big\} \end{aligned} \tag{2.27}$$

Functions (2.26) and (2.27) can be computed by calculating the integral of the third component of the functions. Both functions were estimated via maximum likelihood estimation method. We choose this method because it yields consistent and asymptotically efficient parameter estimates when the assumed distribution is correctly specified. The use of a flexible distribution like the AML is in this regard very important. This distribution captures excess kurtosis and asymmetries which are usual features in the return of financial assets. Other works like Fiorentini *et al* (2003) have relaxed normality using the multivariate Student-t distribution, and although this distribution allows for excess kurtosis it can not deal with skewness, risking a possible inconsistency of parameter estimates. We dedicate entirely Chapter 3 to the analysis of the issue of consistency of estimates when the selected distribution for standardised residuals is the AML.

Finally, the computation of (2.26) and (2.27) is feasible but requires calculation of some functionals and thus it is computationally quite costly. An alternative is the numerical solution suggested in Kotz *et al* (2003).

2.6 Empirical application

This application is intended to provide evidence regarding the superiority of the AML-DCC specification compared to the normal-DCC model. We focus

[6]The parameter estimates in the vector $\widehat{\boldsymbol{\zeta}}$ will still be consistent if the specification for the distribution of the returns is correct

on specification tests for the distribution of standardised residuals and on the features of parameter estimates. We present also a small Value-at-Risk exercise to compare the behaviour of both models. A much more detailed and broad empirical analysis will be presented in Chapter 4 regarding the benefits of our specification for risk management applications.

For this application we employed the sample of 21 FTSE All-World stock indices and 12 bond return indices used in Cappiello, Engle and Sheppard (2004). Weekly returns were calculated through log differences using Friday to Friday closing prices and filtered by removing the mean,

$$r_{jt} = \log\left(\frac{P_{jt}}{P_{jt-1}}\right) - \frac{1}{T}\sum_{i=1}^{T}\log\left(\frac{P_{ji}}{P_{ji-1}}\right), \quad j = 1, ...n \tag{2.28}$$

where P_{jt} is the price of assets j at time t.

We estimate the four models described above: AGDDC (1,1,1), GDCC(1,1), ADCC (1,1,1), and the DCC (1,1)[7]. Tables 2.1a,b report the parameter estimates of the joint GARCH (1,1) processes for the univariate volatilities, and the skewness and kurtosis of the returns standardised by their estimated standard deviation.

To evaluate the parametric assumptions for the univariate case we compare the Kolmogorov-Smirnov distances between residuals standardised by volatilities estimated in the first stage, and the two implicit univariate normal and Laplace distributions. Table 2.2a,b reports the statistics.

[Insert Tables 2.1a,b-2.2a,b here]

In this case there is no apparent advantage in the use of the Laplace distribution. The results for the more relevant multivariate case nonetheless are more encouraging. To evaluate the multivariate distributions we implemented the visual diagnostic proposed in Kawakatsu (2005). The idea is based on the fact that if $r_t \sim N(0, H_t)$, then $r_t H_t^{-1} r_t'$ has a $\chi^2(n)$ distribution. Although we do not know the distribution of $r_t H_t^{-1} r_t'$ when $r_t \sim AML(m, H_t)$ we generated an empirical distribution in order to perform the comparison. Figure 2.7 presents the qq-plots of the sample quantiles of $r_t H_t^{-1} r_t'$ for the four normal

[7]For simplicity from this point we drop the number of lags when we refer to the MGARCH models estimated in the exercise.

models against the quantiles of $\chi^2(n)$, and the qq-plots of the sample quantiles of $r_t H_t^{-1} r_t'$ for the four AML models against the quantiles of the empirical distribution. From a visual inspection it is clear that the assumption of a AML distribution is more appropriate than the one of normality.

[Insert Figure 2.7 here]

To reinforce our findings about the inconvenience of the assumption of multivariate normality we also performed the omnibus test of Doornik and Hansen (1994). Multivariate normality was overwhelmingly rejected for the raw and standardised data after fitting the normal DCC, ADCC, GDCC, and AGDCC models. All p-values are ≈ 0, and thus we do not report them in the paper.

Before estimating the models for the conditional correlation we evaluated the constancy of correlation performing the LM test of Tse (2000) (See Chapter 1). We overwhelmingly reject the null of constant correlation with a p-value $= 0.000$, and in this case too we do not report the results.

Tables 2.3a,b report the correlation parameter estimates for the DCC and ADCC models, and the vector of asymmetry coefficients $\mathbf{b}$.

[Insert Tables 2.3a,b here]

We found that the parameter estimates of the DCC (1,1) model are very similar to those reported in CES. This is not the case of the ADCC model. For this model CES report a much higher level for the persistence parameter (0.94816 for the normal case against 0.5217 for the AML case).

The correlation parameter estimates for the GDCC and AGDCC specifications are reported in Tables 2.4a,b and Tables 2.5a,b. Overall, we found high levels of persistence but not as pronounced as in CES. For the case of normal innovations the range of the beta parameter in the GDCC model goes from 0.9186 (Canada shares) to 0.9759 (Austria bonds), while for the case of AML innovations is much more open; it goes from 0.1764 (Germany shares) to 0.9748 (New Zealand shares). The parameter estimates of the AGDCC model also show a higher degree of heterogeneity across indices when the AML distribution is assumed.

All asymmetric parameters in the AGDCC model were highly significant. Table 2.6 reports the Log-likelihood values for each one of the four models.

[Insert Tables 2.4a,b, 2.5a,b and 2.6 here]

In contrast to the case described in CES where the innovations are assumed normal, the inclusion of asymmetric terms or diagonal components does not increase the log-likelihood.

We follow Engle and Sheppard (2001) and employ the minimum variance portfolio criterion as a specification test of the models. We compare the variance of the portfolios formed by all the securities in the array $\mathbf{X}_t$ estimated with the eight models (four models assuming normality and four models assuming the AML distribution). The weight vector at time t for each one of the portfolios is given by,

$$^{m_i}\mathbf{w}_t = \frac{^{m_i}\mathbf{H}_t^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'^{m_i}\mathbf{H}_t^{-1}\boldsymbol{\iota}} \tag{2.29}$$

where $i = 1, ..., 8$ and $\boldsymbol{\iota}$ is an (n x 1) vector of ones. The variance of each portfolio will be given by $\mathbf{V}_t = \mathbf{w}_t'\mathbf{H}_t\mathbf{w}_t$. If $^{m_i}\mathbf{H}_t$ is accurately specified, then model m_i should give the minimum variance portfolio. Figures 2.8a and 2.8b show the eight series of $^{m_i}\mathbf{V}_t$ and Table 2.7 presents the average portfolio volatilities.

[Insert Table 2.7 here]
[Insert Figures 2.8a-2.8b here]

We also implement a Value-at-Risk (VaR) exercise as specification test for the models. Consider the portfolio return

$$\mathbf{r}_p = \sum_{i=1}^{n} w_i r_i = \mathbf{w}'\mathbf{r} \tag{2.30}$$

where $w_1 + ... + w_n = 1$. The VaR at the α level is the solution to,

$$\alpha = \int_{-\infty}^{VaR} f(\mathbf{r}_p)d\mathbf{r}_p \tag{2.31}$$

where $f(\mathbf{r}_p)$ is the density function of $\mathbf{r}_p$. In the special case where $f(\mathbf{r}_p)$ is the density of the AML distribution the conditional VaR implicit in (2.31) is

$$VaR_t = \left(\mathbf{w}_t'\mathbf{H}_t\mathbf{w}_t\right)^{1/2} L_\alpha \tag{2.32}$$

where L_α is the α-th quantile of the univariate standard Laplace distribution.

We consider three constant vectors of weights $\mathbf{w}$:

$$\mathbf{w^1} = \begin{bmatrix} w_1^1 = 0.035714286 \\ . \\ . \\ . \\ w_{21}^1 = 0.035714286 \\ w_{22}^1 = 0.019230769 \\ . \\ . \\ . \\ w_{34}^1 = 0.019230769 \end{bmatrix} \tag{2.33}$$

$$\mathbf{w^2} = \begin{bmatrix} w_1^2 = 0.029411765 \\ . \\ . \\ . \\ w_{34}^2 = 0.029411765 \end{bmatrix} \tag{2.34}$$

$$\mathbf{w^3} = \begin{bmatrix} w_1^3 = 0.011904762 \\ . \\ . \\ . \\ w_{21}^3 = 0.011904762 \\ w_{22}^3 = 0.057692308 \\ . \\ . \\ . \\ w_{34}^3 = 0.057692308 \end{bmatrix} \tag{2.35}$$

$\mathbf{w}^1$ corresponds to the case where the 21 FTSE All-World indices constitute the 75% of the portfolio and the 13 Bond indices constitute the remaining 25%. $\mathbf{w}^2$ corresponds to the case where the 34 indices have the same weight in the portfolio, and $\mathbf{w}^3$ corresponds to the case where the 21 FTSE All-World indices constitute the 25% of the portfolio and the 13 Bond indices constitute the remaining 75%. Given that we are considering four MGARCH models for the dynamics of $\mathbf{H}_t$ and two different innovation densities (the normal and the AML), we have in total 24 different variance portfolios to analyse.

We computed the conditional VaR for the 24 cases at the 1% level using the entire sample of 785 observations.

To evaluate the goodness of the conditional VaR under the dynamics of the different MGARCH models and alternative distributions we used the duration-based approach proposed by Christoffersen and Pelletier (2004). The Markov tests proposed in this work are designed to detect clustering in the violations of the VaR measures, where a violation is defined as the event where the ex-post portfolio loss exceeds the ex-ante VaR. Clearly, given the parametric model-based nature of the VaR methodology employed in this exercise, a correct dynamic specification of the portfolio volatility and a correct distribution for conditional returns are necessary to secure a right specification of the VaR technique.

We consider the unconditional coverage (*uc*), independence (*ind*), and conditional coverage (*cc*) test of Christoffersen and Pelletier (2004). Consider

the hit sequence of VaR violations defined as,

$$I_t = \begin{cases} 1, \text{if } r_t < -VaR(\alpha) \\ 0, \text{ else} \end{cases} \tag{2.36}$$

In the *uc* test we test the null hypothesis that I_t is i.i.d. Bernoulli with parameter α, against the alternative that the sequence is i.i.d Bernoulli with parameter π, where π is the ratio of the number of violations over the number of observations. If the VaR method is correct the empirical failure rate π must be equal to α.

The *ind* test tests explicitly the assumption of independence of the hit sequence,

$$H_{0,ind:}\pi_{01} = \pi_{11} \tag{2.37}$$

where π_{ij} is the probability of an i on day $t-1$ being followed by a j on day t. Neither the *uc* test nor the *ind* test are complete by their own, the first one test that on a average the coverage implicit by the VaR model is correct, while the second tests the clustering effect on the failures without testing the correct number of failures. The *cc* test combines both tests:

$$H_{0,cc:}\pi_{01} = \pi_{11} = \alpha \tag{2.38}$$

Under the null the likelihood ratio test of unconditional coverage (LR_{uc}) and the likelihood ratio test of independence (LR_{ind}) are χ^2 with one degree of freedom. Under the null the likelihood ratio test of conditional coverage (LR_{cc}) is χ^2 with two degrees of freedom.

Tables 2.8, 2.9, and 2.10 present the failure rates and p-values of the *uc*, *ind*, and *cc* tests for the four MGARCH models and for the three portfolios $\mathbf{w}^1, \mathbf{w}^2$, and $\mathbf{w}^3$.

[Insert Tables 2.8-2.10 here]

The performance of the scalar models (DCC and ADCC) across portfolios is very similar. The estimated VaR models in general capture quite well the clustering of violations. The models with AML innovations are superior to

the models with normal innovations for the cases of the $\mathbf{w}^1$ and $\mathbf{w}^2$ portfolios. For the $\mathbf{w}^1$ portfolio the results for the independence test are quite mixed.

For the $\mathbf{w}^1$ portfolio we found a very poor performance of the models regarding the unconditional and conditional coverage.

Plots with the distribution of conditional correlations for four pairs of correlations as well as descriptive statistics for the estimated series are presented in Figures 2.9-2.12.

[Insert Figures 2.9-2.12 here]

First, we observe that the distribution of the correlation across models changes significantly. Across the four pairs the ADCC model with AML innovations highlights for its extreme level of kurtosis (225.11 for the UK shares-US shares pair, 24.79 for the Japan bonds-UK bonds pair, 30.49 for the UK shares-Mexico shares pair, and 20.92 for the UK bonds-Switzerland bonds pair). This leptokurtosis is a result of very small volatilities (0.84%, 0.72%, 0.82%, and 0.52% respectively) and one single positive jump registered on Black Monday in 1987.

Plots with the conditional correlation series for four pairs of correlations are presented in Figures 2.13-2.16.

[Insert Figures 2.13-2.16 here]

In general, the kurtosis registered for the dynamic correlation estimated assuming an AML distribution is higher than that estimated assuming normality. In Table 2.11 we present a comparison of the levels of kurtosis between correlations assuming the two types of distributions for the asset pairs considered in Figures 2.9 to 2.16.

[Insert Table 2.11 here]

2.7 Conclusions

In this chapter we proposed a multivariate (GARCH) asymmetric generalised dynamic conditional correlation model where the vector of standardised residuals is assumed to follow an asymmetric multivariate Laplace distribution.

This multivariate distribution is able to capture leptokurtosis and asymmetry which characterise returns from financial assets. This is the only distribution (besides the normal) with desirable properties such as additivity and finiteness of moments. In addition, contrary to the majority of (geometric) stable distributions, it has a density function with a closed-form that makes the maximum likelihood estimation method easy to implement. Very importantly, we show that the two-step approach of the DCC model is preserved when innovations are modelled via non-normal multivariate distributions.

The empirical validity of the model we propose is tested by fitting the sample of 21 FTSE All-World stock indices and 12 bond return indices of Cappiello, Engle and Sheppard (2004). We provide clear evidence that this distribution overwhelmingly outperforms the case in which we assume normality of innovations. The empirical validity of this form is also tested in the context of a Value-at-Risk (VaR) model. By performing a conditional-VaR analysis, we obtained mixed results. Though all models capture quite well the clustering of violations of the VaR levels, they performed quite poorly when they were tested for the level of failure rates. But when we evaluate the independence of hit sequences, once again the models with asymmetric multivariate Laplace innovations outperformed models where normality of the innovations is assumed.

The empirical application presented in this chapter was carried out in order to compare and evaluate the behaviour of parameter estimates computed with several specifications. In Chapter 4 we report a more comprehensive set of empirical implementations to risk management to examine the benefits of our proposed framwork.

Table 2.1a: Parameter estimates for the univariate GARCH models, and skewness and kurtosis of the returns standardised by their estimated standard deviation using the AML distribution (shares).

	ϖ	α	β	Stand. Skew.	Stand. Kurt.
Australia	0.000019	0.0296	0.9466	-1.25	11.06
Austria	0.000051	0.1092	0.8452	-0.36	4.21
Belgium	0.000013	0.0497	0.9257	-0.47	4.51
Canada	0.000031	0.0800	0.8650	-1.09	10.77
Denmark	0.000034	0.0771	0.8802	0.05	4.01
France	0.000056	0.0733	0.8487	-0.25	3.58
Germany	0.000019	0.0527	0.9194	-0.56	4.90
H.K.	0.000220	0.1342	0.7470	-1.21	9.30
Ireland	0.000022	0.0395	0.9389	-1.05	10.71
Italy	0.000013	0.0362	0.9590	-0.20	4.37
Japan	0.000151	0.1540	0.7623	0.05	3.90
Mexico	0.000174	0.1015	0.8550	-0.49	6.56
Netherlands	0.000021	0.4681	0.9064	-1.11	9.84
New Zealand	0.000207	0.0380	0.7758	-0.58	6.05
Norway	0.000049	0.0671	0.8923	-1.026	10.82
Singapore	0.000142	0.1282	0.7934	-1.25	11.58
Spain	0.000102	0.0876	0.8068	-0.50	5.77
Sweden	0.000053	0.0787	0.8833	-0.55	5.43
Switzerland	0.000051	0.0340	0.8846	-1.18	13.77
U.K.	0.000015	0.0539	0.9181	-0.99	7.72
U.S.A.	0.000039	0.0308	0.8951	-1.35	14.67

Table 2.1b: Parameter estimates for the univariate GARCH models, and skewness and kurtosis of the returns standardised by their estimated standard deviation using the AML distribution (bonds).

	ϖ	α	β	Stand. Skew.	Stand. Kurt.
Austria	0.000072	0.0818	0.529	0.13	3.62
Belgium	0.000021	0.0705	0.8013	0.12	3.73
Canada	0.000045	0.1466	0.5880	0.00	3.70
Denmark	0.000009	0.0466	0.8975	0.11	3.69
France	0.000042	0.0780	0.6891	0.09	3.27
Germany	0.000062	0.0847	0.5846	0.22	3.86
Ireland	0.000016	0.0653	0.8543	-0.29	4.15
Japan	0.000057	0.0854	0.7169	0.59	5.76
Netherlands	0.000022	0.0495	0.8208	0.26	4.07
Sweden	0.000001	0.0316	0.9638	0.01	3.42
Switzerland	0.000136	0.0638	0.3963	0.24	3.87
U.K.	0.000003	0.0318	0.9507	-0.25	5.113
U.S.A.	0.000020	0.0710	0.4411	0.64	8.342

Table 2.2a. Kolmogorov-Smirnov statistics for shares.

	Normal	Laplace
Australia	0.0529	0.0622
Austria	0.0445	0.0501
Belgium	0.0559	0.0665
Canada	0.0721	0.0577
Denmark	0.0524	0.0485
France	0.0510	0.0717
Germany	0.0557	0.0850
H.K.	0.0744	0.0808
Ireland	0.066	0.0592
Italy	0.029	0.0497
Japan	0.0386	0.0434
Mexico	0.0832	0.0432
Netherlands	0.0758	0.0762
New Zealand	0.0443	0.0425
Norway	0.0661	0.0599
Singapore	0.0690	0.0501
Spain	0.0520	0.0454
Sweden	0.0614	0.0681
Switzerland	0.0786	0.0556
U.K.	0.0965	0.0685
U.S.A.	0.0650	0.0589

Table 2.2b. Kolmogorov-Smirnov statistics for bonds.

	Normal	Laplace
Austria	0.0557	0.0769
Belgium	0.0545	0.0707
Canada	0.0751	0.0483
Denmark	0.0619	0.0600
France	0.0483	0.0658
Germany	0.0490	0.0745
Ireland	0.0709	0.0667
Japan	0.0575	0.0724
Netherlands	0.0489	0.0694
Sweden	0.0602	0.0509
Switzerland	0.0420	0.0616
U.K.	0.0729	0.0527

Table 2.3a: Parameter estimates for the DCC(1,1) model and ADCC(1,1,1) models using the AML distribution.

	DCC(1,1)	ADCC(1,1,1)
α	0.00987	0.0045
β	0.9571	0.5217
γ		0.0863
Shares	β_i	b
Australia	0.2617	0.0000
Austria	-0.3961	0.0000
Belgium	-0.0526	0.0000
Canada	-0.5051	0.0000
Denmark	0.5120	0.0000
France	0.4909	0.0000
Germany	0.0861	0.0000
H.K.	-1.1099	0.0000
Ireland	0.2934	0.0000
Italy	0.4822	0.0000
Japan	0.0931	0.0000
Mexico	0.1047	0.0000
Netherlands	-1.1287	0.0000
New Zealand	0.2909	0.0000
Norway	0.0544	0.0000
Singapore	-0.1068	0.0000
Spain	-0.3626	0.0001
Sweden	0.0493	0.0000
Switzerland	-0.0780	0.0000
U.K.	-0.5037	0.0000
U.S.A.	0.6387	0.0000

Notes to Table 2.3a: DCC(1,1) model: $Q_t = (\overline{\mathbf{Q}} - \boldsymbol{\alpha}\overline{\mathbf{Q}} - \boldsymbol{\beta}\overline{\mathbf{Q}}) + \alpha\varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1}$; ADCC(1,1,1) model: $Q_t = (\overline{\mathbf{Q}} - \boldsymbol{\alpha}\overline{\mathbf{Q}} - \boldsymbol{\beta}\overline{\mathbf{Q}} - \gamma\overline{\mathbf{N}}) + \alpha\varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1} + \gamma\eta_{t-1}\eta'_{t-1}$

Table 2.3b: Parameter estimates for the DCC(1,1) model and ADCC(1,1,1) models using the AML distribution.

	DCC(1,1)	ADCC(1,1,1)
α	0.00987	0.0045
β	0.9571	0.5217
γ		0.0863
Bonds	β_i	b
Austria	-2.3187	0.0000
Belgium	0.9697	0.0000
Canada	0.2407	0.0000
Denmark	0.4694	0.0000
France	-2.641	0.0000
Germany	5.4709	0.0000
Ireland	-3.5647	0.0000
Japan	1.0291	1.1962
Netherlands	0.7855	0.0000
Sweden	0.5523	0.0000
Switzerland	-0.4615	0.0000
U.K.	-0.8344	0.0000
U.S.A.	2.7438	1.8035

Notes to Table 2.3b: DCC(1,1) model: $Q_t = (\overline{\mathbf{Q}} - \boldsymbol{\alpha}\overline{\mathbf{Q}} - \boldsymbol{\beta}\overline{\mathbf{Q}}) + \alpha\varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1}$; ADCC(1,1,1) model: $Q_t = (\overline{\mathbf{Q}} - \boldsymbol{\alpha}\overline{\mathbf{Q}} - \boldsymbol{\beta}\overline{\mathbf{Q}} - \gamma\overline{\mathbf{N}}) + \alpha\varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1} + \gamma\eta_{t-1}\eta'_{t-1}$

Table 2.4a. Parameter estimates for the GDCC (1,1) model using the AML distribution.

	GDCC(1,1)		
Shares	α_i	β_i	b
Australia	0.0130	0.9103	0.0119
Austria	0.0205	0.9793	0.0000
Belgium	0.0002*	0.5437	0.0000
Canada	0.3563	0.6414	0.1949
Denmark	0.0766	0.7250	0.0000
France	0.0760	0.8274	0.0000
Germany	0.0147*	0.1764	0.0000
H.K.	0.0004*	0.4351	0.0072
Ireland	0.0090	0.7565	0.0000
Italy	0.3110	0.5661	0.0000
Japan	0.1250	0.3615	0.0000
Mexico	0.0000*	0.8883	0.0000
Netherlands	0.0000*	0.7923	0.0000
New Zealand	0.0252	0.9748	0.0113
Norway	0.0002	0.6731	0.0128
Singapore	0.0007*	0.8896	0.1428
Spain	0.0150	0.5150	0.0000
Sweden	0.0222	0.7109	0.0000
Switzerland	0.0595	0.2155	0.0000
U.K.	0.0456	0.9498	0.0000
U.S.A.	0.0399	0.9202	0.0000

Notes to Table 2.4a: GDCC model: $Q_t = \left(\overline{\mathbf{Q}} - \mathbf{A}'\overline{\mathbf{Q}}\mathbf{A} - \mathbf{B}'\overline{\mathbf{Q}}\mathbf{B} - \mathbf{G}'\overline{\mathbf{N}}\mathbf{G}\right) + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'Q_{t-1}B$

Table 2.4b. Parameter estimates for the GDCC (1,1) model using the AML distribution.

	GDCC(1,1)		
Bonds	α_i	β_i	b
Austria	0.0735	0.7585	0.0000
Belgium	0.0318	0.8927	0.0000
Canada	0.0208	0.9786	0.4967
Denmark	0.0550	0.8268	0.0000
France	0.0627	0.8444	0.0000
Germany	0.1022	0.7773	0.0000
Ireland	0.1536	0.8069	0.0000
Japan	0.0888	0.8189	0.0756
Netherlands	0.0950	0.7977	0.0000
Sweden	0.0003*	0.8813	0.0000
Switzerland	0.0318	0.9670	0.0000
U.K.	0.0228	0.9203	0.0000
U.S.A.	0.0082	0.9896	3.1433

Notes to Table 2.4a: GDCC model: $Q_t = \left(\overline{\mathbf{Q}} - \mathbf{A}'\overline{\mathbf{Q}}\mathbf{A} - \mathbf{B}'\overline{\mathbf{Q}}\mathbf{B} - \mathbf{G}'\overline{\mathbf{N}}\mathbf{G}\right) + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'Q_{t-1}B$

Table 2.5a. Parameter estimates for the AGDCC (1,1,1) model using the AML distribution.

Shares	α_i	g_i	β_i	b
Australia	0.0006*	0.0568	0.8440	0.0000
Austria	0.0027*	0.0742	0.9231	0.0000
Belgium	0.0002*	0.1990	0.6655	0.0000
Canada	0.0015*	0.2283	0.7701	0.0000
Denmark	0.0050	0.0946	0.9003	0.0000
France	0.0050	0.0864	0.8346	0.0000
Germany	0.0004*	0.1369	0.6508	0.0000
H.K.	0.1014	0.0053*	0.6299	0.0000
Ireland	0.0002*	0.1135	0.6294	0.0000
Italy	0.0025*	0.0309	0.9666	0.0000
Japan	0.0260	0.0000	0.9740	0.5942
Mexico	0.0376*	0.0080	0.7502	0.0000
Netherlands	0.0145	0.0524	0.7463	0.0000
New Zealand	0.0002*	0.0093	0.9905	0.0000
Norway	0.0002	0.0073	0.7881	0.0000
Singapore	0.0518	0.0201	0.9280	0.0000
Spain	0.0237	0.0105*	0.5845	0.0000
Sweden	0.0386	0.0540	0.8854	0.0000
Switzerland	0.0171	0.1301	0.8528	0.0000
U.K.	0.0822	0.0228	0.6062	0.0000
U.S.A.	0.0070*	0.0379	0.8427	0.0000

Note to Table 2.5a: AGDCC (1,1,1) model: $Q_t = \left(\overline{\mathbf{Q}} - \mathbf{A}'\overline{\mathbf{Q}}\mathbf{A} - \mathbf{B}'\overline{\mathbf{Q}}\mathbf{B} - \mathbf{G}'\overline{\mathbf{N}}\mathbf{G}\right) + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'Q_{t-1}B + G'\eta_{t-1}\eta'_{t-1}G$

Table 2.5b. Parameter estimates for the AGDCC (1,1,1) model using the AML distribution.

Bonds	α_i	g_i	β_i	b
Austria	0.1238	0.1482	0.6814	0.0000
Belgium	0.0854	0.1572	0.7059	0.0000
Canada	0.0238	0.0000*	0.9762	0.0000
Denmark	0.0641	0.1767	0.7592	0.0000
France	0.2588	0.0000*	0.6494	0.0000
Germany	0.1305	0.1449	0.6906	0.0000
Ireland	0.0796	0.1684	0.7518	0.0000
Japan	0.0004*	0.0000*	0.5485	0.0000
Netherlands	0.1268	0.1473	0.7247	0.0000
Sweden	0.0003*	0.1929	0.6866	0.0000
Switzerland	0.0036*	0.2820	0.7142	0.0000
U.K.	0.0007*	0.2625	0.7002	0.0000
U.S.A.	0.0009*	0.0345	0.8109	2.1780

Note to Table 2.5b: AGDCC (1,1,1) model used: $Q_t = \left(\overline{\mathbf{Q}} - \mathbf{A}'\overline{\mathbf{Q}}\mathbf{A} - \mathbf{B}'\overline{\mathbf{Q}}\mathbf{B} - \mathbf{G}'\overline{\mathbf{N}}\mathbf{G}\right) + A'\varepsilon_{t-1}\varepsilon'_{t-1}A + B'Q_{t-1}B + G'\eta_{t-1}\eta'_{t-1}G$

Table 2.6 . Log-likelihood values for the four estimated models

Model	Log-likelihood
DCC(1,1)	-24027
ADCC(1,1,1)	-25381
GDCC(1,1)	-25269
AGDCC(1,1,1)	-26808

Table 2.7. Average variance of the portfolios with alternative models

Model	Average Variance
AML-DCC	2.64E-005
AML-ADCC	2.28E-005
AML-GDCC	2.61E-005
AML-AGDCC	2.28E-005

Note to Table 2.7: Average variance of the portfolios formed by all securities in the sample data estimated with the four models assuming the AML distribution.

Table 2.8. VaR Analysis: Portfolio w1

PORTFOLIO W1	Failure rate	uc	ind	cc
Normal-DCC	0.015287	0.16723	0.010686	0.014813
Laplace-DCC	0.0076433	0.48872	0.032332	0.079653
Normal-ADCC	0.014013	0.28667	0.0071606	0.015242
Laplace-ADCC	0.0076433	0.48872	0.032332	0.079653
Normal-GDCC	0.024204	0.00071497	0.079946	0.00070455
Laplace-GDCC	0.011465	0.68686	0.002787	0.010544
Normal-AGDCC	0.015287	0.16723	0.010686	0.014813
Laplace-AGDCC	0.0076433	0.48872	0.032332	0.079653

Note to Table 2.8: Portfolio w1: 75% formed by the 21 FTSE All-World indices and 25% by the 13 Bond indices. VaR(1%) failure rates and p-values for the hit sequence Markov tests: unconditional coverage (uc) test, independence (ind) test, and conditional coverage (cc)

Table 2.9. VaR Analysis: Portfolio w2

PORTFOLIO W2	Failure rate	uc	ind	cc
Normal-DCC	0.012739	0.45939	0.0045904	0.013695
Laplace-DCC	0.0076433	0.48872	0.032332	0.079653
Normal-ADCC	0.012739	0.45939	0.0045904	0.013695
Laplace-ADCC	0.0076433	0.48872	0.032332	0.079653
Normal-GDCC	0.025478	0.00026545	0.098754	0.00033125
Laplace-GDCC	0.012739	0.45939	0.0045904	0.013695
Normal-AGDCC	0.014013	0.28667	0.57605	0.48484
Laplace-AGDCC	0.0076433	0.48872	0.032332	0.079653

Note to Table 2.9: Portfolio w2: The 34 assets have the same weight. VaR (1%) failure rates and p-values for the hit sequence Markov tests: unconditional coverage (uc) test, independence (ind) test, and conditional coverage (cc)

Table 2.10. VaR Analysis: Portfolio w3

PORTFOLIO W3	Failure rate	uc	ind	cc
Normal-DCC	0.012739	0.45939	0.0045904	0.013695
Laplace-DCC	0.0038217	0.046627	0.87941	0.13657
Normal-ADCC	0.012739	0.45939	0.61145	0.66849
Laplace-ADCC	0.0050955	0.12729	0.83959	0.30632
Normal-GDCC	0.017834	0.046926	0.47581	0.10772
Laplace-GDCC	0.0089172	0.75609	0.72265	0.89474
Normal-AGDCC	0.011465	0.68686	0.64774	0.83058
Laplace-AGDCC	0.0050955	0.12729	0.83959	0.30632

Note to Table 2.10: Portfolio w3: 25% formed by the 21 FTSE All-World indices and 75% by the 13 Bond indices. VaR (1%) failure rates and p-values for the hit sequence Markov tests: unconditional coverage (uc) test, independence (ind) test, and conditional coverage (cc)

Table 2.11. Kurtosis of the conditional correlations.

Model	Pair 1	Pair 2	Pair 3	Pair 4
Normal-DCC	3.93	2.15	3.14	2.97
AML-DCC	5.55	2.34	3.58	3.00
Normal-ADCC	11.69	5.66	8.50	6.04
AML-ADCC	225.11	24.79	30.49	20.92
Normal-GDCC	3.42	2.06	2.82	2.93
AML-GDCC	8.79	3.45	7.55	6.20
Normal-AGDCC	5.03	5.29	7.09	4.60
AML-AGDCC	4.97	8.34	6.05	4.12

Note to Table 2.11: Kurtosis of the conditional correlation series created with the eight models for four pairs of assets. Pair 1 corresponds to U.K. shares-U.S.A. shares, Pair 2 to U.K. bonds-Japan bonds, Pair 3 to U.K. sahres-Mexico shares, and Pair 4 to U.K. bonds-Switzerland bonds

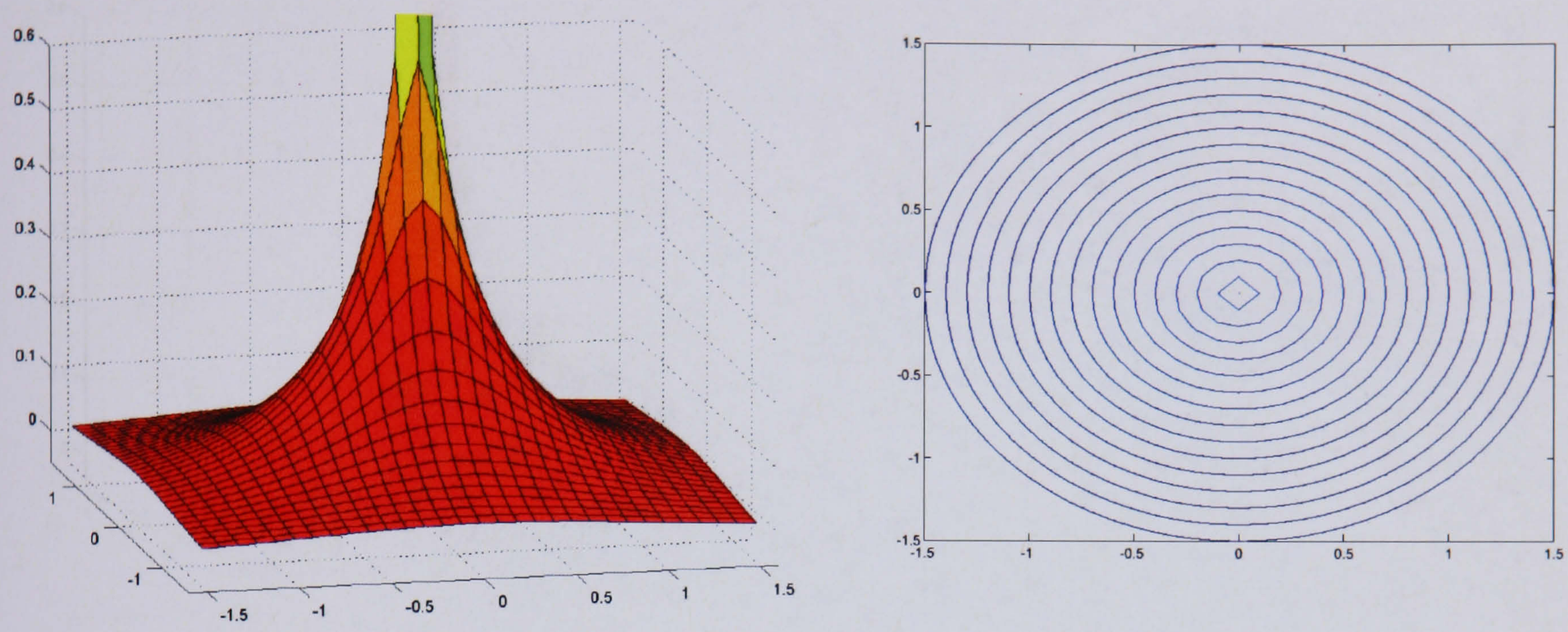


Figure 2.1. Bivariate Asymmetric Laplace density and contours with $m_1 = m_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0$

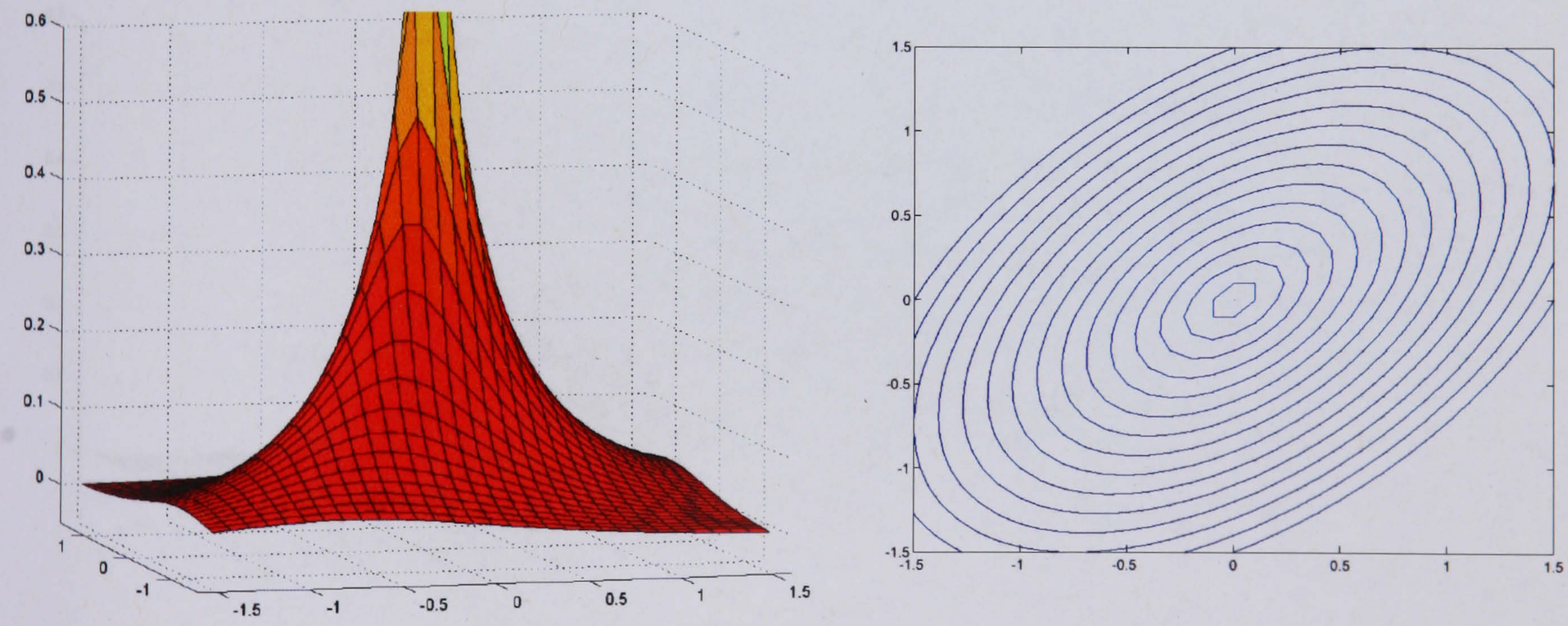


Figure 2.2. Bivariate Asymmetric Laplace density and contours with $m_1 = m_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0.5$

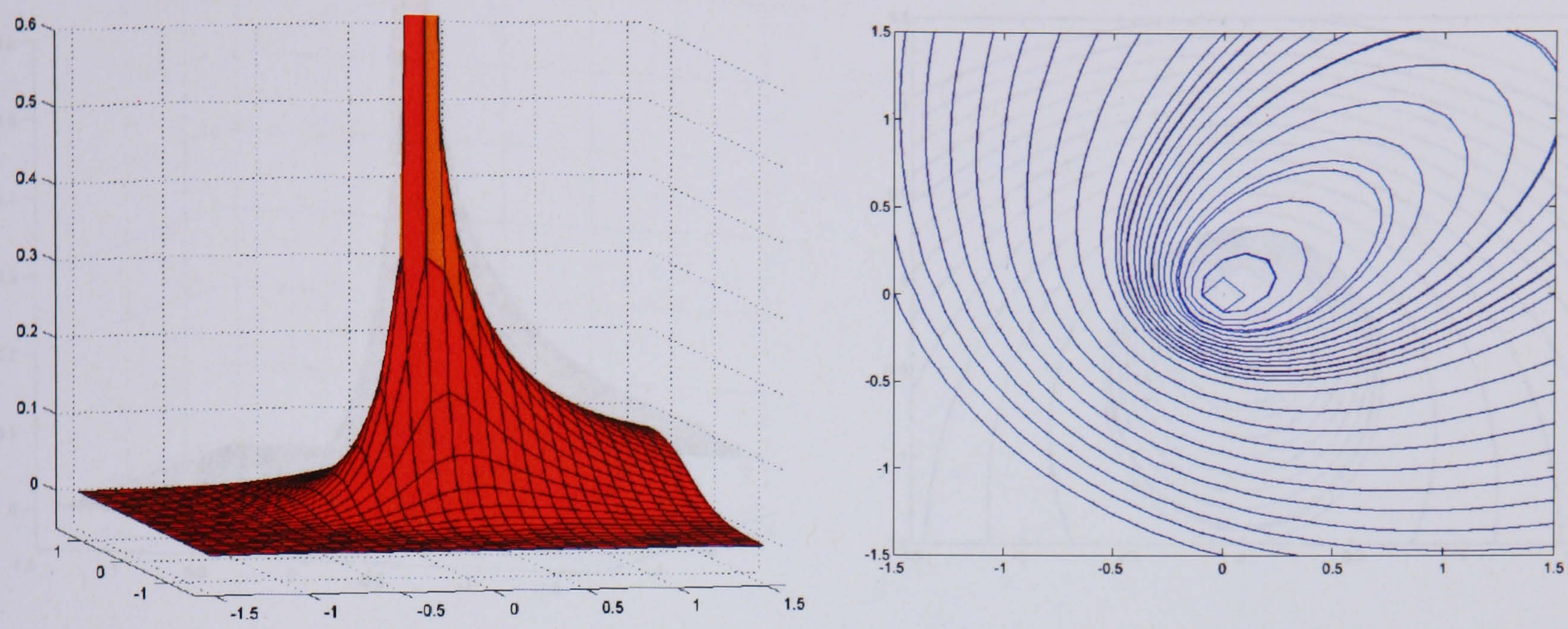


Figure 2.3. Bivariate Asymmetric Laplace density and contours with $m_1 = m_2 = 2$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0$

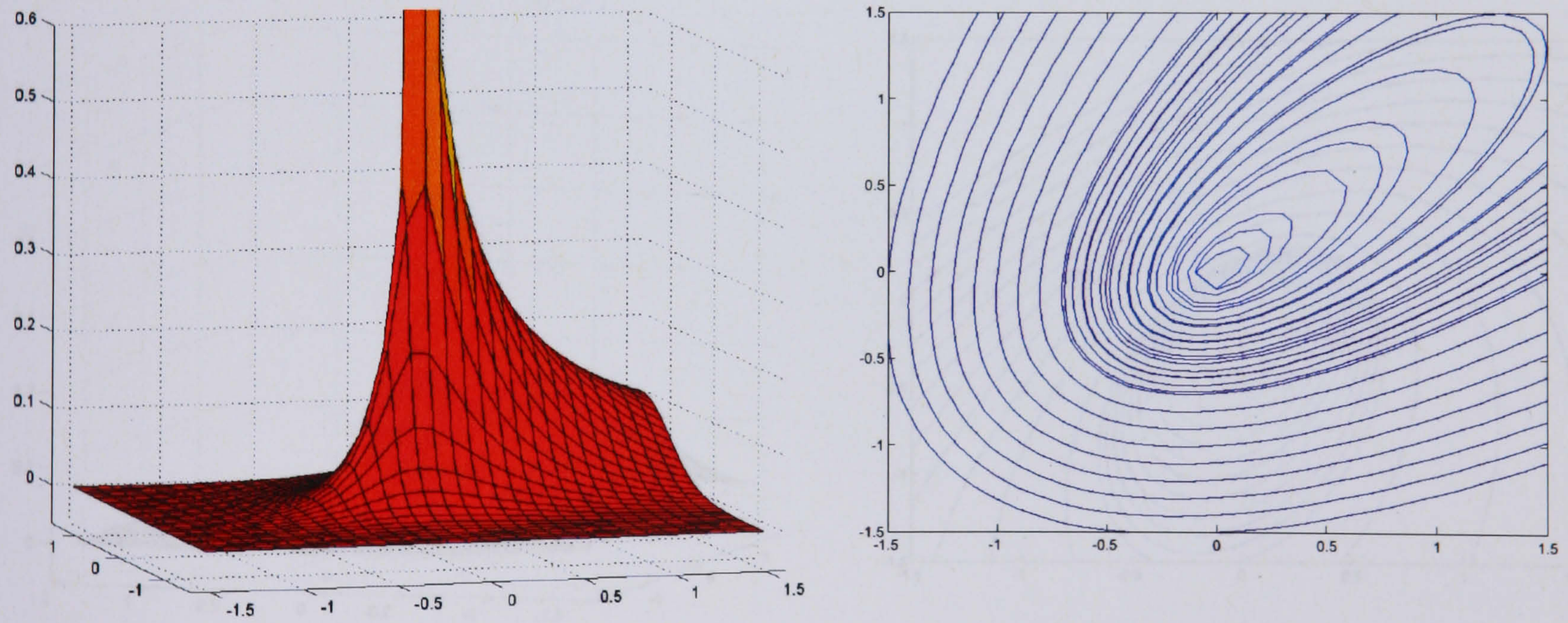


Figure 2.4. Bivariate Asymmetric Laplace density and contours with $m_1 = m_2 = 2$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0.5$

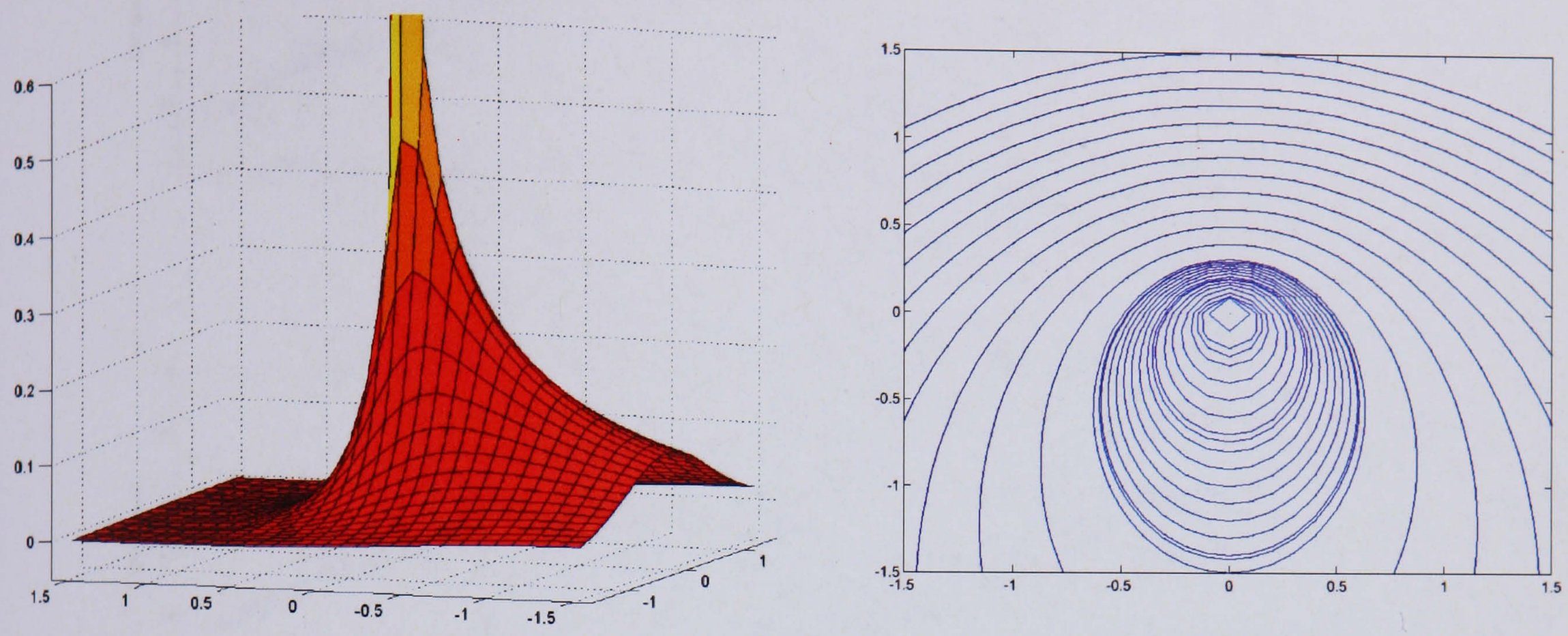


Figure 2.5. Bivariate Asymmetric Laplace density and contours with $m_1 = -2$, $m_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0$

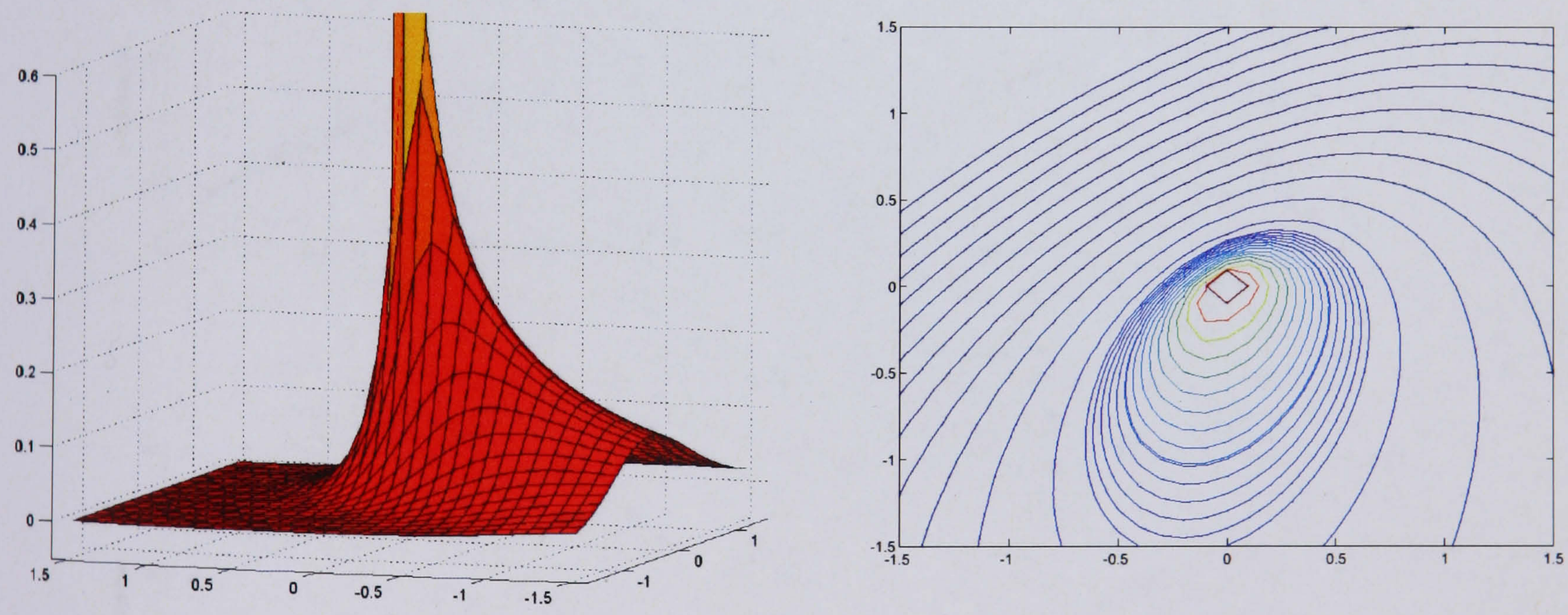


Figure 2.6. Bivariate Asymmetric Laplace density and contours with $m_1 = -2$, $m_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and $\rho = 0.5$

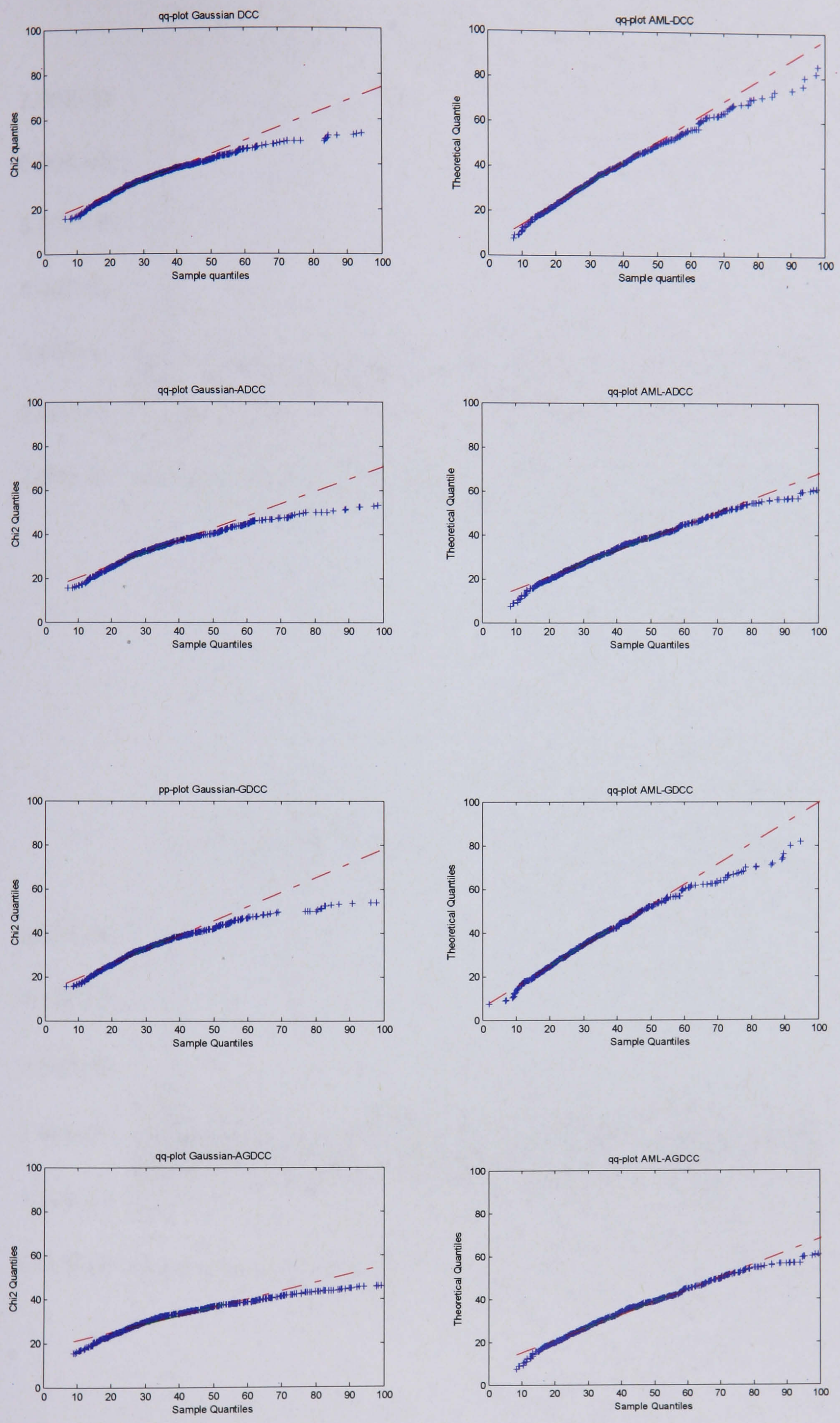
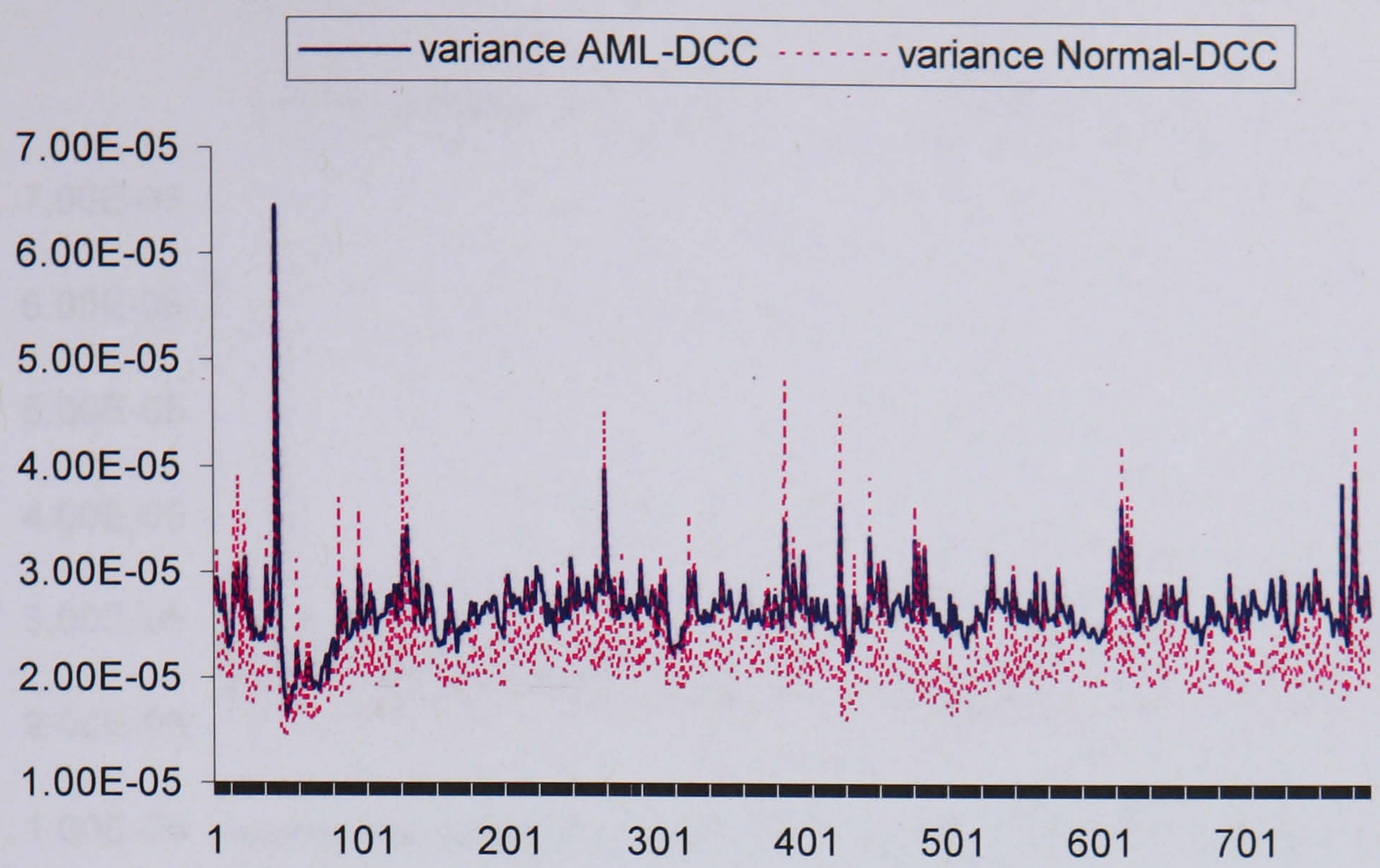


Figure 2.7. qq-plots of the distribution of $r_t H_t^{-1} r'_t$ from the multivariate normal MLE (left hand side plots) and from the AML MLE (right hand side plots) for the four type of DCC models.



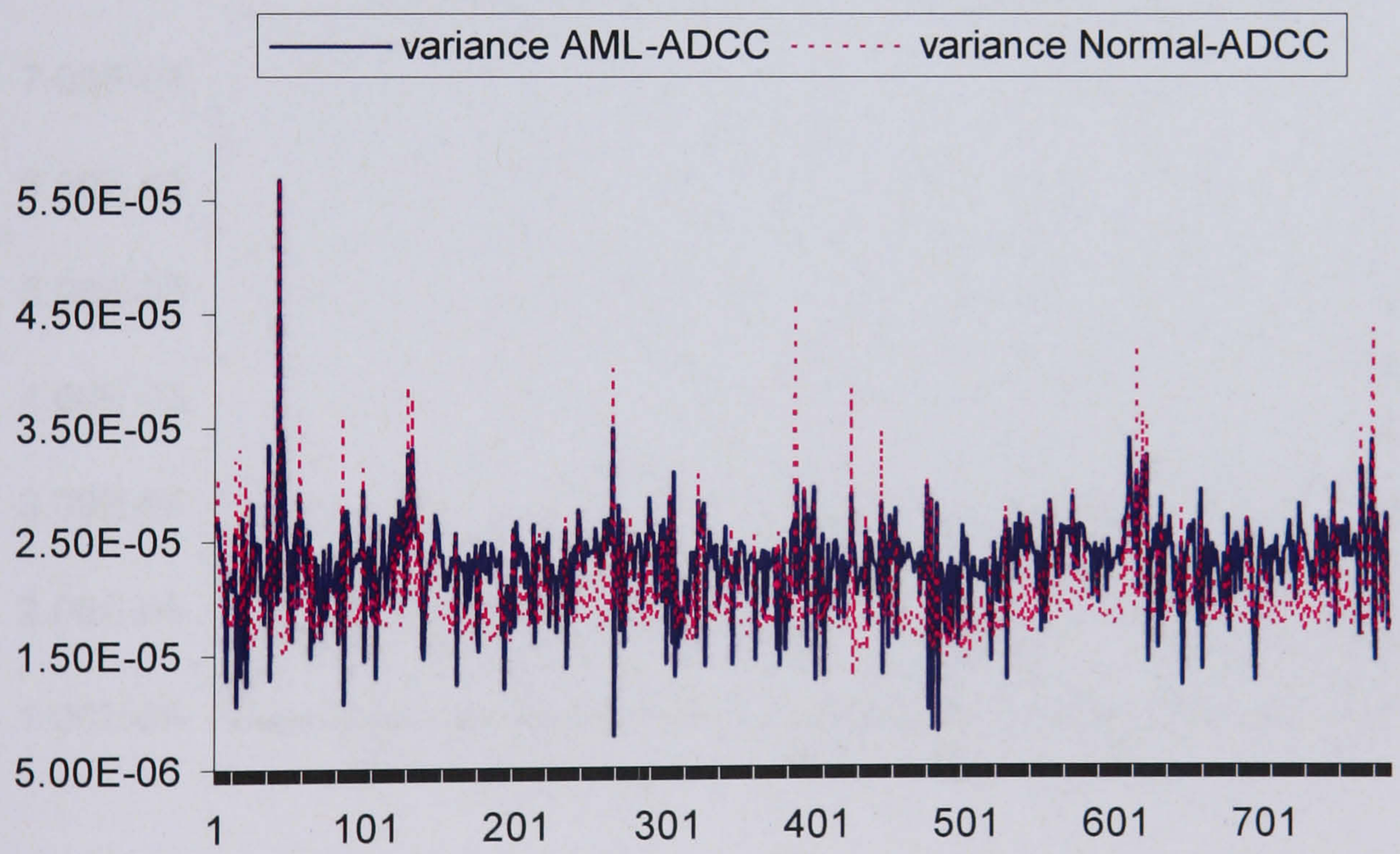
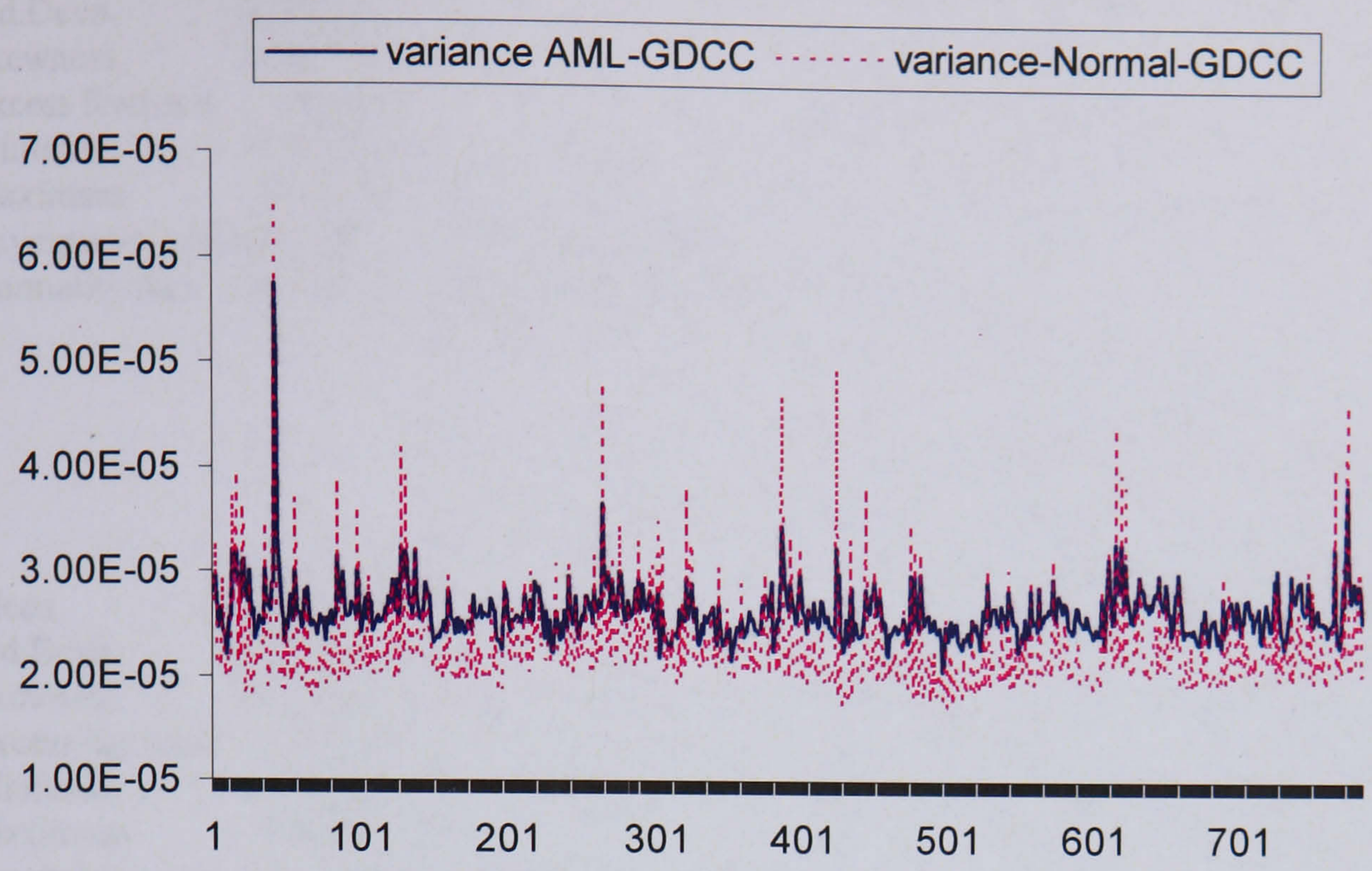


Figure 2.8A. Series of the variances of the portfolios composed of all assets in the sample data for the AML-DCC, Normal-DCC, AML-ADCC and Normal-ADCC models.



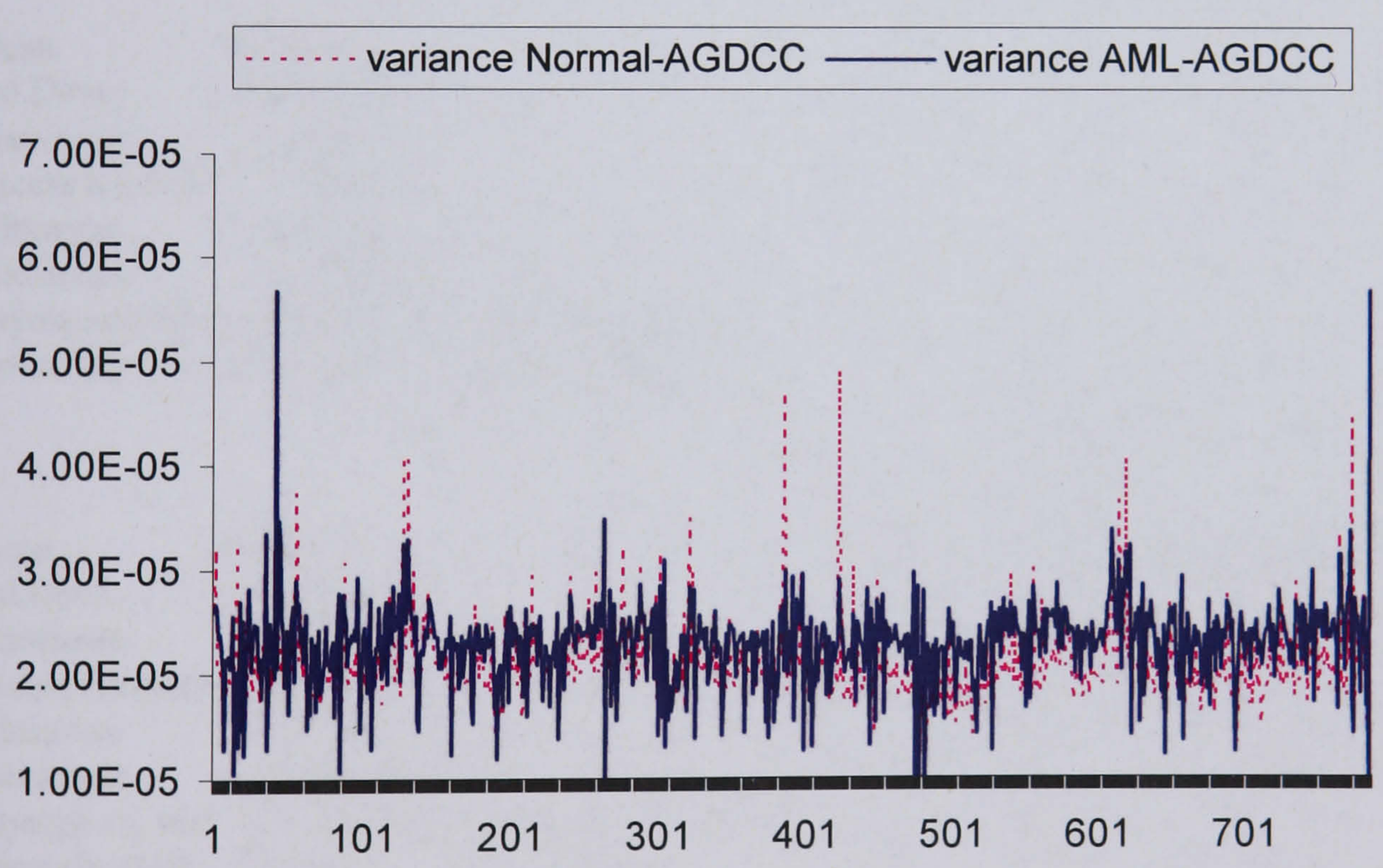
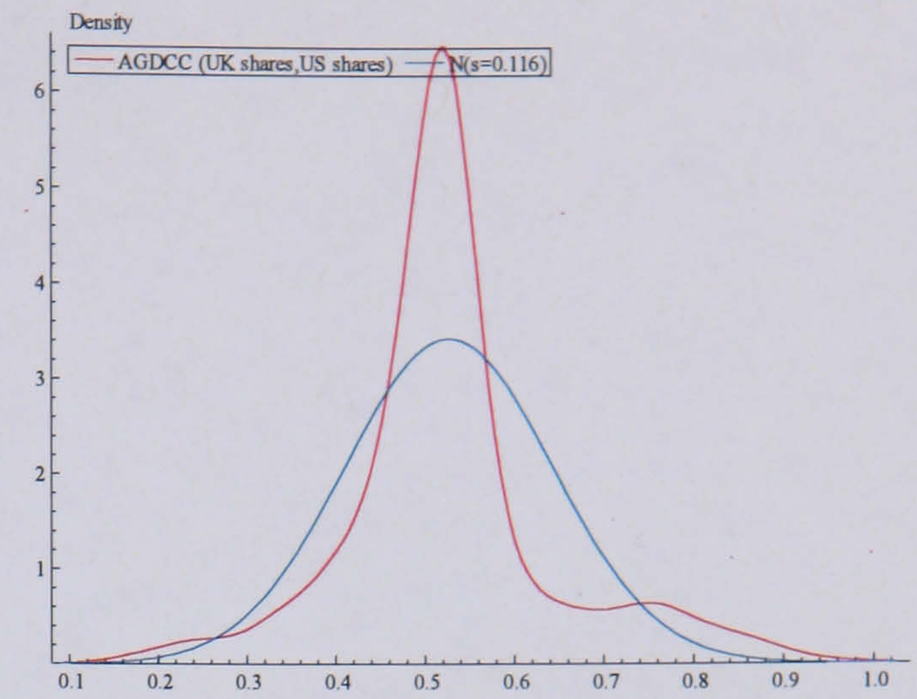
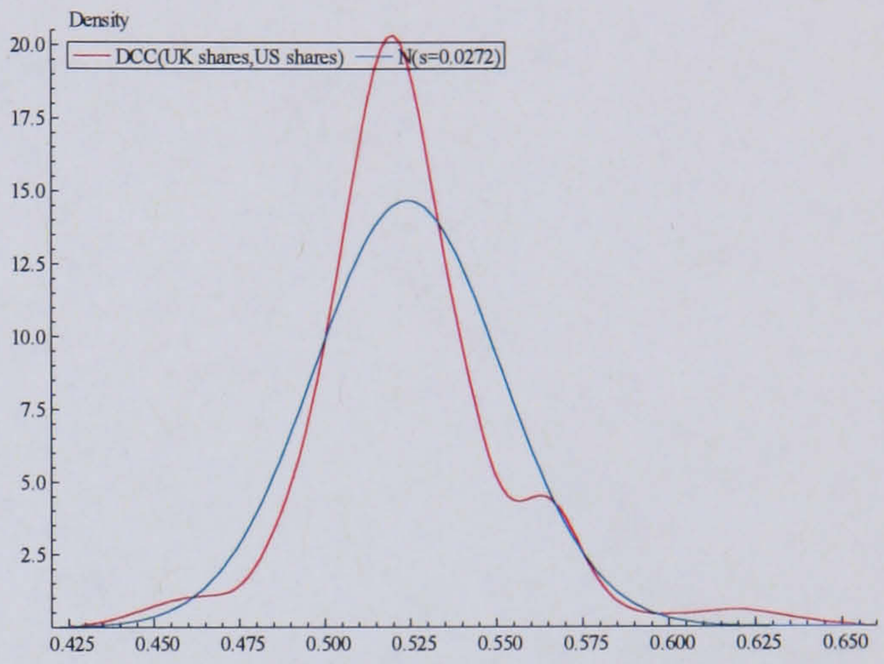


Figure 2.8B. Series of the variances of the portfolios composed of all assets in the sample data for the AML-GDCC, Normal-GDCC, AML-AGDCC and Normal-AGDCC models

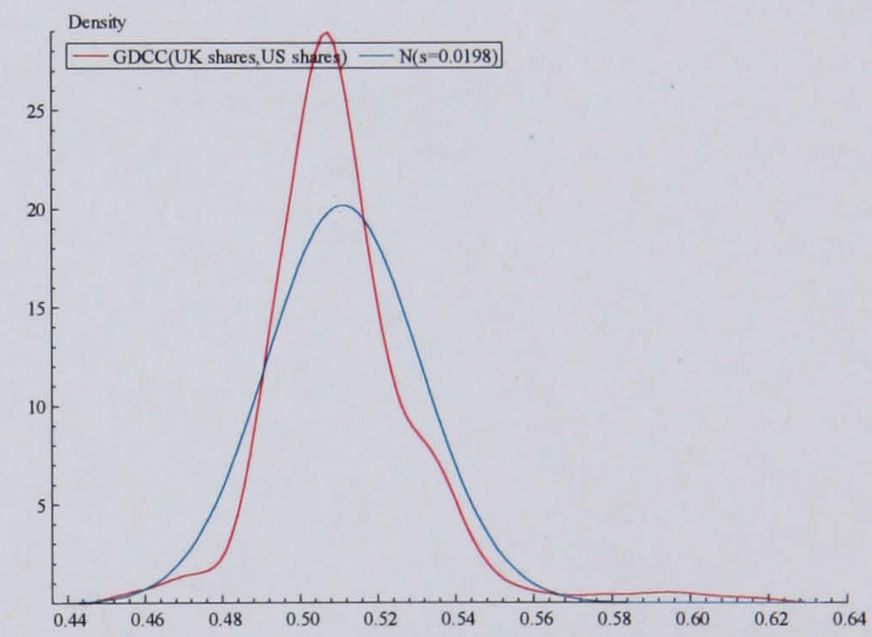
Mean 0.52519
Std.Devn. 0.11643
Skewness 0.59744
Excess Kurtosis 1.9738
Minimum 0.16910
Maximum 0.95840
Asymptotic test: Chi^2(2) = 174.12 [0.0000]**
Normality test: Chi^2(2) = 60.841 [0.0000]**



Mean 0.52390
Std.Devn. 0.027194
Skewness 0.72778
Excess Kurtosis 2.5516
Minimum 0.43860
Maximum 0.64040
Asymptotic test: Chi^2(2) = 282.24 [0.0000]**
Normality test: Chi^2(2) = 78.158 [0.0000]**



Mean 0.51106
Std.Devn. 0.019759
Skewness 1.5059
Excess Kurtosis 5.7961
Minimum 0.45240
Maximum 0.61900
Asymptotic test: Chi^2(2) = 1395.5 [0.0000]**
Normality test: Chi^2(2) = 175.42 [0.0000]**



Mean 0.52361
Std.Devn. 0.0084942
Skewness 10.454
Excess Kurtosis 222.11
Minimum 0.47520
Maximum 0.69710
Asymptotic test: Chi^2(2) =1.6279e+006 [0.0000]**
Normality test: Chi^2(2) = 5200.4 [0.0000]**

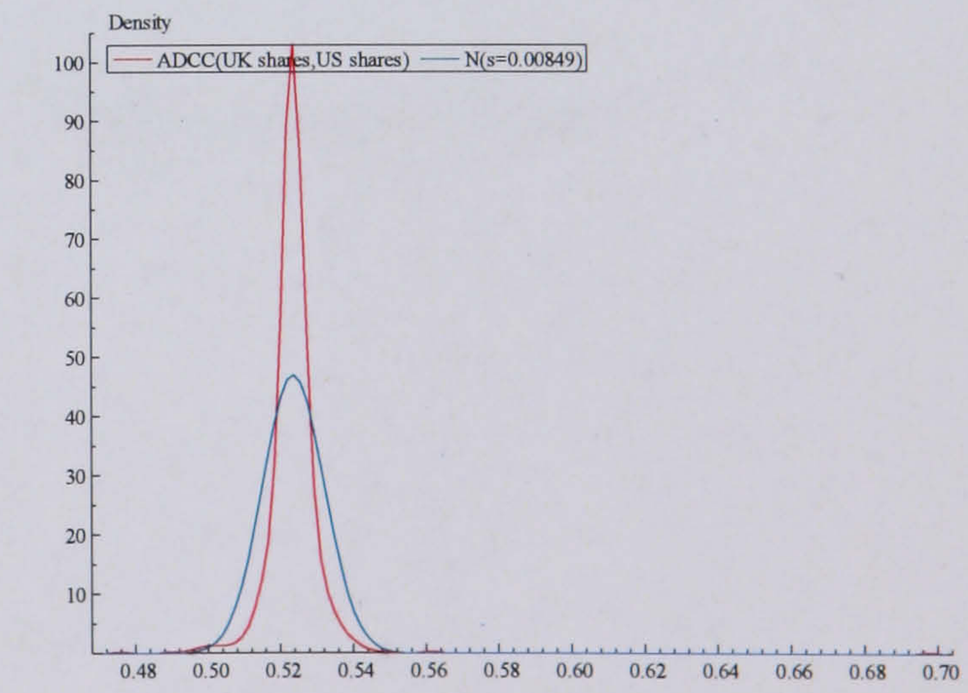
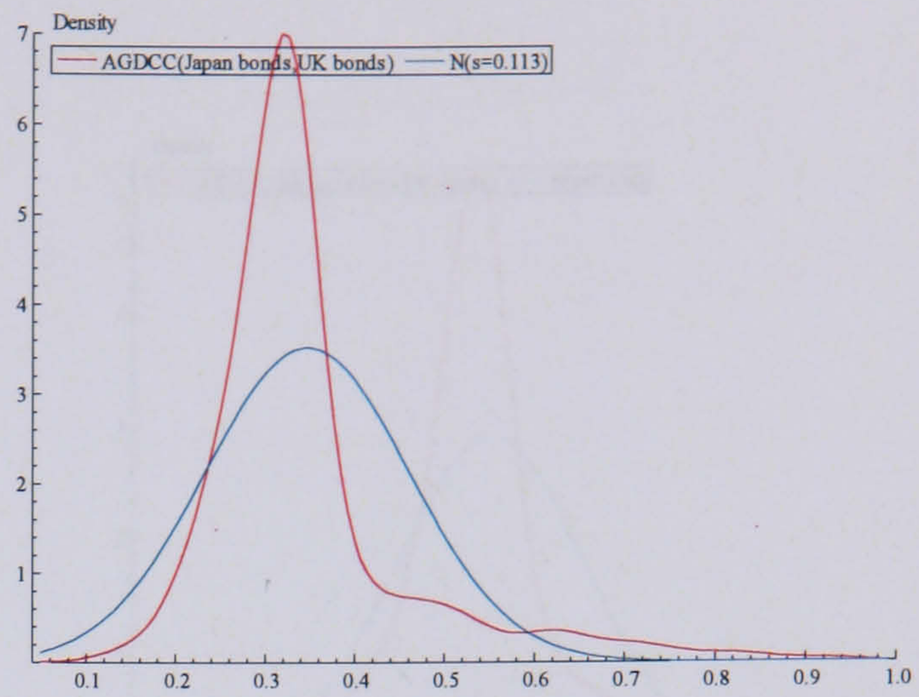
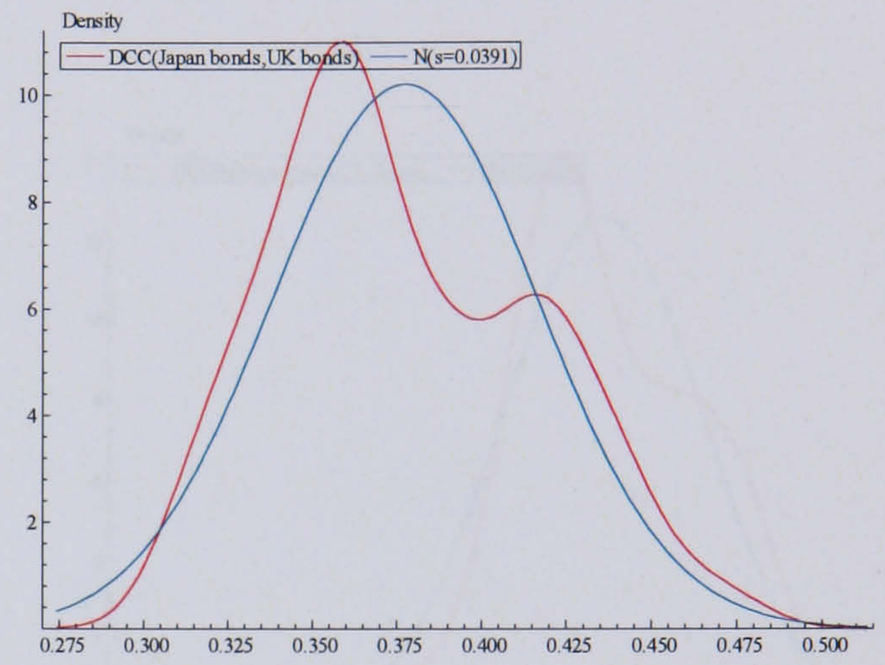


Figure 2.9. Distribution of the dynamic correlation between the return of UK shares and US shares employing the AGDCC, DCC, GDCC, and ADCC models with AML distributions.

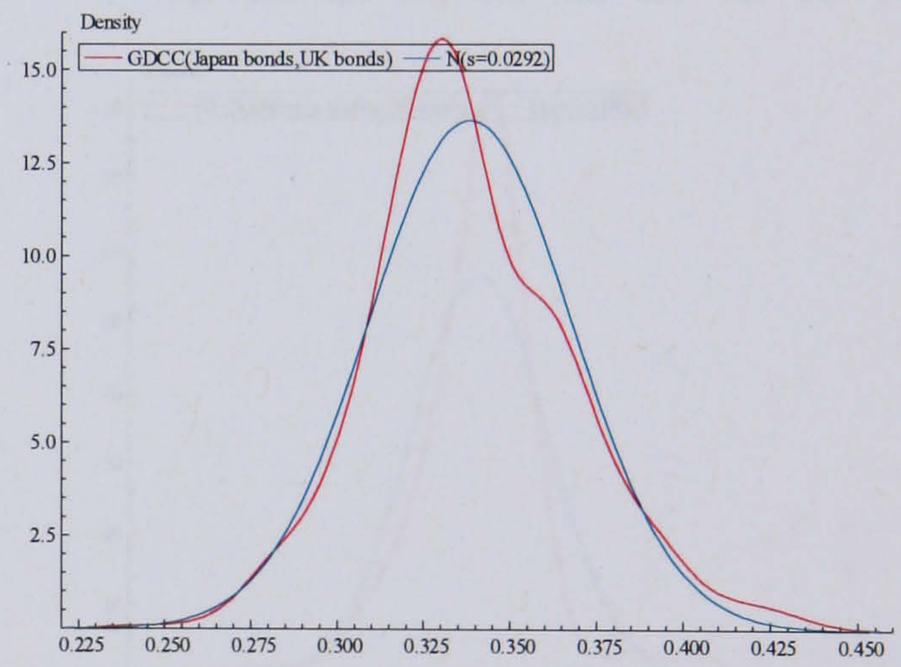
Mean	0.34610
Std.Devn.	0.11333
Skewness	2.0983
Excess Kurtosis	5.3426
Minimum	0.12380
Maximum	0.92800
Asymptotic test:	Chi^2(2) = 1509.7 [0.0000]**
Normality test:	Chi^2(2) = 995.61 [0.0000]**



Mean	0.37703
Std.Devn.	0.039126
Skewness	0.37672
Excess Kurtosis	-0.65933
Minimum	0.30290
Maximum	0.48150
Asymptotic test:	Chi^2(2) = 32.786 [0.0000]**
Normality test:	Chi^2(2) = 66.152 [0.0000]**



Mean	0.33844
Std.Devn.	0.029234
Skewness	0.35843
Excess Kurtosis	0.44919
Minimum	0.23960
Maximum	0.43860
Asymptotic test:	Chi^2(2) = 23.408 [0.0000]**
Normality test:	Chi^2(2) = 16.522 [0.0003]**



Mean	0.38789
Std.Devn.	0.0072434
Skewness	1.7742
Excess Kurtosis	21.791
Minimum	0.34220
Maximum	0.45930
Asymptotic test:	Chi^2(2) = 15944. [0.0000]**
Normality test:	Chi^2(2) = 895.29 [0.0000]**

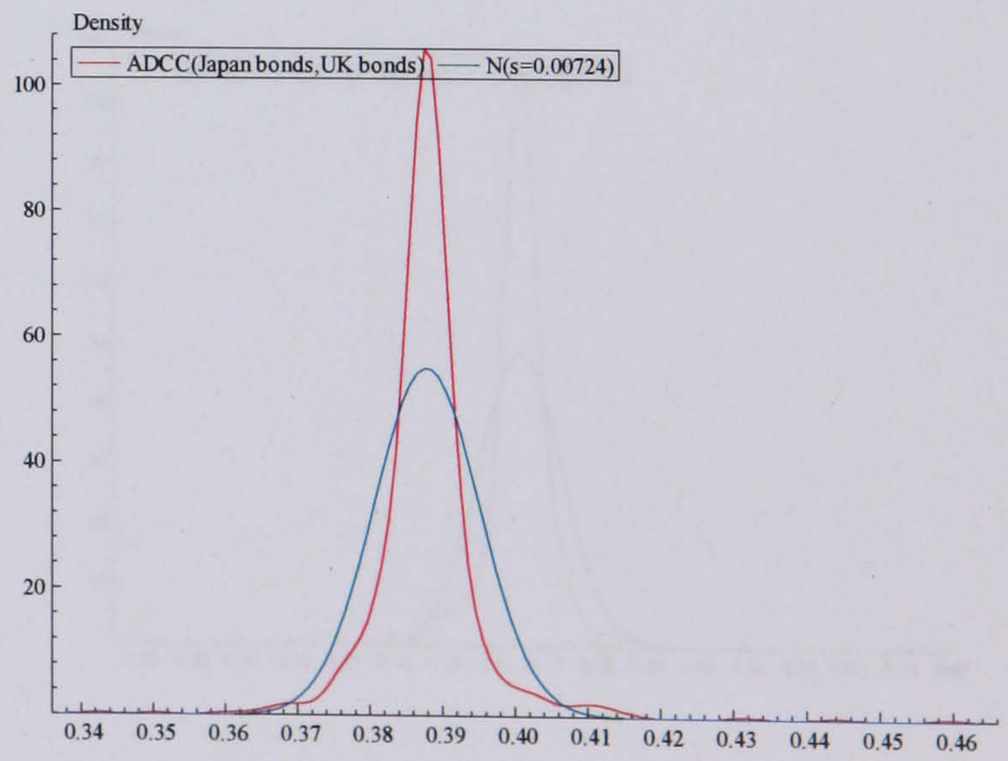
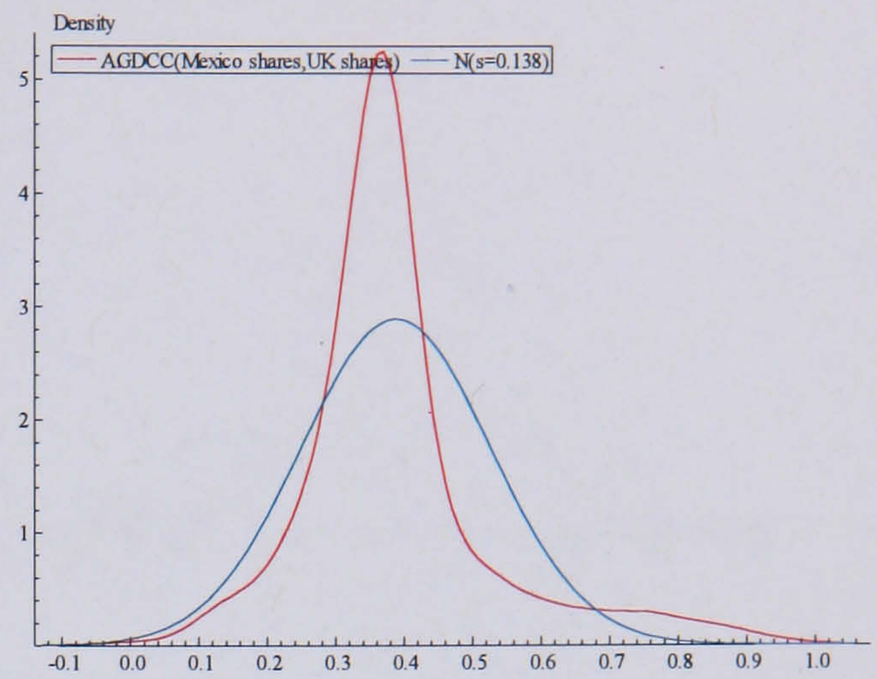
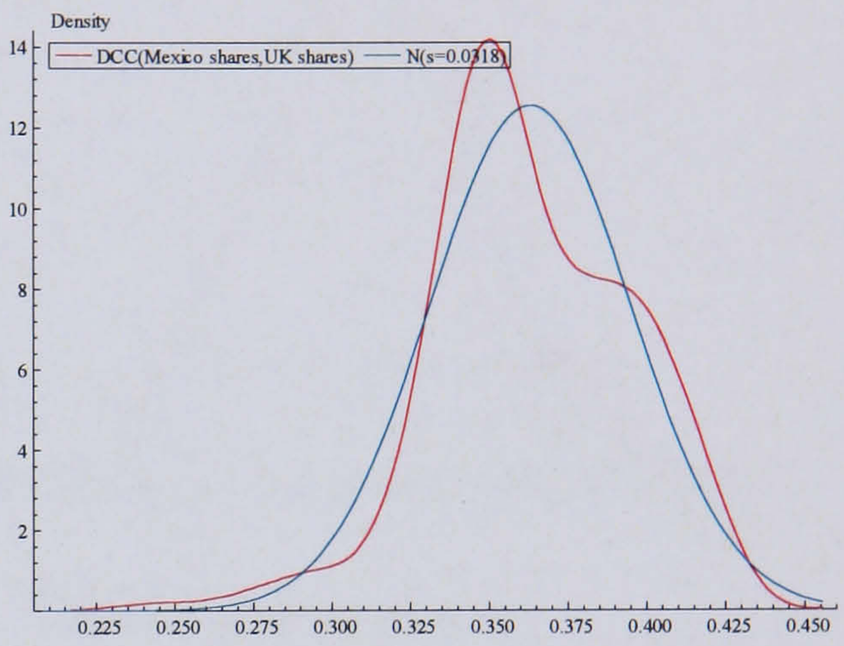


Figure 2.10. Distribution of the dynamic correlation between the return of Japan bonds and UK bonds employing the AGDCC, DCC, GDCC, and ADCC models with AML distributions.

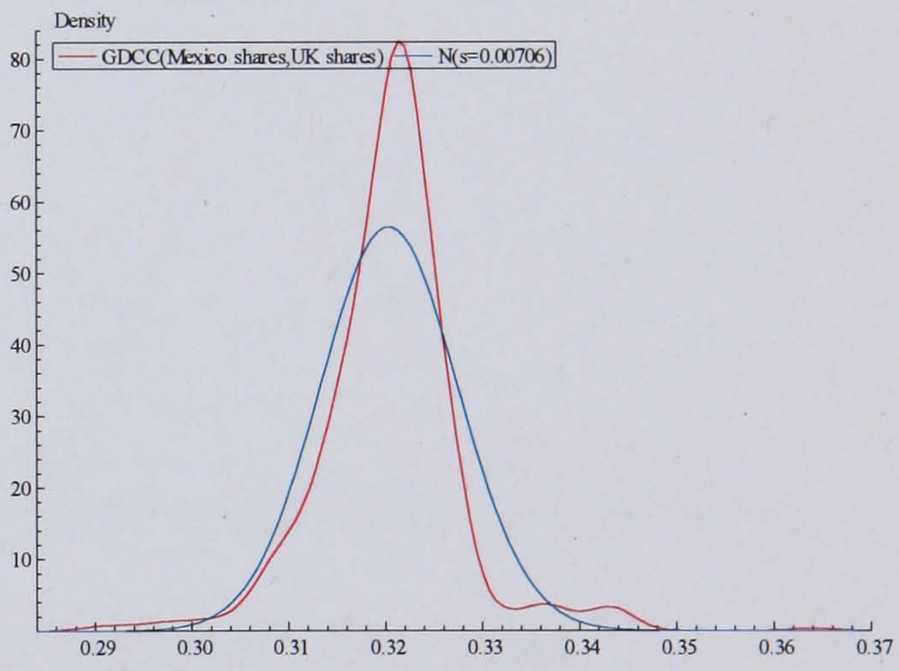
Mean	0.38693
Std.Devn.	0.13806
Skewness	1.3358
Excess Kurtosis	3.0517
Minimum	-0.029900
Maximum	0.94950
Asymptotic test:	Chi^2(2) = 538.06 [0.0000]**
Normality test:	Chi^2(2) = 220.01 [0.0000]**



Mean	0.36282
Std.Devn.	0.031816
Skewness	-0.32559
Excess Kurtosis	0.58357
Minimum	0.23340
Maximum	0.43110
Asymptotic test:	Chi^2(2) = 24.786 [0.0000]**
Normality test:	Chi^2(2) = 16.112 [0.0003]**



Mean	0.32040
Std.Devn.	0.0070628
Skewness	0.23460
Excess Kurtosis	4.5504
Minimum	0.28930
Maximum	0.36390
Asymptotic test:	Chi^2(2) = 684.45 [0.0000]**
Normality test:	Chi^2(2) = 281.82 [0.0000]**



Mean	0.36575
Std.Devn.	0.0082164
Skewness	-0.16322
Excess Kurtosis	27.493
Minimum	0.29340
Maximum	0.44070
Asymptotic test:	Chi^2(2) = 24726. [0.0000]**
Normality test:	Chi^2(2) = 2597.8 [0.0000]**

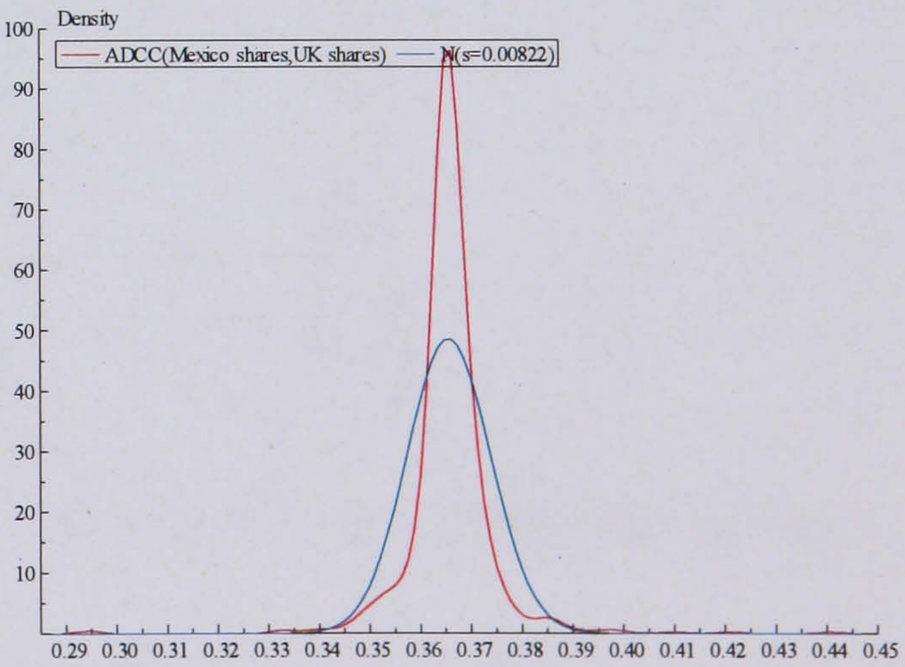
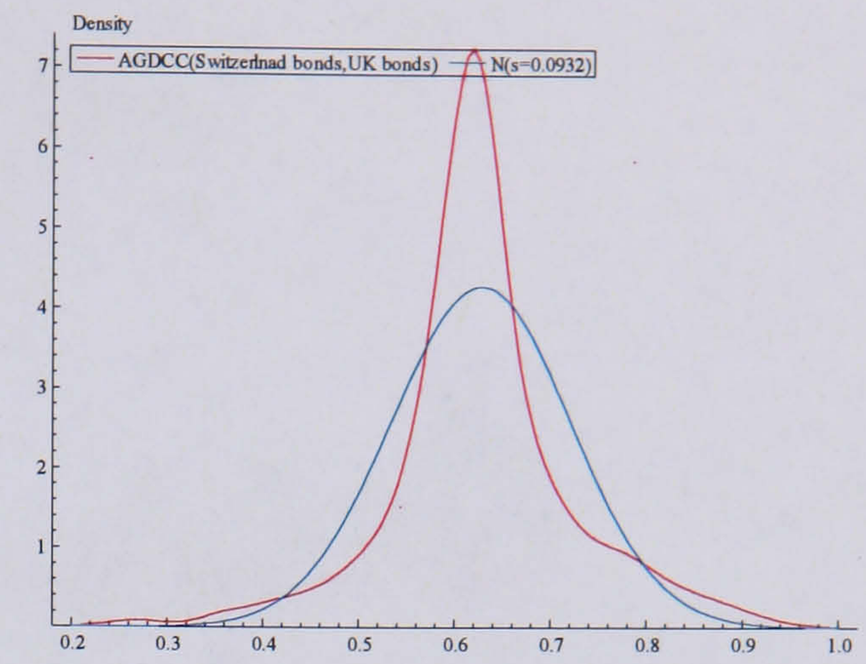
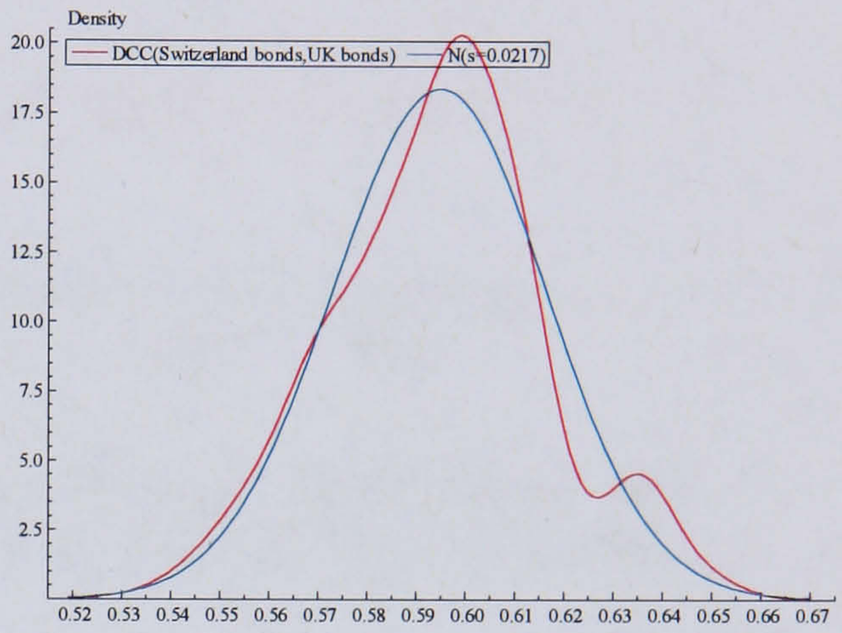


Figure 2.11. Distribution of the dynamic correlation between the return of UK shares and Mexico shares employing the AGDCC, DCC, GDCC, and ADCC models with AML distributions.

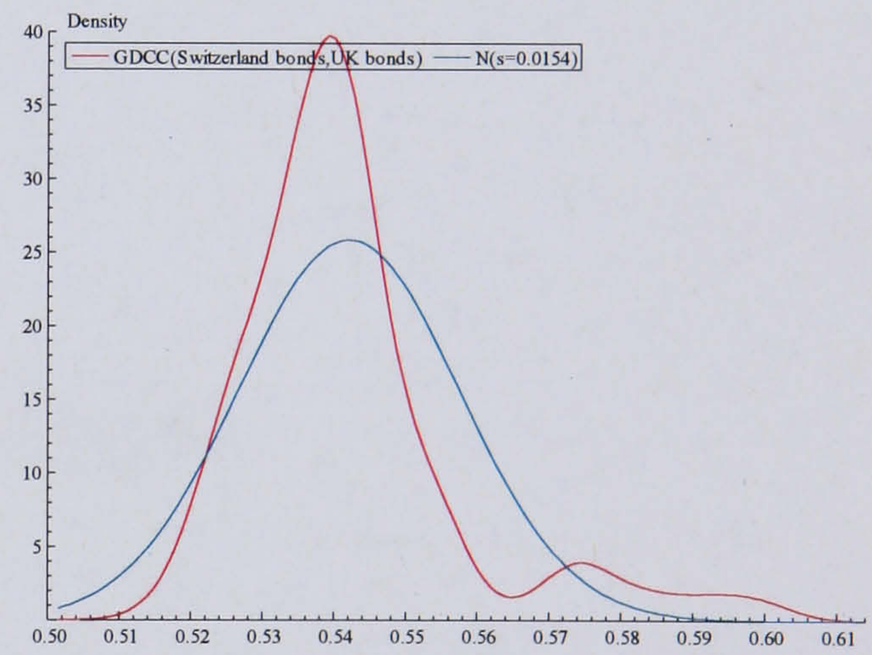
Mean 0.62607
Std.Devn. 0.093218
Skewness 0.050799
Excess Kurtosis 2.1194
Minimum 0.25990
Maximum 0.94750
Asymptotic test: Chi^2(2) = 147.26 [0.0000]**
Normality test: Chi^2(2) = 94.468 [0.0000]**



Mean 0.59432
Std.Devn. 0.021719
Skewness 0.058313
Excess Kurtosis 0.0034808
Minimum 0.53060
Maximum 0.65380
Asymptotic test: Chi^2(2) = 0.44528 [0.8004]
Normality test: Chi^2(2) = 0.47967 [0.7868]



Mean 0.54196
Std.Devn. 0.015424
Skewness 1.6521
Excess Kurtosis 3.2051
Minimum 0.51350
Maximum 0.60200
Asymptotic test: Chi^2(2) = 693.11 [0.0000]**
Normality test: Chi^2(2) = 579.90 [0.0000]**



Mean 0.62043
Std.Devn. 0.0052324
Skewness 0.30366
Excess Kurtosis 17.926
Minimum 0.58740
Maximum 0.67080
Asymptotic test: Chi^2(2) = 10523. [0.0000]**
Normality test: Chi^2(2) = 1609.0 [0.0000]**

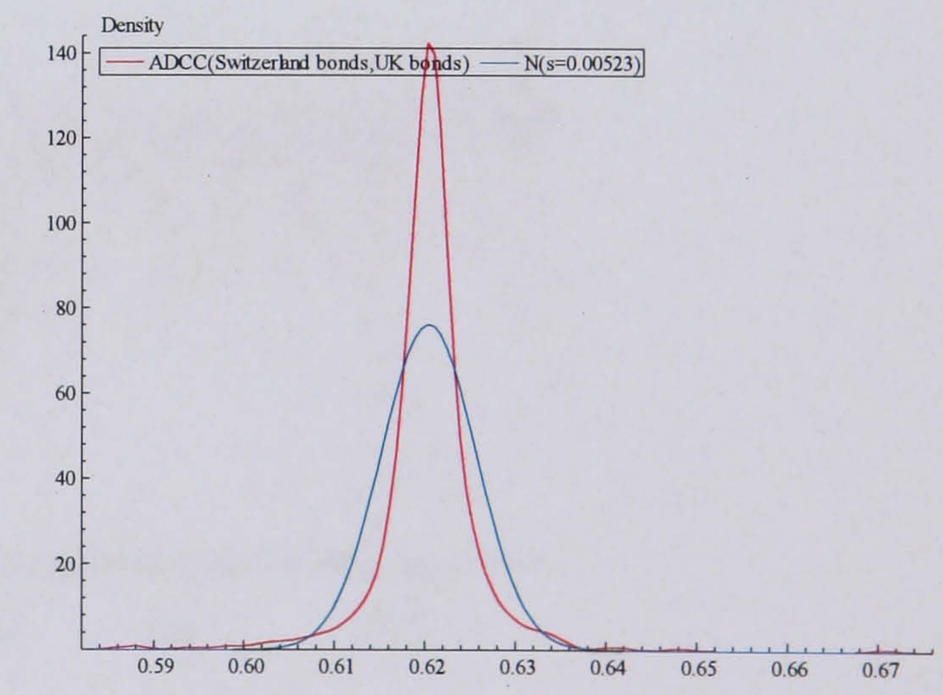


Figure 2.12. Distribution of the dynamic correlation between the return of UK bonds and Switzerland shares employing the AGDCC, DCC, GDCC, and ADCC models with AML distributions.

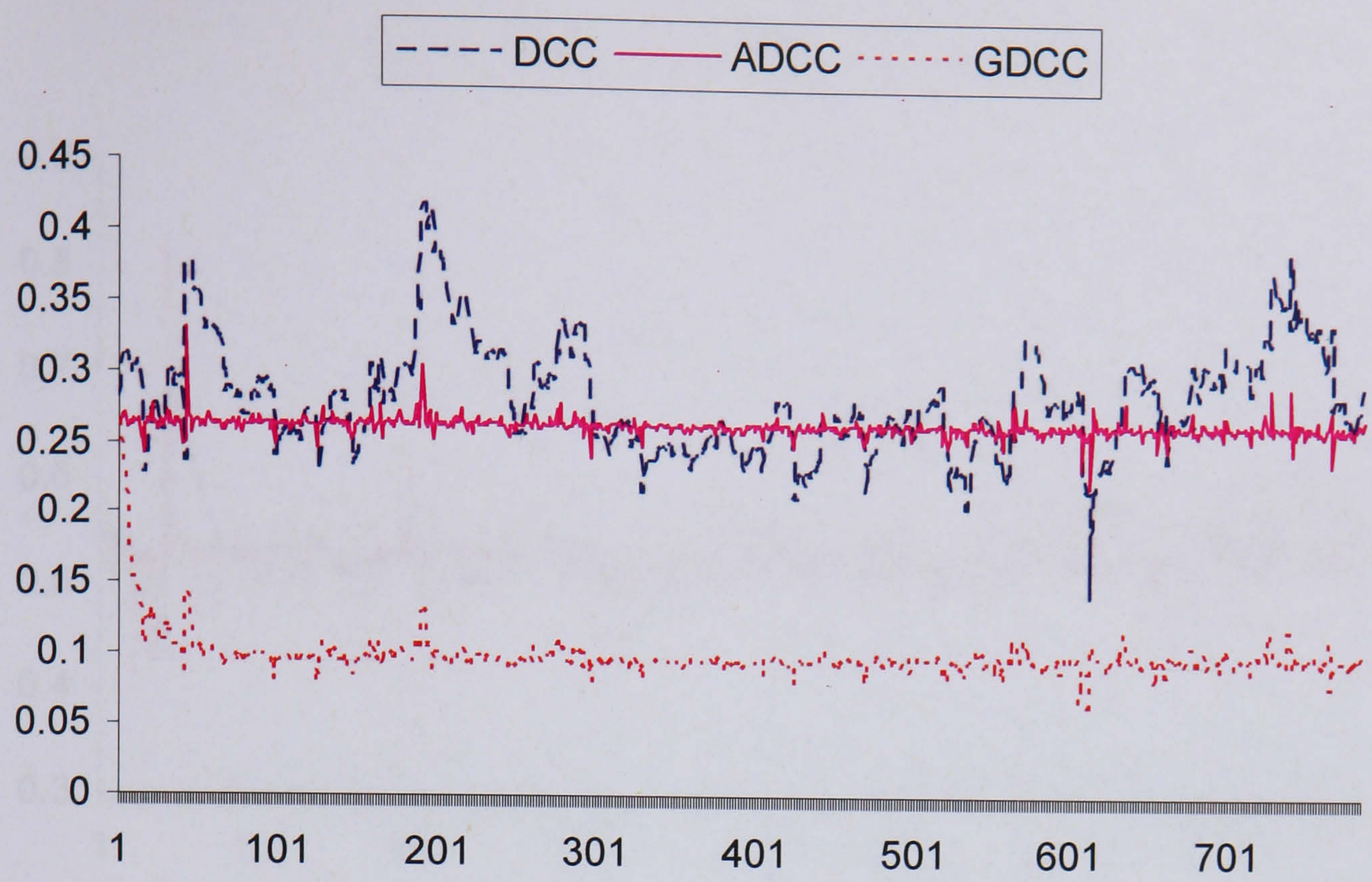


Figure 2.13. Plot of the correlation series between the returns of "Japan shares" and "U.K. shares" estimated with the DCC (1,1), ADCC(1,1,1) and GDCC(1,1) models and with the AML distribution.

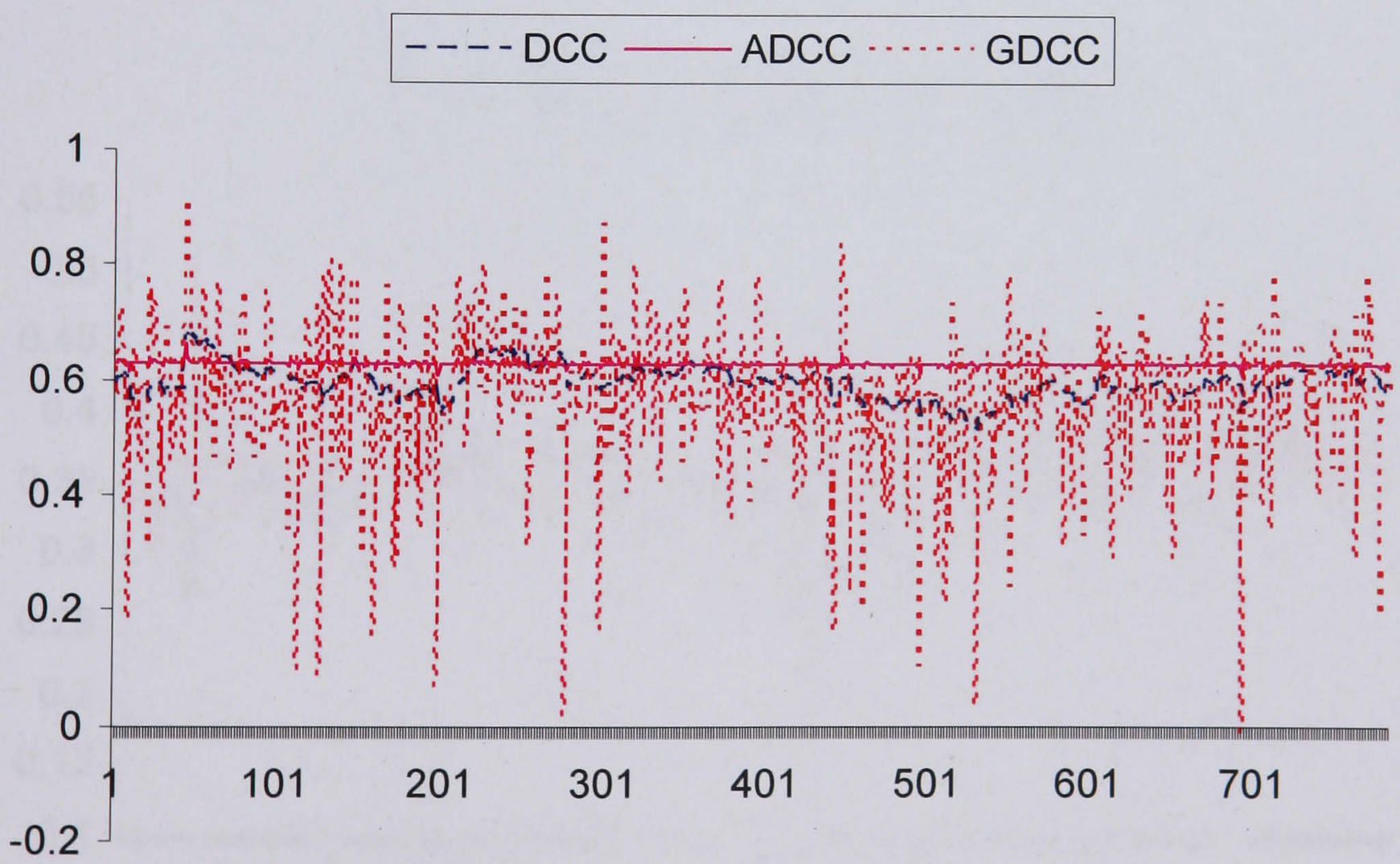


Figure 2.14. Plot of the correlation series between the returns of "Switzerland bonds" and "U.K. bonds" estimated with the DCC (1,1), ADCC(1,1,1) and GDCC(1,1) models and with the AML distribution.

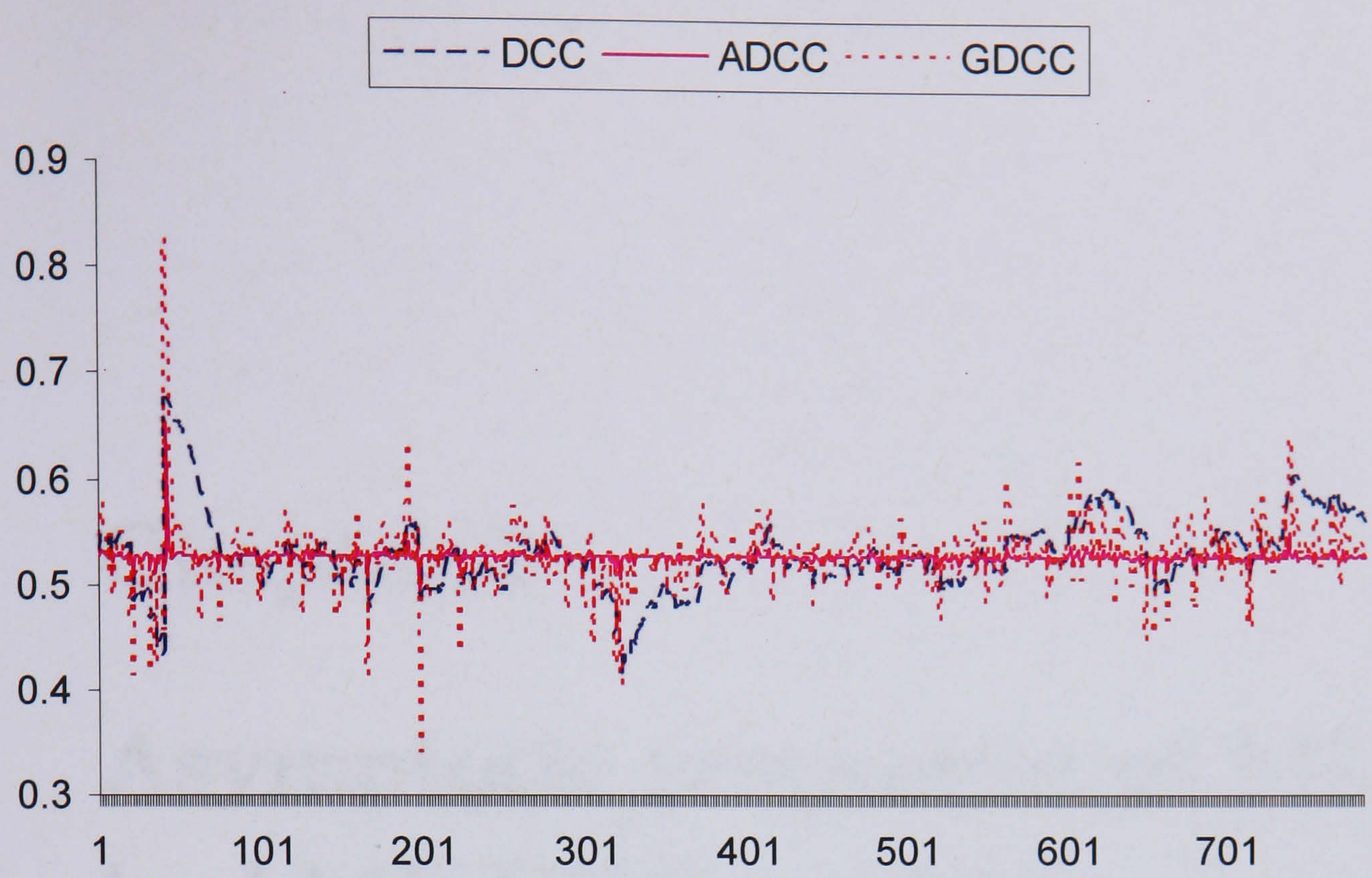


Figure 2.15. Plot of the correlation series between the returns of "U.S. shares" and "U.K. shares" estimated with the DCC (1,1), ADCC(1,1,1) and GDCC(1,1) models and with the AML distribution.

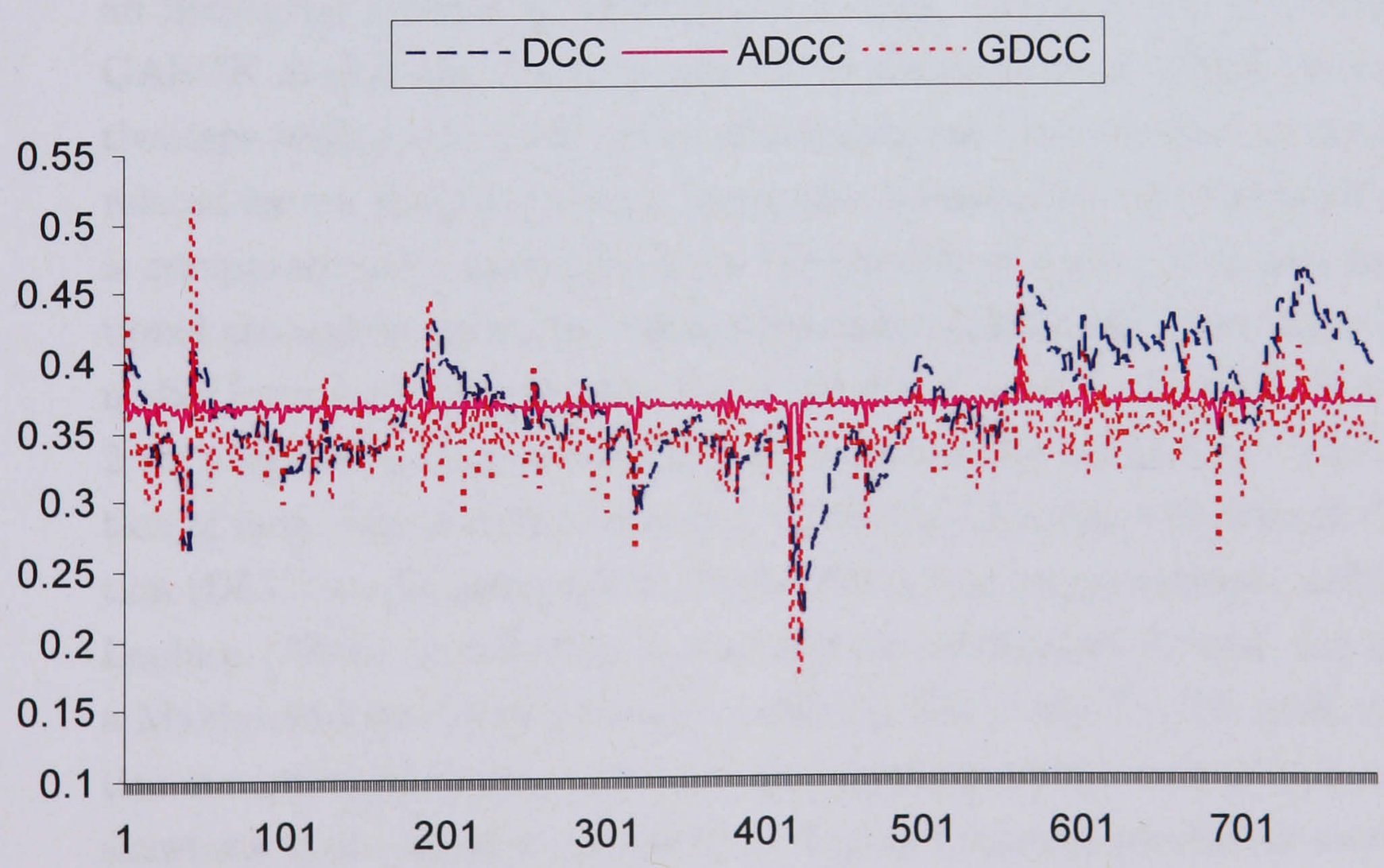


Figure 2.16. Plot of the correlation series between the returns of "Mexico shares" and "U.K. shares" estimated with the DCC (1,1), ADCC(1,1,1) and GDCC(1,1) models and with the AML distribution.

Chapter 3

Asymptotic properties of MLE in AML-DCC models

3.1 Introduction

Conditional heteroskedasticity in financial asset returns is a quite well known phenomenon reported in what is now a countless list of publications. Although the vast majority of papers relates to univariate models, there is now an increasing interest in multivariate models. For a survey on multivariate GARCH models the reader is referred to Bauwens *et al* (2006). In the multivariate setting, the main effort of research has been devoted to estimation-related issues, the main reason being that inference on non-restricted models is computationally unfeasible for a big number of series, and that the traditional estimation technique (Quasi-Maximum-Likelihood Estimation QMLE) makes unrealistic assumptions about the distribution of the data. In Chapter 2 we proposed a framework where these two issues are tackled: The estimation of large sets of data is executed under the Dynamic Conditional Correlation (DCC) model presented in Engle (2002), and an asymmetric multivariate Laplace (AML) distribution is assumed for conditional returns, recurring to a Maximum-Likelihood Estimation (MLE). The choice for the AML distribution is supported also by its many convenient properties such as leptokurtosis, skewness, finite variance, closed-form density function, parsimoniousness, and stability (see Kotz *et al* (2003)).

In this chapter, we prove the asymptotic properties of the MLE estimator as proposed in Chapter 2. Our framework is based on Jeantheau (1998) where the conditions for the strong consistency of the QMLE for multivariate ARCH models are derived. First we provide the conditions for the consistency of the MLE assuming an AML for conditional returns (we refer to this procedure as to AML-MLE) for general multivariate heteroskedastic models, and then we apply this result specifically to the DCC case.

There is a growing body of literature studying the asymptotic properties of parameter estimates in GARCH models. Weiss (1986) proved the consistency and asymptotic normality of the MLE for the ARCH (q) model under several conditions, the most relevant one being the existence of the forth moment of the raw data. This condition is very stringent as estimated parameter values in many applications imply a violation of that moment condition.

Lumsdaine (1996) established consistency and asymptotic normality for the QMLE for GARCH (1,1) and IGARCH(1,1) models. Here the conditions are related not to the raw data but to standardised residuals. A restrictive condition in this proof is that the assumed distribution for standardised residuals must be symmetric.

Lee and Hansen (1994) also prove the consistency and asymptotic normality of the GARCH (1,1) and IGARCH (1,1) models, but relaxing the condition of $i.i.d.$ standardised residuals for the more mild of strictly stationary and ergodic data, and also relaxing the condition for the shape of the distribution. An important remark regarding the methodologies in Lumsdaine (1996) and Lee and Hansen (1994) is that the maximum is local and not global. The log-likelihood function is maximised only in a neighborhood of the population parameter.

Berkes *et al* (2003) derived asymptotic results for a global MLE under weaker conditions. They show that the coefficients in the ARCH(∞) representation of the GARCH(p, q) model satisfy a specific recursion. They use special properties of the solution of this recursion to derive the asymptotic properties of the QMLE. The conditions in this paper are quiet mild and their setting appears to be optimal, but the generalisation to a multivariate scenario of their approach seems extremely difficult.

Kristensen and Rahbek (2005) provided the asymptotic properties of the QMLE for a family of ARCH (q) models. They approach is interesting as they relax the strict stationarity condition, and in special, as they present the only work where the parameter space is not assumed to be compact.

For multivariate models the study of asymptotic properties of GARCH models has been much more limited. Jeantheau (1998) proved the strong consistency of the QMLE for general multivariate ARCH models under the key condition of strict stationarity and ergodicity of the process generated by the model. The proof is for global consistency, and it has the very convenient feature (for multivariate models at least) that the study of the various derivatives of the log-likelihood function is not needed. Jeantheau uses the results to prove the consistency of the Constant Conditional Correlation (CCC) model of Bollerslev (1990). More recently, Comte and Lieberman (2003) established the asymptotic theory of the QMLE for the BEKK (p, q, k) model. They proved consistency by verifying the conditions given by Jeantheau (1998), and asymptotic normality by appealing to conditions provided by Boussama (1998) for general stochastic processes.

In this chapter we appeal, as in Comte and Lieberman (2003), to conditions established by Jeantheau (1998) for consistency. Nonetheless, in our case the model has special features such as a two steps estimation and non-normal distribution for standardised residuals. Because of this, we need to modify some of the conditions in Jeantheau (1998).

The chapter is organised as follows. In Section 3.2 we briefly present the DCC model as specified in Chapter 2. In Section 3.3 we state the necessary conditions for consistency of the estimation method, while in section 3.4 we verify that these conditions hold for our framework. Section 3.5 concludes.

3.2 The model

The specification of the *DCC(p,q)* model is as follows,

$$X_t = H_t^{1/2}\xi_t \tag{3.1}$$

$$H_t = D_t R_t D_t \tag{3.2}$$

$$R_t = (diagQ_t)^{-1/2} Q_t (diagQ_t)^{-1/2} \tag{3.3}$$

$$\mathbf{Q}_t = \left(1 - \sum_{i=1}^{q}\alpha_i - \sum_{j=1}^{p}\beta_j\right)\overline{\mathbf{Q}} + \sum_{i=1}^{q}\alpha_i\xi_{t-i}\xi'_{t-i} + \sum_{j=1}^{p}\beta_j\mathbf{Q}_{t-j} \tag{3.4}$$

where $X_t \in \mathbb{R}^d, D_t \in \mathbb{R}^{d\times d}$ is a diagonal matrix containing univariate volatilities in its diagonal, ξ_t is a martingale difference sequence, $\overline{\mathbf{Q}} \in \mathbb{R}^{n\times n}$ is the unconditional variance-covariance matrix of ξ_t, $R_t \in \mathbb{R}^{d\times d}$ is a conditional correlation matrix, and α_i and β_j are positive-scalar parameters satisfying $\sum_{i=1}^{q}\alpha_i + \sum_{j=1}^{p}\beta_j < 1$.

The specification in (3.4) can be enriched by allowing for asymmetries in conditional correlations as well as for asset-specific correlations. We will refer to this general model as the Asymmetric Generalised Dynamic Conditional Correlation (AGDCC) (p,q,r) model, where q corresponds to the number of autoregressive lags, p corresponds to the number of persistence lags, and r corresponds to the number of asymmetric shock lags:

$$\begin{aligned} Q_t \;=\; & \overline{Q} - \sum_{i=1}^{q} A_i A'_i \circ \overline{Q} - \sum_{j=1}^{p} B_j B'_j \circ \overline{Q} - \sum_{k=1}^{r} G_k G'_k \circ \overline{N} \\ & + \sum_{i=1}^{q} A_i A'_i \circ \xi_{t-i}\xi'_{t-i} + \sum_{j=1}^{p} B_j B'_j \circ Q_{t-j} + \sum_{k=1}^{r} G_k G'_k \circ \eta_{t-k}\eta'_{t-k} \end{aligned} \tag{3.5}$$

where $\eta_t = I\left(\xi_t < 0\right) \circ \xi_t$, I is the indicator function $I : X \to \{0,1\}$ defined as $I(x) = \begin{cases} 1 & \text{if } x \in \Phi \\ 0 & \text{if } x \notin \Phi \end{cases}$ where Φ is a subset of the set X, $A, B, G \in \mathbb{R}^d$, $\overline{N}$ is the unconditional variance-covariance matrix of the vector η_t, and " $\circ$ " is

the Hadamard product. The vectors A, B, and G are of the form

$$S = \begin{bmatrix} \sqrt{s_i} \\ .. \\ . \\ . \\ \sqrt{s_d} \end{bmatrix} \tag{3.6}$$

The model can be estimated in two stages. In the first one univariate volatilities are estimated by assuming zero correlations. In the second stage correlations are estimated once standardised residuals are obtained. We will refer to the process implicit in the first stage as AGDCC*, and to the process implicit in the second stage as AGDCC$^\diamond$.

The AGDCC(p, q, r) model nests the following variations,

- GDCC (p, q) when $G = 0$
- ADCC (p, q, r) when A, B and G are scalars.
- DCC(p, q) when $G = 0$ and A, B are scalars.

For the first two models equation (3.5) is respectively given by

$$\begin{aligned} Q_t^{GDCC(p,q)} &= \overline{Q} - \sum_{i=1}^{q} A_i A_i' \circ \overline{Q} - \sum_{j=1}^{p} B_j B_j' \circ \overline{Q} + \sum_{i=1}^{q} A_i A_i' \circ \xi_{t-i}\xi_{t-i}' \\ &\quad + \sum_{j=1}^{p} B_j B_j' \circ Q_{t-j} \end{aligned} \tag{3.7}$$

$$\begin{aligned} Q_t^{ADCC(p,q,r)} &= \left(1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j\right)\overline{Q} - \left(\sum_{k=1}^{r} g_k\right)\overline{N} \\ &\quad + \sum_{i=1}^{q} a_i \left(\xi_{t-i}\xi_{t-i}'\right) + \sum_{j=1}^{p} b_j Q_{t-j} + \sum_{k=1}^{r} g_k \left(\eta_{t-k}\eta_{t-k}'\right) \end{aligned} \tag{3.8}$$

In what follows we will first adapt the framework of Jeantheau (1998) to our case (Section 3.3) and then in Section 3.4 we will prove the consistency

of the parameter estimates in the AGDCC and nested models when these are estimated by an AML-MLE method.

3.3 Consistency of multivariate ARCH models

3.3.1 Pfanzgal (1969) and Jeantheau (1998)

The methodology that we follow is based on the extension to time series analysis of a theorem proved in Pfanzgal (1969) for the consistency of minimum contrast estimators. Assume first that $\{X_t, t \in Z\}$ is a multivariate ARCH process,

$$X_t = \Upsilon_\theta(\underline{X}_{t-1}) + \Delta_\theta(\underline{X}_{t-1})\varepsilon_t, \tag{3.9}$$

where
$\underline{X}_{t-1} = (X_{t-1}, X_{t-2}, ...)$,
Υ_θ is a measurable function from $(\mathbb{R}^d)^N \to \mathbb{R}^d$,
Δ_θ is a measurable function from $(\mathbb{R}^d)^N \otimes \mathbb{R}^d$ and
θ is a vector of parameters that belong to a parameter space Θ
$\varepsilon_t \sim i.i.d.(0, R)$ where R is a square matrix.

The conditional covariance matrix of the error $\Delta_\theta(\underline{X}_{t-1})\varepsilon_t$ is defined as $H_t(\theta) = \Delta_\theta(\underline{X}_{t-1})R\Delta_\theta(\underline{X}_{t-1})'$. In what follows we will refer to θ_0 as to the true (population) value of θ.

Consider now the following hypotheses:

Hypothesis 1. Θ is a compact, convex parameter space[1]

Hypothesis 2. The stochastic process $\{X_t, t \in Z\}$ is strictly stationary and ergodic.

Hypothesis 3. The function $F(\theta_0, \theta) = E_{\theta_0}(f(\underline{X_1}, \theta)$, has a unique finite minimum at θ_0, where f is a real valued measurable function continuous in θ.

Hypothesis 4. $\forall\theta \in \Theta, E_{\theta_0}(\inf(f_*(\theta, \rho), 0)) > -\infty$, where $f_*(\theta, \rho) = \inf\{f(\underline{X_t}, \theta'), \theta' \in B(\theta, \rho)\}$ where $B(\theta, \rho)$ is a ball of center θ and radius ρ.

[1] A Euclidean space is called compact and convex if it is closed and bounded.

Hypothesis 1 requires the knowledge of bounds of the true parameter value. As discussed in Newey and McFadden (1994) this assumption can be relaxed if the instability of $F_T(\underline{X_t}, \theta)$ when θ is unbounded is not extreme. Hypothesis 3 secures *identification*, i.e. the existence of a unique value of $F(\theta_0, \theta)$ when this is evaluated at the true parameter value. Hypothesis 2 in conjunction with Hypothesis 4 secure uniform convergence of $F_T(\underline{X_t}, \theta)$ in to $F(\theta_0, \theta)$, condition that is also required for consistency (Newey and McFadden (1994)).

In the following theorem Pfanzgal (1969) proofs consistency of the minimum contrast estimator,

Theorem 3.1 *On the sample space* (Ω, F, P), *let* $\{X_t, t \in Z\}$ *be a random process,* $\theta \in \Theta$, *and* $F_T(\underline{X_t}, \theta)$ *be a contrast process such that* $F_T(\underline{X_t}, \theta) = T^{-1}\sum_{t=1}^{T} f(\underline{X_t}, \theta)$. *Under Hypotheses 1 to 4 the minimum contrast estimator* $\widehat{\theta} \overset{a.s}{\rightarrow} \theta_0$.

Proof. See Pfanzgal (1969). ■

Jeantheau (1998) extends the result in Pfanzgal (1969) to the case where the parameter vector θ is estimated by MLE. Though the random variable ε_t is not necessarily assumed to be normal but the optimisation is performed under a Gaussian density method (the so-called QMLE). In order to use QMLE Hypotheses 1-4 in Pfanzgal (1969) are replaced by a set of different assumptions. The introduction of these new assumptions is necessary because the process considered (multivariate ARCH) is more specific than the one studied by Pfanzgal (1969).

The assumptions introduced by Jeantheau (1998) are the following,

- Assumption 1. Θ is a compact, convex parameter space, where $\theta \in \Theta$
- Assumption 2. The process defined by (3.9) is strictly stationary and ergodic.
- Assumption 3. There exists a constant $[c > 0]$ such that $\det(H_t(\theta)) \geq c$ for all t, and $\theta \in \Theta$
- Assumption 4. $\forall \theta \in \Theta, E_{\theta_0}(|\log(\det(H_t(\theta))|) < \infty$, where θ_0 denotes the true parameter value

- Assumption 5. The function $H_t(\theta)$ is such that $\forall\theta \in \Theta, \forall\theta_0 \in \Theta, H_t(\theta) = H_t(\theta_0) \Rightarrow \theta = \theta_0$

- Assumption 6. The function $H_t(\theta)$ is a continuous function of the parameter θ.

As we can see Hypothesis 1 and Hypothesis 2 are replaced by Assumption 1 and Assumption 2. It is easy to verify that under Assumption 3 Hypothesis 4 holds; the negative log-density of a normal distribution[2] with mean zero is given by,

$$f(\underline{X}_t, \theta) = \log(\det H_t(\theta)) + (X_t)' H_t^{-1}(\theta)(X_t) \tag{3.10}$$

Since matrix $H_t^{-1}(\theta)$ is positive definite by definition, Assumption 3 $F_T(\underline{X_t}, \theta) > \log(c)$ and Hypothesis 4 is satisfied. Assumptions 1-5, in addition to the Proposition (3.1) bellow secure the holding of Hypothesis 3. Heuristically, Hypothesis 4 refers to the existence of the first moment of the infimum of $f(\underline{X_t}, \theta')$ when the parameter space of θ' is bounded. In this particular case $f(\underline{X}_t, \theta)$ is the density function of a random variable following a normal distribution and the minimum contrast estimator is obtaining by minimising the sample average of $f(\underline{X_t}, \theta)$ across t. If $E_{\theta_0}(\inf(f_*(\theta, \rho), 0)) = -\infty$ when $F_T(\underline{X_t}, \theta) = T^{-1}\sum_{t=1}^{T} f(\underline{X_t}, \theta)$,then the minimum of $F_T(\underline{X_t}, \theta)$ will not exist and the minimum contrast estimator cannot be obtained. In (3.10) this is secured by Assumption 3. If $c > 0$ and $\det(H_t(\theta)) \geq c$ then $\log(\det H_t(\theta)) > -\infty$ and $E(f(\underline{X}_t, \theta)) > -\infty$.

Proposition 3.1 *Under Assumptions 1-5,* $F_T(\underline{X_t}, \theta) \overset{a.s.}{\to} F(\theta_0, \theta)$, *and this function has a unique minimum at* θ_0.

Proof. See Jeantheau (1998). ■

We turn now to the case of AML.

[2] We consider the negative of the log-density because Theorem 3.1 refers to a *minimum* contrast estimator, defined by $\widehat{\theta} = \arg\min_{\theta} F_T(\underline{Y_T}, \theta)$

3.3.2 The AML case

In this section we follow the strategy in Jeantheau (1998) and we adapt the results in Pfanzgal (1969), given that in our case the maximum likelihood estimation is performed under the assumption of an AML distribution for standardised residuals. We will show how in this case Hypotheses 1-4 can also be replaced by Assumptions 1-6 .

Consider the negative log-density of the AML distribution,

$$\begin{aligned} f(\underline{X}_t, \theta) &= \frac{1}{2}\log(\det H_t(\theta)) - (X_t{}' H_t^{-1}(\theta) m_t) \\ &\quad -\frac{\nu}{2}\left[\log(X_t{}' H_t^{-1}(\theta) X_t) - \log(2 + m_t{}' H_t^{-1}(\theta) m_t)\right] \\ &\quad - \log\left(K_\nu\left(\sqrt{(2 + m_t{}' H_t^{-1}(\theta) m_t)(X_t{}' H_t^{-1}(\theta) X_t)}\right)\right)^3 \end{aligned} \tag{3.11}$$

where $K_\nu(\cdot)$ is the modified Bessel function of the second kind (sometimes called modified Bessel function of the third kind. See Bowman (1958) or Relton (1964)).

In what follows we show how for the specific case where the log-density of X_t is given by (3.11) the four Hypotheses listed by Pfanzgal (1969) can be replaced by Assumptions 1 to 6.

The following Lemma shows under what conditions Hypothesis 4 can be replaced by Assumption 3.

Lemma 3.1 *Hypothesis 4 in Theorem 3.1 holds for $F_T(\underline{X_t}, \theta)$ when X_t is given by (3.11) if $H_t(\theta)$ is a valid variance-covariance matrix and if Assumptions 3 holds.*

Proof. If $H_t(\theta)$ is a valid variance-covariance matrix then $H_t^{-1}(\theta)$ will be positive definite and by Assumption 3 $F_T(\underline{X_t}, \theta) > \log(c)$. Also, given that the probability that all elements in $X_{t=1,i}, i = 1, ..., n$ are equal to zero is insignificant we have,

$$K_\nu\left(\sqrt{(2 + m_t{}' H_t^{-1}(\theta) m_t)(X_t{}' H_t^{-1}(\theta) X_t)}\right) < \infty \tag{3.12}$$

■

Hypothesis 4 is necessary to ensure that the minimum of the log-density function does exist. Assumption 3 is much more easy to verify and Lemma 3.1 provides the connection between the two.

We need the following proposition to show that Hypothesis 3 (identification condition) holds for the case when the AML distribution is assumed.

Proposition 3.2 *If Assumptions 3-5 hold, and if $f(\underline{X}_t, \theta)$ is the density function of ε_t, then $F_T(\underline{X}_t, \theta) \xrightarrow{a.s.} F(\theta_0, \theta)$ and $F(\theta_0, \theta)$ has a unique optimum at θ_0.*

Proof. The first part of the proposition follows from the strict stationarity and ergodicity of the process $\{X_T\}$ (Assumption 2) and the ergodic theorem. To prove that $F_T(\theta_0, \theta)$ has a unique finite optimum in θ_0, we have first that by the concavity of the logarithm function and by Jensen's Inequality

$$E_0\left(\log\left(\frac{L\left(H_t(\theta), m(\theta); r_t\right)}{L\left(H_t(\theta_0), m(\theta_0); r_t\right)}\right)\right) \leq \log E_0\left(\left(\frac{L\left(H_t(\theta), m(\theta); r_t\right)}{L\left(H_t(\theta_0), m(\theta_0); r_t\right)}\right)\right) \tag{3.13}$$

where $L\left(H_t(\theta_0), m(\theta_0); r_t\right)$ is the likelihood function evaluated at the true parameter value, and $L\left(H_t(\theta), m(\theta); r_t\right)$ is the likelihood function evaluated at any other parameter value in the compact (Assumption 1) parameter space Θ. Using the fact that

$$\begin{aligned} E_0\left(\frac{L\left(H_t(\theta), m(\theta); r_t\right)}{L\left(H_t(\theta_0), m(\theta_0); r_t\right)}\right) &= \int\left(\frac{L\left(H_t(\theta), m(\theta); r_t\right)}{L\left(H_t(\theta_0), m(\theta_0); r_t\right)}\right) L\left(H_t(\theta_0), m(\theta_0); r_t\right) dz \\ &= \int L\left(H_t(\theta), m(\theta); r_t\right) dz = 1 \end{aligned} \tag{3.14}$$

we can rewrite equation (3.13) as

$$E_0\left(\log\left(\frac{L\left(H_t(\theta), m(\theta); r_t\right)}{L\left(H_t(\theta_0), m(\theta_0); r_t\right)}\right)\right) \leq 0 \tag{3.15}$$

and

$$E_0\left(\log L(H_t(\theta), m(\theta); r_t)\right) \leq E_0\left(\log L\left(H_t(\theta_0), m(\theta_0); r_t\right)\right) \tag{3.16}$$

The equality holds if and only if $H_t(\theta) = H_t(\theta_0)$ and $m(\theta) = m(\theta_0)$. Therefore, according to Assumption 5, we have $F(\theta_0, \theta) = F(\theta_0, \theta_0)$ if an only if $\theta = \theta_0$. ■

We have shown how under several conditions the Hypotheses in Pfanzgal (1969) can be replaced by Assumption 1 to 6.

The next step is to group these conditions and to prove the consistency of parameter estimates in multivariate heteroskedastic models obtained by AML-MLE. The following Theorem provides the proof of consistency of parameter estimates using the AML-MLE.

Theorem 3.2 *Under assumptions 1 to 6, the AML-Maximum Likelihood Estimator (AML-MLE) for the multivariate heteroskedastic model (3.9) is consistent:* $\widehat{\theta} \xrightarrow{p} \theta_0$.

Proof. The MLE is a minimum contrast estimator when the maximization is taken over the negative-log-likelihood function. The consistency of a minimum contrast estimator was proven by Pfanzgal (1969) under Hypothesis 1 to 4 in Theorem 3.1. Assumptions 1 and 2 replicate Hypothesis 1 and 2. For a process $\{X_t, t \in Z\}$ following the dynamics of process (3.9) and parameters estimated by *AML-MLE* we have that Hypothesis 4 holds if Assumption 3 holds (Lemma 3.1), and that Hypothesis 3 holds if Assumption 1 to 6 hold (Proposition 3.2) ■

3.4 Consistency of estimates in the AGDCC model

In the previous section we provided the conditions for consistency of the AML-MLE for multivariate ARCH models. We will now apply this result to the specific case of the AGDCC model; we will verify under what conditions Assumptions 1 to 6 hold in the case of this particular multivariate ARCH model . As in Section 3.2, we refer to the process defined by the first-step estimation as AGDCC*, and to the process defined by the second-step estimation as AGDCC$^\diamond$.

3.4.1 $AGDCC^*$ model

In this section we will apply the result of consistency derived in Section 3.3 to the particular case when the multivariate ARCH process is of the type AGDCC*. In order to evaluate under what conditions Assumption 2 (stationarity) is valid, and given that for this process $R = I_d$[4], we write (3.3) as

$$X_{it} = \sqrt{h_t}\varepsilon_{it}, \qquad i = 1, ...d \tag{3.17}$$

$$h_{it} = w_i + \sum_{j=1}^{q} \alpha_{ij} X_{it-j}^2 + \sum_{k=1}^{p} \beta_{ik} h_{it-k} \tag{3.18}$$

We can rewrite equation (3.2) as,

$$H_t = \begin{bmatrix} h_{11,t} & \dots & 0 & 0 \\ \vdots & \ddots & h_{(d-1)(d-1),t} & 0 \\ 0 & \dots & 0 & h_{dd,t} \end{bmatrix} \tag{3.19}$$

Let us define $diagH_t' = [h_{1t}, ..., h_{dt}]$ as the vector containing the diagonal elements of H_t. Following specification (3.18) we can write

$$diagH_t = \begin{pmatrix} w_i \\ \vdots \\ w_d \end{pmatrix} + \sum_{j=1}^{q} N_j \begin{pmatrix} X_{t-j,1}^2 \\ \vdots \\ X_{t-j,d}^2 \end{pmatrix} + \sum_{k=1}^{p} M_k diagH_{t-k} \tag{3.20}$$

where N_j and $M_k \in \mathbb{R}^{d\times d}$, and where we assume that all coefficients are positive. Strict stationarity is stated in the following proposition,

Proposition 3.3 *By assuming that θ_0 is such that* $\det(I_d - \sum_{i=1}^{n} (N_i - M_j)\lambda^i)$ *has its roots outside the unit circle the CCC model of Bollerslev (1990) has a unique covariance stationary solution. This solution is unique and is also strictly stationary and ergodic.*

[4] For the process AGDCC* we have that $R_t = I_d$ and the model simplifies to the case of the *CCC* model of Bollerslev (1990) with $R = I_d$

Proof. See Jeantheau (1998). ■

This result is very useful as it shows how for the CCC model strict stationarity follows from covariance stationarity. Covariance stationarity can be easily verified in this case by checking the roots in $\det(I_d - \sum_{i=1}^{n} (N_i - M_j)\, \lambda^i)$.

The next step is to verify conditions for identification. To do this it is convenient first to rewrite Equation (3.18) as

$$P(L) \begin{pmatrix} h_{11,t}(\theta) \\ \vdots \\ h_{dd,t}(\theta) \end{pmatrix} = \begin{pmatrix} w_i \\ \vdots \\ w_d \end{pmatrix} + Q(L) \begin{pmatrix} X_{1,t}^2 \\ \vdots \\ X_{d,t}^2 \end{pmatrix} \tag{3.21}$$

where L is the backshift operator, and P and Q are two matrices with polynomial coefficients such that $P(L) = I_d - \sum_{i=1}^{p} \beta_i L^i$ and $Q(L) = \sum_{i=1}^{q} \alpha_i L^i$ where I_d is the identity matrix. For multivariate GARCH processes the condition of identification is intimately related to the a condition known in the multivariate time series analysis as *minimal.* To understand this concept we need first to consider the following preliminary definitions:

Definition 3.4.1. A polynomial matrix $M(L)$ with degree d_{ij} , i.e. $M_{ij}(L) = \sum_{l=0}^{d_{ij}} a_{ij,l} L^l$, is column reduced if $\det(a_{ij,d_j}) \neq 0$. We define also $d_j(M) = \sup_i d_{ij}$

Definition 3.4.2. Denote by MP the set of matrices with polynomial coefficients. A square matrix $M(L) \in MP$ is unimodular if $\det(M(L)) = \pm 1$ and if $\det(M(L)) \perp L$.

Definition 3.4.3. Let $A, B \in MP$ such that $\det(A) \neq 0$ and $\det(B) \neq 0$. The matrix $D \in MP$ is called the greatest common left divisor of A and B if every left divisor of D is also a left divisor of A and B, and if every left divisor of A and B is also a left divisor of D.

Definition 3.4.4. Two matrices $A, B \in MP$ are coprime if any of their greatest common left divisor is unimodular.

The definition of coprime matrices is more involved than for scalars. In that case two integers are coprime if their greatest common divisor is 1. Likewise, two polynomials are coprime if their greatest common divisor is different from zero.,

The definition of a minimal multivariate GARCH process can now be given,

Definition 3.4.5. The multivariate GARCH (p,q) specification given in (3.21) is minimal if

1. $P(0) = I_d$ and $Q(0) = 0$
2. $det(P) \neq 0$ and $det(Q) \neq 0$
3. P and Q are coprime.
4. $\forall j, 1 \leq j \leq d, d_j(P) = d_j \leq p$ and $d_j(Q) = d_j \leq q$.
5. P or Q is column reduced.

The identification in univariate GARCH models involves less complicated conditions. The assumption of the two polynomials involved in the equation of the conditional variance being coprime is usually enough to secure identification. This is not the case for MGARCH models because the greatest common left divisor is not unique for polynomial matrices. This is the reason way the notion of "column-reduced" matrix must be introduced in Definition 3.4.1.

We introduce now two additional assumptions: Assumption 7 and Assumption 8. The goal is, and given that now the process under analysis is the one defined in (3.17) and (3.18), to express the conditions listed in Assumption 3 and Assumption 5 in a more primitive way.

- Assumption 7. There are two strictly positive constants c_1 and c_2 such that all the w_i elements in (3.18) are greater than $c_1^{1/d}$ and $\det(R) \geq c_2$
- Assumption 8. The formulation at θ_0 for the model (3.17)-(3.18) is minimal.

Assumption 7 identifies primitive conditions in relation to the existence of a positive bound for the determinant of the conditional variance-covariance matrix $H_t(\theta)$ when this is define as in (3.18). By using this assumption we can verify more easily conditions stated in Assumption 3. Assumption 8 is related to the identification condition stated in Assumption 5. With Definition 3.4.5 identification can be verified more easily.

The following theorem establishes consistency of the $AGDCC^*$ model estimated by AML-MLE.

Under assumptions 7 and 8 parameter estimates obtained by AML-MLE for the $AGDCC^*$ model defined by (3.17) and (3.18) are consistent.

Proof. In Proposition 3.3 we showed that the $AGDCC^*$ process in strictly stationary and ergodic when it is covariance-stationary. This satisfies Assumption 2. By definition α_i and β_i are positive, therefore by Assumption 7 we have $\det(D_t) \geq c^{1/2}$ and $\det(H_t) \geq c_1 c_2$, satisfying Assumption 3. Because (3.18) is weakly stationary we know that $E_0(H_{ii,t}) < \infty$. By Jensen's inequality we have $E_0(\log(\det(H_{ii,t}))) < +\infty$. Adding the part of Assumption 7 where $\det(H_t) \geq c_1 c_2$ we get $E_0(|\log(\det(H_{ii,t}))|) < \infty$, satisfying Assumption 4. Jeantheau proved in PROPOSITION 3.4 that under a weakly stationary solution for the process and under Assumption 8, the function H in the CCC model is given such that $\forall\theta \in \Theta, \forall\theta_0 \in \Theta, H_t(\theta) = H_t(\theta_0) \Rightarrow \theta = \theta_0$, satisfying Assumption 5. Finally, Assumption 6 holds obviously. ■

Some comments regarding this Theorem are relevant at this point. First, it is important to notice that the condition in the second part of Assumption 8 is very weak as the determinant of a correlation matrix is zero only in the special case when this is positive-semidefinite. For a positive-definite matrix A we have $|A| > 0$. Secondly, Assumption 8 is necessary in order to satisfy identification, avoiding the case where two (or more) representations of (3.21) are equivalent. We say that two $VARMA$ representations are equivalent if $P(L)^{-1}Q(L)$ results in the same operator $\Psi(L)$ (Dufour and Pelletier (2004)). We need the *minimal* condition to avoid that elements of $P(L)$ and $Q(L)$ "cancel out" when we take $P(L)^{-1}Q(L)$ in (3.21).

3.4.2 $AGDCC^{\Diamond}$ model

So far we have proved consistency of parameter estimates involved in the first-stage process $AGDCC^*$. This was important because, and as explained in the Chapter 2, parameters estimated in the second-stage will be inconsistent if the residuals in the first-stage have been standardised by inconsistent

conditional variances. In this section we will proceed with the analysis of consistency of parameters estimated in the second-step estimation. We will first derive the conditions for the strict stationarity of the AGDCC$^{\diamond}$ process. To this end we will introduced the concept of *top-Lyapounov exponent* and how its behaviour defines the existence of a strict-stationarity solution for the AGDCC$^{\diamond}$ process. We then generalised this result for the models embedded in the AGDCC$^{\diamond}$ process, namely the GDCC$^{\diamond}$ (p, q), ADCC$^{\diamond}$ (p, q, r), and DCC$^{\diamond}(p, q)$ processes. Subsequently we modify Definition 3.4.5 stated in the previous section in order to account for the new dynamics introduced. Finally we present the main theorem of this chapter proving the consistency for the entire process.

Now we verify the main conditions required for consistency, i.e. strict-stationarity and identification. Equations (3.1)-(3.3) for the case of the AGDCC$^{\diamond}$ process can be written as follows

$$X_t = g(D_t, R_t, \xi_t) \tag{3.22}$$

$$R_t = v(Q_t) \tag{3.23}$$

where g and v are measurable functions.

In order to prove the strict stationarity of AGDCC$^{\diamond}$ we make use of the following proposition,

Proposition 3.4 *Consider the process $Z_t = f(\Gamma_t)$ where $f(\cdot)$ is a measurable function. If the process Γ_t is strictly stationary then Z_t is also strictly stationary.*

Proof. The w-moment of Z_t is given by

$$E_t(Z_t^w) = E_t[f(\Gamma_t)^w] \tag{3.24}$$

If the density of Γ_t is defined as $\psi_t(\Gamma_t)$ then

$$E_t(Z_t^w) = \int_{\infty}^{\infty} f_t(\Gamma_t)^w \psi_t(\Gamma_t) d\Gamma_t \tag{3.25}$$

but if Γ_t is strict stationary we have that $\psi_t(\Gamma_t) = \psi(\Gamma_t)$.By definition we also have that $f_t(\cdot) = f(\cdot)$, therefore

$$E_t(Z_t^w) = \int_{\infty}^{\infty} f(\Gamma_t)^w \psi(\Gamma_t) d\Gamma_t = E(Z_t^w) \tag{3.26}$$

i.e. the w-moment of Z_t is time-independent. ■

Proposition 3.4 is useful because it allows us to avoid the proof of strict-stationarity of a non-linear function of a random process. Now we only need to prove the strict-stationarity of equation (3.5) in order to obtain the strict stationary of the AGDCC$^{\diamond}$ process.

Equation (3.5) can be written as

$$U_t = FU_{t-1} + C_t, \qquad t \in Z \tag{3.27}$$

where

$$U_t = \begin{bmatrix} Q_t \\ . \\ . \\ . \\ Q_{t-p-1} \end{bmatrix} \tag{3.28}$$

$$F = \begin{bmatrix} B_1B_1' & B_2B_2' & . & . & . & B_{p-1}B_{p-1}' & B_pB_p' \\ [I_d] & [0] & . & . & . & [0] & [0] \\ [0] & [I_d] & . & . & . & [0] & [0] \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ [0] & [0] & . & . & . & [I_d] & [0] \end{bmatrix} \tag{3.29}$$

$$C_t = \begin{bmatrix} \left\{\overline{Q} - \sum_{i=1}^{q} A_i A_i' \circ \overline{Q} - \sum_{j=1}^{p} B_j B_j' \circ \overline{Q} - \sum_{k=1}^{r} G_k G_k' \circ \overline{N} \right. \\ \left. + \sum_{i=1}^{q} A_i A_i' \circ \xi_{t-i}\xi_{t-i}' + \sum_{k=1}^{r} G_k G_k' \circ \eta_{t-k}\eta_{t-k}' \right\} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{3.30}$$

where the sup-products $(B_i B_i' \circ Q_{t-1})$, $i = 1, ..., p$ in FU_{t-1} are Hadamard. Note that $U_t \in \mathbb{R}^d, B \in \mathbb{R}^{d\times d}$, and that $C_t \in \mathbb{R}^{d\times d}$. This is a valid representation in the sense that a stationary solution is independent of the future at any given time, i.e. for any $\tau \in Z$, U_τ is independent of the variables $\{(F, C_t), t > \tau\}$

The necessary and sufficient conditions for strict stationarity of generalised autoregressive processes of the form $U_t = F_t U_{t-1} + C_t$, $t \in Z$ were derived by Bougerol and Picard (1992). They define the strict stationarity of stochastic recurrence equations in terms of a metric measure of the autoregressive component F_t called the Lyapunov exponent that now we define:

For the process $U_t = F_t U_{t-1} + C_t$, $t \in Z$ the top Lyapounov exponent, when $E\left(\log^+ \|F_1\|\right) < \infty$, is defined by

$$\gamma = \inf(n^{-1} E[\log \|F_1 ... F_t\|], n \in N) \tag{3.31}$$

where $\|\cdot\|$ is any norm on $\mathbb{R}^d$, and the operator norm on $(M(d))$ is given by

$$\|M\| = \sup\left\{\frac{\|Mx\|}{\|x\|}; x \in \mathbb{R}^d, x \neq 0\right\} \tag{3.32}$$

In our case the process for U_t is more simple (the coefficient F is a constant) and the top Lyapounov exponent can be defined as $\gamma = \log \rho(F)$ where

$\rho(M) = \max_{1\leq i\leq d} |\lambda_i|$ is the spectral radious of the matrix $M \in \mathbb{R}^{d\times d}$ and $\lambda_1, ..., \lambda_d$ are the eigenvalues of M.

A theorem in Bougerol and Picard (1992) states the strictly stationary solution of stochastic recurrence equations in terms of its top Lyapounov ex-

ponent. For convenience we report it here,

Theorem 3.3 *Suppose that the stochastic recurrence equation* $U_t = F_t U_{t-1} + C_t$, $t \in Z$ *with an* F_t *i.i.d. coefficient is irreducible and that* $E\left(\log^+ \|F_1\|\right) < \infty$ *and* $E[\log^+ \|C_1\|] < \infty$. *Then* U_t *has a non-anticipative strictly stationary solution if and only if the top Lyapounov exponent* γ *is strictly negative.*

Proof. See Bougerol and Picard (1992). ■

The process considered by Bougerol and Picard (1992) is more general than the process (3.27) that we are analysing. The coefficient matrix F in (3.27) is constant while in Theorem 3.3 F_t is an i.i.d. sequence. For a F constant matrix of a specific form we have the following corollary to Theorem 3.3,

Corollary 3.1 *Consider the autoregressive process* $U_t = FU_{t-1} + C_t, \quad t \in Z$ *given by*

$$\begin{bmatrix} u_t \\ . \\ . \\ . \\ u_{t-p-1} \end{bmatrix} = \begin{bmatrix} f_1 & f_2 & . & . & . & f_{p-1} & f_p \\ 1 & 0 & . & . & . & 0 & 0 \\ 0 & 1 & . & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 & 0 \end{bmatrix} \begin{bmatrix} u_{t-1} \\ . \\ . \\ . \\ u_{t-p} \end{bmatrix} + \begin{bmatrix} c_t \\ 0 \\ . \\ . \\ 0 \end{bmatrix} \tag{3.33}$$

If $E[\log^+ \|C_1\|] < \infty$, $E(U_t^2) < \infty$ *and* U_t *has a non-anticipative weakly stationary solution, then this solution is also strictly stationary.*

Proof. If U_t has a weakly stationary solution then $f_i > 0, \forall i, i = 1, ..., p$, $\sum_{i=1}^{p} f_i < 1$ and the spectral radius of F is given such that $\rho(F) < 1$. For a constant matrix F we know that $\gamma = \log \rho(F)$ and therefore $\gamma < 0$. If we add the condition $E(U_t^2) < \infty$ then the result follows. ■

Corollary 3.1 shows that as in the case of the process AGDCC* covariance stationarity is enough to secure strict stationarity. For the AGDCC$^\diamond$ process the connection relies on the behaviour of the Lyapounov exponent associated to the process; when the roots in $1 - \sum_{i=1}^{p} f_i$ are outside the unit circle the Lyapounov exponent will be extrictly negative.

We now use the results of Corollary 3.1 and Proposition 3.4 to provide the conditions for the strict stationary of the AGDCC$^{\diamond}$ model.

Proposition 3.5 *Consider the matrix* $\Pi = I_d - \sum_{i=1}^{q} A_i A_i' - \sum_{j=1}^{p} B_j B_j' - \sum_{k=1}^{r} G_k G_k'$. *The process* X_t *that follows the* $AGDCC^{\diamond}(p,q,r)$ *model described in (3.5), (3.22) and (3.23), is strictly stationary if* Π *is a positive-definite matrix.*

Proof. If Π is a positive-definite matrix then the AGDCC$^{\diamond}(p,q,r)$ process is weak stationary. The matrix F in (3.29) is constant and of the form (3.33), and also $E(Q_t^2) < \infty$ a.s. because $E(\xi_t \xi_t') < \infty$ and $E(\eta_t \eta_t') < \infty$. By Corollary 3.1 Q_t is strictly stationary and by application of Proposition 3.4 X_t is strictly stationary. ■

The previous result provides the conditions for the strict stationarity of the most general form of the model (asymmetric and diagonal). In the following corollary we extend the results in Proposition 3.5 to the other nested models (the GDCC, ADCC and DCC models described in Chapter 2).

Corollary 3.2 *The* $GDCC^{\diamond}$ (p,q), $ADCC^{\diamond}$ (p,q,r), *and* $DCC^{\diamond}(p,q)$ *processes are strictly stationary if they are weakly stationary.*

Proof. For the GDCC$^{\diamond}$ (p,q) process we have that the vector in (3.30) is substituted by

$$C_t = \begin{bmatrix} \overline{Q} - \sum_{i=1}^{q} A_i A_i' \circ \overline{Q} - \sum_{j=1}^{p} B_j B_j' \circ \overline{Q} + \sum_{i=1}^{q} A_i A_i' \circ \xi_{t-i} \xi_{t-i}' \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{3.34}$$

The matrix Π is in this case given by $\left(I_d - \sum_{i=1}^{q} A_i A_i' - \sum_{j=1}^{p} B_j B_j'\right)$. As in Proposition 3.5, $E[\log^+ \|C_1\|] < \infty$ as $E(\xi_t \xi_t') < \infty$.The conditions for $E\left(\log^+ \|F\|\right) < \infty$ and $\gamma < 0$ are the same as for the AGDCC$^{\diamond}$ (p,q,r) model.

For the ADCC$^{\diamond}$ (p,q,r) process, relations (3.29) and (3.30) are

$$F = \begin{bmatrix} b_1 & b_2 & . & . & . & b_{p-1} & b_p \\ 1 & 0 & . & . & . & 0 & 0 \\ 0 & 1 & . & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & 1 & 0 \end{bmatrix} \tag{3.35}$$

$$C_t = \begin{bmatrix} \left\{ \left(1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j\right) \overline{Q} - \left(\sum_{k=1}^{r} g_k\right) \overline{N} \right. \\ \left. + \sum_{i=1}^{q} a_i \left(\xi_{t-i}\xi'_{t-i}\right) + \sum_{k=1}^{r} g_k \left(\eta_{t-k}\eta'_{t-k}\right) \right\} \\ 0 \\ . \\ . \\ 0 \end{bmatrix} \tag{3.36}$$

For the DCC$^{\diamond}$ (p,q) process, F is given as in (3.35) and the C_t vector is

$$C_t = \begin{bmatrix} \left(1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j\right) \overline{Q} + \sum_{i=1}^{q} a_i \left(\xi_{t-i}\xi'_{t-i}\right) \\ 0 \\ . \\ . \\ 0 \end{bmatrix} \tag{3.37}$$

In these cases Π is not a matrix but a scalar. For the ADCC$^{\diamond}$ (p,q,r) process $\Pi = 1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j - \sum_{k=1}^{r} g_k$ and for the DCC$^{\diamond}$ (p,q) process $\Pi = 1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j$. The conditions for weak stationarity are respectively $\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j + \sum_{k=1}^{r} g_k < 1$ and $\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j < 1$. Again $E[\log^{+} \|C_1\|] < \infty$ as $E\left(\xi_t\xi'_t\right) < \infty$ and $E(\eta_t\eta'_t) < \infty$. The result follows by applying Corollary3.1 and Proposition 3.4. ■

We have established the strict stationarity of the AGDCC$^{\diamond}$ process. The last step of our analysis is to prove the identification of the parameter estimates present in the AGDCC$^{\diamond}$ process.

Let us rewrite (3.5) as

$$
\begin{aligned}
P(L) \circ Q_t &= \overline{Q} - \sum_{i=1}^{q} A_i A_i' \circ \overline{Q} - \sum_{i=1}^{p} B_i B_i' \circ \overline{Q} \\
&\quad - \sum_{i=1}^{r} G_i G_i' \circ \overline{N} + Q(L) \circ \xi_t \xi_t' + S(L) \circ \eta_t \eta_t'
\end{aligned} \tag{3.38}
$$

where $S(L) = \sum_{i=1}^{r} G_i G_i' L^i$. We have now three matrices with polynomial coefficients, and the equivalence of representations explained in the previous section can take place in the two pairs $P(L)^{-1}Q(L)$ and $P(L)^{-1}S(L)$.

In order to accommodate the new matrix S in process (3.38) we extend the definition of *minimal specification* presented in Definition 3.4.5 in the following way:

Definition 3.4.5'. We say that the multivariate GARCH (p,q) specification given in (3.38) is minimal if

1. $P(0) = I_d$, $Q(0) = 0$ and $S(0) = 0$
2. $det(P) \neq 0$, $det(Q) \neq 0$ and $det(S) \neq 0$
3. P and Q are coprime or(and) P and S are coprime
4. $\forall j, 1 \leq j \leq d, d_j(P) = d_j \leq p$, $d_j(Q) = d_j \leq q$, and $d_j(S) = d_j \leq r$.
5. P or Q is column reduced, and P or S is column reduced

Proposition 3.6 *Let (P_1, Q_1) define a minimal formulation of a multivariate GARCH (p,q) model, such that there exists a weakly stationary solution denoted ε_t; then, if ε_t is also a solution of another model written with (P_2, Q_2), there exists j, such that $d_j(P_2) > d_j(P_1)$ or $d_j(Q_2) > d_j(Q_1)$..*

Proof. See Jeantheau (1998) ■

The proposition states that if two processes have the same stationary solution, then the supremes of the degrees of the matrices with polynomial coefficients P_1 and P_2 or Q_1 and Q_2 cannot be both equals. The proposition will be used to facilitate the proof in the last theorem of the chapter.

We have now all the tools to present the final result of this chapter, Theorem 3.4, which establishes consistency of the $AGDCC^{\diamond}$ model estimated by AML-MLE.

Theorem 3.4 *Under the assumption that the formulation at θ_0 for the model (3.5) is minimal, the AML-MLE for the AGDCC model defined by (3.5) is consistent.*

Proof. In Proposition 3.5 we proved the strict stationarity of the process, satisfying Assumption 2. If the process is strictly stationary and all parameters are greater than zero then clearly $\det(Q_t) \geq c$ where $c > 0$, satisfying Assumption 3. For Assumption 4 we can apply the same argument as in Proof of Theorem 3.4.1, i.e. covariance stationary implies $E_0(Q_{ii,t}) < \infty$, Jensen's inequality yields $E_0(\log(\det(Q_{ii,t}))) < +\infty$, and given that $\det(Q_t) \geq c$ we get $E_0(|\log(\det(Q_{ii,t}))|) < \infty$.To verify Assumption 5 consider first that for $Q_{t,0}$=$Q_t(\theta_0)$ we have that (3.5) is

$$\begin{aligned} Q_{t,0} &= \overline{Q} - \sum_{i=1}^{q} A_{i,0}A'_{i,0} \circ \overline{Q} - \sum_{i=1}^{p} B_{i,0}B'_{i,0} \circ \overline{Q} - \sum_{i=1}^{r} G_{i,0}G'_{i,0} \circ \overline{N} \qquad (3.39) \\ &+ \sum_{i=1}^{q} A_{i,0}A'_{i,0} \circ \xi_{t-i}\xi'_{t-i} + \sum_{j=1}^{p} B_{j,0}B'_{j,0} \circ Q_{t-j,0} + \sum_{k=1}^{r} G_{k,0}G'_{k,0} \circ \eta_{t-k}\eta'_{t-k} \end{aligned}$$

If $Q_t = Q_{t,0}$ then

$$\begin{aligned} 0 &= \sum_{i=1}^{q} M_i \circ \xi_{t-i}\xi'_{t-i} + \sum_{j=1}^{p} M_{q+j} \circ Q_{t-j} + \sum_{k=1}^{r} M_{q+p+k} \circ \eta_{t-k}\eta'_{t-k} (3.40) \\ &- \sum_{i=1}^{q} M_i \circ \overline{Q} - \sum_{i=1}^{p} M_{q+i} \circ \overline{Q} - \sum_{i=1}^{r} M_{q+p+i} \circ \overline{N} \end{aligned}$$

where $M_i = A_{i,0}A'_{i,0} - A_iA'_i, M_{q+i} = B_{i,0}B'_{i,0} - B_iB'_i$, and $M_{q+p+i} = G_{i,0}G'_{i,0} - G_iG'_i$. We must prove that all these terms are equal to 0. First, (3.40) yields

$$M_1 \circ \xi_{t-i}\xi'_{t-i} = U \qquad (3.41)$$

where U is a $\mathcal{F}_{t-2}$-measurable matrix. Jeantheau (1998) proved that this implies that both M_1 and U are equal to 0. Because $M_1 = 0$, we have

$$\begin{aligned} M_{q+1} \circ Q_{t-j} &= -\sum_{i=1}^{q} M_i \circ \xi_{t-i}\xi'_{t-i} - \sum_{k=1}^{r} M_{q+p+k} \circ \eta_{t-k}\eta'_{t-k} \\ &\quad + \sum_{i=1}^{q} M_i \circ \overline{Q} + \sum_{i=1}^{p} M_{q+i} \circ \overline{Q} + \sum_{i=1}^{r} M_{q+p+i} \circ \overline{N} \end{aligned} \tag{3.42}$$

Now, when M_{q+1} is different from zero we have that if P is column reduced then $M_{q+1}\det(\beta_{d_j}) \neq 0$ because of Definition 3.4.1. By Proposition 3.6 we have that the l.h.s. of (3.42) must have a formulation with at least one column j with $d_j(P)$ lags. This contradicts (3.42) as the r.h.s. has only $d_j(P)-1$ lags, therefore $M_{q+1} = 0$. The same occurs for $M_2, M_{q+2}, M_3, ..., M_{q+p}, M_{q+p+1}, ..., M_{q+p+r}$. If P is not column reduced but Q and S are, the same demonstration holds ■

3.5 Conclusions

In this chapter, we provided conditions for consistency of the MLE for MGARCH models when an Asymmetric Multivariate Laplace Distribution is assumed for standardised residuals. Further, consistency of parameter estimates is proved valid for a variety of specific cases of DCC models, such as AGDCC, GDCC, ADCC, and original DCC. As in Jeantheau (1998), we employed as base for our analysis results provided in Pfanzgal (1969) where conditions for consistency of a minimum contrast estimator are derived.

In the development of the chapter, we proved the strict stationarity of the DCC model. As far as we are concerned this is the first time were the strict stationarity of this model is provided.

Chapter 4

A risk management application

4.1 Introduction

The measurement of the market risk implicit in the holding and trading of financial assets has been an important area of research for academics, practitioners, and regulators over the last years. The most widely used technique to measure market risk is the quantile-based method Value-at-Risk (VaR). This measure is set such that for a specific amount of risk, the net value of a position at some future pre-determined date is smaller than a $quantile_{alpha}$ with probability *alpha*. In what follows, we will refer to the $quantile_{alpha}$ as the "*alpha*% VaR". A comprehensive overview of Value-at-Risk can be found in Duffie and Pan (1997).

Although the VaR method is used in some degree in every risk department of financial institutions, it presents limits often disregarded by practitioners. To explain why, consider a set $V \subseteq \mathbb{R}^S$, where S represents the number of states of the world, and vectors $A \in V$ and $B \in V$ that denote the probable profit and loss of two different portfolios. According to Artzner *et al* (1999), a risk measure $\varrho(A)$ for A is considered *coherent* if it satisfies the following four conditions (axioms):

1. $\varrho(A+B) \leq \varrho(A) + \varrho(B)$ (sub-additivity);
2. $\varrho(tA) = t\varrho(A)$ (homogeneity);
3. $\varrho(A) \geq \varrho(B)$, if $A \leq B$ (monotonicity);

4. $\varrho(A + r.n) = \varrho(A) - n$ (translation invariance condition).

Sub-additivity means that the risk found in the combination of two portfolios cannot be greater that the sum of the risks of each one of the portfolios. Risk measures that do not comply with this condition are not congruent with the important notion of diversification. Homogeneity states that the risk measure should be a linear function of the size of the portfolio. Monotonicity means that if portfolio B is always worth at least as much as A,then B cannot be riskier than A. Finally, translation invariance condition implies that the measure of risk is expressed in appropriate units.

The VaR measure is not coherent because it fails the axiom of sub-additivity. The computation of VaR cannot be divided into separate computations because of its non-additivity by risk variable or by position. Non-additivity by position means that the sum of two partial VaR's does not equal the total VaR. This implies that every time the position of a portfolio changes, the VaR has to be re-computed. The non-additivity by risk variable means that the sum of the individual VaR's of a structure product does not equal de VaR of the entire structure.

.

A risk measure based on quantiles that instead satisfies the four conditions is the expected shortfall (ES) defined as $E(Loss \mid Loss > alpha\%$ VaR). Intuitively, it quantifies the size of the loss once the VaR level has been broken. This additional information is important for the risk manager, it answers the question "how bad is bad?", while the VaR only answers "is it bad?". To understand why the additional information provided by the ES measure is very relevant for the risk manager, consider the tails of two distributions plotted in Figure 4.1

[Insert Figure 4.1 here]

The 95%VaR given by the two distributions is the same. Nonetheless, it can be seen that a portfolio with returns following Distribution 2 is riskier than a portfolio with returns following Distribution 1 between the 95%VaR value and level a, while the portfolio with Distribution 1 is riskier from level a onwards. The risk manager has no information of the magnitude of risk

involved in the portfolio once the 95%VaR has been reached. This additional information is provided by the ES measure.

The main purpose of this paper is to evaluate the convenience, as well as benefits, of using the AGDCC model estimated with the AML distribution in the computation of market risk measures. We are interested in realistic applications where portfolios are formed by a big number of assets. To this end we employed the data described in Chapter 2 (34 assets), but increasing the number of portfolios to 10 by generating random vectors of weights in order to evaluate a big number of possible portfolio distributions.

Our procedure combines several techniques taken from the financial econometrics literature and from the extreme value theory (EVT) literature. We evaluate the use of the model presented in Chapter 2 in the estimation of conditional heteroskedasticity in the raw vector of financial returns. Other studies (for instance, McNeil and Frey, 2000), have shown the benefits of the estimation of the conditional variance by GARCH models and other volatility models in the performance of VaR calculations, but these have been restricted to the univariate case. Other papers such as Verbeek and Rombouts(2005), Giot and Laurent (2003), and Allen *et al* (2004) also use MGARCH models, but their analysis is restricted to a very small number of assets (less then five).

We enrich the time-varying analysis of the variance by employing the asymmetric multivariate Laplace DCC (AML-DCC) that assumes a more realistic distribution for the series of conditional returns by allowing for asymmetries and leptokurtosis. The inclusion of these two features in the distribution of innovations is of paramount importance. First, given that the VaR and ES are quantile-based measures and that the quantile is normally located on the tail of the distribution, the excess kurtosis effect must be accurately captured. Secondly, because positions in financial assets can be positive or negative it makes a difference if the skewness, which is usually found in the return of financial assets, is ignored.

In the estimation of the VaR and ES measures, is not only important the choice of filter for the estimation of conditional volatilities and covariances, but also the choice of the distribution employed for the estimation of the risk quantile. We extend to a multivariate case the analysis in McNeil and Frey (2000) and consider the generalised Pareto distribution (GPD), a distribution

commonly used in extreme value theory (EVT) for the modeling of the tails of power law distributions.

The chapter is organised as follows. Section 2 briefly describes the models used to filter the raw series of asset returns, and the features of the generalised Pareto distribution that make it appealing for the estimation of risk quantiles. Section 3 explains the estimation of the VaR and ES, and the methodology that we followed for the backtesting analysis. In Section 5 we test the different methodologies by performing an empirical exercise on a large data set. Section 6 concludes.

4.2 The AGDCC model and the generalised Pareto distribution

In this section we describe the models used in the estimation of time-varying variance-covariance matrix of the returns processes, as well as the EVT parametric distribution used for the estimation of VaR and E.S. quantiles.

4.2.1 The MGARCH model

There is empirical evidence that the use of GARCH models enhance the accuracy of risk measures of portfolios composed of financial assets (McNeil and Frey (2000), Allen *et al* (2004)). To this end, we apply the dynamics of the asymmetric multivariate-DCC model (AGDCC (1,1,1)) presented in Chapter 2

$$X_t = H_t^{1/2}\xi_t \tag{4.1}$$

$$H_t = D_t R_t D_t \tag{4.2}$$

$$R_t = (diagQ_t)^{-1/2} Q_t (diagQ_t)^{-1/2} \tag{4.3}$$

$$\begin{aligned}\mathbf{Q}_t &= \left(\overline{\mathbf{Q}} - \mathbf{A}'\overline{\mathbf{Q}}\mathbf{A} - \mathbf{B}'\overline{\mathbf{Q}}\mathbf{B} - \mathbf{G}'\overline{\mathbf{N}}\mathbf{G}\right) + \mathbf{A}'\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}\mathbf{A} + \mathbf{B}'\mathbf{Q}_{t-1}\mathbf{B} \\ &\quad + \mathbf{G}'\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1}\mathbf{G}\end{aligned} \tag{4.4}$$

As we explained in Chapter 2, this representation is very convenient because it allows a two-step estimation of parameters, that allows estimation when a big number of assets is considered. The estimation was carried out assuming the asymmetric multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003). In this Chapter, we explore the possibility of a mixture of distributions in the estimation process. We estimate the parameters in the first stage (when the estimates of univariate volatilities are obtained) assuming a normal distribution, and the parameters in the second stage (parameters in the correlation process) assuming the AML distribution. We do this in order to evaluate the impact of allowing for different specifications in the univariate volatility processes. As explained in Chapter 2, when the AML distribution is assumed the parameters in the univariate volatilities have to be estimated in a single procedure optimasing one single log-likelihood function. The estimation does not only takes more time than when each volatility process is estimated separately, but it also force a specification where all the processes follow the same dynamics. When normality is assumed for the first stage, we can have heterogeneity in the dynamics of univariate volatilities. With respect to the quality of standardised residuals to be used in the second stage, the matching of the first two moments is guaranteed, as can be seen by the representation of the AML distribution as a location-scale mixture of normal distributions

$$\mathbf{X} \stackrel{D}{=} \mathbf{m}\boldsymbol{\xi} + \boldsymbol{\xi}^{1/2}\mathbf{Y} \tag{4.5}$$

where "$\stackrel{D}{=}$" denotes equality in distribution, $\mathbf{Y} \sim N_n(\mathbf{0}, \mathbf{H})$, $\boldsymbol{\mu} \in \mathbb{R}^n, \boldsymbol{\xi} \sim EXP(1)$, and $\mathbf{m}$ is the asymmetry parameter. In the first stage we assume $\mathbf{m} = \mathbf{0}$, therefore, and as in the normal distribution, $E(\mathbf{X}) = \mathbf{0}$ and $Var(\mathbf{X}) = \mathbf{H}$. The trade-off is that leptokurtosis in the residuals obtained in a first stage will not be fully captured.

4.2.2 Extreme value theory: the generalised Pareto distribution

The main purpose of the EVT is to model extreme properties of a random variable Z based on a n-sample of observations $Z_1, ..., Z_n$. The objective is to capture the extreme behaviour by analyzing only the tail of the underlaying distribution of Z, i.e. those realizations of the observed sample that go beyond a specific threshold k. The use of EVT for financial risk management purposes is specially appealing, because one of the main stylized facts in the return of financial assets is the heavy tail phenomenon[1].

The rationale for the preference of EVT distributions in some occasions over the Gaussian distribution for risk management applications can be seen in the following argument. Consider the set of i.i.d. random variables $X_1, X_2, ..., X_n$, $n \in \mathbb{N}$, representing negative returns. If $E(X) = \mu$ and $var(X) = \sigma^2 < \infty$ then by the central limit theorem we have that the partial sums $S_n = X_1 + X_2 + ... + X_n$ satisfy

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) = N(x), \quad x \in \mathbb{R} \tag{4.6}$$

where N is the standard normal cdf. It is clear from (4.6) that the result holds when the variance contributions are asymptotically negligible

The basic notion in EVT is the one of sample maxima. Let $M_n = \max(X_1, X_2, ..., X_n)$, $n > 1$, and $M_1 = X_1$. Then

$$P(M_n \leq x) = P(X_1 \leq x, ..., X_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = F^n(x) \tag{4.7}$$

As we are interested on extremes, the goal is to establish a central limit theorem for the maxima. The question is whether there exists a normalisation, i.e. a nonnegative sequence $\{a_n\}$ and a sequence $\{b_n\}$, so that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} Y \tag{4.8}$$

[1] Embrechts et al. (1997) offer a comprehensive description of the use of EVT in finance and insurance. Kluppelberg (2002) discusses the role of EVT in risk management.

If H is the distribution of Y, this amounts to saying that for every point of continuity of H,

$$\lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n \leq x) = H(x), \quad x \in \mathbb{R} \tag{4.9}$$

If this is the case, we say that the distribution function F belongs to the maximum domain of attraction (MDA) of the distribution function H and we write $F \in MDA(H)$. The Fisher-Tippett theorem (a very basic result in EVT) states that H is one of the following three distribution functions:

- Frechet, $\Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left(-x^{-\alpha}\right), & x > 0, \end{cases}$ for $\alpha > 0$
- Gumbel, $\Lambda(x) = \exp(-e^{-x}), \quad x \in \mathbb{R}$
- Weibull, $\Psi_\alpha(x) = \begin{cases} \exp\left(-(-x)^{\alpha}\right), & x \leq 0, \\ 1, & x > 0, \end{cases}$ for $\alpha > 0$

Many continuous distribution functions are in MDA(H). For example, the exponential, normal, and lognormal distributions are in MDA(Λ), the uniform distribution is in MDA(Ψ_1), and the Pareto distributions in MDA(Φ_α).

The three distributions can be summarised in the generalised extreme value (GEV) distribution

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right), & \xi \neq 0, \\ \exp\left(-e^{-x}\right), & \xi = 0, \end{cases} \tag{4.10}$$

where $1 + \xi x > 0$. The GEV represents all three extreme types: Frechet when $\xi > 0$, Gumbel when $\xi = 0$, and Weibull when $\xi < 0$.

As wee see, EVT offers several alternatives to model distributions with heavy tails. In the case where we are mainly interested in the behaviour of large observations which exceed a high threshold, probably the best approximation is given by the generalised Pareto distribution. To motivate the use of this distribution consider the following result provided by Pickands (1975) and Balkema and de Haan (1974):

Theorem. [Pickands (1975) and Balkema and de Haan (1974)]

$$F \in MDA(H_\xi) \Leftrightarrow \lim_{k \nearrow x_f} \frac{\overline{F}(k + x\beta(k))}{\overline{F}(k)} = \begin{cases} (1+\xi x)^{-1/\xi}, & \xi \neq 0, \\ e^{-x}, & \xi = 0, \end{cases} \tag{4.11}$$

where $\overline{F}(k) = 1 - F(k), 1 + \xi x > 0$, and $\beta(k)$ is some measurable function.

The result states that given a threshold k, for a random variable X with d.f. $F \in MDA(H_\xi)$ we have

$$\lim_{k \uparrow x_f} P\left(\frac{X-k}{\beta(k)} > x \mid X > k\right) = \begin{cases} (1+\xi x)^{-1/\xi}, & \xi \neq 0, \\ e^{-x}, & \xi = 0, \end{cases} \tag{4.12}$$

We are interested in the distribution of points bellow the (finite or infinite) right endpoint of F. This is given by the Generalised Pareto Distribution (GPD), which based on (4.11) has the form

$$G(x) = \begin{cases} 1 - (1 + x\frac{\xi}{\beta})^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-x/\beta} & \text{if } \xi = 0 \end{cases} \tag{4.13}$$

for $1 + \xi x > 0$. ξ is the shape parameter and $\beta > 0$ is the scale parameter. The parameter ξ defines the form of the tails of the distribution. For $\xi > 0$ the distribution obeys a power law for the tails, i.e. heavy tails. For $\xi = 0$ we have instead an exponential law for the tails such as the one followed by the normal distribution. The case $\xi < 0$, which is of no practical relevance for financial applications, groups the light-tailed distributions.

The importance of the Balkema and de Haan (1974) and Pickand (1975) results is that the distribution of excesses may be approximated by the GPD by choosing ξ and β and setting a high threshold k. The estimation of ξ and β can be done by maximum likelihood. Hosking and Wallis (1987) presents regularity conditions for consistency and asymptotic normality of parameter estimates in the GPD when these are obtained by MLE, provided that data after the threshold are i.i.d. and $\xi > -1/2$.

The setting of the threshold k is not straight forward and affects the behaviour of $\widehat{\xi}$ and $\widehat{\beta}$. The selection of k, and henceforth of the number of data points to be considered as extremes, presents the same type of problem as the

selection of the bandwidth in nonparametric estimation, i.e. a bias-variance trade-off in the parameter estimates. We choose k high to reduce the bias (moving toward the centre of the distribution induces bias because the power law intrinsic in the distribution is assumed to hold only in the extreme tail) while keeping the number of extreme points large to reduce the variance of the parameter estimates. McNeil and Frey (2000) present a simulation study of threshold choice for finite samples. They compare the tail estimator obtained by a GPD fitted by MLE, a Hill estimator (the most popular estimation method for extreme values), and an approach based on the empirical distribution function (historical simulation). They estimate bias an MSE against k for various estimators of the 0.99 quantile of a t-distribution with four degrees of freedom based on an i.i.d. sample of 1000 points. They found that for a range of k between 50 and 350 points the GPD approach is the one with the lowest bias and also the most efficient for great part of the variation in k, i.e. the most robust. Their main conclusion is that the choice of the threshold level is not so critical when the GPD method is selected.

4.3 Risk measures: Value-at-Risk and Expected Shortfall

4.3.1 Value-at-Risk (VaR)

There are several techniques to approximate the distribution of returns in the VaR framework: parametric methods, historical simulation, Monte Carlo simulation, and stress-testing. In this chapter we focus on the model-based parametric method were variations in the portfolio are characterized by a parametric distribution. Formally, consider the portfolio return,

$$\mathbf{r}_p = \sum_{i=1}^{n} w_i r_i = \mathbf{w}'\mathbf{r} \tag{4.14}$$

where $w_1 + ... + w_n = 1$. The $alpha$%VaR level is the solution to,

$$alpha\%VaR = \int_{-\infty}^{VaR} f(\mathbf{r}_p)d\mathbf{r}_p \tag{4.15}$$

where $f(\mathbf{r}_p)$ is the density function of $\mathbf{r}_p$. As explained in Chapter 2, stable and geometric stable distributions have the additivity property that allow us to use them in the modelling of portfolio returns. In the special case where $f(\mathbf{r}_p)$ is the density of the AML distribution and assuming that the magnitude of the mean vector $\boldsymbol{\mu}_t$ is considerably less significant than the size of $\boldsymbol{\Sigma}_t^{1/2}$ and, therefore, can be ignored, the conditional VaR implicit in (4.15) reduces to,

$$VaR_t = (\mathbf{w}_t'\boldsymbol{\Sigma}_t\mathbf{w}_t)^{1/2}\, x_q \tag{4.16}$$

where x_q is the risk quantile at the *alpha* level of a specific distribution. In our analysis, we used for the computation of the risk quantile the normal, AML, and GPD distributions.

For the GPD the quantile can be derived from the tail estimator. The distribution of excess values of y over threshold k is defined by the conditional probability

$$F_k(y) = P(X \leq y + k \mid X > k) = \frac{F(y+k) - F(k)}{1 - F(k)}, y > 0 \tag{4.17}$$

Since $x = y + k$ for $X > k$, we have the following representation

$$F(x) = [1 - F(k)]F_k(y) + F(k) \tag{4.18}$$

where $x > k$. Since $F_k(y)$ converges to the GPD for sufficiently large k, we can write

$$F(x) = [1 - F(k)]G_{\xi,\beta(k)}(y) + F(k) \tag{4.19}$$

We can estimate nonparametrically $F(k)$ by means of the empirical cdf

$$\widehat{F}(k) = \frac{n - N_k}{n} \tag{4.20}$$

where N_k is the number of exceedences and n is the sample size. Substi-

tuting (4.20) into (4.19) we get

$$\widehat{F}(x) = 1 - \frac{N_k}{n}\left(1 + \widehat{\xi}\frac{x-k}{\widehat{\beta}}\right)^{-1/\widehat{\xi}} \tag{4.21}$$

The $VaR_{t,q}$ is the qth quantile of the distribution F at time t. Defining F^{-1} as the inverse function of F we have

$$VaR_{t,q} = F^{-1}(q) \tag{4.22}$$

For $q > F(k)$, an estimate of the VaR_q can be obtained from (4.21) by solving for x

$$\widehat{VaR_q} = k + \frac{\widehat{\beta}}{\widehat{\xi}}\left(\left(\frac{1-q}{N_k/n}\right)^{-\widehat{\xi}} - 1\right) \tag{4.23}$$

4.3.2 Expected Shortfall (ES)

The expected shortfall measure also depends upon the choice of the distribution for the risk quantile: $ES = E(Loss \mid Loss > alpha\%VaR)$. For the one-step ahead predictive distribution the conditional expected shortfall of the random variable v can be defined as

$$ES_q^t = \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t E(v \mid v > x_q) \tag{4.24}$$

As in the VaR methodology, the key of the measure is in the specification of the distribution of v .

To estimate this measure we require an estimate of the expected shortfall for the innovation distribution $E(X \mid X > x_q)$. In order to do this, consider the continuous random variable v with probability density function $f(v)$. We can write

$$f(v \mid v > a) = \frac{f(v)}{\Pr(v > a)} \tag{4.25}$$

where a is a constant. The expectation of a truncated random variable is given by

$$E(v \mid v > a) = \int_a^{\infty} v f(v \mid v > a) dv \tag{4.26}$$

therefore,

$$E(v \mid v > a) = \int_a^{\infty} \frac{v f(v)}{\Pr(v > a)} dv \tag{4.27}$$

For the normal distribution the relationship evaluated at the risk quantile level takes the form (known as the Mill's ratio)

$$E(v \mid v > \widehat{x_q}) = \frac{\phi(\widehat{x_q})}{1 - \Phi(\widehat{x_q})} \tag{4.28}$$

where $\phi(v)$ and $\Phi(v)$ are, respectively, the density function and the cumulative distribution function of a standard normal variable v.

For the case of the generalised Pareto distribution with cumulative distribution function

$$G(y) = \begin{cases} 1 - (1 + y\frac{\xi}{\beta})^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-y/\beta} & \text{if } \xi = 0 \end{cases} \tag{4.29}$$

the expected shortfall is given by (McNeil and Frey (2000))

$$E(v \mid v > \widehat{x_q}) = \left(\frac{\widehat{\beta} - \widehat{\xi} x_{(k+1)} + \widehat{x_q}}{1 - \widehat{\xi}} \right) \tag{4.30}$$

where $x_{(k+1)}$ is the quantile defined by the first observation above the threshold k

4.3.3 Backtesting procedures

For the backtesting of the VaR measure, we consider the unconditional coverage (uc), independence (ind), and conditional coverage (cc) test of Christoffersen and Pelletier (2004). Consider the hit sequence of VaR violations defined as,

$$I_t = \begin{cases} 1, \text{if } v_t < alpha\%VaR \\ 0, \text{ else} \end{cases} \tag{4.31}$$

In the *uc* test we test the null hypothesis that I_t is i.i.d. Bernoulli with parameter α, against the alternative that the sequence is i.i.d. Bernoulli with parameter π, where π is the ratio of the number of violations over the number of observations. If the VaR method is correct the empirical failure rate π must be equal to α.

To implement the likelihood ratio test, we have that the likelihood function from a Bernoulli variable z, with parameter p is given by

$$L(z;p) = (1-p)^{t-t1}p^{t1} \tag{4.32}$$

where t is the sample size and $t1$ is the number of VaR violations. The likelihood ratio test of unconditional coverage is then

$$LR_{uc} = 2\left(\ln L(z;\widehat{\pi}_1) - \ln L(z;p)\right) \tag{4.33}$$

where $\widehat{\pi}_1 = t1/t$

The *ind* test tests explicitly the assumption of independence of the hit sequence,

$$H_{0,ind:}\pi_{01} = \pi_{11} \tag{4.34}$$

where π_{ij} is the probability of event i occurring on day $t-1$ and event j occurring on day t. The statistic for this test is

$$LR_{ind} = 2\left(\ln L(z;\widehat{\pi}_{01},\widehat{\pi}_{11}) - \ln L(z;\widehat{\pi}_1)\right) \tag{4.35}$$

where

$$L(z;\pi_{01},\pi_{11}) = (1-\pi_{01})^{t0-t01}\pi_{01}^{t01}(1-\pi_{11})^{t1-t11}\pi_{11}^{t11} \tag{4.36}$$

where tij is the number of observations with a j followed by an i, $\widehat{\pi}_{01} = t01/t0$, and $\widehat{\pi}_{11} = t11/t1$. In the case where $t1 = 0$, we replace equation (4.36) by

$$L(z;\pi_{01},\pi_{11}) = (1-\pi_{01})^{t0-t01}\pi_{01}^{t01} \tag{4.37}$$

Neither the *uc* test nor the *ind* test are complete by their own, the first

one test that on a average the coverage implicit by the VaR model is correct, while the second tests the clustering effect on the failures without testing the correct number of failures. The *cc* test combines both tests:

$$H_{0,cc:}\pi_{01} = \pi_{11} = \alpha \tag{4.38}$$

The likelihood ratio test for conditional coverage is given by

$$LR_{cc} = 2\left(\ln L(z;\widehat{\pi}_{01},\widehat{\pi}_{11}) - \ln L(z;p)\right) \tag{4.39}$$

This is the main test for the evaluation of VaR estimates.

Under the null the likelihood ratio test of unconditional coverage (LR_{uc}) and the likelihood ratio test of independence (LR_{ind}) are χ^2 with one degree of freedom. Under the null the likelihood ratio test of conditional coverage (LR_{cc}) is χ^2 with two degrees of freedom.

While the large-sample distribution of the LR tests described above is theoretically correct, the dearth of violations of 1% VaR or even 5% VaR make the effective sample size rather small, even when the nominal size is large. To overcome this problem and obtain p-values robust to finite sample scenarios we employed as in Christoffersen and Pelletier (2004) the Monte Carlo tests of Dufour (2004). We generate N independent realisations for each one of the three test statistics: $LR_{i,type}, i = 1, ..., N$, $type = uc, ind, cc$. The cases $LR_{o,type}$ correspond to the calculated test statistic. The Monte Carlo p-value $\widehat{p}_N(LR_0)$ is given by

$$\widehat{p}_N(LR_0) = \frac{\widehat{G}_N(LR_0) + 1}{N + 1} \tag{4.40}$$

where

$$\widehat{G}_N(LR_0) = N - \sum_{i=1}^{N} \mathbf{I}(LR_i < LR_0) + \sum_{i=1}^{N} \mathbf{I}(LR_i = LR_0)\mathbf{I}(U_i \geq U_0) \tag{4.41}$$

$\mathbf{I}(\cdot)$ is the indicator function and $U_i, i = 0, ..., N$ are independent realisations of a Uniform distribution on the [0,1] interval.

For the backtesting of the ES we follow the approach proposed by McNeil and Frey (2000). Define the residuals

$$\varkappa_{i,t} = \frac{r_{i,t+1}^{violation} - \widehat{ES}_q^t}{\widehat{\sigma}_{t+1}^p} = \varepsilon_{t+1} - E(\mathbf{r} \mid r > x_q) \tag{4.42}$$

where $r_{i,t+1}^{violation}$ is the portfolio return when a violation occurs, i.e. when $r_{i,t+1}^{violation} > VaR_{q,t}$. Under a correct specification, $\varkappa_{i,t}$ must be i.i.d. with $E(\varkappa_{i,t}) = 0$.

Because the number of VaR violations is very small (specially for the 0.5% VaR) and also because the distribution of residuals is highly non-normal, we construct a test based on a bootstrap procedure described in Efron and Tibshirani (1993). Since the most dangerous scenario is the one where the conditional ES is underestimated, i.e. when $r_{i,t+1}^{violation} > \widehat{ES}_q^t$, the test is settled as

$$H_{0:}\ mean_{boot}(\varkappa_{i,t}) \leq 0 \tag{4.43}$$

$$H_{1:}\ mean_{boot}(\varkappa_{i,t}) > 0 \tag{4.44}$$

where $mean_{boot}(\varkappa_{i,t})$ is the mean obtained by bootstraping $\varkappa_{i,t}$.

4.4 Empirical applications

We estimate both risk management measures, VaR and ES, under a set of hypothetical portfolios and then we proceed to evaluate the results by means of a backtesting analysis. The data are the same used in Chapter 2. We consider shares indices of 21 countries listed in the FTSE All-World Indices and bond indices of 13 countries constructed by Datastream. We refer the interested reader to Cappiello *et al* (2004) for a detailed description of the data[2]. The

[2]We wish to thank Kevin Sheppard for providing us with the dataset.

frequency is weekly and spans over the period 08/01/1987-07/02/2002 (785 observations). The 21 countries of the share indices are: Australia, Austria, Belgium, Canada, Denmark, France, Germany, Hong Kong, Ireland, Italy, Japan, Mexico, Netherlands, New Zealand, Norway, Singapore, Spain, Sweden, Switzerland, United Kingdom, and the United States. The 12 countries of the bond indices are Austria, Belgium, Canada, Denmark, France, Germany, Ireland, Japan, Netherlands, Sweden, Switzerland, and the United Kingdom.

In order to evaluate the behaviour of the different models under a broad choice of distributions for portfolio returns, we generate randomly, and under a uniform distribution, 10 sets of weights for the allocation of individual assets in each one of the hypothetical portfolios

$$r_{portfolio_i} = r_t W_i, \quad i = 1...10 \tag{4.45}$$

$$W_i = \begin{pmatrix} ran_1 \\ . \\ . \\ . \\ ran_{34} \end{pmatrix} \tag{4.46}$$

where $\sum_{j=1}^{34} ran_j = 1$.

As explained in Section 2, we explore the possibility of mixing the normal and AML distributions in the estimation process of the variance-covariance matrices. To this end, we estimate three conditional variance matrices $\mathbf{H}_t^{m_s}$, $s = 1..3$, m_1 =AGDCC (1,1,1) model with first stage normal and second stage normal, m_2 =AGDCC (1,1,1) model with first stage normal and second stage AML, and m_3 =AGDCC (1,1,1) model with first stage AML and second stage AML. For the conditional variance matrices $\mathbf{H}_t^{m_1}$ and $\mathbf{H}_t^{m_2}$ we can estimate the parameters in the first stage with different specifications for univariate volatilities. As in Cappiello *et al* (2004), we select the best specification by means of the Bayesian Information Criterion. The sample of univariate GARCH processes used in the analysis is presented in Table 4.1

The list of models selected for each asset are presented in Tables 4.2a,b.

[Insert Tables 4.1-4.2a,b here]

Following with our parametric methodology for the estimation of the VaR and ES, we form variances for each one of the ten portfolios by computing $(\sigma_t^i)^2 = W_i' \mathbf{H}_t^{m_s} W_i$. This yields a total of 30 variance portfolios.

In the estimation of the risk quantile we employed the normal, asymmetric Laplace, and G.P. distributions. Prior to the estimation of the parameter estimates in the G.P. distribution ($\widehat{\xi}$ and $\widehat{\beta}$), it is necessary to determine to threshold point k. As mentioned by Diebold *et al* (1999) in a study of the pitfalls and advantages in the use of EVT in risk management, the uncertainty in the estimation of k remains as one of the main reasons why the methodology has not been fully implemented in the risk management literature. Because of this, we devote some part of this section to the analysis of the selection of the threshold in the G.P. distribution. First, based on relation (4.23), we present in Figure 4.2 plots showing the variation of the loss estimate based on a 99% quantile with respect to a change in the threshold

[Insert Figure 4.2 here]

Plot A. presents the case where the data have been standardised by the covariance matrix $\mathbf{H}_t^{m_1}$, Plot B. the case where the data have been standardised by the covariance matrix $\mathbf{H}_t^{m_2}$, and Plot C. the case where the data have been standardised by the covariance matrix $\mathbf{H}_t^{m3}$. Dotted lines are upper and lower 95 percent confidence intervals. For this analysis we use a single portfolio where all assets have the same weight. With this figure we get a first indication of the best choice of threshold; the optimum loss level for this quantile is attained at a threshold located between 60 and 90 exceedencies that as in McNeil and Frey (2000) roughly corresponds to 10% of total observations.

In order to analyse the sensitivity of the shape parameter ξ of the G.P. distribution in relation to the value of the threshold, we plot the estimated value of the parameter and 95% confidence interval against values of k ranging from 10 to 500.

[Insert Figure 4.3 here]

Several facts are visible in the plots. First, it is evident the decrease in the variance of the parameter estimate when the number of exceedencies increases. The effect in the bias of $\widehat{\xi}$ is also evident when the number of exceedencies increases; once the number of observations passes an approximated level of one hundred $\widehat{\xi}$ becomes more and more negative, suggesting what would be a light-tailed distribution. This confirms the bias-variance trade-off reported in McNeil and Frey (2000). Nevertheless, we also found that the biasness effect is not uniform; in Plots A and C the value of $\widehat{\xi}$ is very negative for very small number of exceedencies. In Plot C for example, $\widehat{\xi}$ is only positive for a very small interval between 70 and 90 observations. A threshold level where the value of $\widehat{\xi}$ is negative would be problematic because it would imply a platikurtic distribution, which is rarely the case of returns of financial assets.

The results obtained from Figure 4.3 reinforce the idea that the best level is around the 70 exceedencies level. Nonetheless, and in order to produce a fair comparison with the normal and Laplace distributions, we have estimated the G.P. distribution with four different threshold levels: 50, 70, 100, and 150 exceedencies.

To evaluate how well the distribution fits the standardised residuals, we compute the tail estimate over the 70 exceedencies threshold level (Figure 4.4), and construct the QQ-plot of residuals over the same threshold level (Figure 4.5)

[Insert Figures 4.4-4.5 here]

The backtesting results of our exercise for the VaR measure are presented in Tables 4.3a,b,c.

[Insert Table 4.3a,b,c here]

Table 4.3a. presents the results for the unconditional test, Table 4.3b of the independent test, and Table 4.3ac of the conditional test. The label AGDCC_1 corresponds to the Asymmetric-Generalised-DCC model where the first and second stages are estimated assuming a multivariate normal distribution for the standardised residuals, AGDCC_2 corresponds to the case where the first stage is estimated assuming a multivariate normal distribution

and the second stage assuming the asymmetric multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003). Finally, AGDCC_3 corresponds to the case where the first and second stages are estimated assuming a the AML distribution.

In Table 4.3a we can see the mixture of results in terms of best models. At the 0.05 confidence level the best model is AGDCC_1 with a Laplace distribution for the quantile, at the 0.01 level the best is AGDCC_2 with a normal distribution for the quantile, and at the 0.005 level the best is the AGDCC_2 once again but with a G.P. distribution with $k = 70$ for the quantile. It stands out the fact of how critical is the selection of the threshold for the estimation of the G.P. distribution. At the 0.05 and 0.01 confidence levels the performance of the measures is highly affected when the threshold in not optimal. As shown in Figure 4.3, for thresholds far away from the 70 exceedencies level the distribution becomes light-tailed, affecting the performance of the models in terms of capturing the correct number of VaR failures.

In terms of clustering of VaR violations, Table 4.3b shows how the best option is the AGDCC_3 model with a G.P. distribution for the quantile. Table 4.3c is the most important panel as it presents the results of the conditional test, where the number and clustering of VaR failures are considered in a single test. The result is quite mixed: at the 0.05 confidence level the best model is AGDCC_3 with a G.P. with $k = 70$ distribution for the quantile, at the 0.01 level the best is AGDCC_2 with a normal distribution for the quantile, and at the 0.005 level the best is the AGDCC_1 with a G.P. Laplace distribution for the quantile.

The backtesting results for the ES measure are presented in Table 4.4

[Insert Table 4.4 here]

In this case the results are very strong towards a single model: the AGDCC_2 for the variance-covariance matrix and the Laplace distribution for the estimation of the quantile. The big improvement compared to other models comes from the selection of the Laplace distribution for the quantile and from the heterogeneity allowed in the dynamics of univariate volatilities: the second and third best models for each risk-quantile were respectively the AGDCC_1 and AGDCC_3 coupled with a Laplace distribution for the quantile. We also

see that for the E.S. measure the choice of the threshold for the estimation of the G.P. distribution is pretty much irrelevant; the variation of the p-values for the four choices is very small for the three MGARCH models. Also, and in contradiction to McNeil and Frey (2000), we found that the quality of the E.S. measure computed using the normal distribution for the risk quantile is quite acceptable (p-values in all cases are well above 0.05), even more, the difference with the results of the G.P. distribution is rather small.

4.5 Conclusions

In this chapter, we implemented the multivariate GARCH model presented in Chapter 2 and Chapter 3 in a series of risk management applications. We explored an estimation procedure where conditional univariate volatilities are estimated under the assumption of normality and conditional correlations are estimated under the assumption of an asymmetric multivariate Laplace distribution for the returns. For the estimation of risk quantiles in the computation of VaR and expected shortfall we employed the normal, Laplace,and generalised Pareto distributions. In this way, we coupled tools from the financial econometrics literature with Extreme Value Theory. We measured the quality of the risk measures by performing backtesting analysis on both risk measures.

We may summarise the main findings in this Chapter as follows,

- The overall performance of conditional-VaR and conditional-E.S. as risk measures (without considering the specific MGARCH model and quantile distribution used) is satisfactory, as demonstrated by the backtesting results. This coincides with the findings of McNeil and Frey (2000) who performed a similar approach (univariate) on two share indices (S&P and DAX), one stock (BMW), on FX (\$/£) and one commodity (gold).
- No model was identified as unique winner for the VaR measure. In the conditional test we found that for the three confidence levels evaluated (0.05, 0.01, and 0.005), different estimation procedures for the MGARCH model and different quantile distributions outperform in some cases.

- For the expected shortfall measure (which is considered a superior risk measure compared to Value-at_Risk in terms of coherence) we found the AGDCC model estimated in a first stage under normality and under a second one under the AML distribution and coupled with a Laplace distribution for the risk quantile, as the best performing model. This result is encouraging. The performance was much better than the model where the generalised Pareto distribution was used.

- We found the setting of the threshold for the estimation of the G.P. distribution very important for the estimation of the VaR measure but pretty much irrelevant for the computation of the ES measure.

Table 4.1: Sample of univariate-GARCH models used for the estimation of the first stage in the AGDCC (1,1,1) model when multivariate normality is assumed for standardised residuals

GARCH:	$h_t = \varpi + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$
AVGARCH(Absolute Value):	$h_t^{1/2} = \varpi + \alpha\left\|\varepsilon_{t-1}\right\| + \beta h_{t-1}^{1/2}$
NGARCH (Non-linear):	$h_t^{\lambda/2} = \varpi + \alpha\left\|\varepsilon_{t-1}\right\|^{\lambda} + \beta h_{t-1}^{\lambda/2}$
EGARCH (Exponential):	$log(h_t) = \varpi + \alpha\frac{\left\|\varepsilon_{t-1}\right\|}{h_{t-1}^{1/2}} + \lambda\frac{\varepsilon_{t-1}}{h_{t-1}^{1/2}} + \beta\log(h_{t-1})$
TGARCH (Threshold):	$h_t^{1/2} = \varpi + \alpha\left\|\varepsilon_{t-1}\right\| + \lambda I\left[\varepsilon_{t-1} < 0\right]\left\|\varepsilon_{t-1}\right\| + \beta h_{t-1}^{1/2}$
GLR-GARCH:	$h_t = \varpi + \alpha\varepsilon_{t-1}^2 + \lambda I\left[\varepsilon_{t-1} < 0\right]\varepsilon_{t-1}^2 + \beta h_{t-1}$
APARCH (Asym. Power):	$h_t^{\lambda/2} = \varpi + \alpha\left\|\varepsilon_{t-1}\right\|^{\lambda} + \lambda I\left[\varepsilon_{t-1} < 0\right]\left\|\varepsilon_{t-1}\right\|\lambda + \beta h_{t-1}^{\lambda/2}$
AGARCH (Asymmetric):	$h_t = \varpi + \alpha\left(\varepsilon_{t-1} + \lambda\right)^2 + \beta h_{t-1}$
NAGARCH (Nonlinear):	$h_t = \varpi + \alpha\left(\varepsilon_{t-1} + \lambda\sqrt{h_{t-1}}\right)^2 + \beta h_{t-1}$

Table 4.2a. Univariate-GARCH specifications selected for each asset (21 Stocks)

Australia	TARCH
Austria	GARCH
Belgium	GJR-GARCH
Canada	TARCH
Denmark	TARCH
France	GJR-GARCH
Germany	TARCH
Hong Kong	EGARCH
Ireland	EGARCH
Italy	GARCH
Japan	EGARCH
Mexico	GJR-GARCH
Netherlands	EGARCH
New Zealand	GARCH
Norway	AGARCH
Singapore	TARCH
Spain	EGARCH
Sweden	EGARCH
Switzerland	TARCH
United States	TARCH
United Kingdom	TARCH

Table 4.2b. Univariate GARCH specifications selected for each asset (13 Bonds)

Austria	GARCH
Belgium	GARCH
Canada	TARCH
Denmark	AGARCH
France	GARCH
Germany	GARCH
Ireland	GARCH
Japan	NGARCH
Netherlands	GARCH
Sweden	GARCH
Switzerland	EGARCH
United Kingdom	GARCH
United States	GJR-GARCH

Table 4.3a. P-values for the Value-at-Risk backtesting results for the unconditional tests.

Var-Cov	Quantile Distribution	Q-0.95	Q-0.99	Q-0.995
AGDCC_1	Normal	0.7469	0.4444	0.1836
	Laplace	**0.7909**	0.4949	0.6943
	G.P._50	0.0975	0.2646	0.4405
	G.P._70	0.7410	0.4375	0.5545
	G.P._100	0.0431	0.4759	0.4374
	G.P._150	0.0000	0.1806	0.3943
AGDCC_2	Normal	0.2868	**0.8629**	0.3067
	Laplace	0.3692	0.2207	0.3849
	G.P._50	0.2601	0.3856	0.4178
	G.P._70	0.6732	0.6394	**0.8321**
	G.P._100	0.0258	0.4051	0.3236
	G.P._150	0.0000	0.1090	0.3469
AGDCC_3	Normal	0.2376	0.5007	0.1239
	Laplace	0.2940	0.3559	0.3849
	G.P._50	0.2179	0.2598	0.3234
	G.P._70	0.7331	0.5394	0.5224
	G.P._100	0.0659	0.3707	0.3886
	G.P._150	0.0000	0.3000	0.5231

Notes to Table 4.3a. (**1**) The p-values were calculated using three different estimators for the conditional variance-covariance matrix: AGDCC_1 corresponds to the Asymmetric-Generalised-DCC model where the first and second stages are estimated assuming a multivariate normal distribution for the standardised residuals; AGDCC_2 corresponds to the case where the first stage is estimated assuming a multivariate normal distribution and the second stage assuming the asymmetric multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003); AGDCC_3 corresponds to the case where the first and second stages are estimated assuming a the AML distribution. (**2**) For the estimation of the quantile three univariate distributions were used: Normal, Laplace, and Genesalised Pareto (G.P.). (**3**) The parameters in the G.P. distribution were estimated under four different threshold levels: 50, 70, 100, and 150 observations. (**4**) The results are reported for three quantile levels: 95%, 99%, and 99.5%.

Table 4.3b. P-values for the Value-at-Risk backtesting results for the independent tests.

Var-Cov	Quantile Distribution	Q-0.95	Q-0.99	Q-0.995
AGDCC_1	Normal	0.3341	0.1007	0.2429
	Laplace	0.3034	0.4951	0.8716
	G.P._50	0.3635	0.8285	**0.9076**
	G.P._70	0.4864	0.7771	0.8795
	G.P._100	0.5449	0.6150	0.8285
	G.P._150	0.5187	0.5483	0.7519
AGDCC_2	Normal	0.4674	0.5611	0.7535
	Laplace	0.5075	0.8200	0.9075
	G.P._50	0.3457	0.6835	0.7441
	G.P._70	0.4604	0.6013	0.6854
	G.P._100	0.3638	0.5016	0.6836
	G.P._150	0.3651	0.4207	0.5764
AGDCC_3	Normal	0.1037	0.0052	0.0183
	Laplace	0.0894	0.2728	0.7327
	G.P._50	0.4956	**0.8289**	0.9195
	G.P._70	0.4964	0.7537	0.8915
	G.P._100	**0.6263**	0.5935	0.8328
	G.P._150	0.5012	0.5515	0.7705

See Notes to Table 4.3a.

Table 4.3c. P-values for the Value-at-Risk backtesting results for the conditional tests.

Var-Cov	Quantile Distribution	0.95	0.99	0.995
AGDCC_1	Normal	0.5084	0.1393	0.2144
	Laplace	0.5100	0.3935	**0.8759**
	G.P._50	0.1322	0.3982	0.5985
	G.P._70	0.6669	0.6194	0.7666
	G.P._100	0.0850	0.5823	0.6807
	G.P._150	0.0000	0.2420	0.5276
AGDCC_2	Normal	0.3910	**0.7494**	0.5467
	Laplace	0.4761	0.4234	0.6536
	G.P._50	0.2429	0.3648	0.3796
	G.P._70	0.5730	0.6421	0.7370
	G.P._100	0.0494	0.4681	0.4755
	G.P._150	0.0000	0.1871	0.4634
AGDCC_3	Normal	0.1044	0.0126	0.0228
	Laplace	0.1075	0.1477	0.4819
	G.P._50	0.2596	0.3907	0.5490
	G.P._70	**0.6673**	0.7410	0.7597
	G.P._100	0.1427	0.4906	0.6048
	G.P._150	0.0000	0.3971	0.6283

See Notes to Table 4.3a.

Table 4.4: P-values for the Expected Shortfall backtesting tests.

Var-Cov	Quantile Distribution	Q-0.95	Q-0.99	Q-0.995
AGDCC_1	Normal	0.3050	0.2870	0.2066
	Laplace	0.6220	0.5405	0.4686
	G.P._50	0.4050	0.4188	0.2924
	G.P._70	0.4349	0.4416	0.3092
	G.P._100	0.4228	0.4284	0.2973
	G.P._150	0.4399	0.4298	0.3040
AGDCC_2	Normal	0.4654	0.3683	0.2915
	Laplace	**0.8952**	**0.8066**	**0.6977**
	G.P._50	0.3663	0.3375	0.2602
	G.P._70	0.3944	0.3707	0.2892
	G.P._100	0.4045	0.3623	0.2778
	G.P._150	0.4248	0.3755	0.2814
AGDCC_3	Normal	0.3710	0.2586	0.1523
	Laplace	0.7057	0.5196	0.3941
	G.P._50	0.3764	0.4277	0.3163
	G.P._70	0.3982	0.4312	0.3160
	G.P._100	0.4051	0.4325	0.3160
	G.P._150	0.4254	0.4381	0.3181

Notes to Table 4.4. (**1**) The p-values were obtained using three alternative estimators for the conditional variance-covariance matrix. AGDCC_1 corresponds to the Asymmetric-Generalised-DCC model where the first and second stages are estimated assuming a multivariate normal distribution for the standardised residuals; AGDCC_2 corresponds to the case where the first stage is estimated assuming a multivariate normal distribution and the second stage assuming the asymmetric multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003); AGDCC_3 corresponds to the case where the first and second stages are estimated assuming a the AML distribution. (**2**) For the estimation of the quantile three univariate distributions are being used: Normal, Laplace, and Genesalised Pareto (G.P.). (**3**) The parameters in the G.P. distribution were estimated under four different threshold levels: 50, 70, 100, and 150 observations. (**4**) The results are reported for three quantile levels: 95%, 99%, and 99.5%.

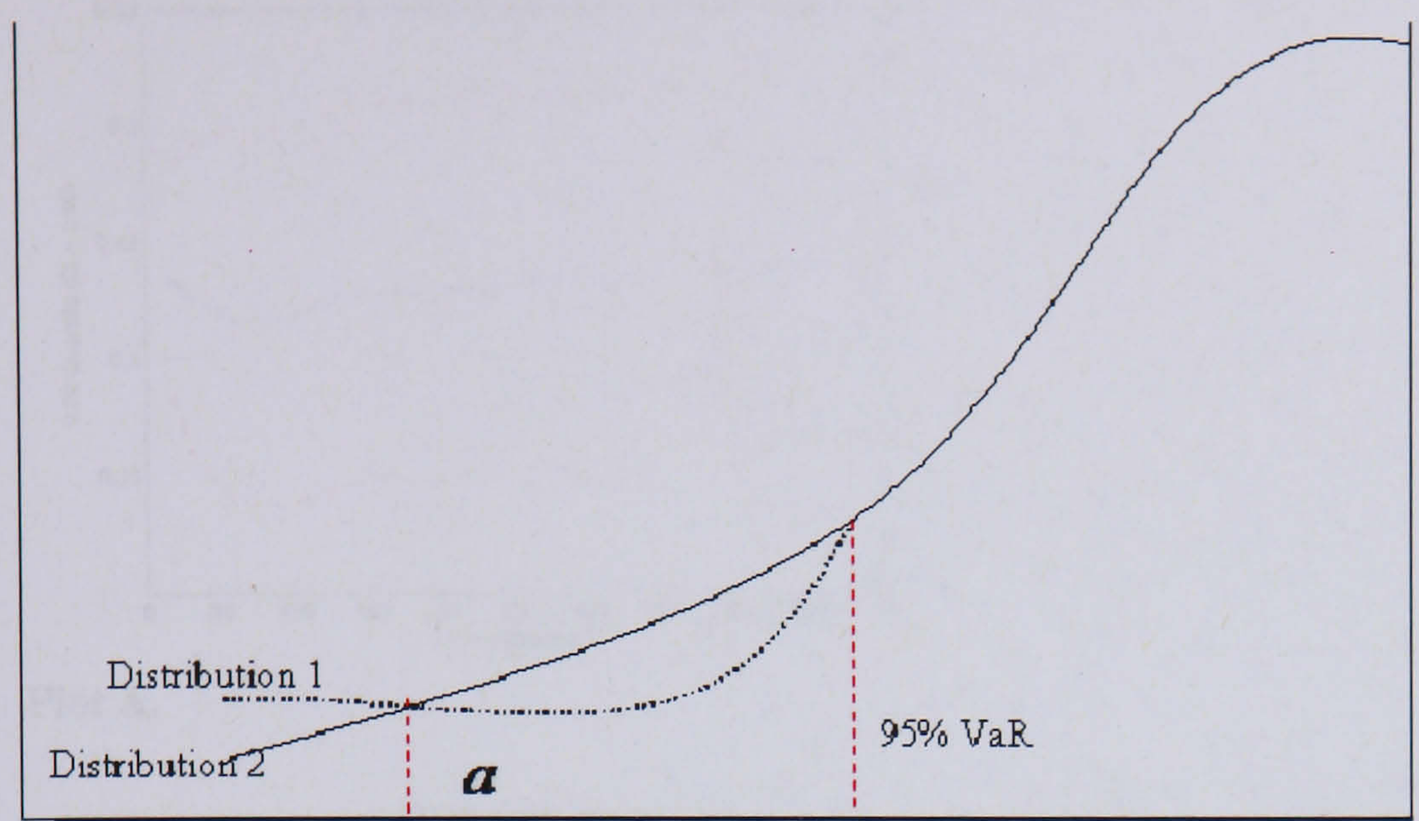
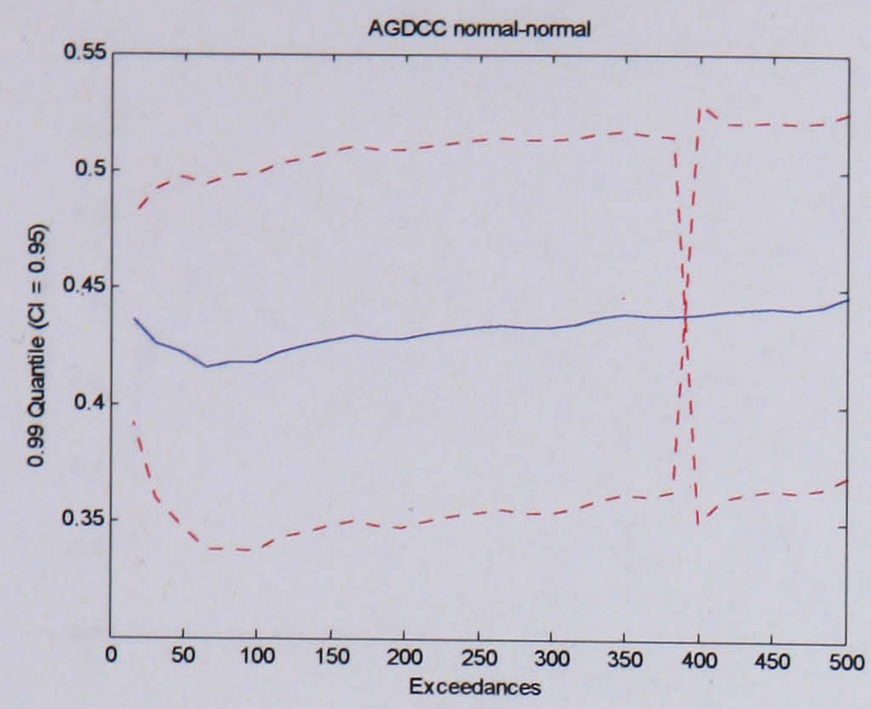
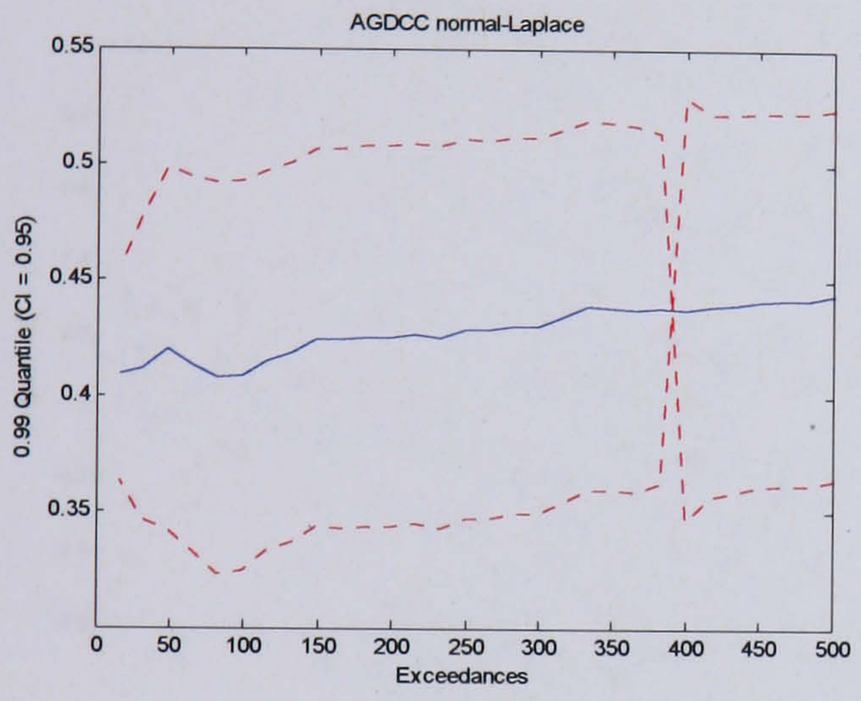


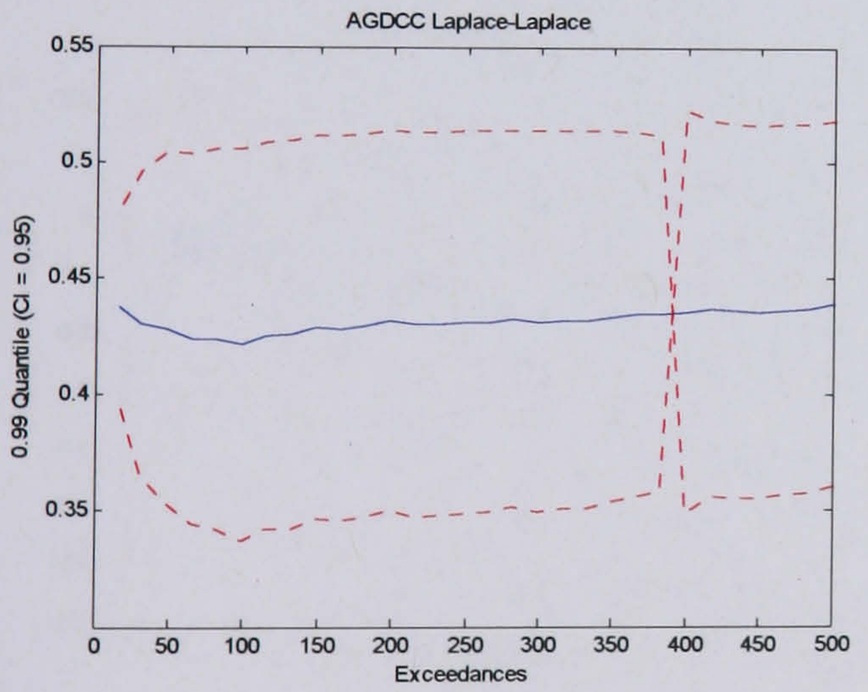
Figure 4.1. Value-at-Risk from two different probability distributions. The confidence level is 0.5%



Plot A.

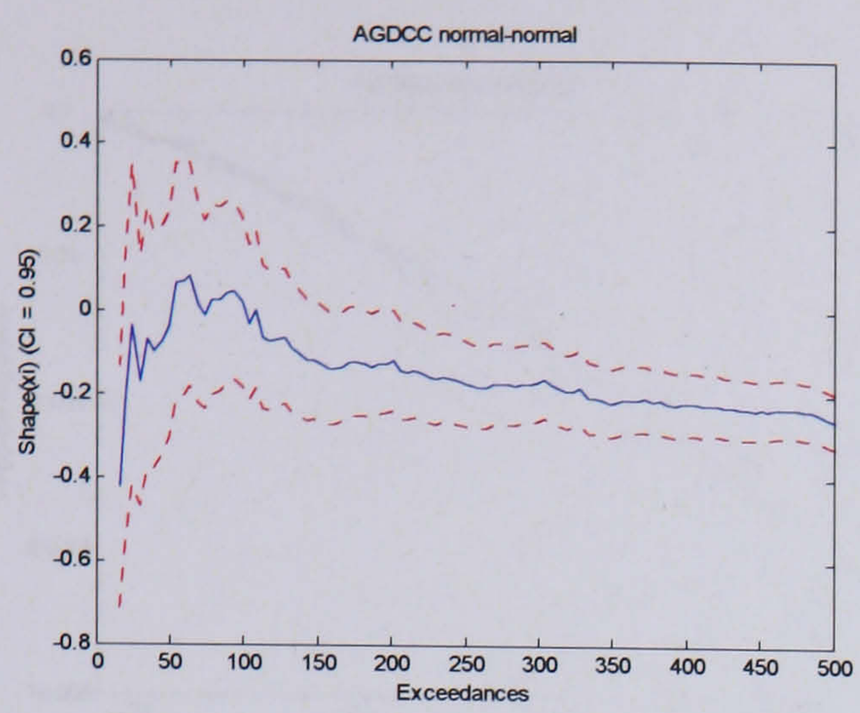


Plot B.

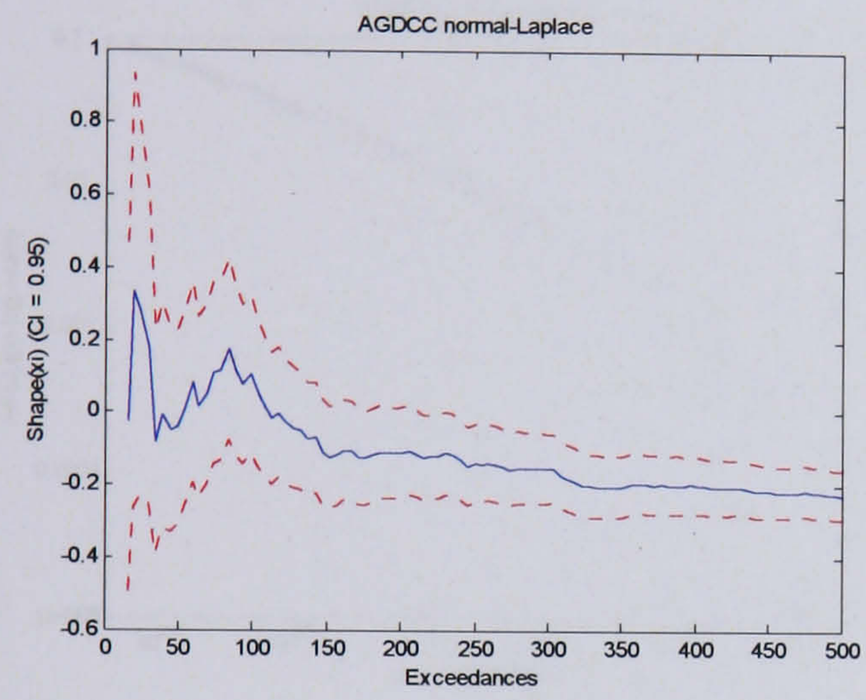


Plot C.

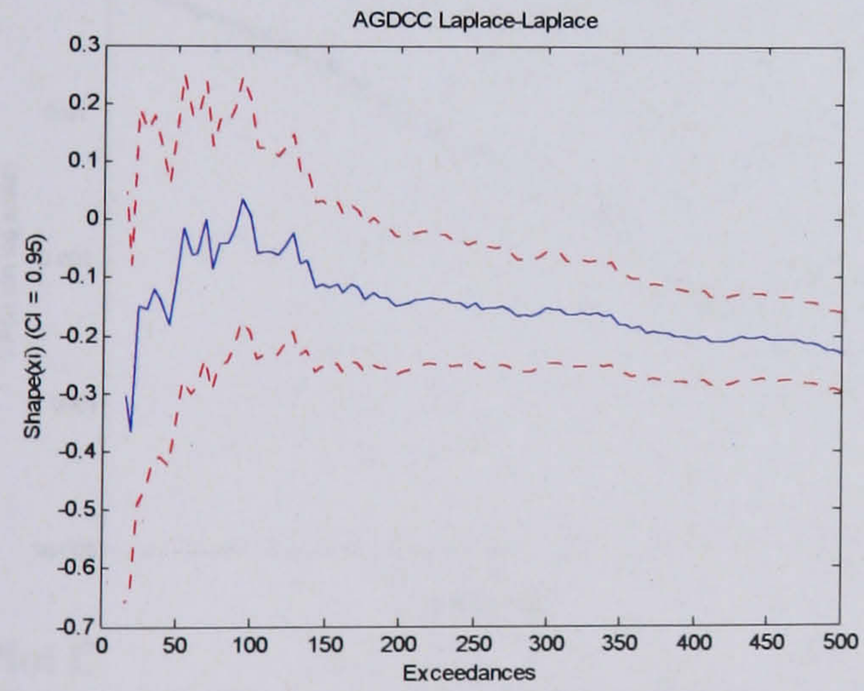
Figure 4.2. Loss estimates of quantile 99 as a function of exceedences over the threshold. Plot A corresponds to the case where the data was filtered by an Asymmetric-Generalised-DCC model where the first and second stages are estimated assuming a multivariate normal. Plot B corresponds to the case where the first stage is estimated assuming a multivariate normal distribution and the second stage assuming the asymmetric multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003). Plot C corresponds to the case where the first and second stages are estimated assuming a the AML distribution. Dotted lines are upper and lower 95 percent confidence intervals.



Plot A.

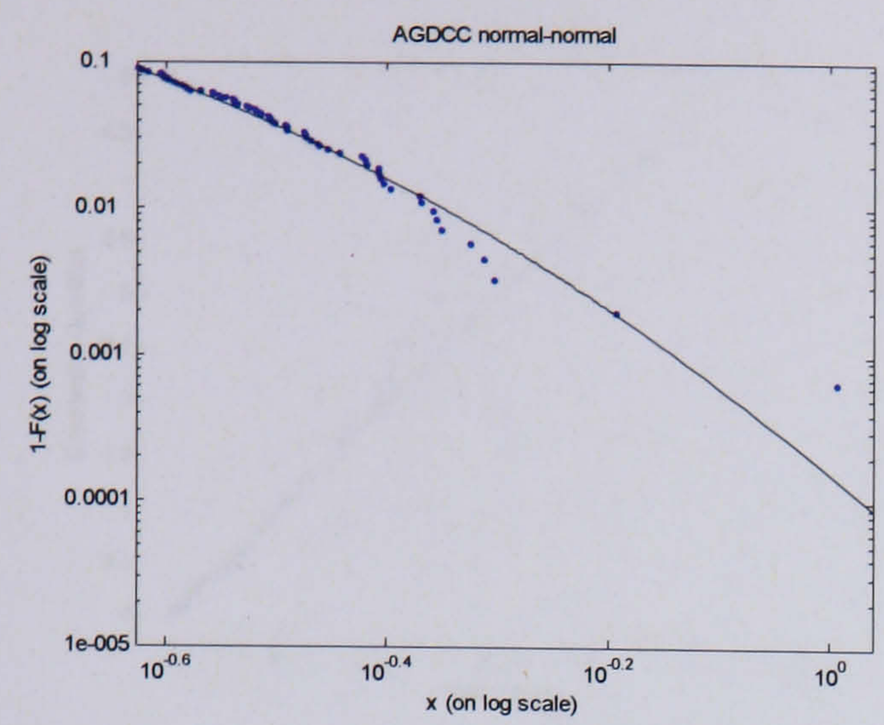


Plot B.

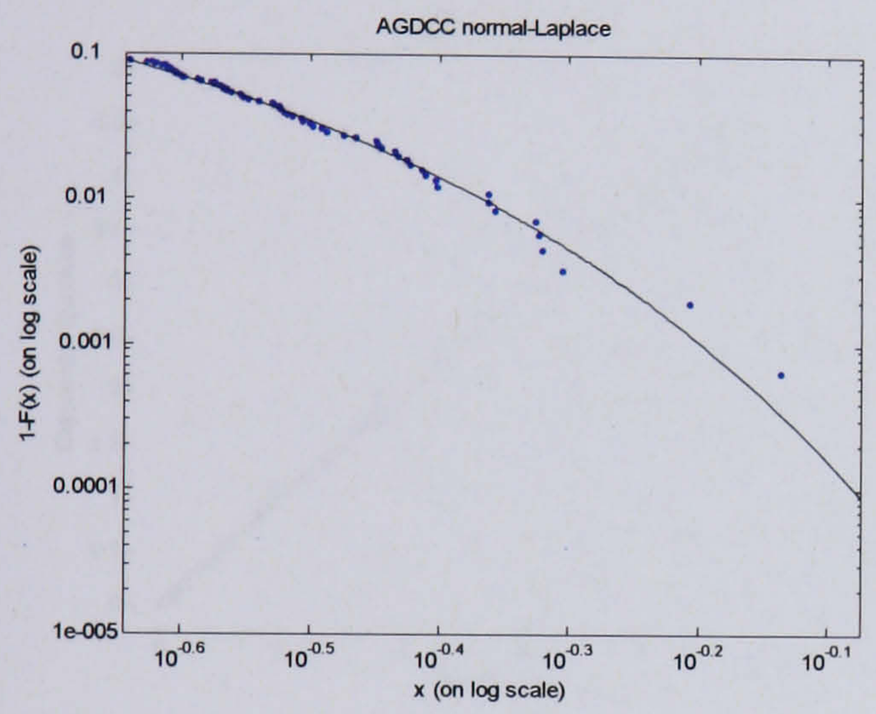


Plot C.

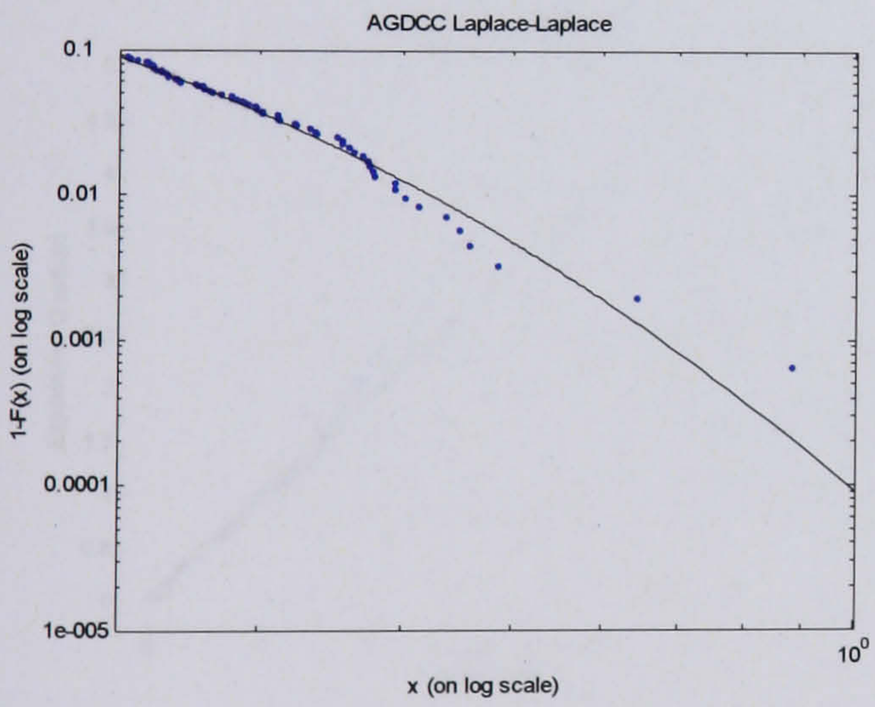
Figure 4.3. Estimates of shape parameter ξ of the Generalised Pareto distribution at different number of exceedences. Plot A corresponds to the case where the data was filtered by an Asymmetric-Generalised-DCC model where the first and second stages are estimated assuming a multivariate normal. Plot B corresponds to the case where the first stage is estimated assuming a multivariate normal distribution and the second stage assuming the asymmetric multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003). Plot C corresponds to the case where the first and second stages are estimated assuming a the AML distribution. Estimates were obtained by 100 different models. Dotted lines are upper and lower 95 percent confidence intervals.



Plot A

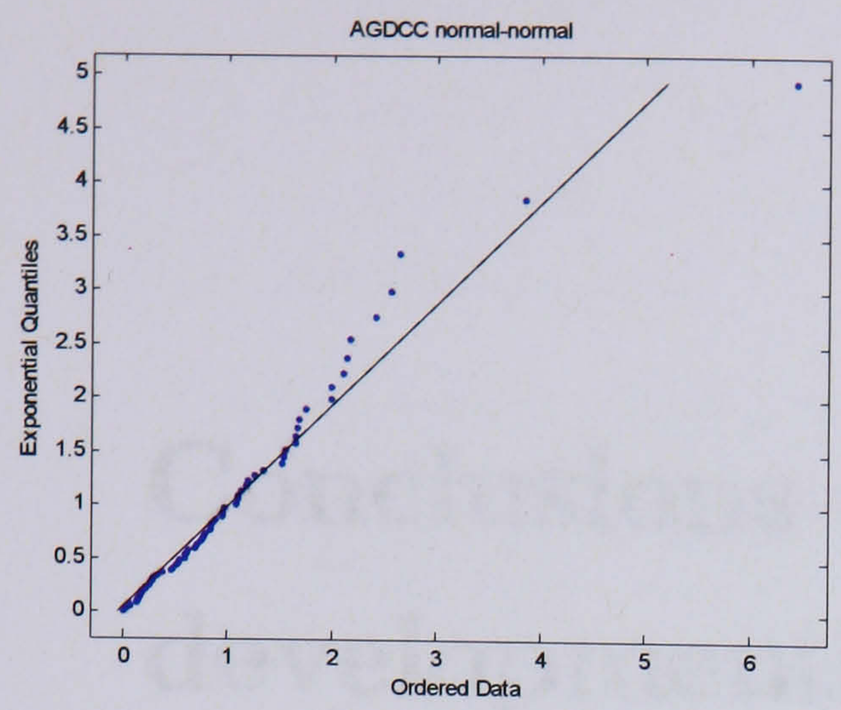


Plot B

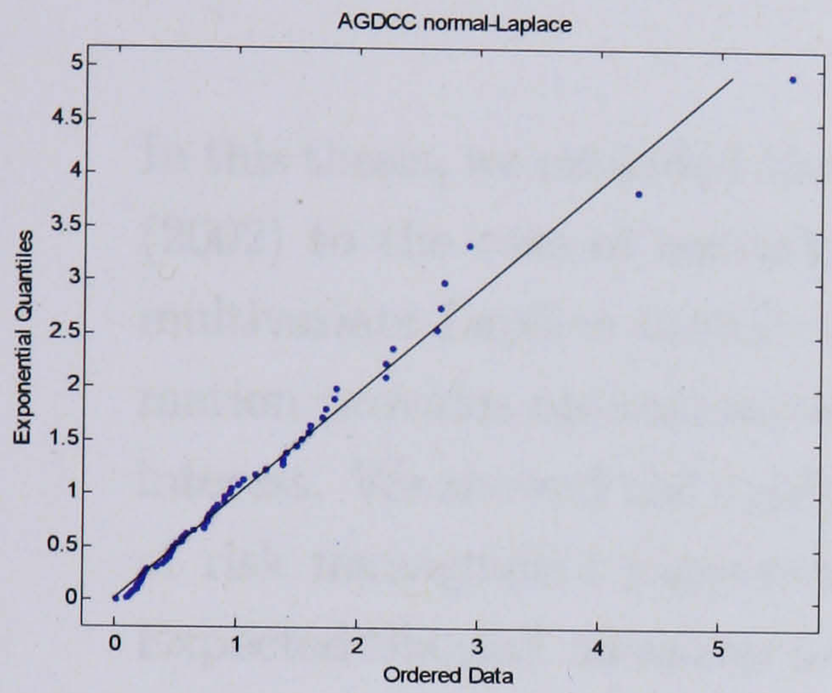


Plot C

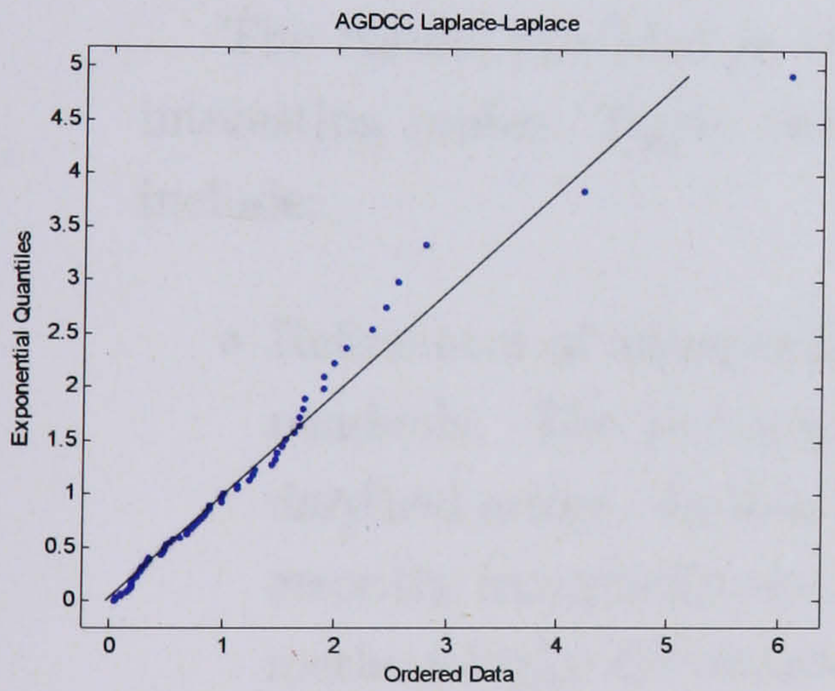
Figure 4.4. Tail estimate for the filtered data over the 70 exceedences threshold in Generalised Pareto distribution. The estimated tail is plotted as a solid line while the actual data in circles. Plot A corresponds to the case where the data was filtered by an Asymmetric-Generalised-DCC model where the first and second stages are estimated assuming a multivariate normal. Plot B corresponds to the case where the first stage is estimated assuming a multivariate normal distribution and the second stage assuming the asymmetric multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003). Plot C corresponds to the case where the first and second stages are estimated assuming a the AML distribution. The left axis indicates the tail probabilities. Both axes are on logarithmic scale.



Plot A



Plot B



Plot C

Figure 4.5. QQ-plot of residuals from the GPD fit to the filtered data over the selected threshold. Plot A corresponds to the case where the data was filtered by an Asymmetric-Generalised-DCC model where the first and second stages are estimated assuming a multivariate normal. Plot B corresponds to the case where the first stage is estimated assuming a multivariate normal distribution and the second stage assuming the asymmetric multivariate Laplace (AML) distribution of Kotz, Kozubowsky and Podgorski (2003). Plot C corresponds to the case where the first and second stages are estimated assuming a the AML distribution

Conclusions and further developments

In this thesis, we extended the dynamic conditional correlation model of Engle (2002) to the case of conditional returns supposed to follow an asymmetric multivariate Laplace distribution. We proved that maximum likelihood estimation provides optimal asymptotic properties of estimates of parameters of interest. We showed the applicability of our approach in a comprehensive set of risk management implementations where we compute Value-at-Risk and Expected-Shorfall measures for portfolios composed by a large number of assets

The results provided in this dissertation can be extended along various interesting routes. Topics that we think may be worth investigating further include:

- Refinement of mis-specification tests for the distribution of standardised residuals. The Kolmogorov-Smirnov test is based on estimated standardised errors. To take into account the effect of parameter estimation error the marginalisation approach of Bai (2003) could be applied. Other methodologies for distributional testing, including the cross-correlation approach of Hong and Li (2002) and the simulation based approach of Corradi and Swanson (2005), can be analysed to find if they can be implemented in a multivariate-GARCH model framework.

- Other applications of the model to finance can be implemented such as testing CAPM under AML distribution and conditional CAPM under a DCC specification.

- Derivation of the primitive conditions for the verification of the asymptotic normality of parameter estimates in the AML-MLE method for the AGDCC model

- Comparison of forecasting capabilities with MGARCH models recently proposed such as Audrino and Trojani (2004), Ledoit *et al* (2003), Palandri (2005), Lanne and Saikkonen (2006)

References

Alexander, C. O. and A. Chibumba (1996), "Multivariate Orthogonal Factor GARCH", *Emerging Markets Investor*, 2-6.

Allen, D., M. McAleer and B. Veiga (2004), "Modelling and Forecasting Dynamic VaR Thresholds for Risk Management and Regulation", mimeo.

Ang, A. and J. Chen (2002), "Asymmetric Correlations of Equity Portfolios", *Journal of Financial Economics*, 63, 443-494

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath (1999). "Coherent Measures of Risk", *Mathematical Finance*, 9, 203-228.

Audrino, F. and G. Barone-Adesi (2003), "Semiparametric Multivariate GARCH Models for Volatility Asymmetries and Dynamic Correlations", mimeo, University of Southern Switzerland.

Audrino, F. and P. Buhlmann (2003), "Volatility Estimation with Functional Gradient Descent for Very High-Dimensional Financial Time Seires", *Journal of Computational Finance*, 6, 65-89.

Audrino, F. and F. Trojani (2004), " A General Multivariate Threshold GARCH Model with Dynamic Conditional Correlations", mimeo, University of Lugano and University of St. Gallen - Department of Economics.

Audrino, F. and F. Trojani (2006), "Estimating and Predicting Multivariate Volatility Thresholds in Global Stock Markets", *Journal of Applied Econometrics*, 21: 345–369.

Bai, J. (2003), "Testing Parametric Conditional Distributions of Dynamic Models", *Review of Economics and Statistics,* 85, 531-549.

Bai, J. and S. Ng (2005), "Tests for Skewness, Kurtosis, and Normality for Time Series", *Journal of Business and Economic Statistics*, 23, 49-60.

Balkema, A. and L. de Haan (1974), "Residual Life Time at Great Age".

Annals of Probability, 2, 792-804.

Ball, C. A. and W. N. Torous (2002), "Stochastic Correlation Across International Stock Markets", *Journal of Empirical Finance*, 7, 373-388

Barndorff-Nielsen, O. (1977), "Exponentially Decreasing Distributions for the Logarithm of Particle Size", *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 353, 401-419.

Barndorff-Nielsen, O. and N. Shephard (2004), "Econometric Analysis of Realised Covariation: High Frequency Based Covariance, Regression and Correlation in Financial Economics", *Econometrica,* 72, 885-925.

Baur, D. (2002), "A Flexible Dynamic Correlation Model", mimeo.

Bauwens, L. and S. Laurent (2004), "A New Class of Multivariate Skew Densities, with Application to GARCH Models", mimeo.

Bauwens, L., S. Laurent and J. Rombouts (2006), "Multivariate GARCH Models: a Survey", *Journal of Applied Econometrics*, 21, 79-109.

Bera, A. K., P. Garcia and J. Roh (1997), "Estimation of Time-Varying Hedging Ratios for Corns and Soybeans: BGARCH and Random Coefficient Approaches", *Sankhya*, 59, 346-368.

Bera, A. K. and S. Kim (2002), "Testing Constancy of Correlation and Other Specifications of the BGARCH Model with an Application to International Equity Returns", *Journal of Empirical Finance*, 9, 171-195.

Bekaert, G., C. R. Harvey and A. Ng (2003), "Market Integration and Contagion", *The Journal of Business*, 78, 39–69

Berkes, I., L. Horvath and P. Kokoszka (2003), "GARCH Processes: Structure and Estimation", *Bernoulli*, 9, 201-227.

Bertero, E. and C. Mayer (1990), "Structure and Performance: Global Interdependence of Stock Markets Around the Crash of October 1987", *European Economic Review*, 34, 1155-1180.

Billio, M., M. Caporin and M. Gobbo (2004), "Flexible Dynamic Conditional Correlation Multivariate GARCH for Asset Allocation", GRETA Working Paper, n.04.03.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity", *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. (1990), "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate GARCH", *Review of Economics and Statistics*,

72, 498-505.

Bollerslev, T. (2001), "Financial Econometrics: Past Developments and Future Challenges", *Journal of Econometrics*, 100, 41-51.

Bollerslev, T., R. Engle and J. M. Wooldridge (1988), "A Capital Asset Pricing Model with Time Varying Covariances", *Journal of Political Economy*, 96, 116-131.

Bollerslev, T. and J. M. Wooldridge (1992), "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances", *Econometric Reviews,* 11, 143-172.

Bougerol, P. and N. Picard (1992), "Stationarity of GARCH Processes and Some Nonnegative Time Series", *Journal of Econometrics*, 52, 115-127.

Boussama F. (1998), "Ergodicity, Mixing and Estimation in GARCH Models", Ph.D. Dissertation, University of Paris 7.

Bowman, F. C. (1958), *Introduction to Bessel Functions*, Dover Publications, New York.

Boyer, B. H., M. S. Gibson and M. Loretan (1997), "Pitfalls in Tests for Changes in Correlations", Board of Governors of the Federal Reserve System International Finance Discussion Papers Number 597.

Butler, K. C. and D. C. Joaquin (2002), "Are the Gains from International Portfolio Diversification Exaggerated? The Influence of Downside Risk in Bear Markets", *Journal of International Money and Finance*, 21, 981-1011.

Campbell, R. A. J., K. C. Koedijc and P. Kofman (2000), "Increased Correlation in Bear Markets", *Financial Analysts Journal*, 58, 87-94

Cappiello, L., R. Engle and K. Sheppard (2004), "Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns", ECB Working Paper No. 204.

Chan-Lau, J. A., D. J. Mathieson and J. Y. Yao (2004), "Extreme Contagion in Equity Markets", *IMF Staff Papers, International Monetary Fund*, 51, 386-408

Christoffersen, P. and D. Pelletier (2004), "Backtesting Value-at-Risk: A Duration-Based Approach", *Journal of Financial Econometrics*, 2, 84-108.

Comte, F., and Lieberman, O. (2003), "Asymptotic Theory for Multivariate GARCH Processes", *Journal of Multivariate Analysis*, 84, 61,84.

Corradi, V. and N. R. Swanson (2005), "Bootstrap Conditional Distribu-

tion Test in the Presence of Dynamic Misspecification", in *Journal of Econometrics* (in press).

Diebold, F., T. Schermann and J.D. Stroughair (1999), "Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management", *Advances in Computational Finance*, Kluwer Academic Publishing.

Ding, Z., and R. Engle (2001), "Large Scale Conditional Covariance Matrix Modelling, Estimation, and Testing", mimeo.

Doornik, J. A. and H. Hansen (1994), "An Omnibus Test for Univariate and Multivariate Normality", mimeo.

Duffie, D., and J. Pan (1997), "An overview of Value at Risk", *Journal of Derivatives*, 4, 7-49.

Dufour, J-M. (2004), "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics", mimeo, Universite de Montreal.

Dufour, J-M and D. Pelletier(2004), "Linear Estimation of Weak VARMA Models With a Macroeconomic Application", mimeo.

Dumas, B., R. Harvey and P. Ruiz. (2003), "Are Correlation of Stock Returns Justified by Subsequent Changes in National Outputs?", *Journal of International Money and Finance*, 22, 777-811.

Efron, B., and R, Tibshirani (1993), *An introduction to the Bootstrap*, Chapman and Hall, New York.

Embrechts, P., C. Kluppelberg and T. MiKosch (1997), *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Embrechts, P., A. McNeil and D. Straumann (1999), "Correlation: Pitfalls and Alternatives ", *Risk*, 12, 69-71

Engle, R. (1982), "ARCH with Estimates of the Variance of United Kingdom Inflation", *Econometrica*, 50, 987-1007

Engle, R. (2001), "Financial Econometrics. A New Discipline with New Methods", *Journal of Econometrics*, 100, 53-56.

Engle, R. (2002), "Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models", *Journal of Business and Economic Statistics*, 20, 339- 350.

Engle, R., and R. Colacito (2006), "Testing and Valuing Dynamic Correlations for Asset Allocation", *Journal of Business and Economic Statistics*,

24, 238-253

Engle, R. and K. Kroner (1995), "Multivariate Simultaneous Generalised ARCH", *Econometric Theory*, 11, 122-150.

Engle, R. and V. K. Ng (1993), "Measuring and Testing the Impact of News on Volatility", *Journal of Finance*, 48, 1749-1778.

Engle, R., V. K. Ng and M. Rothschild (1990), "Asset Pricing with a Factor ARCH Covariance Structure: Empirical Estimates for Treasury Bills", *Journal of Econometrics*, 45, 213- 238.

Engle, R. and K. Sheppard (2001), "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH", NBER Working Paper No.8554.

Fiorentini, G., G. Calzolari and E. Sentana (2003), "Maximum Likelihood Estimation and Inference in Multivariate Conditionally Heteroskedastic Dynamic Regression Models with Student t Innovations", *Journal of Business and Economic Statistics,* 21, 532-546.

Forbes, K. J. and R. Rigobon (2002), "No Contagion, Only Interdependence: Measuring Stock Market Comovements", *Journal of Finance*, 57, 2223-2261

Giot, P. and S. Laurent (2003), "Value-at-Risk for long and Short Trading positions," *Journal of Applied Econometrics*, 18, 641-663

Goetzmann, W. N., L. Li and K. G. Rouwenhorst (2001), "Long-Term Global Market Correlations", NBER Working Paper No. W8612.

Granger, C. W. J., T. Teräsvirta and A. Patton (2006), "Common Factors in Conditional Distributions for Bivariate Time Series", *Journal of Econometrics*, 132, 43-57.

Hafner, C. M. and H. P. Franses (2004), "A Generalized DCC Model for Many Asset Returns", mimeo Econometric Institute, Erasmus University Rotterdam.

Hafner, C. M. and H. Herwartz (1998), "Time-Varying Market Price of Risk in the CAPM. Approaches, Empirical Evidence and Implications", *Finance*, 19, 93-112.

Hanson, F. B. and Z. Zhu (2004), "Comparison of Market Parameters for Jump-Diffusion Distributions Using Multinomial Maximum Likelihood Estimation", mimeo.

Hartmann, P., S. Straetmans and C. De Vries. (2001), "Asset Market Linkages in Crisis Periods," Working Paper Series 071, European Central Bank.

Heyde, C. C. and S. G. Kou (2004), "On the Controversy Over Tailweight of Distributions", *Operations Research Letters*, 32, 399-408.

Hong, Y. and H. Li (2002), "Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates", mimeo, Department of Economics, Cornell University.

Hosking, J. R. and J. R. Wallis (1987) "Parameter and Quantile Estimation for the Generalized Pareto Distribution", *Technometrics*, 29, 339-349.

Jacquier, E., and A. Marcus (2001), "Asset Allocation Models and Market Volatility ", *Financial Analysts Journal*, March-April.

Jeantheau, T. (1998), "Strong Consistency of Estimators for Multivariate ARCH models", *Econometrics Theory*, 14, 70-86.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

Johnson, N. L. and S. Kotz (1972), *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York.

Kaplanis, E. C. (1988), "Stability and Forecasting of the Co-Movement Measures of International Stock Market Return", *Journal of International Money and Finance*, 8, 63-75

Kawakatsu, H. (2005), "Matrix Exponential GARCH", in *Journal of Econometrics* (in press).

Karolyi, G. A. and R. M. Stulz (1995), "Why Do Markets Move Together? An Investigation of U.S.-Japan Stock Return Comovements using ADRS", Research in Financial Economics No 9603, Ohio State University.

Kearney, C. and A. Patton (2000), "Multivariate GARCH Modelling of Exchange Rate Volatility Transmission in the European Monetary System", *Financial Review*, 41, 29-48.

Khindanova, I., S. Rachev and E. Schwartz (2001), "Stable Modeling of Value at Risk", *Mathematical and Computer Modeling,* 34, 1223-1259.

King, M., and S. Wadhwani (1992), "Transmission of Volatility Between Stock Markets", *Review of Financial Studies*, 3, 5-33.

Kluppelberg, C. (2002), "Risk Management with Extreme Calue Theory", working paper, Munich Unverstiy of Technology.

Koch, P. D. and T. W. Koch (1991), "Evolution in Dynamic Linkages Across National Stock Indexes", *Journal of International Money and Finance*, 10, 231-251.

Komunjer, I. (2005), "Quasi-Maximum Likelihood Estimation for Conditional Quantiles", *Journal of Econometrics,* 128, 137-164

Kotz, S., T. J. Kozubowski and K. Podgorski (2003), "An Asymmetric Multivariate Laplace Distribution", mimeo.

Kou, S. G. and H. Wang (2001), "Option Pricing Under a Double Exponential Jump Diffusion Model", mimeo.

Kristensen, D., and A. Rahbek (2005), "Asymptotics of the QMLE for General ARCH(q) Models", mimeo.

Kroner, K. F. and V. K. Ng (1998), "Modelling Asymmetric Comovements of Asset Returns", *Review of Financial Studies*, 11, 817-844.

Kozubowski, T. J., and K. Podgorski (2001), "Asymmetric Laplace Laws and Modeling Financial Data", *Mathematical and Computer Modelling*, 34, 1003-1021.

Lanne, M., and P. Saikkonen (2006), "A Multivariate Generalized Orthogonal Factor GARCH Model", forthcoming in the *Journal of Business and Economic Statistics.*

Ledoit, O., P. Santa-Clara, P. and M. Wolf (2003), " Flexible Multivariate GARCH Modeling with an Application to International Stock Markets ", *Review of Economics and Statistics*, 85, 735-747.

Lee, S-W, and B. E. Hansen (1994), "Asymptotic Theory for the GARCH (1,1) Quasi-Maximum Likelihood Estimator", *Econometric Theory*, 10, 29-52.

Li, L. (2002), "Macroeconomic Factors and the Correlation of Stock and Bond Returns", mimeo.

Linden, M. (2001), "A Model for Stock Return Distribution", *International Journal of Finance and Economics*, 6, 159-169.

Longin, F. N. and B. Solnik (1995), "Is the Correlation in International Equity Returns Constant: 1960—1990?", *Journal of International Money and Finance*, 14, 3-26.

Loretan, M., and W. B. English (2000), "Evaluating Correlation Breakdowns During Periods of Market Volatility", Board of Governors of the Federal

Reserve System International Finance Working Paper No. 658

Lumsdaine, R. L. (1996), "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models", *Econometrica*, 64, 575-596.

Madan, D. B. and E. Seneta (1990), "The Variance Gamma (V.G.) Model for Share Market Returns", *Journal of Business*, 63, 511-524.

Madan, D. B., P. P. Carr and E. C. Chang. (1998), "The Variance Gamma Process and Option Pricing", *European Finance Review*, 2, 79-105.

McNeil, A. and R. Frey (2000), "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach", *Journal of Empirical Finance,* 7, 271-300

Mencia, F. J., and E. Sentana (2005), "Estimation and Testing of Dynamic models with Generalised Hyperbolic Innovations", CEMFI Working Paper No. 0411

Newey, W., and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing,", *Handbook of Econometrics*, (eds.) R. Engle and D. McFadden, 4, 2113-2245.

Orme, C. (1990), "The Small Sample Performance of the Information Matrix Test", *Journal of Econometrics*, 46, 309-331.

Palandri, A. (2005), "Sequential Conditional Correlations: Inference and Evaluation.", paper presented at the Conference on "Changing Structures in International and Financial Markets and the Effects on Financial Decision Making", Venice, Italy, June 2-3.

Pelletier, D. (2006), "Regime Switching for Dynamic Correlations", *Journal of Econometrics*, 131, 445-473.

Pickands, J. (1975), "Statistical Inference Using Extreme Order Statistics", *The Annals of Statistics*, 3, 119-131.

Pfanzgal, J. (1969), "On the Measurability and Consistency of Minimum Contrast Estimates", *Metrika*, 14, 249-272.

Ratner, M. (1992), "Portfolio Diversification and the Inter-Temporal Stability of International Indices", *Global Finance Journal*, 3, 67-78.

Relton, F. E. (1965), *Applied Bessel Functions*, Dover Publications, New York.

Ronn, E. I. (1998), "The Impact of Large Changes in Asset Prices on Intra-

Market Correlations in the Stock and Bond Markets", mimeo, University of Texas at Austin.

Rigobon, R. (2003), "On the Measurement of the International Propagation of Shocks: Is it Stable?", *Journal of International Economics*, 61, 261-283.

Schweizer, B. and A. Sklar (1983), *Probabilistic Metric Spaces*, North-Holland/Elsevier, New York.

Sepp, A. (2004), "Analytical Valuation of Lookback Options in a Double Exponential Jump-Diffusion Model", mimeo.

Starica, C. (1999), "Multivariate Extremes for Models with Constant Conditional Correlations", *Journal of Empirical Finance*, 6, 515-553.

Tse, I. K. (2000), "A Tests for Constant Correlations in a Multivariate GARCH Model", *Journal of Econometrics*, 98, 107-127.

Van der Weide, R. (2002), "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model", *Journal of Applied Econometrics*, 17, 549,564.

Verbeek, M. and J. Rombouts (2005) "Evaluating Portfolio Value-at-Risk using Semi-Parametric GARCH Models," Computing in Economics and Finance 2005 40, Society for Computational Economics.

Von Furstenberg, G. M. and B. N. Jeon (1989), "International Stock Prices Movements: Links and Messages", *Brookings Papers on Economic Activity*, 1, 125-179.

Weiss, A. A. (1986), "Asymptotic Theory for ARCH Models: Estimation and Testing", *Econometric Theory*, 2, 107-131.