university of groningen

faculty of science and engineering

Modelling the dependence structure of financial time series: considering copula-MGARCH models

Master Project Mathematics

May 2016

Student: D.H. Chu

First supervisor: Prof. dr. E.C. Wit

Second supervisor: Dr. M.A. Grzegorczyk

Modelling the dependence structure of financial time series: considering copula-MGARCH models

Hung Chu (S2348594)

MSc. Mathematics
University of Groningen, the Netherlands
Faculty of Science and Engineering
Master thesis

June 12, 2017

Abstract

In this thesis we study multivariate models using twenty financial time series. The models are obtained by combining copula functions with multivariate GARCH models. The advantage is that these models can account for the volatile behaviour of the time series through the multivariate GARCH specification, while allowing the copula functions to explicitly account for the underlying dependence structure. The dependence structure of the time series will be described by a rank correlation and tail-dependence coefficients. The main objective of this thesis is to identify adequate models for fitting and forecasting of multivariate time series data and how these models describe the underlying dependence structure. We achieve this by evaluating the models using three different tests: one for model selection, one for model validation and one for model ranking. For comparison, we also evaluate traditional multivariate models. By combining the findings of the three tests, we conclude that the DCC-ARMA-GJR model incorporated with the t copula is the most appropriate. The main finding is that financial crises do not necessarily affect the degree of correlation of pairs of stock indices. The change in correlation can differ significantly per event and can be relatively limited. Furthermore, many stock markets have little to no tendency to crash or boost together, sometimes even in case of relatively high correlation.

JEL classification. C22; C51; C52; C53; G17; G32.
Keywords. Copulas; Dependence structure; Model evaluation; Multivariate GARCH models; Value-at-risk.

Acknowledgement

First and foremost I would like to express my sincere gratitude to my first supervisor Prof. dr. E.C. Wit, also known as Ernst. I am extremely thankful for his aspiring guidance, creative ideas, valuable comments and patience. He gave me the opportunity to work on my own on a wonderful topic, but he always steered me in the right direction whenever I needed it. It was a great honor and pleasure to have Ernst as my first supervisor.

Second of all, I would like to express my warm thanks to my second supervisor Dr. M.A. Grzegorczyk, also known as Marco. I am very thankful for his clear and profound comments. He was accessible for assistance at all times and I truly admire his open attitude.

Finally, I would like to thank my family and friends who provided me with continuous support and encouragement throughout my years of study. My accomplishments would not have been possible without them. Thank you all!

Hung

Contents

1 Introduction

The Basel Committee on Banking Supervision (BCBS) concerns itself with the health of financial institutions. The BCBS requires financial institutions to hold a minimum amount of capital to cover potential losses. It is of interest of financial institutions to constantly review their investment strategies and risk profiles to make sure that they still comply the regulations, but at the same time stay competitive in the market.

The calculation of the minimum required amount of capital for the market risk is currently based on the so-called *Value-at-Risk* (VaR) risk measure. The VaR is widely considered in risk management and currently it is practised by the BCBS in Basel III and by the European Union in Solvency II. The VaR is a quantile of the return (or loss) distribution over a given time period. The regulators of the BCBS and the European Union allow financial institutions to use their own models to come up with estimates for the VaR and consequently the minimum required amount of capital. However, they have to convince the regulators that their model performs the estimation properly. Otherwise the regulators will determine this amount, which could be relatively large. An overestimation of the VaR implies that the financial institution has to hold a relatively large amount of capital aside, which could lead to lower overall profits. On the other hand, an underestimation implies that the financial institution has a larger exposure to risks, which could lead to a default. Thus it is important for financial institutions to use models that are able to accurately estimate the risk measure. Related to this, it is important that these models are able to accurately forecast future values.

The *Vector Autoregressive* (VAR) model is a traditional model that investigates the linear interdependencies among multiple time series. In this model each variable at a certain time t, such as the returns of an asset, is explained as a linear combination of its own lags and the lags of other variables, such as the returns of a different asset at time $t - 1$. A shortcoming of this model is that it does not explicitly account for *heteroskedasticity*, also known as the *time-varying variance* or *volatility*, and therefore lacks the ability to describe the volatile behaviour of the time series.

As financial time series generally change quickly through time, it is important to model the volatility accurately. Modelling the volatility of returns has attracted a lot of attention ever since the introduction of the *Autoregressive Conditional Heteroskedasticity* (ARCH) and the *Generalized ARCH* (GARCH) model, respectively developed by Engle (1982) and Bollerslev (1986), where the latter is currently the most popular one. Both models have in common that they explicitly account for the volatile behaviour of time series, thus providing a solution to the shortcoming of the VAR model. As it is of great interest to understand the comovements of financial returns, it is suggested to extend the consideration to *Multivariate*

GARCH (MGARCH) models. Many different MGARCH models have been proposed and can be found in the literature, see Bauwens et al. (2006) and Silvennoinen & Terävirta (2008) for an overview. Empirically, many MGARCH models have shown to be applicable for financial time series modelling (Billio & Caporin 2006, Cappiello et al. 2006, Engle 2002) and therefore are widely studied in financial econometrics.

In recent years, a lot of attention has been paid to measuring the interdependence of financial time series data. Many of these studies are based on the concept of *copulas*. Copulas are very popular and useful tools for multivariate modelling in many fields, such as actuarial science, finance, biomedical studies and engineering (Yan 2007). Empirically, copulas have shown to be powerful tools for modelling the dependence structure in financial risk management (Naifar 2016, McNeil et al. 2015, Aloui et al. 2011, Ghorbel & Trabelsi 2009). An advantage of copulas, which is particularly desirable for financial institutions, is that they can be used to estimate the VaR risk measure. Another advantage is that they can effectively be used in combination with time series models such as the GARCH models. Furthermore, copulas have the nice property that they do not require the multivariate joint distribution and the margins to be of the same type of distribution.

This thesis provides an thorough overview of the performance and features of multivariate GARCH models. Moreover, this thesis can be used as a reference in risk management for selecting adequate multivariate models.

Research question

The distribution of asset log losses are often assumed to be normally distributed for simplicity. This is not in agreement with the stylized facts that these distributions are fat-tailed and leptokurtic (McNeil et al. 2015). As a result, the models based on the assumption of normality tend to underestimate the risk. Also, many multivariate time series models, such as the traditional MGARCH models, use the conventional Pearson correlation coefficient as a basis for measuring the overall dependence. As mentioned by Embrechts et al. (2001), this has some serious drawbacks and could lead to a mismeasure of the dependence structure. To overcome this, Embrechts et al. (2001) suggest to use rank correlations instead. Additionally, we describe the extremal dependence of pairs of time series by tail-dependence coefficients.

For most sets of time series data it is not directly clear which model is convenient for describing certain stylized facts, such as the interdependence. In this thesis we will therefore study a number of different MGARCH-type models. We consider data of twenty different financial stock indices worldwide: seven from Asia, seven from Europe, two from North-America and six from South-America. Each of the MGARCH-type models is combined with a particular copula to try to improve the estimation of the dependence structure. These models are also

known as copula-MGARCH models. The copula allows us to effectively use different distributions for the log losses and besides can be used to estimate rank correlations and tail-dependence coefficients. The fit of the copula-MGARCH and traditional MGARCH (without copula modelling) models are evaluated by in-sample fit and out-of-sample forecast tests and compared to each other. Using the model with the overall optimal in-sample fit and out-of-sample forecast performance, we try to describe the dependence structure of the time series as accurately as possible by rank correlations and tail-dependence coefficients. These results provide proxies for (not) taking on a position of a particular stock market to hedge against other markets. Hence, the research questions are formulated as follows.

According to the criteria of the test procedures, do copulas improve the fit and the forecast of the traditional MGARCH models? Also, do the models that are based on non-normality assumptions perform better than the models that are based on normality assumptions? How does the optimal model describe the dependence structure of the financial time series data?

In order to answer the research questions, we formulate the following sub-questions:

1. How do we test the in-sample fit performance of the models?
2. How do we test the out-of-sample forecast performance of the models?
3. How do the copula-MGARCH models perform the fit and forecast of the financial time series data relative to the traditional MGARCH models?
4. How do we obtain estimates of the rank correlations and tail-dependence coefficients in order to describe the dependence structure?

This thesis contributes to the literature on multivariate modelling by incorporating copula theory into MGARCH models and additionally evaluating the performance of these models using different in-sample and out-of-sample tests. The remainder of the thesis is organized as follows. Section 2 introduces the basics of time series analysis, the GARCH-type models of interest, copula theory and lastly the model that is a combination of GARCH-type models and copulas. The empirical data and descriptive statistics are presented in Section 3. The result of our study is shown in Section 4. Section 6 concludes the thesis and Section 7 discusses our methodology and consequently provides suggestions for further studies.

2 Theoretical background

In this section we will introduce the basic definitions in time series analysis as well as the (M)GARCH models of interest and copula theory. In the end of this

section we will propose hybrid models, that are constructed by combining the concept of the MGARCH models and copulas. The proposed models enjoy the advantages from the separate concepts and corrects for the disadvantages. The theory in this section is adopted from McNeil et al. (2015), if not mentioned differently.

2.1 Time series analysis

We start by introducing the main definitions and concepts in time series analysis that are required for our study of the GARCH model.

A time series model for a risk factor, e.g. the daily (percentage) losses of a market index, is a discrete-time process $\{X_t\}_{t\in\mathbb{Z}}$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Time series models are generally used for *forecasting* future values.

Moment functions. Since time series processes are series of random variables, we are dealing with moment functions. The *mean function* $\mu(t)$ and the *autocovariance function* $\gamma(t,s)$ of a time series model $\{X_t\}_{t\in\mathbb{Z}}$ are defined as

$$\begin{aligned} \mu(t) &= E[X_t], && t \in \mathbb{Z}, \\ \gamma(t,s) &= E[(X_t - \mu(t))(X_s - \mu(s))], && t,s \in \mathbb{Z}, \end{aligned}$$

provided that the functions exists. Similarly as for the case without time-dependence, the autocovariance at two same time period is the variance, written as $\gamma(t,t) = \text{var}(X_t)$, $t \in \mathbb{Z}$.

Stationarity. The time series models considered in this thesis deal with one of the following two types of stationarity of the time series.

Definition 2.1. The time series $\{X_t\}_{t\in\mathbb{Z}}$ is (*strictly*) *stationary* if, for all $t_1, \ldots, t_n, k \in \mathbb{Z}$ and all $n \in \mathbb{N}$,

$$(X_{t_1}, \ldots, X_{t_n})' \overset{\text{d}}{=} (X_{t_1+k}, \ldots, X_{t_n+k})',$$

where $\overset{\text{d}}{=}$ means that the two series have the same distribution.

In other words, a time series is (strictly) stationary if the distribution of the observations in any two equally-spaced time periods are the same. The next type of stationarity is closely related to strictly stationarity.

Definition 2.2. The time series $\{X_t\}_{t\in\mathbb{Z}}$ is *covariance stationary* if it satisfies the following two properties

$$\mu(t) = \mu, \qquad t \in \mathbb{Z}, \tag{1}$$

$$\gamma(t,s) = \gamma(t+k, s+k), \quad t,s,k \in \mathbb{Z}, \tag{2}$$

provided that the first two moments, $E[X_t]$ and $E[X_t^2]$, exist.

The meaning of the first equation speaks for itself. The second equation implies that the "movements" of the time series is similar in any two equally-spaced. Generally speaking, both types of stationarities tell us that the time series behave similarly in any two equally-spaced time periods. It can be shown that a strictly stationary time series with finite variance is covariance stationary.

From (2) we have that the autocovariance function between X_t and X_s of a covariance stationary process only depends on the lag $|t - s|$, i.e.

$$\gamma(t - s, 0) = \gamma(t, s) = \gamma(s, t) = \gamma(s - t, 0).$$

Therefore, the autocovariance function of a covariance stationary process can be written as a function of one variable:

$$\gamma(t) \equiv \gamma(t, 0), \quad \forall t \in \mathbb{Z},$$

and so that $\gamma(0) = \text{var}(X_t)$ for all $t \in \mathbb{Z}$. Using this notation, we define the following function.

Definition 2.3. For a covariance stationary process $\{X_t\}_{t\in\mathbb{Z}}$, the *autocorrelation function* (ACF) $\rho(t)$, or the *autocorrelation* at lag t, is defined as

$$\rho(t) = \rho(X_t, X_0) = \gamma(t)/\gamma(0), \quad \forall t \in \mathbb{Z}.$$

In other words, autocorrelation is the correlation of a time series with its own past and future values. The ACF plays an important role in the construction of the time series models.

In financial time series analysis the ACF is often considered. The ACF can be used to detect whether or not the data is correlated, and if so, at which lags. Put differently, it can be used to detect randomness in the data. For independent data, the autocorrelations should be close to zero for any lags. Otherwise, for dependent data, one or more of the autocorrelations should differ significantly from zero. However, note that if the autocorrelations do not differ significantly from zero, then the data can still have dependence. The result of the ACF is just a sufficient test of independence. There are formal numerical tests to check for independence, also known as *portmanteau tests*. This includes the Ljung–Box test, see Appendix A

White noise processes. The *white noise* (WN) process is the building block of a classical time series model, namely the ARMA models. A WN process is a stationary process without autocorrelation, defined as follows.

Definition 2.4. The time series $\{X_t\}_{t\in\mathbb{Z}}$ is a WN process if it is covariance stationary with autocorrelation function

$$\rho(h) = \begin{cases} 1, & h \neq 0, \\ 0, & h = 0. \end{cases}$$

A WN process with mean μ and variance σ^2 is denoted as WN(μ, σ^2). An example of a WN process is the so-called *strict white noise* (SWN) process, defined as follows.

Definition 2.5. A discrete-time stochastic process $\{Z_t\}_{t\in\mathbb{Z}}$ is a SWN process if it is a series of independent and identically-distributed (i.i.d.) and finite variance random variables.

A SWN process with mean μ and variance σ^2 is denoted as SWN(μ, σ^2). The SWN process serves as a building block for the GARCH models. It represents the so-called *innovations* (or *standardized residuals*) of the GARCH models.

At last, we introduce the *Autoregressive-Moving-Average* (ARMA) model, which is widely used in financial risk management. The main property of the ARMA model is that it describes the autocorrelation in the time series. This model will play an important role in our analysis: it will be used in combination with the standard GARCH model.

Definition 2.6. Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be WN(0, σ_ε^2). The process $\{X_t\}_{t\in\mathbb{Z}}$ is an ARMA(p, q) process if it is covariance stationary and it satisfies the following difference equation.

$$\begin{aligned} X_t &= \mu_t + \varepsilon_t, \\ \mu_t &= \mu + \sum_{i=1}^{p} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q} \lambda_j \varepsilon_{t-j}, \quad \forall t \in \mathbb{Z}. \end{aligned}$$

In this model we have that the value of the demeaned process, i.e. $X_t - \mu$, is expressed as a function of its own lagged values and the lagged values of some error terms. This parametrization allows us to explicitly describe the autocorrelation of the time series, i.e. how the current value is related to the values in past.

2.2 Risk measure

In risk management, the upper tail of the loss distribution is commonly considered to identify the potential extreme losses. This risk can be expressed numerically by risk measures. One such risk measure is the VaR, which is currently widely used in risk management. There are unconditional and conditional variants of the VaR. We first introduce the unconditional variant in order to get a better understanding of the generally concept of the VaR.

Let $\{X\}$ be a series representing the daily losses and denote its distribution function by $F_X(x) = P(X \leq x)$.

Definition 2.7. For a given confidence level (or quantile level) $\alpha \in (0,1)$, the *unconditional VaR* at the confidence level of α is defined as

$$\mathrm{VaR}_\alpha = \inf\{x \in \mathbb{R} \mid F_X(x) \geq \alpha\}. \tag{3}$$

In risk management, $\alpha \geq 0.95$ is usually considered for losses time series. The unconditional VaR, at a confidence level $\alpha = 0.95$, is illustrated in Figure 1. From the figure we can see that the unconditional VaR is actually the α-quantile of the loss distribution function. So it answers the following question:

What value will our losses not exceed with a certain probability?

For example, the VaR for a one-day period at a confidence level of 95% is the worst loss that would be expected to occur in a single day over the next 20 days, under the assumption that the loss distribution is identical every day.

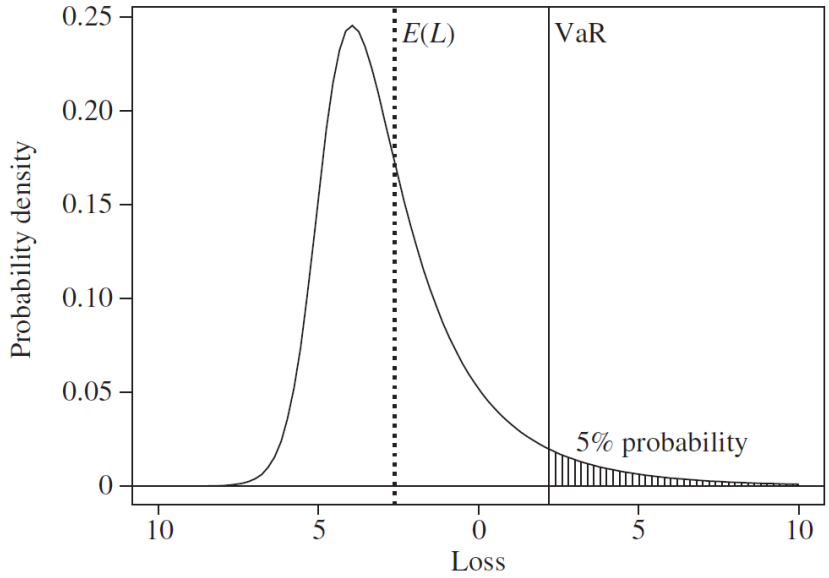


Figure 1: 95% unconditional VaR illustration (McNeil et al. 2015).

The unconditional VaR does not take into account the information at a specific time period t. As it is likely that the value of the loss at time period t is influenced by the value of the loss at time period $t-1$ to a greater extend than the values of the losses at time periods $t-2, t-3, \ldots$, we would like to account for the development of the time series through time. For this, we define the conditional variant of the VaR.

Let $\{X\}$ be a strictly stationary time series of daily losses and let us denote the information that is available at time period t by $\mathcal{F}_t$. Further, we denote the conditional distribution function of the losses by $F_{X_{t+1}|\mathcal{F}_t}(x) = P(X_{t+1} \leq x \mid \mathcal{F}_t)$.

Definition 2.8. For a given confidence level $\alpha \in (0,1)$, the (*conditional*) *VaR* over the next time period at the confidence level α is defined as

$$\mathrm{VaR}_\alpha^t = \{x \in \mathbb{R} \mid F_{X_{t+1}|\mathcal{F}_t(x)} \geq \alpha\}. \tag{4}$$

As the conditional variant of the VaR is of interest for our study, we will from now on mean the conditional VaR if we talk about the VaR. [1]

In risk management it is of interest to study the accuracy of the VaR estimation of a model. The optimal estimated VaRs are the ones that are sufficiently high so that in periods of high losses, the financial institution has enough capital/reserves to cover the losses. However, these estimated VaRs should not be too low either, because the saved capital is costly in the sense that the financial institution is not able to invest it. The evaluation of the estimated VaRs can be done by different backtests. In general, by backtesting we assess the out-of-sample forecast performance of the model. Backtests are in fact used for model validation.

For the backtests of the VaR we are interested in the number of *exceptions* (or *violations*), where an exception is said to occur if $x_t > \mathrm{VaR}_\alpha^{t-1}$. Thus the number of exceptions over the time period $T+1, \ldots, T+n$ and for a given confidence level $\alpha \in (0,1)$ is equal to

$$I_\alpha^n = \sum_{t=T+1}^{T+n} \mathbb{1}_{\{x_t > \mathrm{VaR}_\alpha^{t-1}\}},$$

where $\mathbb{1}_{\{x_t > \mathrm{VaR}_\alpha^{t-1}\}} = 1$ if $x_t > \mathrm{VaR}_\alpha^{t-1}$ and 0 otherwise.

The following three VaR backtests are commonly employed.

1. *Unconditional coverage backtest*. This is a test to examine the accuracy of the VaR estimation. The null hypothesis states that the observed number of exceptions is equal to the expected number of exceptions.

2. *Independence backtest*. This is a test to examine the clustering of the exceptions. The null hypothesis states that the exceptions occur independently throughout the backtesting period.

3. *Conditional coverage backtest*. This is a test to examine simultaneously the accuracy of the VaR estimation and the clustering of the exceptions. The null hypothesis states that the observed number of exceptions is equal to the expected number of exceptions and that the exceptions occur independently throughout the backtesting period.

[1] Some literature use the terminology of 'conditional VaR (CVaR)' for the definition of a different risk measure that is also known as the *unconditional Expected Shortfall* (ES). The conditional VaR in this thesis defines the conditional variant of the VaR and it is by no means equal to the CVaR or unconditional ES.

See Appendix B for a detailed description of these backtests. As we can see in this appendix, each backtests is a likelihood ratio test. Thus, the VaR estimation performance of the models can be evaluated by p-values. In all cases we prefer not to reject the null hypothesis, which corresponds to a p-value greater or equal to 0.05.

The conditional coverage backtest is the most comprehensive among the three as it is a compound of the other two backtests. To specify the VaR estimation we will therefore focus more on the results for this test than the other two tests. For a complete overview of the VaR estimation performance of the models, we will also perform the other two backtests and compare the accuracy, independence and joint performance.

2.3 GARCH models

In general it is more difficult to forecast a time series accurately if it is heteroskedastic than if it is homoskedastic (i.e. constant volatility). Since most, if not all, financial time series exhibit heteroskedasticity, it is essential for financial institutions to make use of models that can describe the heteroskedasticity. One such model is the GARCH model.

Definition 2.9. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be SWN(0,1). The process $\{X_t\}_{t\in\mathbb{Z}}$ is a GARCH(p,q) process if it is strictly stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $\{\sigma_t\}$, the following equations.

$$X_t = \sigma_t Z_t, \tag{5}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{6}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, p$, and $\beta_j \geq 0$, $j = 1, \ldots, q$.

In equation (6) we see that the volatility, σ_t, is expressed as a function of its own lagged values and lagged values of the process. This parametrization allows us to explicitly describe the process by its volatility. It can be shown that the necessary and sufficient condition for covariance stationarity of the GARCH model is $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. In that case the variance is finite and given by $\alpha_0/(1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j)$.

An important aspect in GARCH modelling is the assumption that we make about the distribution of the innovations $\{Z_t\}_{t\in\mathbb{Z}}$. Moreover, the assumption is crucial for the goodness-of-fit of the model. In this thesis we restrict ourselves to the normal, t and skewed t distribution for the innovations distribution.

2.3.1 Extensions of the GARCH model

In the literature there is a large number of extensions of the GARCH model, see Teräsvirta (2009) for a thorough overview. Each GARCH-type model is constructed for a particular purpose, mainly with the idea to improve older GARCH-type models. For our study, we consider two extensions of the GARCH model. The first extension is a combination of the ARMA and GARCH model, defined as follows.

Definition 2.10. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a SWN(0,1) process. The process $\{X_t\}_{t\in\mathbb{Z}}$ is an ARMA(p_1, q_1)-GARCH(p_2, q_2) process if it is covariance stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $\{\sigma_t\}_{t\in\mathbb{Z}}$, the following equations.

$$X_t = \mu_t + \varepsilon_t, \tag{7}$$

$$\mu_t = \mu + \sum_{i=1}^{p_1} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \lambda_j \varepsilon_{t-j}, \tag{8}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p_2} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2, \tag{9}$$

where $\varepsilon_t = \sigma_t Z_t$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, p_2$, $\beta_j \geq 0$, $j = 1, \ldots, q_2$ and $\sum_{i=1}^{p_2} \alpha_i + \sum_{j=1}^{q_2} \beta_j < 1$.

This model can be seen as an ARMA(p_1, q_1) process with GARCH(p_2, q_2) errors, i.e. where $\{\varepsilon_t\}_{t\in\mathbb{Z}} = \{\sigma_t Z_t\}_{t\in\mathbb{Z}}$ is a WN$(0, \sigma_\varepsilon^2)$ process. This implies that the GARCH errors, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, should be covariance stationary, because WN processes are defined as covariance stationary processes. Since the necessary and sufficient condition for covariance stationarity of the GARCH model is $\sum_{i=1}^{p_2} \alpha_i + \sum_{j=1}^{q_2} \beta_j < 1$, we build this condition into the definition of the ARMA-GARCH model. Observe that for $p_1 = q_1 = 0$ we have the standard GARCH model as in Definition 2.9.

The ARMA-GARCH model has the advantage that it is able to describe two important stylized facts of financial time series, namely the autocorrelation and the heteroskedasticity. In practice, low-order ARMA and GARCH models like the ARMA(p_1, q_1) with $p_1, q_1 \in \{0, 1, 2\}$, and GARCH$(1, 1)$ models are commonly used owing to its parsimony and the interpretability of the variables, see for example Tang et al. (2015), Weiß (2013), Ghorbel & Trabelsi (2009) and Huang et al. (2009). For that reason we will consider the ARMA(p_1, q_1)-GARCH$(1, 1)$ models, where we choose the optimal values for $p_1, q_1 \in \{0, 1, 2\}$: the pair (p_1, q_1) that results the lowest *Akaike Information Criterion* (AIC) value is considered optimal. [2]

[2] AIC $= 2k - 2\ln(L)$, where k is the number of parameters and L is the value of the likelihood function of the fitted model.

The GARCH(p, q) model is symmetric in the sense that negative and positive shocks in the previous time periods $t-1, \ldots, t-p$ have the same impact on the volatility at time t. According to economic theory, negative shocks tend to affect the volatility more than positive shocks, especially for stock returns. This asymmetry is called the *leverage effect*. Glosten et al. (1993) proposed the *Glosten-Jagannathan-Runkle-GARCH* (GJR) model that extends the GARCH model by incorporating the leverage effect. More precisely, the model allows the volatility to respond differently to negative and positive innovations in the past. The GJR model can be extended in the similar way as the GARCH model in the ARMA-GARCH model. The second extension that we consider is a combination of the ARMA and GJR model, defined as follows (Würtz et al. 2009).

Definition 2.11. Let $\{Z_t\}_{t\in\mathbb{Z}}$ be a SWN(0,1) process. The process $\{X_t\}_{t\in\mathbb{Z}}$ is an ARMA(p_1, q_1)-GJR(p_2, q_2) process if it is covariance stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $\{\sigma_t\}_{t\in\mathbb{Z}}$, the following equations.

$$X_t = \mu_t + \varepsilon_t, \tag{10}$$

$$\mu_t = \mu + \sum_{i=1}^{p_1} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \lambda_j \varepsilon_{t-j}, \tag{11}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p_2} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}\right)^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2, \tag{12}$$

where $\varepsilon_t = \sigma_t Z_t$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\gamma_i \in [-1, 1]$ is the leverage term, $i = 1, \ldots, p_2$, $\beta_j \geq 0$, $j = 1, \ldots, q_2$ and $\sum_{i=1}^{p_2} \alpha_i \mathrm{E}[(|Z| + \gamma_i Z)^2] + \sum_{j=1}^{q_2} \beta_j < 1$.

For the same reason as for the ARMA-GARCH models, we will consider low-ordered ARMA(p_1, q_1)-GJR(1, 1) models, where the optimal values for $p_1, q_1 \in \{0, 1, 2\}$ are chosen based on the AIC values.

2.3.2 Parameter estimation

The parameters of the two extensions, namely the ARMA-GARCH and ARMA-GJR model, can be estimated by the *maximum likelihood* (ML) estimation procedure. Since for our study we consider dynamic models, we denote the information that is available at observation (here: time period) t by $\mathcal{F}_t$ and recall that the conditional distribution function of the observations is given by

$$F_{X_t|\mathcal{F}_{t-1}}(x_t) = P(X_t \leq x_t \mid \mathcal{F}_{t-1}).$$

Now we show that the distribution of the observations is characterized by the innovations distribution. For time series of the form

$$X_t = \mu_t + \sigma_t Z_t,$$

we have

$$F_{X_t|\mathcal{F}_{t-1}}(x_t) = P(\mu_t + \sigma_t Z_t \leq x_t \mid \mathcal{F}_{t-1}) \tag{13}$$
$$= P(Z_t \leq (x_t - \mu_t)/\sigma_t \mid \mathcal{F}_{t-1}) \tag{14}$$
$$= F_Z((x_t - \mu_t)/\sigma_t), \tag{15}$$

where F_Z is the distribution function of the innovations $\{Z_t\}_{t\in\mathbb{Z}}$. So we can write the conditional density function of the observations as

$$\begin{aligned} f_{X_t|\mathcal{F}_{t-1}}(x_t) &= \frac{\partial}{\partial x_t} F_{X_t|\mathcal{F}_{t-1}}(x_t) \\ &= \frac{\partial}{\partial x_t} F_Z((x_t - \mu_t)/\sigma_t) \\ &= \frac{1}{\sigma_t} f_Z((x_t - \mu_t)/\sigma_t), \end{aligned}$$

where f_Z is the density function of the innovations. Thus, for a sample of T observations and vector of parameters $\boldsymbol{\theta}$, the conditional likelihood function is

$$\begin{aligned} L_f(\boldsymbol{\theta}) &= \prod_{t=1}^{T} f_{X_t|\mathcal{F}_{t-1}}(x_t) \\ &= \prod_{t=1}^{T} \frac{1}{\sigma_t} f_Z\left(\frac{X_t - \mu_t}{\sigma_t}\right), \end{aligned}$$

where μ_t and σ_t are respectively given by the ARMA and GARCH specification in equations (11) and (12). Yet we require some initial values for the conditional mean, volatility and error term, μ_0, σ_0 and ε_0, since they are unobserved. For μ_0 and σ_0 we use the theoretical unconditional values. That is, we let μ_0 be the sample mean and let σ_0 be the sample standard deviation. In like manner, we let ε_0 be the true mean value of $\{\varepsilon_t\}$, i.e. $\varepsilon_0 = 0$.

We have just seen that the distribution of the innovations characterizes the underlying structure of the model and thus can have a great impact on the fitting and forecasting performance of the model. For our study we assume that the innovations either follow a standard normal (N), standard Student's t (t) or standard skewed Student's t (Skt) distribution. [3] Using the corresponding densities, the conditional log-likelihood function for the innovations are respectively given as follows (Peters 2001).

$$\ell_N(\boldsymbol{\theta}) = -\frac{1}{2}\sum_{t=1}^{T}\left[\ln(2\pi) + \ln(\sigma_t^2) + \frac{(X_t - \mu_t)^2}{\sigma_t^2}\right] \tag{16}$$

$$\ell_t(\boldsymbol{\theta}) = T\ln\left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}\right] - \frac{1}{2}\sum_{t=1}^{T}\left[\ln\sigma_t^2 + (\nu+1)\ln\left(1 + \frac{(X_t - \mu_t)^2}{\sigma_t^2(\nu-2)}\right)\right] \tag{17}$$

[3] This is because of the fact that the **R** package `rmgarch` (v.1.3-0) by Ghalanos 2015*a* only supports these three type of margins for the estimation of copula-MGARCH-type models.

$$\ell_{Skt}(\boldsymbol{\theta}) = T \ln \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right) \ln\left(\frac{2s}{\xi+\frac{1}{\xi}}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)} \right] - \frac{1}{2} \sum_{t=1}^{T} \left[\ln \sigma_t^2 + (\nu+1) \ln \left(1 + \frac{s(X_t - \mu_t) + \sigma_t m}{\sigma_t(\nu-2)} \xi^{-\mathbb{1}_t} \right) \right], \tag{18}$$

where ν is the degrees of freedom (df), Γ is the gamma function, ξ is the asymmetry (or skew) parameter,

$$\begin{aligned} m &= \frac{\Gamma\left(\frac{\nu+1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(\xi - \frac{1}{\xi}\right), \\ s &= \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}, \quad \text{and} \\ \mathbb{1}_t &= \begin{cases} 1 & \text{if } x_t \geq -\frac{\sigma_t m}{s} + \mu_t, \\ -1 & \text{if } x_t < -\frac{\sigma_t m}{s} + \mu_t. \end{cases} \end{aligned}$$

Using equations (16) – (18), the conditional log-likelihood function can be written as

$$\begin{aligned} \ell_f(\boldsymbol{\theta}) &= \ln L_f(\boldsymbol{\theta}) && (19) \\ &= -\sum_{t=1}^{T} \ln(\sigma_t) + \sum_{t=1}^{T} \ln \left[f_Z \left(\frac{X_t - \mu_t}{\sigma_t} \right) \right] && (20) \\ &= -\sum_{t=1}^{T} \ln(\sigma_t) + \ell_j(\boldsymbol{\theta}), && (21) \end{aligned}$$

where $j = \{N, t, Skt\}$. [4] The ML estimate $\hat{\boldsymbol{\theta}}$ is obtained by maximizing the conditional log-likelihood function subject to $\sigma_t > 0$ (and $\nu > 2$ in case of t and skewed t distributed innovations). The ML estimation procedure of GARCH-type models is implemented in the **R** package `rugarch` (v.1.3-1) by Ghalanos (2015*b*).

2.3.3 VaR estimation

From equations (13) – (15) we have

$$F_{X_{t+1}|\mathcal{F}_t}(x) = F_Z((x - \mu_{t+1})/\sigma_{t+1}),$$

and so

$$z_\alpha = F_Z^{-1}(F_{X_{t+1}|\mathcal{F}_t}(x)) = F_Z^{-1}(F_Z((x - \mu_{t+1})/\sigma_{t+1})) = (x - \mu_{t+1})/\sigma_{t+1},$$

[4] Notice that the vector of parameters $\boldsymbol{\theta}$ is different for every other distributed innovations.

where z_α is the α-quantile of the innovations Z_t. By the definition of the VaR, see (4), we can solve this equation for x to obtain the following expression for VaR_α^t:

$$\mathrm{VaR}_\alpha^t = \mu_{t+1} + \sigma_{t+1} z_\alpha.$$

Consequently, this VaR is estimated by

$$\widehat{\mathrm{VaR}}_\alpha^t = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} z_\alpha, \tag{22}$$

where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ are ML estimates following respectively the ARMA and GARCH (or GJR) specification.

If the innovations are assumed to follow a normal, t or skewed t distribution, then the VaR estimate is obtained by (22) with z_α respectively given by

$$z_\alpha = \begin{cases} \Phi^{-1}(\alpha), & \text{normal} \\ t_\nu^{-1}(\alpha), & t \\ Skt_{\nu,\xi}^{-1}(\alpha), & \text{skewed } t \end{cases}$$

where Φ is the standard normal distribution, t_ν is the standard t distribution function with ν df and $Skt_{\nu,\xi}$ is the standard skewed t distribution function with ν df and asymmetry parameter ξ.

2.4 MGARCH model

Multivariate time series models are useful for the understanding of the comovements of multiple time series. The definitions for univariate time series analysis, such as the mean function, covariance function, stationarities, autocorrelation function and white noise processes, are straightforwardly extended to multivariate time series analysis using vectors and matrices.

For financial time series it is likely that the correlation between the time series varies over time. Therefore we consider the *Dynamic Conditional Correlation GARCH* (DCC-GARCH) model, proposed by Engle (2002), that captures the dynamic conditional correlation of the multivariate time series. The DCC-GARCH-type model extends the univariate GARCH-type model of Section 2.3.1 to the multivariate GARCH-type model. Before defining these DCC-GARCH-type models, we introduce the following operator on a (covariance) matrix $\mathbf{\Sigma} \in \mathbb{R}^{n\times n}$ of an n-dimensional random vector $\boldsymbol{X} = (X_1, \ldots, X_n)'$:

$$\varrho(\mathbf{\Sigma}) = \mathrm{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{nn}})^{-1} \cdot \mathbf{\Sigma} \cdot \mathrm{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{nn}})^{-1},$$

where $\mathrm{diag}(\sqrt{\sigma_{11}}, \ldots, \sqrt{\sigma_{nn}})$ is the $(n \times n)$ diagonal matrix of standard deviations of X_i, $i = 1, \ldots, n$. Observe that $\varrho(\mathbf{\Sigma})$ is actually the correlation matrix of $\boldsymbol{X}$.

Definition 2.12. Let $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$ be SWN$(\boldsymbol{0}, \boldsymbol{I}_n)$. The process $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is a DCC(p,q)-ARMA-GARCH process if it is covariance stationary and if it satisfies, for all $t \in \mathbb{Z}$, the equation

$$\boldsymbol{X}_t = \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t^{1/2}\boldsymbol{Z}_t, \tag{23}$$

where $\boldsymbol{\Sigma}_t^{1/2} \in \mathbb{R}^{n\times n}$ is the factor of the $\mathcal{F}_{t-1}$-measurable conditional covariance matrix $\boldsymbol{\Sigma}_t = \Delta_t \boldsymbol{P}_t \Delta_t$ with the conditional correlation matrix $\boldsymbol{P}_t$ satisfying the equation

$$\boldsymbol{P}_t = \varrho\left(\left(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\right)\boldsymbol{P}_c + \sum_{i=1}^{p}\alpha_i \boldsymbol{Y}_{t-i}\boldsymbol{Y}_{t-i}^T + \sum_{j=1}^{q}\beta_j \boldsymbol{P}_{t-j}\right), \tag{24}$$

where $\alpha_i \geq 0$, $i = 1,\ldots,p$, $\beta_j \geq 0$, $j = 1,\ldots,q$, $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1$ and, for $k = 1,\ldots,n$,

1. $\boldsymbol{\mu}_t = (\mu_{1,t},\ldots,\mu_{n,t})'$ is the mean vector with elements $\mu_{k,t}$ satisfying

$$\mu_{k,t} = \mu_k + \sum_{i=1}^{p_{1k}}\phi_{ki}(X_{k,t-i} - \mu_k) + \sum_{j=1}^{q_{1k}}\lambda_{kj}(X_{k,t-j} - \mu_{k,t-j}), \tag{25}$$

2. $\Delta_t = \text{diag}(\sigma_{1,t},\ldots,\sigma_{n,t})$ is the diagonal volatility matrix with elements $\sigma_{k,t}$ satisfying

$$\sigma_{k,t}^2 = \alpha_{k0} + \sum_{i=1}^{p_{2k}}\alpha_{ki}(X_{k,t-i} - \mu_{k,t-i})^2 + \sum_{j=1}^{q_{2k}}\beta_{kj}\sigma_{k,t-j}^2, \tag{26}$$

   where $\alpha_{k0} > 0$, $\alpha_{ki} \geq 0$, $i = 1,\ldots,p_{2k}$, $\beta_{kj} \geq 0$, $j = 1,\ldots,q_{2k}$, and $\sum_{i=1}^{p_{2k}}\alpha_{ki} + \sum_{j=1}^{q_{2k}}\beta_{kj} < 1$,

3. $\boldsymbol{P}_c$ is a constant positive-definite correlation matrix,

4. $\boldsymbol{Y}_t = \Delta_t^{-1}(\boldsymbol{X}_t - \boldsymbol{\mu}_t)$ is the *devolatized process*.

Similarly as for the univariate GARCH models, we should make an assumption about the joint multivariate distribution function of the innovations $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$.

Observe that the devolatized process, $\boldsymbol{Y}_t$, can be written as

$$\boldsymbol{Y}_t = \Delta_t^{-1}\boldsymbol{\Sigma}_t^{1/2}\boldsymbol{Z}_t = \Delta_t^{-1}(\Delta_t \boldsymbol{P}_t^{1/2})\boldsymbol{Z}_t = \boldsymbol{P}_t^{1/2}\boldsymbol{Z}_t,$$

which is a SWN$(\boldsymbol{0}, \boldsymbol{P}_t)$ process.

The DCC-ARMA-GARCH model fits a univariate ARMA-GARCH model to each time series. So the orders p_{1k}, q_{1k}, p_{2k} and q_{2k} in (25) and (26) correspond to the orders of the univariate ARMA(p_{1k}, q_{1k})-GARCH(p_{2k}, q_{2k}) model that is fitted to the k^{th} time series, $k = 1,\ldots,n$. In our case, we have $p_{1k}, q_{1k} \in \{0,1,2\}$ and $p_{2k} =$

$q_{2k} = 1$ for all $k = 1, \ldots, n$. Furthermore, observe that each element of the mean vector $\boldsymbol{\mu}_t$ and the diagonal volatility matrix Δ_t respectively follow the ARMA and GARCH specification of the univariate ARMA-GARCH model, see (8) and (9). The DCC-ARMA-GJR model is defined likewise, where the mean vector $\boldsymbol{\mu}_t$ is defined with elements satisfying (25) and the diagonal volatility matrix Δ_t is instead defined with elements $\sigma_{k,t}$ satisfying

$$\sigma_{k,t}^2 = \alpha_{k0} + \sum_{i=1}^{p_{2k}} \alpha_{ki} \left(|(\boldsymbol{\Sigma}_{t-i}^{1/2} \boldsymbol{Z}_{t-i})_k| - \gamma_{ki} (\boldsymbol{\Sigma}_{t-i}^{1/2} \boldsymbol{Z}_{t-i})_k \right)^2 + \sum_{j=1}^{q_{2k}} \beta_{kj} \sigma_{k,t-j}^2, \quad k = 1, \ldots, n,$$

where $\gamma_{ki} \in [-1, 1]$, $i = 1, \ldots, p_{2k}$, and $\sum_{i=1}^{p_{2k}} \alpha_{ki} \mathrm{E}[(|\tilde{\boldsymbol{Z}}_{k,\bullet}| + \gamma_{ki} \tilde{\boldsymbol{Z}}_{k,\bullet})^2] + \sum_{j=1}^{q_{2k}} \beta_{kj} < 1$ with $\tilde{\boldsymbol{Z}} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_t)$, $k = 1, \ldots, n$.

It is common to use low-order DCC-GARCH-type models, also for the reason of parsimony of the model and interpretability of the variables as in the univariate case. Next to that, it can be computational expensive to estimate a DCC-GARCH-type model for a large number of time series. Therefore we will consider DCC(1, 1)-GARCH-type models, i.e. the DCC(1, 1)-ARMA(p_{1k}, q_{1k})-GARCH(1, 1) and DCC(1, 1)-ARMA(p_{1k}, q_{1k})-GJR(1, 1) models, $p_{1k}, q_{1k} \in \{0, 1, 2\}$, $k = 1, \ldots, n$. Again, the optimal values for p_{1k} and q_{1k} are chosen based on the AIC values.

2.4.1 Parameter estimation

The devolatized process is estimated by

$$\hat{\boldsymbol{Y}}_t = \hat{\Delta}_t^{-1} (\boldsymbol{X}_t - \hat{\boldsymbol{\mu}}_t), \tag{27}$$

where $\hat{\Delta}_t$ and $\hat{\boldsymbol{\mu}}_t$ are obtained from the parameter (ML) estimates of the univariate ARMA-GARCH or ARMA-GJR model. The constant, positive-definite correlation matrix $\boldsymbol{P}_c$ is estimated by the sample correlation of the devolatized process:

$$\hat{\boldsymbol{P}}_c = \mathrm{cor}(\hat{\boldsymbol{Y}}). \tag{28}$$

The remaining parameters of the DCC-GARCH-type model can also be estimated by the ML estimation procedure. The construction of the conditional log-likelihood function is analogous to the univariate case. Namely, for time series of the form

$$\boldsymbol{X}_t = \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{Z}_t$$

we have that the conditional distribution function of the observations can be written as

$$\begin{aligned} F_{\boldsymbol{X}_t | \mathcal{F}_{t-1}} &= P(\boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{Z}_t \leq \boldsymbol{x}_t \mid \mathcal{F}_{t-1}) \\ &= P(\boldsymbol{Z}_t \leq \boldsymbol{\Sigma}_t^{-1/2} (\boldsymbol{x}_t - \boldsymbol{\mu}_t) \mid \mathcal{F}_{t-1}) \\ &= F_{\boldsymbol{Z}} (\boldsymbol{\Sigma}_t^{-1/2} (\boldsymbol{x}_t - \boldsymbol{\mu}_t)) \end{aligned}$$

where $F_{\boldsymbol{Z}}$ is the multivariate distribution function of the innovations $\{\boldsymbol{Z}_t\}_{t\in\mathbb{Z}}$. As a result, the corresponding conditional multivariate density function of the observations can be written as

$$\begin{aligned} f_{\boldsymbol{X}_t|\mathcal{F}_{t-1}}(\boldsymbol{x}_t) &= \frac{\partial}{\partial \boldsymbol{x}_t} F_{\boldsymbol{X}_t|\mathcal{F}_{t-1}}(\boldsymbol{x}_t) \\ &= \frac{\partial}{\partial \boldsymbol{x}_t} F_{\boldsymbol{Z}}(\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu}_t)) \\ &= |\boldsymbol{\Sigma}_t|^{-1/2} f_{\boldsymbol{Z}}(\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu}_t)), \end{aligned}$$

where $f_{\boldsymbol{Z}}$ is the multivariate density function of the innovations and $|\cdot|$ is the determinant. For a sample of T observations and vector of parameters $\boldsymbol{\theta}$, the conditional likelihood function is

$$L_f(\boldsymbol{\theta}) = \prod_{t=1}^{T} f_{\boldsymbol{X}_t|\mathcal{F}_{t-1}}(\boldsymbol{x}_t) \tag{29}$$

$$= \prod_{t=1}^{T} |\boldsymbol{\Sigma}_t|^{-1/2} f_{\boldsymbol{Z}}(\boldsymbol{\Sigma}_t^{-1/2}(\boldsymbol{x}_t - \boldsymbol{\mu}_t)). \tag{30}$$

The conditional log-likelihood function is based on the conditional joint density of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T$, given initial values $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$ and $\boldsymbol{Z}_0$. For $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ we use the theoretical unconditional values. That is, let $\boldsymbol{\mu}_0$ be the vector of the sample mean and let $\boldsymbol{\Sigma}_0$ be the sample covariance matrix. For $\boldsymbol{Z}_0$ we use the true mean value of the innovations, i.e. a vector of zeros $\boldsymbol{Z}_0 = \boldsymbol{0}$.

The ML estimates of the remaining parameters are obtained by maximizing the conditional log-likelihood function subject to $\alpha_i \geq 0$, $i = 1, \ldots, p$, $\beta_j \geq 0$, $j = 1, \ldots, q$, and $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1$. This estimation procedure of MGARCH-type models is implemented in the **R** package `rmgarch` (v.1.3-0) by Ghalanos (2015*a*).

Observe that the variables $\boldsymbol{X}_t, \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t, \boldsymbol{Z}_t$ and $\boldsymbol{P}_t$ are expressed by values from the past. So we can simply plug their past estimates, up to some time t, into the expression of the variables to forecast the values of the variables at time period $t + 1$. [5] This method enables us to perform the out-of-sample forecast of these type of models.

2.4.2 Rank correlations

The disadvantage of most, if not all, MGARCH models is that these models use the *Pearson's linear correlation coefficient* as a standard measure for describing the dependence structure. This has a number of limitations, as described by Embrechts et al. (2001). Instead they suggest to use *rank correlations* for describing the dependence structure.

[5] In fact, in this way we can also forecast future values of time period $t + 2$, $t + 3$, etc.

The two main varieties of rank correlations are the *Kendall's tau* and *Spearman's rho*. As Kendall & Gibbons (1990) argue, the confidence intervals for the Kendall's tau parameters are more reliable and more interpretable than the confidence intervals for the Spearman's rho parameters. For that reason we exclusively consider the Kendall's tau measure.

Two points in $\mathbb{R}^2$, say $(x_1, x_2)'$ and $(y_1, y_2)'$, are concordant if $(x_1 - y_1)(x_2 - y_2) > 0$ and discordant if $(x_1 - y_1)(x_2 - y_2) < 0$. The Kendall's tau is a measure based on the probability of *concordance* and *discordance* of bivariate random vectors. The Kendall's tau, $\rho_\tau \in [-1, 1]$, is defined as follows.

Definition 2.13. For a random vector $(X_1, X_2)'$, the Kendall's tau is

$$\rho_\tau(X_1, X_2) = P((X_1 - Y_1)(X_2 - Y_2) > 0) - P((X_1 - Y_1)(X_2 - Y_2) < 0),$$

where the random vector $(Y_1, Y_2)'$ is an independent copy of $(X_1, X_2)'$. [6]

The Kendall's tau can also be written as an expectation:

$$\rho_\tau(X_1, X_2) = \mathrm{E}\left[\mathrm{sign}((X_1 - Y_1)(X_2 - Y_2))\right],$$

where $\mathrm{sign}(x) = \mathbb{1}_{\{x>0\}} - \mathbb{1}_{\{x<0\}}$. So, consequently, in the multivariate setting we have that the Kendall's tau matrix of a random vector $\boldsymbol{X} = (X_1, \ldots, X_n)'$ can be written as

$$\rho_\tau(\boldsymbol{X}) = \mathrm{cov}(\mathrm{sign}(\boldsymbol{X} - \boldsymbol{Y})),$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ is an independent copy of $\boldsymbol{X}$. Notice that the elements of the random vectors are componentwise connected, i.e. element (i, j) of the Kendall's tau matrix is the Kendall's tau for the random vector $(X_i, X_j)'$:

$$\rho_\tau(\boldsymbol{X})_{i,j} = \rho_\tau(X_i, X_j) = \mathrm{E}\left[\mathrm{sign}((X_i - Y_i)(X_j - Y_j))\right].$$

As we will see later, it is possible to obtain a dynamic conditional Kendall's tau matrix from a dynamic conditional correlation matrix, $\boldsymbol{P}_t$. [7] In that manner the Kendall's tau measure can be used to describe the dependence structure of time series.

2.5 Copulas

A copula is a function that connects univariate marginal distributions (or *margins*) to a multivariate distribution (Bouye et al. 2000). Copula theory gives practitioners the possibility to decompose any multivariate joint distribution into

[6] The random vector $(Y_1, Y_2)'$ is an independent copy in the sense that it has the same distribution as $(X_1, X_2)'$, but the two random vectors are independent.

[7] See Section 4.3.1.

margins and a copula, and vice versa. Moreover, it gives us the possibility to first specify the distribution of each marginal and then choose a copula function to describe the dependence structure between the time series (Aloui et al. 2011). Additionally it allows us to describe more complex multivariate dependence structures, such as non-linear and tail-dependence (Hürlimann 2004).

For the understanding of copula theory, we are required to be familiar with some basics of statistics and statistical operations, such as the *probability* and *quantile transformations*. The basics are presented in the following proposition from Hult et al. (2012), where $\mathcal{U}(a, b)$ denotes the uniform distribution on the interval $[a, b]$.

Proposition 2.1. Let $F : \mathbb{R} \to [0, 1]$ be a distribution function and let F^{-1} be the corresponding quantile function.

1. $u \leq F(x)$ if and only if $F^{-1}(u) \leq x$.
2. $F(F^{-1}(u)) = u$, provided that F is continuous.
3. (Quantile transformation). If $U \sim \mathcal{U}(0, 1)$, then $P(F^{-1}(U) \leq x) = F(x)$.
4. (Probability transformation). If F is a continuous distribution function of a random variable X, then $F(X) \sim \mathcal{U}(0, 1)$.

Definition 2.14. (Ghorbel & Trabelsi 2009) An n-dimensional copula is a multivariate distribution function $C : [0, 1]^n \to [0, 1]$ with standard uniform margins $u_1, \ldots, u_n$, satisfying the following properties:

1. $C(u_1, \ldots, u_n) = 0$ if there exists $u_i = 0$, $i \in \{1, \ldots, n\}$.
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ for all $i \in \{1, \ldots, n\}$, $u_i \in [0, 1]$.
3. C is grounded and n-increasing.

Notice that a copula can only be used with margins that are transformed to standard uniform. There is no mathematical reason for this transformation, but it may be convenient from a statistical point of view (Embrechts 2009). The third condition ensures that for a random vector $\boldsymbol{U} = (U_1, \ldots, U_n)'$ with distribution function C, we have $P(a_1 \leq U_1 \leq b_1, \ldots, a_n \leq U_n \leq b_n) \geq 0$, where $a_i, b_i \in [0, 1]$, $a_i \leq b_i$ for $i = 1, \ldots, n$ (McNeil et al. 2015).

Copulas express the dependence on a quantile scale, which is in particular useful for examining the interdependence of extreme values. This is one of the reason that attracted risk managers to use copulas. The following theorem is due to the seminal work of Sklar (1959).

Theorem 2.1. (Sklar 1959). For any n-dimensional distribution function F with margins $F_1, \ldots, F_n$, there exists an n-dimensional copula C such that

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)), \tag{31}$$

for all $x_1, \ldots, x_n$ in $\bar{\mathbb{R}} \in [-\infty, \infty]$. If $F_1, \ldots, F_n$ are continuous, then C is unique; Otherwise it is uniquely determined on $F_1(\bar{\mathbb{R}}) \times \cdots \times F_n(\bar{\mathbb{R}})$. Conversely, for any n-dimensional copula C and univariate distribution functions $F_1, \ldots, F_n$, the function F in (31) is an n-dimensional distribution function with margins $F_1, \ldots, F_n$.

The first statement of Sklar's Theorem tells us that we can decompose any multivariate distribution function with arbitrary margins into its margins and a copula, i.e.

$$F \Rightarrow C, F_1, \ldots, F_n.$$

The second, converse statement tells us that copulas can be used in combination with arbitrary univariate distribution functions to construct multivariate distributions, i.e.

$$F \Leftarrow C, F_1, \ldots, F_n.$$

In other words, the second statement tells us that we can construct a joint distribution with arbitrary margins, using any copula. Both statements of the Sklar's Theorem are of main importance for our study. Namely, we consider the second statement for the construction of the joint distribution. After obtaining the joint distribution, we perform the in-sample fit test and consider the first statement to estimate the VaR by performing *Monte Carlo* simulations.

The modelling of multivariate distributions using a copula can be done by a 2-step approach:

1. Identify the margins. Choose the distribution for the margins that is most appropriate.
2. Given the identification of the margins, construct a copula to describe the dependence structure

A copula can be seen as a tool to join arbitrary margins to form a joint distribution. It has the consequence that the dependence structure is exclusively determined by the copula, while the structure of the joint distribution is exclusively determined by the margins. On top of that, a copula does also not restrict us to assume that all the data comes from the same type of univariate distribution. A copula may therefore be more suitable for capturing the dependence structure than the traditional multivariate distributions.

2.5.1 Traditional copulas

In the literature we can find a number of different copulas. In this section we introduce two traditional copulas, which perhaps are the most popular ones. [8]

[8] The **R** package `rmgarch` (v.1.3-0) by Ghalanos (2015*a*) only supports these two copulas for the estimation of copula-MGARCH-type model.

Before we introduce these copulas, let us denote the realizations of the random vectors $\boldsymbol{X}$ and $\boldsymbol{U}$ by respectively $\boldsymbol{x} = (x_1, \ldots, x_n)'$ and $\boldsymbol{u} = (u_1, \ldots, u_n)'$. For $i = 1, \ldots, n$, set $F_i(X_i) = U_i$ (or equivalently $X_i = F_i^{-1}(U_i)$, where F_i^{-1} is the quantile function). Then, for a distribution function F_{U_i} of U_i we derive:

$$\begin{aligned} F_{U_i}(u_i) &= P(U_i \leq u_i) \\ &= P(F_i(X_i) \leq u_i) \\ &= P(X_i \leq F_i^{-1}(u_i)) \\ &= F_i(F_i^{-1}(u_i)) \\ &= u_i, \end{aligned}$$

i.e. $u_i \in [0, 1]$ is from the standard uniform distribution. This is the so-called the *Probability-Integral Transformation* (PIT) (Ghorbel & Trabelsi 2009). Also, the density of the copula is defined as follows.

$$c(\boldsymbol{u}) = \frac{\partial^n C(\boldsymbol{u})}{\partial \boldsymbol{u}} = \frac{\partial^n C(u_1, \ldots, u_n)}{\partial u_1 \ldots \partial u_n}.$$

Gaussian copula. (Bouye et al. 2000). Let Φ be the standard normal distribution function and let Φ_P^n be the n-dimensional standardized normal distribution function with correlation matrix $\boldsymbol{P}$ (see Appendix C). [9] The Gaussian copula and its density function are respectively given by

$$C^{Ga}(\boldsymbol{u}; \boldsymbol{P}) = \Phi_P^n(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)) \tag{32}$$

$$= P(X_1 \leq \Phi^{-1}(u_1), \ldots, X_n \leq \Phi^{-1}(u_n)), \tag{33}$$

$$c^{Ga}(\boldsymbol{u}; \boldsymbol{P}) = |\boldsymbol{P}|^{-1/2} \exp\left\{-\frac{1}{2}\tilde{\boldsymbol{u}}^T(\boldsymbol{P}^{-1} - \boldsymbol{I}_n)\tilde{\boldsymbol{u}}\right\}, \tag{34}$$

where $\tilde{\boldsymbol{u}} = (\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n))'$ and $\boldsymbol{I}_n$ is the $(n \times n)$ identity matrix.

The Gaussian copula belongs to the family of *elliptical* copulas (Jondeau & Rockinger 2006). As we can see, the Gaussian copula is closely related to the multivariate normal distribution. In fact, if each marginal is assumed to follow a univariate standard normal distribution, then the Gaussian copula coupled with the normal margins is identical to the multivariate normal distribution.

Empirically the financial loss distribution show fatter tails than the normal distributions. To overcome this problem we also consider the following copula.

t **copula** (Bouye et al. 2000). Let t_ν be the standard t distribution function with ν df and let $t_{\boldsymbol{P},\nu}^n$ be the n-dimensional standardized t distribution function with ν df and correlation matrix $\boldsymbol{P}$ (see Appendix C). The t copula and its density

[9] The element (i, j) of the correlation matrix $\boldsymbol{P}$, say $\rho_{i,j} \in [-1, 1]$, is the Pearson's linear correlation coefficient between the i^{th} and j^{th} time series, $i, j \leq n$.

function are respectively given by

$$C^t(\boldsymbol{u}; \boldsymbol{P}, \nu) = t^n_{\boldsymbol{P},\nu}(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_n)) \tag{35}$$

$$= P(X_1 \leq t_\nu^{-1}(u_1), \ldots, X_n \leq t_\nu^{-1}(u_1)), \tag{36}$$

$$c^t(\boldsymbol{u}; \boldsymbol{P}, \nu) = |\boldsymbol{P}|^{-1/2} \frac{\Gamma(\frac{\nu+n}{2})\Gamma(\frac{\nu}{2})^n}{\Gamma(\frac{\nu+1}{2})^n \Gamma(\frac{\nu}{2})} \cdot \frac{(1 + \frac{1}{\nu}\tilde{\boldsymbol{u}}^T \boldsymbol{P}^{-1}\tilde{\boldsymbol{u}})^{-\frac{\nu+n}{2}}}{\prod_{i=1}^n \left(1 + \frac{\tilde{u}_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}}, \tag{37}$$

where $\tilde{\boldsymbol{u}} = (\tilde{u}_1, \ldots, \tilde{u}_n)' = (t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_n))'$.

Also the t copula belongs to the family of elliptical copulas (Jondeau & Rockinger 2006). As we can see, the t copula is closely related to the multivariate t distribution. In fact, if each marginal is assumed to follow a univariate standard t distribution with df ν (so all margins have the same df ν), then the t copula coupled with the t margins is identical to the multivariate t distribution.

2.5.2 Meta-copula distributions

According to the second statement of Sklar's Theorem, we can construct a joint distribution with arbitrary margins and any copula. To see that the margins can be arbitrary, observe from Section 2.5.1 that for the Gaussian copula we can write $\tilde{\boldsymbol{u}} = (\Phi^{-1}(F_1(x_1)), \ldots, \Phi^{-1}(F_n(x_n)))'$. This can be done in a similar manner for any other copula, such as the t copula.

For a copula C and arbitrary margins $F_1, \ldots, F_n$, the multivariate distribution function $F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$ is called a *meta-copula* distribution. For example, the multivariate distribution function constructed from a 3-dimensional Gaussian copula with arbitrary margins, such as the normal, t and skewed t marginal distribution, is called a meta-Gaussian distribution. These type of distributions are of interest for our study. Let us illustrate the idea behind copulas by the following example.

Example 2.1. In this example we will illustrate how any multivariate distribution function with arbitrary margins can be decomposed into its margins and a copula. Furthermore, we will also see how copulas can be used in combination with arbitrary univariate distribution functions to construct multivariate distributions.

Let us denote a bivariate random vector by $\boldsymbol{X} = (X_1, X_2)'$. Moreover, assume that X_1 and X_2 are correlated, without loss of generality assume positively correlated. Furthermore, assume that $X_1 \sim F_1$, $X_2 \sim F_2$ and that they are jointly normally distributed.

The first step in copula modelling is to transform the realizations of X_1 and X_2 to uniform using the arbitrary (or rather: desired) distribution functions for the

margins, F_1 and F_2:

$$u_1 = F_1(x_1) \quad \text{and} \quad u_2 = F_2(x_2).$$

After this transformation the uniform margins still hold the positive correlation of X_1 and X_2. As a second step we perform another transformation, which is required to model the joint distribution of X_1 and X_2. Since we assumed the random variables to be jointly normally distributed, we use the standard normal cumulative distribution function for this transformation:

$$\tilde{u}_1 = \Phi^{-1}(u_1) = \Phi^{-1}(F_1(x_1)) \quad \text{and} \quad \tilde{u}_2 = \Phi^{-1}(u_2) = \Phi^{-1}(F_2(x_2)).$$

By this step we obtained (transformed) realizations from a joint normal distribution with arbitrary margins. Performing the two transformations consecutively has the consequence that the rank order of the variables is maintained. As a result, this complete procedure does not alter the original dependence structure. Accordingly, these realizations can be fitted to the bivariate normal distribution. This distribution is in fact the bivariate Gaussian copula, also known as the meta-Gaussian distribution. This shows that any multivariate distribution function with arbitrary margins can be decomposed into its margins and a copula.

By going the opposite direction of the procedure above, we can transform the realizations $\tilde{u}_1$ and $\tilde{u}_2$ from the bivariate Gaussian copula to (transformed) realizations from a multivariate distribution, without altering the original dependence structure. This is also done by using the standard normal cumulative distribution function and arbitrary distribution functions F_1 and F_2:

$$x_1 = F_1^{-1}(\Phi(\tilde{u}_1)) \quad \text{and} \quad x_2 = F_2^{-1}(\Phi(\tilde{u}_2)).$$

Accordingly, these realizations can be fitted to a multivariate distribution. This shows that copulas can be used in combination with arbitrary univariate distribution functions to construct multivariate distributions.

■

2.5.3 Parameter estimation

The parameters of the copulas can be estimated by the ML estimation procedure. For observations $\boldsymbol{x} = (x_1, \ldots, x_n)'$, margins $F_1, \ldots, F_n$ and $\boldsymbol{u} = (u_1, \ldots, u_n)' = (F_1(x_1), \ldots, F_n(x_n))'$, the log-likelihood function of the copula is given by

$$\ell_c(\boldsymbol{\phi}, \boldsymbol{\theta}) = \ln c(\boldsymbol{u}), \tag{38}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are the vectors of respectively the copula and marginal parameters. In our case we have that $c(\boldsymbol{u})$ is given by either $c(\boldsymbol{u}) = c^{Ga}(\boldsymbol{u}; \boldsymbol{P})$ (see (34)) or $c(\boldsymbol{u}) = c^{t}(\boldsymbol{u}; \boldsymbol{P}, \nu)$ (see (37)).

In this thesis we make the assumption that the regularity conditions for asymptotic ML theory are satisfied for the multivariate model (i.e. copula) and its margins. Under these regularity conditions, the MLE exists and moreover it is consistent and asymptotically efficient (Huang et al. 2009).

The estimation of the copula in this static setting is not of our most interest, although it is useful to get an insight about the estimation of the copula at a certain stage in time. The estimation of the copula that is of our interest is explained in Section 2.6, where the dynamic setting is employed by the MGARCH specification.

2.5.4 tail-dependence

Correlation coefficients do not focus on the *tail-dependence*, or *extreme comovements*, between two univariate data sets. In order to describe this dependency, we introduce *tail-dependence coefficients*. tail-dependence coefficients quantify the dependence in the upper and lower tails of a bivariate distribution with continuous margins. More explicitly, in terms of losses, the upper (respectively lower) tail-dependence coefficient describes to what extent large drops (respectively rises) of one time series is related to large drops (respectively rises) of another time series. As a result it can be helpful in the financial market to measure the tendency of markets to crash or boost together (Aloui et al. 2011). Conveniently, copulas could be used to estimate these coefficients.

Mathematically, the upper tail-dependence coefficient is the probability that a random variable X_2 exceeds its q-quantile, given that X_1 exceeds its q-quantile, where q is considered as a limit going to one.

Definition 2.15. Let $X_1 \sim F_1$ and $X_2 \sim F_2$ be random variables. The *upper tail-dependence coefficient* of X_1 and X_2 is given by

$$\lambda_u(X_1, X_2) = \lim_{q \to 1^-} P(X_2 > F_2^{-1}(q) \mid X_1 > F_1^{-1}(q)),$$

provided that the limit exists.

If $\lambda_u = 0$, then X_1 and X_2 are said to be *asymptotically independent* in the upper tail. If $\lambda_u \in (0, 1]$, then X_1 and X_2 are said to exhibit upper tail-dependence. If X_1 and X_2 represent the losses (instead of returns) of two different time series, then the upper tail-dependence coefficient is used to measure the tendency of the two time series to crash together. The higher the value of λ_u, the stronger this extremal dependence.

In a similar way we can define the lower tail-dependence coefficient, where in this case q is considered as a limit going to zero.

$$\lambda_l(X_1, X_2) = \lim_{q \to 0^+} P(X_2 \leq F_2^{-1}(q) \mid X_1 \leq F_1^{-1}(q)),$$

provided that the limit exists.

Likewise, if X_1 and X_2 represent the losses, then the lower tail-dependence coefficient is used to measure the tendency of two time series to boost together.

If the distribution functions of the random variables are continuous, we can express the upper and lower tail-dependence coefficients in terms of copulas. Before we derive these expressions, let us introduce *survival copulas*. For a random vector $\boldsymbol{X}$ with multivariate survival function $\bar{F}$, marginal survival functions $\bar{F}_i = 1 - F_i$, $i = 1, \ldots, n$, we have

$$\bar{F}(x_1, \ldots, x_n) = \bar{C}(\bar{F}_1(x_1), \ldots, \bar{F}_n(x_n)), \tag{39}$$

where $\bar{C}$ is called the survival copula. Notice that this is in fact the Sklar's Theorem applied to survival functions.

Using (31) and (39), we derive

$$\begin{aligned}
\lambda_u &= \lim_{q \to 1^-} \frac{P(X_2 > F_2^{-1}(q)\ ,\ X_1 > F_1^{-1}(q))}{P(X_1 > F_1^{-1}(q))} \\
&= \lim_{q \to 1^-} \frac{P(1 - X_2 \leq 1 - F_2^{-1}(q)\ ,\ 1 - X_1 \leq 1 - F_1^{-1}(q))}{P(1 - X_1 \leq 1 - F_1^{-1}(q))} \\
&= \lim_{q \to 1^-} \frac{\bar{C}(1-q, 1-q)}{1-q}, \qquad (40) \\
\lambda_l &= \lim_{q \to 0^+} \frac{P(X_2 \leq F_2^{-1}(q)\ ,\ X_1 \leq F_1^{-1}(q))}{P(X_1 \leq F_1^{-1}(q))} \\
&= \lim_{q \to 0^+} \frac{C(q, q)}{q}. \qquad (41)
\end{aligned}$$

McNeil et al. (2015) show that the Gaussian copula is asymptotically independent in both tails (i.e. $\lambda_u = \lambda_l = 0$), while the t copula exhibit both upper and lower tail-dependence. Put differently, the Gaussian copula does not exhibit tail-dependence, while the t copula does. Due to radial symmetry of the t copula, its dependence coefficients are equal (i.e. $\lambda_u = \lambda_l \in (0, 1]$). [10] Moreover, in case of negative or no dependence, the two tail-dependence coefficients are (almost) zero. In case of positive dependence, the value of the tail-dependence coefficients depends on the copula. In this thesis we let the tail index (TI) denote the value of the upper/lower tail-dependence coefficient, as these two tail-dependence coefficients are always equal in our case.

[10] Let copula C be the distribution function of the n-dimensional random vector $\boldsymbol{U}$ and denote $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^n$. Then C is said to be *radially symmetric* if $\boldsymbol{U} \overset{d}{=} \mathbf{1} - \boldsymbol{U}$.

2.5.5 Exploratory analysis

Generally it is a major challenge to analyse a large amount of information. It is convenient to provide visualizations of large datasets in order to study these datasets. In this section we will shortly describe how we perform an exploratory analysis in copula theory.

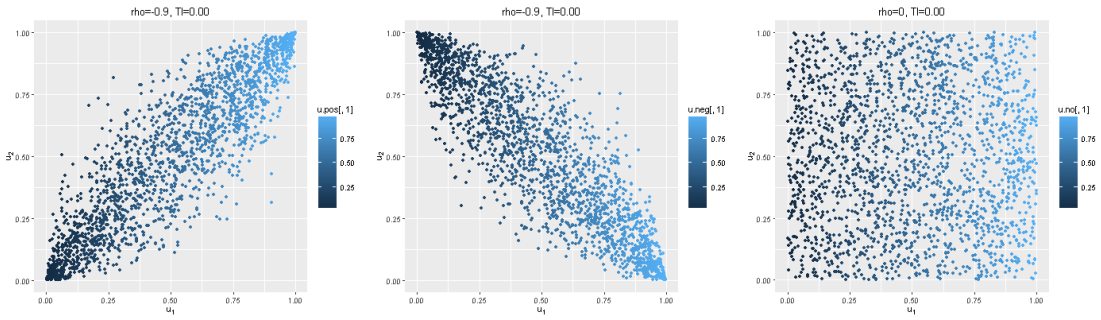
As an example we consider 2,000 simulated random values from bivariate Gaussian and t copulas with different Pearson's correlation coefficient ρ, and additionally in the case of t copula, different dfs. More explicitly, we consider $\rho = 0.9$ (positive dependence), -0.9 (negative dependence) and 0 (no dependence) and df of 2 (fat-tailed) and 10 (thin-tailed).

To start with, we analyse a common visualization tool in copula theory (and data analysis in general), namely the 2D scatter plot. A scatter plot displays values for a pair of variables of a dataset. Such a plot can be useful for the study of the dependence properties between the variables. The scatter plots of the simulated random values from the copulas are shown in Figure 2, where the ρ and df (if applicable) of the corresponding copula and TI are indicated on top of each scatter plot. We identify the following from the scatter plot.

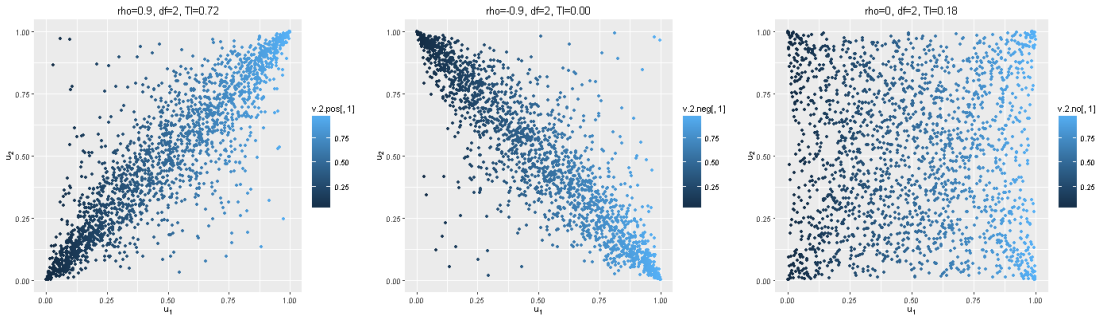
- For positive dependence we see the tendency of small (respectively large) values of one variable to be related to small (respectively large) values of the other variable. For negative dependence this relation is vice versa. For no dependence the paired values lie independently in the unit square. This is in accordance with the definition of the dependencies. However, the simulated values from the two t copulas have some tendency to lie in the corners of the unit square.

- The scatter plots of the t copula with the largest df, say the t_{10} copula, is almost similar to the ones (with the same ρ) of the Gaussian copula. This is in accordance with the fact that the t distribution approaches the Gaussian distribution as the df increases.

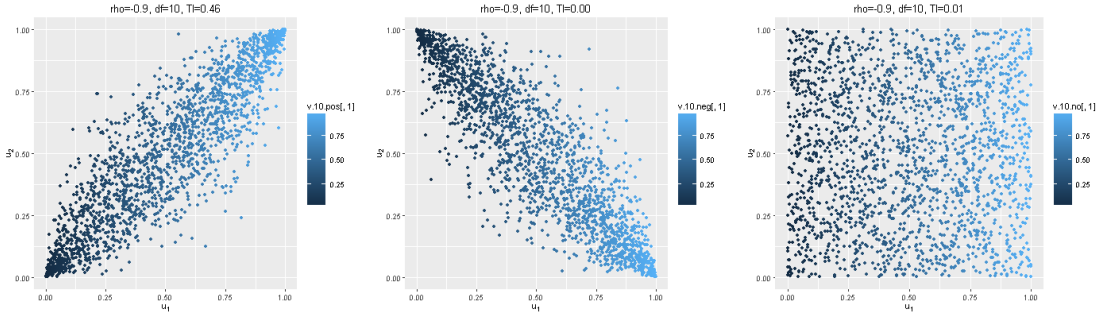
- The dots in the scatter plots of the t copula with the lowest df have a greater tendency to lie close to the corners of the unit square than the ones from t_{10} copula (with the same ρ). This indicates that the simulated values from the t_2 copula have a stronger tail-dependence (i.e. larger TIs) than the ones from the t_{10} copula. This is in accordance with the fact that the t distribution has fatter tails for lower df.



(a) Scatter plots of simulations from the Gaussian copula with $\rho = 0.9$ (left), -0.9 (middle) and 0 (right).



(b) Scatter plots of simulations from the t copula with 2 df and $\rho = 0.9$ (left), -0.9 (middle) and 0 (right).



(c) Scatter plots of simulations from the t copula with 10 df and $\rho = 0.9$ (left), -0.9 (middle) and 0 (right).

Figure 2

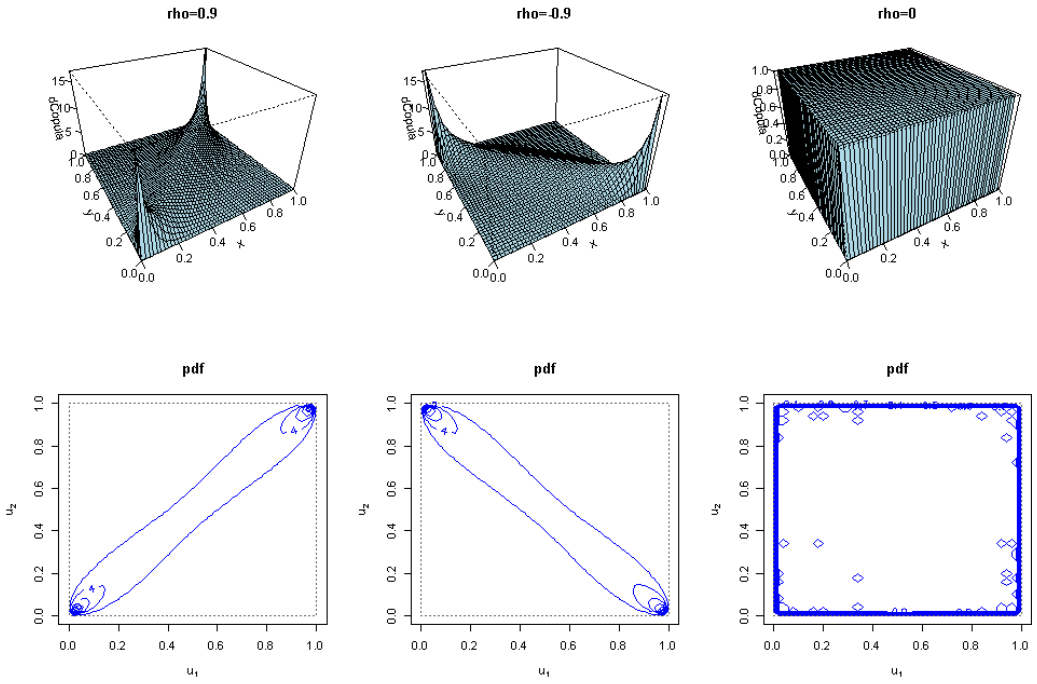
It may be difficult to see from the scatter plots how many dots are (or how much mass is) focussed around a certain area of the unit square, especially at dense areas such as the corners. By this way it can therefore be challenging to determine the fatness of the tails of the distribution of the simulated values. Density and contour plots provide a solution to this issue. The density and contour plots of the simulated values from the copulas are shown in Figure 3. We identify the following from these plots.

- The structure of the density and contour plots are conform to the structure of the corresponding scatter plots. For example, in the case of positive dependence, most mass covers the diagonal with positive slope in the

(u_1, u_2)-plane. An analogous argument applies for the case of negative dependence.

- The density of the simulated values from the Gaussian copula with $\rho = 0$ corresponds to that of the uniform distribution. This can be explained by fact that the Gaussian distribution has thin tails and so in the case of no dependence there is not tendency of the dots to be focussed in the tails. On the other hand, the density of the simulated values from the t_2 and t_{10} copulas with $\rho = 0$ show existence of tails. This can be explained by fact that the t distribution has fatter tails.

- The tails of the density plots exhibit most mass for the simulated values from the t_2 copula, followed by the t_{10} copula and then the Gaussian copula (with the same ρ). [11] This is in accordance with the fact that the t distribution has fatter tails for lower df.

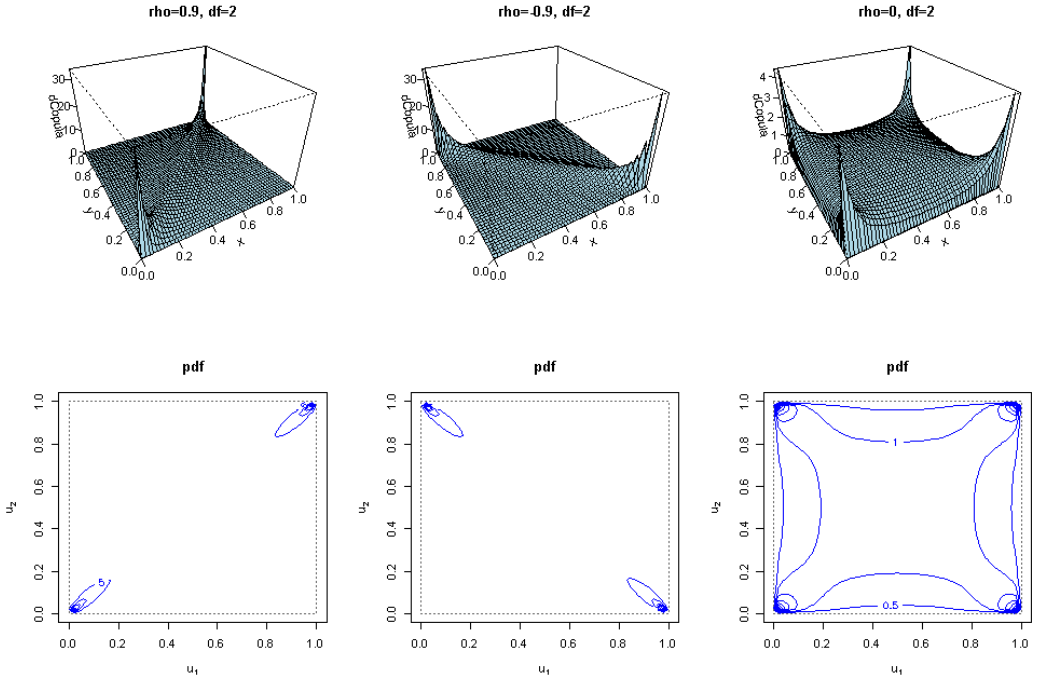


(a) Density (above) and contour (below) plots of simulations from the Gaussian copula with $\rho = 0.9$ (left), -0.9 (middle) and 0 (right).

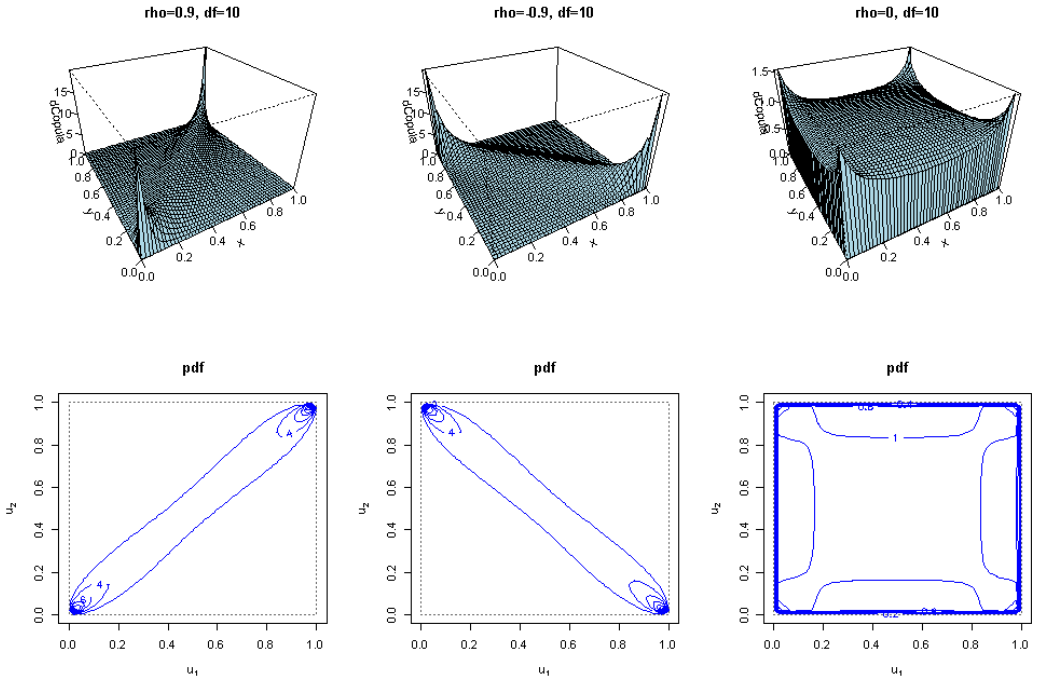
2.6 Copula-MGARCH model

In this section we will propose the hybrid models, which are combinations of MGARCH models and copulas. These copula-MGARCH models are defined

[11] Take note that the range of the density functions are different.



(b) Density (above) and contour (below) plots of simulations from the t copula with 2 df and $\rho = 0.9$ (left), -0.9 (middle) and 0 (right).



(c) Density (above) and contour (below) plots of simulations from the t copula with 10 df and $\rho = 0.9$ (left), -0.9 (middle) and 0 (right).

Figure 3

similarly as the (traditional) MGARCH models (see Definition 2.12), but where the multivariate distribution function of the innovations is a copula function. By doing so we allow for different types of univariate distributions for the marginal innovations. Therefore we are able to propose a number of these models, each with a different underlying structure.

The copula theory is defined in the static setting. By considering the GARCH-type models, we have to deal with the dynamic setting. The copula theory in Section 2.5 can be adjusted straightforwardly be changing the subscript 'i' into 'i,t' for the variables. That is, for time period $t \in \{1,\dots,T\}$, denote the realizations of the random vectors $\boldsymbol{X}_t$ and $\boldsymbol{U}_t$ by respectively $\boldsymbol{x}_t = (x_{1,t},\dots,x_{n,t})'$ and $\boldsymbol{u}_t = (u_{1,t},\dots,u_{n,t})'$. For $i = 1,\dots,n$, set $F_i(X_{i,t}) = U_{i,t}$ (or equivalently $X_{i,t} = F_i^{-1}(U_{i,t})$, where F_i^{-1} is the quantile function) and so $u_{i,t} \in [0,1]$ is from the standard uniform distribution.

Patton (2006) defined the *conditional copulas* and introduced the conditional extension of the Sklar's Theorem. This extension allows us to combine the copula with the MGARCH model.

Theorem 2.2. (Patton 2006). Let $\mathcal{F}_{t-1}$ be the information that is available at time period $t-1$. For any n-dimensional conditional distribution function $F(\cdot|\mathcal{F}_{t-1})$ with conditional margins $F_1(\cdot|\mathcal{F}_{t-1}),\dots,F_n(\cdot|\mathcal{F}_{t-1})$, there exists an n-dimensional conditional copula $C_t(\cdot|\mathcal{F}_{t-1})$ such that

$$F_t(x_{1,t},\dots,x_{n,t} \mid \mathcal{F}_{t-1}) = C_t(F_1(x_{1,t} \mid \mathcal{F}_{t-1}),\dots,F_n(x_{n,t} \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}), \tag{42}$$

for all $x_{1,t},\dots,x_{n,t}$ in $\bar{\mathbb{R}} \in [-\infty,\infty]$. If $F_1(\cdot|\mathcal{F}_{t-1}),\dots,F_n(\cdot|\mathcal{F}_{t-1})$ are continuous, then $C_t(\cdot|\mathcal{F}_{t-1})$ is unique; Otherwise it is uniquely determined on $F_1(\bar{\mathbb{R}} \mid \mathcal{F}_{t-1}) \times \cdots \times F_n(\bar{\mathbb{R}} \mid \mathcal{F}_{t-1})$. Conversely, at time period t, for any n-dimensional conditional copula $C_t(\cdot|\mathcal{F}_{t-1})$ and conditional univariate distribution functions $F_1(\cdot|\mathcal{F}_{t-1}),\dots,F_n(\cdot|\mathcal{F}_{t-1})$, the conditional function $F(\cdot|\mathcal{F}_{t-1})$ in (42) is an n-dimensional conditional distribution function with conditional margins $F_1(\cdot|\mathcal{F}_{t-1}),\dots,F_n(\cdot|\mathcal{F}_{t-1})$.

2.6.1 Traditional conditional copulas

For the conditional copula we consider a time-varying correlation matrix $\boldsymbol{P}_t$ instead of the constant correlation matrix $\boldsymbol{P}$ of the (unconditional) copula. For the copula-DCC-GARCH models, $\boldsymbol{P}_t$ is the dynamic conditional correlation matrix following the DCC-GARCH model specification (see (24)).

For the Gaussian copula we have that the conditional density function is given by

$$c_t^{Ga}(\boldsymbol{u}_t|\boldsymbol{P}_t) = |\boldsymbol{P}_t|^{-1/2} \exp\left\{-\frac{1}{2}\tilde{\boldsymbol{u}}_t{}^T(\boldsymbol{P}_t^{-1} - \boldsymbol{I}_n)\tilde{\boldsymbol{u}}_t\right\}, \tag{43}$$

where $\tilde{\boldsymbol{u}}_t = (\Phi^{-1}(u_{1,t}), \ldots, \Phi^{-1}(u_{n,t}))'$ with $u_{i,t} = F_i(x_{i,t} \mid \mathcal{F}_{t-1})$, $i = 1, \ldots, n$, and $\boldsymbol{I}_n$ is the $(n \times n)$ identity matrix.

For the t copula we have that the conditional density function is given by

$$c_t^t(\boldsymbol{u}_t|\boldsymbol{P}_t, \nu) = |\boldsymbol{P}_t|^{-1/2} \frac{\Gamma(\frac{\nu+n}{2})\Gamma(\frac{\nu}{2})^n}{\Gamma(\frac{\nu+1}{2})^n \Gamma(\frac{\nu}{2})} \cdot \frac{(1 + \frac{1}{\nu}\tilde{\boldsymbol{u}}_t{}^T \boldsymbol{P}_t^{-1} \tilde{\boldsymbol{u}}_t)^{-\frac{\nu+n}{2}}}{\prod_{i=1}^n \left(1 + \frac{\tilde{u}_{i,t}^2}{\nu}\right)^{-\frac{\nu+1}{2}}}, \tag{44}$$

where $\tilde{\boldsymbol{u}}_t = (\tilde{u}_{1,t}, \ldots, \tilde{u}_{n,t})' = (t_\nu^{-1}(u_{1,t}), \ldots, t_\nu^{-1}(u_{n,t}))'$ with $u_{i,t} = F_i(x_{i,t} \mid \mathcal{F}_{t-1})$, $i = 1, \ldots, n$.

2.6.2 Parameter estimation

For convenience, we will from now on write the conditional joint distribution function as F_t instead of $F_t(\cdot|\mathcal{F}_{t-1})$, and the conditional margins as F_i instead of $F_i(\cdot|\mathcal{F}_{t-1})$, $i = 1, \ldots, n$.

For an n-times differentiable conditional joint distribution function F, the corresponding conditional joint density function at time period t is given by

$$f_t(x_{1,t}, \ldots, x_{n,t} \mid \mathcal{F}_{t-1}) = \frac{\partial^n F_t(x_{1,t}, \ldots, x_{n,t} \mid \mathcal{F}_{t-1})}{\partial x_{1,t} \ldots \partial x_{n,t}}. \tag{45}$$

Using the conditional extension of Sklar's Theorem we can write the conditional joint density in (45) as

$$\begin{aligned} f_t(x_{1,t}, \ldots, x_{n,t} \mid \mathcal{F}_{t-1}) &= f_t(F_1^{-1}(u_{1,t} \mid \mathcal{F}_{t-1}), \ldots, F_n^{-1}(u_{n,t} \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}) && (46) \\ &= \frac{\partial^n C_t(u_{1,t}, \ldots, u_{n,t} \mid \mathcal{F}_{t-1})}{\partial u_{1,t} \ldots \partial u_{n,t}} \cdot \prod_{i=1}^n f_i(F_i^{-1}(u_{i,t} \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}) && (47) \\ &= c_t(u_{1,t}, \ldots, u_{n,t} \mid \mathcal{F}_{t-1}) \cdot \prod_{i=1}^n f_i(F_i^{-1}(u_{i,t} \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}), && (48) \end{aligned}$$

where $f_i = \frac{\partial F_i(x_i|\mathcal{F}_{t-1})}{\partial x_i}$ is the conditional marginal density function of the i^{th} sample. Recall that for the conditional margins $F_1, \ldots, F_n$ we have $F_i(x_{i,t} \mid \mathcal{F}_{t-1}) = u_{i,t}$, $i = 1, \ldots, n$. In other words, for a normally distributed margin F_i we have $u_{i,t} = F_i(x_{i,t} \mid \mu_{i,t}, \sigma_{i,t})$, while for the (skewed) t distributed margin we additionally have conditional on ν_i (and ξ_i). [12]

Considering all observations, $t = 1, \ldots, T$, and by taking the natural logarithm of both sides in equation (48), we obtain the conditional log-likelihood function

[12] Notice that $\mu_{i,t}, \sigma_{i,t}, \nu_i, \xi_i \in \mathcal{F}_{t-1}$, because $\mu_{i,t}$, $\sigma_{i,t}$, ν_i and ξ_i (for the (skewed) t distributed margins) are obtained from the estimates of the data up until time period $t-1$. For the same reason, in the multivariate setting (e.g. joint distribution and copulas) we have $\boldsymbol{P}_t, \boldsymbol{\Sigma}_t, \nu \in \mathcal{F}_{t-1}$.

of the conditional joint distribution:

$$\ell(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{t=1}^{T} \ln c_t(u_{1,t}, \ldots, u_{n,t} \mid \mathcal{F}_{t-1}) + \sum_{t=1}^{T} \sum_{i=1}^{n} \ln f_i(F_i^{-1}(u_{i,t} \mid \mathcal{F}_{t-1}) \mid \mathcal{F}_{t-1}) \tag{49}$$

$$\overset{13}{=} \sum_{t=1}^{T} \ell_{c_t}(\boldsymbol{\phi}_t, \boldsymbol{\theta}) + \sum_{i=1}^{n} \sum_{t=1}^{T} \ell_{f_i}(\boldsymbol{\theta}_{i,t}), \tag{50}$$

where $\boldsymbol{\phi} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_T)'$ is the vector of the copula parameters and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T)'$ is the vector with elements of the marginal parameters $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_{1,t}, \ldots, \boldsymbol{\theta}_{n,t})'$, $t = 1, \ldots, T$. The parameters that are included in $\boldsymbol{\theta}_{i,t}$ depend on the assumption that we make about the innovations distribution, e.g. $\boldsymbol{\theta}_{i,t}$ includes ν_i for the t distribution, while it includes ν_i and ξ_i for the skewed t distribution.

As we can see in equation (50), the conditional log-likelihood function of the joint distribution can be decomposed into a margins part and a copula part. The decomposition of the log-likelihood function allows us to perform the *method of inference function for margins* (IFM) (Liu & Luger 2009). This is a 2-step estimation method that proceeds as follows.

1. Maximize the conditional log-likelihood function of the i^{th} margin at time t, $\sum_{i=1}^{T} \ell_{f_i}(\boldsymbol{\theta}_{i,t})$, to obtain a ML estimate of the marginal parameters, $\boldsymbol{\theta}_i$, for $i = 1, \ldots, n$. See equations (16) – (21).

2. Given the ML estimate $\hat{\boldsymbol{\theta}}$ of the marginal parameters from Step 1, maximize the conditional log-likelihood function of the copula, $\sum_{t=1}^{T} \ell_{c_t}(\boldsymbol{\phi}, \hat{\boldsymbol{\theta}})$, to obtain a ML estimate of the copula parameters, $\boldsymbol{\phi}$. See equations (38), (43) and (44).

As explained by Liu & Luger (2009), this 2-step procedure is less computational intensive than the 1-step estimation procedure of estimating all parameters, $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, simultaneously. This makes the estimation also feasible for a large number of time series, such as twenty in our case. This method is implemented in the **R** package `rmgarch` (v.1.3-0) by Ghalanos (2015*a*).

2.6.3 Methodology

So far we have combined the theory of GARCH models and copulas to formulate the theory of copula-MGARCH models. In this section we will provide a more practical view of estimating these models. The estimation procedure of the copula-DCC(1, 1)-ARMA-GARCH(1, 1) model is described by the following steps.

[13] Since $\sum_{t=1}^{T} \sum_{i=1}^{n} |\ell_{f_i}(\boldsymbol{\theta}_{i,t})| < \infty$, we have $\sum_{t=1}^{T} \sum_{i=1}^{n} \ell_{f_i}(\boldsymbol{\theta}_{i,t}) = \sum_{i=1}^{n} \sum_{t=1}^{T} \ell_{f_i}(\boldsymbol{\theta}_{i,t})$.

1. For the i^{th} time series, $i = 1, \ldots, n$, estimate a univariate ARMA(p_i, q_i)-GARCH$(1,1)$ model, where for the model we make the assumption that the innovations either follow a normal, t or skewed t distribution. Optimal values for $p_i, q_i \in \{0,1,2\}$ and the most appropriate innovations distribution are chosen based on the AIC values: we choose the combination of the pair (p_i, q_i) and the innovations distribution that yields the lowest AIC value (and hence is assumed to provide the optimal fit of the i^{th} time series). The most appropriate innovations distribution will be used for the construction of the copula margins in Step 3. Recall that margins do not all have to be of the same class.

2. Transform the innovations of the fitted univariate models from Step 1 to uniform $(0,1)$ using PIT, i.e. F_i(innovations).

3. Fit the copula to the transformed innovations by IFM and consequently estimate the remaining parameters of the copula-DCC$(1,1)$-ARMA-GARCH$(1,1)$ model by ML (see (30)), where the joint distribution of the innovations $f_{\mathbf{Z}}$ follows a meta-Gaussian or meta-t distribution with the margins and transformed innovations being obtained in Steps 1 and 2 respectively.

By this procedure we basically model the multivariate distribution function of the innovations by a copula. A very similar description of the procedure applies for the copula-DCC$(1,1)$-ARMA-GJR$(1,1)$ model.

The models that we will consider and their abbreviations are given in Table 1. Recall that the margins (or rather their innovations) of the copula-MGARCH-type models will either each be assumed to follow a normal, t or skewed t distribution and also the margins do not necessarily have to be of the same class. The distribution assumption for each margin will be determined in Section 4.

Table 1: All the models considered for the study.

Model	Innovations distribution	MGARCH-type	Margins
Ga-GARCH	meta-Gaussian	DCC-ARMA-GARCH	N, t and/or Skt
Ga-GJR	meta-Gaussian	DCC-ARMA-GJR	N, t and/or Skt
t-GARCH	meta-t	DCC-ARMA-GARCH	N, t and/or Skt
t-GJR	meta-t	DCC-ARMA-GJR	N, t and/or Skt
GARCH-N	Multivariate normal	DCC-ARMA-GARCH	N
GJR-N	Multivariate normal	DCC-ARMA-GJR	N
GARCH-t	Multivariate t	DCC-ARMA-GARCH	t
GJR-t	Multivariate t	DCC-ARMA-GJR	t

The performance of each of the eight models is evaluated by the in-sample fit test, that is based on four different information criteria, and the out-of-sample forecast test, that is based on VaR estimates and statistical loss functions. The models that are optimal, according to the in-sample and out-of-sample test criteria, will be used to describe the dependence structure of the data. Observe

from the table that we will also consider the traditional DCC-GARCH-type models, i.e. without copulas modelling. By considering these traditional MGARCH-type models we can check how well the copula-MGARCH-type models perform relative to these traditional models and, more importantly, whether or not the copula modelling improves the fit and forecast (according to the criteria of the test procedures).

3 Data description

We analyse the dependence structure of the daily losses of twenty different stock index time series worldwide, where the daily losses of each stock index is calculated in terms of percentage differenced logarithmic value series as follows:

$$l_t = -\ln\left(\frac{P_t}{P_{t-1}}\right) \cdot 100, \tag{51}$$

with P_t denoting the daily adjusted closing price of a stock index at time t. [14] For every stock index, Yahoo! Finance has provided us with daily adjusted closing prices from January 1, 2000 to July 31, 2016. See Table 2 for the descriptive statistics of the losses for each stock index. We observe that all loss distributions are skewed and most exhibit *excess kurtosis*. [15] This indicates that the losses are non-normally distributed.

In the remainder of this section we will perform visual and numerical tests on the losses of each stock index. To illustrate the idea behind these tests, we will only discuss the results of one stock index, namely the *Euronext Amsterdam* (AEX). The analysis of the remaining stock indices apply in a similar manner.

3.1 Visual tests

We start by investigating the overall structure of the loss distribution of the AEX using a histogram. A histogram namely represents the distribution of losses, by plotting the frequency (or probability) against the losses.

The histogram of the AEX losses is depicted in Figure 4, where the red line in the histogram refers to an estimate of the normal density. From the histogram we

[14] Log losses (or log returns), instead of raw losses $l_t = -\frac{P_t - P_{t-1}}{P_{t-1}}$ (or raw returns), are commonly used in quantitative finance. The reasons are that it is mathematical convenient (e.g. it allows for time additivity) and it provides more meaningful and robust results.

[15] The excess kurtosis, $\kappa_{exc.}$, is the difference between the sample kurtosis, κ, and the theoretical kurtosis of the normal distribution, i.e. $\kappa_{exc.} = \kappa - 3$. A distribution is said to exhibit excess kurtosis if $\kappa_{exc.} > 0$

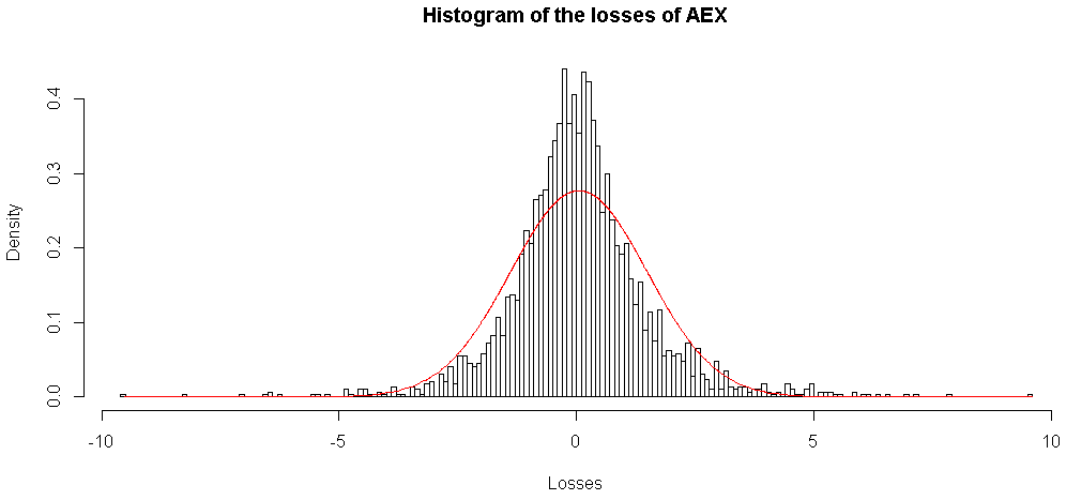


Figure 4: Daily closing price (left) and daily losses (right) of the AEX stock index from January 1, 2000 to July 31, 2016.

observe that the distribution of the AEX losses has long and fat tails compared to the estimated normal distribution. In addition there is more mass around the center than the estimated normal distribution. This suggests that it is unlikely that the losses of AEX follow a normal distribution. See Figure 16 in Appendix E for the histogram of each stock index.

Next we investigate the presence of heteroskedasticity in the losses time series using plots that show the changes in the time series. The time series of the AEX stock index' closing price P_t and the stock losses l_t are plotted in Figure 5. From the figure we observe for a long period of time relatively high volatility around 2002, 2008 and 2015, and relatively low volatility around the remaining time periods. That is to say that the losses time series exhibits *volatility clustering*, i.e. large losses (either with positive or negative sign) tend to be followed by large losses (either with positive or negative sign) for a long period of time, and also for small losses (McNeil et al. 2015). The heteroskedasticity, and in particular volatility clustering, suggest that a GARCH-type model is appropriate for modelling the losses time series of the AEX. See Figure 17 in Appendix E for similar plots of each stock index. From this figure we observe that most stock indices show large volatility during the 2008 financial crisis (also known as the global financial crisis). Many stock indices also show large volatility during the dot-com bubble of $1995 - 2001$ (also known as the tech bubble or Internet bubble) and the Russian financial crisis of 2014.

Now we investigate the heavy-tailedness of the loss distribution using a quantile-quantile (Q-Q) plot. A Q-Q plot is a graphical tool that compares the sample distribution with a certain reference distribution by plotting their quantiles against each other. With this plot we try to visually assess whether or not the sample

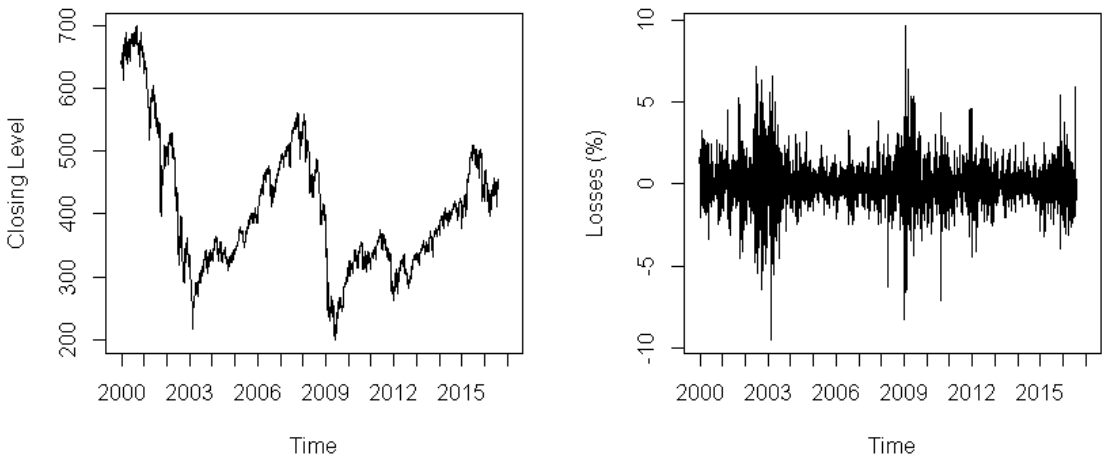


Figure 5: Daily closing price (left) and daily losses (right) of the AEX stock index from January 1, 2000 to July 31, 2016.

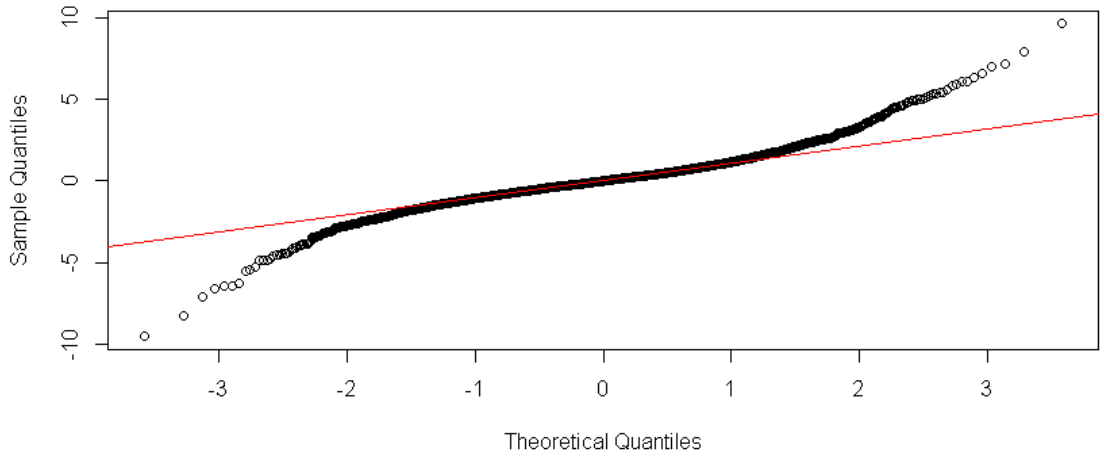


Figure 6: Q-Q plot of the losses of the AEX stock index from January 1, 2000 to July 31, 2016. The red line refers to the normal distribution.

distribution comes from that reference distribution. In our case we compare the quantiles of the loss distribution with that of the normal reference distribution. Thus the observations in the Q-Q plot should lie on a straight line if the sample distribution is the same as the normal distribution.

The Q-Q plot for the AEX losses is depicted in Figure 6, where the red line in the Q-Q plot refers to the normal reference distribution. We observe that the low and high quantiles of the losses greatly deviates from that of the normal distribution. The inverted S-shaped curve of the observations suggests that the loss distribution exhibit heavier tails than that of the normal distribution (McNeil et al. 2015). See Figure 18 in Appendix E for the Q-Q plots of each stock index.

We end by investigating the presence of autocorrelation and volatility clustering

in the losses time series using *correlograms*. A correlogram is a graphical tool that plots the sample autocorrelation $\rho(h)$ against the time lag h. The correlograms up to a 12th lag length for the (raw) losses and absolute losses can be found in Figure 7, where the blue dashed lines indicate the 95% confidence band. For both raw and absolute losses we observe that more than 5% of the estimated correlations lie outside the confidence band. Thus, the correlograms of the raw and absolute losses respectively suggest that autocorrelation is present, which supports the ARMA modelling, and again that volatility clustering is present. See Figure 19 in Appendix E for the correlograms of each stock index.

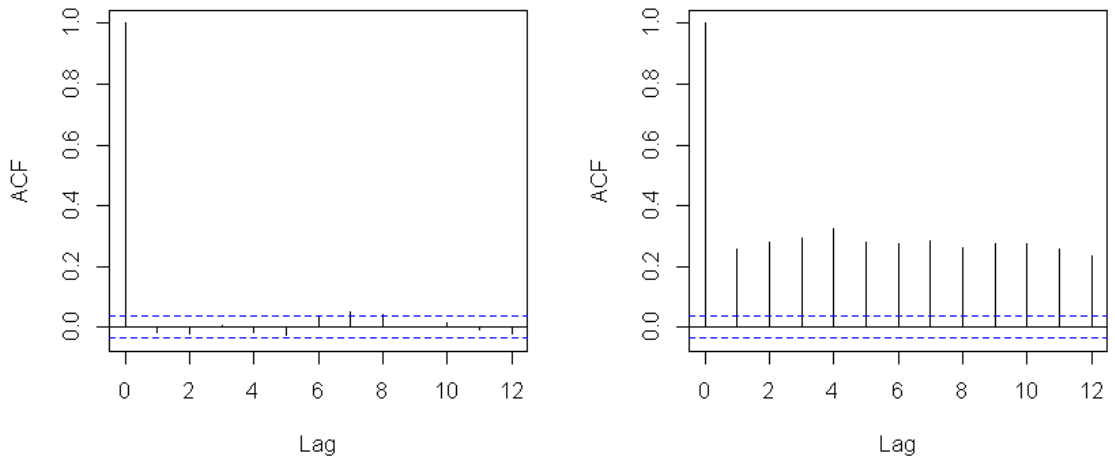


Figure 7: Correlograms of the raw losses (left) and the absolute losses (right) of the AEX stock index from January 1, 2000 to July 31, 2016.

3.2 Numerical tests

So far we investigated the data graphically. In order to approve the use of the ARMA-GARCH-type (i.e. ARMA-GARCH and ARMA-GJR) models, we will perform the following three different numerical tests to the losses time series of the stock indices.

1. *Jarque-Bera (J-B) test (for normality of the time series).* The null hypothesis states that the data follows a normal distribution. In case of non-normality, the losses (and hence the innovations) distribution is heavy-tailed and/or leptokurtic. This motivates a long-tailed loss distribution, e.g. the t or skewed t distribution.

2. *Ljung-Box Q(12)-test (for autocorrelation in the time series up to a 12th lag length).* The null hypothesis states that the data exhibits no autocorrelation, or similarly that the data is independently distributed. Applying this test to raw and absolute losses, we can examine whether or not the data exhibits autocorrelation and volatility clustering. In other words, whether or not respectively the ARMA and GARCH modelling are appropriate.

3. *Augmented Dickey-Fuller (ADF) unit root test (for non-stationarity of the time series)*. The null hypothesis states that the data is non-stationary. The stationarity condition is required for the estimation of the ARMA-GARCH-type models.

See Appendix B for a detailed description of these tests. The results of the formal tests on the losses of the stock indices can be found in Table 2. Notice that the estimation of a MGARCH-type model is only possible if every time series has an equally number of observations. Therefore we only consider the time periods for which a loss observation is available for every stock index, yielding a total of 2,909 loss observations. The remaining loss observations are left out of the analysis. The most recent period of time of the samples becomes July 27, 2016. The results of the formal tests can be summarized as follows.

1. From the J-B test we always reject the null hypothesis of normality at the 5% significance level, which suggests for every stock index that the loss distribution follows a non-normal distribution. In combination with the excess kurtosis, which is the case for eighteen stock indices, we assert that the loss distribution has heavy tails.

2. From the Ljung-Box Q(12) test for the raw losses we most often reject the null hypothesis of no autocorrelation at the 1% or 5% significance level. In the cases of rejecting the null hypothesis (at either significance levels) we have evidence that the losses time series exhibits autocorrelation and so the non-zero-ordered ARMA modelling is likely to be appropriate. Otherwise, it is likely that the zero-ordered ARMA modelling is appropriate.

3. From the Ljung-Box Q(12) test for the absolute losses, referred by |Q(12)|, we always reject the null hypothesis of no autocorrelation at the 1% significance level, which implies for every stock index that the losses time series exhibits volatility clustering. This result implies that the GARCH modelling is appropriate.

4. From the ADF test we always reject the null hypothesis of unit root at the 1% significance level, which implies that the losses time series is stationary. This allows the use of the ARMA-GARCH-type models for the losses time series of every stock index.

These results (including from the visual tests) imply that the ARMA-GARCH-type models are appropriate and that the corresponding assumptions are well satisfied.

Table 2: Descriptive and test statistics of the daily losses l_t for the twenty stock indices (from January 1, 2000 to July 31, 2016)

Dataset	Obs.	Mean	Median	Max.	Min.	Variance	Skewness	Kurtosis	J-B	Q(12)	|Q(12)|	ADF
Asia												
Hang Seng (Hong Kong)	2,909	0.0137	−0.0035	9.0513	−13.4068	2.0539	0.0506	6.0138	4,393.9*	19.2***	1,399.1*	−13.6*
KLSE (Malaysia)	2,909	−0.0108	−0.0290	9.9785	−4.5027	0.6980	1.0183	12.0808	18,223.2*	53.1*	1,007.4*	−13.6*
Nikkei 225 (Japan)	2,909	0.0137	−0.0146	12.1110	−13.2346	2.3741	0.3677	6.1400	4,644.5*	37.8*	970.4*	−14.2*
S&P/ASX (Australia)	2,909	−0.0061	−0.0349	7.3144	−5.6282	0.9737	0.4128	4.3643	2,396.7*	31.5*	1,605.2*	−13.6*
BSE Sensex 30 (India)	2,909	−0.0219	−0.0864	11.8092	−15.9900	2.4300	0.1686	8.0871	7,955.9*	19.0***	1,922.3*	−13.9*
SSE (China)	2,909	−0.0106	−0.0188	9.2562	−9.4008	2.5980	0.1828	5.1693	3,262.2*	15.9	819.2*	−13.2*
TSEC (Taiwan)	2,909	−0.0123	−0.0424	6.9123	−6.1721	1.8648	0.2117	2.6696	888.1*	34.3*	1,120.5*	−14.5*
Europe												
AEX (the Netherlands)	2,909	0.0507	−0.0025	9.5903	−9.5169	2.0725	0.2802	4.6227	2,634.1*	23.4**	2,647.9*	−12.7*
BFX (Belgium)	2,909	0.0240	−0.0189	7.1179	−9.3340	1.6035	0.0737	5.4016	3,546.7*	31.8*	2,169.9*	−12.6*
BIST 100 (Turkey)	2,909	−0.0344	−0.0521	19.9785	−17.7736	4.6755	0.0035	7.4847	6,803.2*	7.5	1,016.8*	−15.0*
CAC 40 (France)	2,909	0.0433	−0.0028	9.4715	−9.2208	2.1640	0.2830	3.4156	1,456.5*	16.6	1,604.4*	−13.0*
DAX (Germany)	2,909	0.0377	−0.0543	7.3355	−10.6851	2.3250	0.2025	3.0872	1,178.3*	8.5	2,217.2*	−13.2*
IBEX (Spain)	2,909	0.0267	−0.0536	13.1852	−13.4836	2.3008	0.0887	5.6479	3,878.4*	18.4	1,308.3*	−13.7*
SMI (Switzerland)	2,909	0.0190	−0.0291	6.3161	−6.4517	1.3704	0.2405	4.0186	1,990.2*	25.1**	2,229.0*	−13.7*
North-America												
Dow 30 (United States)	2,909	0.0117	−0.0363	8.2005	−10.3259	1.3659	0.1972	6.2332	4,737.7*	52.0*	2,833.7*	−14.9*
S&P/TSX (Canada)	2,909	0.0125	−0.0353	9.7879	−9.3702	1.2629	0.3689	8.6730	9,200.4*	33.7*	2,414.3*	−14.6*
South-America												
Ibovespa (Brazil)	2,909	0.0126	0.0000	12.0961	−12.5982	3.2898	0.1183	2.9447	1,060.8*	22.8**	741.2*	−15.0*
IPC (Mexico)	2,909	−0.0140	−0.0603	7.2661	−9.8126	1.7742	0.0700	3.8851	1,836.3*	29.8*	1,592.2*	−14.6*
MERV (Argentina)	2,909	−0.0390	−0.0659	12.9516	−16.1165	4.6568	0.1857	4.3594	2,325.6*	6.4	818.6*	−14.4*
S&P/BVL (Peru)	2,909	−0.0342	−0.0248	9.8764	−8.4271	1.8810	0.3977	7.3985	6,724.3*	79.0*	1,983.2*	−12.4*

*Statistically significant at the 1% level. **Statistically significant at the 5% level. ***Significant at the 10% significance level.

4 Empirical results

In our study we estimate the univariate GARCH-type and copula-MGARCH-type models using respectively the **R** packages *rugarch* (Ghalanos 2015*b*) and *rmgarch* (Ghalanos 2015*a*).

4.1 Margins selection

Before a copula can be incorporated in the modelling, we have to choose the most appropriate specification for the univariate conditional heteroskedasticity (Aloui et al. 2013). This is achieved by fitting the losses time series of every stock index to the univariate ARMA(p_i, q_i)-GARCH(1, 1)-type model, with ARMA orders $p_i, q_i \in \{0, 1, 2\}$, $i = 1, \ldots, 20$, under the assumption of normal, t or skewed t distributed innovations. For each stock index, the result of the AIC value and the corresponding optimal orders (p_i, q_i) can be found in Table 3, where for both the ARMA-GARCH and ARMA-GJR model we indicate the lowest AIC value and the corresponding optimal order (p_i, q_i) by a bold.

Based on AIC, it is notable that the losses time series of most stock indices provide the optimal fit under the assumption of skewed t distributed innovations, while under the assumption of normally distributed innovations they always fits worse. Also, the ARMA-GJR model always fits better than the ARMA-GARCH model (with identical innovations distribution). This result is in line with the stylized fact of fat-tailed loss distribution and the existence of leverage effect, as discussed in Section 2.3. The exhibition of fat-tails of the loss distributions is also supported by the Q-Q plots (see Figure 18 in Appendix E).

Recall from Section 2.3 that the innovations should follow a SWN(0,1) process and so they should behave like an i.i.d. sample (with finite variance). In other words, the innovations should not exhibit autocorrelation. For the AEX index, Figure 8 depicts the time series of the innovations (left) and the correlogram (right) from the fitted univariate ARMA-GARCH model. See Figure 20 and 21 in Appendix F for similar plots for every stock index, where the innovations are respectively from the ARMA-GARCH and ARMA-GJR model. From the innovations plot it looks like the innovations are indeed independently distributed. In fact, from the correlogram we can see that all of the estimated correlations (with lag > 0) lie outside the 5% confidence band. These results suggest that the innovations do not exhibit autocorrelation and thus are likely to be realizations of a SWN process.

Table 3: AIC values and the corresponding optimal orders (p_i, q_i) of the fitted univariate ARMA(p_i, q_i)-GARCH$(1,1)$-type models.

Dataset	GARCH-N	(p_i, q_i)	GARCH-*t*	(p_i, q_i)	GARCH-*st*	(p_i, q_i)	GJR-N	(p_i, q_i)	GJR-*t*	(p_i, q_i)	GJR-*st*	(p_i, q_i)
Asia												
Hang Seng	9,570.58	(0, 0)	9,470.37	(2, 2)	**9,466.21**	**(2, 2)**	9,531.66	(2, 2)	9,448.20	(2, 2)	**9,443.23**	**(2, 2)**
KLSE	6,327.87	(1, 0)	6,098.01	(1, 0)	**6,096.79**	**(1, 0)**	6,297.71	(1, 0)	6,089.35	(1, 0)	**6,088.49**	**(1, 0)**
Nikkei 225	10,238.26	(0, 1)	10,141.56	(0, 1)	**10,129.74**	**(0, 1)**	10,193.96	(2, 2)	10,116.10	(1, 2)	**10,105.30**	**(0, 1)**
S&P/ASX	7,362.04	(0, 0)	7,289.31	(2, 2)	**7,278.41**	**(0, 0)**	7,290.22	(0, 0)	7,240.05	(0, 0)	**7,222.74**	**(0, 0)**
BSE Sensex 30	9,930.83	(0, 1)	9,784.38	(0, 2)	**9,775.57**	**(0, 2)**	9,891.23	(0, 1)	9,738.36	(2, 2)	**9,734.46**	**(0, 2)**
SSE	10,364.87	(0, 0)	10,051.02	(2, 1)	**10,045.41**	**(2, 1)**	10,364.53	(0, 0)	10,050.86	(2, 1)	**10,044.91**	**(2, 1)**
TSEC	9,331.82	(0, 2)	9,217.68	(0, 2)	**9,207.86**	**(0, 2)**	9,304.21	(2, 2)	9,202.52	(0, 2)	**9,191.00**	**(0, 2)**
Europe												
AEX	9,226.94	(2, 2)	9,177.76	(0, 0)	**9,161.22**	**(0, 0)**	9,140.22	(0, 0)	9,104.27	(0, 0)	**9,072.20**	**(2, 2)**
BFX	8,618.33	(0, 0)	8,504.68	(0, 0)	**8,496.21**	**(0, 0)**	8,535.96	(0, 1)	8,450.97	(0, 0)	**8,443.20**	**(0, 0)**
BIST 100	12,003.02	(0, 0)	**11,808.09**	**(0, 0)**	11,808.25	(0, 0)	11,982.90	(0, 0)	11,789.86	(0, 0)	**11,789.45**	**(0, 0)**
CAC 40	9,754.29	(1, 1)	9,695.57	(1, 1)	**9,672.84**	**(1, 1)**	9,705.22	(0, 0)	9,638.81	(1, 1)	**9,619.22**	**(1, 1)**
DAX	9,838.31	(0, 0)	9.792.91	(0, 0)	**9,779.44**	**(1, 1)**	9,783.81	(0, 0)	9,736.71	(0, 0)	**9,722.46**	**(0, 0)**
IBEX	9,972.03	(2, 1)	9,857.02	(0, 0)	**9,841.98**	**(1, 1)**	9,933.48	(0, 0)	9,806.08	(0, 0)	**9,792.01**	**(0, 0)**
SMI	8,258.48	(0, 2)	8,178.05	(0, 0)	**8,164.15**	**(0, 2)**	8,196.68	(0, 2)	8,120.92	(0, 0)	**8,105.99**	**(2, 0)**
North-America												
Dow 30	8,068.70	(1, 1)	7,993.43	(1, 1)	**7,979.49**	**(2, 2)**	8,003.29	(2, 2)	7,933.70	(1, 1)	**7,913.61**	**(1, 1)**
S&P/TSX	7,722.67	(0, 0)	7,664.79	(2, 1)	**7,632.23**	**(2, 1)**	7,691.78	(0, 0)	7,641.51	(0, 0)	**7,610.09**	**(0, 0)**
South-America												
Ibovespa	11,300.50	(2, 1)	11,275.68	(2, 1)	**11,268.42**	**(2, 2)**	11,256.83	(2, 2)	11,240.10	(0, 0)	**11,235.17**	**(0, 0)**
IPC	9,121.41	(2, 2)	9,043.79	(2, 2)	**9,026.66**	**(2, 2)**	9,061.15	(1, 0)	8,994.39	(2, 2)	**8,982.11**	**(2, 2)**
MERV	12,186.20	(1, 0)	11,972.98	(2, 2)	**11,966.81**	**(2, 2)**	12,141.62	(2, 0)	11,953.45	(2, 2)	**11,945.47**	**(2, 2)**
S&P/BVL	8,954.92	(2, 1)	**8,736.65**	**(2, 1)**	8,738.34	(2, 1)	8,931.12	(2, 1)	**8,729.70**	**(2, 1)**	8,731.39	(2, 1)

For the ARMA-GARCH and ARMA-GJR models, the lowest AIC value and the corresponding, optimal order are indicated by a bold.

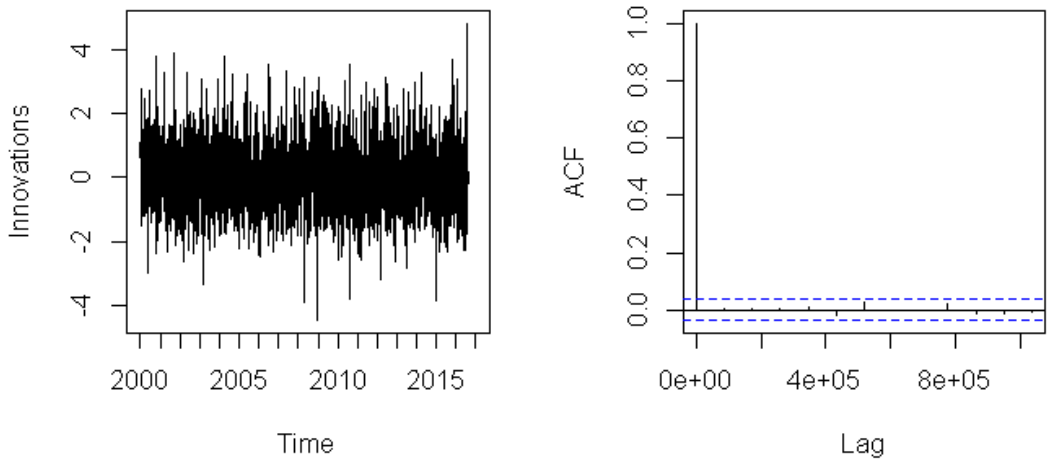


Figure 8: Innovations plot (left) and correlogram (right) from the fitted univariate ARMA-GARCH for the AEX, from January 1, 2000 to July 31, 2016.

In addition we perform the Ljung-Box Q(12)-test to formally examine whether or not the innovations of the margins exhibit autocorrelation. The result of this test can be found in Table 4. From the table we see that for every marginal model the null hypothesis of the Q(12)-test is not rejected at the 5% (and so also 1%) significance level. In other words, we have evidence that the innovations of the univariate ARMA-GARCH and ARMA-GJR models do not exhibit autocorrelation. Thus, we have evidence that the innovations are i.i.d. and so indeed follow a SWN process, as desired. This result is confirmed by the previous exploratory analysis of the innovations.

4.2 Copula-MGARCH model adequacy

In Section 4.1 we determined the most appropriate univariate ARMA-GARCH and ARMA-GJR model for each stock index. Further, we confirmed that the innovations of each of these models indeed follow a SWN process. With this knowledge we can now construct the multivariate copula-MGARCH models, namely the copula-DCC-ARMA-GARCH (using the most appropriate ARMA-GARCH model for each margin) and copula-DCC-ARMA-GJR model (using the most appropriate ARMA-GJR model for each margin), where the copula in each of the copula-MGARCH models is either a Gaussian or t copula. The models are evaluated in terms of their in-sample fit and out-of-sample forecast performance. For comparison we also perform these evaluations for the traditional DCC-ARMA-GARCH and DCC-ARMA-GJR models, with univariate and mul-

Table 4: Ljung-Box Q(12)-test result of the innovations from the most appropriate fitted ARMA-GARCH (left) and ARMA-GJR (right) models.

Stock index (innovations dist.)	Q(12)	Stock index (innovations dist.)	Q(12)
Hang Seng (*st*)	14.2	Hang Seng (*st*)	13.5
KLSE (*st*)	18.3	KLSE (*st*)	16.1
Nikkei 225 (*st*)	12.5	Nikkei 225 (*st*)	10.9
S&P/ASX (*st*)	14.1	S&P/ASX (*st*)	16.9
BSE Sensex 30 (*st*)	7.0	BSE Sensex 30 (*st*)	7.4
SSE (*st*)	10.8	SSE (*st*)	10.4
TSEC (*st*)	11.0	TSEC (*st*)	9.4
AEX (*st*)	6.5	AEX (*st*)	4.4
BFX (*st*)	14.0	BFX (*st*)	11.2
BIST 100 (*t*)	5.6	BIST 100 (*st*)	6.0
CAC 40 (*st*)	9.3	CAC 40 (*st*)	4.9
DAX (*st*)	9.0	DAX (*st*)	6.6
IBEX (*st*)	7.3	IBEX (*st*)	6.2
SMI (*st*)	6.8	SMI (*st*)	3.2
Dow 30 (*st*)	6.0	Dow 30 (*st*)	5.9
S&P/TSX (*st*)	14.3	S&P/TSX (*st*)	14.0
Ibovespa (*st*)	10.1	Ibovespa (*st*)	12.2
IPC (*st*)	15.0	IPC (*st*)	15.6
MERV (*st*)	8.9	MERV (*st*)	10.6
S&P/BVL (*t*)	20.1	S&P/BVL (*t*)	20.1

*Statistically significant at the 1% level. **Statistically significant at the 5% level.

tivariate normal or t distribution for the innovations. These evaluations are used for model selection, model validation and model ranking.

4.2.1 How do we test the in-sample fit of the models?

In this section we will perform model selection by testing the in-sample fit of all eight models. To test the in-sample fit, we start by fitting each model to the whole sample and then compute the log-likelihood ($\ln(L)$) value and the Akaike, corrected Akaike (AICc), Bayesian (BIC), Shibata (SIC) and Hannan-Quinn (HQIC) information criteria values. The last four information criteria are defined as

$$\begin{aligned}
\text{AICc} &= \text{AIC} - \frac{2k(k+1)}{n-k-1}, \\
\text{BIC} &= k\ln(n) - 2\ln(L), \\
\text{SIC} &= n\ln\left(\frac{n+2k}{n}\right) - 2\ln(L), \\
\text{HQIC} &= 2k\ln(\ln(n)) - 2\ln(L),
\end{aligned}$$

where n is the number of observations, k is the number of parameters and L is the value of the likelihood function of the fitted model (Akaike 1974, Ghalanos

2015*b*). [16] These five information criteria enable model selection by penalizing the overfitting at different rates (Ghalanos (2015*b*)). In other words, they deal with the trade-off between the goodness-of-fit of the model (see the log-likelihood part) and the complexity of the model (see the remainder part).

The fit of a model is assumed to be better for higher values of the (log-)likelihood. In a similar way as for the AIC, the fit of the model is assumed to be better for lower values of the AICc, BIC, HQIC and SIC. The outcomes of the in-sample fit test can be found in Table 5. The table reports the ln(L), AIC, AICc, BIC, SIC and HQIC values for each of the eight fitted models, where the highest value of the log-likelihood and the lowest value of each information criterion are indicated by a bold.

Table 5: The ln(L), AIC, AICc BIC, SIC and HQIC values of the fitted copula-MGARCH and traditional MGARCH models for the whole sample are given. The highest value of the log-likelihood and the lowest value of each information criterion are indicated by a bold.

Model	ln(L)	AIC	AICc	BIC	SIC	HQIC
Ga-GARCH	−73327.26	146,980.13	147,077.95	147,954.65	146,962.68	147,322.12
t-GARCH	−72965.02	146,258.70	146,357.11	**147.239.04**	146,241.25	**146,610.69**
GARCH-N	−74571.74	149,746.59	149,816.33	151,544.36	149,691.32	150,392.39
GARCH-*t*	−73431.18	147,524.12	147,609.40	149,502.24	147,460.12	148,236.82
Ga-GJR	−73279.58	146,910.32	147,015.38	147,954.65	146,889.96	147,285.58
t-GJR	**−72943.39**	**146,238.34**	**146,344.02**	147,291.40	**146,217.98**	146,616.51
GJR-N	−74424.28	149,493.51	149,574.48	151,425.09	149,432.42	150,188.76
GJR-*t*	−73406.99	147,503.75	147,596.90	149,566.24	147,433.94	148.245.55

The following conclusions are all drawn according to the log-likelihood and the information criteria.

- The copula-based models provide better fits than non-copula-based models. To stress this, based on the log-likelihood and each information criterion value, the four models with the lowest values are all copula-based models.
- Considering the Gaussian-based models only, we notice that the ones following the GJR specification are fitting at least as well as the ones following the standard GARCH specification.
- Considering the *t*-based models only, we notice that the ones following the GJR specification do not necessarily fit at least as good as the ones following the standard GARCH specification. This holds for both the copula-based and non-copula-based models. Therefore we have evidence that the *t*-based models perform generally well compared to the Gaussian-based

[16] The Bayesian information criterion is also known as the Schwarz information criterion and hence is sometimes abbreviated by SIC. This should not be confused with the Shibata information criterion.

models and so in that case, the choice of the GARCH specification (either GJR or standard GARCH) does not have a big impact. This highlights the relatively strong performance of the t-based models.

- Most importantly, the t-GARCH and t-GJR models have the two highest values of the log-likelihood and the two lowest values of each information criterion. As we can see from the table, the t-GARCH model provides the best fit according to the BIC and HQIC, whereas the t-GJR model provides the best fit according to the $\ln(L)$, AIC, AICc and SIC.

That the t-GJR model performs the fitting relatively well could be expected from what we have seen in Section 4.1. Namely the fact that the ARMA-GJR marginal models provide better in-sample fits than the ARMA-GARCH marginal models (with identical innovations distribution), according to the AIC values. In Section 4.1 we also found that the innovations distributions that exhibit fat-tails always provide better fits than the ones with normal-tails. Concluding, the t-GJR model has the best in-sample fit performance. [17]

4.2.2 How do we test the out-of-sample forecast of the models?

In this section we will perform model validation and model ranking by testing the out-of-sample forecast of all eight models. Thus in this thesis we perform two different out-of-sample forecast evaluation procedures. Specifically we are interested in the one-day-ahead risk and volatility forecast performance of the models.

The procedure for model validation, which is a common method in risk management, is the VaR backtest to check how well the models estimate the VaR. Before we describe this procedure, let us introduce *portfolios*. Let us denote the weight and the loss at time t of the i^{th} stock index respectively by $w_i^t \in [0,1]$ and $l_i^t = \mu_{i,t} + \sigma_{i,t} z_{i,t}$, $i = 1, \ldots, 20$. Then the portfolio loss at time t is given by $l_{port}^t = \boldsymbol{w}^{t\prime} \boldsymbol{l}^t$, where $\boldsymbol{l}^t = (l_1^t, \ldots, l_{20}^t)'$ and $\boldsymbol{w}^t = (w_1^t, \ldots, w_{20}^t)'$ with the restriction that $\sum_{i=1}^{20} w_i^t = 1$, $\forall t$. In this thesis we only consider the equally-weighted portfolio, i.e. $w_i^t = 1/20$ for $i = 1, \ldots, 20$ and all t.

The VaR backtest procedure is based on a series of rolling windows, where each window consists of 800 observations. The procedure for the copula-MGARCH models is described by the following steps, inspired by Weiß (2013) and Ghorbel & Trabelsi (2009).

[17] It should be mentioned that the t-GARCH model performs the fitting very much alike the t-GJR model: the difference in the value of the log-likelihood and each information criterion between the two models is relatively small.

1. Let $k = 1$ and define

$$T_1 = \begin{cases} 2{,}109 & \text{if } k = 423 \\ 1 + 5(k-1) & \text{otherwise,} \end{cases}$$

$$T_2 = \begin{cases} 2{,}908 & \text{if } k = 423 \\ 800 + 5(k-1) & \text{otherwise.} \end{cases}$$

2. Consider the losses observations of each stock index for time periods $T_1, \ldots, T_2$: $\mathbb{L}_i = (l_{i,T_1}, l_{i,T_1+1} \ldots, l_{i,T_2})'$, $i = 1, \ldots, 20$.

3. Fit the copula-MGARCH models, with the most appropriate margins, to $\mathbb{L} = (\mathbb{L}_1, \ldots, \mathbb{L}_{20})$, see Section 2.6.3.

4. Forecast μ_{i,T_2+1} and σ_{i,T_2+1} using the estimated variables up to time T_2 and denote them respectively by $\hat{\mu}_{i,T_2+1}$ and $\hat{\sigma}_{i,T_2+1}$, $i = 1, \ldots, 20$. Further, denote $\hat{\boldsymbol{\mu}}_{T_2+1} = (\hat{\mu}_{1,T_2+1}, \ldots, \hat{\mu}_{20,T_2+1})'$ and $\hat{\Delta}_{T_2+1} = \text{diag}(\hat{\sigma}_{1,T_2+1}, \ldots, \hat{\sigma}_{20,T_2+1})$.

5. Perform *Monte Carlo simulations*: simulate $M = 5{,}000$ times from the (20-dimensional) fitted copula in order to obtain $(M \times 1)$ vectors of simulated marginal samples $\boldsymbol{U}_{i,T_2+1} = (u^{(1)}_{i,T_2+1}, \ldots, u^{(M)}_{i,T_2+1})'$, $i = 1, \ldots, 20$.

6. For $i = 1, \ldots, 20$, $m = 1, \ldots, M$, transform the simulated marginal samples $\boldsymbol{U}_{i,T_2+1}$ to the original innovations scale using the quantile function of the corresponding distribution, i.e. $F_i^{-1}(u^{(m)}_{i,T_2+1}) = z^{(m)}_{i,T_2+1}$. Denote the (20×1) vectors of simulated innovations by $\boldsymbol{Z}^{(m)}_{T_2+1} = (z^{(m)}_{1,T_2+1}, \ldots, z^{(m)}_{20,T_2+1})'$.

7. Forecast M one-day-ahead losses for every stock index by

$$\hat{\boldsymbol{l}}^{(m)}_{T_2+1} = \hat{\boldsymbol{\mu}}_{T_2+1} + \hat{\Sigma}^{1/2}_{T_2+1} \boldsymbol{Z}^{(m)}_{T_2+1}.$$

8. For quantile level $\alpha \in \{0.90, 0.95, 0975, 0.99, 0.995\}$, estimate the portfolio VaR by

$$\widehat{\text{VaR}}^{\alpha}_{port,T_2} = \text{quantile}_{\alpha}\{\boldsymbol{w}' \hat{\boldsymbol{l}}^{(m)}_{T_2+1}\}_{m=1}^{M}$$

   where $\boldsymbol{w} = (1/20, \ldots, 1/20)'$.

9. Repeat Steps 2 – 8 until $k = \lceil (2{,}909 - 800)/5 \rceil + 1 = 423$.

10. Consider the exceptions $\{l^{t+1}_{port} > \widehat{\text{VaR}}^{\alpha}_{port,t}\}$ to determine the value of the backtest's test statistic and obtain the p-value.

The series of rolling windows used in the VaR backtest procedure is depicted in Figure 9.

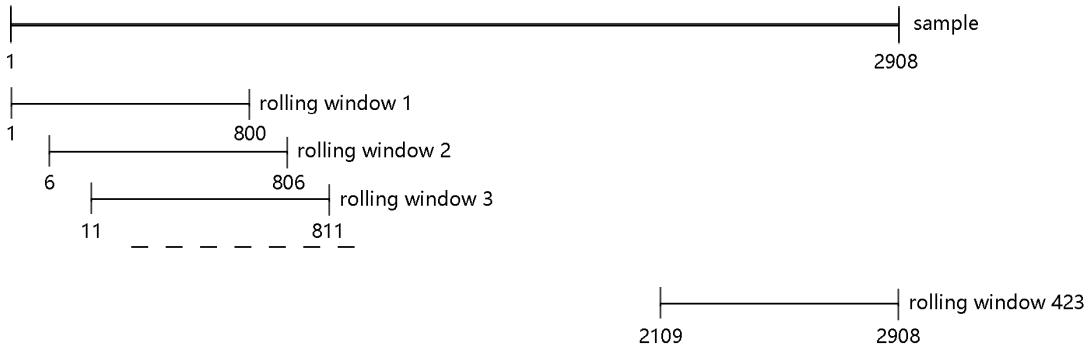


Figure 9: Series of rolling windows.

Thus the estimation process is repeated once in every five observations. This is because of the high computational cost that comes from the re-estimation of (20-dimensional) models with a large set of parameters. For a low number of different time series, usually less than or equal to five, it is common to repeat the process once in every observation.

Observe that we take $T_1 = 2{,}109$ and $T_2 = 2{,}908$ if $k = 423$ (as opposed to $T_1 = 2{,}110$ and $T_2 = 2{,}909$ in the in-sample fit evaluation procedure). The reason for this is that the last losses, $l_{i,2909}$, are used to determine whether or not the exception $\{l_{port}^{2909} > \widehat{\text{VaR}}_{port,2908}^{\alpha}\}$ occurred.

Concretely, for $i = 1, \ldots, 20$, $t = 1, \ldots, T_2$, we firstly model the marginal losses by $l_{i,t} = \mu_{i,t} + \sigma_{i,t} z_{i,t}$, secondly transform the innovations to uniform by $F_i(z_{i,t}) = u_{i,t}$, thirdly fit the copula to $u_{i,t}$, fourthly simulate M times from the fitted copula to obtain simulated marginal samples $u_{i,T_2+1}^{(m)}$, $m = 1, \ldots, M$, fifthly reintroduce the forecast of the conditional variance-covariance matrix and mean vector to forecast a set of losses at time T_2+1 by $\hat{\boldsymbol{l}}_{T_2+1}^{(m)} = \hat{\boldsymbol{\mu}}_{T_2+1} + \hat{\Sigma}_{T_2+1}^{1/2} \boldsymbol{Z}_{T_2+1}^{(m)}$ and lastly estimate the portfolio VaR by the α-quantile of $\{\boldsymbol{w}' \hat{\boldsymbol{l}}_{T_2+1}^{(m)}\}_{m=1}^{M}$. [18] The Monte Carlo simulation is performed in order to obtain the quantile of the joint innovations distribution indirectly. We compare the estimated portfolio VaRs with the observed next-day portfolio loss to determine whether or not an exception occurred.

The above procedure only applies to the copula-MGARCH models. For the traditional MGARCH models there exists analytical methods to estimate the VaR. To be explicit, under the assumption of multivariate normal or t distributed innovations, we can simply consider the quantile function of the standardized normal or t distribution. The VaR backtest procedure for the traditional MGARCH models is described by the following steps, inspired by Huang et al. (2009).

[18] Recall from Section 2.4.1 that the past estimates of the variables, up to some time t, are used to forecast the values of the variables at time period $t+1$. Therefore we use the estimated parameters of the in-sample fitted models (see Section 4.2.1) to forecast the one-day out-of-sample values.

1. Let $k = 1$ and define

$$T_1 = \begin{cases} 2{,}109 & \text{if } k = 423 \\ 1 + 5(k-1) & \text{otherwise,} \end{cases}$$

$$T_2 = \begin{cases} 2{,}908 & \text{if } k = 423 \\ 800 + 5(k-1) & \text{otherwise.} \end{cases}$$

2. Consider the losses observations of each stock index for time periods $T_1, \ldots, T_2$: $\mathbb{L}_i = (l_{i,T_1}, l_{i,T_1+1} \ldots, l_{i,T_2})'$, $i = 1, \ldots, 20$.

3. Fit the traditional MGARCH models to $\mathbb{L} = (\mathbb{L}_1, \ldots, \mathbb{L}_{20})$, see Section 2.4.1.

4. Forecast μ_{i,T_2+1} and σ_{i,T_2+1} using the estimated variables up to time T_2 and denote them respectively by $\hat{\mu}_{i,T_2+1}$ and $\hat{\sigma}_{i,T_2+1}$, $i = 1, \ldots, 20$. Further, denote $\hat{\boldsymbol{\mu}}_{T_2+1} = (\hat{\mu}_{1,T_2+1}, \ldots, \hat{\mu}_{20,T_2+1})'$ and $\hat{\Delta}_{T_2+1} = \text{diag}(\hat{\sigma}_{1,T_2+1}, \ldots, \hat{\sigma}_{20,T_2+1})$.

5. Estimate $\boldsymbol{Y}_t$, $t = 1, \ldots, T_2$, (see (27)) $\boldsymbol{P}_c$ (see (28)) and consequently $\boldsymbol{P}_{T_2+1}$ (see (24)). Estimate $\boldsymbol{\Sigma}_{T_2+1}$ by $\hat{\Delta}_{T_2+1}\hat{\boldsymbol{P}}_{T_2+1}\hat{\Delta}_{T_2+1}$.

6. For quantile level $\alpha \in \{0.90, 0.95, 0975, 0.99, 0.995\}$, estimate the portfolio VaR by

$$\widehat{\text{VaR}}^{\alpha}_{port,T_2} = \hat{\mu}_{port,T_2+1} + \hat{\sigma}_{port,T_2+1} z_\alpha$$

   where $\hat{\mu}_{port,T_2+1} = \boldsymbol{w}'\hat{\boldsymbol{\mu}}_{T_2+1}$ and $\hat{\sigma}_{port,T_2+1} = \sqrt{\boldsymbol{w}'\hat{\boldsymbol{\Sigma}}_{T_2+1}\boldsymbol{w}}$, $z_\alpha = \Phi^{-1}(\alpha)$ under the assumption of multivariate normal distribution and $z_\alpha = t_\nu^{-1}(\alpha)$ under the assumption of multivariate t distribution.

7. Repeat Steps 2 – 6 until $k = \lceil (2{,}909 - 800)/5 \rceil + 1 = 423$.

8. Consider the exceptions $\{l^{t+1}_{port} > \widehat{\text{VaR}}^{\alpha}_{port,t}\}$ to determine the value of the backtest's test statistic and obtain the p-value.

For each model, the number of exceptions for each quantile α can be found in Table 6 as well as the p-value of the test statistics, where a p-value less than or equal to 0.05 is indicated by a bold.

From the p-value we can evaluate the significance of the test statistic of a backtest. If the p-value is larger than the 5% significance level, then the test statistic does not differ significantly from zero and consequently we do not reject the corresponding null hypothesis. As the null hypotheses of the backtests indicate that the VaRs are statistically well-estimated, we opt for the model that has the largest number of p-values greater than 0.05.

Table 6: VaR backtest result, where 423 windows are considered.

	Expected	Ga-GARCH	*t*-GARCH	GARCH-N	GARCH-*t*	Ga-GJR	*t*-GJR	GJR-N	GJR-*t*
$\alpha = 0.90$									
Number of exceptions	42	49	49	46	44	49	50	38	39
LR_{uc}		0.288	0.288	0.554	0.784	0.288	0.224	0.479	0.588
LR_{ind}		0.740	0.740	0.994	0.832	0.403	0.660	0.799	0.719
LR_{cc}		0.538	0.538	0.839	0.942	0.401	0.433	0.753	0.810
$\alpha = 0.95$									
Number of exceptions	21	25	26	26	24	23	25	23	23
LR_{uc}		0.403	0.295	0.295	0.533	0.684	0.403	0.684	0.684
LR_{ind}		0.665	0.746	0.746	0.729	0.512	0.665	0.512	0.512
LR_{cc}		0.642	0.549	0.549	0.776	0.743	0.642	0.743	0.743
$\alpha = 0.975$									
Number of exceptions	11	15	15	16	16	14	14	13	13
LR_{uc}		0.194	0.194	0.116	0.116	0.309	0.309	0.466	0.466
LR_{ind}		0.293	0.293	0.261	0.261	0.327	0.327	0.363	0.363
LR_{cc}		0.248	0.248	0.155	0.155	0.369	0.369	0.507	0.507
$\alpha = 0.99$									
Number of exceptions	4	7	7	12	10	6	6	11	10
LR_{uc}		0.216	0.216	**0.002**	**0.017**	0.416	0.416	**0.006**	**0.017**
LR_{ind}		0.627	0.627	0.402	0.486	0.677	0.677	0.443	0.486
LR_{cc}		0.413	0.413	**0.006**	**0.044**	0.659	0.659	**0.017**	**0.044**
$\alpha = 0.995$									
Number of exceptions	2	3	4	8	6	3	4	8	8
LR_{uc}		0.566	0.248	**0.002**	**0.029**	0.566	0.248	**0.002**	**0.002**
LR_{ind}		0.836	0.782	0.578	0.677	0.836	0.782	0.578	0.578
LR_{cc}		0.830	0.493	**0.007**	0.084	0.830	0.493	**0.007**	**0.007**

A p-value less than or equal to 0.05 is indicated by a bold.

We make the following observations from the VaR backtest.

- The number of exceptions of the models is in all cases higher than the expected number of exceptions, except for the GJR-N and GJR-t model for $\alpha = 0.90$.
- The models following the GJR specification generally have equal or lower number of exceptions than their counterparts following the standard GARCH specification.
- For each of the three tests (LR_{uc}, LR_{ind} and LR_{cc}), every model has all VaR estimates being statistically significant at the 5% level for quantile level $\alpha \in \{0.90, 0.95, 0.975\}$.
- For each of the three tests, the copula-MGARCH models have all VaR estimates being statistically significant at the 5% level.

The first observation implies that the MGARCH models are generally optimistic: higher number of exceptions is related to generally lower VaR estimates. Furthermore in terms of number of exceptions, the copula-MGARCH models perform relatively well for higher quantile levels, while the traditional MGARCH models perform relatively well for lower quantile levels. The second observation implies that the models following the GJR specification generally estimate the volatility higher than their GARCH counterparts, and hence the former models are generally more pessimistic. From the third observation we have evidence that each MGARCH-type model is adequate for VaR estimation for $\alpha \in \{0.90, 0.95, 0.975\}$. This indicates why MGARCH-type models are popular in risk management, because $\alpha = 0.95$ and 0.975 are commonly considered for estimating risk measures. From the last observation we have evidence that the copula-MGARCH models are adequate for VaR estimation for even any considered α. Concluding, the copula-MGARCH models are all optimal for the estimation of the VaR. Moreover, these models outperform the traditional MGARCH models.

Recall from Section 2.2 that the VaR backtest is only used for model validation, but not for model selection or ranking. So from the first out-of-sample forecast procedure, namely the VaR backtest, we only have evidence that every copula-MGARCH-type model performs sufficiently well on estimating the VaR. We perform a second procedure, which is used to rank the models based on their out-of-sample volatility forecast performance. Moreover, this procedure is based on statistical loss functions, namely Mean Squared Error (MSE), Mean Absolute Deviation (MAD), Logarithmic Loss (LL), Heteroskedasticity-Adjusted MSE (HMSE) and a loss function implicit in the Gaussian quasi-maximum likelihood function (GMLE) (Lopez 2001): [19]

[19] The MAD is defined identically to the Mean Absolute Error (MAE). The latter description is used by Lopez (2001).

$$\text{MSE} = \frac{1}{Tn}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sigma_{i,t+1|t} - \hat{\sigma}_{i,t+1|t}\right)^2, \tag{52}$$

$$\text{MAD} = \frac{1}{Tn}\sum_{i=1}^{n}\sum_{t=1}^{T}|\sigma_{i,t+1|t} - \hat{\sigma}_{i,t+1|t}|, \tag{53}$$

$$\text{LL} = \frac{1}{Tn}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\ln(\sigma^2_{i,t+1|t}) - \ln(\hat{\sigma}^2_{i,t+1|t})\right), \tag{54}$$

$$\text{HMSE} = \frac{1}{Tn}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\frac{\sigma^2_{i,t+1|t}}{\hat{\sigma}^2_{i,t+1|t}} - 1\right)^2, \tag{55}$$

$$\text{GMLE} = \frac{1}{Tn}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\ln(\hat{\sigma}^2_{i,t+1|t}) + \frac{\sigma^2_{i,t+1|t}}{\hat{\sigma}^2_{i,t+1|t}}\right)^2, \tag{56}$$

where $\sigma_{i,t+1|t}$ is the latent volatility and $\hat{\sigma}_{i,t+1|t}$ is the i^{th} one-day-ahead volatility forecast at current time t. We use the squared non-standardized residuals, $\hat{\epsilon}^2_{i,t+1|t} \equiv (l_{i,t+1} - \hat{\mu}_{i,t+1|t})^2$, as a proxy for the latent squared volatility, $\sigma^2_{i,t+1|t}$, inspired by Kosapattarapim et al. (2011). This is supported by the fact that

$$\sigma^2_{i,t+1|t} = \text{E}[\sigma^2_{i,t+1}|\mathcal{F}_t] = \text{E}[\epsilon^2_{i,t+1}Z^2_{i,t+1}|\mathcal{F}_t] = \text{E}[\epsilon^2_{i,t+1}|\mathcal{F}_t] = \text{E}[(l_{i,t+1} - \mu_{i,t+1})^2|\mathcal{F}_t].$$

Notice that this evaluation procedure particularly focusses on the volatility forecast instead of the losses forecast as opposed to the VaR backtesting procedure. Additionally, this procedure is used for model ranking, where the one-day-ahead forecast of a model is assumed to be more accurate if it yields lower values for the statistical loss functions.

The values of the statistical loss functions are determined from the estimated and forecasted variables of the VaR backtesting procedure. Thus, again $T = 423$ is the total number of observations, $n = 20$ is the number of stock indices and define $\tau = 800 + 5(t-1)$ so that the subscript "$t+1|t$" of the variables in (52) – (56) becomes "$\tau+1|\tau$". The outcomes are reported in Table 7, where for each statistical loss function the lowest value is indicated by a bold and the ranking is given in parenthesis.

From the table it is noteworthy that there is no model that provides the optimal value for each statistical loss function. In fact, the ranking of the models greatly differs for each statistical loss function. It seems that the choice of the statistical loss function affects the result of the volatility forecast evaluation. For instance, it seems that MSE and MAD favor a given traditional MGARCH model over its copula-MGARCH counterpart, whereas it seems that HMSE and GMLE favor a given copula-MGARCH model over its traditional MGARCH counterpart. Further, it also seems that LL, HMSE and GMLE favor the models following the GJR specification over the models following the standard GARCH specification.

This result is in accordance with previous studies on forecast evaluation. Kosapattarapim et al. (2011) and Shamiri & Isa (2009) empirically show that the best fitted GARCH model, based on AIC, is not necessarily the model that provides the best volatility forecast in terms of MSE and MAD. [20] An empirical example by Lopez (2001) confirms that the choice of the statistical loss functions in (52) − (56) directly affects the forecast evaluation results. According to Laurent et al. (2013), the results of the forecast evaluation are also dependent on the choice of the proxy for the latent squared volatility.

Kosapattarapim et al. (2011) suggest to evaluate the difference of the values of the statistical loss functions between the best fitted model (here: t-GJR) and the best performing model (i.e. the model with lowest statistical loss function value) by the Percent Error (PE). The PE is defined as

$$\text{PE} = \frac{A - B}{|A|},$$

where A is the statistical loss function value of the best fitted model and B is the lowest statistical loss function value (see bold values in Table 7). The PE values are also reported in Table 7. All PE values are less than 0.05, which indicates that the statistical loss function values of the t-GJR model are not statistically different from the best forecast performing model. This suggests that the t-GJR model is reliable for volatility forecasting, although it does not yield the optimal value for each statistical loss function.

Table 7: Statistical loss functions.

	MSE	MAD	LL	HMSE	GMLE
Ga-GARCH	0.9478051 (2)	0.7435966 (3)	−1.556842 (7)	4.527939 (6)	1.299437 (6)
t-GARCH	0.9478312 (3)	0.7435934 (2)	−1.556953 (6)	4.527936 (5)	1.299381 (5)
GARCH-N	**0.9439955** (1)	**0.7416926** (1)	−1.551210 (8)	4.580340 (8)	1.300505 (7)
GARCH-t	0.9522066 (4)	0.7462446 (4)	−1.567432 (4)	4.536987 (7)	1.301881 (8)
Ga-GJR	0.9935522 (7)	0.7549792 (7)	−1.584870 (2)	4.226730 (2)	**1.287352** (1)
t-GJR	0.9935280 (6)	0.7549756 (6)	−1.584739 (3)	**4.226617** (1)	1.287356 (2)
GJR-N	0.9773505 (5)	0.7512461 (5)	−1.561576 (5)	4.370069 (4)	1.291647 (4)
GJR-t	0.9973043 (8)	0.7573616 (8)	**−1.598582** (1)	4.277152 (3)	1.290476 (3)
PE	0.0498551	0.0175939	0.0087352	0.0000000	0.0000031

4.3 Describing the dependence structure

By describing the dependence structure of the stock indices, we could identify useful proxies for (not) taking on a position of a particular stock market to hedge against other markets. Based on the conducted test criteria of the in-sample fit

[20] Like Lopez (2001), both Kosapattarapim et al. (2011) and Shamiri & Isa (2009) also use the name MAE instead of MAD.

test and the two out-of-sample forecast tests we conclude that the t-GJR model is optimal. Because of this we expect that the t-GJR model is able to describe the dependence structure of the stock indices most accurate out of the eight models studied. Therefore we will exclusively consider this model in this section.

It should be mentioned that the tests also show that the t-GARCH model performs the fit and forecast relatively well, in fact almost as well as the t-GJR model. Namely, the difference in the value of the log-likelihood and each information criterion between the two models is relatively small and all of their VaR estimates are significant at the 5% level. We expect that the t-GARCH model would yield similar results as the t-GJR model.

4.3.1 How do we obtain estimates of the rank correlations and tail-dependence coefficients in order to describe the dependence structure?

In Section 2.4.2 we mentioned that we prefer to use the Kendall's tau over the Pearson's correlation coefficient for the estimation of the correlation between time series. For a Gaussian and t copula it can be shown that the Kendall's tau of a bivariate random vector $(X_i, X_j)'$ can be written as:

$$\rho_\tau(X_i, X_j) = \frac{2}{\pi} \arcsin(\rho(X_i, X_j)), \tag{57}$$

where $\rho(X_i, X_j)$ is the Pearson's correlation coefficient for $(X_i, X_j)'$ from the copula (Demarta & McNeil 2004). Thus the dynamic conditional Kendall's tau matrix, denoted by $\boldsymbol{P}_{\tau,t}$, is obtained by applying the transformation in (57) to every Pearson's correlation coefficient of the estimated dynamic conditional correlation matrix $\hat{\boldsymbol{P}}_t$.

A correlation matrix is required to be positive-definite. However, according to McNeil et al. (2015) the transformation in (57) does not necessarily yield a matrix that is positive-definite. If the dynamic conditional Kendall's tau matrix is not positive definite, we could perform Algorithm 7.57 (eigenvalue method) of McNeil et al. (2015) in order to obtain a positive-definite matrix that is close to the original non-positive-definite dynamic conditional Kendall's tau matrix.

We can approve that the matrix is positive-definite using the following theorem in Leon (2008):

Theorem 4.1. A symmetric matrix M is positive-definite if and only if all of its eigenvalues are positive.

Next to estimating the correlation matrix for describing the dependence structure, we also estimate the TIs of every pair of stock indices. The estimates of the TIs are obtained from fitted unconditional bivariate t copulas using equations

(40) and (41). [21] These bivariate copulas are fitted on the $M = 5,000$ simulated innovations of each of two corresponding stock indices generated from the fitted t-GJR model.

4.3.2 Using the t-GJR model to describe the dependence structure

Recall that we found the t-GJR model being the most appropriate model and thus this model will be used to describe the dependence structure of the stock indices. The estimated parameters of the margins in the t-GJR model are reported in Table 8, where the asterisks indicate whether or not a parameter is statistically significant and if so, at which level. The t-statistic value of each estimated parameter is left out due to the lack of space. The result can be summarized as follows.

- For each stock index the variables α_{k1} and β_{k1} are statistically significant at the 1% level, which implies that the conditional volatility of the losses are significantly explained by lagged squared residual and lagged conditional variance. The latter result indicates that the volatility is highly persistent, i.e. large (respectively small) changes in the conditional variance are followed by other large (respectively small) changes (Aloui et al. 2013).
- The use of the asymmetric specification of GJR is supported by the fact that the leverage terms γ_k are all statistically significant at the 1% level, except for KLSE (for which it is statistically significant at the 5% level) and SSE (for which it is non-significant). This verifies the existence of asymmetric response of volatility to market shocks (Aloui et al. 2013).
- The non-normality assumption for the loss distributions is supported by the statistical significance at the 1% level of the skew (if applicable) and shape parameters. The positive skewness and relatively low shape parameter of the margins indicate that the tail of the loss distribution on the right side is longer and fatter than the left side. That is, positive extreme losses are higher and they are more likely to occur than negative extreme losses.
- Finally, the DCC parameters, $\alpha_1 = 0.005$ and $\beta_1 = 0.988$ (not given in Table 8), are both statistically significant at the 1% level, which indicates that the conditional correlation process between the stock indices is persistent. Therefore the assumption of constant conditional correlation is not supported and so we have evidence that the DCC specification is more realistic for the sample portfolios than Bollerslev's Constant Conditional Correlation (CCC) specification, which assumes $\alpha_1 = \beta_1 = 0$ in equation (24) (McNeil et al. 2015). This is in agreement with previous studies in financial stock markets, see for instance Isogai (2015) and Padhi & Lagesh (2012).

[21] In order to estimate the TIs we have to use unconditional copulas, instead of the conditional copula from the fitted t-GJR model, due to the lack of time-dependence property of the TIs.

Table 8: Estimated parameters of the margins in the t-GJR model.

k	μ_k	ϕ_{k1}	ϕ_{k2}	λ_{k1}	λ_{k2}	α_{k0}	α_{k1}	β_{k1}	γ_k (leverage)	ξ_k (skew)	ν_k (shape)
HS	0.006	−0.710*	−0.971*	0.711*	0.955*	0.019**	0.092*	0.931*	−0.067*	1.068*	7.664*
KLSE	−0.016	0.074*	–	–	–	0.012**	0.151*	0.870*	−0.067**	1.044*	5.352*
NIKKEI	−0.003	–	–	−0.036**	–	0.076*	0.149*	0.869*	−0.107*	1.105*	9.674*
ASX	−0.016	–	–	–	–	0.011*	0.128*	0.918*	−0.121*	1.124*	11.046*
BSE	−0.039***	–	–	0.031***	−0.042**	0.065*	0.190*	0.855*	−0.149*	1.098*	7.049*
SSE	0.029	0.991*	0.005	−0.982*	–	0.038*	0.081*	0.922*	−0.026	1.066*	3.912*
TSEC	−0.032***	–	–	0.006	−0.050**	0.009*	0.074*	0.945*	−0.052*	1.098*	7.620*
AEX	0.019	−0.570*	−1.000*	0.568*	0.100*	0.027*	0.178*	0.885*	−0.159*	1.113*	11.725*
BFX	−0.018	–	–	–	–	0.033*	0.203*	0.853*	−0.159*	1.089*	9.014*
BIST	−0.051	–	–	–	–	0.115*	0.143*	0.876*	−0.090*	1.041*	6.220*
CAC	0.011	0.632*	–	−0.672*	–	0.036*	0.143*	0.899*	−0.125*	1.136*	10.806*
DAX	0.005	–	–	–	–	0.035*	0.155*	0.896*	−0.139*	1.107*	11.741*
IBEX	0.011	–	–	–	–	0.025*	0.161*	0.897*	−0.132*	1.112*	9.423*
SMI	−0.002	−0.011	−0.038***	–	–	0.035*	0.200*	0.854*	−0.168*	1.118*	9.712*
DOW	−0.004	−0.817*	–	0.789*	–	0.025*	0.203*	0.873*	−0.192*	1.126*	8.283*
TSX	−0.006	–	–	–	–	0.016*	0.136*	0.895*	−0.100*	1.166*	11.130*
BVSP	0.070	–	–	–	–	0.070*	0.112*	0.914*	−0.104*	1.073*	15.332*
IPC	−0.030	0.572*	−0.987*	−0.560*	0.985*	0.023*	0.136*	0.905*	−0.112*	1.104*	9.080*
MERV	−0.027	−0.008**	−0.989*	0.006*	0.998*	0.194*	0.178*	0.844*	−0.122*	1.082*	5.484*
BVL	−0.043	1.099*	−0.106*	−0.983*	–	0.058*	0.212*	0.806*	−0.084*	–	5.330*

*Statistically significant at the 1% level. **Statistically significant at the 5% level. ***Statistically significant at the 10% level.

The estimated dynamic conditional Pearson's correlation matrix $\hat{\boldsymbol{P}}_T$ and the corresponding dynamic conditional Kendall's tau matrix, which is denoted by $\hat{\boldsymbol{P}}_{\tau,T}$, for the last period of time of the samples (i.e. July 27, 2016) can be found in Table 9 and 10 respectively. As the matrices are symmetric, we only show the correlation coefficients in the upper triangular part of the matrix. Furthermore, on the top of each tables indicate the stock indices by only their starting letter due to a lack of space.

The minimum eigenvalue of $\hat{\boldsymbol{P}}_{\tau,T}$ is $\lambda_{min} = 0.2158 > 0$ and so by Theorem 4.1 we deduce that the matrix in positive-definite, as required.

We make the following observations from the dynamic conditional Kendall's tau matrix $\hat{\boldsymbol{P}}_{\tau,T}$. [22]

- All correlation coefficients are positive.
- For a given stock index from a certain continent (e.g. Europe), its correlation with another stock index from the same continent is generally higher than the correlation between it and a stock index from another continent (e.g. Asia).
- The Hang Seng (HS, Asia) is relatively strong correlated with other Asian stock indices, with a minimum correlation coefficient of 0.28.
- The correlations between SSE (Asia) and the other Asian stock indices are relatively strong compared to those between SSE and the non-Asian stock indices.
- Four Asian stock indices, namely Hang Seng, Nikkei 225 (NIKKEI), S&P/ASX (ASX), BSE Sensex 30 (BSE), have correlation coefficients greater than 0.20 with the European stock indices in 26 out of 28 cases. On the other hand, these four Asian stock indices have correlation coefficients lower than or equal to 0.20 with the North-America and South-America stock indices for 22 out of 24 cases.
- Each Asian stock index has lowest correlation with a stock index from North-America or South-America, except for SSE and SMI (Europe), where the correlation is equally low as the correlation between SSE and Dow 30 (DOW, North-America).
- The BIST 100 (BIST, Turkey) stock index shows significantly less correlation with the other European stock indices than the other European stock indices with each other.
- The only two considered stock indices from North-America are seemingly stronger correlated with each other than each of them with any other stock index.

[22] Since the function $\arcsin(x)$ is strictly monotonically increasing in $x \in [-1, 1]$, similar observations can be made when we consider the dynamic conditional correlation matrix $\hat{\boldsymbol{P}}_T$. The only difference would be the correlation values.

Table 9: Estimated dynamic conditional correlation matrix $\hat{\boldsymbol{P}}_T$ of the t-GJR model, where T is the last period of time of the samples (T: July 27, 2016).

	H	K	N	A	B	S	T	A	B	B	C	D	I	S	D	T	B	I	M	B
HS	1.00	0.46	0.56	0.57	0.50	0.42	0.57	0.42	0.40	0.32	0.42	0.40	0.39	0.40	0.23	0.31	0.27	0.34	0.21	0.28
KLSE		1.00	0.43	0.40	0.31	0.19	0.44	0.23	0.22	0.26	0.22	0.21	0.22	0.23	0.14	0.18	0.21	0.27	0.13	0.19
NIKKEI			1.00	0.54	0.37	0.23	0.52	0.35	0.35	0.24	0.36	0.34	0.37	0.38	0.19	0.24	0.21	0.27	0.17	0.20
ASX				1.00	0.37	0.19	0.48	0.35	0.33	0.25	0.34	0.32	0.32	0.36	0.18	0.28	0.22	0.29	0.17	0.25
BSE					1.00	0.21	0.41	0.39	0.40	0.35	0.40	0.40	0.40	0.37	0.30	0.28	0.26	0.33	0.20	0.27
SSE						1.00	0.27	0.13	0.13	0.12	0.13	0.12	0.12	0.10	0.10	0.13	0.13	0.13	0.10	0.16
TSEC							1.00	0.31	0.29	0.28	0.31	0.30	0.30	0.29	0.20	0.25	0.23	0.29	0.21	0.26
AEX								1.00	0.88	0.40	0.93	0.90	0.85	0.82	0.59	0.54	0.39	0.55	0.36	0.37
BFX									1.00	0.38	0.88	0.85	0.83	0.79	0.55	0.50	0.35	0.53	0.32	0.34
BIST										1.00	0.40	0.39	0.40	0.34	0.28	0.26	0.31	0.30	0.25	0.31
CAC											1.00	0.93	0.89	0.82	0.58	0.52	0.37	0.54	0.35	0.36
DAX												1.00	0.84	0.81	0.59	0.51	0.36	0.54	0.34	0.36
IBEX													1.00	0.74	0.56	0.50	0.41	0.52	0.36	0.38
SMI														1.00	0.50	0.46	0.31	0.49	0.29	0.28
DOW															1.00	0.67	0.53	0.66	0.45	0.38
TSX																1.00	0.53	0.60	0.46	0.46
BVSP																	1.00	0.57	0.54	0.43
IPC																		1.00	0.44	0.40
MERV																			1.00	0.37
BVL																				1.00

Table 10: Estimated dynamic conditional Kendall's tau matrix $\hat{\boldsymbol{P}}_{\tau,T}$ of the t-GJR model, where T is the last period of time of the samples (T: July 27, 2016).

	H	K	N	A	B	S	T	A	B	B	C	D	I	S	D	T	B	I	M	B
HS	1.00	0.30	0.38	0.39	0.33	0.28	0.39	0.28	0.26	0.21	0.28	0.26	0.26	0.26	0.15	0.20	0.18	0.22	0.13	0.18
KLSE		1.00	0.28	0.26	0.20	0.12	0.29	0.15	0.14	0.17	0.14	0.13	0.14	0.15	0.09	0.12	0.14	0.17	0.09	0.12
NIKKEI			1.00	0.37	0.24	0.15	0.35	0.23	0.23	0.15	0.24	0.22	0.24	0.25	0.12	0.15	0.14	0.17	0.11	0.13
ASX				1.00	0.24	0.12	0.32	0.23	0.22	0.16	0.22	0.21	0.21	0.23	0.12	0.18	0.14	0.18	0.11	0.16
BSE					1.00	0.14	0.27	0.26	0.26	0.23	0.26	0.26	0.26	0.24	0.19	0.18	0.17	0.21	0.13	0.17
SSE						1.00	0.17	0.08	0.08	0.08	0.08	0.08	0.08	0.06	0.06	0.08	0.08	0.08	0.07	0.10
TSEC							1.00	0.20	0.19	0.18	0.20	0.19	0.20	0.19	0.13	0.16	0.15	0.18	0.13	0.17
AEX								1.00	0.68	0.26	0.77	0.71	0.65	0.62	0.40	0.37	0.26	0.37	0.23	0.24
BFX									1.00	0.25	0.68	0.65	0.62	0.58	0.37	0.33	0.23	0.35	0.21	0.22
BIST										1.00	0.26	0.25	0.26	0.22	0.18	0.17	0.20	0.19	0.16	0.20
CAC											1.00	0.76	0.69	0.62	0.39	0.35	0.24	0.37	0.23	0.24
DAX												1.00	0.64	0.60	0.40	0.34	0.24	0.36	0.22	0.23
IBEX													1.00	0.53	0.38	0.34	0.27	0.35	0.24	0.25
SMI														1.00	0.33	0.30	0.20	0.32	0.19	0.18
DOW															1.00	0.46	0.36	0.46	0.30	0.24
TSX																1.00	0.36	0.41	0.31	0.30
BVSP																	1.00	0.39	0.36	0.28
IPC																		1.00	0.29	0.26
MERV																			1.00	0.24
BVL																				1.00

- Each non-Asian stock index has the lowest correlation with the SSE.
- The three highest correlations are between AEX − CAC (0.77), CAC − DAX (0.76) and AEX − DAX (0.71).
- The three lowest correlations are between SSE − SMI (0.06), SSE − DOW (0.06) and SSE − MERV (0.07).

Thus the dynamic volatilities in the losses of the stock indices show positive interdependence. This effect is in particular strong between stock indices from the same continent. The relatively low correlation between BIST and the other European stock indices could be explained from a geographical and political point of view (Kirisci 2005, Corke et al. 2014). [23] Namely, BIST is the only East-European stock index considered in our study and on top of that, Turkey has a relatively poor (political) relationship with the other (West-)European countries for a period of time. [24]

For the stock indices that correspond to at least one of the three highest correlations (i.e. AEX, CAC and DAX), the multiplots of the losses and the relating scatter, density and contour plots can be found in Figure 10 and Figure 11, respectively. [25] The plots in Figure 11 are created from the innovations of the fitted t-GJR model, after transforming to uniform. The strong correlation between these three stock indices indicates that their losses time series have relatively many correspondences. This can informally be confirmed by comparing their losses time series. In Figure 10 we can see that their losses time series indeed evolve similarly. From the plot in Figure 11 we can confirm that the losses time series are closely related. Moreover, the plots also show the existence of strong dependency in the tails. These results do not come as a complete surprise, because these stock indices are from West-European countries that together have a relatively strong political relationship.

How do the correlations between these stock indices change over time? In Figure 12 we can see the time series of the estimated dynamic conditional Kendall's taus for AEX − CAC, CAC − DAX and AEX − DAX over the time period January 1, 2000 to July 31, 2016. In each of the three plots, we recognize abrupt decreases and increases in correlation around the period of 2001. This is likely to be the result of the dot-com bubble (1995 − 2001). During that period the value of many stock markets from Internet-related sectors increased rapidly at a certain time period, but these increases were shortly afterwards followed by rapid decreases. In addition, there are some more than usual increases and decreases around the period of 2007 (except for CAC − DAX) and 2014, but overall the degree of the correlations stays rather stable after the dot-com bubble. So interestingly, there is no clear evidence that the Russian financial crisis (2014)

[23] BIST has also relatively low correlations with non-European stock indices.
[24] Actually, Turkey is a transcontinental country in Eurasia.
[25] See Figure 17 in Appendix E for single plots of the losses of the stock indices.

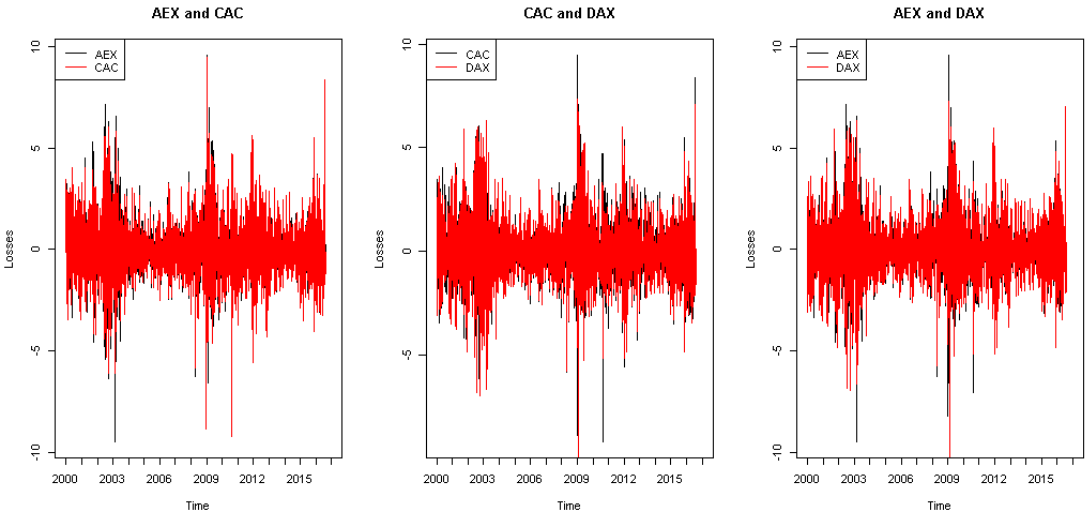


Figure 10: Multiplots of the losses of the AEX, CAC and DAX.

and, in particular, the global financial crisis (2007 – 2008) have a large effect on the degree of correlations between these stock indices.

With this result we may approve the general claim that the correlation between stock indices increases (respectively decreases) after a rapid decrease (respectively increase) of stock market values. However, the degree to which this effect is the case differs per event and can even be relatively limited.

Diversification of a portfolio is a general investment strategy of composing a portfolio of assets with little to no correlation (Hull 2015). Consider stock indices, in this way a large decline in one stock index will most likely not be followed by a (large) decline in one or more of the other stock indices in the portfolio. This strategy has become a standard in investing as it often achieves the long-term financial goals, while minimizing risk. Thus, by following the diversification strategy, it is not recommended to invest in a number of highly correlated stock indices (e.g. both AEX and CAC). Though it is interesting to notice that the correlation between these stock indices may not be as extreme as is implicated by the Pearson's correlation coefficients, see Table 9. Hence, the Kendall's tau indicates that the degree of diversity between these stock indices is larger than is expected when considering the Pearson's correlation coefficient.

For the stock indices that correspond to at least one of the three lowest correlations (i.e. SSE, SMI, DOW and MERV), the multiplots of the losses and the relating scatter, density and contour plots can be found in Figure 13 and Figure 14, respectively. [26] The plots in Figure 14 are created from the innovations of the fitted t-GJR model, after transforming to uniform. The low (and almost zero) correlation coefficients for SSE with SMI (Europe), SSE with DOW (North-America) and SSE with MERV (South-America) indicate that the losses time se-

[26] See Figure 17 in Appendix E for single plots of the losses of the stock indices.

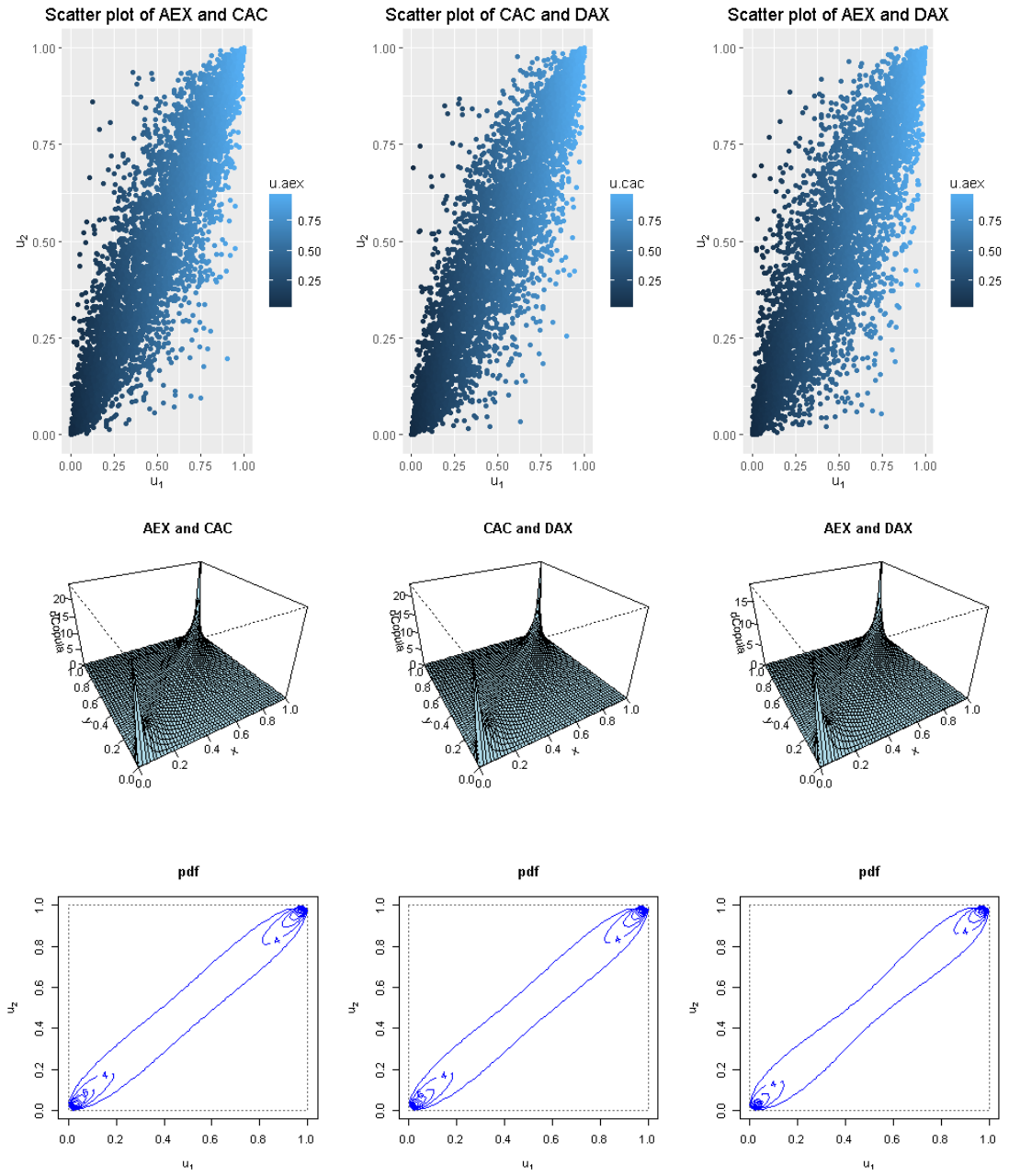


Figure 11: Scatter, density and contour plots of AEX − CAC, CAC − DAX and AEX − DAX. The plots are created from the innovations of the fitted *t*-GJR model, after transforming to uniform.

ries of these pairs of stock indices have nearly no correspondences. This can informally be confirmed by comparing their losses time series, see Figure 13. From the plots in Figure 14 we can confirm that the losses time series are poorly related. Moreover, the plots also show that there is very little dependency in the tails.

It is noteworthy that the SSE is involved in each of the three low correlations. This is in line with earlier claims that the Chinese stock market is weakly cor-

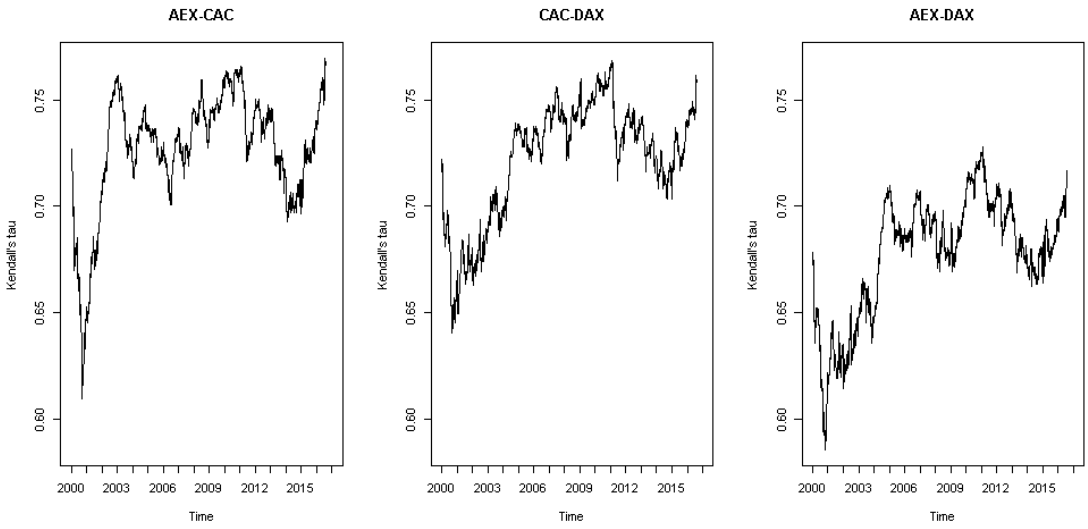


Figure 12: Estimated dynamic conditional Kendall's taus for AEX − CAC, CAC − DAX and AEX − DAX over the time period January 1, 2000 to July 31, 2016.

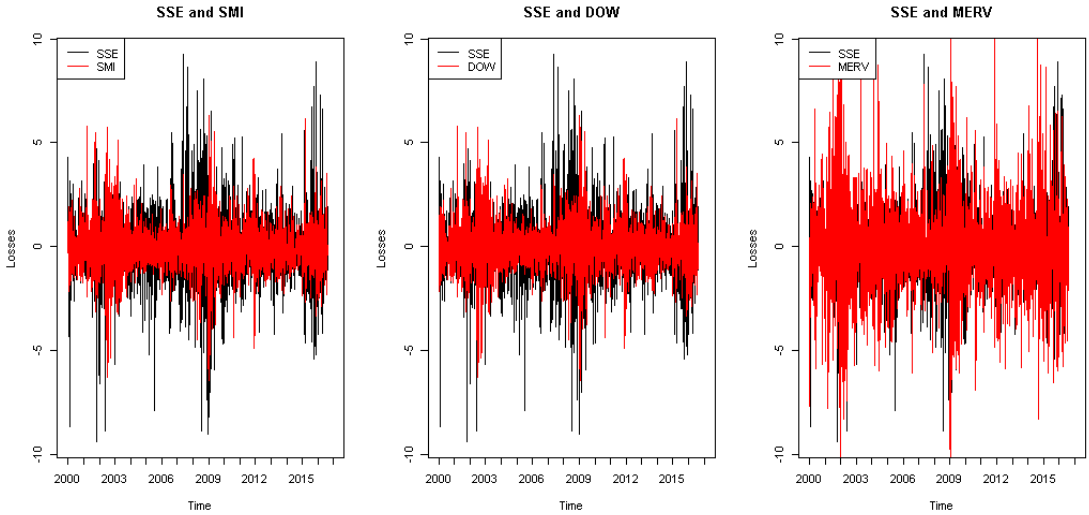


Figure 13: Multiplots of the losses of the SSE, SMI, DOW and MERV.

related with non-Chinese stock markets, see Valukonis (2013) and Labuszewski (2014). According to Valukonis (2013), this could be explained by the fact that the Chinese stock market was completely isolated from the global market for a time period. Currently, the strict rules by the Chinese mainland authorities causes the Chinese stock market to still bear with a low integrity into the global market. Accordingly, by following the diversification strategy it is suggested to invest in the Chinese stock market in order to hedge against other (non-Chinese) markets (S.R. and C.C.W. 2014).

For the three strongest correlated stock indices (according to the estimated dynamic conditional Kendall's tau as of July 31, 2016), namely AEX, CAC and DAX,

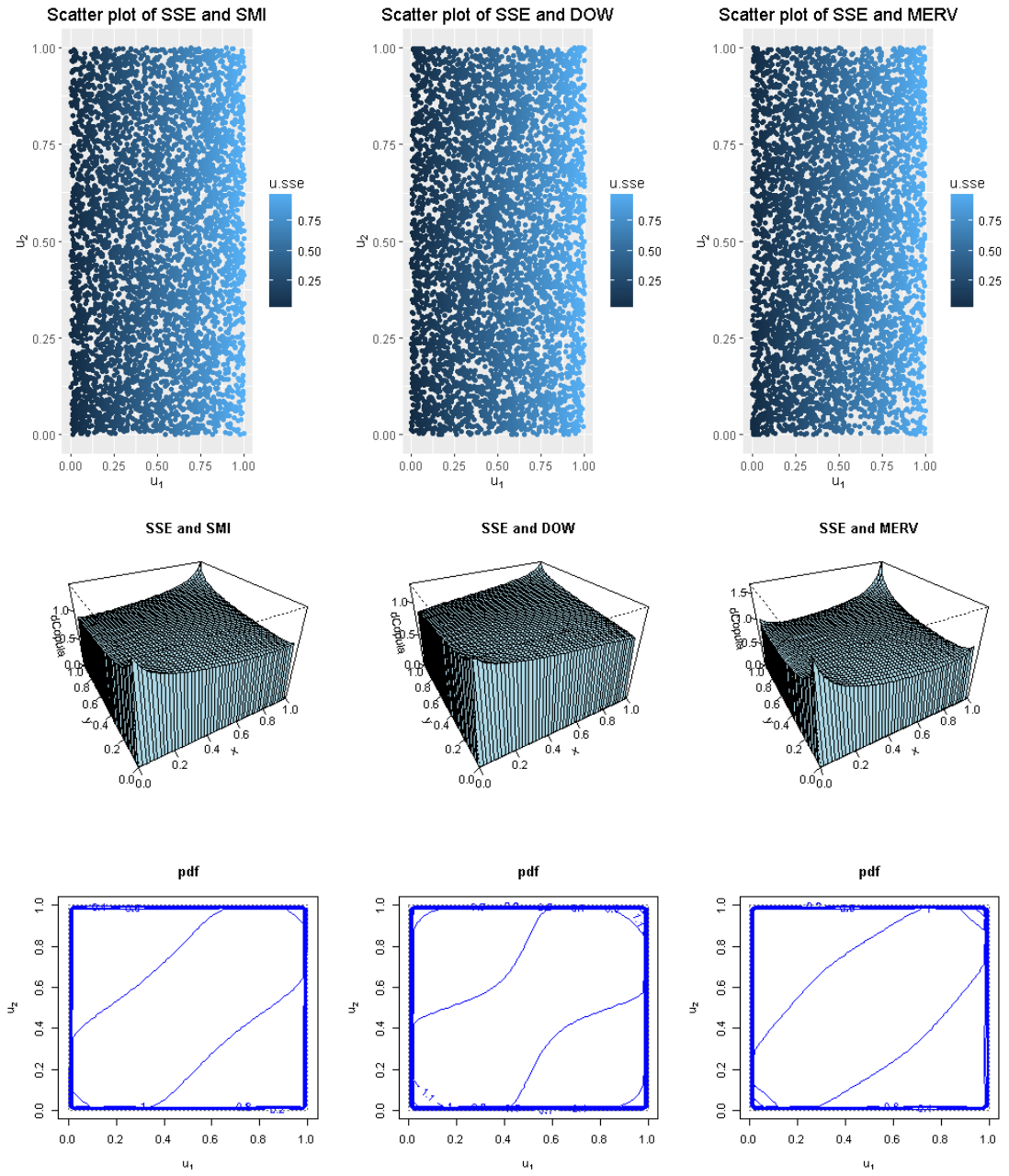


Figure 14: Scatter, density and contour plots of SSE – SMI, SSE – DOW and SSE – MERV. The plots are created from the innovations of the fitted t-GJR model, after transforming to uniform.

we have seen that their correlation fluctuated heavily around the period of 2001. After this period the correlations are rather stable until July 31, 2016, but with some more than usual increases and decreases around 2007 and 2014. How does the correlation between SSE – SMI, SSE – DOW and SSE – MERV change over time? In Figure 15 we can see the time series of the estimated dynamic conditional Kendall's taus for these three pairs over the time period January 1, 2000 to July 31, 2016.

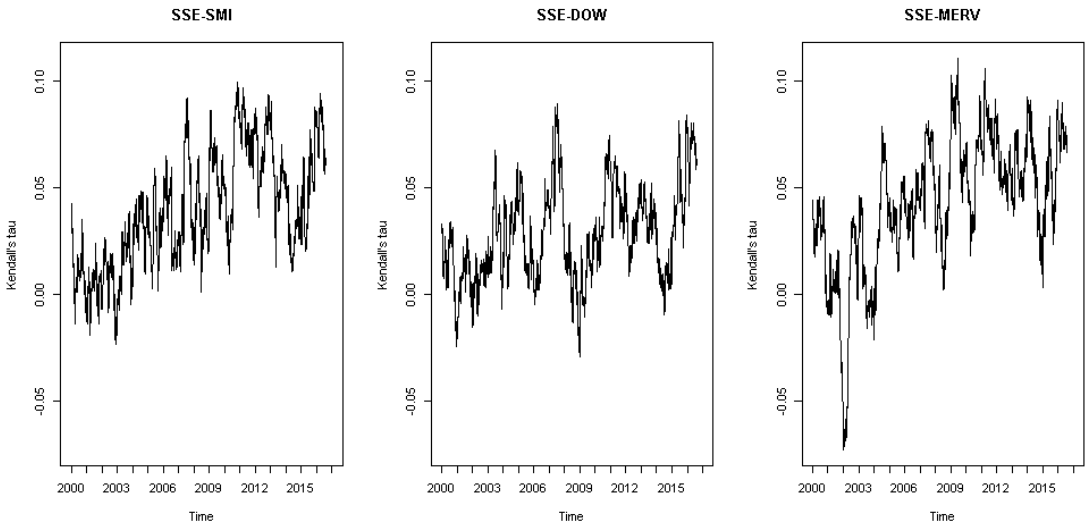


Figure 15: Estimated dynamic conditional Kendall's taus for SSE − SMI, SSE − DOW and SSE − MERV over the time period January 1, 2000 to July 31, 2016.

For SSE − SMI there are no large changes in the correlation during the whole time period, and so surprisingly also not during the dot-com bubble, global financial crisis and Russian financial crisis. For SSE − DOW there are rapid decreases during the financial crises, in particular during the global financial crisis. This is in line with the general claim that the correlation between financial assets decreases during financial crises due to the decline in value of the assets. Furthermore, it is noteworthy that SSE and DOW are negatively correlated in the period of the financial crises. Lastly, for SSE − MERV we notice some very large fluctuations in the correlation, but these only arise for a short period of time. Moreover, these fluctuations in particular happen around the period of the financial crises.

With these results we could infer that the change in correlation over time is different for strongly correlated stock indices and poorly correlated stock indices. Actually, we even observe that the change in correlation can be significantly different for different pairs of stock indices. More explicitly, the change in correlation of a pair of stock indices with a certain degree of correlation can be significantly different for another pair of stock indices with (almost) the same degree of correlation. Moreover, there is no clear evidence that a financial crisis necessarily affects the degree of correlation.

So far we only studied the dynamic conditional correlation coefficients for describing the dependence structure of the stock indices. Now, let us study the extremal dependence of the stock indices by estimating the TIs (also known as upper/lower tail-dependence coefficients) for every pair of stock indices. Recall that a TI is a probability that a large drop (or large rise) of one time series is followed by large drops (or large rise) of another time series. And therefore, this

quantity is used to measure the tendency of markets to crash or boost together. The estimated TIs can be found in Table 11. Not too surprisingly, there are some correspondences with the estimated dynamic conditional matrices $\hat{\boldsymbol{P}}_T$ and $\hat{\boldsymbol{P}}_{\tau,T}$ in respectively Table 9 and 10. We make the following observations from the table of estimated TIs.

- For a given stock index from a certain continent (e.g. Europe), its TI together with another stock index from the same continent is generally higher than that with a stock index from another continent (e.g. Asia).
- The TIs of the SSE together with any other stock index is at most 0.03.
- The BIST 100 (BIST, Turkey) stock index has significantly lower TIs together with the other European stock indices than the other European stock indices with each other.
- The only two considered stock indices from North-America have a larger TI together than each of them with any other stock index.
- Every non-Asian stock index has (almost) no tail-dependence with each Asian stock index. The maximum TI between a European stock index and an Asian stock index is 0.02, while the maximum TI between an (North and South) American stock index and an Asian stock index is 0.03.
- The three highest TIs are between AEX − CAC (0.60), CAC − DAX (0.59) and BFX − CAC (0.51).

The TIs of SSE together with the other stock indices (including the other Asian stock indices) is at most 0.03. This highlights the very low integrity of the Chinese stock market into the global market. The relatively low TIs of BIST with the other European stock indices demonstrate the poor integration of the Turkish stock market into the West-European stock market. [27] These results are in line with the conclusion that we drawn from the estimated dynamic conditional correlations. Most noticeable is the amount of TIs that are (close to) zero: 153 out of 190 TIs with two different stock indices are less than or equal to 0.05. This shows that many markets have little to no tendency to crash or boost together.

It is important to remark that a TI does not explicitly take into consideration the development through time, as opposed to the dynamic conditional correlations. That is, it does not take into account the time periods *when* large drops (or large rise) of one stock index is followed by a large drop (or large rise) by another stock index, but rather the *frequency* of these events. Therefore it only measures the tendency of these events happening on the same day, but not within the next few days. So given two stock indices with a high TI, it may well be the case that one of the two stock indices follows the other stock index perfectly in terms of the extreme comovements, but with some time lag.

[27] BIST has also relatively low TIs with non-European stock indices.

Table 11: Estimated tail indices using innovations from the *t*-GJR model.

	H	K	N	A	B	S	T	A	B	B	C	D	I	S	D	T	B	I	M	B
HS	1.00	0.03	0.02	0.11	0.02	0.01	0.12	0.02	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00
KLSE		1.00	0.05	0.01	0.01	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NIKKEI			1.00	0.16	0.01	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ASX				1.00	0.02	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BSE					1.00	0.00	0.02	0.00	0.01	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.03
SSE						1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
TSEC							1.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
AEX								1.00	0.50	0.05	0.60	0.50	0.39	0.40	0.12	0.02	0.05	0.00	0.08	0.01
BFX									1.00	0.04	0.51	0.43	0.37	0.37	0.07	0.06	0.03	0.01	0.04	0.02
BIST										1.00	0.05	0.04	0.02	0.03	0.01	0.01	0.01	0.02	0.00	0.00
CAC											1.00	0.59	0.47	0.45	0.13	0.02	0.04	0.02	0.05	0.01
DAX												1.00	0.39	0.36	0.15	0.03	0.04	0.00	0.04	0.01
IBEX													1.00	0.25	0.09	0.01	0.03	0.01	0.04	0.00
SMI														1.00	0.04	0.03	0.02	0.00	0.02	0.02
DOW															1.00	0.16	0.13	0.11	0.08	0.01
TSX																1.00	0.06	0.06	0.11	0.07
BVSP																	1.00	0.12	0.14	0.02
IPC																		1.00	0.08	0.05
MERV																			1.00	0.06
BVL																				1.00

5 Extension

By modelling the joint distribution with copulas it is often of interest to choose the copula function, among many other copula functions, that provides the best fit to the given data. This can be achieved with the help of the so-called goodness-of-fit (GOF) tests. With this argument we extend our study by performing GOF tests in order to determine the joint distribution, in the form of a static copula, that fits the losses of the stock indices best. In particular, we consider three different GOF statistics. We additionally consider four Archimedean copulas, namely the Clayton, Frank, Gumbel and Joe copula, next to the Gaussian and t copula. Details about the Archimedean copulas can be found in Appendix D.

Let us motivate the general idea behind the extension. For this, let us briefly recall the methodology of fitting the copula-MGARCH model to the n-dimensional dataset of stock index losses (see Section 2.6.3). In the first step we determine for every stock index the most appropriate innovations distribution that will be used for the construction of the copula margins. This is supported by the fact that from equations (13) – (15) the loss distribution of the stock indices is characterized by the corresponding innovations distribution. Then in the second step we transform the innovations of the fitted univariate GARCH-type models to standard uniform using PIT. Lastly, in the third step these transformed uniform innovations, $u_{i,t} \in [0,1]$, $i = 1, \dots, n$, $t = 1, \dots, T$, would then be used to estimate the parameters of the copula-MGARCH model. Now for the GOF tests, instead of doing this third step, we directly fit the transformed uniform innovations obtained from Step 1 and 2 to an n-dimensional copula. For the given dataset, the fit of this copula can be tested by the GOF tests for copulas.

We could use the information of the GOF tests to decide for the most appropriate joint distribution (or rather say copula) for the construction of the copula-MGARCH model. Notice that to perform these GOF tests we do not need to take into account the DCC specifications. Thus this extension can be considered as an intermediate step in the fitting methodology of the copula-MGARCH model in Section 2.6.3. It provides practitioners the option to choose for an appropriate copula function before applying it in the full copula-MGARCH model. We should make the remark that the GOF tests do not necessarily guarantee that the appropriate copulas for fitting are also appropriate for forecasting.

5.1 Goodness-of-fit tests for copulas

The theory in this section is adopted from Genest et al. (2009), if not mentioned differently.

For the GOF tests of copulas we are interested in testing the null hypothesis that

an underlying copula C belongs to some class of copulas $\mathscr{C}_0 = \{C_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{O}\}$, for some open subset $\boldsymbol{O} \subseteq \mathbb{R}^p$, $p \in \mathbb{N}$:

$$H_0 : C \in \mathscr{C}_0 \quad \text{vs.} \quad H_1 : C \notin \mathscr{C}_0. \tag{58}$$

Recall from Section 2.6 that for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, the transformed innovations are denoted by $U_{i,t} = F_i(X_{i,t})$, which follows a standard uniform distribution. A natural estimator of the underlying copula C is the *empirical copula*, which is nonparametric and defined as:

$$C_T(\boldsymbol{u}) = \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(U_{1,t} \leq u_1, \ldots, U_{n,t} \leq u_n) \tag{59}$$

for all $\boldsymbol{u} = (u_1, \ldots, u_n) \in [0,1]^n$.

A well-known GOF test for copulas compares the distance between the empirical copula C_T, which is a non-parametric estimate of C, and an estimated parametric copula $C_{\hat{\theta}}$ from an estimated parameter $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, obtained under H_0 (Genest et al. 2013). More explicitly, the GOF test is based on

$$\mathbb{C}_T(\boldsymbol{u}) = \sqrt{T}(C_T(\boldsymbol{u}) - C_{\hat{\theta}}(\boldsymbol{u})).$$

Furthermore, one of the three GOF statistics of our interest is based on the L_2-norm between C_T and $C_{\hat{\theta}}$ (Genest et al. 2013):

$$\begin{aligned} S_T &= \int_{[0,1]^n} \mathbb{C}_T(\boldsymbol{u})^2 dC_T(\boldsymbol{u}) \\ &= T \int_{[0,1]^n} \left(C_T(u_1, \ldots, u_n) - C_{\hat{\theta}}(u_1, \ldots, u_n)\right) dC_T(u_1, \ldots, u_n) \\ &= \sum_{t=1}^{T} \left(C_T(U_{1,t}, \ldots, U_{n,t}) - C_{\hat{\theta}}(U_{1,t}, \ldots, U_{n,t})\right)^2, \end{aligned} \tag{60}$$

which is related to the Cramér-von Mises statistics.

For this GOF statistic we approximate the p-values by a parametric bootstrap approach. The p-value can be used to test H_0 in (58). For a given copula, the parametric bootstrap approach proceeds as follows (see Appendix A in Genest et al. (2009)):

1. Using $\boldsymbol{U}_t = (U_{1,t}, \ldots, U_{n,t})$, $t = 1, \ldots, T$, estimate the copula parameter $\boldsymbol{\theta}$, i.e. $\hat{\boldsymbol{\theta}}$, and compute the empirical copula C_T as of (59) and $C_{\hat{\theta}}$ (e.g. fitted Gaussian copula).

2. Using the computed C_T and $C_{\hat{\theta}}$, compute the GOF statistic S_T as of (60).

3. (*Bootstrap*) Choose some large integer M, which corresponds to the number of bootstraps. For $m = 1, \ldots, M$, perform the following steps:

   (a) Simulate samples $\boldsymbol{U}_{1,m}, \ldots, \boldsymbol{U}_{T,m}$ from the n-dimensional fitted copula $C_{\hat{\boldsymbol{\theta}}}$.

   (b) Using these simulated samples, estimate the copula parameter $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}^{(m)}$, and compute the empirical copula C_T as of (59) and $C_{\hat{\boldsymbol{\theta}}}$. Denote them respectively by $C_{T,m}$ and $C_{\hat{\boldsymbol{\theta}},m}$.

   (c) Using the computed $C_{T,m}$ and $C_{\hat{\boldsymbol{\theta}},m}$, compute the GOF statistic S_T as of (60). Denote this by $S_{T,m}$.

4. Approximate the p-value for testing H_0 in (58) by $\frac{1}{M}\sum_{m=1}^{M} \mathbb{1}(S_{T,m} > S_T)$.

Other GOF tests for copulas is based on a probability integral transformation by Rosenblatt (1952). This probability integral transformation of a copula C (with parameter $\boldsymbol{\theta}$) is the mapping $\mathcal{R}_{\boldsymbol{\theta}} : (0,1)^n \to (0,1)^n$ for which

$$\mathcal{R}_{\boldsymbol{\theta}}(u_1, \ldots, u_n) = (e_1, \ldots, e_n),$$

where $e_1 = u_1$ and

$$e_i = \frac{\partial^{i-1} C(u_1, \ldots, u_i, 1, \ldots, 1)}{\partial u_1 \cdots \partial u_{i-1}} \Bigg/ \frac{\partial^{i-1} C(u_1, \ldots, u_{i-1}, 1, \ldots, 1)}{\partial u_1 \cdots \partial u_{i-1}}$$

for $i = 2, \ldots, n$. According to Genest et al. (2009), this mapping decomposes a random vector with a certain distribution into mutually independent components that are standard uniformly distributed. Furthermore, an important consequence is that $\boldsymbol{U} = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_T)'$ is distributed as a copula C if and only if $\mathcal{R}_{\boldsymbol{\theta}}(\boldsymbol{U})$ follows the distribution of the n-dimensional *independence copula*, where the independence copula is defined as

$$C_{ind}(e_1, \ldots, e_n) = e_1 \times \cdots \times e_n \tag{61}$$

for all $e_1, \ldots, e_n \in [0,1]$. In other words:

$$\boldsymbol{U} \sim C \iff \mathcal{R}_{\boldsymbol{\theta}}(\boldsymbol{U}) \sim C_{ind}.$$

As a result, the null hypothesis in (58) can be restated as

$$H_0 : \mathcal{R}_{\boldsymbol{\theta}}(\boldsymbol{U}) \sim C_{ind} \quad \text{vs.} \quad H_1 : \mathcal{R}_{\boldsymbol{\theta}}(\boldsymbol{U}) \nsim C_{ind} \tag{62}$$

for some $\boldsymbol{\theta} \in \boldsymbol{O}$. Therefore, given observations $\boldsymbol{U}_t = (U_{1,t}, \ldots, U_{n,t})$, $t = 1, \ldots, T$, under H_0 we can assume that the pseudo-observations $\boldsymbol{E}_1 = \mathcal{R}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{U}_1), \ldots, \boldsymbol{E}_T = \mathcal{R}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{U}_T)$ are a sample from the independence copula C_{ind}.

Genest et al. (2009) introduced two GOF statistics that are constructed equivalently as the GOF statistic S_T before. Namely, a natural estimator of the independence copula C_{ind} is the empirical distribution function

$$D_T(\boldsymbol{u}) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(\boldsymbol{E}_t \leq \boldsymbol{u}) \tag{63}$$

for all $\boldsymbol{u} = (u_1, \ldots, u_n) \in [0,1]^n$. Accordingly, the two GOF statistics for copulas by Genest et al. (2009) compares the distance between the empirical distribution function D_T and the independence copula C_{ind}:

$$\begin{aligned} S_T^{(B)} &= T \int_{[0,1]^n} \left(D_T(\boldsymbol{u}) - C_{ind}(\boldsymbol{u})\right)^2 d\boldsymbol{u} \qquad (64)\\ &= \frac{T}{3^n} - \frac{1}{2^{n-1}} \sum_{t=1}^{T} \prod_{k=1}^{n} \left(1 - E_{k,t}^2\right) + \frac{1}{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \prod_{k=1}^{n} \left(1 - \max(E_{k,t}, E_{k,s})\right), \\ S_T^{(C)} &= T \int_{[0,1]^n} \left(D_T(\boldsymbol{u}) - C_{ind}(\boldsymbol{u})\right)^2 dD_T(\boldsymbol{u}) \qquad (65)\\ &= \sum_{t=1}^{T} \left(D_T(\boldsymbol{E}_t) - C_{ind}(\boldsymbol{E}_t)\right)^2. \end{aligned}$$

For these two GOF statistics we approximate the p-values also by a parametric bootstrap approach. The p-value can be used to test H_0 in (62). For a given copula, the parametric bootstrap approach proceeds as follows (see Appendix D in Genest et al. (2009)):

1. Using $\boldsymbol{U}_t = (U_{1,t}, \ldots, U_{n,t})$, $t = 1, \ldots, T$, estimate the copula parameter $\boldsymbol{\theta}$, i.e. $\hat{\boldsymbol{\theta}}$, and compute the pseudo-observations $\boldsymbol{E}_t = \mathcal{R}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{U}_t)$, the independence copula C_{ind} as of (61), the empirical distribution function D_T as of (63) and $C_{\hat{\boldsymbol{\theta}}}$ (e.g. fitted Gaussian copula).

2. Using the computed C_{ind} and D_T, compute the GOF statistic $S_T^{(B)}$ as of (64) (resp. $S_T^{(C)}$ as of (65)).

3. (*Bootstrap*) Choose some large integer M, which corresponds to the number of bootstraps. For $m = 1, \ldots, M$, perform the following steps:

   (a) Simulate samples $\boldsymbol{U}_{1,m}, \ldots, \boldsymbol{U}_{T,m}$ from the n-dimensional fitted copula $C_{\hat{\boldsymbol{\theta}}}$.

   (b) Using these simulated samples, estimate the copula parameter $\boldsymbol{\theta}$, say $\hat{\boldsymbol{\theta}}^{(m)}$, and compute the pseudo-observations $\boldsymbol{E}_{t,m} = \mathcal{R}_{\hat{\boldsymbol{\theta}}^{(m)}}(\boldsymbol{U}_{t,m})$, the independence copula C_{ind} as of (61), the empirical distribution D_T as of (63) and $C_{\hat{\boldsymbol{\theta}}}$. Denote them respectively by $C_{ind,m}$, $D_{T,m}$ and $C_{\hat{\boldsymbol{\theta}},m}$.

   (c) Using the computed $C_{ind,m}$ and $D_{T,m}$, compute the GOF statistic $S_T^{(B)}$ as of (64) (resp. $S_T^{(C)}$ as of (65)).

4. Approximate the p-value for testing H_0 in (62) by $\frac{1}{M}\sum_{m=1}^{M}\mathbb{1}(S_{T,m}^{(B)} > S_T^{(B)})$ (resp. $\frac{1}{M}\sum_{m=1}^{M}\mathbb{1}(S_{T,m}^{(C)} > S_T^{(C)})$).

5.2 Illustration

The proposed extension can be performed on any d-dimensional dataset, $2 \leq d \leq n$, and so is not restricted to the complete n-dimensional dataset. This is because the transformed innovations $u_{i,t} \in [0,1]$ from the stock index loss $l_{i,t}$ are obtained from an univariate distribution function, i.e. because the extension is performed before the multivariate modelling. In this section we illustrate the extension by considering a 2-, 3- and 4-dimensional dataset instead of the 20-dimensional dataset. More explicitely, we consider the transformed uniform innovations of the lowest correlated pair of stock indices (i.e. SSE and DOW), the three highest (positively) correlated stock indices (i.e. AEX, CAC and DAX, see Section 4.3.2) and the four stock indices corresponding to the three lowest correlations (i.e. DOW, MERV, SMI and SSE, see Section 4.3.2). Moreover, in the parametric bootstrap approach of the GOF tests we perform $M = 1,000$ bootstraps. The reason that we do not perform the extension on the 20-dimensional dataset is due to the extremely high computational costs. This is because the number of parameters to be estimated for the Gaussian and t copula increases quadratically with an increase in the dimension of the dataset. In combination with the parametric bootstrap this leads to extremely high computational time.

In our main study we obtained transformed uniform innovations $u_{i,t} \in [0,1]$ from the stock index loss $l_{i,t}$ using the corresponding most appropriate univariate distribution function found in Section 4.1. The result of the GOF tests for copulas using the transformed innovations from SSE and DOW can be found in Table 12, from AEX, CAC and DAX in Table 13, and from DOW, MERV, SMI and SSE in Table 14. For each of the six copulas, the values of the three computed GOF statistics are given in the table as well as whether the result is significant or not. The computed GOF statistics for which the p-value is greater than or equal to 0.05 are indicated by a bold.

Table 12: Result of the GOF tests for copulas using transformed uniform innovations of SSE and DOW.

GOF statistic	Gaussian	t	Clayton	Frank	Gumbel	Joe
S_T	**0.0455**	**0.0129**	**0.0173**	**0.0115**	**0.0140**	**0.0232**
$S_T^{(B)}$	**0.0133**	**0.0135**	**0.0197**	**0.0127**	**0.0139**	**0.0216**
$S_T^{(C)}$	**0.0141**	**0.0143**	**0.0192**	**0.0134**	**0.0153**	**0.0247**

The GOF statistics for which the p-value is greater than or equal to 0.05 are indicated by a bold.
*Significant at the 1% significance level. **Significant at 5% significance level. ***Significant at 10% significance level.

In Section 4.3 we found that SSE and DOW are poorly correlated and have (almost) no tail-dependence. According to this result we would expect that each

of the six copulas are appropriate for fitting the joint loss distribution of these stock indices, because these copulas can capture zero tail-dependence. [28] Indeed, from Table 12 we see that each copula have all the GOF tests being not statistical significant at the 5% level. Thus from the GOF test result we have evidence that each of these copulas are appropriate for fitting the joint loss distribution of SSE and DOW.

Table 13: Result of the GOF tests for copulas using transformed uniform innovations of AEX, CAC and DAX.

GOF statistic	Gaussian	t	Clayton	Frank	Gumbel	Joe
S_T	**0.0595**	**0.0451**	6.4450*	**0.4370*****	0.4957**	4.0932*
$S_T^{(B)}$	0.2939*	**0.0416**	2.7129*	1.1635*	1.9635*	6.3440*
$S_T^{(C)}$	0.3279*	**0.0391**	5.6033*	1.5573*	2.0564*	7.8088*

The GOF statistics for which the p-value is greater than or equal to 0.05 are indicated by a bold.
*Significant at the 1% significance level. **Significant at 5% significance level. ***Significant at 10% significance level.

Table 13 shows that only the t copula have all GOF tests being not statistical significant at the 5% level. Hence we have evidence that the t copula is appropriate for fitting the joint distribution of the stock index losses of AEX, CAC and DAX. This result could be expected, because the t copula is able to capture high correlation and both the upper and lower tail-dependence. It is noteworthy that the result of S_T indicates that also the Gaussian and Frank copula are appropriate for fitting the joint loss distribution of AEX, CAC and DAX. These three copulas are in fact the only considered ones for which the upper and lower tail-dependence coefficients are equal. Moreover, this could suggest that the joint loss distribution is radially symmetric.

Table 14: Result of the GOF tests for copulas using transformed uniform innovations of DOW, MERV, SMI and SSE.

GOF statistic	Gaussian	t	Clayton	Frank	Gumbel	Joe
S_T	**0.0382**	**0.0304**	1.1067*	0.6374*	0.8301*	1.5421*
$S_T^{(B)}$	**0.0433**	**0.0335**	0.8822*	0.6102*	0.7758*	1.4241*
$S_T^{(C)}$	**0.0479**	**0.0356**	0.9545*	0.6049*	0.8146*	1.5741*

The GOF statistics for which the p-value is greater than or equal to 0.05 are indicated by a bold.
*Significant at the 1% significance level. **Significant at 5% significance level. ***Significant at 10% significance level.

Due to the low correlation between SSE − SMI, SSE − DOW and SSE − MERV we would expect a relatively low upper and lower tail-dependence overall. Though, any other pairwise combination of these four stock indices is also considered by the copulas. From the GOF test result in Table 14 we have evidence that only the Gaussian and t copula are appropriate for the fitting of the joint loss distribution of these stock indices. Notice that for only SSE and DOW we found in Table 12 that each of the six copulas is appropriate for fitting the joint loss distribution. However, for the 4-dimensional dataset, which includes SSE and DOW, only two

[28] The considered copulas, except the Gaussian copula, can capture positive tail-dependence. However, these copulas can also capture zero tail-dependence.

copulas are appropriate. The reason that not each of the copulas is appropriate is, again, because any other pairwise combination of these four stock indices is also taken into account for the copula fit.

Finally, we would like to make the remark that for some datasets it may happen that all p-values are less than 0.05, i.e. the GOF tests are statistically significant at the 5% level. In that case we do not know from the GOF tests which copula is appropriate for fitting the joint distribution of the dataset. However this is not an issue if we would use the copula mainly for forecasting, as the GOF tests are only to test for the fitting of the copula. Else, we may choose the copula that is capable to describe a certain effect of interest in the dataset, e.g. (only) upper extremal dependence.

6 Conclusion

The aim of this thesis is to describe the dependence structure of twenty stock indices worldwide. This is done with the help of comprehensive time series models. Empirically, MGARCH-type models and copula functions have respectively shown to be suitable for modelling the dynamics and the dependence structure of multivariate time series data. By combining a MGARCH-type model with a copula function, we allow the margins of the univariate time series to model the structure of the joint distribution, while allowing the copula to model the dependence structure.

In this study we evaluated the performance of four different copula-MGARCH models. For comparison, we also considered four traditional MGARCH models. Firstly, from the in-sample fit test we found that the t-GJR model was optimal, but closely followed by the t-GARCH model. This implied that the in-sample fit of the models with t copula modelling are superior to the other models. Secondly, from the VaR backtest we found all copula-MGARCH models to perform optimal. This is advantageous for financial risk management, where it is of importance to estimate the conditional VaR as accurate as possible. Lastly, based on the estimated statistical loss functions there was not a particular model outperforming the other models. Interestingly, the ranking of the models could be extremely different for different statistical loss functions. The ranking from this test depends on the choice of statistical loss functions and the choice for the volatility proxy. In fact the best fitting model, i.e. the t-GJR model, did not perform statistically worse than any other model.

To answer our research question in particular, from the results of these three tests we found that the copulas improve the fit and VaR estimation of the traditional MGARCH models. Also, the models that are based on the non-normality assumption generally perform better than their counterparts that are based on

the normality assumption. By combining the findings of the tests we concluded that the t-GJR model is the most appropriate model. Therefore we used this model to estimate the rank correlation and the tail-dependencies of the financial time series data.

The Kendall's taus show that the correlations between the stock indices are not as extreme as implied by the Pearson's correlation coefficients. Based on the time series of the estimated dynamic conditional Kendall's taus, we could infer that a financial crisis does not necessarily affects the degree of correlation of a pair of stock indices. Though large fluctuations in the correlation during financial crises occur more often for highly correlated pairs of stock indices than for poorly correlated pairs of stock indices. In fact, the change in correlation differs per event and can be relatively limited. Furthermore, the change in correlation of a pair of stock indices with a certain degree of correlation can be significantly different for another pair of stock indices with (almost) the same degree of correlation.

As the estimated correlation coefficients generally do not provide accurate measurements for the tail-dependency, we consider tail-dependence coefficients. According to the estimated tail-dependence coefficients, many stock indices have little to no tendency to crash or boost together, sometimes even if these stock indices have relatively high correlation. Thus we infer that correlation coefficients do actually not provide accurate measures for the tail-dependence.

Goodness-of-fit tests for copulas primary provide information about the underlying joint distribution of multivariate data. We proposed an extension of our main study that incorporates goodness-of-fit tests for copulas before DCC modelling. The extension provides the option to choose for an appropriate copula function before applying it in the full copula-MGARCH model. We illustrated the extension for the lowest correlated pair of stock indices (i.e. SSE and DOW), the three highest positively correlated stock indices (i.e. AEX, CAC and DAX) and the four stock indices corresponding to the three lowest correlations (i.e. DOW, MERV, SMI and SSE). Among the Gaussian, t and four Archimedean copulas, we found that each of the six copulas is appropriate for fitting the joint loss distribution of SSE and DOW. For the AEX, CAC and DAX this was only the case for the t copula. Lastly, the Gaussian and t copula are appropriate for fitting the joint loss distribution of DOW, MERV, SMI and SSE.

7 Discussion

Multivariate analysis with a large number of time series, where all these time series are used simultaneously, results in a relatively low number of data. As we have seen in this thesis, this is because we can only consider the data at time periods for which every time series has an observation. Furthermore, a down-

side of using a high-dimensional copula model is that its simulated values may not be very reliable due to the large number of parameters in the model. Also, the multivariate t copula uses a unique degrees of freedom parameter for the high-dimensional data. A possible solution to these limitations is to perform separate bivariate modellings. This does not only increase the number of time periods, but it also increases the reliability of the simulations from the fitted copula. An alternative solution is to apply vine copulas, which are closely related to performing separate bivariate copula modelling. Namely, a vine copula is a joint distribution that is constructed from bivariate and conditional bivariate copulas. The vine copula constructed from bivariate t copulas should allow for greater modelling flexibility than the standard multivariate t copula.

Longin & Solnik (2001) suggest that financial data might well show significant non-zero tail-dependence. This tail-dependence may be higher or lower in the upper tail than in the lower tail of the loss distribution. So it is preferable to allow for the existence of asymmetric tail-dependence when specifying a copula model. Unfortunately, neither the Gaussian copula nor the t copula have this feature. Vine copulas, on the other hand, do allow for the flexibility of asymmetric tail-dependence. Further, to our best knowledge there is no definition of conditional tail-dependence coefficients, i.e. tail-dependence coefficient with time-dependence. Such coefficients could explicitly take into account the time periods of large or low fluctuations in the time series. Accordingly, these coefficients would be more reliable than the standard tail-dependence coefficients.

In the in-sample fit and out-of-sample forecast tests we performed the processes once in every five observations due to the high computational cost of re-estimating a model with many parameters. If the computational cost is not a big issue, we suggest to perform the processes once in every observation so that as much as possible data is used.

Future studies are highly encouraged to extend our approach by considering a larger set of candidate distributions for the margins (such as the Generalized Hyperbolic Distribution and the Generalized Hyperbolic Skew Student Distribution) and different copulas (such as Archimedean, Extreme Value and vine copulas). Additionally, the approach can be combined with Extreme Value Theory in which the tails of the margins are fitted to the Generalized Pareto Distribution. Also recall that we restricted our study to MGARCH orders of $(1, 1)$. Future studies could extend this by allowing for different orders, similar as we have done for the ARMA specification of the margins. So the orders could for example be chosen based on the AIC or any other information criterion. Finally, it may be interesting to do the analysis (also) on time series of assets, such as the price of gold and oil.

Appendix

A Formal tests

Jarque-Bera (J-B) test (Bera & Jarque 1981). This is a test for normality of a sample distribution. The null hypothesis for a given sample $X = (X_1, \ldots, X_T)'$ with mean $\mu = \mathrm{E}[X]$ and variance $\sigma^2 = \mathrm{var}(X)$ is given by

$$H_0 : X \sim \mathcal{N}(\mu, \sigma^2) \quad \text{vs.} \quad H_A : X \nsim \mathcal{N}(\mu, \sigma^2).$$

The Jarque-Bera test simultaneously checks whether or not the *skewness* and *kurtosis* of the sample distribution is consistent with that of a normal distribution. The theoretical kurtosis and skewness of a sample X are respectively defined as

$$\begin{aligned} \beta &= \mathrm{E}[(X - \mu)^3/\sigma^3] \quad \text{and} \\ \kappa &= E[(X - \mu)^4/\sigma^4]. \end{aligned}$$

The Jarque-Bera test statistic is given by

$$\lambda_{JB} = \frac{1}{6} T \left(\beta^2 + \frac{1}{4}(\kappa - 3)^2 \right) \overset{a}{\sim} \chi^2(2),$$

i.e. the test statistic is asymptotically $\chi^2(2)$-distributed under H_0.

The skewness and kurtosis of a normal distribution are equal to 0 and 3, respectively. So if the skewness and/or the kurtosis differ greatly from 0 and 3, respectively, then it is likely that the null hypothesis of normality of the sample distribution will be rejected. In that case we have evidence that the sample distribution follows a fat- or thin-tailed and/or skewed distribution. The rejection of the null hypothesis is based on the p-value: reject the null hypothesis if the p-value is less than the chosen significance level. Otherwise, do not reject the null hypothesis.

Ljung-Box Q(h)-test (Ljung & Box 1978). This is a *portmanteau test* for autocorrelation of a time series model. Consider a time series model $\{X_t\}_{t \in \mathbb{Z}}$ with the first h autocorrelations denoted by $\rho(1), \ldots, \rho(h)$. The null hypothesis is given by

$$H_0 : \rho(1) = \ldots = \rho(h) = 0 \quad \text{vs.} \quad H_A : \rho(j) \neq 0 \text{ for some } j \in \{1, \ldots, h\},$$

i.e. the Ljung-Box Q(h) test simultaneously checks for autocorrelation at multiple lags. The Ljung-Box test statistic is given by

$$Q_{LB}(h) = T(T+2) \sum_{j=1}^{h} \frac{\hat{\rho}(j)^2}{T - j} \overset{a}{\sim} \chi^2(h),$$

i.e. the test statistic is asymptotically $\chi^2(h)$-distributed under H_0. Commonly the value for h is chosen to be 12.

At a certain significance level, if we do not reject the null hypothesis, then we have evidence that the time series is independently distributed. Otherwise we have evidence that it is not independently distributed.

Augmented Dickey-Fuller (ADF) test (Dickey & Fuller 1981). This is a unit root test of a time series model $\{X_t\}_{t\in\mathbb{Z}}$. Consider the so-called *augmented Dickey-Fuller regression*:

$$X_t = \alpha + \beta t + \rho X_{t-1} + \phi_1 \Delta X_{t-1} + \phi_2 \Delta X_{t-2} + \ldots + \phi_{p-1} \Delta X_{t-p+1} + \varepsilon_t, \qquad (66)$$

where α is a constant, β is a time trend coefficient, p is the lag order and Δ is the difference operator, i.e. $\Delta X_t = X_t - X_{t-1}$. The null hypothesis is given by

$$H_0 : \rho = 1 \quad \text{vs.} \quad \rho < 1.$$

The ADF test statistic is a t statistic:

$$\tau_{ADF} = \frac{\hat{\rho} - 1}{\text{s.e.}(\hat{\rho})},$$

where $\hat{\rho}$ is estimated by ordinary least squares and s.e.$(\hat{\rho})$ is the standard error of the estimate.

At a certain significance level, if we do not reject the null hypothesis, then we have evidence that the process $\{X_t\}_{t\in\mathbb{Z}}$ has a unit root and so not stationary, but it is *difference stationary*, i.e. $\{\Delta X_t\}_{t\in\mathbb{Z}}$ is stationary. [29] Otherwise we have evidence that $\{X_t\}_{t\in\mathbb{Z}}$ has no unit root and so stationary.

[29] For $\rho = 1$, the regression in (66) can be written as an ARMA$(p, 0)$ process of $\{\Delta X_t\}_{t\in\mathbb{Z}}$:

$$\Delta X_t = \alpha + \beta t + \phi_1 \Delta X_{t-1} + \phi_2 \Delta X_{t-2} + \ldots + \phi_{p-1} \Delta X_{t-p+1} + \varepsilon_t.$$

B Backtests

Unconditional coverage backtest.
This backtest is proposed by Kupeic (1995) and it is a likelihood ratio test to examine whether or not the VaR estimation yields an accurate number of exceptions.

Recall that for a confidence level $\alpha \in (0,1)$ we expect that the loss at time period t does not exceed the estimated VaR at time period $t-1$ with $(1-\alpha)100\%$. In other words, over a time period $T+1,\dots,T+n$ we expect that the number of exceptions, I_α^n, is equal to $n(1-\alpha)$. The null hypothesis is stated as follows.

$$H_0 : \mathrm{E}[I_\alpha^n] = n(1-\alpha) \quad \text{vs.} \quad H_A : \mathrm{E}[I_\alpha^n] \neq n(1-\alpha).$$

The corresponding test statistic is given by

$$\mathrm{LR}_{uc} = 2\ln[(1-m/n)^{n-m}(m/n)^m] - 2\ln[\alpha^{n-m}(1-\alpha)^m] \sim \chi^2(1),$$

where $1-\alpha$ is the theoretical probability of exceptions and $m = I_\alpha^n$ is the observed number of exceptions over n trading days.

Independence backtest
This backtest is proposed by Christoffersen (1998) and it is a likelihood ratio test to examine whether or not the occurrence of exceptions are evenly spread throughout the backtesting period.

For financial institutions it is important that their models estimate the VaRs so that the exceptions occur independently (or non-clustered), as motivated by Hull (2015). For example, if the $n(1-\alpha)$ number of exceptions over a time period of $T+1,\dots,T+n$ occur all in a row, then this has the consequence that the additional losses will be accumulated and the financial institution is exposed to a high risk of default. Clustering of the exceptions could be a result of poorly measured heteroskedasticity. Let n_{ij} denote the number of pairs (i,j) where $i = \mathbb{1}_{\{l_t > \mathrm{VaR}_\alpha^{t-1}\}}$ and $j = \mathbb{1}_{\{l_{t+1} > \mathrm{VaR}_\alpha^{t}\}}$, $t = T+1,\dots,T+n$, and

$$\pi_{01} = \frac{n_{01}}{n_{00}+n_{01}},$$
$$\pi_{11} = \frac{n_{11}}{n_{10}+n_{11}}.$$

Observe that π_{11} (respectively π_{01}) is the probability that an exception occurs at time $t+1$ given that (respectively no) exception occurred at time t. The null hypothesis is stated as follows.

$$H_0 : \pi_{01} = \pi_{11} \quad \text{vs.} \quad H_A : \pi_{01} \neq \pi_{11}.$$

In other words, the null hypothesis states that the probability of an exception at time $t+1$ given that there was no exception at time t is equal to the probability of an exception at time $t+1$ given that there was an exception at time t.

The corresponding test statistic is given by

$$\mathrm{LR}_{ind} = 2\ln\left[(1-\pi_{01})^{n_{00}}\pi_{01}^{n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{n_{11}}\right] - 2\ln\left[(1-\pi)^{n_{00}+n_{10}}\pi^{n_{01}+n_{11}}\right] \sim \chi^2(1),$$

where $\pi = (n_{01} + n_{11})/(n_{00} + n_{01} + n_{10} + n_{11})$.

Conditional coverage backtest.
This backtest is proposed by Christoffersen (1998) and it is a combination of the unconditional coverage and independence backtest. So it simultaneously tests the null hypothesis of accurate number of exceptions and independent occurrence of exceptions. The corresponding likelihood ratio test statistic is given by

$$\mathrm{LR}_{cc} = \mathrm{LR}_{uc} + \mathrm{LR}_{ind} \sim \chi^2(2).$$

C Multivariate distribution functions

Multivariate normal distribution.
A random vector $\boldsymbol{U} = (U_1, \ldots, U_n)'$ following an n-dimensional normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$ and variance-covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is denoted by $\boldsymbol{U} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The corresponding multivariate density function is given by

$$f(\boldsymbol{u}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{u} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{u} - \boldsymbol{\mu}) \right\}, \tag{67}$$

where $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ is the vector of realizations of $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$ is a symmetric positive-definite matrix (Hofert 2013). For $n = 1$, (67) becomes the density function of the univariate normal distribution.

The multivariate density function can also be written as

$$f(\boldsymbol{u}) = (2\pi)^{-n/2} |\boldsymbol{P}|^{-1/2} \left(\prod_{i=1}^{n} \sigma_i \right)^{-1} \exp\left\{ -\frac{1}{2} \boldsymbol{x}' \boldsymbol{P}^{-1} \boldsymbol{x} \right\},$$

where σ_i is the standard deviation of U_i, $\boldsymbol{x} = (x_1, \ldots, x_n)' \in \mathbb{R}^n$ with elements $x_i = \frac{u_i - \mu_i}{\sigma_i}$, $i = 1, \ldots, n$, and $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ is the correlation matrix. This representation is particularly useful for the construction of the Gaussian copula density function (see (34)).

Multivariate t distribution.
A random vector $\boldsymbol{U} = (U_1, \ldots, U_n)'$ following an n-dimensional t distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)' \in \mathbb{R}^n$, variance-covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ and ν df is denoted by $\boldsymbol{U} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The corresponding multivariate density function is given by

$$f(\boldsymbol{u}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\pi\nu)^{n/2} \Gamma\left(\frac{\nu}{2}\right)} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{1}{\nu} (\boldsymbol{u} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{u} - \boldsymbol{\mu}) \right)^{-\frac{\nu+n}{2}}, \tag{68}$$

where $\boldsymbol{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$ is the vector of realizations of $\boldsymbol{U}$ and $\boldsymbol{\Sigma}$ is a symmetric positive-definite matrix (Hofert 2013). For $n = 1$, (68) becomes the density function of the univariate t distribution.

The multivariate density function can also be written as

$$f(\boldsymbol{u}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{(\pi\nu)^{n/2} \Gamma\left(\frac{\nu}{2}\right)} |\boldsymbol{P}|^{-1/2} \left(\prod_{i=1}^{n} \sigma_i \right)^{-1} \left(1 + \frac{1}{\nu} \boldsymbol{x}' \boldsymbol{P}^{-1} \boldsymbol{x} \right)^{-\frac{\nu+n}{2}},$$

where σ_i is the standard deviation of U_i, $\boldsymbol{x} = (x_1, \ldots, x_n)' \in \mathbb{R}^n$ with elements $x_i = \frac{u_i - \mu_i}{\sigma_i}$, $i = 1, \ldots, n$, and $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ is the correlation matrix. This representation is particularly useful for the construction of the t copula density function (see (37)).

D Archimedean copulas

The theory in this appendix is adopted from Hofert et al. (2012), if not mentioned differently.

Archimedean copulas are a class of copulas. Archimedean copulas allow modelling dependence in arbitrarily high dimensions with just one dependency parameter. Furthermore, the Kendall's tau rank correlation coefficient and upper and lower tail-dependence coefficients can be computed from explicit formulas for each Archimedean copula.

An n-dimensional copula is called *Archimedean* if it is of the form

$$C(u_1, \ldots, u_n|\theta) = \psi(\psi^{-1}(u_1|\theta) + \cdots + \psi^{-1}(u_n|\theta)|\theta), \tag{69}$$

where θ is the Archimedean copula parameter defined on some parameter space Θ and $\psi : [0, \infty] \times \Theta \to [0, 1]$ is the *Archimedean generator* with inverse $\psi^{-1} : [0, 1] \times \Theta \to [0, \infty]$. Moreover, ψ is a continuous, decreasing function satisfying:

$$\begin{aligned} \psi(0|\theta) &= 1 \\ \psi(\infty|\theta) &= \lim_{t\to\infty} \psi(t|\theta) = 0 \\ \psi^{-1}(0|\theta) &= \inf\{t : \psi(t|\theta) = 0\} \end{aligned}$$

and which is strictly decreasing on $[0, \inf\{t : \psi(t|\theta) = 0\}]$.

An Archimedean generator uniquely defines the structure of an Archimedean copula. Among the most-applied Archimedean copulas are the Clayton, Frank, Gumbel and Joe copula. Conveniently, many Archimedean copulas allow for greater flexibility in measuring the tail-dependence than elliptical copulas like the Gaussian and t copula. More explicitly, many Archimedean copulas can capture different upper and lower tail-dependence coefficient (i.e. $\lambda_u \neq \lambda_l$), contrary to elliptical copulas where the upper and tail-dependence coefficient are equal.

The parameter space Θ, Archimedean generator $\psi(t|\theta)$ and the explicit formulas for the upper and lower tail-dependence coefficient λ_u and λ_l of these copulas are given in the following table.

Table 15: Archimedean copulas

Copula	Θ	$\psi(t\|\theta)$	λ_u	λ_l
Clayton	$(0, \infty)$	$(1+t)^{\frac{-1}{\theta}}$	0	$2^{-1/\theta}$
Frank	$(0, \infty)$	$-\frac{1}{\theta}\ln(1-(1-e^{-\theta})e^{-t})$	0	0
Gumbel	$[1, \infty)$	$e^{-t^{\frac{1}{\theta}}}$	$2-2^{1/\theta}$	0
Joe	$[1, \infty)$	$1-(1-e^{-t})^{\frac{1}{\theta}}$	$2-2^{1/\theta}$	0

For a given Archimedean generator in Table 15 we can straightforwardly obtain the mathematical expression of the corresponding Archimedean copula using (69). These expressions are given below for convenience.

The Clayton copula can be written as:

$$C(u_1, \ldots, u_n|\theta) = \left(\sum_{i=1}^{n} u_i^{-\theta} - n + 1 \right)^{-1/\theta},$$

where $\theta \in (0, \infty)$.

The Frank copula can be written as:

$$C(u_1, \ldots, u_n|\theta) = \ln \left(1 + \frac{\prod_{i=1}^{p} (\theta^{u_i} - 1)}{(\theta - 1)^{n-1}} \right),$$

where $\theta \in (0, \infty)$.

The Gumbel copula can be written as:

$$C(u_1, \ldots, u_n|\theta) = \exp \left(- \left(\sum_{i=1}^{n} (-\ln(u_i))^{\theta} \right)^{\frac{1}{\theta}} \right),$$

where $\theta \in [1, \infty)$.

The Joe copula can be written as:

$$C(u_1, \ldots, u_n|\theta) = 1 - \left(1 - \prod_{i=1}^{n} (1 - (1 - u_i)^{\theta}) \right)^{\frac{1}{\theta}},$$

where $\theta \in [1, \infty)$.

E Plots from the visual tests

Histograms.

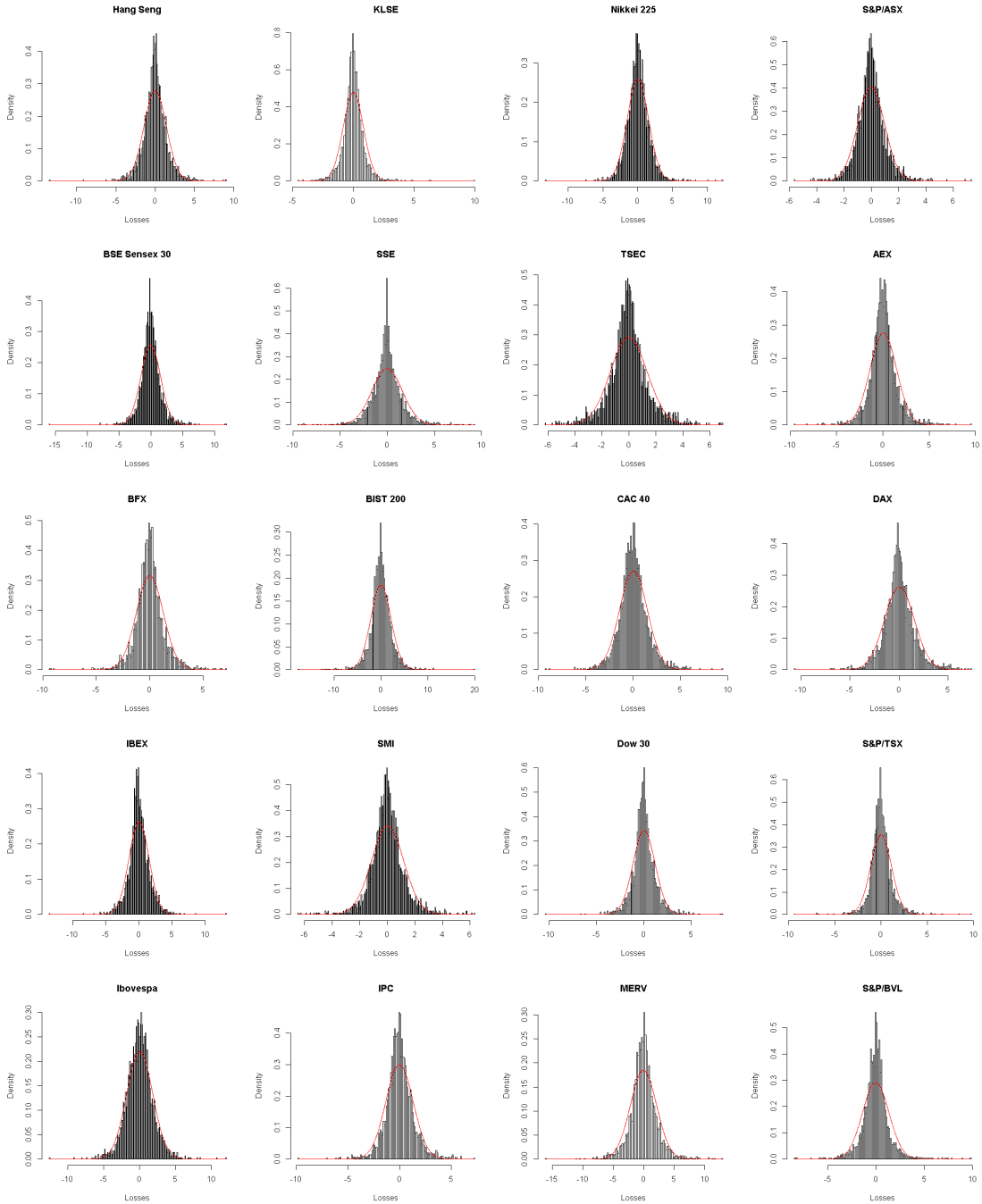
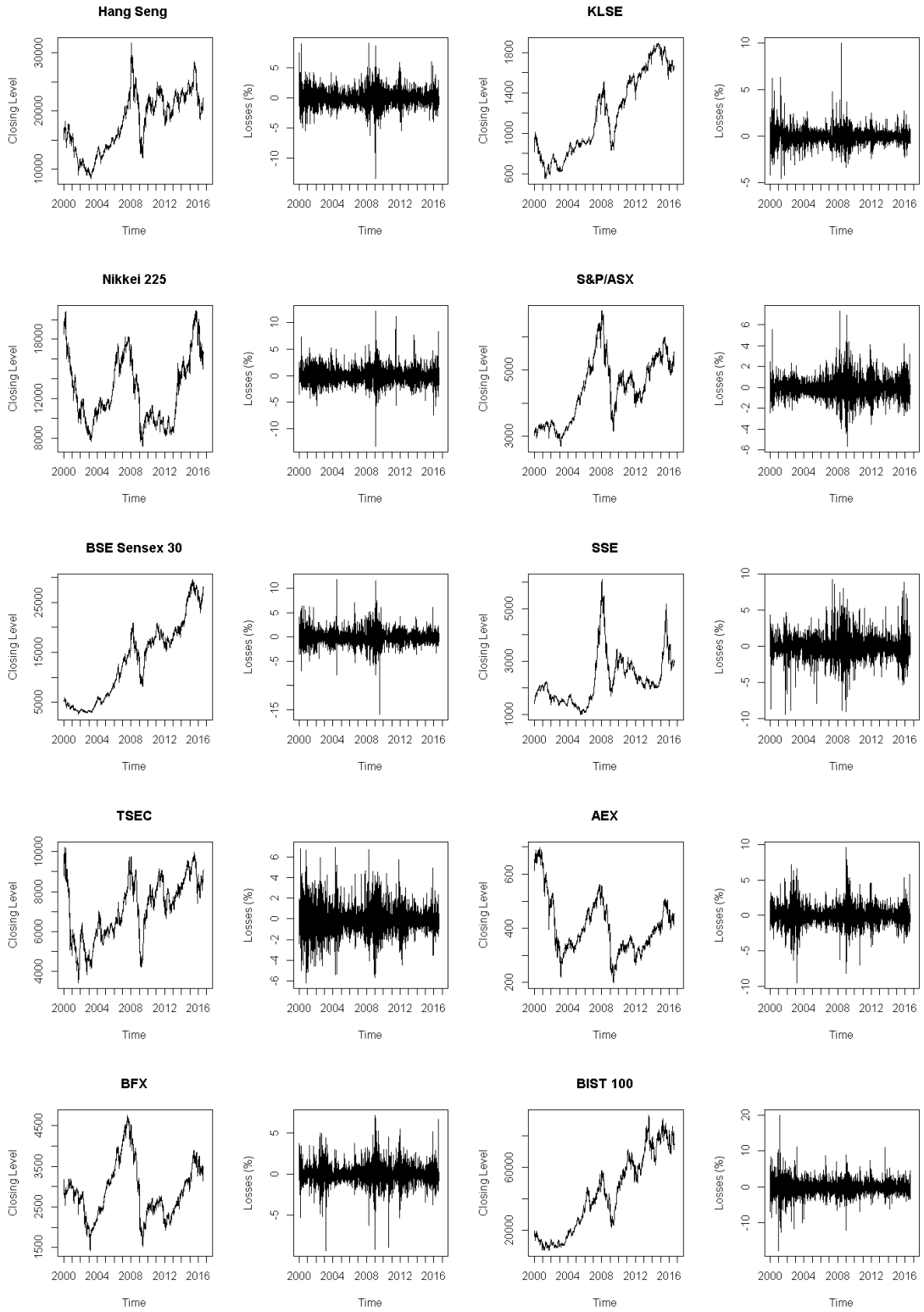


Figure 16: Histogram of the losses of every stock index from January 1, 2000 to July 31, 2016. The red line refers to an estimate of the normal density.

Plots of the daily closing price (left) and daily losses (right).



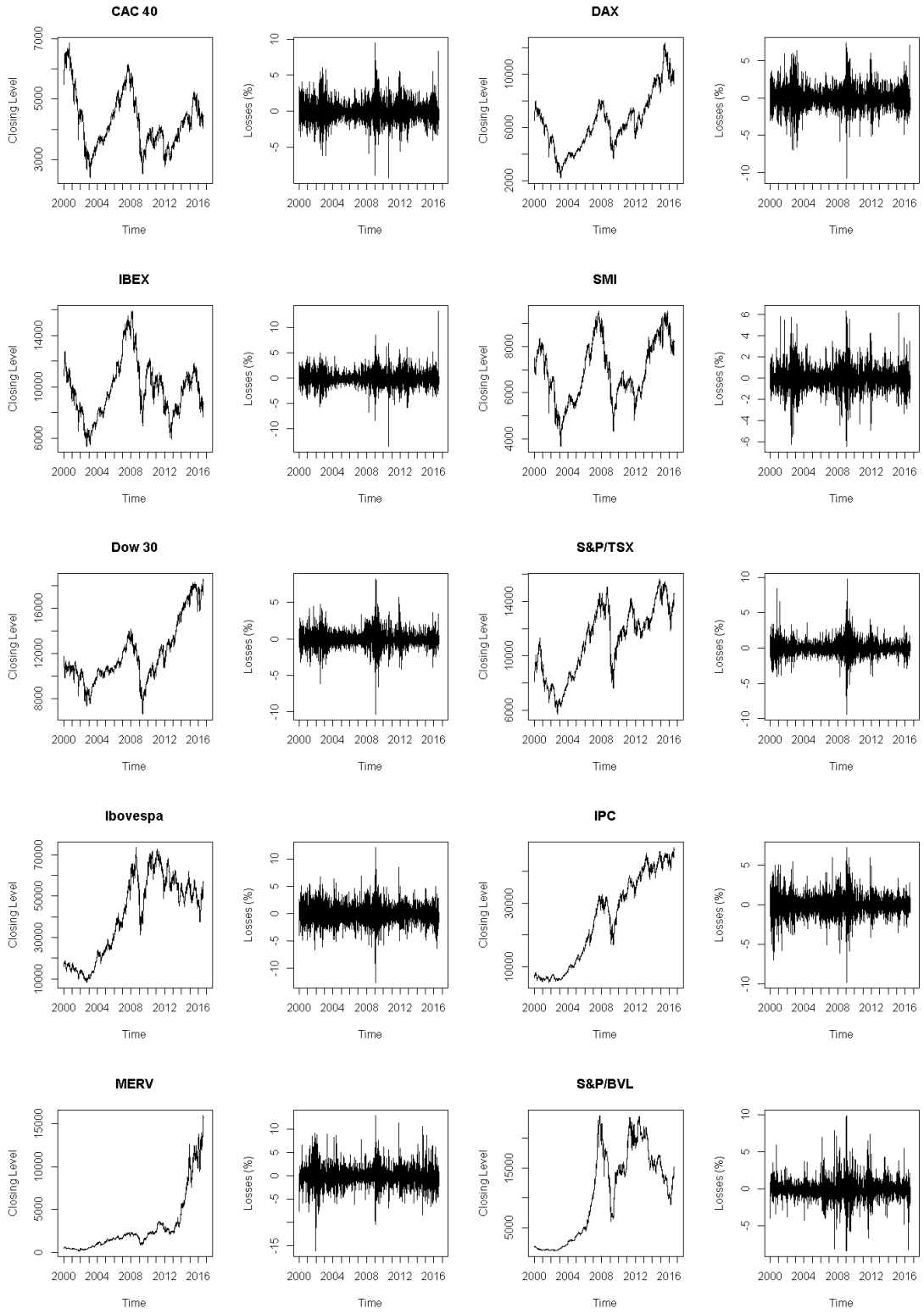


Figure 17: Daily closing price (left) and daily losses (right) of every stock index from January 1, 2000 to July 31, 2016.

Q-Q plots.

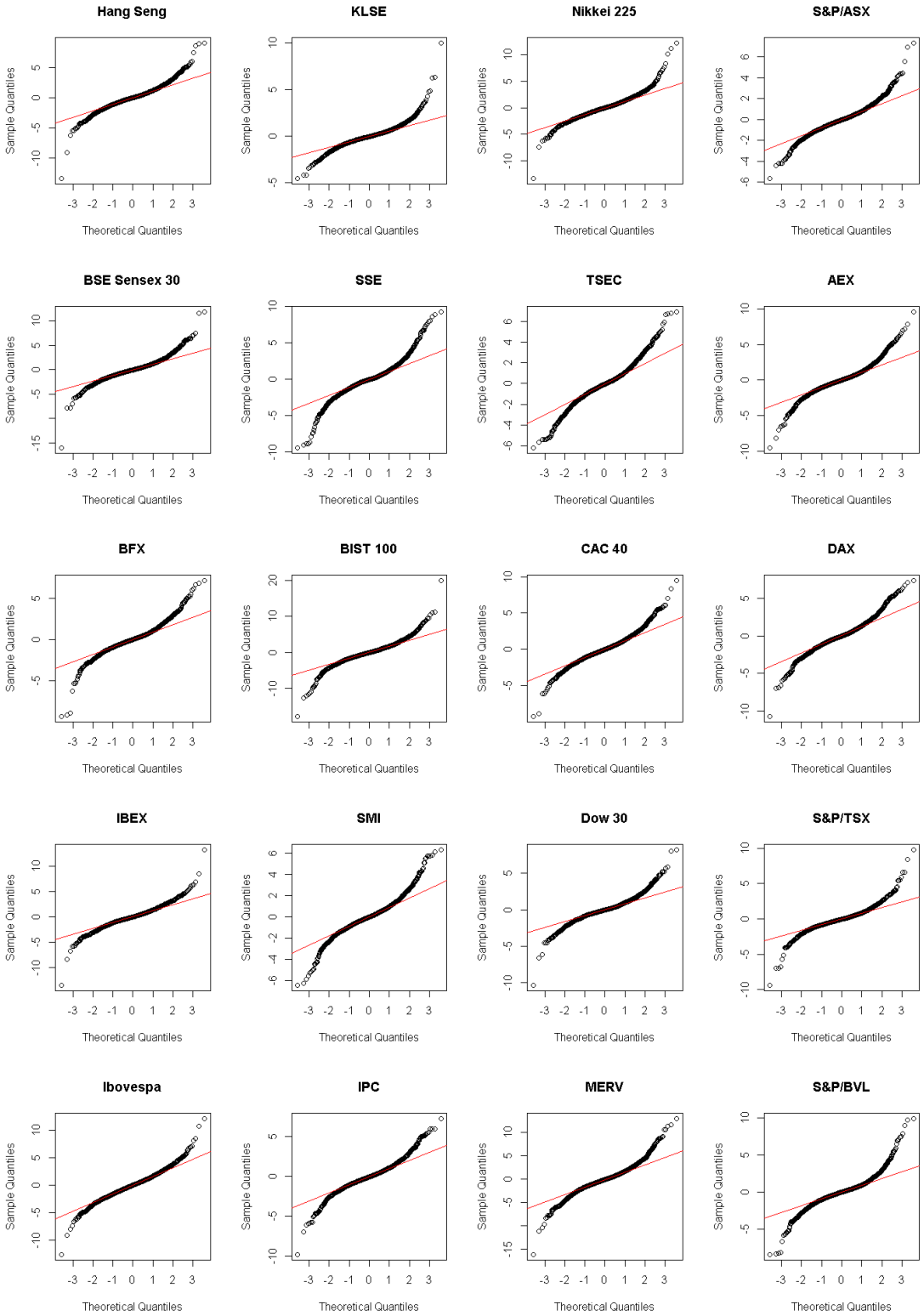
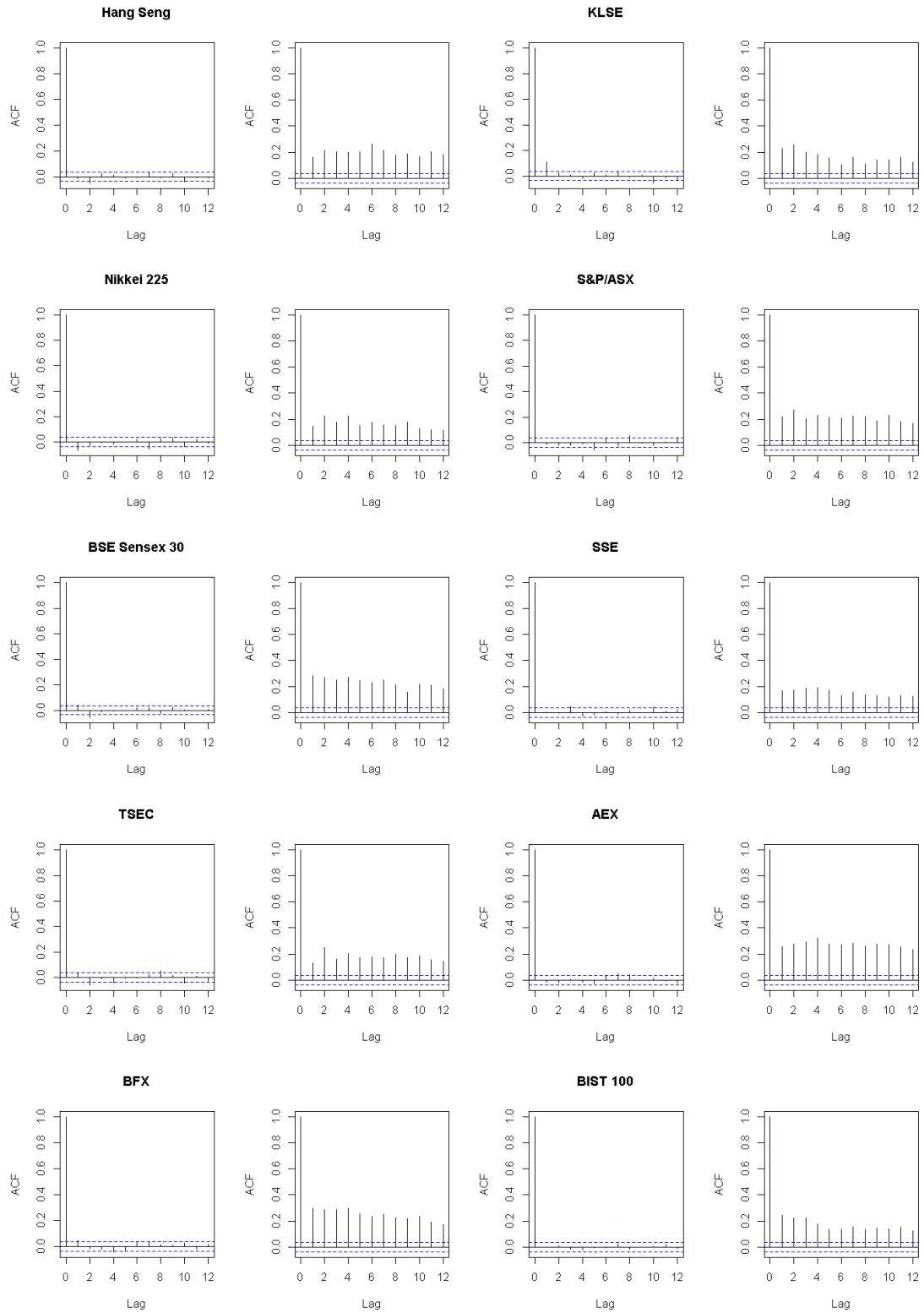


Figure 18: Q-Q plot of the losses of every stock index from January 1, 2000 to July 31, 2016. The red line refers to the normal distribution.

Correlograms of the raw losses (left) and absolute losses (right).



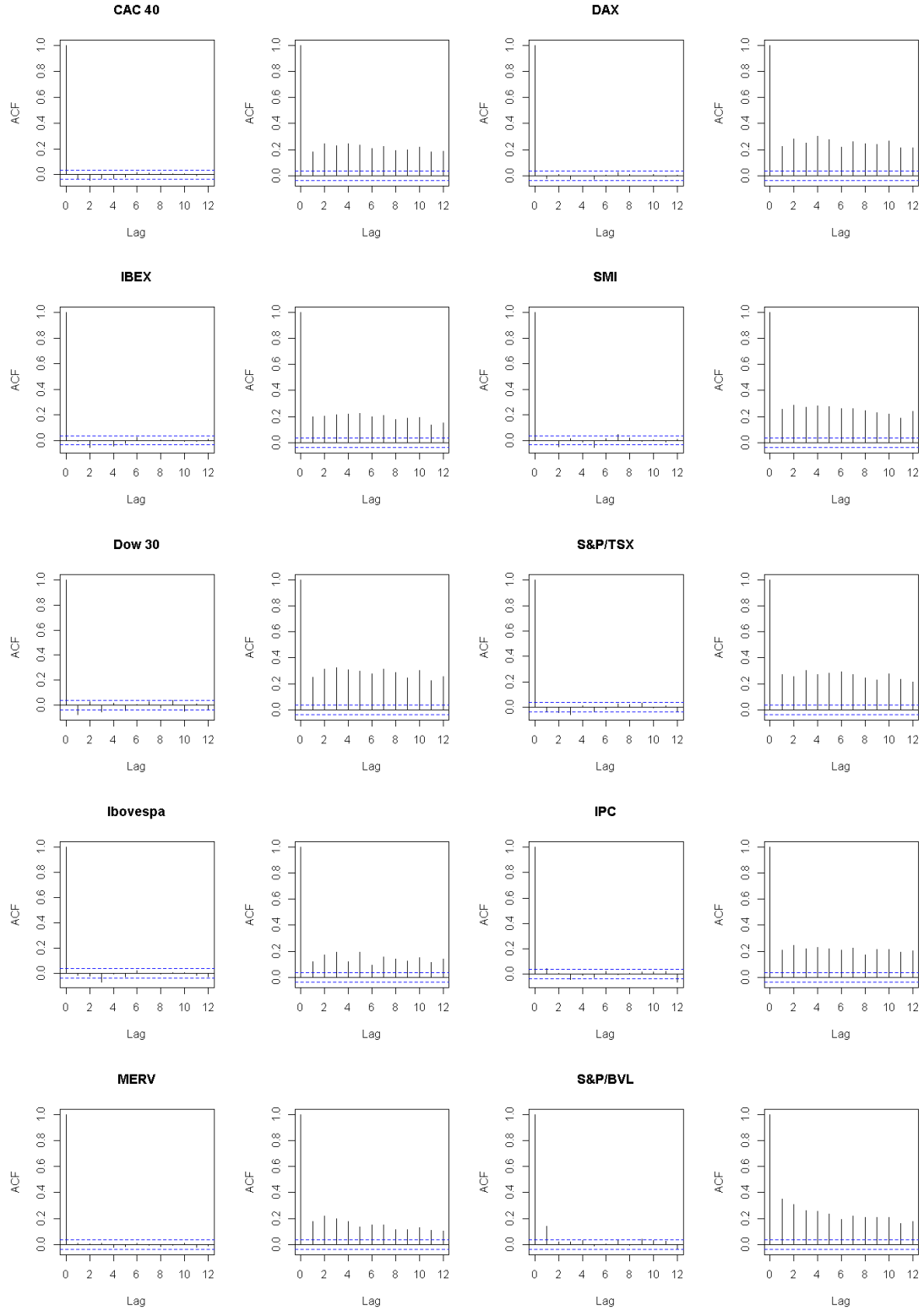
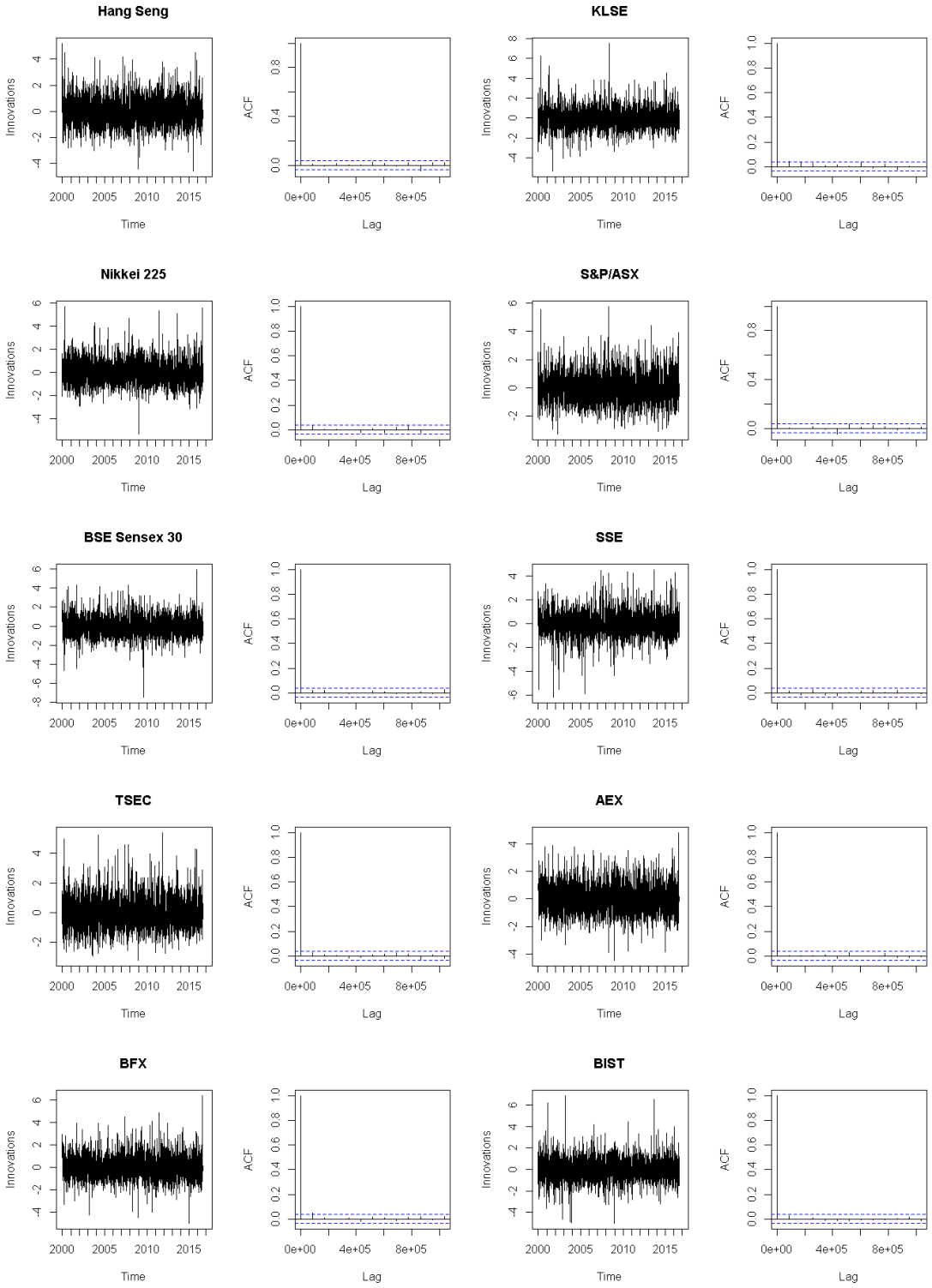


Figure 19: Correlograms of the raw losses (left) and the absolute losses (right) of every stock index from January 1, 2000 to July 31, 2016.

F Plots from the empirical analysis

Innovations plot (left) and correlogram (right) of the ARMA-GARCH model.



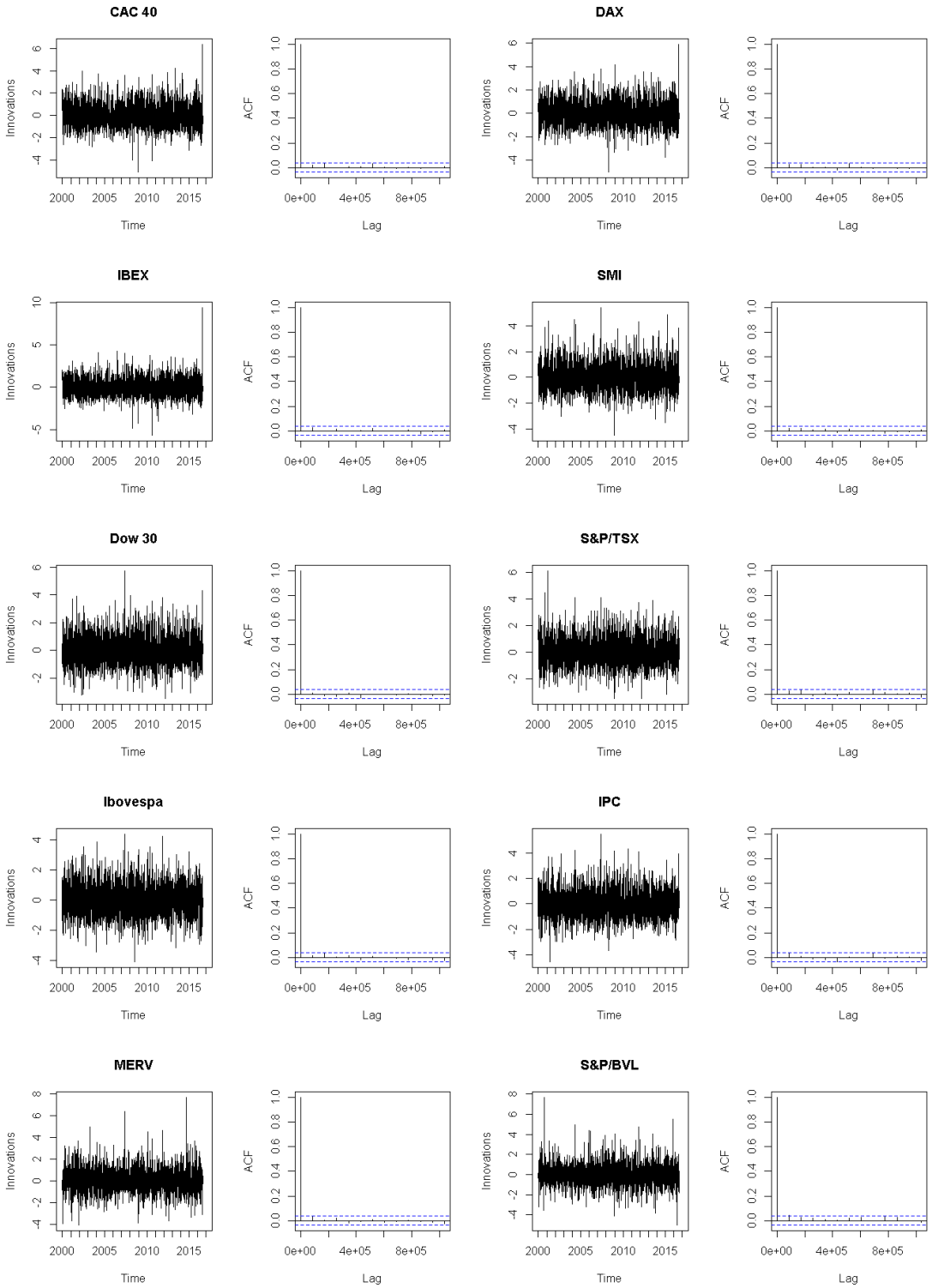
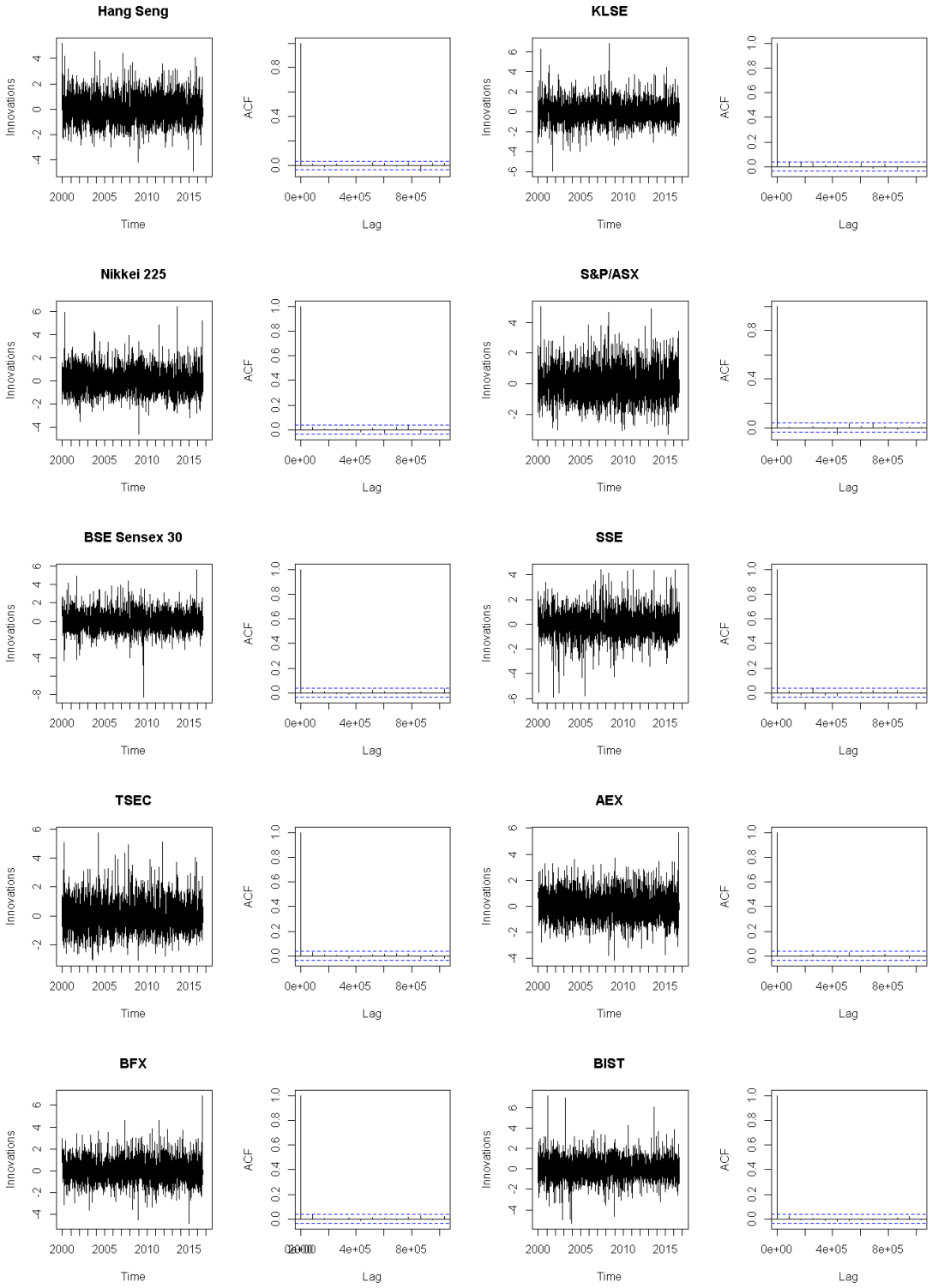


Figure 20: Innovations plot (left) and correlogram (right) of the ARMA-GARCH model for every stock index, from January 1, 2000 to July 31, 2016.

Innovations plot (left) and correlogram (right) of the ARMA-GJR model.



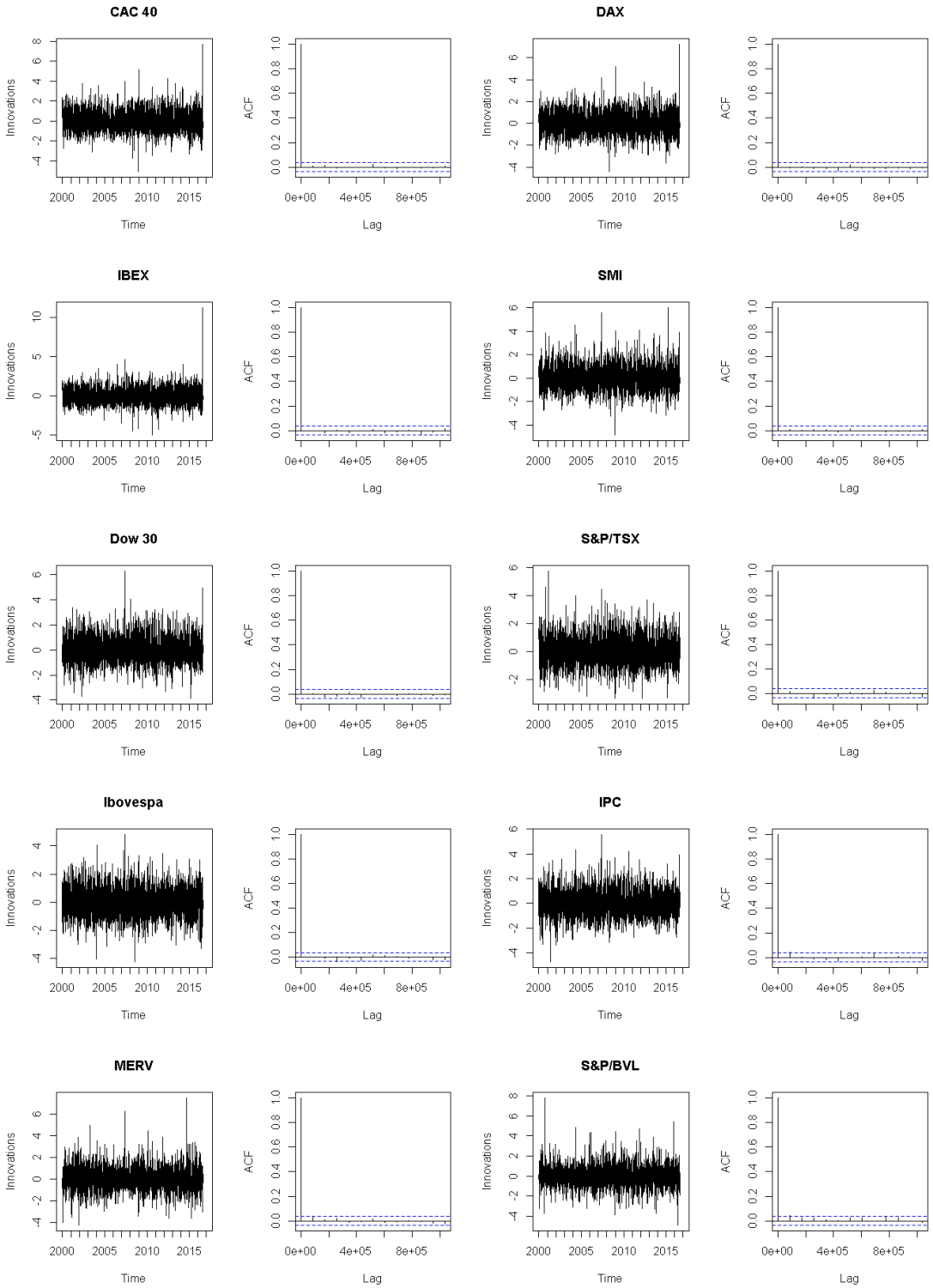


Figure 21: Innovations plot (left) and correlogram (right) of the ARMA-GJR model for every stock index, from January 1, 2000 to July 31, 2016.

References

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Aloui, R., Aïssa, M. S. B. & Nguyen, D. K. (2011), 'Global financial crisis, extreme interdependence, and contagion effects: The role of economic structure?', *Journal of Banking & Finance* **35**, 130–141.

Aloui, R., Aïssa, M. S. B. & Nguyen, D. K. (2013), 'Conditional dependence structure between oil prices and exchange rates: A copula-GARCH approach', *Journal of International Money and Finance* **32**, 719–738.

Bauwens, L., Laurent, S. & Rombouts, J. V. K. (2006), 'Multivariate GARCH models: A Survey', *Journal of Applied Econometrics* **21**, 79–109.

Bera, A. & Jarque, C. (1981), 'Efficient tests for normality, heteroskedasticity and serial independence of regression residuals: Monte Carlo evidence', *Economics Letter* **7**, 313–318.

Billio, M. & Caporin, M. (2006), A generalized dynamic conditional correlation model for portfolio risk evaluation, Technical report, Ca' Foscari University of Venice, Department of Economics. unpublished manuscript.

Bollerslev, T. (1986), 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bouye, E., Durrleman, V., Nikeghbali, A., Riboulet, G. & Roncalli, T. (2000), 'Copulas for Finance: A Reading Guide and Some Applications'. Groupe de Recherche Opérationnelle, Crédit Lyonnais. Working Paper.

Cappiello, L., Engle, R. F. & Sheppard, K. (2006), 'Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns', *Journal of Financial Econometrics* **4**, 537–572.

Christoffersen, P. (1998), 'Evaluating Interval Forecasts', *International Economic Review* **39**, 841–862.

Corke, S., Finkel, A., Kramer, D. J., Robbins, C. A. & Schenkkan, N. (2014), *Democracy in Crisis: Corruption, Media and Power in Turkey*, A Freedom House Special Report. `https://freedomhouse.org/sites/default/files/Turkey%20Report%20-%202-3-14.pdf`.

Demarta, S. & McNeil, A. J. (2004), 'The *t* Copula and Related Copulas', *International Statistical Review* **73**(1), 111–129.

Dickey, D. A. & Fuller, W. A. (1981), 'Distribution of the estimators for autoregressive time series with a unit root', *Econometrica* **49**, 1057–1072.

Embrechts, P. (2009), 'Copulas: A personal view', *Journal of Risk & Insurance* **76**, 639–650.

Embrechts, P., McNeil, A. & Straumann, D. (2001), Correlation and Dependence in Risk Management: Properties and Pitfalls, *in* M. Dempster & H. K. Moffatt, eds, 'Risk Management: Value at Risk and Beyond', Cambridge University Press.

Engle, R. (2002), 'Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models', *Journal of Business and Economic Statistics* **20**, 339–350.

Engle, R. F. (1982), 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica* **50**(4), 987–1008.

Genest, C., Huang, W. & Dufour, J.-M. (2013), 'A regularized goodness-of-fit test for copulas', *Journal de la Société Francaise de Statistique* **154**(1), 64–77.

Genest, C., Rémillard, B. & Beaudoin, D. (2009), 'Goodness-of-fit tests for copulas: A review and a power study', *Insurance: Mathematics and Economics* **44**, 199–213.

Ghalanos, A. (2015*a*), *rmgarch: Multivariate GARCH models.* R package version 1.3-0.

Ghalanos, A. (2015*b*), *rugarch: Univariate GARCH models.* R package version 1.3-6.

Ghorbel, A. & Trabelsi, A. (2009), 'Measure of financial risk using conditional extreme value copulas with EVT margins', *The Journal of Risk* **11**(4), 51–85.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), 'On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks', *The Journal of Finance* **48**(5), 1779–1801.

Hofert, M. (2013), 'On Sampling from the Multivariate *t* Distribution', *The R Journal* **5**(2), 129–136.

Hofert, M., Mächler, M. & McNeil, A. J. (2012), 'Likelihood inference for Archimedean copulas in high dimensions under known margins', *Journal of Multivariate Analysis* **110**, 113–150.

Huang, J.-J., Lee, K.-J., Liang, H. & Lin, W.-F. (2009), 'Estimating value at risk of portfolio by conditional copula-GARCH method', *Insurance: Mathematics and Economics* **45**, 315–324.

Hull, J. (2015), *Risk Management and Financial Institutions*, Vol. 4th Revised edition, John Wiley & Sons, Inc., Hoboken, New Jersey.

Hult, H., Lindskog, F., Hammerlid, O. & Rehn, C. J. (2012), *Risk and Portfolio Analysis*, 1st ed., Springer-Verlag New York.

Hürlimann, W. (2004), 'Fitting bivariate cumulative returns with copulas', *Computational Statistics and Data Analysis* **45**, 355–372.

Isogai, T. (2015), 'An empirical study of the dynamic correlation of Japanese stock returns', *Bank of Japan Working Paper Series* **7**(15).

Jondeau, E. & Rockinger, M. (2006), 'The Copula-GARCH model of conditional dependencies: An international stock market application', *Journal of International Money and Finance* **25**, 827–853.

Kendall, M. G. & Gibbons, J. D. (1990), *Rank Correlation Methods*, 5th ed., New York: Oxford University Press.

Kirisci, K. (2005), 'Turkey and the European Union: The Domestic Politics of

Negotiating Pre-Accession', *Macalester International Journal* **15**.

Kosapattarapim, C., Lin, Y.-X. & McCrae, M. (2011), 'Evaluating the Volatility Forecasting Performance of Best Fittin GARCH Models in Emerging Asian Stock Markets', *Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper* **13-11**.

Kupeic, P. (1995), 'Techniques for Verifying the Accuracy of Risk Management Models', *Journal of Derivatives* **3**, 73–84.

Labuszewski, J. W. (2014), Asian stock index correlations, Technical report, CME Group. unpublished manuscript.

Laurent, S., Rombouts, J. V. K. & Violante, F. (2013), 'On loss functions and ranking forecasting performances of multivariate volatility models', *Journal of Econometrics* **173**, 1–10.

Leon, S. (2008), *Linear Algebra with Applications*, International ed. of 8th revised ed., Pearson Education (Us).

Liu, Y. & Luger, R. (2009), 'Efficient estimationof copula-GARCH models', *Computational Statistics and Data Analysis* **53**, 2284–2297.

Ljung, G. M. & Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**(2), 297–303.

Longin, F. & Solnik, B. (2001), 'Extreme Correlation of International Equity Markets', *Journal of Finance* **56**, 649–676.

Lopez, J. A. (2001), 'Evaluating the Predictive Accuracy of Volatility Models', *Journal of Forecasting* **20**(2), 87–109.

McNeil, A. J., Frey, R. & Embrechts, P. (2015), *Quantitative Risk Management: concepts, techniques and tools*, Princeton Series in Finance, Princeton University Press, Princeton (N.J.).

Naifar, N. (2016), 'Modeling dependence structure between stock market and sukuk yields: A nonlinear study in the case of Saudi Arabia', *Borsa Istanbul Review* **16**, 157–166.

Padhi, P. & Lagesh, M. A. (2012), 'Volatility spillover and time-varying correlation among the Indian, Asian and US stock markets', *Journal of Quantitative Economics* **10**(2), 78–90.

Patton, A. J. (2006), 'Modelling Assymetric Exchange Rate Dependence', *International Economic Review* **47**(2).

Peters, J.-P. (2001), 'Estimating and forecasting volatility of stock indices using assymetric GARCH models and (Skewed) Student-t densities'. University of Liege, Belgium. Preprint.

Rosenblatt, M. (1952), 'Remarks on a multivariate transformation', *The Annals of Mathematical Statistics* **23**, 470–472.

Shamiri, A. & Isa, Z. (2009), 'Modeling and forecasting volatility of Malaysian stock markets', *Journal of Mathematics and Statistics* **3**, 234–240.

Silvennoinen, A. & Terävirta, T. (2008), 'Multivariate GARCH models', *SSE/EFI*

Working Paper Series in Economics and Finance **669**, 1–25.

Sklar, A. (1959), 'Fonctions de répartition á *n* dimensions et leurs marges', *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.

S.R. and C.C.W. (2014), 'China's stockmarket: A market unto itself', `http://www.economist.com/blogs/freeexchange/2014/10/chinas-stockmarket`.

Tang, J., Zhou, C., Yuan, X. & Sriboonchitta, S. (2015), 'Estimating Risk of Natural Gas Portfolios by using GARCH-EVT-Copula model', *The Scientific World Journal* .

Teräsvirta, T. (2009), *An Introduction to Univariate GARCH Models*, Handbook of Financial Time Series, Springer-Verlag Berlin Heidelberg.

Valukonis, M. (2013), 'China's stock market trends and their determinant analysis under market indices', *Economics and Management* **18**(4), 651–660.

Weiß, G. N. F. (2013), 'Copula-MGARCH versus dynamic conditional correlation: an empirical study on VaR and ES forecasting accuracy', *Rev Quant Finan Acc* **41**, 179–202.

Würtz, D., Chalabi, Y. & Luksan, L. (2009), 'Parameter Estimation of ARMA Models with GARCH/APARCH Errors: An R and SPlus Software Implementation', *Journal of Statistical Software* . forthcoming.

Yan, J. (2007), 'Enjoy the Joy of Copulas: With a Package copula', *Journal of Statistical Software* **21**(4), 1–21.