Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

Milan Maďar

Modelování ve finanční analýze

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. RNDr. Jan Hurt CSc.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

Praha 2011

Poďakovanie
Na tomto mieste by som rád poďakoval svojmu vedúcemu doc. RNDr. Janovi Hurtovi CSc., za ochotu pri vedení práce, svojmu školitelovi Prof.dr. A.W. van der Vaartovi z VU University Amsterdam za veľmi zajímavú tému a nápady pre ďalší rozvoj práce.

Ďalej by som chcel poďakovať Jozefovi Maďarovi za morálnu podporu behom štúdií.

Prehlasujem, že som túto diplomovú prácu vypracoval samostatne a výhradne s použitím citovaných prameňov, literatúry a ďalších odborných zdrojov.

Beriem na vedomie, že sa na moju prácu vzťahujú práva a povinnosti vyplývajúce zo zákona č. 121/2000 Sb., autorského zákona v platnom znení, najmä skutočnosť, že Univerzita Karlova v Praze má právo na uzavrení licenčnej zmluvy o použití tejto práce ako školného diela podľa §60 odst. 1 autorského zákona.

V Praze dne 05.12.2011 podpis autora

Contents

Názov práce: Modelování ve finanční analýze
Autor: Milan Maďar
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedúci diplomovej práce: doc. RNDr. Jan Hurt CSc.
Abstrakt: V predloženej práci študujeme regionálne a globálne väzby, ako dôkaz integrácie akciových trhov vo Frankfurte, Amsterdame, Prahe a USA rovnako dynamiku prenosu volatility medzi súvisiacimi devízovými kurzami použitím viacrozmerných GARCH modelov. U každého modelu je uvedený popis základných definícií, vlastností a postup odhadu parametrov. Praktická časť práce ilustruje použitie jednotlivých modelov na reálnych dátach. Práca sleduje dva rôzne ciele, jednak charakteristiku a popis existencie regionálnych a globálnych väzieb na rôznych akciových trhoch, ale aj vzájomné porovnanie jednotlivých viacrozmerných GARCH modelov na vzorke dát. Výsledkom je, že odhady podmienených korelácii závislých na čase naznačujú obmedzenú integráciu medzi trhmi z čoho vyplýva, že investori môžu využiť diverzifikáciu portfólia medzi rôzne akciové trhy a tým znížiť rizikovosť investície obzvlášť v dobe krízy.
Kľúčové slová: Markets linkages, multivariate GARCH, VECH, BEKK, O-GARCH, GO-GARCH, CCC, DCC

Title: Modelling in Financial Analysis
Author: Milan Maďar
Department: Katedra pravděpodobnosti a matematické statistiky
Supervisor: doc. RNDr. Jan Hurt CSc.
Abstrakt: In this thesis we study the regional and global linkages as evidence of markets integration of the stock markets in Frankfurt, Amsterdam, Prague the U.S. and the dynamics of volatility transmission of related foreign exchange rates using multivariate GARCH approach. For each of the model classes, a theoretical review, basic properties and estimation procedure are provided. We illustrate approach by applying the models to daily market data. Our two main aims are discussing and report the existence of regional and global stock markets linkages and provide comparison of such multivariate GARCH models on the data sample. We find out that the estimated time-varying conditional correlations indicate limited integration among the markets which implies that investors can benefit from the risk reduction by investigating in the different stock markets especially during the crisis.
Keywords: Markets linkages, multivariate GARCH, VECH, BEKK, O-GARCH, GO-GARCH, CCC, DCC

Chapter 1

Introduction

In the face of globalization, it is important to document developments and linkages in global as well as in local markets. An accumulation of such information will provide a platform for determining the integration of global markets. Understanding of linkages and volatility transmission in stock and foreign exchange markets‘ returns and correlation of such returns will help investors and fund managers better manage their investment portfolios. They want to diversify their portfolios as much as possible on account of risk minimization however, the diversification benefits of investigating in the different markets depend on the extend of the linkages between the markets, or we can ask how strong are the markets integrated which requires a good understanding of the underlying foreign exchange volatility. Only when market returns are less than perfectly correlated, is risk reduction possible. Indeed with the current crisis of confidence in risk management and the requirements of regulators, there is a requirement for GARCH modelling to take explicitly into account multivariate issues. When we are in multivariate framework, we are always balancing between two difficulties. The number of parameters in the model increases quickly with the dimension of the model resulting estimation problems overleaf simple models may not be able to capture the relevant dynamics in the structure. A lot of multivariate GARCH models have been developed and we survey three basic approaches of constructing. Then there exists a broad literature on the research done on markets return, volatility and even integration in the stock markets all around the world. Most of them have focused on national and regional stock and foreign exchange markets only. The comprehensive analysis of more multivariate GARCH models is missing. Therefore, motivated by the impact of the recent crisis authors own contribution was primarily into the summarizing the particular component parts of multivariate GARCH models into the one complex work thereof, to show robustness of the GARCH methods and show how models are used in practice. Thesis contains two main aims, analyze the dynamics of volatility transmission in foreign exchange markets, examining the stock market linkages from a very representative global perspective and then comparing such basic types of multivariate GARCH models on the data.

The rest of the thesis is organized as follows. The next chapter gives some basic details about the used data and presents some of the stylized features of financial data, which need to be taken into account when writing down models. The Chapter 3 presents a theoretical survey of univariate GARCH models, while Chapter 4 collects theoretical survey of multivariate GARCH framework, containing the following models: VECH, BEKK, O-GARCH, GO-GARCH, CCC and DCC. For each class of the model, a theoretical review, basic properties and estimation procedure are provided. Chapter 5 presents the findings and analysis from applying three multivariate models namely BEKK, GO-GARCH and DCC on the data containing data description, estimation results and models comparison. The thesis is concluded in Chapter 6.

Chapter 2

Preliminary analysis

2.1 Data

In this thesis, we used data which can be divided into the two major groups. On one hand, our data consists of the daily closing spot prices for the Czech koruna and Euro versus the U.S. dollar from the Bloomberg research database. The daily series represents changes between business days with no adjustment for holidays. On the other hand, data used in the study consist of time series of daily stock market indices at the closing values of the markets in Prague (PX), Amsterdam (AEX), Frankfurt (DAX) and the U.S. (DJIA). The stock indices are based in the local currency terms and their changes are thus restricted to the movements in the stock process, avoiding any distortions included by the currency exchange rates devaluations of the countries.

The investigated currencies, U.S. dollar and Euro, constitute the largest foreign exchange markets in the world measured in terms of turnover, are highly liquid, and have low transaction costs. Trading also occurs on a 24 hour basis, with almost instantaneous transmission of news items to market participants using computerised technology and on-line broking services. Consequently these markets are as close to the efficient market ideal as is currently possible. It is clear that because of the market in Prague, we selected third currency Czech koruna. We assume that the reader is familiar with background of this three widely accepted currencies. We therefore focus more on stock market issues. The stock market in Prague represents the emerging markets in Central and Eastern Europe, Amsterdam represents the market situation in Western Europe, Frankfurt which is one of the biggest stock markets in Europe represents the market situation in Europe and finally we used data from U.S. given by Dow Jones index which generally reflects the financial situation in this part of the world. The stock markets in Frankfurt and the

U.S. are considered to serve well as leaders for the regional and global developed markets respectively and are expected to play an influential role in the markets in Central and Western Europe. The inclusion of the stock markets in Frankfurt and the U.S. therefore permits us to investigate the regional and global linkages between markets.

The indices used in this thesis are the widely accepted benchmark indices for the stock markets. Because in different countries holidays, no trading days fall on different dates, we have removed the data of those dates, when any series has a missing value due to no trading. Thus all data are collected for the same dates across the stock markets.

The DAX (Deutscher Aktien-Index (German stock index)) is a blue chip stock market index consisting of the 30 major German companies trading on the Frankfurt Stock Exchange. Prices are taken from the electronic Xetra trading system. According to Deutsche Boerse, the operator of Xetra, DAX measures the performance of the Prime Standard's 30 largest German companies in terms of order book volume and market capitalization. The base date for the DAX is 30 December, 1987 with a base value of 1,000. The Xetra system calculates the index every second since January 1, 2006.

The AEX index, derived from Amsterdam Exchange index, is a stock market index composed of Dutch companies that are traded on Euronext Amsterdam, formerly known as the Amsterdam Stock Exchange. Started on 3 January 1983 from a base level of 100 index points, the index is composed of a maximum 25 of the most actively traded securities on the exchange. The AEX is a market value-weighted index. The index comprises a basket of shares, the numbers of which are based on the constituent weights and index value at the time of readjustment. The value of the index is calculated by multiplying the price (in Euros) of each of the stocks by the number of shares that are trading in the basket, then summing the resulting numbers and dividing by 100.

The PX index (until March 2006 the PX 50) is an index of major stocks that trade on the Prague Stock Exchange. Selected as the starting exchange day (a benchmark date) for the Index PX 50 was 5 April 1994 and its opening value was fixed at 1,000 points. At this time the index included 50 companies traded on the Prague Stock Exchange, accordingly named PX 50. Frequency of calculation is every 15 seconds.

The DJIA index derived from Dow Jones Industrial Average, also referred to as the Industrial Average, the Dow Jones, the Dow 30, or simply the Dow, is one of several stock market indices created by Wall Street Journal editor and Dow Jones & Company co-founder Charles Dow. It is an index that shows how 30 large, publicly owned companies based in the United States have traded during a standard trading session in the stock market. It is the second oldest U.S. market index after the Dow Jones Transportation Average, which Dow also created. Dow is among the most closely-watched benchmark indices tracking targeted stock market activity.

The studied period is between 1 January 2000 and 30 December 2009, and the data in this study are downloaded from the website Yahoo Finance [1], Prague Stock Exchange and Bloomberg. Figures 2.1 and 2.2 presents the time plots of the time series, which fluctuate on a daily basis.

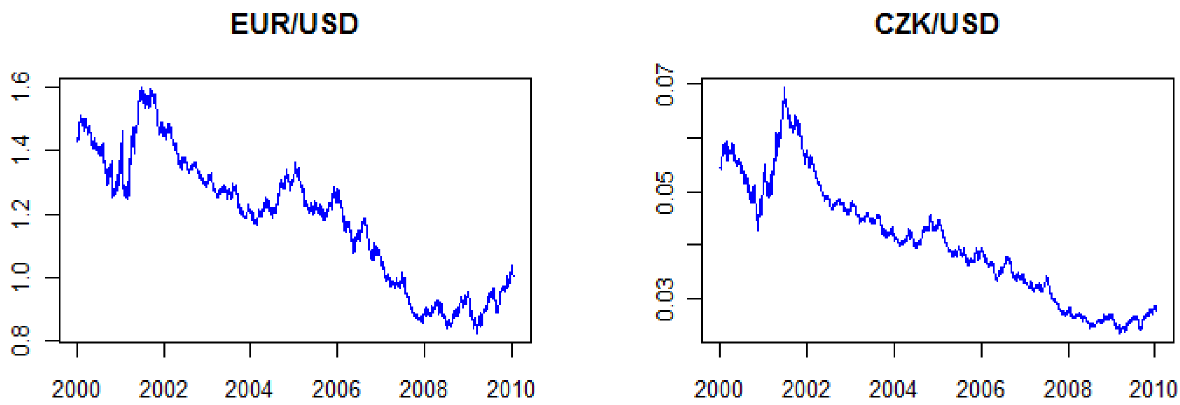

Figure 2.1: FX rates EUR/USD and CZK/USD during January 2000 and December 2009.

Note that we denote successive price observations made at time t and $t-1$ as P_t and P_{t-1}, respectively, then transformation a price series P_t into a log return or simply return series r_t

$$r_t = \log \frac{P_t}{P_{t-1}} = \log P_t - \log P_{t-1}.$$

Note that, r_t represents the interest or percentage yield obtained within period $t-1$ to t. Plots of returns computed from our data can be found in Figures 2.3 and 2.4.

2.2 Stylized features of financial data

When the statistical models are developed to describe financial data, it is often useful to have some directives which describe the most important characteristic features of the data which the models should have to consider. These directives are referred to as "stylized

[1] http://finance.yahoo.com

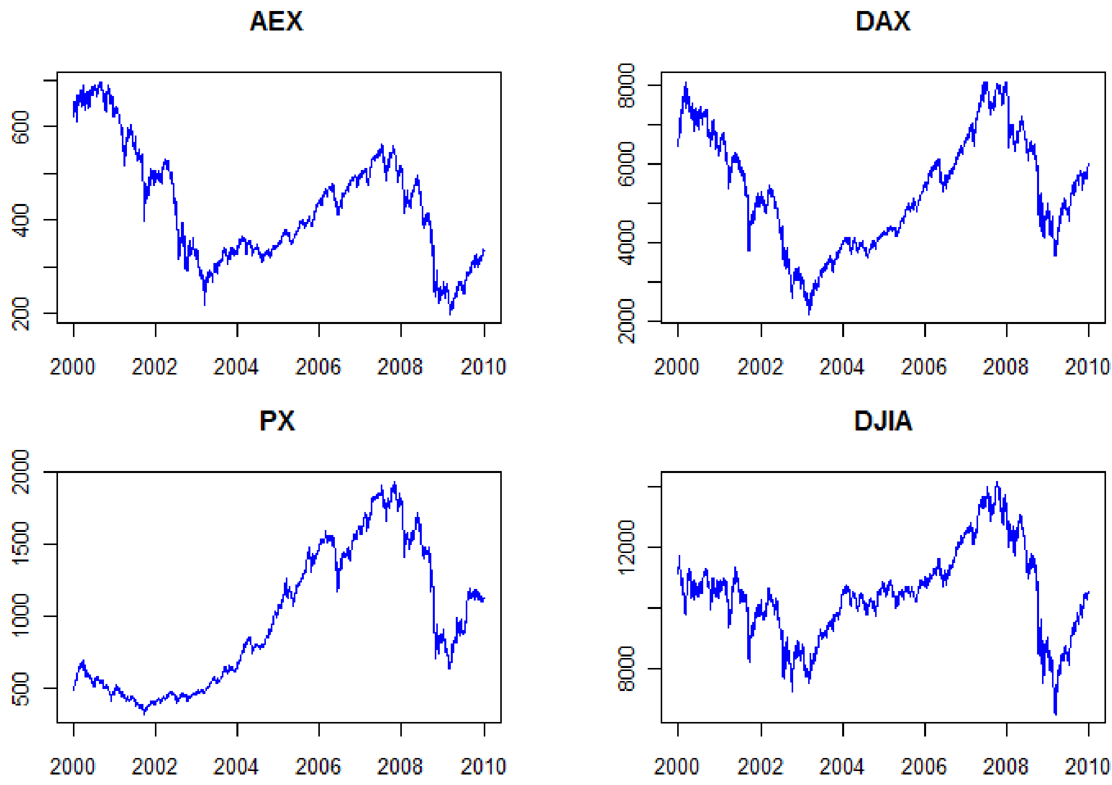

Figure 2.2: Stock indices of AEX, DAX, PX, DJIA correspond, to the stock markets in Amsterdam, Frankfurt, Prague and the U.S. during January 2000 and December 2009.

features" or "stylized facts". Taylor [28] mentioned that "General properties that are expected to be present in any set of returns are called stylized facts." Stylized features are the result of more than half of the century empirical studies on financial time series, examines their properties from a statistical point of view. Let us start by stating a set of stylized facts which are common to a wide set of financial analysis. Let us explain number of them.

2.2.1 Volatility clustering

Volatility clustering refers to observation, as noted by Mandelbrot [24], that large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes what is a well-known stylized fact in financial markets. In simple terms, volatility clustering manifests itself as quiet periods interrupted by volatile periods called turbulence. As can be seen from Figures 2.3 and 2.4 the changes between large return changes and relatively silent phases of small price activity is a slow process and does not indicate any significant autocorrelation. A look at the autocorrelation function Figure 2.5

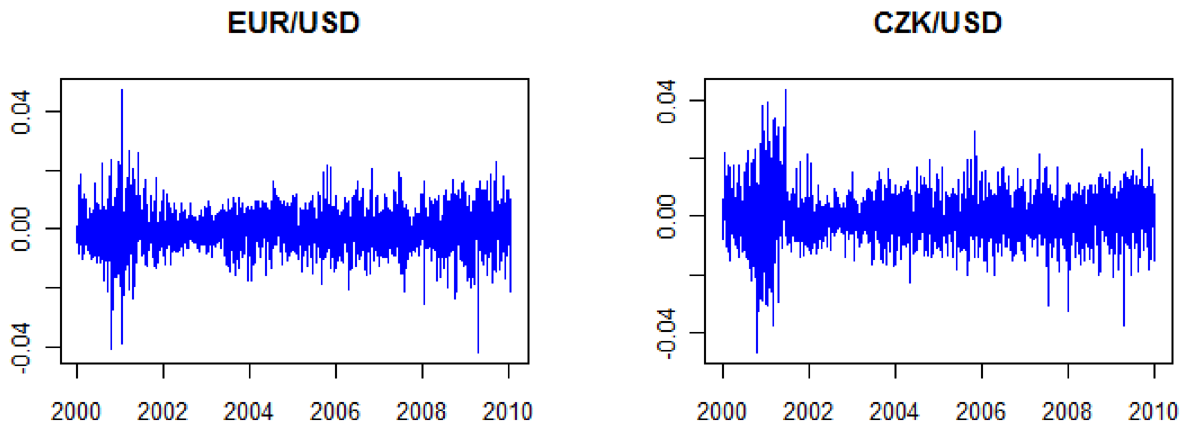

Figure 2.3: Return Series of FX rates EUR/USD and CZK/USD during January 2000 and December 2009.

of the realization shows a rapid decay of the autocorrelations of price changes. We can also see that absence of autocorrelation in returns does not imply the independence of the increments. Simple nonlinear function of returns, such as squared returns or absolute returns, show significant positive autocorrelation.

2.2.2 Heavy tails

The observation of time series have a distribution, which is often assumed to be a normal (Gaussian). However, empirical studies of any financial time series shows, that this is not quite correct. Mandelbrot was the first to show that returns on financial markets are not Gaussian, but exhibit excess kurtosis. Heavy tails distribution means that the unconditional price or return distributions tend to have fatter (leptokurtic) tails then the normal distribution. In terms of shape, as we can see from Figure 2.6 a leptokurtic distribution has a more acute peak around the mean (that is, a higher probability than a normally distributed variable near the mean) and fatter tails (that is, higher probability for extreme events than in normally distributed data). Measure of fatness of the tails of a random variable X_t distribution is kurtosis defined as $\kappa_4(X) = \mathbb{E}(X - \mathbb{E}X)^4/(varX)^2$. For normally distributed variable is equal to 3.

2.2.3 Aggregational Gaussianity

By aggregational Gaussianity we mean the fact that long term aggregation of returns, in the sense of assuming the returns over longer periods, will lead to approximately normally distributed variables.

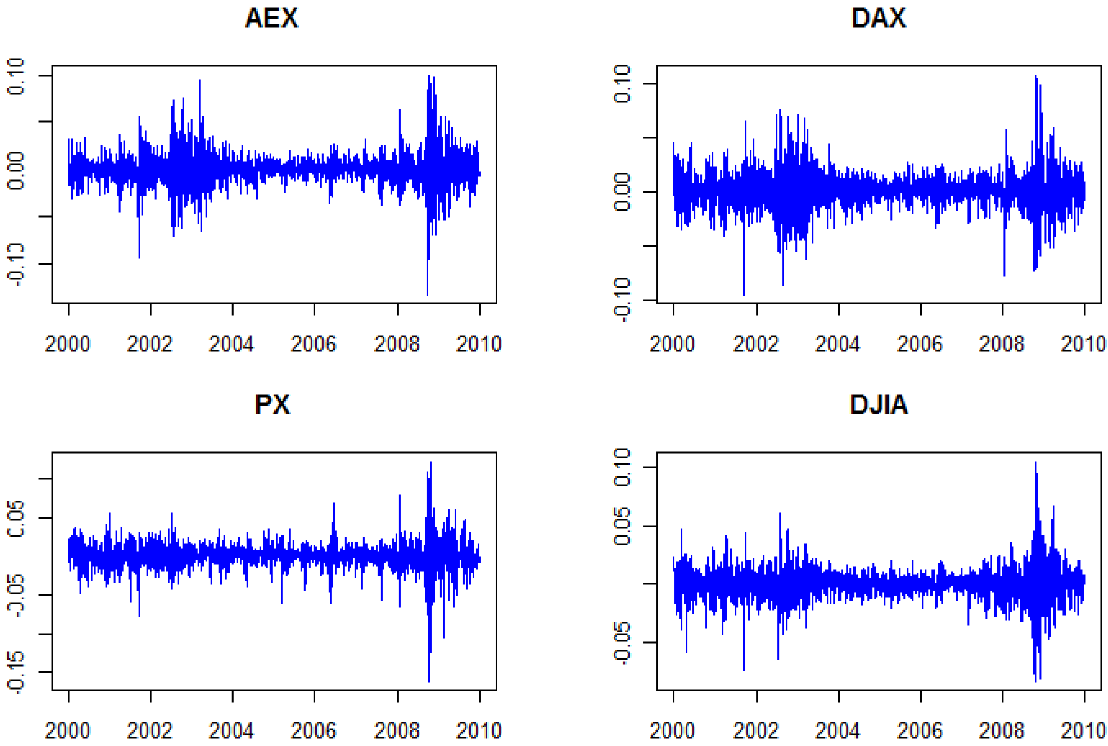

Figure 2.4: Return Series of AEX, DAX, PX, DJIA correspond, to the stock markets in Amsterdam, Frankfurt, Prague and the U.S. during January 2000 and December 2009.

2.2.4 Leverage effect

The volatility tend to be larger for the price falls, than for price rises, when the magnitude of the price rise and fall, is identical. This is asymmetric influence of negative and positive information on future level of volatility.

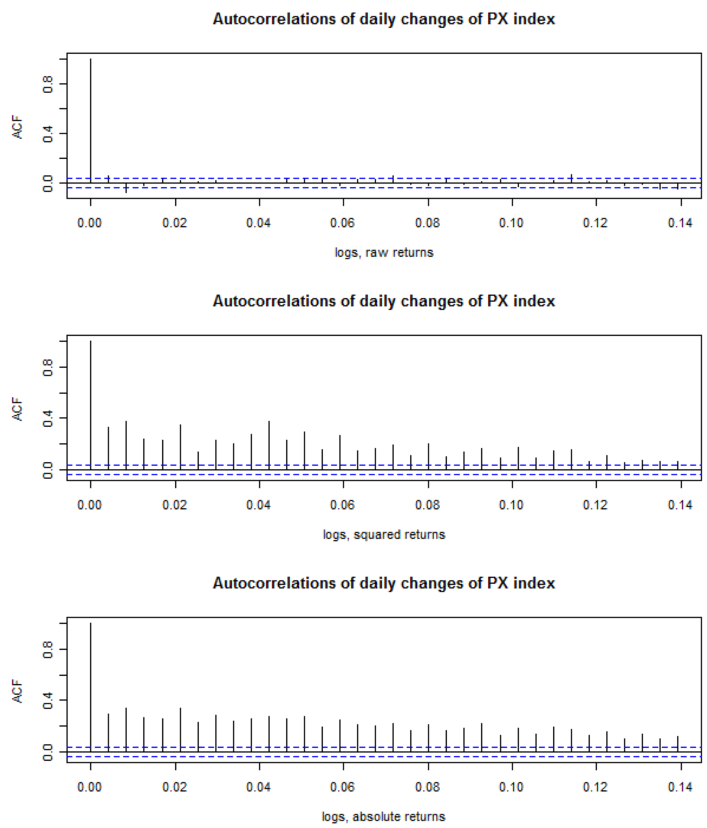

Figure 2.5: Autocorrelations of daily changes of PX index computed by R programming software.

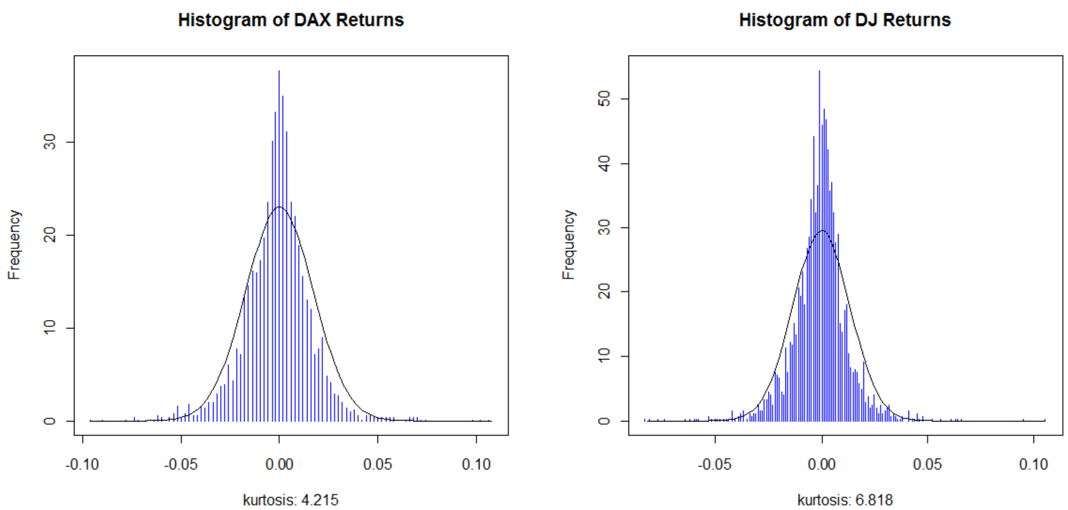

Figure 2.6: Histogram of price increments Dow Jones and Prague stock indices during January 2000 and December 2009 computed by R programming software, solid line represent density function of normal distribution.

Chapter 3

GARCH

We start with basic univariate GARCH framework. In 1982 Engle[1] introduced a volatility process with time varying conditional variance, known as the Autoregressive conditional heteroskedasticity (ARCH) process. The popularity of this class of models can be inferred, that several hundred research papers using this model have appeared in the decade since its introduction. Detailed discussion, technical conditions and statistical properties this type of models have been studied for example in Weiss [33]. However, in many of the financial applications with the ARCH models empirical works shows that high ARCH order has to be selected to catch the dynamic of the conditional variance of the financial time series. The high order of the model of course implies that many parameters have to be estimated and this is also difficult for computation. Another practical difficulty is that with high order of the model estimation will often lead to the violation of the non-negativity constraints that are needed to ensure that the conditional variance is always positive. Four years after, Bollerslev [7] introduced the generalized version of the model namely GARCH model as a natural solution of the high ARCH orders problem. Bollerslev‘s model is based on an infinite ARCH and reduces the number of parameters that needs to be estimated from infinite number to just a few. The main principle of modelling time series using GARCH is that a large movements in period ("bursts of activity") increase the variance of the movements in the following periods. This constructs a feedback mechanism whereby a single univariate series determines both, the time series and its conditional variance structure. Alternating between volatile and quiet periods as we mentioned before is called volatility clustering.

In finance GARCH-type processes became very favourite to model returns of stocks,

[1]The winner of the 2003 Nobel Memorial Prize in Economic Sciences for his work Engle, R. F. : Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.

exchange rates, stock indices and other series observed at equidistant time points. They have been designed to capture so-called stylized facts of such data which are, as we said before volatility clustering and others like dependence without correlation and tail heaviness. There exist many types of GARCH processes and the linear ones were the earliest.

Before we start with the definition of GARCH model we introduce some of the basic building blocks of time series analysis, which we will often use in the next parts. First are the white noise series which can be defined as a doubly infinite sequence of random variables Z_t with mean zero and finite variances. Special types of white noise series are independent and identically distributed series. The i.i.d. white noise series ourselves are not so interesting, but are important to construct other series, for instance series where the random variables are dependent, so that the future can be predicted from past. We shall speak about a heteroscedastic white noise processes if the autocovariances at nonzero lags vanish, but the variances are still time-dependent. A related concept is a martingale difference sequences. *Filtration $\mathcal{F}_t$ is a non decreasing collection of σ-fields $\dots \subset \mathcal{F}_{-1} \subset \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots$. A martingale difference sequence relative to the filtration $\mathcal{F}_t$ is a time series X_t such that X_t is $\mathcal{F}_t$-measurable and $\mathbb{E}(X_t|\mathcal{F}_{t-1}) = 0$ almost surely for every t.* This implies that any martingale difference sequence X_t with finite second moments is a white noise series with zero first moment given the past. However conversely not every white noise series is a martingale difference sequence relative to a natural filtration.

3.1 Linear GARCH

In this section we closely follow van der Vaart [30], chapter 8, whereas there exists plenty of possible definitions of GARCH process. We chose this interpretation, because we can easily move into multivariate framework.

Definition 1. *A GARCH (p,q) process is a martingale difference sequence X_t relative to a given filtration $\mathcal{F}_t$, whose conditional variances $\sigma_t^2 = \mathbb{E}(X_t^2|\mathcal{F}_{t-1})$ satisfy, for every $t \in \mathbb{Z}$ and given constants α, $\phi_1, \dots, \phi_p$, $\theta_1, \dots, \theta_q$,*

$$\sigma_t^2 = \alpha + \sum_{i=1}^{p} \phi_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \theta_i X_{t-i}^2, \tag{3.1}$$

where

$p \geq 0, \quad q \geq 0, \quad \alpha > 0,$

$\phi_i \geq 0, \quad i = 1, \ldots, p,$
$\theta_i \geq 0, \quad i = 1, \ldots, q.$

To understand properties of GARCH processes, it is informative to use the following representation. We can rewrite equation (3.1) for the conditional variance σ_t^2 using lag or backshift operator B, defined as $BX_t = X_{t-1}$ and convention that $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ and $\theta(z) = \theta_1 z + \cdots + \theta_q z^q$. Then we can rewrite (3.1) as

$$\phi(B)\sigma_t^2 = \alpha + \theta(B)X_t^2.$$

Note that for $p = 0$, i.e. if the coefficients $\phi_1, \ldots, \phi_p$ all vanish, then the process of σ_t^2 reduces to a pure ARCH (q) process, and for $p = q = 0$ the process of σ_t^2 reduces to the white noise. In the ARCH (q) process the conditional variance σ_t^2 is modelled as linear function of $X_{t-1}^2, \ldots, X_{t-q}^2$, when as the GARCH ($p, q$) allows lagged conditional variances to enter as well. If we assume $\sigma_t > 0$, then we can define $Z_t = X_t/\sigma_t$. The martingale difference property of X_t with the definition (3.1) of σ_t^2 as conditional variance implies that

$$\mathbb{E}(Z_t|\mathcal{F}_{t-1}) = 0, \qquad \mathbb{E}(Z_t^2|\mathcal{F}_{t-1}) = 1. \tag{3.2}$$

We can also define the GARCH process X_t starting with given martingale difference process Z_t and a process σ_t that is $\mathcal{F}_{t-1}$ measurable and then $X_t = \sigma_t Z_t$. If (3.1) is valid then σ_t is the conditional variance of X_t. In the most cases we assume that the variables Z_t are i.i.d.. Then Z_t is independent of $\mathcal{F}_{t-1}$. This assumption is equivalent to assuming that conditional law of the variables $Z_t = X_t/\sigma_t$ given $\mathcal{F}_{t-1}$ is a given distribution, for instance standard normal distribution and then we can write [2]

$$X_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2). \tag{3.3}$$

We now move on to the stationary condition of the GARCH processes. Consider the following construction. Let Z_t be a martingale difference sequence such that $\mathbb{E}(Z_t|\mathcal{F}_{t-1}) = 0$, $\mathbb{E}(Z_t^2|\mathcal{F}_{t-1}) = 1$, defined on a fixed probability space. Then we construct a GARCH process such that $X_t = \sigma_t Z_t$ by first defining the process of squares σ_t^2 in terms of the Z_t. If the coefficients α, ϕ_i, θ_i are nonnegative we obtain a stationary solution if $\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i < 1$. Note that the non-negativity of α, ϕ_i, θ_i is necessary condition for the non-negativity of σ_t^2. We can state the following theorem see van der Vaart [30].

Theorem 1. *Let $\alpha > 0$, let $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ be non-negative numbers, and let Z_t be a martingale difference sequence satisfying (3.2) relative to a filtration $\mathcal{F}_t$. Then there exist*

[2] In assuming the conditional distribution to be normal we follow Engle [15], but other distributions could be applied as well.

a unique stationary GARCH process X_t such that $X_t = \sigma_t Z_t$, where $\sigma_t^2 = \mathbb{E}(X_t^2|\mathcal{F}_{t-1})$, if and only if $\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i < 1$. This unique process among the GARCH processes X_t such that $X_t = \sigma_t Z_t$ has bounded second moments and $\mathbb{E}(X_t) = 0$, $var(X_t) = \alpha[1 - \sum_{i=1}^{p}\phi_i - \sum_{i=1}^{q}\theta_i]^{-1}$.

Proof. Assume that $\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i < 1$. Using substitution we get

$$\begin{aligned}
\sigma_t^2 &= \alpha + \sum_{i=1}^{p} \phi_i \sigma_{t-i}^2 + \sum_{i=1}^{q} \theta_i Z_{t-i}^2 \sigma_{t-i}^2 \\
&= \alpha + \sum_{j=1}^{p} \phi_j \left(\alpha + \sum_{i=1}^{p} \phi_i \sigma_{t-i-j}^2 + \sum_{i=1}^{q} \theta_i Z_{t-i-j}^2 \sigma_{t-i-j}^2 \right) + \\
&\sum_{j=1}^{q} \theta_j Z_{t-j}^2 \left(\alpha + \sum_{i=1}^{p} \phi_i \sigma_{t-i-j}^2 + \sum_{i=1}^{q} \theta_i Z_{t-i-j}^2 \sigma_{t-i-j}^2 \right) \\
&\vdots \\
&= \alpha \sum_{k=0}^{\infty} M(t,k),
\end{aligned} \tag{3.4}$$

where $M(t,k)$ are all the terms of the form

$$\prod_{i=1}^{p} \phi_i^{a_i} \prod_{j=1}^{q} \theta_j^{b_j} \prod_{l=1}^{n} Z_{t-S_l}^2,$$

for

$$\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j = k, \qquad \sum_{i=1}^{q} b_i = n,$$

and

$$1 \le S_1 < S_2 < S_3 < \cdots < S_n \le \max\{kq, (k-1)q + p\}.$$

Since $\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i < 1$, then series $\alpha \sum_{k=0}^{\infty} M(t,k)$ converges and thus

$$\begin{aligned}
M(t,0) &= 1, \\
M(t,1) &= \sum_{i=1}^{p} \phi_i + \sum_{i=1}^{q} \theta_i Z_{t-i}^2, \\
M(t,2) &= \sum_{j=1}^{p} \phi_j \left(\sum_{i=1}^{p} \phi_i + \sum_{i=1}^{q} \theta_i Z_{t-i-j}^2 \right) + \sum_{j=1}^{q} \theta_j Z_{t-j}^2 \left(\sum_{i=1}^{p} \phi_i + \sum_{i=1}^{q} \theta_i Z_{t-i-j}^2 \right),
\end{aligned}$$

and in general

$$M(t,k+1) = \sum_{i=1}^{p} \phi_i M(t-i,k) + \sum_{i=1}^{q} \theta_i Z_{t-i}^2 M(t-i,k). \tag{3.5}$$

Since Z_t^2 is i.i.d., the moments of $M(t,k)$ do not depend on t, and in particular

$$\mathbb{E}(M(t,k)) = \mathbb{E}(M(s,t)), \tag{3.6}$$

for all k, t, s. From (3.5) and (3.6) we get

$$\begin{aligned} \mathbb{E}(M(t,k+1)) &= \left(\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i\right)\mathbb{E}(M(t,k)) \\ &\vdots \\ &= \left(\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i\right)^{k+1}\mathbb{E}(M(t,0)) \\ &= \left(\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i\right)^{k+1}. \end{aligned} \tag{3.7}$$

Finally by (3.5), (3.6) and (3.7),

$$\begin{aligned} \mathbb{E}(X_t^2) = \alpha\mathbb{E}\left(\sum_{k=0}^{\infty} M(t,k)\right) &= \alpha\sum_{k=0}^{\infty}\mathbb{E}(M(t,k)) \\ &= \alpha\left[1 - \sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i\right]^{-1}, \end{aligned} \tag{3.8}$$

if and only if

$$\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i < 1,$$

and X_t^2 converges almost surely. $\mathbb{E}(X_t) = 0$ and $cov(X_t, X_s) = 0$ for $t \neq s$ follows immediately by symmetry. □

Note that in practice, one observes a sample $X_1, \ldots, X_T$ and also in this situation we assume that this vector comes from a stationary model. In particular, we assume that X_0 and σ_t have the stationary initial distribution. The previous theorem implies that the condition $\sum_{i=1}^{p}\phi_i + \sum_{i=1}^{q}\theta_i < 1$, is necessary for existence of a stationary GARCH process with bounded second moments, but strong than necessary if we are interested in a strictly stationary solution to the GARCH equations with possibly infinite second moments. Conditions for strict stationarity we will specify later.

The GARCH process has a close relation to widely known Box-Jenkins ARMA processes. Consider $\eta_t = X_t^2 - \sigma_t^2$ so that $\sigma_t^2 = X_t^2 - \eta_t$. Rearranging the terms in (3.1) process can be interpreted as an ARMA process

$$X_t^2 = \alpha + \sum_{i=1}^{\max(p,q)}(\theta_i + \phi_i)X_{t-i}^2 + \eta_t - \sum_{i=1}^{p}\phi_i\eta_{t-i}. \tag{3.9}$$

It is easy to check that η_t is a martingale difference series (i.e. $\mathbb{E}(\eta_t) = 0$ and $cov(\eta_t, \eta_{t-i}) = 0$ for $i \geq 1$) and a white noise sequence if its second moments exist and are independent of t. However η_t in general is not an i.i.d. sequence. Equation (3.9) is an characterizing equation for ARMA process X_t^2 of orders $m = max(p, q)$ and p with AR parameters $\phi(B) + \theta(B)$, MA parameters $-\theta(B)$ and uncorrelated innovation sequence $\{X_t^2 - \sigma_t^2\}$.

Furthermore, it may be seen that the GARCH model is based on a infinite dimensional ARCH (∞) model. To ensure a well-defined process all the parameters in the infinite dimensional ARCH representation

$$\sigma_t^2 = \frac{\alpha}{1-\phi(1)} + \frac{\theta(B)}{1-\phi(B)} X_t^2, \tag{3.10}$$

must be nonnegative. This assumed that all the roots of the polynomial $1 - \phi(B) = 0$ lie outside the unit circle.

In Chapter 2 we introduced some of the stylized facts of financial time series and let now discuss how the GARCH models consider these features. Volatility clustering is one of the features that is always presented in financial time series and it is completely captured by GARCH models. It is because large absolute values of a GARCH series at time $t-1, \ldots, t-q$ in the past lead, through the GARCH equation (3.1), to a large conditional variance σ_t^2 at time t, and hence the value $X_t = \sigma_t Z_t$ of the time series at time t tends to be large. By equation we can see that a large σ_{t-1}^2 or X_{t-1}^2 gives rise to a large σ_t^2. So then large σ_{t-1}^2 tends to be followed by another large σ_t^2, generating the behavior of volatility clustering.

Another stylized fact that may be very often observed in financial time series are leptokurtic tails of the marginal distribution. As we mentioned before a quantitative measure of fatness of the tails distribution of a random variable X is the kurtosis defined as $\kappa_4(X) = \mathbb{E}(X - \mathbb{E}X)^4/(varX)^2$. And is equal to 3 for a normally distributed variable. If $X_t = Z_t \sigma_t$, where σ_t is $\mathcal{F}_{t-1}$ measurable and Z_t is independent of $\mathcal{F}_{t-1}$ with mean zero and variance one, then

$$\mathbb{E}X_t^4 = \mathbb{E}\sigma_t^4 \mathbb{E}Z_t^4 = \kappa_4(Z_t)\mathbb{E}(\mathbb{E}(X_t^2|\mathcal{F}_{t-1}))^2 \geq \kappa_4(Z_t)(\mathbb{E}\mathbb{E}(X_t^2|\mathcal{F}_{t-1}))^2 = \kappa_4(Z_t)(\mathbb{E}X_t^2)^2.$$

Dividing the left and right sides by $(\mathbb{E}X_t^2)^2$, we can see immediately that $\kappa_4(X_t) \geq \kappa_4(Z_t)$. A soon as the variance of the random variable $\mathbb{E}(X_t^2|\mathcal{F}_{t-1})$ is large the difference can be significant. Taking the difference of the left and right sides gives

$$\kappa_4(X_t) = \kappa_4(Z_t)\left(1 + \frac{var\mathbb{E}(X_t|\mathcal{F}_{t-1})}{(\mathbb{E}X_t^2)^2}\right).$$

Consequently, the tail distribution of a GARCH process is heavier than a normal distribution. If we use a Gaussian process Z_t, then the kurtosis of the observed series X_t is always bigger than 3. It follows that the GARCH structure is able to capture some of the observed leptokurtosis of financial time series.

Next stylized fact observed in financial time series are positive auto-correlations for the sequence of squares X_t^2. The auto-correlation function of the squares of a GARCH series will exist under appropriate additional conditions on the coefficients and the noise process Z_t. We can compute auto-correlation function of this using the ARMA relation (3.9) for the square process X_t^2 and using formulas for the auto-correlation function of an ARMA process. Note that the process η_t in (3.9) is defined through X_t and hence its variance depends on the parameters in the GARCH relation.

3.2 GARCH (1,1) process

The most simple example of GARCH processes is GARCH $(1,1)$ process in which conditional variances are given by

$$\sigma_t^2 = \alpha + \phi\sigma_{t-1}^2 + \theta X_{t-1}^2, \qquad \alpha > 0, \quad \phi \geq 0, \quad \theta \geq 0.$$

From Theorem 1 $\phi + \theta < 1$ suffices wide-sence stationary. If we assume stationarity of the process X_t then expectation of σ_t^2 does not depend on t and is equal

$$\mathbb{E}\sigma^2 = \mathbb{E}X_t^2 = \frac{\alpha}{1-\phi-\theta}.$$

By squaring the GARCH equation we can find

$$\sigma_t^4 = \alpha^2 + \phi^2\sigma_{t-1}^4 + \theta^2 X_{t-1}^4 + 2\alpha\phi\sigma_{t-1}^2 + 2\alpha\theta X_{t-1}^2 + 2\phi\theta\sigma_{t-1}^2 X_{t-1}^2.$$

If Z_t is independent of $\mathcal{F}_{t-1}$, then $\mathbb{E}\sigma_t^2 X_t^2 = \mathbb{E}\sigma_t^2$ and $\mathbb{E}X_t^4 = \kappa_4(Z_t)\mathbb{E}\sigma_t^4$. If we assume that moments exist and are independent of t, then

$$\mathbb{E}(X_t^4) = \mathbb{E}(Z_t^4)\mathbb{E}(\sigma_t^4) = 3\alpha^2(1+\phi+\theta)[(1-\phi-\theta)(1-\phi^2-2\phi\theta-3\theta^2)]^{-1}.$$

Since the marginal kurtosis is given by

$$\kappa_4 = \frac{\mathbb{E}(X_t^4)}{[\mathbb{E}(X_t^2)]^2},$$

from previous calculus it immediately follows that

$$\kappa_4 = \frac{3(1+\phi+\theta)(1-\phi-\theta)}{(1-\phi-\theta)(1-\phi^2-2\phi\theta-3\theta^2)}.$$

A little calculus shows

$$\begin{aligned} 3var(\sigma_t^2) &= \mathbb{E}(X_t^4) - 3\left[\mathbb{E}(X_t^2)\right]^2 \\ &= \frac{3\alpha^2(1+\phi+\theta)}{(1-\phi-\theta)(1-\phi^2-2\phi\theta-3\theta^2)} - 3\left[\frac{\alpha}{1-\phi-\theta}\right]^2 \\ &= \frac{3\alpha^2}{(1-\phi-\theta)^2}\frac{2\theta^2}{(1-\phi^2-2\phi\theta-3\theta^2)}. \end{aligned}$$

Since from the assumptions we have that $\alpha > 0$, $1-\phi-\theta > 0$ and $1-\phi^2+2\phi\theta-3\theta^2 < 1$, it follows that all the factors in are positive so we conclude that the GARCH $(1,1)$ process is leptokurtic.

3.3 Estimation of the GARCH model

Existence of stationary solution for the GARCH process is the key ingredient to derive the estimation procedure and asymptotic theory. Consider the GARCH (p,q) model defined as before. The existence of unique and strictly stationary solution to the GARCH equations is that if and only if the sequence of matrices A_t, where

$$A_t = \begin{pmatrix} \phi_1+\theta_1 Z_{t-1}^2 & \phi_2 & \cdots & \phi_{p-1} & \phi_p & \theta_2 & \dots & \theta_{q-1} & \theta_q \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 \\ Z_{t-1}^2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

has a strictly negative top Lyapunov exponent $\gamma < 0$, where

$$\gamma = \inf_{T\in N^*} \frac{1}{T}\mathbb{E}\log\|A_{-1}A_{-2}\dots A_{-T}\| = \lim_{n\to\infty}\frac{1}{T}\log\|A_{-1}A_{-2}\dots A_{-T}\|, \quad a.s. \tag{3.11}$$

Here $\|.\|$ may be any matrix norm because the definition of γ does not depend on the choice of a norm. The Lyapunov exponent in general cannot be calculated explicitly for the model under study, but it can be estimated numerically for a given sequence Z_t. Existence of top Lyapunov exponent γ is guaranteed by the inequality $\mathbb{E}(\log^+\|A_1\|) \le \mathbb{E}\|A_1\| < \infty$.

Let $Y_t = (\sigma_t^2, \dots, \sigma_{t-p+1}^2, X_{t-1}^2, \dots, X_{t-q+1}^2)' \in \mathbb{R}^{p+q}$ and $b = (\alpha, 0, \dots, 0)' \in \mathbb{R}^{p+q}$. Then GARCH equation can be equivalently rewritten as the system of equations

$$Y_t = A_t Y_{t-1} + b, \tag{3.12}$$

and if $\gamma < 0$, the unique strictly stationary solution is given by

$$Y_t = b + \sum_{k=1}^{\infty} A_t A_{t-1} \dots A_{t-k+1} b. \tag{3.13}$$

Theorem 2. *Let $\alpha > 0$, and $\phi_1, \dots, \phi_p$, $\theta_1, \dots, \theta_q$ be a nonnegative numbers, and let Z_t be an i.i.d. sequence with mean zero and unit variance. There exists a strictly stationary GARCH process X_t such that $X_t = \sigma_t Z_t$, where $\sigma_t^2 = \mathbb{E}(X_t^2 | \mathcal{F}_{t-1})$ and $\mathcal{F}_t = \sigma(Z_t, Z_{t-1}, \dots)$, if and only if the top Lyapounov coefficient of the random matrices A_t given by (3.11) is strictly negative. For this process $\sigma(X_t, X_{t-1}, \dots) = \sigma(Z_t, Z_{t-1}, \dots)$.*

Proof. We give a short proof for more details see van der Vaart [30]. Let first define b as $b = \alpha e_1$, where e_i is the ith unit vector in $\mathbb{R}^{p+q+1}$. If γ' is strictly larger than the top Lyapounov exponent γ, then as $T \to \infty$

$$\|A_t A_{t-1} \dots A_{t-T+1} b\| < e^{\gamma' T} \|b\|, \qquad a.s.$$

If $\gamma' < 0$ then $\sum_T \|A_t A_{t-1} \dots A_{t-T+1} b\| < \infty$ almost surely. Then the series on the righ hand side of (3.13) converges almost surely and defines a process Y_t. Next step is define processes σ_t and X_t by setting $\sigma_t = \sqrt{Y_t}$ and $X_t = \sigma_t Z_t$. It follows from (3.12) that this equation satisfy the GARCH relation. And because the process is a fixed measurable transformation of $(Z_t, Z_{t-1}, \dots)$ for each t, then the process (σ_t, X_t) is strictly stationary. Now we have to prove that $\sigma(X_t, X_{t-1}, \dots) = \sigma(Z_t, Z_{t-1}, \dots)$. We can see immediately that X_t is $\sigma(Z_t, Z_{t-1}, \dots)$-measurable because of construction of X_t. For second implication Z_t is $\sigma(X_t, X_{t-1}, \dots)$-measurable we use relation $(\phi - \theta)(B) X_t^2 = \alpha + \phi(B)\eta_t$, for $\eta_t = X_t^2 - \sigma_t^2$. We conclude that η_t is $\sigma(X_t^2, X_{t-1}^2, \dots)$-measurable, if the polynomial ϕ has no zeros on the unit disc.

Finally, we show the necessity of the top Lyapounov exponent being negative. If there exists a strictly stationary solution to the GARCH equations, then by the non-negativity of the coefficients

$$\sum_{i=1}^{T} A_0 A_{-1} \dots A_{-T+1} b \leq Y_0,$$

for every T, and then

$$A_0 A_{-1} \dots A_{-T+1} b \to 0, \qquad a.s. for T \to \infty$$

By the definition of b the last equation is equivalent to $A_0 A_{-1} \dots A_{-T+1} e_1 \to 0$. Using the structure of the matrices A_t we next see that $A_0 A_{-1} \dots A_{-T+1} \to 0$ in probability as $T \to \infty$. Because the matrices A_t are independent and the event where $A_0 A_{-1} \dots A_{-T+1} \to 0$ is a tail event, this event must have probability one. This is possible only if the top Lyapounov exponent γ of the matrices A_t is strictly negative. □

The estimation of GARCH models is usually carried out using maximum likelihood estimation. However, obtaining a likelihood function is not straightforward. We have data $X_1, \ldots, X_T$ assumes to be random observations which are given from a distribution $F_X(x;\theta)$ and we denote joint probability density $(x_1, \ldots, x_T) \mapsto p_{T,\theta}(x_1, \ldots, x_T)$ of these observations that depends on a unknown parameter θ from the parameter space Θ. The stochastic process defined by

$$L(\theta) \mapsto p_{T,\theta}(X_1, \ldots, X_T) = p_\theta(x_1)p_\theta(x_2|x_1)\ldots p_\theta(x_T|x_{T-1}, \ldots, x_1),$$

is the likelihood function. If the observations $X_1, \ldots, X_T$ are i.i.d. then the likelihood function is the product of the likelihood functions of the individual observations. We may extend the conditioning in each term to include the whole past, yielding the quasi (pseudo) likelihood

$$L(\theta) = p_\theta(x_1|x_0, x_{-1}, \ldots)p_\theta(x_2|x_1, x_0, \ldots)\ldots p_\theta(x_T|x_{T-1}, x_{T-2}, \ldots).$$

Observe that the formula of the quasi likelihood function requires to know all values $X_{T-1}, \ldots, X_0, X_{-1}, \ldots$ but in practice the variables $X_0, X_{-1}, \ldots$ are not observed. However the past observations $X_0, X_{-1}, \ldots$ do not play an important role in defining quasi likelihood because the likelihood does not change much if the conditioning in each term is limited to a fixed number of variable in the past, and most of the terms of the product will take almost a common form. Similarly, if the time series is AR of order p, i.e. $p(x_t|x_{t-1}, x_{t-2}, \ldots)$ depends only on $x_t, x_{t-1}, \ldots, x_{t-p}$, then the two likelihoods differ only in p terms. This should be negligible when T is large relative to p. In GARCH (p, q) situation a practical implementation is to define $\sigma_0^2, \ldots, \sigma_{1-p}^2$ and $X_0^2, \ldots, X_{1-q}^2$ to be zero, and compute next $\sigma_1^2, \sigma_2^2, \ldots$ recursively using observation $X_1, \ldots, X_T$.

Now, suppose that we have GARCH process $X_t = \sigma_t Z_t$ with the noise process Z_t and given observations $X_1, \ldots, X_T$. A common practice in estimation of the GARCH models is to assume Z_t to be Gaussian when deriving the likelihood and this is a basic estimation method for classic GARCH models. The vector of parameters is

$$\theta = (\theta_1, \ldots, \theta_{p+q+1})' = (\alpha, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)'$$

and belongs to a parameter space Θ. We denote true unknown parameter value by $\theta_0 = (\alpha_0, \phi_{01}, \ldots, \phi_{0p}, \theta_{01}, \ldots, \theta_{0q})'$, where $\theta_0 \in \Theta$. Then

$$\begin{aligned} &p_\theta(x_1|x_0, x_{-1}, \ldots)p_\theta(x_2|x_1, x_0, \ldots)\ldots p_\theta(x_T|x_{T-1}, x_{T-2}, \ldots) \\ &= \prod_{t=1}^{T} \frac{1}{\sigma_t(\theta)} f_z\left(\frac{X_t}{\sigma_t(\theta)}\right). \end{aligned} \tag{3.14}$$

Note that X_t given the whole past $X_{t-1}, X_{t-2}, \ldots$, and conditioning argument yields the density function p_θ of $X_1, \ldots, X_T$ through the conditional densities of the X_t's given $X_1 = x_1, \ldots, X_T = x_T$. Assuming that Z_t is Gaussian then conditionally on initial values, the quasi log-likelihood function for a GARCH (p, q) process is given by (ignoring some constants)

$$L_t(\theta) = L_t(\theta; X_1, \ldots, X_T) = \frac{1}{2T}\sum_{t=1}^{T} l_t(\theta), \tag{3.15}$$

where

$$l_t(\theta) = -\left(\log \sigma_t^2(\theta) + \frac{X_t^2}{\sigma_t^2(\theta)}\right). \tag{3.16}$$

The quasi maximum likelihood estimator of θ is defined as any measurable solution $\hat{\theta}_T$ that maximizes the likelihood function within parameter space Θ, i.e.,

$$\hat{\theta}_T = arg \max_{\theta \in \Theta} L_t(\theta). \tag{3.17}$$

The resulting value is the quasi maximum likelihood estimator of the parameters of a GARCH (p, q) process. However, as we can see there are obvious problems with this procedure. Most controversial assumption is that Z_t as Gaussian noise process. Although this is not the most realistic assumption, it gives nice results such as $\sqrt{T}$-consistency (consistency with $\sqrt{T}$-rate) and $\sqrt{T}$-normality. Theoretical works (see references below) shows that asymptotic properties remain valid for large number of noise distributions. Attempts to replace the Gaussian density of the Z_t's by a more realistic density for example t-density can lead to non-consistency of the QMLE. If one wants to achieve consistency, the exact density of underlying Z_t need to be known. But when dealing with data one can never rely on this assumption.

There exist various papers dealing with the asymptotic properties of the quasi MLE and have been studied initially by Weiss [33] but only for pure ARCH (q) processes and fourth-order moment conditions on the process. The problem of finding weak assumptions for the consistency and asymptotic normality of the QMLE in GARCH models solved Lee and Hansen [22] and Lumsdaine [23] for the GARCH $(1, 1)$ case. The asymptotic properties of QMLE for the GARCH (p, q) models have been studied by, amongst others, Francq and Zakoïan [19] and Berkes et al. [5]. We introduce convergence and asymptotic normality under the conditions presented in Francq and Zakoïan. For more details and proofs of the theorems see Francq and Zakoïan.

Assume that Z_t is i.i.d. and the QMLE $\hat{\theta}_T$ maximizes the likelihood under Θ. Let $\mathcal{A}_{\theta_0}(z) = \sum_{i=1}^{q} \theta_i z^i$ and $\mathcal{B}_{\theta_0}(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ with the convention $\mathcal{A}_{\theta_0}(z) = 0$ if $q = 0$

and $\mathcal{B}_{\theta_0}(z) = 1$ if $p = 1$. To show strong consistency the following assumptions will be made.

Assumption 1. *The parameter space* Θ *is compact.*

Assumption 2. $\gamma < 0$ *and* $\forall \theta \in \Theta$, $\sum_{i=1}^{p} \phi_i < 1$.

Assumption 3. Z_t^2 *has a non-degenerate distribution with* $\mathbb{E}Z_t^2 = 1$.

Assumption 4. *If* $p > 0$ *then* $\mathcal{A}_{\theta_0}(z)$ *and* $\mathcal{B}_{\theta_0}(z)$ *have no common root,* $\mathcal{A}_{\theta_0}(z) \neq 0$, *and* $\phi_{0p} + \theta_{0q} \neq 0$.

We are now in a position to state the following consistency theorem.

Theorem 3 (Strong consistency)**.** *Let* $(\hat{\theta}_T)$ *be a sequence of QML estimators satisfying (3.17). Then, under assumptions 1-4*

$$\hat{\theta}_T \to \theta_0, \qquad \textit{almost surely when } T \to \infty. \tag{3.18}$$

Proof. See Francq and Zakoïan [19]. □

Theorem shows that there exists a consistent root of the likelihood equation.

To establish the asymptotic normality we require the following additional assumptions.

Assumption 5. $\theta_0 \in \Theta^c$, *where* Θ^c *denotes the interior of* Θ.

Assumption 6. $\kappa := \mathbb{E}Z_t^4 < \infty$.

Theorem 4 (Asymptotic normality)**.** *Under the assumptions of Theorem 3 and assumptions 5 and 6* $\sqrt{T}(\hat{\theta}_T - \theta_0)$ *converges in distribution to* $N(0, (\kappa - 1)J^{-1})$, *where*

$$J := \mathbb{E}_{\theta_0}\left(\frac{\partial \ell_t(\theta_0)}{\partial \theta \partial \theta'}\right) = \mathbb{E}_{\theta_0}\left(\frac{1}{\sigma_t^4(\theta_0)} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'}\right). \tag{3.19}$$

Proof. See Francq and Zakoïan [19]. □

The results above shows that the quasi likelihood estimate $\hat{\theta}_T$ is $\sqrt{T}$-consistent for the true parameter values and $\sqrt{T}$-asymptotically normal with mean θ_0 and covariance matrix specified before. However, in the presence of non-Gaussian innovations, this estimator can fail to produce asymptotically efficient estimates. Given the results of Theorem 2 and Theorem 3 and mild regularity conditions on the innovation terms we can construct semiparametric estimators which are asymptotically more efficient than the QMLE.

Chapter 4

Multivariate GARCH

Nowadays globalization has resulted in higher international economics integration, investors and also financial institutions are interested in knowing financial markets integration and how financial volatilities together move over time across several markets. Empirical results show that working with separate univariate models is much less relevant than multivariate modelling framework. Cross market effects capturing returns linkage, transmission of stocks and volatility spillover effects are used to indicate markets integration.

Multivariate generalized autoregressive conditional heteroscedasticity (MGARCH) models were initially developed in the late of 1980s and the first half of the 1990s. The most common application of these class of models is to estimate the volatility spillover effects among different markets. When we are in multivariate framework, we are always balancing between two expected difficulties, on one hand as the number of parameters in MGARCH model often increases very quickly with the dimension of the model, the specification of the model should be parsimonious enough to allow for relatively easy estimation of the model and also allow for easy interpretation of the model parameters. On the other hand parsimony means simplification and models with only a few parameters may not be able to capture the relevant dynamics in the covariance structure. Another feature that needs to be taken into account in the specification of the model is imposing positive definiteness (as covariance matrix need, by definition, to be positive definite). One possibility is to derive conditions under which the conditional covariance matrices implied by the model are positive definite, but this is often infeasible in practice. An alternative is to formulate the model in a way that positive definiteness is implied by the structure (in addition to some simple constraints).

We review different specifications of conditional covariance matrices in the following subsections. We distinguish three approaches for constructing multivariate GARCH models

- Generalizations of the univariate GARCH model
- Linear combinations of univariate GARCH models
- Nonlinear combinations of univariate GARCH models

Before we start with the definitions we introduce some basic blocks of multivariate framework concerned the GARCH models. Consider a stochastic vector process X_t with dimension n. Let $\mathcal{F}_t$ be the non decreasing collection of σ-fields generated by past of the series X_t, i.e. $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$.Assume that conditional covariance matrix H_t of X_t is measurable with respect to $\mathcal{F}_{t-1}$. The multivariate GARCH framework is then given by

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

where $H_t = [h_{ij}]_t$ is $n \times n$ symmetric positive definite matrix for all t. $H_t^{1/2}$ may be obtained by Cholesky factorization of H_t and Z_t is a n dimensional i.i.d. vector process with zero mean and unit variance. Hence Z_t is independent of $\mathcal{F}_{t-1}$, it follows that $cov(Z_t|\mathcal{F}_{t-1}) = cov(Z_t) = I_n$. The process X_t is then a n dimensional vector martingale difference sequence

$$\begin{aligned} \mathbb{E}(X_t|\mathcal{F}_{t-1}) &= 0, \\ cov(X_t|\mathcal{F}_{t-1}) &= H_t^{1/2} cov(Z_t|\mathcal{F}_{t-1}) H_t^{1/2} = H_t. \end{aligned} \tag{4.2}$$

The information set $\mathcal{F}_t$ contains both lagged values of the squares and cross-product of X_t and elements of the conditional covariance matrices up to time t. The challenge in multivariate GARCH modelling is to find a parameterization of H_t as a function of $\mathcal{F}_{t-1}$ that is fairly general while feasible in terms of estimation.

4.1 Generalizations of the univariate GARCH models

The extension from a univariate GARCH model to an n-variate model (multivariate) requires considering n-dimensional stochastic process with zero mean random variables X_t and covariance matrix H_t.

4.1.1 VECH model

VECH model proposed by Bollerslev, Engle and Wooldridge [10] is straightforward generalization of the univariate GARCH model. Every conditional variance and covariance is a function of all lagged conditional variances and covariances, as well as lagged squared returns and cross products of returns. The VECH (p,q) model can be defined using Bollerslev interpretation as follows.

Definition 2. *A VECH (p,q) process is a martingale difference sequence X_t relative to a given filtration $\mathcal{F}_t$, whose conditional covariance matrix $H_t = cov(X_t|\mathcal{F}_{t-1})$ satisfy, for every $t \in \mathbb{Z}$*

$$vech(H_t) = c + \sum_{i=1}^{q} A_i vech(X_{t-i}X'_{t-i}) + \sum_{i=1}^{p} G_i vech(H_{t-i}), \tag{4.3}$$

where vech[1](.) is the operator that stacks the lower triangular portion of a symmetric square $n \times n$ matrix into a $(n(n+1)/2)$-dimensional vector, c is an $(n(n+1)/2)$-dimensional vector, and A_i, G_i are square parameter matrices of order $(n(n+1)/2)$.

For illustration we consider bivariate VECH $(1,1)$ model and denote $h_t = vech(H_t)$ then (4.3) becomes

$$\begin{aligned} h_t &= \begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} \\ &= \begin{bmatrix} c_{1,t} \\ c_{2,t} \\ c_{3,t} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} X^2_{1,t-1} \\ X_{1,t-1}X_{2,t-1} \\ X^2_{2,t-1} \end{bmatrix} + \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix}. \end{aligned}$$

Notice that, here we can immediately see equivalency VEC and VECH representation. In VEC[2] representation all covariance equations appear twice, because there is an equation for $h_{ij,t}$ as well as for $h_{ji,t}$. All the off diagonal terms appear twice within each equation (i.e. both of the terms $X_{i,t-1}X_{j,t-1}$ and $X_{j,t-1}X_{i,t-1}$ and both of the terms $h_{ij,t-1}$

[1]In many of the literature about the multivariate GARCH models we can find different notation, VEC (p,q) instead of VECH (p,q). It is because we can use two possible vectorization of covariance matrix H_t. The difference between vec and vech operator is that for a symmetric matrix A, the vector vec (A) contains more information than is strictly necessary, since the matrix is completely determined by the symmetry together with the lower triangular portion, that is, the $(n(n+1)/2)$ entries on and below the main diagonal.

[2]using vec vectorization in definition of VECH model.

and $h_{ji,t-1}$ appear in each equation). These redundant terms we can remove without affecting the model. If we do that, dimensions of our matrices A_i and G_i will be $n(n+1)/2$ instead of n^2 as we mentioned before.

This model is very general, flexible and we can also directly interpret the coefficients, however it brings two major disadvantages in applications. The number of parameters in the model equals $(p+q)(n(n+1)/2)^2 + n(n+1)/2$, which makes this model practicable in practice only in the bivariate case. Second is that there exist only sufficient conditions on the parameters to ensure that conditional variance matrices H_t are positive definite almost surely for all t.

Bollerslev, Engle, and Wooldridge [10] introduced restriction of the model such that, each component of the covariance matrix H_t depends only on its own past and past values of $X_t X_t'$. In other words in the diagonal representation, is assumed that the matrices A_i and G_i are diagonal. This so called diagonal VECH model (DVECH) reduces the number of parameters to $(p+q+1)n(n+1)/2$ and in this case is also possible to obtain conditions for positive definiteness of H_t for all t. However, DVECH representation seems to be too restrictive since no interaction is allowed between the different conditional variances and covariances.

Here we derive a sufficient condition for diagonal VECH model for H_t to be positive definite. Then the diagonal VECH model can be written in matrix representation as follows

$$H_t = \tilde{C} + \tilde{A} \odot X_{t-1}X_{t-1}' + \tilde{G} \odot H_{t-1}, \tag{4.4}$$

where the symbol $\odot$ represents Hadamard[3] product of the two matrices, $\tilde{C}$, $\tilde{A}$ and $\tilde{G}$ are all $n \times n$ parameter matrices. Using Cholesky decomposition of the parameter matrices and from properties of Hadamard product can be seen that usual matrix multiplication will be carried out first, hence $\tilde{A}\tilde{A}' \odot X_{t-1}X_{t-1}'$ should be interpreted as $(\tilde{A}\tilde{A}') \odot (X_{t-1}X_{t-1}')$. Then

$$H_t = \tilde{C}\tilde{C}' + \tilde{A}\tilde{A}' \odot X_{t-1}X_{t-1}' + \tilde{G}\tilde{G}' \odot H_{t-1}, \tag{4.5}$$

since $\tilde{C}\tilde{C}'$, $\tilde{A}\tilde{A}'$ and $\tilde{G}\tilde{G}'$ are all positive semi-definite, H_t will be positive definite for all t as far as the initial covariance matrix H_0 is positive definite. If sample covariance is used for H_0 then H_t will always be positive definite.
So each conditional covariance depends on its own past values. The difference between

[3]The Hadamard product $A \odot B$ of two matrices of the same dimensions is a matrix of the same dimensions with elements given by $(A \odot B)_{ij} = A_{i,j}.B_{i,j}$.

this representation and Bollerslev diagonal VECH representation is that the parameterization used here imposed restrictions implicitly among different parameters to ensure that the parameter matrix is positive semidefinite, and which further assure the conditional covariance matrices are positive definite. By writing the parameter matrices in the form of $\tilde{C}\tilde{C}'$, $\tilde{A}\tilde{A}'$ and $\tilde{G}\tilde{G}'$ instead of just $\tilde{C}$, $\tilde{A}$ and $\tilde{G}$ the positive semi-definiteness is guaranteed in estimation without imposing any further restrictions.

Let define the backshift operator L such that $L^K X_t = X_{t-k}$ and convention that $A(L) = A_1 L + A_2 L^2 + \cdots + A_q L^q$ and $G(L) = G_1 L + G_2 L^2 + \cdots + G_p L^p$. Let Z_t be an n dimensional i.i.d. vector process with mean zero and unit variance. Hence Z_t is independent of $\mathcal{F}_{t-1}$, it follows that $cov(Z_t|\mathcal{F}_{t-1}) = cov(Z_t) = I_n$. There exists a VECH process X_t such that $X_t = H_t^{1/2} Z_t$, where $H_t = cov(X_t|\mathcal{F}_{t-1})$ and $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$. Assuming that X_t is doubly infinite sequence, we can rewrite equation for conditional covariance matrix (4.3) as

$$vech(H_t) = \sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)vech(X_t X_t')]. \tag{4.6}$$

The following computation shows that this parameterization gives indeed VECH model

$$\begin{aligned} vech(H_t) &= c + A(L)vech(X_t X_t') + \sum_{i=2}^{\infty} G(L)^{i-1}[c + A(L)vech(X_t X_t')] \\ &= c + A(L)vech(X_t X_t') + G(L)\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)vech(X_t X_t')] \\ &= c + A(L)vech(X_t X_t') + G(L)vech(H_t). \end{aligned} \tag{4.7}$$

Note that the backshift operator L works also with vech operator such that shifts indices in both X_t's and H_t. We can state the following stationary theorem of Bollerslev, Engle and Wooldridge [10].

Theorem 5. *Let c be an $(n(n+1)/2)$-dimensional vector and A_i, G_i are square parameter matrices of order $(n(n+1)/2)$. Let Z_t be an i.i.d. vector process with mean zero and unit variance. Hence Z_t is independent of $\mathcal{F}_{t-1}$, it follows that $cov(Z_t|\mathcal{F}_{t-1}) = cov(Z_t) = I_n$. Then there exists a covariance stationary VECH process X_t such that $X_t = H_t^{1/2} Z_t$, where $H_t = cov(X_t|\mathcal{F}_{t-1})$ and $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$ if and only if all the eigenvalues of $A(1) + G(1)$ are less than one in modulus.*

Proof. Assuming that X_t is doubly infinite sequence we can use the VECH representation for conditional covariance matrix given in (4.6) and for simplicity we denote

$\eta_t = vech(X_t X_t')$ and $h_t = vech(H_t)$. Then second step is to define $\mathbb{E}_{t-1}$ to be expectations operator, conditioned on the information set $\mathcal{F}_{t-1}$. Then we can compute

$$\begin{aligned}
\mathbb{E}_{t-1}\eta_t &= \sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i}] \\
\mathbb{E}_{t-2}\eta_t &= \mathbb{E}_{t-2}\mathbb{E}_{t-1}\eta_t \\
&= \mathbb{E}_{t-2}\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i}] \\
&= \sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\mathbb{E}_{t-2}\eta_{t-i}] \\
&= c + A(L)\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-1}] + \sum_{i=2}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i}] \\
&= c + A(L)\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-1}] + G(L)\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-1}] \\
&= c + [A(L) + G(L)]\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-1}]
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{t-3}\eta_t &= \mathbb{E}_{t-3}\mathbb{E}_{t-2}\eta_t \\
&= c + [A(L) + G(L)]\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\mathbb{E}_{t-3}\eta_{t-i-1}] \\
&= c + [A(L) + G(L)](c + A(L)h_{t-2}) + [A(L) + G(L)]\sum_{i=2}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-1}] \\
&= c + [A(L) + G(L)]c + [A(L) + G(L)]A(L)\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-2}] \\
&\quad + [A(L) + G(L)]\sum_{i=2}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-1}] \\
&= c + [A(L) + G(L)]c + [A(L) + G(L)]A(L)\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-2}] \\
&\quad + [A(L) + G(L)]G(L)\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-2}] \\
&= c + [A(L) + G(L)]c + [A(L) + G(L)]^2\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-2}] \\
&\ \ \vdots \\
\mathbb{E}_{t-\tau}\eta_t &= [I + (A(L) + G(L)) + \cdots + (A(L) + G(L))^{\tau-2}]c \\
&\quad + [A(L) + G(L)]^{\tau-1}\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)\eta_{t-i-\tau+1}].
\end{aligned}$$

Next we use some knowledge from matrix theory about asymptotic properties of square matrix U. If all the eigenvalues of matrix U are less than one in modulus then $U^\tau \to 0$ as $\tau \to \infty$. Hence the eigenvalues of U are less than one in modulus if and only if $[I + U + U^2 + \dots] \to (I - U)^{-1}$. Therefore $\mathbb{E}_{t-\tau}\eta_t$ converges in probability to $[I - A(1) - G(1)]^{-1}c$ as $\tau \to \infty$ if and only if the eigenvalues of $(A(1) + G(1))$ are less than one in modulus. Also, $\mathbb{E}(X_t X'_{t+\gamma}) = \mathbb{E}[\mathbb{E}(X_t X'_{t+\gamma})] = 0$ for all $\gamma \neq 0$. Then $\mathbb{E}(X_t X'_{t+\gamma})$ exists and depends only on γ for all t. □

For any parameterization necessary and sufficient conditions on the parameters we have to ensure that conditional covariance matrices H_t are positive definite. This can be difficult to check for given parameters, Engle and Kroner [17] propose a new parameterization for H_t that easily imposes these restrictions.

4.1.2 BEKK model

We consider BEKK (p, q, K) model proposed by Baba, Engle, Kraft and Kroner [4] defined as follows.

Definition 3. *A BEKK (p, q) process is a martingale difference sequence X_t relative to a given filtration $\mathcal{F}_t$, whose conditional covariance matrix $H_t = cov(X_t|\mathcal{F}_{t-1})$ satisfy, for every $t \in \mathbb{Z}$*

$$H_t = CC' + \sum_{k=1}^{K}\sum_{i=1}^{q} A_{ik}^{*\prime} X_{t-i} X'_{t-i} A^*_{ik} + \sum_{k=1}^{K}\sum_{i=1}^{p} G_{ik}^{*\prime} H_{t-i} G^*_{ik}, \tag{4.8}$$

where C is a upper triangular $n \times n$ matrix, A^*_{ik} and G^*_{ik} are $n \times n$ parameter matrices and summation limit K determines the generality of the process.

The decomposition of the constant term into a product of two triangular matrices ensures positive definiteness of H_t. A property of BEKK model is that conditional covariance matrices H_t are positive definite by construction. A sufficient condition for positivity is for example that at least one of the matrices C or G^*_{ik} have full rank and the matrices $H_0, \dots, H_{1-p}$ are positive definite.

Let now turn to investigate the relationship between the BEKK and VECH parameterizations. Relationship between this two parameterizations can be found by vectorizing[4]

[4]Recognizing that $vec(ABC) = (C' \otimes A)vec(B)$, where $\otimes$ denote the Kronecker product that is an operation on two matrices of arbitrary size resulting in a block matrix.

of equation (4.8)

$$
\begin{aligned}
vec(H_t) = (C \otimes C)'vec(I_n) + \sum_{k=1}^{K}(A_{1k}^{*} \otimes A_{1k}^{*})'vec(X_{t-1}X_{t-1}') \\
+ \sum_{k=1}^{K}(G_{1k}^{*} \otimes G_{1k}^{*})'vec(H_{t-1}).
\end{aligned} \tag{4.9}
$$

Hence

$$
A_1 = \sum_{k=1}^{K}(A_{1k}^{*} \otimes A_{1k}^{*})' \tag{4.10}
$$

and

$$
G_1 = \sum_{k=1}^{K}(G_{1k}^{*} \otimes G_{1k}^{*})', \tag{4.11}
$$

which leads to the following representation theorem Engle and Kroner [17] that establishes the equivalence of DVEC models that have positive definite covariance matrices and general diagonal BEKK models.

Theorem 6. *The VECH and BEKK parameterizations are equivalent if and only if there exist c, A_{ik}^{*} and G_{ik}^{*} such that*

$$
\begin{aligned}
c &= (C^{*} \otimes C^{*})'vech(I_n), \\
A_i &= \sum_{k=1}^{K}(A_{ik}^{*} \otimes A_{ik}^{*})', \\
G_i &= \sum_{k=1}^{K}(G_{ik}^{*} \otimes G_{ik}^{*})'.
\end{aligned} \tag{4.12}
$$

Proof. Without lost of generality we prove this theorem only for $p = q = 1$. Recognizing that $\eta_t = vech(X_t X_t')$ and $h_t = vech(H_t)$, then the VECH $(1,1)$ becomes

$$
h_t = c + A_1\eta_{t-1} + G_1 h_{t-1} \tag{4.13}
$$

and the BEKK $(1,1)$ becomes

$$
H_t = CC' + \sum_{k=1}^{K} A_{1k}^{*\prime} X_{t-1}X_{t-1}' A_{1k}^{*} + \sum_{k=1}^{K} G_{1k}^{*\prime} H_{t-1} G_{1k}^{*}. \tag{4.14}
$$

Vectorizing equation (4.14) gives

$$\begin{aligned} vech(H_t) &= vech(CC') + vech\left(\sum_{k=1}^{K} A_{1k}^{*'} X_{t-1} X_{t-1}' A_{1k}^{*}\right) + vech\left(\sum_{k=1}^{K} G_{1k}^{*'} H_{t-1} G_{1k}^{*}\right) \\ h_t &= vech(CC') + \sum_{k=1}^{K} (A_{1k}^{*} \otimes A_{1k}^{*})' vech(X_{t-1}X'_{t-1}) + \sum_{k=1}^{K} (G_{1k}^{*} \otimes G_{1k}^{*})' vech(H_{t-1}) \\ &= (C \otimes C)' vech(I_n) + \sum_{k=1}^{K} (A_{1k}^{*} \otimes A_{1k}^{*})' \eta_{t-1} + \sum_{k=1}^{K} (G_{1k}^{*} \otimes G_{1k}^{*})' h_{t-1}. \end{aligned} \tag{4.15}$$

Now if (4.12) hold, then last term in (4.15) becomes

$$h_t = c + A_1 \eta_{t-1} + G_1 h_{t-1},$$

which is exactly (4.13) and then we proved sufficiency. Next step is to show that relations (4.13) and (4.15) hold for all X_{t-1}, proving necessity. So by appropriate choice of X_{t-1}, each column of A_1 can be equated individually with each column of $\sum_{k=1}^{K}(A_{1k}^{*} \otimes A_{1k}^{*})'$. For instance, letting $X_{t-1}' = (1, 0, \ldots, 0)$ establishes equality of the first column of A_1 with the first column of $\sum_{k=1}^{K}(A_{1k}^{*} \otimes A_{1k}^{*})'$. The rest of relations (4.12) can be done in the same way. □

Conclusion of theorem is that each of the BEKK models implies a unique VECH model, which then generates positive definite conditional covariance matrices, while the converse implication is not true. To show that converse implication is not true we simply distinguish that for a given A_1 the choice of A_{1k}^{*} is not unique. This can be seen by recognizing that $(A_{1k}^{*} \otimes A_{1k}^{*}) = (-A_{1k}^{*} \otimes -A_{1k}^{*})$, so while $A_1 = \sum_{k=1}^{K}(A_{ik}^{*} \otimes A_{ik}^{*})'$ is unique the choice of A_{ik}^{*} is not unique. Note that from relations (4.12) is obvious that DVECH is returned from the BEKK parameterization if and only if each of the A_{ik}^{*} and G_{ik}^{*} matrices are diagonal. It can be also shown that the BEKK model eliminates few, if any of the interesting positive definite models permitted by the VECH model. All positive definite DVECH models can be written in the BEKK framework, so that if one restricts the focus to diagonal models, the BEKK model is equally general as the VECH model.

Now we are going to discuss necessary and sufficient conditions for covariance stationarity of the BEKK process. Let L be backshift operator such that $L^K X_t = X_{t-k}$ and convention that $A(L) = \sum_{k=1}^{K}(A_{ik}^{*} \otimes A_{ik}^{*})'L + \cdots + \sum_{k=1}^{K}(A_{qk}^{*} \otimes A_{qk}^{*})'L^q$, $G(L) = \sum_{k=1}^{K}(G_{1k}^{*} \otimes G_{1k}^{*})'L + \cdots + \sum_{k=1}^{K}(G_{pk}^{*} \otimes G_{pk}^{*})'L^p$ and $c = (C^* \otimes C^*)' vech(I_n)$. Let Z_t be an i.i.d. process with mean zero and unit variance. There exists a BEKK process X_t such that $X_t = H_t^{1/2} Z_t$, where $H_t = cov(X_t|\mathcal{F}_{t-1})$ and $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$. Assuming that X_t is doubly infinite sequence we can rewrite equation for conditional covariance matrix (4.8) using vectorization as

$$vech(H_t) = \sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)vech(X_t X_t')]. \tag{4.16}$$

By the Theorem 6 this parameterization gives BEKK model because

$$\begin{aligned} vech(H_t) &= c + A(L)vech(X_t X_t') + \sum_{i=2}^{\infty} G(L)^{i-1}[c + A(L)vech(X_t X_t')] \\ &= c + A(L)vech(X_t X_t') + G(L)\sum_{i=1}^{\infty} G(L)^{i-1}[c + A(L)vech(X_t X_t')] \\ &= c + A(L)vech(X_t X_t') + G(L)vech(H_t). \end{aligned} \tag{4.17}$$

Notice that the parameterization in (4.16) nests both the VECH and the BEKK models.

Theorem 7. *Let C is a upper triangular $n \times n$ matrix and A_{ik}^*, G_{ik}^* are $n \times n$ parameter matrices. Let Z_t be an i.i.d. process with mean zero and unit variance. Hence Z_t is independent of $\mathcal{F}_{t-1}$, it follows that $cov(Z_t|\mathcal{F}_{t-1}) = cov(Z_t) = I_n$. There exists a covariance stationary BEKK process X_t such that $X_t = H_t^{1/2} Z_t$, where $H_t = cov(X_t|\mathcal{F}_{t-1})$ and $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$ if and only if all the eigenvalues of $A(1) + G(1)$ are less than one in modulus.*

Proof. The proof for the BEKK model is analogous as in the VECH model, except that we substitute relations (4.12) into proof. □

So BEKK model is covariance stationary if and only if all the eigenvalues of $\sum_{i=1}^{q}\sum_{k=1}^{K}(A_{ik}^* \otimes A_{ik}^*) + \sum_{i=1}^{p}\sum_{k=1}^{K}(G_{ik}^* \otimes G_{ik}^*)$ are less than one in modulus. Then the unconditional covariance matrix, when it exists, is given for $K = 1$

$$\mathbb{E}(vech(X_t X_t')) = [I - (A_{11}^* \otimes A_{11}^*) - (G_{11}^* \otimes G_{11}^*)]^{-1} vech(C'C). \tag{4.18}$$

Estimation of multivariate GARCH models is troublesome, since the number of parameters may be large also for relative small vector dimension n. Let us assume through this chapter that Z_t are i.i.d. $N(0, I_n)$. The conditional covariance matrices H_t are modelled as (4.8). Let p be a density function, θ be the vector of parameters that are needed to parameterize.

Suppose that there is an underlying data generating process characterized by the unknown parameter vector θ_0 which one wants to estimate using a given sample of T

observations. Hence the joint distribution of $(X_1, X_2, \ldots, X_T)$ where T is the number of observations, need not to be multivariate normally distributed. But the joint density is the product of all the conditional densities, so the log-likelihood function of the joint distribution is the sum of all the log-likelihood functions of the conditional distributions. Thus under the assumption that Z_t are i.i.d. conditionally on initial values, the quasi log-likelihood function is given by

$$L_t(\theta) = \sum_{t=1}^{T} \ell_t(\theta) \tag{4.19}$$

where

$$\ell_t(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|H_t| - \frac{1}{2}X_t' H_t^{-1} X_t. \tag{4.20}$$

The quasi maximum likelihood estimator of θ is defined as any measurable solution $\hat{\theta}_n$ that maximizes the likelihood function with respect to these parameters. A reasonable set of assumptions on initial conditions is that all presentable data has been fixed at their unconditional expectation. For example $X_0 X_0'$ is assumed to equal its unconditional expectation, given in (4.18). Note that because no reference is made to the functional form chosen for the conditional covariance matrix, we may apply the result of this section whether the VECH or BEKK parameterization is chosen. In either case, however, the models are large and complex, leading one to question how flat the likelihood function is with respect to many of the parameters in the model such as the diagonal model or the BEKK model with $K = 1$ and then use Lagrange multiplier test to examine the validity of the restriction.

Statistical properties of multivariate GARCH models are only partially known. For development of statistical estimation, it would be desirable to have conditions for strict stationarity and ergodicity of a multivariate GARCH processes, as well as conditions for consistency and asymptotic normality of quasi maximum likelihood estimator.

Comte and Lieberman [14] study asymptotic properties of the quasi[5] maximum likelihood estimator. They provide conditions for strong consistency and asymptotic normality of the quasi maximum likelihood estimator $\hat{\theta}$. In addiction they give, through survey of asymptotic results published so far for univariate as well as multivariate GARCH processes. Let us state for completeness two of their main theorems for strong consistency and asymptotic normality of the QML estimator. For the proofs and further details refer to

[5] Note that in Comte and Lieberman [14] $\hat{\theta}$ is presented as quasi MLE, since they do not assume that Z_t's are Gaussian, but work with the Gaussian log-likelihood function.

Comte and Lieberman [14].
Consider the BEKK (p, q) model as defined by (4.8) then we can write the following two theorems.

Theorem 8 (Consistency of quasi MLE). *For the MGARCH (p, q) process defined by (4.8) with $Z_t \sim i.i.d.(0, I_n)$ and for $\hat{\theta}_T$, the quasi maximum likelihood estimate obtained from a sample of length T, and the true parameter $\theta_0 \in \Theta$, assume that*

- *Θ is compact, C, A_i and G_i, are continuous functions of θ, and there exists a $c > 0$ such that $\inf_{\theta \in \Theta} \det(C(\theta)) \geq c > 0$,*
- *model is identifiable in the sence of Engle and Kroner [17],*
- *rescaled errors Z_t admit a density absolutely continuous with respect to the Lebesgue measure and positive in a neighbourhood of the origin,*
- *for all $\theta \in \Theta$, $\rho(\sum_{i=1}^{q} A_i(\theta) + \sum_{i=1}^{p} G_i(\theta)) < 1$. Where ρ returns the largest modules of the eigenvalues.*

Then $\hat{\theta}_T$ is strongly consistent that is, $\hat{\theta}_T \to \theta_0$ almost surely for $T \to \infty$.

Theorem 9 (Asymptotic Normality of quasi MLE). *Under the assumptions*

- *assumptions from Theorem 8 and C, A_i G_i, admit continuous derivatives up to order 3 on Θ,*
- *components of Z_t are independent,*
- *X_t admits bounded moments of order 8,*
- *the initial states of the process H_t are fixed stationary.*

Then the quasi MLE $\hat{\theta}_T$ given the initial state is strongly consistent and

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \to N(0, C_1^1 C_0 C_1^1), \tag{4.21}$$

where $C_1 = \mathbb{E}\left(\left(\frac{\partial^2 \ell_t(\theta_0)}{\partial \theta_i \theta_j}\right)_{1 \geq i,j \geq r}\right)$, $C_0 = \mathbb{E}\left(\frac{\partial \ell_t(\theta_0)}{\partial \theta} \frac{\partial \ell_t(\theta_0)'}{\partial \theta}\right)$ and r is the length of the parameter vector θ.

The BHHH (Berndt Hall and Hall and Hausman) iterative algorithm is useful in practice to obtain the optimal values of parameters by utilizing the following equation proposed by Engle and Kroner [17]

$$\theta^{(i+1)} = \theta^{(i)} + \lambda_i \left(\left(\frac{\partial \ell_t}{\partial \theta}\right)' \frac{\partial \ell_t}{\partial \theta}\right)^{-1} \left(\frac{\partial \ell_t}{\partial \theta}\right)' \tag{4.22}$$

where $\theta^{(i)}$ denotes the parameter estimate after the i^{th} iteration; $\frac{\partial \ell_t}{\partial \theta}$ is evaluated at $\theta^{(i)}$ and λ is a variable step length chosen to maximize the likelihood function in the given direction, which is calculated from a least squares regression of a $T \times 1$ vector of ones on $\frac{\partial \ell_t}{\partial \theta}$.

4.2 Linear combinations of univariate GARCH models

In this section we introduce somewhat different approach of multivariate GARCH models. One can assume that the observed data can be linearly transformed into a set of components by means of an matrix i.e. model tries to express multivariate GARCH by means of univariate GARCH models. This approach has been proposed initially by Alexander and Chibumba [2] and called the Orthogonal GARCH (O-GARCH). Clearly one of the restrictions imposed by the O-GARCH model is that, it requires the matrix that is assumed to link the components with the observed variables, to be orthogonal. This restriction has great computational properties and analytical tractability, so that O-GARCH models have found many applications in finance. However orthogonal matrices are very special and they only reflect a very small subset of all possible invertible linear maps. The generalized O-GARCH model allows the linkage to be given by any possible invertible matrix and was proposed by van der Weide [31], as a generalization of the orthogonal GARCH model.

The O-GARCH model is also known to suffer from identification problem, mainly because estimation of the matrix is entirely based on unconditional information (the sample covariance matrix). For example, when the data exhibits weak dependence, the model has substantial difficulties to identify a matrix that is truly orthogonal.

4.2.1 O-GARCH model

Consider a vector process X_t representing n different returns. Letting $\mathcal{F}_t$ denote the filtration generated by X_t and denote V_t conditional covariance matrix of X_t such that $V_t = cov(X_t|\mathcal{F}_{t-1})$. The data are commonly normalized, so that every series has unit sample variance and zero mean. The vector process X_t can be represented as linear combination of n uncorrelated univariate GARCH processes Y_t with unconditional variances of one.

Definition 4. *The O-GARCH (p, q) process is a vector process X_t defined as*

$$X_t = MY_t, \tag{4.23}$$

where M is a $n \times n$ orthogonal[6] matrix and Y_t is n vector process with the components y_{it} which satisfy

$$\mathbb{E}(y_{it}|\mathcal{F}_{t-1}) = 0, \quad var(y_{it}|\mathcal{F}_{t-1}) = h_{it}, \quad cov(y_{it}, y_{jt}|\mathcal{F}_{t-1}) = 0, \quad i \neq j = 1, \ldots, n, \tag{4.24}$$

such that the components of Y_t are conditionally uncorrelated and each component is modelled as a univariate GARCH process

$$\begin{aligned} y_{it}|\mathcal{F}_{t-1} \sim & N(0, h_{it}), \\ h_{it} = \alpha_i + \sum_{j=1}^{q} \theta_{ji} y_{ji,t-1}^2 + & \sum_{j=1}^{p} \phi_{ji} h_{ji,t-1} \qquad for \quad i = 1, \ldots, n. \end{aligned} \tag{4.25}$$

The conditional covariances of X_t are given by

$$V_t = MH_tM', \qquad H_t = diag(h_{1t}, \ldots, h_{nt}). \tag{4.26}$$

We assume that Y_t and hence X_t is covariance stationary, such that the unconditional variances $H = var(Y_t)$ and $V = var(X_t) = MHM'$ exist. The parameters for O-GARCH (p, q) model are all ϕ_{ji}, all θ_{ji}, M and V. The number of parameters to be estimated in this model is equal $(p + q)n(n + 5)/2$.

Denote P the orthogonal matrix of eigenvectors of V, and Λ the diagonal matrix containing the corresponding eigenvalues, such that $V = P\Lambda P'$. Then Y_t satisfies $H = var(Y_t) = PVP' = \Lambda$, such that the components of Y_t are unconditionally uncorrelated. This property is then amplified by assumption that the Y_t are conditionally uncorrelated and then H_t is diagonal. After we have estimated all the parameters, the conditional covariance matrix of the original series is simply

$$V_t = \mathbb{E}_{t-1} X_t X_t' = \mathbb{E}_{t-1} M Y_t Y_t' M' = P H_t P'. \tag{4.27}$$

The advantage of the model is that only a few principle components are enough to explain most of variability in the system which suggest this model applicable in large dimension models. However when the data exhibits weak dependence, the O-GARCH model is not always able to identify the orthogonal matrix M which lead to the development of more general model.

[6] We follow Alexander [1], however there exist other possibilities for instance Vrontos et al. [32] restrict M to be lower triangular, which is not without loss of generality.

4.2.2 GO-GARCH model

Generalized O-GARCH model was proposed by van der Weide [31] and in this section we closely follow his work. In the GO-GARCH model the components of X_t do not have to be standardized as in the O-GARCH model. The starting point of the GO-GARCH model is the assumption defined as follows.

Assumption 7. *The observed process X_t is defined by a linear combination of conditionally uncorrelated components Y_t*

$$X_t = MY_t \tag{4.28}$$

where Y_t is a n vector process with the components y_{it} of which satisfy

$$\mathbb{E}(y_{it}|\mathcal{F}_{t-1}) = 0, \quad var(y_{it}|\mathcal{F}_{t-1}) = h_{it}, \quad cov(y_{it}, y_{jt}|\mathcal{F}_{t-1}) = 0, \quad i \neq j = 1, \ldots, n. \tag{4.29}$$

The linear map M that links the unobserved components with the observed variables is assumed to be constant over time, and invertible.

Hence the GO-GARCH model can be defined as follows.

Definition 5. *The GO-GARCH (p, q) process is a vector process X_t defined as*

$$X_t = MY_t. \tag{4.30}$$

Where each of the component process y_{it} is modelled as a univariate GARCH process and then

$$\begin{aligned} y_{it}|\mathcal{F}_{t-1} \sim & N(0, h_{it}), \\ h_{it} = \alpha_i + \sum_{j=1}^{q} \theta_{ji} y_{ji,t-1}^2 + \sum_{j=1}^{p} \phi_{ji} h_{ji,t-1} & \qquad for \quad i = 1, \ldots, n. \end{aligned} \tag{4.31}$$

Hence, the conditional covariances of X_t see van der Weide [31] are given by

$$V_t = MH_tM', \qquad H_t = diag(h_{1t}, \ldots, h_{nt}). \tag{4.32}$$

Note that we impose without lost of generality that each of the unobserved components y_{it} have unit variance so that $V = MM'$.
If we consider the singular value decomposition of M

$$M = P\Lambda^{1/2}U' \tag{4.33}$$

where P and Λ denote the matrices with, respectively, the orthogonal eigenvectors and the eigenvalues of $V = MM'$, then U is the orthogonal matrix of eigenvectors of MM'.

The matrices P and Λ will be estimated directly by means of unconditional information, as they will be extracted from the sample covariance matrix V, the main task for inference on the loading matrix M is to identify the orthogonal matrix U. The O-GARCH model corresponds then to the particular choice $U = I_n$. Van der Weide express U as the product of $n(n-1)/2$ rotation matrices

$$U = \prod_{i<j} G_{ij}(\delta_{ij}), \qquad -\pi \leq \delta_{ij} \leq \pi, \qquad i, j = 1, 2, \ldots, n, \tag{4.34}$$

where $G_{ij}(\delta_{ij})$ performs a rotation in the plane spanned by the ith and jth vectors of the canonical basis of R over an angle δ_{ij}. For example in the trivariate case

$$G_{12} = \begin{pmatrix} \cos\delta_{12} & \sin\delta_{12} & 0 \\ -\sin\delta_{12} & \cos\delta_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix}, G_{13} = \begin{pmatrix} \cos\delta_{13} & 0 & -\sin\delta_{13} \\ 0 & 1 & 0 \\ -\sin\delta_{12} & 0 & \cos\delta_{13} \end{pmatrix}. \tag{4.35}$$

As rotation angles are most often used the Euler angles which can be estimated by means of maximum likelihood. It is obvious that the (G)O-GARCH model is covariance stationary if the n univariate GARCH processes are themselves stationary.

GO-GARCH model has property that it can be nested as a more general BEKK model. To keep it simple we focus on the GO-GARCH $(1,1)$ model only, but it can be verified that the results also hold for the more general GO-GARCH (p,q) model. Consider the following BEKK representation

$$V_t = C + \sum_{i=1}^{n} A_i X_{t-1} X'_{t-1} A'_i + B V_{t-1} B', \tag{4.36}$$

where C is a positive definite $n \times n$ matrix[7], A_i and B are $n \times n$ matrices. The following theorem show relationship between models.

Theorem 10. *Let the matrices $\{A_i\}_{i=1}^m$ and B be restricted to have identical eigenvector matrix M, where the eigenvalues of A_i are all zero except for the ith one. Moreover, assume that C can be decomposed as MD_CM', where D_C is some positive definite diagonal matrix. Then the associated BEKK parameterization, given in (4.36) is a GO-GARCH process with GARCH $(1,1)$ component where the M reflects the linkage between the conditionally uncorrelated components and the observed process.*

Proof. The matrices $\{A_i\}_{i=1}^n$ and B are assumed to have identical eigenvector matrix M. So they can be diagonalized as follows

$$A_i = MD_{A_i}M^{-1} \qquad \textit{and} \qquad B = MD_BM^{-1}, \tag{4.37}$$

[7]In order to guarantee positive definiteness of V_t for all t.

where $\{D_{A_i}\}$ and D_B denote diagonal eigenvalue matrices. Note that all element of the matrix D_{A_i} are zero except for its i'th diagonal element, which represents the only non-zero eigenvalue of A_i and will be denoted as a_i. By substitution we have

$$V_t = MD_CM' + \sum_{i=1}^{n} MD_{A_i}M^{-1}X_{t-1}X'_{t-1}M^{-1\prime}D_{A_i}M' + MD_BM^{-1}V_{t-1}(M^{-1})'D_BM',$$
$$V_t = M(D_C + \sum_{i=1}^{n} D_{A_i}M^{-1}X_{t-1}X'_{t-1}(M^{-1})'D_{A_i} + D_BM^{-1}V_{t-1}(M^{-1})'D_B)M'. \tag{4.38}$$

By definition we have $X_t = MY_t$. Then $Y_t = M^{-1}X_t$ represent the unobserved components in the GO-GARCH framework. Let $H_t = M^{-1}V_tM^{-1}$ denote the conditional covariance matrix of Y_t. Rearranging terms in (4.38) we can find that

$$H_t = D_C + \sum_{i=1}^{n} D_{A_i}Y_{t-1}Y'_{t-1}D_{A_i} + D_BH_{t-1}D_B. \tag{4.39}$$

By the properties of the matrices $\{D_{A_i}\}$ it follows that the sum can be rewritten using Hadamard product as

$$\sum_{i=1}^{n} D_{A_i}Y_{t-1}Y'_{t-1}D_{A_i} = D_A \odot Y_{t-1}Y'_{t-1}, \tag{4.40}$$

where $D_A = diag\{a_1, \ldots, a_n\}$. Then D_C, D_B and $D_A \odot Y_{t-1}Y'_{t-1}$ are all diagonal, and the conditional covariance matrix of Y_t, denoted by H_t, is also diagonal. Therefore, equation (4.39) implies univariate GARCH $(1,1)$ specifications for the components of Y_t, as it is assumed by the GO-GARCH model. □

The parameter estimation of the GO-GARCH model is carried out as usual with maximum likelihood estimation. We show before that GO-GARCH can be nested as more general BEKK model, so that most of the theory of maximum likelihood estimation available for the BEKK models can be applied for GO-GARCH models. The parameters that need to be estimated include the vector θ of rotation coefficients that will identify the invertible matrix M, and the parameters (ϕ_i, θ_i) for the n univariate GARCH models. The quasi log-likelihood function $L_t(\theta)$ for the GO-GARCH with a given sample of T observations is given by

$$\begin{aligned} L_t(\theta) &= -\frac{1}{2}\sum_{t=1}^{T} n\log(2\pi) + \log|V_t| + X'_tV_t^{-1}X_t \\ &= -\frac{1}{2}\sum_{t=1}^{T} n\log(2\pi) + \log|Z_\theta H_tM_\theta| + Y'_tM'_\theta(M_\theta H_tM_\theta)^{-1}Y_tM_\theta \\ &= -\frac{1}{2}\sum_{t=1}^{T} n\log(2\pi) + \log|M_\theta M'_\theta| + \log|H_t| + Y'_tH_t^{-1}Y_t, \end{aligned} \tag{4.41}$$

where $M_\theta M_\theta' = P\Lambda P'$ is independent of θ. Even in high-variate cases, when the covariance matrices are very large, it should not be a problem to maximize the quasi log-likelihood function over the $n(n-1)/2 + 2n$ parameters. However practical power of GO-GARCH model lies in its two-step estimation procedure. For this method it is necessary that link matrix M is orthogonal. In first step, matrices P and Λ are estimated directly by means of unconditional information as they will be extracted from sample covariance matrix V_t. This involves only solving an eigenvalue problem. In the second step the conditional information is used to estimate rotation coefficients of U and all θ_i and ϕ_i of n factors. This separation shows that a two-step estimation procedure is feasible and that variances and correlations can be estimated separately. The two-step approach mainly has the advantage that the dimensionality of the maximization problem is reduced accelerating the maximization process.

The problems of maximizing the multivariate likelihood function for high dimensions lead to development of three-step procedure. This method was proposed by Boswijk and van der Weide [12]. First step is the same as before and the second step of the two-step procedure is divided into two steps. This allows to separate the estimation of a part of link matrix U from univariate GARCH parameters. The tree-step procedure tries to identify U from the autocorrelation structure of $s_t s_t'$ where $s_t = \Lambda^{-1/2} P' X_t$. They obtain estimate for $B = UA'U$ by regressing the following model

$$s_t s_t' - I_m = B(s_{t-1}s_{t-1}' - I_m)B + \Gamma_t, \qquad \mathbb{E}(\Gamma_t) = 0, \tag{4.42}$$

using non-linear least squares method. Estimate for U may be obtained as the eigenvector matrix of B as A is diagonal matrix. The three-step procedure is not only more practical in terms of implementation but also is less prone to convergence problems. However the main disadvantage is loss of efficiency. The two-step as well as the three-step procedure seem to be too slow when dimension of the model is high. Broda and Paolella [13] introduced a new two-step procedure for estimation of GO-GARCH model. They use independent component analysis as the main tool for the decomposition of a high-dimensional problem into a set of univariate models. The algorithm maximizes the conditional heteroscedasticity of the estimated components. Their method is called CHICAGO (Conditionally Heteroscedastic Independent Component Analysis of Generalized Orthogonal GARCH models). Their procedure allows them to apply non-Gaussian innovations. For more details see Broda and Paolella [13].

As we discussed before in Theorem 10, the GO-GARCH model is a special case of the

BEKK model of Engle and Kroner [17], and as such the general results of Comte and Lieberman [14] concerning consistency and asymptotic normality of maximum likelihood estimators can be directly applied. Conditions for strong consistency of the maximum likelihood estimator for BEKK model are given by Theorem 8 and conditions for the asymptotic normality are given in Theorem 9 and then strong consistency and asymptotic normality of the quasi MLE for GO-GARCH can therefore be established by appealing this conditions. For initial value we choose the unconditional covariance matrix.

4.3 Nonlinear combinations of univariate GARCH models

This section collects such models that may be viewed as nonlinear combinations of univariate GARCH models. The models in this category are based on the idea of modelling the conditional variances and correlations instead of straight forward modelling the conditional covariance matrix. In most of the literature about multivariate GARCH models these models can be found as models of conditional variances and correlations. This class of models includes Constant Conditional Correlation Model (CCC, Bollerslev [8]), and Dynamic Conditional Correlation Models (DCC models of Tse and Tsui [29], and Engle [16]).

4.3.1 CCC model

The conditional correlation matrix in this class of models is time invariant. Conditional covariance matrix thus can be specified in a hierarchical way. First, one chooses a GARCH-type model for each conditional variance. Second, based on the conditional variances, one models the conditional correlation matrix (imposing its positive definiteness $\forall t$). Since conditional correlation matrix is time invariant, the conditional covariances are proportional to the product of the corresponding conditional standard deviations. Let us formalize our assertions.

Definition 6. *The CCC (p, q) process is a martingale difference sequence X_t relative to a given filtration $\mathcal{F}_t$, whose conditional covariance matrix $H_t = cov(X_t|\mathcal{F}_{t-1})$ satisfy*

$$H_t = D_t R D_t = (\rho_{ij}\sqrt{h_{iit}h_{jjt}}), \tag{4.43}$$

where

$$D_t = diag(h_{11t}^{1/2} \dots h_{nnt}^{1/2}) \tag{4.44}$$

and

$$R = (\rho_{ij}) \tag{4.45}$$

is a symmetric positive definite matrix with $\rho_{ii} = 1, \forall i$. *Then the off-diagonal elements of the conditional covariance matrix are defined as* $[H_t]_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij}$ *for* $i \neq j$, $1 \leq i, j \leq n$. h_{iit} *is defined as univariate GARCH*(p, q) *model*[8]

$$h_t = c + \sum_{i=1}^{q} A_i X_{t-i}^2 + \sum_{i=1}^{p} G_i h_{t-i}, \tag{4.46}$$

where c *is* $n \times 1$ *vector,* A_i *and* G_i *are diagonal* $n \times n$ *matrices.*

Time invariant $n \times n$ symmetric matrix R with unit diagonal elements, containing the constant conditional correlations ρ_{ij}. If the elements of c and the diagonal elements of A_i and G_i are positive, and the conditional correlation matrix R is positive definite, then the conditional covariance matrix H_t is positive definite. Positivity of the diagonal elements of A_i and G_i is not however necessary for R to be positive definite unless $p = q = 1$. This CCC model contains $n(n+5)/2$ parameters.

The CCC model was first introduced by Bollerslev [8]. Although the CCC model because of simplicity and attractive parameterization has been very popular in practice, empirical studies have suggested that the assumption of constant conditional correlations, and thus the conditional covariances may be too restrictive and unrealistic. A sufficient condition for strict stationarity and the existence of fourth-order moment of the CCC (p, q) is established in Aue, Hormann, Horvath, and Reimherr [3]. Existence of stationary solution is the key ingredient for estimation, so we first state necessary and sufficient conditions for strict stationarity of the model.

Technique is very similar as in the univariate GARCH models so we state only consequential points. We can write

$$X_t = D_t \eta_t, \qquad \eta_t = R^{1/2} Z_t, \tag{4.47}$$

then we define matrix Ω as follows

$$\Omega_t = \begin{pmatrix} \eta_{1t}^2 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & \\ 0 & \dots & & \eta_{nt}^2 \end{pmatrix}.$$

[8]Can be defined as any univariate GARCH model, because of simplicity we choose the simplest one.

And let define the $(p+q)n \times (p+q)n$ matrix

$$C_t = \begin{pmatrix} \Omega_t A_1 & \Omega_t A_2 & \cdots & \Omega_t A_q & \Omega_t G_1 & \Omega_t G_2 & \dots & \Omega_t G_p \\ I_n & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & I_n & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & I_n & \dots & 0 & 0 & \dots & 0 \\ A_1 & A_2 & \dots & A_q & G_1 & G_2 & \dots & G_p \\ 0 & 0 & \dots & 0 & I_n & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & I_n & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & I_n & 0 \end{pmatrix}.$$

Theorem 11. *Let c_i be an n-dimensional vector R be a time invariant $n \times n$ symmetric matrix and let A_{ij}, G_{ij}, are square diagonal matrices of order n. Let Z_t be an i.i.d. vector process with mean zero and unit variance. Hence Z_t is independent of $\mathcal{F}_{t-1}$, it follows that $cov(Z_t|\mathcal{F}_{t-1}) = cov(Z_t) = I_n$. There exists stationary CCC process X_t such that $X_t = H_t^{1/2} Z_t$, where $H_t = cov(X_t|\mathcal{F}_{t-1})$ and $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$ if and only if $\gamma(C_0) < 0$, where $\gamma(C_0)$ is the top Lyapunov exponent of the sequence $C_0 = \{C_t, t \in \mathbb{Z}\}$. This stationary solution, when $\gamma(C_0) < 0$, is unique and ergodic.*

Proof. The proof similar to that given for univariate GARCH (p, q) models. Existence of top Lyapunov exponent γ is guaranteed by the condition $\mathbb{E}(\log^+ \|C_0\|) < \infty$. Then for $\gamma(C_0) < 0$ the series

$$Y_t = b + \sum_{k=1}^{\infty} C_t C_{t-1} \dots C_{t-k+1} b \tag{4.48}$$

converges almost surely for all t. A strictly stationary and ergodic solution is obtained as $X_t = \{diag(Y_{q+1,t})\}^{1/2} R^{1/2} Z_t$ where $Y_{q+1,t}$ denotes the $(q+1)$th subvector of size n of Y_t. The proof of the uniqueness is exactly the same as in the univariate case.

Now we show the necessity of the top Lyapounov exponent being negative. It suffices to show that

$$\lim_{k\to\infty} C_0 C_{-1} \dots C_{-k+1} e_i \to 0, \qquad a.s. \qquad for \quad 1 \le i \le p+q. \tag{4.49}$$

Existence of a strictly stationary solution implies as $k \to \infty$

$$C_0 C_{-1} \dots C_{-k+1} b \to 0 \quad a.s. \tag{4.50}$$

Using the relation $b = e_1\Omega_{-k}c + e_{q+1}c$ we get

$$\lim_{k\to\infty} C_0 \dots C_{-k+1}e_1\Omega_{-k}c = 0, \qquad \lim_{k\to\infty} C_0 \dots C_{-k+1}e_{q+1}c = 0, \quad a.s.$$

Since components of c are strictly positive (4.49) thus holds for $i = q + 1$. Using

$$C_{-k+1}e_{q+i} = \Omega_{-k+1}G_i e_1 + G_i e_{q+1} + e_{q+i+1}, \qquad i = 1, \dots, p \tag{4.51}$$

with convention $e_{p+q+1} = 0$, for $i = 1$ we obtain

$$0 = \lim_{t\to\infty} C_0 \dots C_{-k+1}e_{q+1} \geq \lim_{k\to\infty} C_0 \dots C_{-k+1}e_{q+2} \geq 0.$$

Therefore (4.49) holds for $i = q+2$ and by induction, for $i = q+j$, $j = 1, \dots, p$. Moreover, $C_{-k+1}e_q = \Omega_{-k+1}A_q e_1 + A_q e_{q+1}$ so (4.49) holds for $i = q$ and thus we can conclude that for other values of i (4.49) holds using recursion. □

Estimation of the CCC models is as usual carried out using maximum likelihood estimator. Let $(X_1, \dots, X_T)$ be an observation of length T of the unique and strictly stationary solution X_t of model (4.43). Conditionally on initial values we can write quasi likelihood function as

$$L(\theta) = L(\theta; X_1, \dots, X_T) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{n/2}} |H_t(\theta)|^{1/2} \exp\left(-\frac{1}{2} X_t' H_t^{-1}(\theta) X_t\right), \tag{4.52}$$

and the corresponding quasi log-likelihood

$$L_t(\theta) = \frac{1}{t} \sum_{t=1}^{T} \ell_t, \tag{4.53}$$

where

$$\ell_t = \log |H_t(\theta)| + X_t' H_t^{-1}(\theta) X_t. \tag{4.54}$$

A QML estimator of θ is defined as any measurable solution $\hat{\theta}_t$ of

$$\hat{\theta}_t = arg \max_{\theta\in\Theta} L_t(\theta). \tag{4.55}$$

Asymptotic properties of QMLE were developed by Francq and Zakoïan [20]. They proved $\sqrt{n}$-consystency and $\sqrt{n}$-normality under the similar assumptions as they introduced in the univariate case.

The following assumptions will be used to establish the strong consistency of the QMLE. Assume that Z_t is i.i.d. and the QMLE $\hat{\theta}_t$ maximizes the quasi log-likelihood under Θ. Let $\mathcal{A}_{\theta_0}(z) = \sum_{i=1}^{q} \theta_i z^i$ and $\mathcal{B}_{\theta_0}(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ with the convention $\mathcal{A}_{\theta_0}(z) = 0$ if $q = 0$ and $\mathcal{B}_{\theta_0}(z) = 1$ if $p = 1$. To show strong consistency the following assumptions will be made.

Assumption 8. *$\theta_0 \in \Theta$ and Θ is compact.*

Assumption 9. *$\gamma(C_0) < 0$ and $\forall \theta \in \Theta$, $|\mathcal{B}_\theta(z)| = 0 \Rightarrow |z| > 1$.*

Assumption 10. *The components of Z_t are independent and their squares have non degenerate distributions.*

Assumption 11. *If $p > 0$, then $\mathcal{A}_{\theta_0}(z)$ and $\mathcal{B}_{\theta_0}(z)$ are left coprime and $M_1(\mathcal{A}_{\theta_0}, \mathcal{B}_{\theta_0})$ has full rank n.*

Assumption 12. *R is a positive definite correlation matrix for all $\theta \in \Theta$.*

Theorem 12 (Strong consistency)**.** *Let $\hat{\theta}_t$ be a sequence of QML estimators satisfying (4.55). Then, under assumptions 8-12*

$$\hat{\theta}_t \to \theta_0, \qquad \textit{almost surely when } n \to \infty. \tag{4.56}$$

Proof. See Francq and Zakoïan [20]. □

Theorem shows that there exist a consistent root of the likelihood equation.

To establish the asymptotic normality, we require the following additional assumptions.

Assumption 13. *$\theta_0 \in \Theta^c$, where Θ^c denotes the interior of Θ.*

Assumption 14. $\mathbb{E}\|\eta_t \eta_t'\|^2 < \infty$.

Theorem 13 (Asymptotic normality)**.** *Under the assumptions of Theorem 12 and assumptions 13 and 14 $\sqrt{n}(\hat{\theta}_t - \theta_0)$ converges in distribution to $N(0, J^{-1}IJ^{-1})$, where J is a positive-definite matrix and I is a semi positive-definite matrix, defined by*

$$I = \mathbb{E}_{\theta_0}\left(\frac{\partial \ell_t(\theta_0)}{\partial \theta}\frac{\partial \ell_t(\theta_0)}{\partial \theta'}\right), \qquad J = \mathbb{E}_{\theta_0}\left(\frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'}\right). \tag{4.57}$$

Proof. See Francq and Zakoïan [20]. □

Note that when $n = 1$, results becomes to the univariate setting. In particular, no assumption is made concerning the existence of moments of the observed process.

4.3.2 DCC model

A new class of multivariate models called dynamic conditional correlation (DCC) model was proposed by Engle and they are the generalization of the CCC model by making the conditional correlation matrix time-dependent. These models are flexible like univariate GARCH and parsimonious parametric models for the correlations.

Definition 7. *The Dynamic Conditional Correlation (DCC) process of Engle [2002] is a martingale difference sequence X_t relative to a given filtration $\mathcal{F}_t$, whose conditional covariance matrix $H_t = cov(X_t|\mathcal{F}_{t-1})$ satisfy*

$$H_t = D_t R_t D_t, \tag{4.58}$$

where

$$D_t = diag(h_{1t}^{1/2} \dots h_{nt}^{1/2}) \tag{4.59}$$

and R_t is $n \times n$ time varying correlation matrix of X_t. h_{it} is defined as univariate GARCH(p, q) model[9]

$$h_{it} = c_i + \sum_{j=1}^{q_i} \theta_{ij} X_{t-j}^2 + \sum_{j=1}^{p_i} \phi_{ij} h_{t-j}, \tag{4.60}$$

where c_i, θ_{ij} and ϕ_{ij} are nonnegative parameters for $i = 1, \dots, n$, with the usual GARCH restrictions for non-negativity and stationarity being imposed, such as non-negativity of variances and $\sum_{j=1}^{p_i} \phi_{ij} + \sum_{j=1}^{q_i} \theta_{ij} < 1$.

Note that the univariate GARCH models can have different orders. The number of parameters to be estimated equals $(n + 1)(n + 4)/2$ in bivariate case and is quite large when the n is large. There exists different forms of R_t. When specifying a form of R_t, two requirements have to be considered. First H_t has to be positive definite, because it is a covariance matrix. To ensure H_t to be positive definite, R_t has to be positive definite (D_t is positive definite since all the diagonal elements are positive). Second is that all the elements in the correlation matrix R_t have to be equal to or less than one by definition. To ensure both of these requirements in the model, R_t is decomposed into

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1} \tag{4.61}$$

and Q_t has the following dynamics

$$Q_t = (1 - a - b)\bar{Q} + a\eta_{t-1}\eta'_{t-1} + bQ_{t-1}, \tag{4.62}$$

[9] Can be defined as any univariate GARCH model, we chose the simplest one.

where $\eta_t = D_t^{-1}X_t$ and $\bar{Q} = cov[\eta_t\eta_t'] = \mathbb{E}[\eta_t\eta_t'] = R$ is the the unconditional covariance matrix of the standardized errors η_t. $\bar{Q}$ can be estimated as

$$\bar{Q} = \frac{1}{T}\sum_{t=1}^{T}\eta_t\eta_t',$$

where parameters a and b are scalars, and Q_t^* is diagonal matrix with square root of the diagonal elements of Q_t at the diagonal. Q_t has to be positive definite to ensure R_t to be positive definite. There are also some criterions on the parameters a and b to guarantee H_t to be positive definite such as $a \geq 0$, $b \geq 0$ and $a + b < 1$. In addition the starting value of Q_t, has to be positive definite to ensure H_t to be positive definite.

The correlation structure can be extended to the general DCC (M, N) model

$$Q_t = \left(1 - \sum_{i=1}^{M} a_m - \sum_{n=1}^{N} b_n\right)\bar{Q}_t + \sum_{m=1}^{M} a_m\eta_{t-1}\eta_{t-1}' + \sum_{n=1}^{N} b_n Q_{t-1}. \tag{4.63}$$

In this thesis only the DCC $(1, 1)$ will be studied. For more details see Engle [16].

Suppose now that process Z_t is multivariate Gaussian distributed such that $\mathbb{E}Z_t = 0$ and $\mathbb{E}[Z_tZ_t'] = I_n$, Engle proposed the estimation of the DCC model by two-step procedure. This is possible as the conditional variance $H_t = D_tR_tD_t$ can be divided into volatility part and correlation part. Instead of using the likelihood function for all the coefficients he suggested replacing R_t by the identity matrix which leads to a quasi log-likelihood function that is the sum of likelihood functions of n univariate models. In the second step Engle estimates parameters of R_t. Method produces consistent but not efficient estimators. In order to estimate the parameters of H_t, the following log-likelihood function L can be used

$$\begin{aligned} L &= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + \log(|H_t|) + X_t'H_t^{-1}X_t) \\ &= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + \log(|D_tR_tD_t|) + X_t'D_t^{-1}R_t^{-1}D_t^{-1}X_t) \\ &= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + 2\log(|D_t|) + \log(|R_t|) + X_t'D_t^{-1}R_t^{-1}D_t^{-1}X_t). \end{aligned} \tag{4.64}$$

In the first step the likelihood involves replacing R_t with the identity matrix I_n in (4.64). Let the parameters of the model θ be written in two groups $(\psi, \theta) = (\psi_1, \ldots, \psi_T, \theta)$, where the elements of ψ_i correspond to the parameters of the univariate GARCH model for the

ith returns, $\psi_i = (c, \phi_{1i}, \ldots, \phi_{p_i i}, \theta_{1i}, \ldots, \theta_{q_i i})$. Lets call the first step quasi log-likelihood function $L_1(\psi)$ defined as

$$\begin{aligned} L_1(\psi) &= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + 2\log(|D_t|) + \log(|I_n|) + X_t' D_t^{-1} R_t^{-1} D_t^{-1} X_t) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\big(n\log(2\pi) + 2\log(|D_t|) + X_t' D_t^{-1} R_t^{-1} D_t^{-1} X_t\big) \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left(n\log(2\pi) + \sum_{i=1}^{n}\left[\log(h_{it}) + \frac{X_{it}^2}{h_{it}}\right]\right) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\left(T\log(2\pi) + \sum_{t=1}^{T}\left[\log(h_{it}) + \frac{X_{it}^2}{h_{it}}\right]\right), \end{aligned} \tag{4.65}$$

which is the sum of the log-likelihoods of the univariate GARCH processes of n returns. Hence the parameters of the different univariate models can be determined separately. The result of first step is the estimator of parameter ψ. Then also the conditional variance h_{it} is estimated for each returns $i = 1, \ldots, n$ and then $\eta_t = D_t^{-1/2} X_t$ and $\bar{Q} = \mathbb{E}[\eta_t \eta_t']$ can be estimated as well.

In the second step, $\theta = (a, b)$ is estimated, given the estimated parameters from step one. Second step quasi log-likelihood is defined as follows

$$\begin{aligned} L_2(\theta|\hat{\psi}) &= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + 2\log(|D_t|) + \log(|R_t|) + X_t' D_t^{-1} R_t^{-1} D_t^{-1} X_t) \\ &= -\frac{1}{2}\sum_{t=1}^{T}(n\log(2\pi) + 2\log(|D_t|) + \log(|R_t|) + \eta_t' R_t^{-1} \eta_t). \end{aligned} \tag{4.66}$$

Since we are conditioning on $\hat{\psi}$, the D_t terms are constant and we can exclude that terms and maximize

$$L_2^*(\theta|\hat{\psi}) = -\frac{1}{2}\sum_{t=1}^{T}(\log(|R_t|) + \eta_t' R_t^{-1} \eta_t).$$

Asymptotic properties of the two-step estimation procedure have been studied in Engle and Schepard [18]. They introduced assumptions for consistency and asymptotic normality of the parameter estimates for DCC models.

Chapter 5

Empirical application

This chapter contains empirical application of the multivariate GARCH models proposed in the previous chapter. The models used in the empirical application were these three: BEKK model proposed by Baba, Engle, Kraft and Kroner, GO-GARCH model of van der Weide and Dynamic conditional correlation (DCC) of Engle. The used data are described in Chapter 2, however let us it briefly recapitulate for completeness.

The data consists of daily foreign exchange returns of Euro/U.S. dollar and Czech koruna/U.S. dollar pairs and as mentioned before, further we have returns of 4 stock market indices. Namely AEX, DAX, PX and DJIA correspond to the Amsterdam, Frankfurt, Prague stock market indices and Dow Jones Industrial Average respectively. This chapter is organized as follows. In the first part data descriptions are summarized, then in the second part we take a look at the dynamics of estimated conditional volatilities using all three models. We study volatility dynamics of the returns by utilizing multivariate GARCH models and then we report statistically significant cross market effects as evidence of linkages and measure the extent of the linkages by the estimated time-varying correlations. The next part of this chapter is focused on comparison of the multivariate GARCH models. The focus of reporting results will therefore be on conditional correlations implied by the estimated models. This chapter will end up with diagnostic checking of our results.

All the estimations in this thesis were performed by the R programming software which is freely available[1]. The main three packages used are: mgarchBEKK developed by Schmidbauer and Tunalioglu (2006), ccgarch developed by Nakatani (2009) and gogarch developed by Pfaff (2009).

[1] http://www.r-project.org/

Table 5.1: Descriptive Statistics of the foreign exchange returns.

	EUR/USD	CZK/USD
Mean	-1.3992e-4	-7.1043e-5
Std. Dev.	0.005696	0.0066392
Skewness	-0.095439	0.086989
Kurtosis	5.108289	5.072628
Jarque-Bera	4378.421	4316.572
P-value	0.000	0.000

Table 5.2: Descriptive Statistics of the Indices.

	AEX	DAX	PX	DJIA
Mean	-2.6797e-4	-3.6927e-5	3.5270e-4	-2.2371e-5
Std. Dev.	0.01694524	0.01725464	0.01662702	0.01346429
Skewness	-0.1711321	0.01974198	-0.6053644	-0.00931465
Kurtosis	6.668711	4.215467	11.64303	6.81868
Jarque-Bera	4412.066	1759.293	13554.24	4600.591
P-value	0.000	0.000	0.000	0.000

5.1 Data description

Table 5.1 provides a summary of the descriptive statistics of the returns for the two currencies measured against the dollar. For the standard normal distribution, the skewness and kurtosis have values of 0 and 3, respectively. As can be observed from the Table both series have a relatively high kurtosis greater than 3 indicating that the series is non-symmetric with higher peaks than the normal distribution.

Situation is similar in stock markets data. In Table 5.2 we report the descriptive statistics for each index. It can be observed that the standard deviation of the daily returns shows little variation across the indices. We find out that the European indices are maybe a bit more volatile than the U.S. Dow Jones index. The least volatile European index is the Prague PX index. This may be, because it is the smallest from our sample in terms of market capitalisation. However, there is still significant difference between PX

Table 5.3: Unconditional Correlation coefficients of the returns series.

	AEX	DAX	PX	DJIA
AEX	1			
DAX	0.8568444	1		
PX	0.5330840	0.4924072	1	
DJIA	0.5630591	0.6086716	0.3260289	1

and Dow Jones indices. More apparent differences between the indices concern skewness and kurtosis. From descriptive statistics, we read that the empirical densities associated with the PX and Dow Jones indices exhibit the most substantial heavy tails, however all series exhibit heavy tailed distributions. The unconditional correlations for each pair are displayed in the Table 5.3.

More importantly, we need to find out the dynamics of the correlations, so in the next sections we take a look at the estimation results.

5.2 Estimation results

In this section we report the results for each estimated multivariate GARCH model. We start with the BEKK (p, q) model. Estimated coefficients of the parameters of the BEKK model for both data samples exchange rates and stock market indices respectively can be found in Tables in the Appendix. The order of the model was estimated as BEKK $(1, 1)$ with $K = 1$ and the method for the estimation of parameters was maximum log-likelihood. To illustrate the time commitment of the estimation of the complex BEKK model we mention that it took 1 hour and 45 minutes. A plot of the estimated conditional volatilities of the series in Figure 5.1 reveals that the volatility dynamics of foreign exchanges is similar e contra to Figure 5.4, which implies that European stock indices have always been more volatile than the U.S. Dow Jones especially during the two financial crises in 2002 and 2008 respectively, which is not surprising. However we can clearly see quite similar dependence in conditional volatilities for each series.

The second model which we consider in our empirical application is the GO-GARCH (p, q) model. Components were estimated by maximum likelihood with formula for unobserved components as GARCH $(1, 1)$. Estimation of the parameters included two parts. The estimation of inverse matrix of the linear map M given in Table 7.5, and the estimation of the parameters of the GARCH models of the unobserved components. Coefficients of the estimated parameters can be found in Table 7.6 in Appendix. Estimated conditional volatilities for each of the series based on the GO-GARCH model are displayed in Figures 5.5 and 5.3. Here we can also see quite similar dependence in conditional volatilities for all stock market series except PX. The PX index differs in its estimated conditional volatility during the crisis in 2002, which is much smaller compared to the other series.

The last model which we consider was the DCC model. The DCC estimates of the conditional correlations between the volatilities and also estimates of the GARCH parameters are presented in Table 7.7. As the estimates of both a, the impact of past shocks on current conditional correlations, and b the impact of previous dynamic conditional correlations, are statistically significant, this clearly indicates that the conditional correlations are not constant. The estimate of a is generally low and close to zero, whereas the estimate b is extremely high and close to unity. The conditional correlations between the indices are dynamic. These findings are consistent with the plots of dynamic correlations between the index pairs in Figures 5.7 - 5.12 which change over time. Figures 5.6 and 5.2 displays the estimated volatilities based on the DCC model. At first sight, all tree methods seem to imply very similar volatilities.

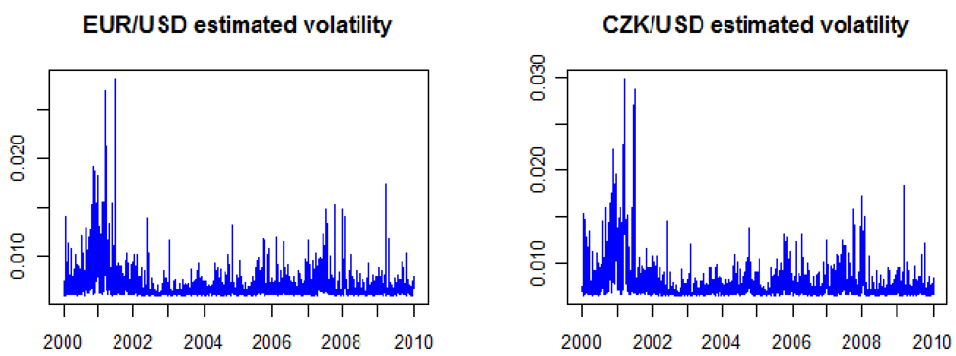

Figure 5.1: Estimated conditional volatility of the foreign exchange rates during January 2000 and December 2009, using the BEKK model computed in R programming software.

Figures 5.1, 5.2 and 5.3 provides a general view of the dynamics of the conditional volatility over the entire sample. We can clearly see from the graphs that as soon as in the foreign exchange rates the GO-GARCH model provides smoother conditional volatilities in comparison to the BEKK. In estimates in stock market data, the smoother volatilities are provided by DCC model. In general, BEKK estimates are more volatile than the other multivariate models. In 2000, volatility of the Euro/U.S. dollar averaged about 8

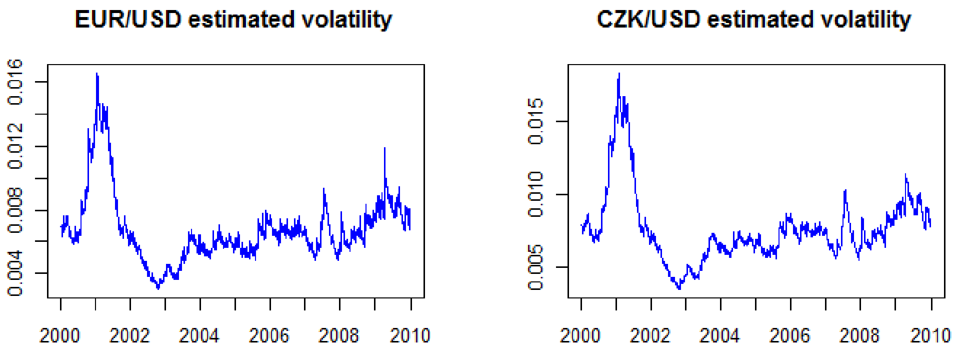

Figure 5.2: Estimated conditional volatility of the foreign exchange rates during January 2000 and December 2009, using the DCC model computed in R programming software.

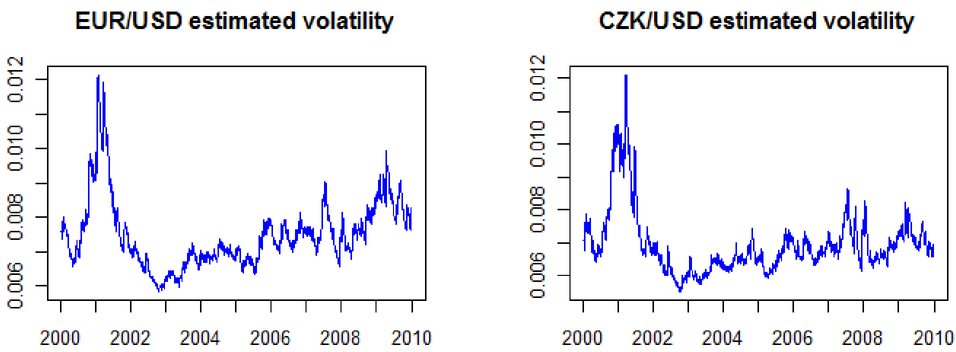

Figure 5.3: Estimated conditional volatility of the foreign exchange rates during January 2000 and December 2009, using the GO-GARCH model computed in R programming software.

percent, before beginning a steady increase over subsequent year. In 2001, volatility in the Euro/U.S. dollar averaged about 15 percent. In general volatility strongly increases during the financial crises, which is not surprising. The first crisis known as dot-com bubble[2] started in 2001 and lasted up to second half of 2003. We also mention the September 11th 2001 terrorist attack into the Twin Towers of the World Trade Center in New York City. In 2003, it has been just above 6 percent however we can see slowly decline over the subsequent years. Again, looking at average volatility does not convey the full story. Whereas spikes were common in the 1990s, they have been few and far between more recently. Even dramatic events such as widely known financial crisis of 2007-present, a crisis triggered by a liquidity crisis in the United States banking system. Volatility is

[2]The "dot-com bubble" was a speculative bubble covering roughly 1995 - 2000 during which stock markets in industrialized nations saw their equity value rise rapidly from growth in the more recent Internet sector and related fields.

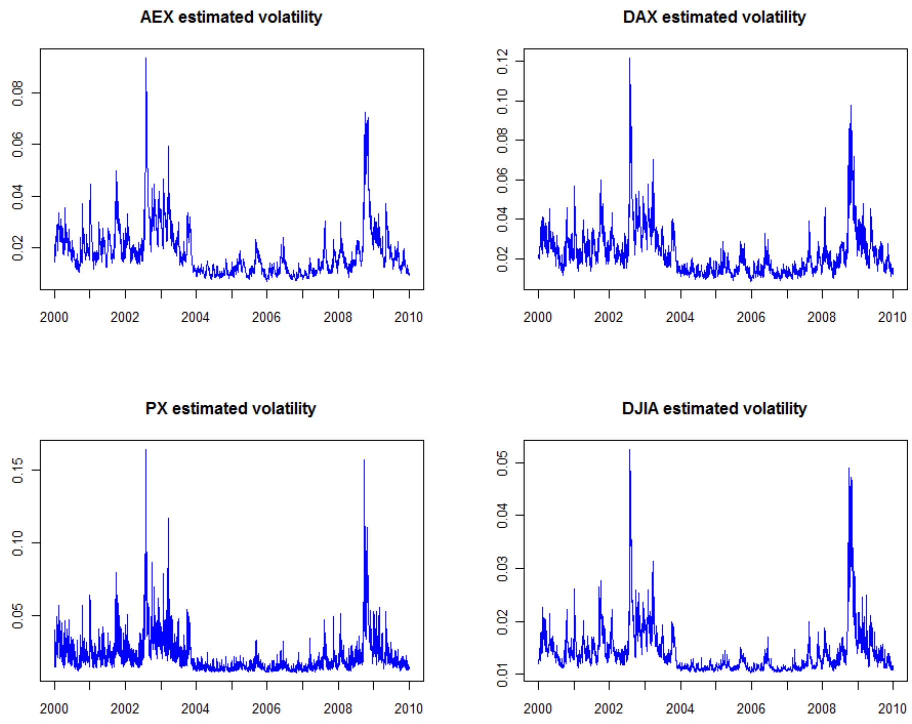

Figure 5.4: Estimated conditional volatility of stock indices AEX, DAX, PX, DJIA (corresponding, to the stock markets in Amsterdam, Frankfurt, Prague and the U.S.) during January 2000 and December 2009, using the BEKK model computed in R programming software.

somewhere near the peaks reached in the mid-1990s. This increase is observable on all financial markets and is captured by all models. We observe the same phenomenon in the Czech koruna/U.S. dollar currency pair. Other papers find that this longer term phenomenon in many emerging market currencies as well, including some that experienced currency crises in the recent past. In recent years, a number of countries have adopted more flexible exchange rate regimes, often after being unable to maintain fixed or very narrow bands.

Much more interesting results of this investigation are the estimated conditional correlations between the stock markets presented in Figures 5.7 - 5.12. Such previous findings are robust across models and are valid also in stock markets. It is clear that correlations have changed substantially over the 10 year period and exhibit time-dependence.

We now move to the investigation of market linkages between the European indices and the globally leading developed market index Dow Jones. We can see from Figures

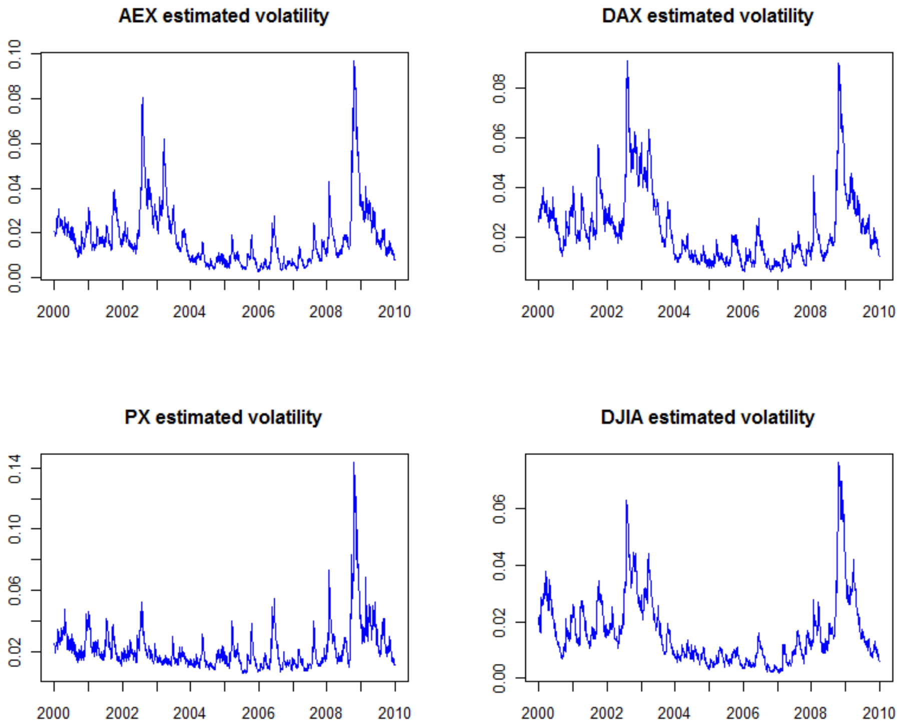

Figure 5.5: Estimated conditional volatility of stock indices AEX, DAX, PX, DJIA (corresponding, to the stock markets in Amsterdam, Frankfurt, Prague and the U.S.) during January 2000 and December 2009, using the GO-GARCH model computed in R programming software.

5.10 and 5.11 quite similar level of dependence between DAX & DJIA and AEX & DJIA respectively. The correlations between DAX & DJIA are quite strongly positively correlated and seem to be stable during the financial crises. High conditional correlation represents high financial integration, however economic and political developments of the different regions plays significant role especially during the relatively quiet periods without financial crises. A bit different situation is displayed in Figure 5.12 which represents the conditional correlation between DJIA & PX. Correlations seem to be stable in this case, however BEKK line exhibits big volatility. From our sample pair DJIA & PX represents for investors the best ability of portfolio diversification. In Europe the most developed stock market is Frankfurt. As we expect the correlations between DAX & AEX are highest and smooth. These results suggest that the regional developed market in Frankfurt is influential in the pricing process of the markets in Amsterdam and Prague, and there is a close relationship between the stock markets in Amsterdam and Frankfurt in particular. Given this interdependence, investors may perceive the stock markets in Frankfurt

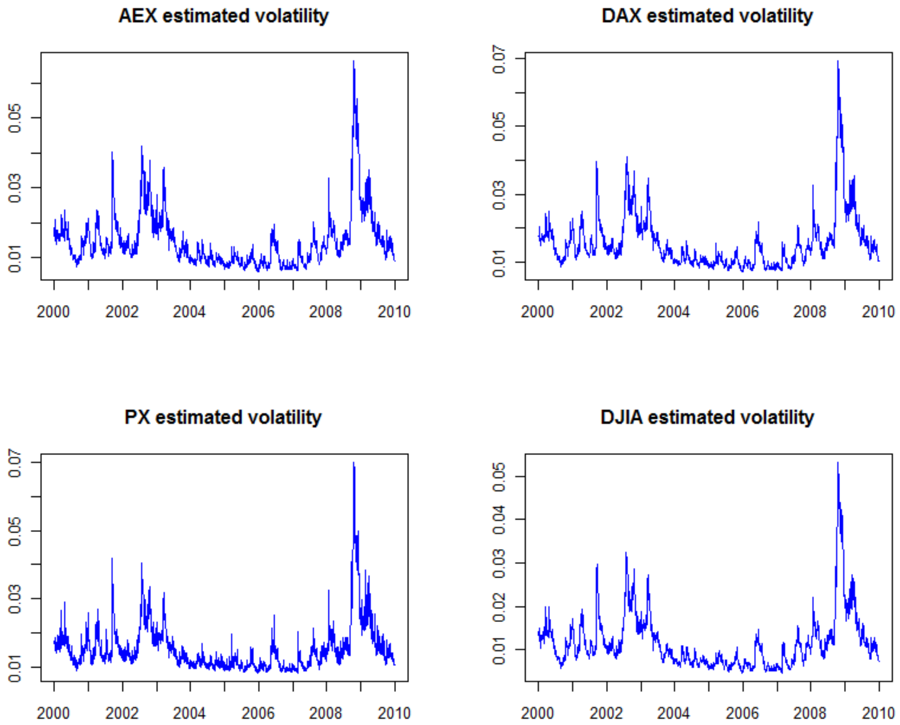

Figure 5.6: Estimated conditional volatility of stock indices AEX, DAX, PX, DJIA (corresponding, to the stock markets in Amsterdam, Frankfurt, Prague and the U.S.) during January 2000 and December 2009, using the DCC model computed in R programming software.

and Amsterdam as one investment opportunity instead of two separate classes of assets. The emerging stock market in Prague is much more affected by the local political and economic decisions in the Czech Republic. The evidence in this study seems to confirm some findings in earlier studies, which that suggest that stock market movements in one country can significantly affect stock market movements in another country via a transmission mechanism that exists because global markets are now more closely integrated. It is possible that changes in the U.S. stock returns do indeed influence those of other markets. However emerging markets of the world are responding to economic and political developments in their regions as well. These results indicate further opportunities for global portfolio diversification.

5.3 Model comparison

We have observed a number of apparent results. Whereas estimated conditional volatility for each model seems to be similar, for the conditional correlations the differences between the three methods are more pronounced. This is obvious from Figures 5.7 - 5.12, where we have depicted the estimated conditional correlation series of each stock market pairs in a separate plot. The most obvious difference between the BEKK correlations and the other two specifications is the range in which they vary. This feature of BEKK model is useful, however for this generality indeed we need to pay that a lot of parameters have to be estimated and then, as we mentioned before, this model cannot be used in high dimensional systems. We also see that the GO-GARCH and DCC correlation patterns are similar, and that correlation series behaves like a smoothed version of the BEKK correlations. However DCC provides a little underestimate correlations comparing to the GO-GARCH, most in the beginning of the period. It is quite debatable whether the short periods of very low correlation implied by the BEKK model are genuine, they may be fully driven by the volatility patterns in those periods, and in that case the less volatile behavior of the GO-GARCH and DCC correlations may provide a better indication of the actual correlation between the pairs of the stock markets indices.

The BEKK model is not very convenient for investigating conditional covariances in high-dimensional systems, because it has huge time commitment. Thus we can use BEKK model without any restrictions on the parameters in long term technical analysis on stock markets, but it is useless in short decision processes such as algorithmic trading. However, we have a direct interpretation of the parameters which explain such of an interesting information. The off-diagonal elements of the matrices A capture the cross-market shock effects among the four pairs. The off-diagonal elements of matrices G capture the cross-market volatility spillovers. Although the estimated models do not display fully identical correlations, the general message in them remains more or less the same. It is up to the user to select the model he wants to use in portfolio management.

Finally we present in our empirical application small diagnostic checking on the standardized residuals. Figures 5.13 - 5.15 represent Q-Q plots which are Quantile-Quantile plots where we can visually check for fit of a theoretical distribution to the observed data. The observed values are plotted against the theoretical quantiles. A good fit of the theoretical distribution to the observed values would be indicated by this plot if the plotted values fall onto a straight line. Except for a few points we can see in our figures that

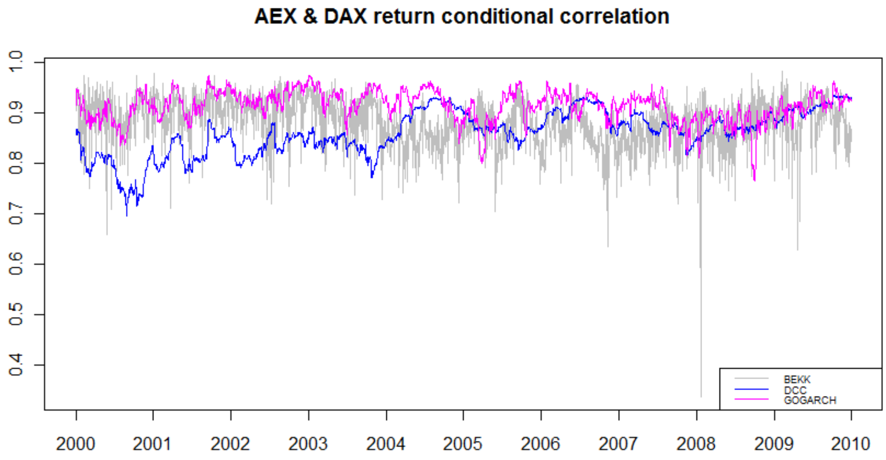

Figure 5.7: Estimated conditional correlation of stock indices AEX & DAX computed in R programming software.

points lie more or less on a straight line and so we can conclude that our models fit well.

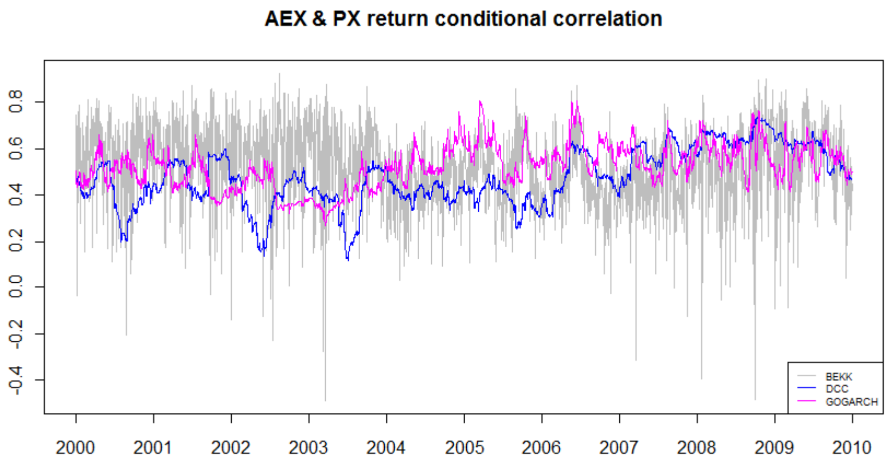

Figure 5.8: Estimated conditional correlation of stock indices AEX & PX computed in R programming software.

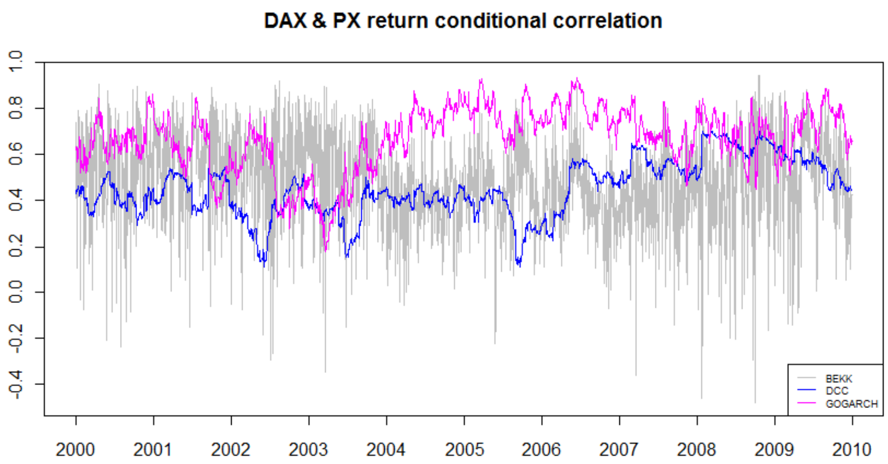

Figure 5.9: Estimated conditional correlation of stock indices DAX & PX computed in R programming software.

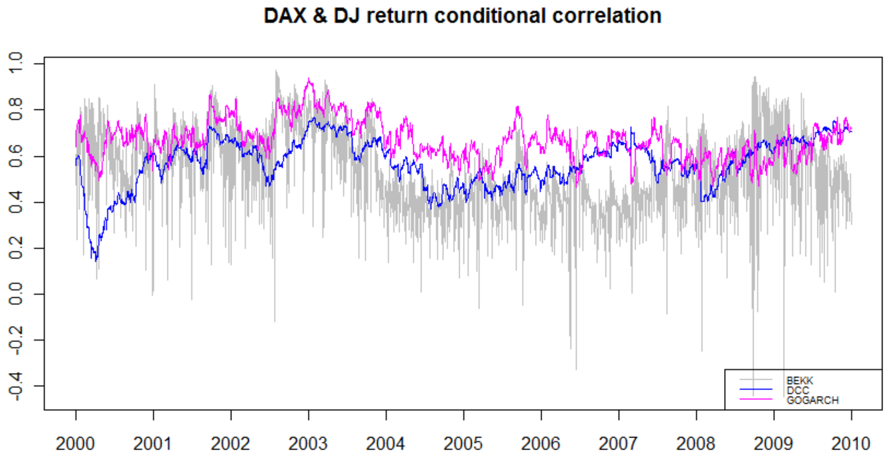

Figure 5.10: Estimated conditional correlation of stock indices DAX & DJIA computed in R programming software.

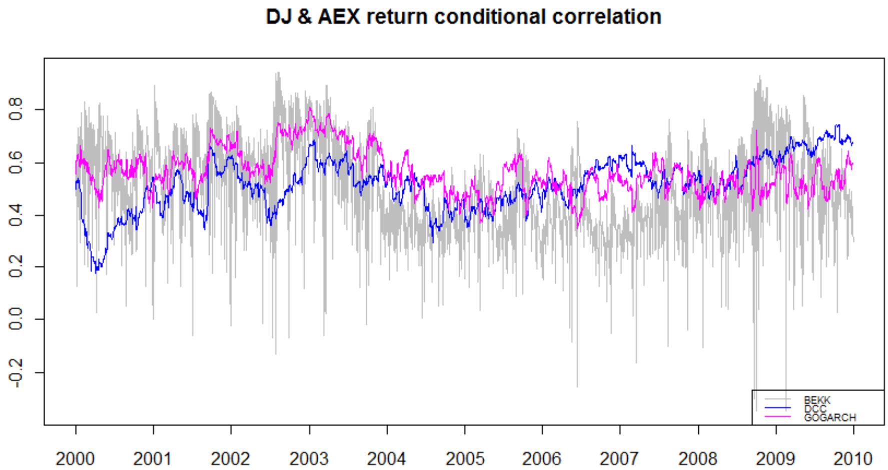

Figure 5.11: Estimated conditional correlation of stock indices DJIA & AEX computed in R programming software.

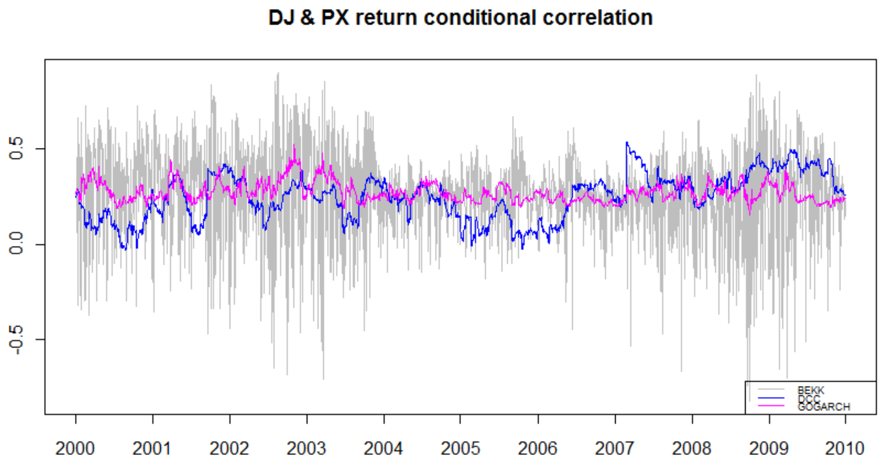

Figure 5.12: Estimated conditional correlation of stock indices DJIA & PX computed in R programming software.

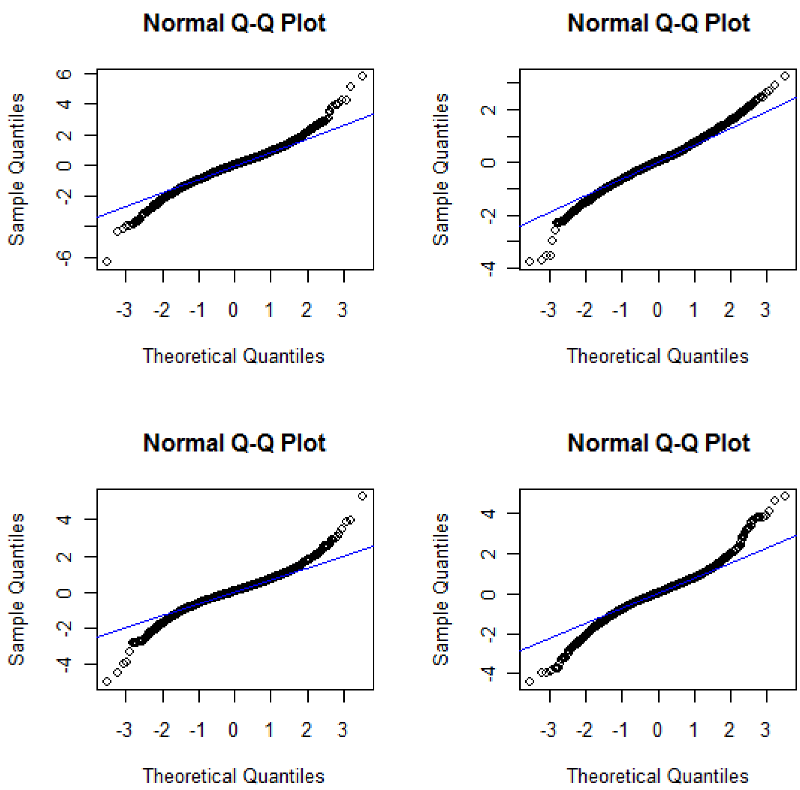

Figure 5.13: The QQ-plot of BEKK standardized residuals for each of the series plotted against normal distribution.

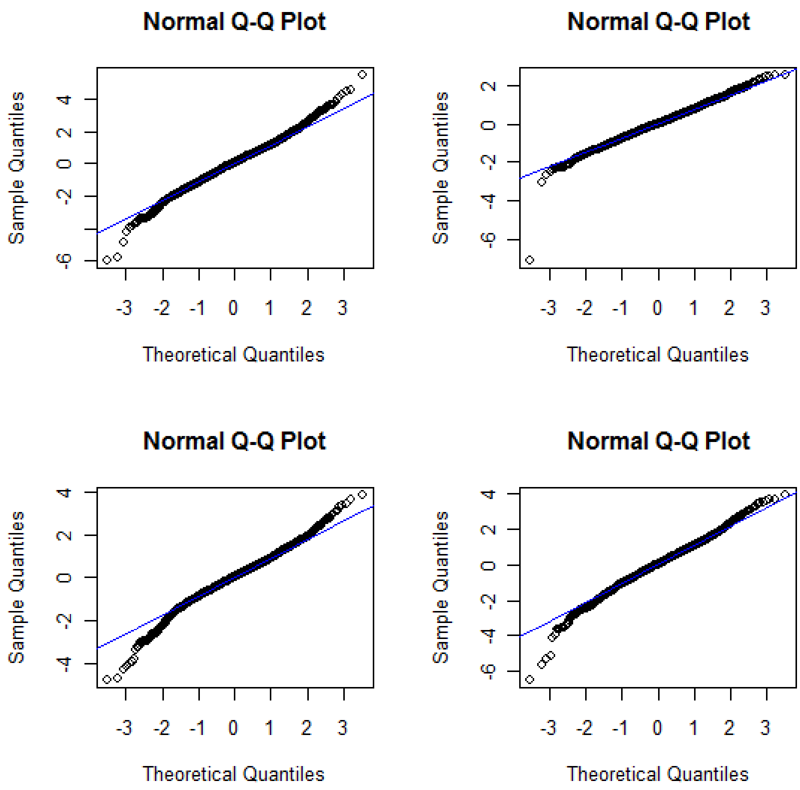

Figure 5.14: The QQ-plot of GO-GARCH standardized residuals for each of the series plotted against normal distribution.

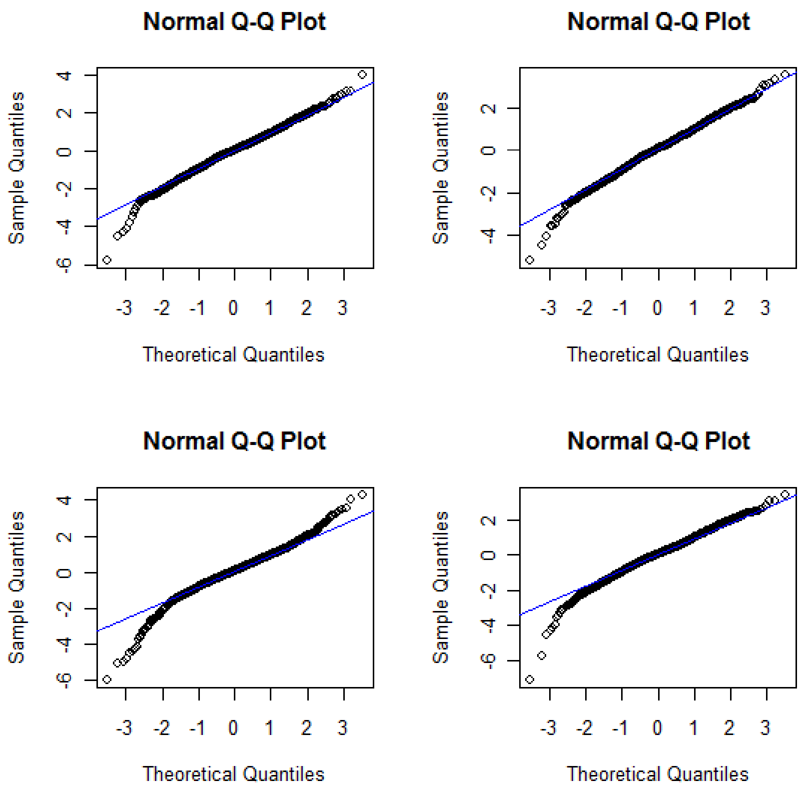

Figure 5.15: The QQ-plot of DCC standardized residuals for each of the series plotted against normal distribution.

Chapter 6

Conclusion

Volatilities and correlations among market returns are widely used in asset pricing, risk management, etc. The correct estimation in financial asset risk has important implications for investors using standard asset pricing models. Although researches have built many multivariate models, we still face the problems of curse of dimension due to the number of parameters and the restrictions on the parameters to ensure the positive definiteness of the covariance matrix. In this thesis we presented a summary of theoretical and empirical modelling with multivariate GARCH models and highlighted their features. There exist a lot of types of multivariate GARCH models, we give a survey into a basic construction. For the empirical work, BEKK, GO-GARCH and DCC models are considered and we used multistep maximum likelihood estimation procedures to estimate the models. One of the main findings is that conditional correlations exhibit significant changes over time so we concluded that despite the impact of globalization there still exist opportunities to maximize global portfolio returns through diversification. Our comparison of the models shows that the best model is BEKK, because contains the most information and is most general. However it can be used only in small dimensional systems and long term technical analysis. The differences between DCC and GO-GARCH models are not very significant and it is up to the user to select the model that he wants to use in portfolio management. Our data sample contains only 4 stock markets and 2 exchange rates all over the world. Indeed one of the challenges for the future may be including many more stock markets representatives with related exchange rates such as stock markets in New York, London, Brussels, Vienna, Warsaw or markets in Russia, Asia etc. Up to now we are unable to estimate the covariance matrix for more than 20 series keeping the flexibility of the correlations.

Chapter 7

Appendix

Table 7.1: Estimated parameters of the BEKK model for the foreign exchange rates using R programming software.

C estimates:			ARCH estimates:			GARCH estimates:		
	[, 1]	[, 2]		[, 1]	[, 2]		[, 1]	[, 2]
[1,]	0.004022	0.003747	[1,]	-0.213836	0.402224	[1,]	1.137612	1.330774
[2,]	0.000000	0.003085	[2,]	-0.024131	-0.647911	[2,]	-1.319593	-1.364835

Table 7.2: Estimation of the unobserved components of the GO-GARCH model for the foreign exchange rates using R programming software.

component GARCH model of y1			component GARCH model of y2		
	Estimate	Std. Error		Estimate	Std. Error
α_1	0.006876	0.002361	α_2	0.004719	0.002403
θ_1	0.066163	0.013278	θ_2	0.035643	0.005629
ϕ_1	0.923888	0.015038	ϕ_2	0.961683	0.005971
	inverse of linear map M				
	$[,1]$	$[,2]$			
$[1,]$	1.106201	-1.760648			
$[2,]$	-1.565876	0.758795			

Table 7.3: Estimation of the coefficients of the DCC model for the foreign exchange rates using R programming software.

	c_1	c_2	θ_{11}	θ_{21}	θ_{21}	θ_{22}
Estimate	1.77e-09	1.93e-09	0.044451	0.010989	0.004733	2.01e-02
Std.Error	1.50e-06	3.59e-02	0.019356	1.014234	0.784739	4.78e-07
	ϕ_{11}	ϕ_{21}	ϕ_{12}	ϕ_{22}	dcc a	dcc b
Estimate	0.067246	0.244156	0.688066	0.782775	0.049417	0.926287
Std.Error	0.035229	0.013626	1.054878	0.805877	0.077242	0.121174

Table 7.4: Estimated coefficients of the BEKK model for the stock indices using R programming software.

		C estimates:		
	$[,1]$	$[,2]$	$[,3]$	$[,4]$
$[1,]$	-0.001328252	0.0007815360	-0.007136690	0.0009434768
$[2,]$	0.000000000	-0.0002523830	0.001593616	0.0077046831
$[3,]$	0.000000000	0.0000000000	0.004083134	0.0009110713
$[4,]$	0.000000000	0.0000000000	0.000000000	0.0054894023
		ARCH estimates:		
	$[,1]$	$[,2]$	$[,3]$	$[,4]$
$[1,]$	0.4498350	0.1717677	1.81319949	-0.34512093
$[2,]$	-0.5521095	-0.2778714	-1.75350218	0.32867268
$[3,]$	0.2539265	0.4869067	0.29606270	-0.09401716
$[4,]$	-0.1322980	-0.4634222	-0.02202411	-0.18845023
		GARCH estimates:		
	$[,1]$	$[,2]$	$[,3]$	$[,4]$
$[1,]$	0.53359550	1.67104581	0.630420825	0.555606895
$[2,]$	0.14676640	-0.53467291	0.330673484	0.004454358
$[3,]$	0.27559500	0.07941991	-0.511893705	0.011039569
$[4,]$	0.01806348	0.02735065	-0.007910602	-0.176833176

Table 7.5: The inverse linear map M of the GO-GARCH model for the stock indices using R programming software.

	$[,1]$	$[,2]$	$[,3]$	$[,4]$
$[1,]$	-1.5790340	0.54437287	0.5986121	-0.01659185
$[2,]$	1.2247263	-1.85931672	0.1977904	-0.01368647
$[3,]$	0.1504665	0.62499901	-0.3200783	-1.19868540
$[4,]$	0.1740508	-0.06600478	-0.9525356	0.41074408

Table 7.6: Estimation the unobserved components of the GO-GARCH model for the stock indices using R programming software.

component GARCH model of y1			component GARCH model of y2		
	Estimate	Std. Error		Estimate	Std. Error
α_1	0.002453505	0.0007907027	α_2	0.01169125	0.003377802
θ_1	0.095984809	0.0113428725	θ_2	0.07564997	0.009729530
ϕ_1	0.903295047	0.0103996197	ϕ_2	0.91696444	0.010068098
component GARCH model of y3			component GARCH model of y4		
	Estimate	Std. Error		Estimate	Std. Error
α_3	0.005753897	0.001574812	α_4	0.02364814	0.005380396
θ_3	0.087195479	0.010800894	θ_4	0.10975311	0.012080695
ϕ_3	0.904130614	0.011275317	ϕ_4	0.87279734	0.012526196

Table 7.7: Estimation the coefficients of the DCC model for the stock indices using R programming software.

	c_1	c_2	c_3	c_4
Estimate	2.150578e-09	2.172862e-08	4.036082e-05	3.736150e-07
Std.Error	1.165675e-05	3.476810e-02	3.304652e-02	1.731100e-02
	θ_{11}	θ_{21}	θ_{31}	θ_{41}
Estimate	0.05819915	0.0001371262	0.001034935	0.03017585
Std.Error	0.04722802	0.4794529685	0.314904224	0.22678224
	θ_{12}	θ_{22}	θ_{32}	θ_{42}
Estimate	0.05649081	9.987891e-03	0.009407108	0.001248209
Std.Error	0.64837345	1.546184e-05	0.033259931	0.040022438
	θ_{13}	θ_{23}	θ_{33}	θ_{43}
Estimate	0.02124878	0.01411880	0.07519828	0.00875847
Std.Error	0.01663747	0.04481412	0.56621089	0.43600906
	θ_{14}	θ_{24}	θ_{34}	θ_{44}
Estimate	0.0448448	0.02122279	1.777048e-01	0.06470316
Std.Error	0.3357809	0.80412176	2.355803e-05	0.03407691
	ϕ_{11}	ϕ_{21}	ϕ_{31}	ϕ_{41}
Estimate	0.26481339	0.55286028	0.55360694	0.03337047
Std.Error	0.03238008	0.02720216	0.06119448	0.54261979
	ϕ_{12}	ϕ_{22}	ϕ_{32}	ϕ_{42}
Estimate	0.05516633	0.1419639	0.06464883	2.664884e-02
Std.Error	0.40661244	0.4919522	1.24803387	5.204570e-06
	ϕ_{13}	ϕ_{23}	ϕ_{33}	ϕ_{43}
Estimate	0.09009870	0.28060490	0.001208547	0.002628470
Std.Error	0.02081805	0.01545874	0.009197382	0.031374674
	ϕ_{14}	ϕ_{24}	ϕ_{34}	ϕ_{44}
Estimate	0.6312651	0.1156965	0.05845889	0.7690477
Std.Error	0.2305305	0.1964639	0.15454294	0.2726603
	dcc a	dcc b		
Eestimate	0.01614616	0.97938917		
Std.Error	0.01585987	0.02006798		

Bibliography

[1] ALEXANDER, C. O. (2001): Orthogonal GARCH, *Mastering Risk*, Vol. 2, edited by C.O. Alexander, 21-38, London: Financial Times-Prentice Hall.

[2] ALEXANDER, C. O. AND CHIBUMBA, A.M. (1995): Multivariate Orthogonal Factor Garch, University of Sussex Discussion Papers in Mathematics.

[3] AUE, A., HORMANN, S., HORVATH, L. AND REIMHERR, M. (2009): Break detection in the covariance structure of multivariate time series models, *Annals of Statistics*, 37, 4046-4087.

[4] BABA, Y., ENGLE, R. F., KRAFT, D. AND KRONER, K. (1990): Multivariate simultaneous generalized ARCH, *mimeo*, UCSD.

[5] BERKES, I., HORVATH, L. AND KOKOSZKA, P. S. (2003): GARCH processes: structure and estimation. *Bernoulli*, 9, 201-227.

[6] BAUWENS, L., LAURENT, S. AND ROMBOUTS, J. V. K. (2006): Multivariate GARCH models: A survey, *Journal of Applied Econometrics*, 21, 79-109.

[7] BOLLERSLEV, T. (1986): Generalized Autoregressive Conditional Heteroskedasticity, *Journal of GARCH*, 31, 307-327.

[8] BOLLERSLEV, T. (1990): Modelling The Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model, *Review of Economics and Statistics*, 72, 498-505.

[9] BOLLERSLEV, T., ENGLE, R. F. AND NELSON, D. B. (1994): ARCH models, *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, 4, 2959-3038. Elsevier Science, Amsterdam.

[10] BOLLERSLEV, T., ENGLE, R. F. AND WOOLDRIDGE, J. M. (1988): A capital asset pricing model with time-varying covariances, *The Journal of Political Economy*, 96, 116-131.

[11] Boswijk, H. P. and van der Weide, R. (2008): Testing the number of factors in GO-GARCH models, *2nd Annual Conference on Bootstrap and Numerical Methods in Time Series*, Granger Centre, University of Nottingham.

[12] Boswijk, H. P. and van der Weide, R. (2006): Wake me up before you GO-GARCH, Tinbergen Institute Discussion Papers 06-079/4, Tinbergen Institute.

[13] Broda, S. A. and Paolella, M.S. (2008): CHICAGO: A fast and accurate method for portfolio risk calculation, *Journal of Financial Econometrics*, 7, 412-436.

[14] Comte, F. and Lieberman, O. (2003): Asymptotic theory for multivariate GARCH processes, *Journal of Multivariate Analysis*, 84, 61-84.

[15] Engle, R. F. (1982): Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1006.

[16] Engle, R. F. (2002): Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics*, 20, 339-350.

[17] Engle, R. F. and Kroner, K. F. (1995): Multivariate simultaneous generalized ARCH, *Econometric Theory*, 11, 122-150.

[18] Engle, R. F. and Sheppard, K. (2001): Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH, NBER Working Papers 8554, National Bureau of Economic Research, Inc.

[19] Francq, C. and Zakoian, J-M. (2004): Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10, 605-637.

[20] Francq, C. and Zakoian, J-M. (2010): QML estimation of a class of multivariate GARCH models without moment conditions on the observed process. *MPRA Paper No. 20779*, Unpublished.

[21] Hong, L. and Majerowska, E. (2008): Testing stock market linkages for Poland and Hungary: A multivariate GARCH approach, *Research in International Business and Finance*, 22, 247-266.

[22] Lee, S. and Hansen, B. (1994): Asymptotic theory for the GARCH (1,1) quasi maximum likelihood estimator, *Econometric Theory*, 10, 29-52.

[23] Lumsdaine, R. (1996): Consistency and asymptotic normality of the quasi maximum estimators in conditional heteroskedasticity models, *Econometrica*, 65, 587-599.

[24] Mandelbrot, B. (1963): The variation of certain speculative prices, *The Journal of Business of the University of Chicago*, 36, 394-419.

[25] Sheppard, K. (2003): Multi-step estimation of Multivariate GARCH models, Proceedings of the International ICSC Symposium: Advanced Computing in Financial Markets, University of California at San Diego, revised 25 Oct 2003.

[26] Scheicher, M. (2001): The comovements of stock markets in Hungary, Poland and the Czech Republic, *International Journal of Finance and Economics*, 6, 27-39.

[27] Silvennoinen, A. and Terasvirta, T. (2008): Multivariate GARCH models, Working Paper Series in Economics and Finance 669, Stockholm School of Economics, revised 18 Jan 2008.

[28] Taylor, S. J. (2005): Asset Price Dynamics, Volatility, and Prediction, Princeton University Press.

[29] Tse, Y. K. and Tsui, K. C. (2002): A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations, *Journal of Business and Economic Statistics*, 20, 351-362.

[30] van der Vaart, A. W. (2008): Time series, Lecture notes, Vrije Universiteit Amsterdam.

[31] van der Weide, R. (2002): GO-GARCH: A multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics*, 17, 549-564.

[32] Vronos, I. D., Dellaportas, P. and Politis, D. N. (2003): A full-factor multivariate GARCH model, *Econometrics Journal*, 6, 312-334.

[33] Weiss, A. A. (1986): Asymptotic theory for ARCH models: Estimation and testing, *Econometric Theory*, 2, 107-136.