



DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics: Data Engineering and Analytics

Backtesting Value-at-Risk of Financial Data Using Vine Copulas

Mohamed Maarouf





DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics: Data Engineering and Analytics

Backtesting Value-at-Risk of Financial Data Using Vine Copulas

Backtesting des Value-at-Risk von Finanzdaten mit Vine Copulas

Author:	Mohamed Maarouf
Supervisor:	Claudia Czado
Advisor:	Karoline Bax
Submission Date:	August 28 2021

I confirm that this master's thesis in informatics:data engineering and analytics is my own work and I have documented all sources and material used.

Munich, August 28 2021

Contents

1	Introduction	3
2	Background	4
2.1	Asset Returns	4
2.2	Stationarity	5
2.3	Modelling Returns	5
2.4	AR Model	6
2.4.1	Properties of AR Models	6
2.4.2	Parameter Estimation	8
2.4.3	Forecasting	9
2.5	MA Model	9
2.5.1	Properties of MA Models	9
2.5.2	Parameter Estimation	10
2.5.3	Forecasting	10
2.6	ARMA Model	11
2.7	ARCH Model	11
2.7.1	Properties of ARCH Models	13
2.7.2	Parameter Estimation	14
2.7.3	Model Checking	15
2.7.4	Forecasting	15
2.8	GARCH Model	16
2.8.1	Parameter Estimation	16
2.8.2	Model Checking	16
2.8.3	Forecasting	17
2.9	The Ljung-Box Test	17
2.10	Estimation of ARMA-GARCH models	18
2.10.1	A Two-pass estimation method	18
2.10.2	Joint estimation	18
2.11	Copulas	19
2.11.1	Conditioning	20
2.11.2	Dependence Measure	20
2.11.3	Copula Families	21
2.11.4	Estimation	21
2.12	Vine Copulas	22
2.12.1	Regular Vine Copulas	23
2.13	Value-at-Risk (VaR)	25
2.13.1	Estimating the VaR	26
2.14	Backtesting Value-at-Risk	27
2.14.1	Unconditional Coverage Tests	28

2.14.2	Conditional Coverage Tests	29
2.14.3	Pinball Loss	31
3	A Vine Copula based ARMA-GARCH Model	32
3.1	Single day VaR	32
3.2	Backtesting the model	36
4	R-package	43
5	Conclusion	56

1 Introduction

The high volatility of price movements in the financial markets is rare, but very important. The stock market crashes that have taken places over the past few decades brought about considerable attention to market risk management researchers, practitioners as well as regulators. More precisely, this attention has been manifested by the Basel II accord (Jiménez-Martín et al., 2009) which has published recommendations on banking regulations. These recommendations are generally enforced by the member countries which include the G20 countries in addition to others. Basel II explicitly recognizes the Value-at-Risk(VaR) as a market risk measure that must be implemented by banks and financial institutes to determine their risk exposure and the amount of capital that is subject to regulation. Therefore, VaR has been adopted to be the main risk management measures in the field of Risk Management.

In this thesis, we are interested in estimating the VaR of a portfolio of financial assets. More precisely, we are interested in applying copula based models estimating the VaR of large portfolios that are too complicated and computationally expensive to work with using conventional multivariate models (McAleer, 2009). Our approach is based on utilizing univariate ARMA-GARCH models to capture the trend and volatility of each component of the portfolio while using vine copula models to capture the cross dependence structure among the different components.

In the following, Section 2 will give a theoretical background on financial time series modelling, ARMA-GARCH models, vine copula models, VaR estimation, and backtesting VaR. Section 3.1 illustrate our method in utilizing ARMA-GARCH models with vine copula based dependence structure in estimating the VaR for a single day. Section 3.2 extends the single day model to estimate the VaR for an interval of days which can be used to backtest the model. Finally, Section 4 shows an example of using VaRVine, an R-package that we have developed to implement our approach.

2 Background

2.1 Asset Returns

Let P_t be the closing price of an asset on day t , then we can define two types of asset returns.

Simple Returns

By holding an asset from day $t - 1$ to day t , then the simple return over this 1 day period is defined as:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

In our analysis, we consider the continuously compounded returns (or log returns) which are defined next.

Continuously Compounded Returns (Log Returns)

For illustration, assume that a bank pays 10% annual interest on a deposit of \$1000. If the bank pays interest once a year, then the net value of the deposit in one year is $\$1000(1 + 0.1) = \1100 . However, if the bank pays interest twice a year, each being $\frac{10\%}{2} = 5\%$, then the net value after one year is $\$1000(1 + 0.05)^2 = 1102.5$. More generally, if the bank pays interest m times a year, then the net value after one year is $\$1000(1 + \frac{0.1}{m})^m$.

So, if the initial investment is P and the interest rate r is payed m times a year, then the net value of the investment after one year A can be calculated as

$$A = P \left(1 + \frac{r}{m}\right)^m.$$

Following the same reasoning we can derive the continuous compounding net value of an asset with price P over a period t :

$$\begin{aligned} A &= \lim_{m \rightarrow \infty} P \left(1 + \frac{r}{m}\right)^{mt} \\ &= P \lim_{m \rightarrow \infty} \exp\left(mt \ln\left(1 + \frac{r}{m}\right)\right) \\ &\approx P \exp(rt). \end{aligned} \tag{1}$$

Where t can be in days, months, years, etc. In our analysis will consider t to be in days since we are interested in daily returns.

From Equation (1), we can derive the daily log returns:

$$P_t = P_{t-1} \exp(r) \implies$$

$$r = \ln \frac{P_t}{P_{t-1}}$$

We usually use r_t instead of r to denote the daily return on a certain day t .

2.2 Stationarity

Before explaining modelling asset returns, we need to define the concept of *weak stationarity*. We say that a time series $\{r_t\}$ is weakly stationary if:

label=() The unconditional mean $E(r_t) = \mu$ is time invariant.

lbbel=() The covariance between $\text{cov}(r_t, r_{t-\ell}) = \gamma_\ell$ depends only on the lag ℓ .

In practice, assume we have T observations $\{r_t, t = 1, \dots, T\}$, then weak stationarity implies that the time plot of $\{r_t\}$ should oscillate with a constant variation around a fixed point which is the mean as shown in Figure 1.

In the literature, it is often assumed that the return series of an asset is weakly stationary (Tsay, 2010, p. 29), and we will make the same assumption in the following sections.

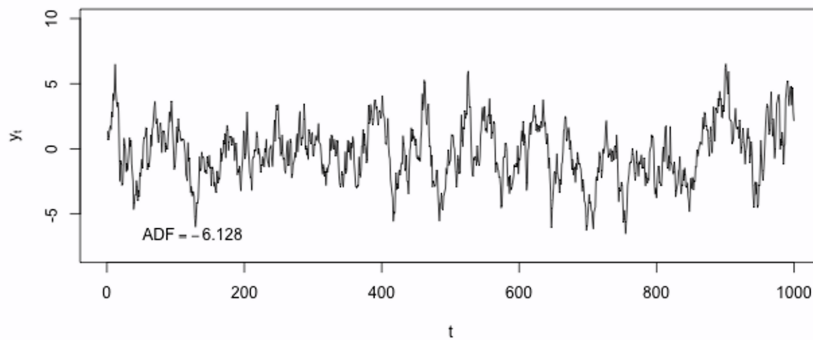


Figure 1: An example of a weakly stationary time series

2.3 Modelling Returns

Let $\{r_t\}$ be a time series representing the log returns of an asset. Then, we model $\{r_t\}$ as the following:

$$r_t = \mu_t + \epsilon_t, \quad t = 1, \dots, T \quad (2)$$

and we consider the following conditional mean and conditional variance equations:

$$\begin{aligned}\mu_t &= E(r_t|F_{t-1}) \\ \sigma_t^2 &= \text{Var}(r_t|F_{t-1}) = E((r_t - \mu_t)^2|F_{t-1}) = E(\epsilon_t^2|F_{t-1}),\end{aligned}\tag{3}$$

where F_{t-1} denotes the information set available at time $t - 1$.

In the following, we discuss two time series models. Namely, The autoregressive moving average (ARMA) model (Tsay, 2010, p. 64), and the generalized autoregressive conditional heteroskedasticity (GARCH) model (Tsay, 2010, p. 142). The former assumes that the time series exhibits a constant variance across time. It is mainly used to model the conditional mean $\{\mu_t\}$. On the other hand, the GARCH model assumes that the variance changes with time, and hence, it is mainly used to model the conditional variance $\{\sigma_t^2\}$. Furthermore, we discuss combining the two model into an ARMA-GARCH, model which we use to model $\{r_t\}$ assuming a non-constant variance.

First, we describe the ARMA model by explaining its two components, the Autoregressive (AR) model and the Moving Average (MA) mode:

2.4 AR Model

In an AR(p) (Autoregressive model of order p), Equation (2) becomes:

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + \epsilon_t,$$

where the shocks (or innovations) $\{\epsilon_t\}$ are assumed to be independent and identically distributed (i.i.d) with zero mean and constant variance σ^2 .

For simplicity, we will only consider the case of AR(1) with normally distributed innovations:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \epsilon_t,\tag{4}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. However, all the results stated can be easily generalized to describe an AR(p) with different assumption on the distribution of ϵ_t .

2.4.1 Properties of AR Models

Comparing Equations (4) and (3), we can write the equations for the conditional mean and conditional variance of an AR(1) model:

$$\begin{aligned}
E(r_t|F_{t-1}) &= \phi_0 + \phi_1 r_{t-1} \\
\text{Var}(r_t|F_{t-1}) &= E(\epsilon_t^2|F_{t-1}) = \text{Var}(\epsilon_t|F_{t-1}) = \sigma^2.
\end{aligned}$$

The unconditional mean can be calculated by taking the expectation of Equation (4):

$$E(r_t) = \phi_0 + \phi_1 E(r_{t-1}).$$

Assuming weak stationarity, $E(r_t) = E(r_{t-1}) = \mu$, then

$$E(r_t) = \mu = \frac{\phi_0}{1 - \phi_1}.$$

Using $\phi_0 = (1 - \phi_1)\mu$, we can rewrite Equation (4) as:

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \epsilon_t. \quad (5)$$

The unconditional variance can be calculated by taking the expectation of the square of Equation (5)

$$\begin{aligned}
\text{Var}(r_t) &= E[(r_t - \mu)^2] \\
&= E[\phi_1^2(r_{t-1} - \mu)^2 + \epsilon_t^2 + 2\phi_1(r_{t-1} - \mu)\epsilon_t] \\
&= E[\phi_1^2(r_{t-1} - \mu)^2 + \epsilon_t^2] \\
&= \phi_1^2 \text{Var}(r_{t-1}) + \text{Var}(\epsilon_t) \\
&= \phi_1^2 \text{Var}(r_{t-1}) + \sigma^2,
\end{aligned}$$

where in the second step, we used that $E[(r_{t-1} - \mu)\epsilon_t] = 0$ which can be seen from the fact that ϵ_t does not depend on past events, in particular ϵ_t does not depend on r_{t-1} and hence $\text{cov}(r_{t-1}, \epsilon_t) = 0 = E[(r_{t-1} - \mu)(\epsilon_t - 0)]$.

Once again, assuming weak stationarity, $\text{Var}(r_t) = \text{Var}(r_{t-1})$, then

$$\text{Var}(r_t) = \frac{\sigma^2}{1 - \phi_1^2}.$$

2.4.2 Parameter Estimation

For an AR(1) model described by Equation (4), the **conditional**-maximum likelihood estimation method that starts with the 2nd observation can be used to estimate the parameters ϕ_0 , ϕ_1 , and σ .

Assume we have T observations of $\{r_t\}$, then the **conditional**-likelihood function can be expressed as:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}|r_2, \dots, r_T; F_1) &= f(r_2, \dots, r_T|\boldsymbol{\alpha}, F_1) \\ &= f(r_3, \dots, r_T|\boldsymbol{\alpha}, F_2)f(r_2|\boldsymbol{\alpha}, F_1) \\ &= f(r_4, \dots, r_T|\boldsymbol{\alpha}, F_3)f(r_3|\boldsymbol{\alpha}, F_2)f(r_2|\boldsymbol{\alpha}, F_1) \\ &\vdots \\ &= \prod_{t=2}^T f(r_t|\boldsymbol{\alpha}, F_{t-1}),\end{aligned}$$

where $\boldsymbol{\alpha} = (\phi_0, \phi_1, \sigma)^t$ is a vector of the parameters to be estimated and F_t is as defined in Equation (3).

Now, assuming that $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ for some unknown parameter σ that need to be estimated, then we can write using Equation (4):

$$\begin{aligned}\epsilon_t \sim \mathcal{N}(0, \sigma^2) &\iff \\ r_t|F_{t-1} \sim \mathcal{N}(\phi_0 + \phi_1 r_{t-1}, \sigma^2).\end{aligned}$$

Using the above we can rewrite the **conditional**-likelihood function:

$$\mathcal{L}(\boldsymbol{\alpha}|r_2, \dots, r_T; F_1) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(r_t - (\phi_0 + \phi_1 r_{t-1}))^2}{2\sigma^2}\right], \quad (6)$$

where $\phi_0 + \phi_1 r_{t-1}$ is the conditional mean of r_t (see Equation (3)).

We want to maximize (6) with respect to $\boldsymbol{\alpha}$, which is usually done by maximizing its logarithm:

$$\begin{aligned}\ell(\boldsymbol{\alpha}|r_2, \dots, r_T; F_1) &= \ln \mathcal{L}(\boldsymbol{\alpha}|r_2, \dots, r_T; F_1) \\ &= \sum_{t=2}^T \left[-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \frac{(r_t - (\phi_0 + \phi_1 r_{t-1}))^2}{\sigma^2} \right].\end{aligned}$$

We maximize the expression above to get an estimate $\hat{\boldsymbol{\alpha}} = (\hat{\phi}_0, \hat{\phi}_1, \hat{\sigma})$. i.e

$$\hat{\boldsymbol{\alpha}} = \max_{\boldsymbol{\alpha}} \ell(\boldsymbol{\alpha}|r_2, \dots, r_T; F_1).$$

2.4.3 Forecasting

The forecasts of an AR(1) model can be obtained using Equation (4). Assuming that we have fitted an AR(1) model and obtained the fitted parameters $\hat{\mathbf{a}}$, then the 1-step ahead forecast of r_t assuming we are at time $t = h$ is:

$$\hat{r}_h(1) = \hat{\phi}_0 + \hat{\phi}_1 r_h.$$

Similarly, the ℓ step ahead forecast can be calculated recursively:

$$\hat{r}_h(\ell) = \hat{\phi}_0 + \hat{\phi}_1 \hat{r}_h(\ell - 1),$$

where $\ell \geq 2$.

2.5 MA Model

The Moving Average (MA) model is the second component of the ARMA model.

The equation MA(q) (Moving average model of order q) model for a time series $\{r_t\}$ is:

$$r_t = c_0 + \epsilon_t - \sum_{i=1}^q \theta_i \epsilon_{t-i},$$

where $c_0, \theta_i; i \in \{1, \dots, q\}$ represent the parameters of the model, and once again the innovation series $\{\epsilon_t\}$ is assumed to be i.i.d with zero mean and constant variance σ^2 .

Similar to the AR model, we will limit the discussion to the MA(1) model.

$$r_t = c_0 + \epsilon_t - \theta_1 \epsilon_{t-1}, \quad (7)$$

where we assume that $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$. Again, all the results stated can be easily generalized to describe an MA(q) model.

2.5.1 Properties of MA Models

The conditional mean and conditional variance of MA(1) model:

$$\begin{aligned} E(r_t | F_{t-1}) &= c_0 - \theta_1 \epsilon_{t-1} \\ \text{Var}(r_t | F_{t-1}) &= \text{Var}(\epsilon_t) = \sigma^2. \end{aligned}$$

The unconditional mean can be calculated by taking the expectation of Equation (7):

$$E[r_t] = c_0.$$

The unconditional variance can be calculated by taking the variance of Equation (7):

$$\text{Var}(r_t) = \sigma^2 + \theta_1^2 \sigma^2 = (1 + \theta_1^2) \sigma^2.$$

where we used the assumption that ϵ_t and ϵ_{t-1} are not correlated.

2.5.2 Parameter Estimation

Similar to Section 2.4.2, we can write the **conditional**-likelihood function:

$$\mathcal{L}(\boldsymbol{\alpha} | r_2, \dots, r_T; F_1) = \prod_{t=2}^T f(r_t | \boldsymbol{\alpha}, F_{t-1}), \quad (8)$$

where $\boldsymbol{\alpha} = (c_0, \theta_1, \sigma)^t$ is a vector of the parameters to be estimated, and F_t , as mentioned above, denotes the information available at time t .

Using the assumption that $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$, we can rewrite Equation (8):

$$\mathcal{L}(\boldsymbol{\alpha} | r_2, \dots, r_T; F_1) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(r_t - (c_0 - \theta_1\epsilon_{t-1}))^2}{2\sigma^2},$$

where $c_0 - \theta_1\epsilon_{t-1}$ is the conditional mean and ϵ_t can be expressed recursively as follows:

1. Assume $\epsilon_0 = 0$.
2. $\epsilon_1 = r_1 - c_1$.
3. $\epsilon_t = r_t - (c_t - \theta_t\epsilon_{t-1})$ for $t \geq 2$.

Again, we usually maximize the log likelihood rather than the likelihood:

$$\begin{aligned} \ell(\boldsymbol{\alpha} | r_2, \dots, r_T; F_1) &= \ln \mathcal{L}(\boldsymbol{\alpha} | r_2, \dots, r_T; F_1) \\ &= \sum_{t=2}^T \left[-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \frac{(r_t - (c_0 - \theta_1\epsilon_{t-1}))^2}{\sigma^2} \right]. \end{aligned}$$

2.5.3 Forecasting

The forecasts of an MA(1) model can be obtained using Equation (7). Assuming that we have fitted an MA(1) model and obtained the fitted parameters $\hat{\boldsymbol{\alpha}}$, then the 1-step ahead forecast of r_t assuming that we are at time $t = h$ is:

$$\hat{r}_h(1) = E[r_{h+1} | F_h] = \hat{c}_0 - \hat{\theta}_1 \hat{\epsilon}_h.$$

For $\ell \geq 2$, the ℓ -ahead step forecast is:

$$\hat{r}_h(\ell) = E[r_{h+\ell}|F_h] = \hat{c}_0,$$

where we have used the fact that $E[\epsilon_t|F_{t-2}] = 0$.

2.6 ARMA Model

Now we can combine the Autoregressive model with the Moving Average model to get the Autoregressive Moving Average (ARMA) model, which is defined as:

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} - \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t. \quad (9)$$

Equation (9) describes an ARMA(p, q) model where we assume that $\{\epsilon_t\}$ is i.i.d with mean zero and variance σ^2 .

Once again, we will only consider the case of ARMA(1, 1) model:

$$r_t = \phi_0 + \phi_1 r_{t-1} - \theta_1 \epsilon_{t-1} + \epsilon_t.$$

The properties, estimation process, and forecasting of an ARMA(1, 1) model can be directly deduced from Sections 2.4, and 2.5 and therefore, will be omitted.

2.7 ARCH Model

Let $\{r_t\}$ be the log return series of a an asset. The idea behind volatility modeling is that $\{r_t\}$ is either serially uncorrelated or with minor lower order serial correlation; however, it is dependent.

To measure serial correlation in a time series, we use the autocorrelation function (ACF) which indicates, as a function of ℓ , the correlation coefficient between r_t and $r_{t-\ell}$. More precisely:

$$\begin{aligned} ACF_\ell = \rho_\ell &= \frac{\text{Cov}(r_t, r_{t-\ell})}{\sqrt{\text{Var}(r_t)\text{Var}(r_{t-\ell})}} \\ &= \frac{\text{Cov}(r_t, r_{t-\ell})}{\text{Var}(r_t)}, \end{aligned}$$

where in the last step, we used $\text{Var}(r_t) = \text{Var}(r_{t-\ell})$ under the weak stationarity assumption. We say that a time series is serially uncorrelated if $\rho_\ell = 0$ for all $\ell > 0$ as shown in the left panel of Figure 2.

Figure 2 shows the monthly log returns of IBM between January of 1974 and December of 2008. Figure 23 (a) shows the ACF plot of the log returns and Figure 23 (b) shows the ACF plot of the squared log returns. It can be seen from the left panel that the monthly log returns are not serially correlated. However, the second panel indicates that there exists a higher order dependence in the return sequence that could be captured in a higher order model like ARCH and GARCH models as will be illustrated in the following (Tsay, 2010, p. 111).

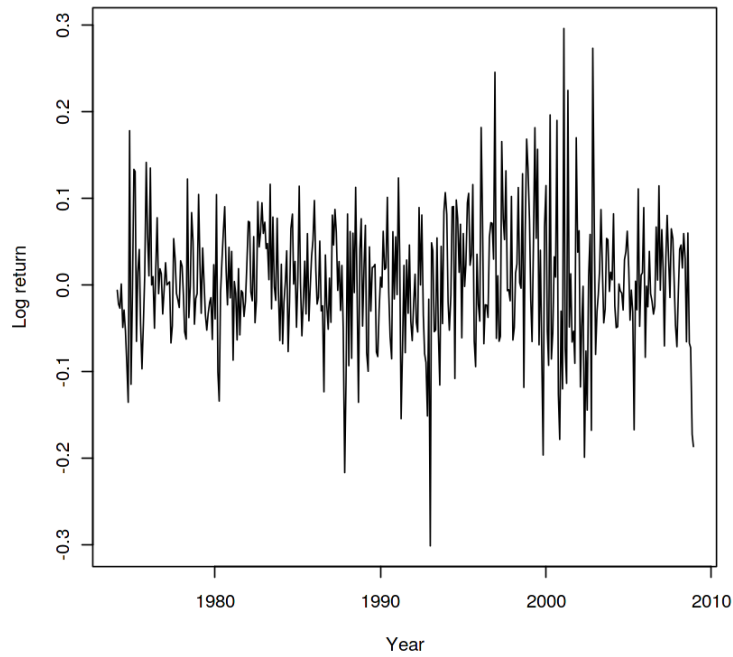


Figure 2: Monthly log returns of IBM from 1974-01-01 to 2008-12-31

The idea behind ARCH models is the following:

1. The shock (or innovation) ϵ_t of an asset return is serially uncorrelated but dependent.
2. This dependence can be described by a quadratic function of its lagged values

More specifically, an ARCH(m) model assumes that:

$$\begin{aligned} \epsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^m \alpha_i \epsilon_{t-i}^2, \end{aligned} \tag{10}$$



Figure 3: Sample ACF of IBM monthly log returns from January 1974 to December 2008: (a) ACF of the log returns, (b) ACF of the squared log returns

where Z_t is an i.i.d sequence of random variables with mean zero and variance one.

In the following, we will only focus on the ARCH(1) model, with normally distributed standardized innovations (i.e. $Z_t \sim \mathcal{N}(0, 1)$).

$$\begin{aligned}\epsilon_t &= \sigma_t Z_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2.\end{aligned}\tag{11}$$

However, all the results stated can be easily generalized to describe an ARCH(m) model.

2.7.1 Properties of ARCH Models

The conditional mean of an ARCH(1) model can be derived from Equation (11):

$$E(\epsilon_t | F_{t-1}) = E(\sigma_t Z_t | F_{t-1}) = E(\sqrt{\alpha_0 + \alpha_1 \epsilon_{t-1}^2} Z_t | F_{t-1}) = 0,$$

where we used the fact that $\epsilon_{t-1} | F_{t-1}$ is known.

Similarly, the conditional variance of an ARCH(1) model can be derived as following:

$$\begin{aligned}
\text{Var}(\epsilon_t|F_{t-1}) &= E(\epsilon_t^2|F_{t-1}) \\
&= E(\sigma_t^2 Z_t^2|F_{t-1}) \\
&= E((\alpha_0 + \alpha_1 \epsilon_{t-1}^2) Z_t^2|F_{t-1}) \\
&= (\alpha_0 + \alpha_1 \epsilon_{t-1}^2) E(Z_t^2|F_{t-1}) \\
&= \sigma_t^2(1) \\
&= \sigma_t^2,
\end{aligned} \tag{12}$$

where in the forth step, we again used the fact that $\epsilon_{t-1}|F_{t-1}$ is known.

The unconditional mean of ϵ_t is zero.

$$E(\epsilon_t) = E[E(\epsilon_t|F_{t-1})] = E[\sigma_t E(z_t)] = 0.$$

The unconditional variance of ϵ_t can be calculated as:

$$\begin{aligned}
\text{Var}(\epsilon_t) &= E(\epsilon_t^2) \\
&= E[E(\epsilon_t^2|F_{t-1})] \\
&= E[\text{Var}(\epsilon_t|F_{t-1})] \\
&= E[\sigma_t^2] && \text{(Using Equation (12))} \\
&= E[\alpha_0 + \alpha_1 \epsilon_{t-1}^2] && \text{(Using Equation (11))} \\
&= \alpha_0 + \alpha_1 E(\epsilon_{t-1}^2)
\end{aligned}$$

Assuming ϵ_t is weakly stationary , then $\text{Var}(\epsilon_t) = \text{Var}(\epsilon_{t-1}) = E(\epsilon_{t-1}^2)$, thus we get

$$\text{Var}(\epsilon_t) = \frac{\alpha_0}{1 - \alpha_1}.$$

2.7.2 Parameter Estimation

The derivation of the **conditional**-likelihood function of an ARCH(1) model is identical to that of an AR(1) model (see Section 2.4.2).

$$\mathcal{L}(\boldsymbol{\alpha}|\epsilon_2, \dots, \epsilon_T; F_1) = \prod_{t=2}^T f(\epsilon_t|\boldsymbol{\alpha}, F_{t-1}). \tag{13}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \sigma_t)^t$ represents the model parameters.

The above equation depends on the distribution assumption of z_t . Following our assumption that $z_t \sim \mathcal{N}(0, 1)$, then we have:

$$\begin{aligned}\epsilon_t | \epsilon_{t-1} &\sim \mathcal{N}(0, \sigma_t^2) \\ &= \mathcal{N}(0, \alpha_0 + \alpha_1 \epsilon_{t-1}^2).\end{aligned}\tag{14}$$

And hence we can rewrite the **conditional**-likelihood equation:

$$\mathcal{L}(\boldsymbol{\alpha} | \epsilon_2, \dots, \epsilon_T; F_1) = \prod_{t=2}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma_t^2}\right).$$

Once again we consider maximizing the conditional log-likelihood function:

$$\ell(\boldsymbol{\alpha} | \epsilon_1, \dots, \epsilon_T) = \sum_{t=2}^T \left(-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma_t^2 - \frac{1}{2} \frac{\epsilon_t^2}{\sigma_t^2} \right).\tag{15}$$

2.7.3 Model Checking

For a fitted ARCH model based on observations $\{\epsilon_1, \dots, \epsilon_T\}$, the standardized residuals can be calculated as

$$\hat{z}_t = \frac{\epsilon_t}{\hat{\sigma}_t},$$

for $t = 1, \dots, T$, where $\hat{\sigma}_t$ can be estimated recursively using Equation (11):

1. $\hat{\sigma}_2 = \hat{\alpha}_0 + \hat{\alpha}_1 \epsilon_1$.
2. $\hat{\sigma}_3 = \hat{\alpha}_0 + \hat{\alpha}_1 \epsilon_2$.
3. Similarly for $t > 3$, $\hat{\sigma}_t = \hat{\alpha}_0 + \hat{\alpha}_1 \epsilon_{t-1}$.

If the ARCH model is fitted properly, then $\{\hat{z}_t\}$ form an approximate i.i.d sequence. So to test if the model adequately captures the volatility, we verify that the sequence $\{\hat{z}_t^2\}$ is not serially correlated which can be tested using the Ljung-Box test (see Section 2.9).

2.7.4 Forecasting

The forecasts of an ARCH(1) model can be obtained recursively using Equation (11). Assume that we are at time $t = h$, then the 1-step ahead forecast of $\hat{\sigma}_{h+1}^2$ is

$$\hat{\sigma}_h^2(1) = \alpha_0 + \alpha_1 \hat{\sigma}_h^2.$$

Similarly the ℓ step forecast is

$$\hat{\sigma}_h^2(\ell) = \alpha_0 + \alpha_1 \hat{\sigma}_h^2(\ell - 1) \quad (\ell \geq 2).$$

2.8 GARCH Model

The GARCH model (or Generalized ARCH) is a generalization of the ARCH model by modifying Equation (10):

$$\begin{aligned}\epsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^m \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^s \beta_i \sigma_{t-i}^2.\end{aligned}$$

This gives the equation of a GARCH(m, s) model. This generalization is due to the fact that an ARCH(m) model often requires a large order m to model the volatility in the return series while, on the other hand, the GARCH(m, s) model requires lower values of m and s to fit well (Tsay, 2010, p. 131).

As before, in the following we will only consider the case of a GARCH(1, 1) model which can be easily generalized to a GARCH(m, s) model:

$$\begin{aligned}\epsilon_t &= \sigma_t z_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.\end{aligned}$$

2.8.1 Parameter Estimation

The estimation of a GARCH(1, 1) model is very similar to an ARCH(1) model. Now Equation (14) looks like:

$$\begin{aligned}\epsilon_t | \epsilon_{t-1}, \sigma_{t-1} &\sim \mathcal{N}(0, \sigma_t^2) \\ &= \mathcal{N}(0, \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2).\end{aligned}$$

And the conditional-likelihood function is identical to that in Equation (13) and (15). The only difference is now we calculate σ_t^2 recursively as follows:

1. Assume we are at a certain iteration in the optimization process where we have intermediate values of $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1)$.
2. Initialize $\hat{\sigma}_1$ to be the sample standard deviation of the observed sequence $\{\epsilon_t\}$.
3. Using Equation (2.8) we estimate σ_2^2 by $\hat{\sigma}_2^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \epsilon_1^2 + \hat{\beta}_1 \hat{\sigma}_1^2$
4. Similar to step 3, we can estimate σ_t^2 by $\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \epsilon_{t-1}^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2$ for $t \in \{3, \dots, T\}$.

2.8.2 Model Checking

Model checking is identical to Section 2.7.3

2.8.3 Forecasting

Similar to Section 2.7.4 the 1-step ahead forecast of σ_{h+1}^2 assuming we are at time $t = h$ is

$$\hat{\sigma}_h^2(1) = \hat{\alpha}_0 + \hat{\alpha}_1 \epsilon_h^2 + \hat{\beta}_1 \hat{\sigma}_h^2.$$

For the 2-step ahead forecast, we use $\epsilon_t = \sigma_t z_t$ and rewrite the volatility equation

$$\begin{aligned} \sigma_{h+2}^2 &= \alpha_0 + \alpha_1 \sigma_{h+1}^2 z_{h+1}^2 + \beta_1 \sigma_{h+1}^2 \\ &= \alpha_0 + (\alpha_1 + \beta_1) \sigma_{h+1}^2 + \alpha_1 \sigma_{h+1}^2 (z_{h+1} - 1). \end{aligned}$$

Notice that $E(z_{h+1}^2 - 1 | F_h) = 0$, then the 2-step ahead forecast is

$$\hat{\sigma}_h^2(2) = \hat{\alpha}_0 + (\hat{\alpha}_1 + \hat{\beta}_1) \hat{\sigma}_h^2(1).$$

And in general for $\ell > 1$ the ℓ -step ahead forecast is

$$\hat{\sigma}_h^2(\ell) = \hat{\alpha}_0 + (\hat{\alpha}_1 + \hat{\beta}_1) \hat{\sigma}_h^2(\ell - 1).$$

2.9 The Ljung-Box Test

The Ljung-Box test can be used to test for the existence of auto-correlations in a time series. In our case, it is used to test the adequacy of the mean equation by testing for the existence of auto correlations in the residuals sequence $\{\hat{z}_t\}$. In addition, it is used to test the adequacy of the volatility equation by testing for the existence of auto-correlations in the squared residuals sequence $\{\hat{z}_t^2\}$.

More precisely, we can state the null and alternative hypotheses for a random sequence $\{X_t\}, t = 1, \dots, T$ for some chosen number of lags m :

1. H_0 : The random sequence $\{X_t\}$ does not exhibit any serial correlations (i.e. $\rho_1 = \rho_2 = \dots = \rho_m = 0$).
2. H_a : $\{X_t\}$ exhibit serial correlation.

Here ρ_k is the auto-correlation at lag k :

$$\rho_k = \text{cor}(X_t, X_{t+k}).$$

The test statistic based on the observed sequence $\{x_t, t = 1, \dots, T\}$ is:

$$Q(m) = T(T+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T-k}$$

where $\hat{\rho}_k$ is the empirical auto-correlation at lag k and is given by:

$$\hat{\rho}_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Under the null hypothesis, $Q(m)$ follows a $\chi^2_{(m)}$ distribution (Tsay, 2010), and hence we reject H_0 at the significance level of α if $Q > \chi^2_{1-\alpha, m}$. Here $\chi^2_{\beta, m}$ denotes the β -quantile of $\chi^2_{(m)}$.

2.10 Estimation of ARMA-GARCH models

Again, we will consider the simplest case of an ARMA(p, q)-GARCH(m, s) model, namely the ARMA(1, 1)-GARCH(1, 1) model. Assume we have the log return series r_t , then an ARMA(1, 1)-GARCH(1, 1) model assumes the following:

$$r_t = \phi_0 + \phi_1 r_{t-1} - \theta_1 \epsilon_{t-1} + \epsilon_t \quad (16a)$$

$$\epsilon_t = \sigma_t z_t \quad (16b)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (16c)$$

Where ϕ_0, ϕ_1, θ_1 are the parameters of the mean equation (16a), while $\alpha_0, \alpha_1, \beta_1$ are the parameter of the volatility equation (16c). We can estimate those parameters using the following two approaches.

2.10.1 A Two-pass estimation method

This method will first estimate the mean equation without any ARCH effects. (i.e. assuming constant variance). Then we use the residuals of the fitted mean model $\{\hat{\epsilon}_t = r_t - (\hat{\phi}_0 + \hat{\phi}_1 r_{t-1} - \hat{\theta}_1 \hat{\epsilon}_{t-1})\}$ as an observation sequence to fit the GARCH model specified in Equations (16b) & (16c) as described in Sections 2.7.2 and 2.8.1. Note that the GARCH estimation in this case is only an approximation; however, it is often found to be a good approximation to the true parameters, when the sample size is large (Tsay, 2010, p. 140).

2.10.2 Joint estimation

Another approach is to estimate the parameters of the mean and volatility equations jointly. From Equation (16), we can show that

$$\epsilon_t | F_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$$

which implies that

$$r_t|F_{t-1} \sim \mathcal{N}(\phi_0 + \phi_1 r_{t-1} - \theta_1 \epsilon_{t-1}, \sigma_t^2)$$

Then similar to Equation (13), we can write the the conditional-likelihood function:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}|r_2, \dots, r_T; F_1) &= \prod_{t=2}^T f(r_t|\boldsymbol{\alpha}, F_{t-1}) \\ &= \prod_{t=2}^T \frac{1}{\sqrt{(2\pi\sigma_t^2)}} \exp\left(-\frac{\epsilon_t}{2\sigma_t^2}\right) \\ &= \prod_{t=2}^T \frac{1}{\sqrt{(2\pi\sigma_t^2)}} \exp\left(-\frac{(r_t - (\phi_0 + \phi_1 r_{t-1} - \theta_1 \epsilon_{t-1}))^2}{2(\alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2)}\right), \end{aligned}$$

where $\boldsymbol{\alpha} = (\phi_0, \phi_1, \theta_1, \alpha_0, \alpha_1, \beta_1)$ are the model parameters to be estimated. Here ϵ_t and σ_t are estimated recursively as follows:

1. Initialize $\hat{\sigma}_1$ to be the sample standard deviation of the observed sequence $\{r_t\}$.
2. Initialize $\hat{\epsilon}_1 = 0$.
3. Assume we are at a certain iteration in the optimization process where we have intermediate values of $(\hat{\phi}_0, \hat{\phi}_1, \hat{\theta}_1, \hat{\alpha}_0, \hat{\alpha}_1, \beta_1)$.
4. Using Equation (16c), we estimate σ_2^2 by $\hat{\sigma}_2^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\epsilon}_1^2 + \hat{\beta}_1 \hat{\sigma}_1^2$.
5. Using Equation (16a), we estimate ϵ_2 by $\hat{\epsilon}_2 = r_2 - (\hat{\phi}_0 + \hat{\phi}_1 r_1 - \hat{\theta}_1 \hat{\epsilon}_1)$.
6. Similarly, we estimate σ_t^2 by $\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\epsilon}_{t-1}^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2$ for $t \in \{3, \dots, T\}$.
7. Likewise, we estimate ϵ_t by $\hat{\epsilon}_t = r_t - (\hat{\phi}_0 + \hat{\phi}_1 r_{t-1} - \hat{\theta}_1 \hat{\epsilon}_{t-1})$ for $t \in \{3, \dots, T\}$.

2.11 Copulas

Copulas are used to model the dependence structure among random variables separate from the univariate margins. Specifically, a copula C of dimension d is a multivariate cumulative distribution function on the hypercube $[0, 1]^d$ with uniformly distributed margins. Sklar's theorem (Czado, 2019, p. 13) describes copulas in more precise terms:

Theorem. Let \mathbf{X} be a d -dimensional random vector with joint distribution function F and marginal distribution functions $F_i, i = 1, \dots, d$, then the joint distribution function can be expressed as:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$$

with associated density function:

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d))f_1(x_1), \dots, f_d(x_d). \quad (17)$$

Here c denotes the associated density to the distribution function C .

We will only consider the case of bivariate copulas (i.e. $d = 2$) since they are the building blocks of vine copulas which will be explained in the next section.

2.11.1 Conditioning

From Sklar's theorem we can write the conditional density $f_{1|2}$ and conditional distribution function $F_{1|2}$:

$$\begin{aligned} f_{1|2}(x_1|x_2) &= c_{12}(F_1(x_1), F_2(x_2))f_2(x_2) \\ F_{1|2}(x_1|x_2) &= \frac{\partial}{\partial F_2(x_2)} C_{12}(F_1(x_1), F_2(x_2)). \end{aligned} \quad (18)$$

2.11.2 Dependence Measure

Kendall's τ , or Spearman's ρ are usually used to measure the dependence of the copula variables. Both of them are invariant to marginal transformations which is why they are used rather than Pearson's correlation.

The dependence measure Kendall's τ between two random variables X_1 and X_2 is defined as:

$$\tau(X_1, X_2) = P((X_{11} - X_{21})(X_{12} - X_{22}) > 0) - P((X_{11} - X_{21})(X_{12} - X_{22}) < 0),$$

where $(X_{11}, X_{12})^t, (X_{21}, X_{22})^t$ are independent and identically distributed copies of $(X_1, X_2)^t$.

On the other hand, Spearman's correlation is defined as:

$$\rho_s(X_1, X_2) = \text{cor}(F_1(X_1), F_2(X_2)),$$

where "cor" refers to Pearson's correlation which is given by:

$$\text{cor}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)\text{Var}(X_2)}.$$

2.11.3 Copula Families

There are many bivariate copula families which differ in their dependence properties. Some of the common ones include the Gaussian copula, the Student t copula, the Gumbel copula, and the Clayton copula. In order to visualize the differences among the different families, it is useful to define 3 scales:

label=() The original scale (X_1, X_2) .

lbbel=() The copula scale (or U-scale) $(U_1, U_2) = (F_1(X_1), F_2(X_2))$.

lcbel=() The z-scale $(Z_1, Z_2) = (\Phi^{-1}(U_1), \Phi^{-1}(U_2))$.

where Φ is the distribution of a $\mathcal{N}(0, 1)$ random variable.

Figure 4 shows the contours of the associated bivariate density of (Z_1, Z_2) for different copula families.

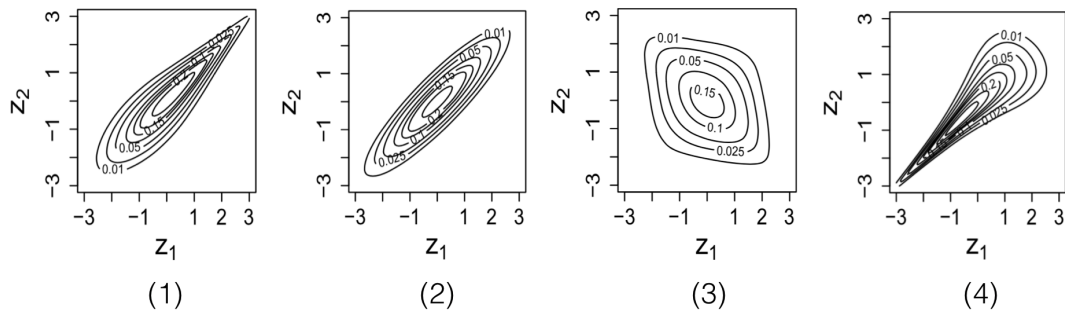


Figure 4: Normalized contour plots of some bivariate copula families: (1) Gumbel with $\tau = 0.7$, (2) Gaussian with $\tau = 0.7$, (3) Student t with 4 degrees of freedom and $\tau = -0.2$, (4) Clayton with $\tau = 0.7$ (Czado, 2019, p. 60–61).

2.11.4 Estimation

Given i.i.d data (x_{i1}, x_{i2}) for $i = 1, \dots, n$, we want to fit a bivariate copula to the data. To do that we first need to transform the data from the original scale to the copula scale. In particular, the pseudo copula data is given by:

$$(u_{i1}, u_{i2}) = (\hat{F}_1(x_{i1}), \hat{F}_2(x_{i2}))$$

where \hat{F}_j , $j \in \{1, 2\}$ is the estimated marginal distribution of X_j based on the observations (x_{1j}, \dots, x_{nj}) .

Assuming that a parametric model has been used to estimate F_1 and F_2 , then we use inference for margins (IFM) method for estimating the copula parameters (Joe, 2005). More precisely, we maximize the likelihood function:

$$\ell(\theta; \mathbf{u}) = \prod_{i=1}^n c(u_{i1}, u_{i2}; \theta),$$

where θ is the copula parameter to be estimated.

2.12 Vine Copulas

The idea behind vine copulas is to construct a multivariate copula by only using bivariate copulas as building blocks. In the following, we will show how to construct a multivariate copula of dimension $d = 3$ using bivariate copulas. Using Bayes rule we can write $f(x_1, x_2, x_3)$ as:

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2)f_{2|1}(x_2|x_1)f(x_1) \quad (19)$$

where $f_{j|D}$ denotes the conditional density function of X_j given $\mathbf{X}_D = \mathbf{x}_D$.

Now we want to express each term in terms of bivariate copulas. First let's consider $f_{3|12}(x_3|x_1, x_2)$:

$$\begin{aligned} f_{3|12}(x_3|x_1, x_2) &= \frac{f_{13|2}(x_1, x_3|x_2)}{f_{1|2}(x_1|x_2)} && \text{(Bayes rule)} \\ &= \frac{c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))f_{1|2}(x_1|x_2)f_{3|2}(x_3|x_2)}{f_{1|2}(x_1|x_2)} \\ &= c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2))f_{3|2}(x_3|x_2) \end{aligned} \quad (20)$$

where in Step 2, we used Sklar's theorem (Equation (17)). $F_{j|D}$ is the conditional distribution function of X_j given \mathbf{X}_D , and $c_{ij;k}$ is the copula density corresponding to the conditional distribution of (X_i, X_j) given $X_k = x_k$.

Using Equation (18), we can write:

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2))f_2(x_2) \quad (21)$$

$$f_{3|2}(x_2|x_3) = c_{23}(F_2(x_2), F_3(x_3))f_3(x_3) \quad (22)$$

Substituting (20), (21), and (22) into (19):

$$f(x_1, x_2, x_3) = c_{13,2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{12}(F_1(x_1), F_2(x_2)) c_{23}(F_2(x_2), F_3(x_3)) f(x_1) f_2(x_2) f_3(x_3) \quad (23)$$

which describe the joint density of the three variables using only bivariate copulas and the marginal densities.

2.12.1 Regular Vine Copulas

Equation (23) does not uniquely represent the joint density of X_1, X_2 , and X_3 . Namely, we could have alternatively wrote the following equivalent expressions:

$$\begin{aligned} f(x_1, x_2, x_3) &= c_{13,2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) c_{12}(F_1(x_1), F_2(x_2)) \\ &\quad c_{23}(F_2(x_2), F_3(x_3)) f(x_1) f_2(x_2) f_3(x_3) \\ &= c_{12,3}(F_{1|3}(x_1|x_3), F_{2|3}(x_2|x_3); x_3) c_{13}(F_1(x_1), F_3(x_3)) \\ &\quad c_{23}(F_2(x_2), F_3(x_3)) f_1(x_1) f_2(x_2) f_3(x_3) \\ &= c_{23,1}(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1); x_1) c_{12}(F_1(x_1), F_2(x_2)) \\ &\quad c_{13}(F_1(x_1), F_3(x_3)) f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned}$$

So as we can see, the construction of the joint density using bivariate copulas is not unique and depends on the conditioning order. Here comes the role of regular vines that allows for arbitrary dimensions as well as for general conditioning orders. More precisely, a regular vine on d variables is defined by a sequence of trees $\{T_j = (V_j, E_j)\}, j = 1, \dots, d-1$ where V_j represents the vertex set of T_j , while E_j represents the edge set of T_j . This sequence represents a valid regular vine, if it satisfies the following conditions:

1. The vertex set of T_1 is $V_1 = 1, \dots, d$.
2. For $j \geq 2, V_j = E_{j-1}$.
3. Proximity condition: if edge $(a, b) \in E_j$, then $|a \cap b| = 1$ for $j = 2, \dots, d-1$.

The last condition (proximity condition) ensures that for an edge to be part of a tree T_j , then the corresponding edges in tree T_{j-1} share a node.

In a vine distribution model, each edge of $T_j, j \in \{1, \dots, d\}$ represents a bivariate copula. To illustrate, consider an example of 5 random variables X_1, X_2, X_3, X_4, X_5 , then a regular

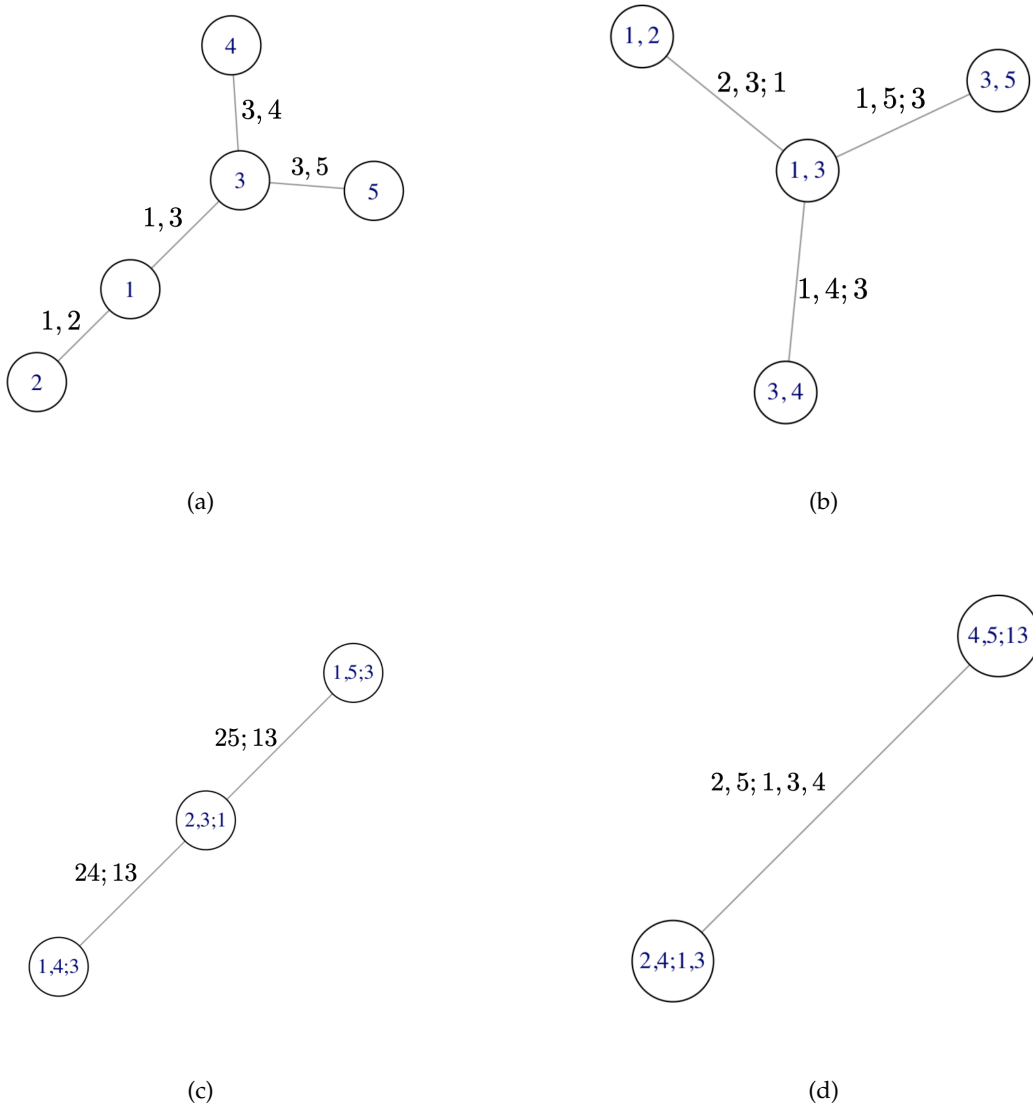


Figure 5: Example of a regular vine tree sequence of dimension 5

vine tree structure for $F(X_1, X_2, X_3, X_4, X_5)$ is shown in Figure 5, where an edge $i, j; \mathbf{D}$ corresponds to the bivariate copula:

$$C(F_{i|\mathbf{D}}(x_i|\mathbf{x}_D), F_{j|\mathbf{D}}(x_j|\mathbf{x}_D)),$$

with the corresponding copula density:

$$c(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D)),$$

such that the conditional bivariate distribution of X_i, X_j given $\mathbf{X}_D = \mathbf{x}_D$ is given by:

$$F(x_i, x_j|\mathbf{x}_D) = C(F_{i|D}(x_i|\mathbf{x}_D), F_{j|D}(x_j|\mathbf{x}_D))$$

For example, the edge (1,5;3) in tree (b) of Figure 5 indicates the bivariate copula

$$C(F_{1|3}(x_1|x_3), F_{5|3}(x_5|x_3)).$$

Notice that the joint density function $f(x_1, x_2, x_3, x_4, x_5)$ can be fully expressed by only using the bivariate copulas corresponding to the edges of Figure 5 alongside the marginal densities. The joint density can be derived similar to how we derived Equation (23). See (Czado, 2019, p. 105) for an example.

Finally, as a remark, a regular vine distribution is considered a regular vine copula if all the margins follow a uniform distribution on $[0, 1]$.

2.13 Value-at-Risk (VaR)

Given a single asset A with log-return series $\{r_t^A\}, t = 1, \dots, T$, we can define the VaR of A , on day t , at a certain level $\alpha \in (0, 1)$, by the largest number r^* such that the probability of r_t^A falling below r^* is not larger than α . More formally:

$$VaR_\alpha^{A,t} = \sup\{r^* \in \mathcal{R} : P(r_t^A < r^*) \leq \alpha\} = \sup\{r^* : F_{r_t^A}(r^*) \leq \alpha\}.$$

Thus $VaR_\alpha^{A,t}$ is the α -quantile of the r_t^A distribution as shown in Figure 6

The above approach can be easily extended to define the VaR of a portfolio. A portfolio is a weighted collection of assets. As an example, consider a certain portfolio Ω that is built using 3 assets A, B, C with weights w_A, w_B, w_C respectively such that $w_A + w_B + w_C = 1$, and $0 < w_i < 1, i \in \{A, B, C\}$. In addition, assume that the log return series for each of the assets are given by $\{r_t^A\}, \{r_t^B\}, \{r_t^C\}, t = 1, \dots, T$. Then we define the log return of the portfolio Ω on day t as ¹:

$$r_t^\Omega := w_A r_t^A + w_B r_t^B + w_C r_t^C, t = 1, \dots, T. \quad (24)$$

We can use r_t^Ω to define the VaR for the portfolio:

$$VaR_\alpha^{\Omega,t} = \sup\{r^* \in \mathcal{R} : P(r_t^\Omega < r^*) \leq \alpha\} = \sup\{r^* : F_{r_t^\Omega}(r^*) \leq \alpha\}.$$

¹Equation (24) is only an approximation for the log-returns of a portfolio; however, in practice, it is often used (Tsay, 2010, p. 5).

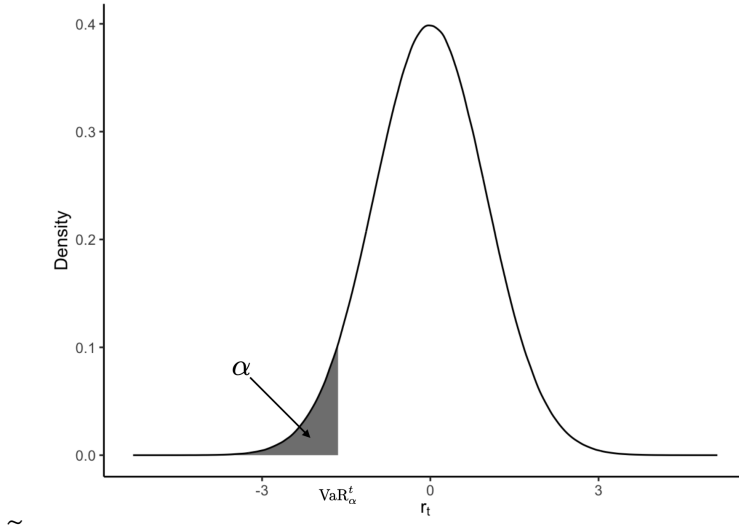


Figure 6: The distribution of the returns of an asset on day t . Shows the VaR at level α

2.13.1 Estimating the VaR

There are several methods to evaluate the VaR. However, in our analysis we will focus on the Monte Carlo method.

To illustrate the Monte Carlo method for estimating the VaR, consider the case of a single asset A with log return series $\{r_t^A\}$. Now, assume that we know the the distribution of $r_{t_0}^A$ on a specific day t_0 , then using the Monte Carlo method, we simulate a random sample from the distribution of $r_{t_0}^A$ and then we take the empirical α -quantile of the sample to be our estimate of $\text{VaR}_\alpha^{A,t_0}$.

Now, we give a more practical example, on how to estimate the VaR of a single asset using an ARMA-GARCH model to fit the return series, while using the aforementioned Monte Carlo method to forecast the VaR. Let the asset of interest be A with its log return series $\{r_t^A\}, t = 1, \dots, T$. We are interested in estimating the VaR of A on day $T + 1$ for level α (i.e. $\text{VaR}_\alpha^{A,T+1}$):

1. Fit an ARMA(1, 1)-GARCH(1, 1) model on $\{r_t^A\}, t = 1, \dots, T$ as described in Section 2.10.
2. Forecast the mean $\hat{\mu}_{T+1}^A$ (see Sections 2.4, and 2.5).
3. Forecast the volatility $\hat{\sigma}_{T+1}^A$ as described in Section 2.8.3.
4. Assuming the standardized residuals of the fitted ARMA-GARCH model follow

a $\mathcal{N}(0, 1)$, then we generate a sample $\hat{z}_1, \dots, \hat{z}_k$ from $\mathcal{N}(0, 1)$ for some large k (e.g. $k = 1000$).

5. $\hat{z}_1, \dots, \hat{z}_k$ can be used to get an approximate sample of the returns on day $T + 1$ $\{\hat{r}_{T+1}^{A,1}, \dots, \hat{r}_{T+1}^{A,k}\}$ using $\hat{r}_{T+1}^{A,i} = \hat{\mu}_{T+1}^A + \hat{\sigma}_{T+1}^A \hat{z}_i$ for $i = 1, \dots, k$.
6. Then, the empirical α -quantile of the sample $\{\hat{r}_{T+1}^{A,1}, \dots, \hat{r}_{T+1}^{A,k}\}$ is an estimate of $\text{VaR}_\alpha^{A,T+1}$.

In Section 3, we will extend the above approach to estimate the VaR of a portfolio of assets by using a separate ARMA-GARCH model to fit each asset returns, while utilizing a vine copula model to capture the cross-sectional dependence among the different components of the portfolio.

2.14 Backtesting Value-at-Risk

Assume we have a model that is used to predict the VaR, then the idea behind backtesting is to diagnose if the model is correctly fitted by comparing its output to historical data.

Many backtesting methods are based on the concept of what is called a *violation process* which is explained in the following.

Consider an asset A with log return series $\{r_t^A\}$, then as described in the previous section, $\text{VaR}_\alpha^{A,t}$ could be interpreted such that the probability of r_t^A falling below $\text{VaR}_\alpha^{A,t}$ on a given day t is α . More formally:

$$P(r_t^A < \text{VaR}_\alpha^{A,t}) = \alpha.$$

Thus, we can define the violation process on a given day t at level α :

$$I_t(\alpha) = \mathbb{1}_{\{r_t^A < \text{VaR}_\alpha^{A,t}\}}, \quad (25)$$

where $\mathbb{1}$ is the indicator function which evaluates to 1 when the condition inside the curly brackets is true and evaluates to 0 otherwise.

Now assume we have the return sequence $\{r_t\}$ $t \in \{1, \dots, T, \dots, T + m\}$; such that we fit a VaR forecast model using $\{r_t\}$ $t \in \{1, \dots, T\}$, while we use $\{r_{T+1}, \dots, r_{T+m}\}$ to backtest the fitted model. To that end, we need to build the sequence $\{I_t(\alpha)\}_{t=T+1}^{T+m}$. We build this sequence by comparing the observed sequence against the predicted VaR for each day $t \in \{T + 1, \dots, T + m\}$ as shown in Equation (25). Now if our model is fitted correctly, we would expect that:

$$P(I_t(\alpha) = 1) = \alpha$$

$$P(I_t(\alpha) = 0) = 1 - \alpha.$$

Following this reasoning, we can distinguish between two types of coverage tests:

2.14.1 Unconditional Coverage Tests

We say that a VaR_α model has a correct unconditional coverage if

$$P(I_t(\alpha) = 1) = E[I_t(\alpha)] = \alpha.$$

This basically suggested that unconditional coverage indicates that the proportion of losses exceeding the VaR predictions is not significantly different than α . More formally, the null hypothesis for an unconditional coverage test is:

$$H_0 : E[I_t(\alpha)] = \alpha.$$

Using this assumption, it is enough to verify that the number of violations x follows a binomial distribution with success probability α :

$$f(x) = \binom{m}{x} \alpha^x (1 - \alpha)^{(m-x)},$$

where $x = 0, \dots, m$.

As the number of observations increase, we can approximate the binomial distribution with a normal distribution, so for testing the null hypothesis, we can use:

$$\frac{x - \alpha m}{\sqrt{\alpha(1 - \alpha)m}} \approx \mathcal{N}(0, 1),$$

where αm is the expected number of violation and $\sqrt{\alpha(1 - \alpha)m}$ is its variance.

The Kupiec Test

One of the commonly used unconditional coverage tests is the Kupiec test (Roccioletti, 2015, p. 49) with the null hypothesis defined as:

$$H_0 : \alpha = \frac{x}{m}.$$

The idea behind it is to find out if there is a significant discrepancy between the observed proportion of violations $\frac{x}{m}$ and the theoretical one α . The test statistic used is the likelihood ratio between the observed and the theoretical proportions:

$$LR_{uc}^x = -2 \ln \left(\frac{\alpha^x (1 - \alpha)^{m-x}}{\left(\frac{x}{m}\right)^x \left(1 - \frac{x}{m}\right)^{m-x}} \right),$$

where the numerator is the likelihood under the null hypothesis using the theoretical violation factor α , while the denominator is the likelihood using the observed proportion $\frac{x}{m}$.

Under the null hypothesis (the assumption of a correct model) LR_{uc}^x is asymptotically χ_1^2 distribution with one degree of freedom (Roccioletti, 2015, p. 49). So, we reject the null hypothesis if the value of LR_{uc}^x is greater than the critical value of χ_1^2 and the model is considered incorrect.

2.14.2 Conditional Coverage Tests

As discussed in Section 2.7, the volatility of daily returns is time varying. So, if a VaR forecast model fails to detect the time varying aspect of volatility, then the VaR forecasts will exhibit a delay in adapting to the changing market conditions and violations will appear to be clustered. Therefore, we need to address the independence of violations problem: *In a correct VaR forecast model, the violations on day t should not depend on whether or not a violation has occurred on day $t - 1$.*

In the following, we discuss the *Christoffersen's test* (Roccioletti, 2015, p. 51) which along with testing for the correct violation proportion, it also tests for the independence of the violations.

Christoffersen's Interval Forecast Test

One of the commonly used conditional coverage tests is the Christoffersen's Interval Forecast Test. Christoffersen's test assumes that the dependence structure of the sequence of violations $\{I_t\}$ can be described as a first order Markov chain with 2 states and the following transition matrix:

$$A_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix},$$

where we interpret the matrix elements as follows:

- π_{01} is the probability of a violation following a non-violation.
- π_{11} is the probability of a violation following a violation.

- $1 - \pi_{01}$ is the probability of a non-violation following a non-violation.
- $1 - \pi_{11}$ is the probability of a non-violation following a violation.

Given the sample of observation $\{I_{T+1}, \dots, I_{T+m}\}$, we can write the likelihood function of the first order Markov process as:

$$L(A_1) = (1 - \pi_{01})^{N_{00}} \pi_{01}^{N_{01}} (1 - \pi_{11})^{N_{10}} \pi_{11}^{N_{11}},$$

where N_{ij} is the number of consecutive observations with j following i for $i, j = 1, 2$. More formally:

$$N_{ij} = \sum_{k=T+2}^{T+m} \mathbb{1}_{\{I_{k-1}=i \wedge I_k=j\}}$$

To maximize the likelihood, we take the first derivative of the likelihood function with respect to π_{01} and π_{11} and set it to zero. We get the following:

$$\hat{\pi}_{01} = \frac{N_{00}}{N_{00} + N_{01}}$$

$$\hat{\pi}_{11} = \frac{N_{11}}{N_{10} + N_{11}}$$

To test the independence of violations, we are interested in finding out whether $\hat{\pi}_{01}$ is statistically different than $\hat{\pi}_{11}$. More specifically, we are interested in the case of $\hat{\pi}_{11}$ being significantly larger than $\hat{\pi}_{01}$ as this would imply that it is more likely to have 2 consecutive violations than having a violation following a non-violation. We define the independence null hypothesis:

$$H_0 : \pi_{01} = \pi_{11},$$

which is tested using the likelihood ratio test:

$$LR_{ind} = -2 \ln \left(\frac{(1 - \pi)^{N_{00} + N_{10}} \pi^{N_{01} N_{11}}}{(1 - \pi_{01})^{N_{00}} \pi_{01}^{N_{01}} (1 - \pi_{11})^{N_{10}} \pi_{11}^{N_{11}}} \right),$$

where $\pi = \frac{N_{01} + N_{11}}{N_{00} + N_{01} + N_{10} + N_{11}}$, and LR_{ind} follows a χ_1^2 with one degree of freedom (Roccioletti, 2015, p. 52).

Finally the Christoffersen's coverage test simultaneously tests for both properties: a good violation rate plus the independence of violations:

$$LR_{cc} = LR_{uc} + LR_{ind},$$

where $LR_{cc} \sim \chi_2^2$ with 2 degrees of freedom.

2.14.3 Pinball Loss

One common method to assess the accuracy of a quantile forecast is to use the *pinball loss* (or quantile loss). So, it is appropriate to use it to measure the accuracy of a VaR forecast model since $\text{VaR}_\alpha^{A,t}$ is the α -quantile of the return series of asset A on day t as described in Section 2.13.

The pinball loss of a VaR forecast for an asset A on day t is defined as:

$$L_\alpha(r_t^A, \text{VaR}_\alpha^{A,t}) = \begin{cases} (r_t^A - \text{VaR}_\alpha^{A,t})\alpha & r_t^A \geq \text{VaR}_\alpha^{A,t} \\ (\text{VaR}_\alpha^{A,t} - r_t^A)(1 - \alpha) & r_t^A < \text{VaR}_\alpha^{A,t} \end{cases} \quad (26)$$

where the first case corresponds to under-predicting, while the second case corresponds to over-predicting. It is worth noting that when α is small (which is usually the case for VaR models), over-predicting is penalized more than under-predicting since $(1 - \alpha)$ is large. This means that a violation with distance d from the VaR forecast is penalized more than a non-violation with the same distance d as shown in Figure 7.

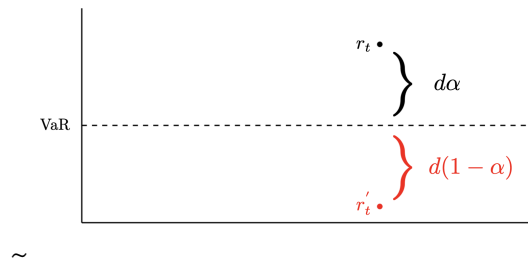


Figure 7: Shows a violation r'_t and a non-violation r_t , both having the same distance d from the VaR forecast. For small α , the r'_t contributes more than r_t to the pinball loss

The above gave a definition for the pinball loss of a single forecast. To define the overall pinball loss of the model, assume that we have fitted a VaR_α model on the return series $\{r_t^A\}, t = 1, \dots, T$ of an asset A . In addition, we have used the model to get the VaR forecasts of days $T + 1, \dots, T + m$, namely, we forecasted $\text{VaR}_\alpha^{A,T+1}, \dots, \text{VaR}_\alpha^{A,T+m}$, then from Equation (26), we can define the overall pinball loss of the model:

$$L_\alpha(r^A, \text{VaR}_\alpha^A) = \frac{1}{m} \sum_{t=T+1}^{T+m} L_\alpha(r_t^A, \text{VaR}_\alpha^{A,t}),$$

where it can be shown that a model with lower $L_\alpha(r^A, \text{VaR}_\alpha^A)$ implies a more accurate model (Koenker and Hallock, 2001).

3 A Vine Copula based ARMA-GARCH Model

Now we turn to our main problem: Given a portfolio Ω consisting of d assets A_1, \dots, A_d with weights w_1, \dots, w_d and log returns $r_t^{A_i}, t = 1, \dots, T, i = 1, \dots, d$, we are interested in developing a model that forecast the VaR of Ω .

The next two sections address this problem. The first section describes a model for forecasting the VaR on day $T + 1$, i.e. estimating $\text{VaR}_\alpha^{\Omega, T+1}$. On the other hand, the second section tackles backtesting the model by describing a rolling window based method for forecasting the a multi-period VaR rather than a single day forecast.

3.1 Single day VaR

The following steps are an extension of the method for estimating the VaR for a single asset which is described in Section 2.13.1. It generalizes it for a portfolio of assets by utilising a vine copula model.

1. Modeling the margins

First, we use each of the observed return series $\{r_t^{A_i}\}_{t=1}^T, i = 1, \dots, d$, to fit an ARMA(1, 1)-GARCH(1, 1) model ² as described in Sections 2.10. Hence, we get d models M_1, \dots, M_d as shown in Figure 8.

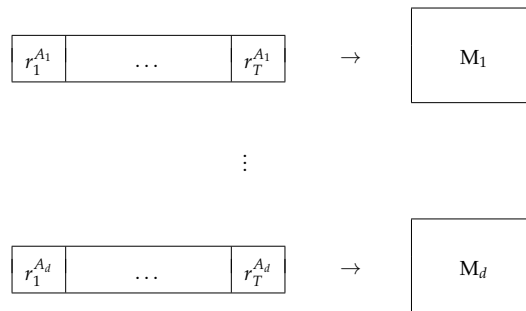


Figure 8: Using d different return series to fit d ARMA(1, 1)-GARCH(1, 1) models.

²We could have chosen here a general ARMA(p, q)-GARCH(m, s) model, and the subsequent steps would have been very similar to the ones described.

2. Forecasting the conditional mean and conditional variance

We use M_1, \dots, M_d to forecast:

1. The conditional mean on day $T + 1$ for each return series: $\hat{\mu}_{T+1}^{A_1}, \dots, \hat{\mu}_{T+1}^{A_d}$ as described in Sections 2.4 and 2.5.
2. The conditional volatility on day $T + 1$ for each return series: $\hat{\sigma}_{T+1}^{A_1}, \dots, \hat{\sigma}_{T+1}^{A_d}$ as described in Section 2.8.3,

as shown in Figure 9

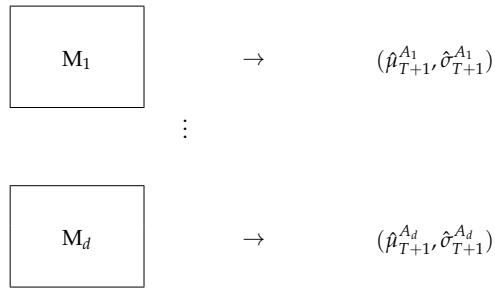


Figure 9: Using each of the d fitted ARMA-GARCH models to forecast the conditional mean and conditional volatility.

3. Calculating the standardized residuals

Next we calculate the standardized $\hat{z}_t^{A_i}$ for $t = 1, \dots, T, i = 1, \dots, d$ using:

$$\hat{z}_t^{A_i} = \frac{r_t^{A_i} - \hat{\mu}_t^{A_i}}{\hat{\sigma}_t^{A_i}}, \quad (27)$$

where $\hat{\mu}_t^{A_i}, \hat{\sigma}_t^{A_i}$ are estimated as described in Section 2.10.

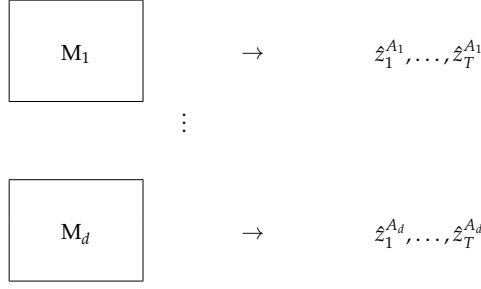


Figure 10: Extracting the standardized residuals for the fitted ARMA-GARCH models.

4. Fitting a vine copula model

We want to use the standardized residuals that we got in the previous step to capture the cross-sectional dependence among the d assets. As described in Section 2.10, we assume that the standardized innovations $\{z_t^{A_i}\}_{t=1}^T$ is an i.i.d sequence that follow some pre-selected distribution F_i with mean 0 and variance 1. Therefore, if our marginal models have been fitted correctly in Step 1, then for each $i = 1, \dots, d$, $\{\hat{z}_t^{A_i}\}_{t=1}^T$ is an approximate i.i.d sequence that follows F_i .

Now we want to utilize a regular vine copula model to capture the dependence among the standardized residuals. As described in Section 2.11.4, we first need to transform the residuals to the copula scale, in particular, we want to calculate $u_t^{A_i}$ as follows:

$$u_i^{A_i} = F_i(\hat{z}_t^{A_i}), \quad (28)$$

for $t = 1, \dots, T$ and $i = 1, \dots, d$. Then we can use the transformed residuals as pseudo multivariate observations to fit a regular vine copula model V as shown in Figure 13

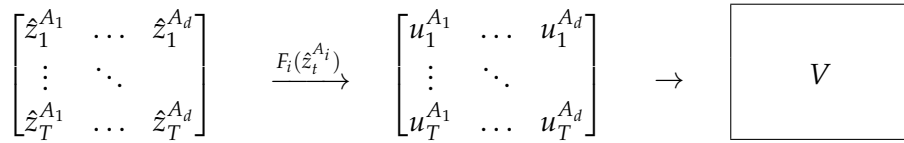


Figure 11: Shows the workflow of fitting the regular vine copula model V using the standardized residuals of the fitted ARMA-GARCH models.

5. Simulating returns on day $T + 1$

Next we use the fitted vine model V to simulate a multivariate sample of copula data with sample size k large enough (e.g. $k \geq 1000$). In particular, we simulate $\hat{\mathbf{u}}_i = (\hat{u}_i^{A_1}, \dots, \hat{u}_i^{A_d})^t, i = 1, \dots, k$. Then we transform the simulated data to the original scale:

$$z'_i{}^{A_j} = F_j^{-1}(\hat{u}_i^{A_j}), \quad (29)$$

for $i = 1, \dots, k$, and $j = 1, \dots, d$.

We use the simulated data on the original scale to get a sample of the returns on day $T + 1$ using:

$$\hat{r}_{T+1}^{A_j, i} = \hat{\mu}_{T+1}^{A_j} + \hat{\sigma}_{T+1}^{A_j} z'_i{}^{A_j}, \quad (30)$$

where $i = 1, \dots, k, j = 1, \dots, d$, and $\hat{\mu}_{T+1}^{A_j}, \hat{\sigma}_{T+1}^{A_j}$ were estimated in Step 2.

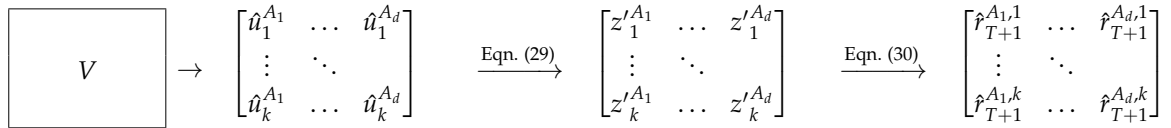


Figure 12: Shows the workflow of simulating a sample of returns on day $T + 1$ using the simulated data from the vine copula model V

6. Estimating the VaR on day $T + 1$

Now that we have an approximate sample of returns on day $T + 1$ for the different assets, we can get an approximate sample of the portfolio returns on day $T + 1$. In particular we want to calculate:

$$\hat{r}_{T+1}^{\Omega, i} = \sum_{j=1}^d w_j \hat{r}_{T+1}^{A_j, i},$$

where $i = 1, \dots, k$, and w_1, \dots, w_d are the weights of the assets as defined before.

Finally, the VaR estimate of the portfolio on day $T + 1$ at level α is the empirical α -quantile of the sample $\{\hat{r}_{T+1}^{\Omega, 1}, \dots, \hat{r}_{T+1}^{\Omega, k}\}$.

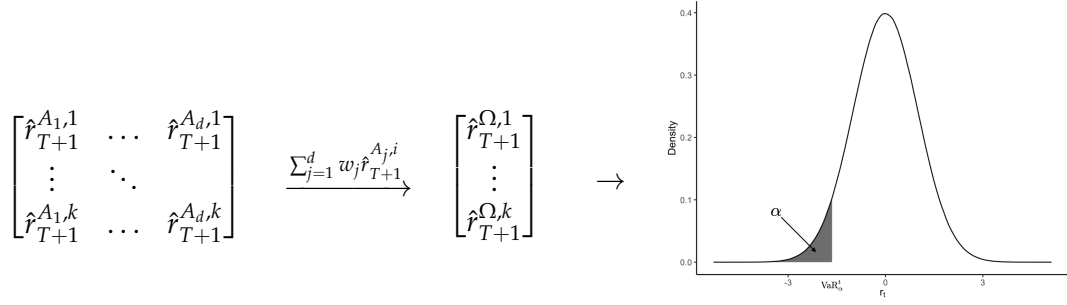


Figure 13: Shows the process of estimating $\text{VaR}_\alpha^{\Omega, T+1}$ using the approximate sample of returns of the d assets.

3.2 Backtesting the model

To tackle backtesting the model described in the previous section, we consider the same portfolio Ω ; however, we express the return series as $r_1^{A_1}, \dots, r_\Gamma^{A_1}, \dots, r_\Gamma^{A_d}, \dots, r_T^{A_1}, \dots, r_T^{A_d}, i = 1, \dots, d$ where $\Gamma < T \in \mathbb{N}$. We are interested in using a one day ahead forecast model to forecast the VaR for an interval rather than just a single day. In particular, we want to forecast $\text{VaR}_\alpha^{\Omega, \Gamma+1}, \dots, \text{VaR}_\alpha^{\Omega, T}$ which will be used to backtest the model as described in Section 2.14.

It is worth noting that the model described in the previous section can be used directly to forecast the VaR for the interval of interest as follows:

1. Fit the model on the observations $r_1^{A_1}, \dots, r_\Gamma^{A_1}, i = 1, \dots, d$ and use the model to forecast $\text{VaR}_\alpha^{\Omega, \Gamma+1}$.
2. Fit the model on the observations $r_2^{A_1}, \dots, r_{\Gamma+1}^{A_1}, i = 1, \dots, d$ and use the model to forecast $\text{VaR}_\alpha^{\Omega, \Gamma+2}$.
3. Repeat until we have the observations $r_{T-\Gamma}^{A_1}, \dots, r_{T-1}^{A_1}$ which we use to fit the model which is used to forecast $\text{VaR}_\alpha^{\Omega, T}$.

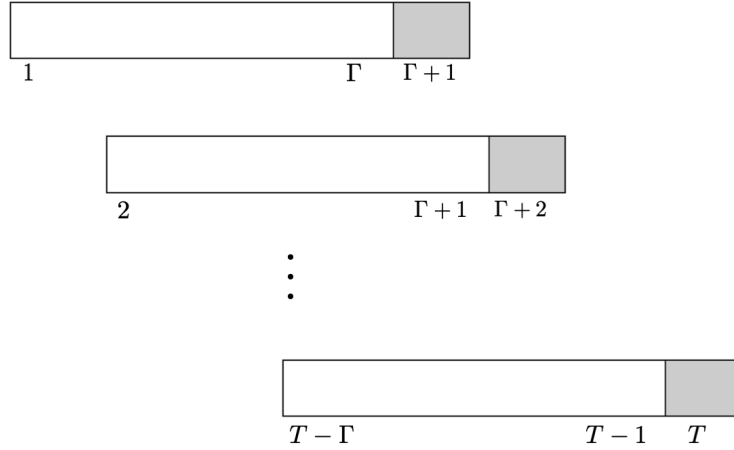


Figure 14: Fitting the model introduced in Section 3.1 ($T - \Gamma$) times.

As d gets larger, the above method becomes computationally very expensive. This is why we are interested in modifying it in such way that avoids refitting the marginal models as well as the vine copula model for each point forecast which will be described next.

A rolling window forecast method

First we define some parameters that will control the frequency of refitting the marginal models as well as the vine copula model.

- $\Gamma < T \in \mathbb{N}$: (mentioned above) used to define the number of observations used to for fitting the marginal models (i.e. the ARMA-GARCH models). Note that $T - \Gamma$ is the length of the forecasted sequence which will be used for backtesting.
- $\gamma \leq (T - \Gamma) \in \mathbb{N}$: used to define the length of the marginal forecast for each window. Equivalently, can be defined as the frequency of refitting the marginal models.
- $\Psi \leq \Gamma \in \mathbb{N}$: used to define the number of standardized residuals of the marginal models to use for fitting the vine copula model. In particular, $\hat{z}_{\Gamma-\Psi+1}^{A_i}, \dots, \hat{z}_{\Gamma}^{A_i}, i = 1, \dots, d$ will be used for fitting the vine copula model.
- $\kappa \leq \gamma \in \mathbb{N}$: used to define the length of the forecast sequence of the vine copula model. Equivalently, can be defined as the frequency of refitting the vine copula model.

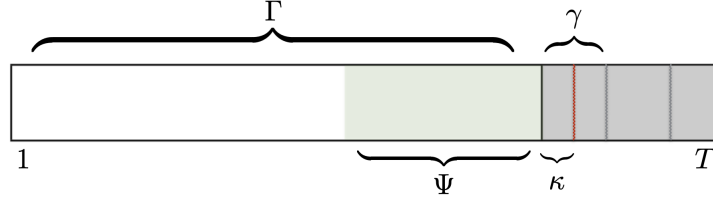


Figure 15: Example of the rolling window parameters.

In the following, we discuss the steps of the rolling window forecast method:

1. Marginal models

This step is similar to Step 1 and 2 of the method described in Section 3.1. We basically do the following:

1. Fit an ARMA(1, 1)-GARCH(1, 1) for each return series $r_1^{A_i}, \dots, r_\Gamma^{A_i}, i = 1, \dots, d$.
2. Use the fitted models to forecast the conditional means and variances for the interval $\Gamma + 1, \dots, \Gamma + \gamma$ using the following procedure:
 - (a) Forecast $\hat{\mu}_{\Gamma+1}^{A_1}, \dots, \hat{\mu}_{\Gamma+1}^{A_d}$ as described in Sections 2.4 and 2.5. More precisely, we use the following equation:

$$\hat{\mu}_{\Gamma+1}^{A_i} = \hat{\phi}_0 + \hat{\phi}_1 r_\Gamma^{A_i} - \theta \hat{\epsilon}_\Gamma,$$

for $i = 1, \dots, d$.

- (b) Forecast $\hat{\sigma}_{\Gamma+1}^{A_1}, \dots, \hat{\sigma}_{\Gamma+1}^{A_d}$ as described in Section 2.8.3. More precisely, we use the following equation:

$$\hat{\sigma}_{\Gamma+1}^{2A_i} = \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\epsilon}_\Gamma + \hat{\beta}_1 \hat{\sigma}_\Gamma^{2A_i},$$

for $i = 1, \dots, d$.

- (c) For $\Gamma + 2 \leq t \leq \Gamma + \gamma$, use the following:

$$\begin{aligned} \hat{\mu}_t^{A_i} &= \hat{\phi}_0 + \hat{\phi}_1 r_{t-1}^{A_i} - \theta \hat{\epsilon}_{t-1}^{A_i} \\ \hat{\sigma}_t^{2A_i} &= \hat{\alpha}_0 + \hat{\alpha}_1 \hat{\epsilon}_{t-1}^{A_i} + \hat{\beta}_1 \hat{\sigma}_{t-1}^{2A_i}, \end{aligned}$$

for $i = 1, \dots, d$. Here, $\hat{\epsilon}_t^{A_i}$ are the forecasted residuals which can be calculated as:

$$\hat{\epsilon}_t = r_t^{A_i} - \hat{\mu}_t^{A_i},$$

where $\hat{\mu}_t^{A_i}$ are the forecasted residuals that are calculated in Step (a).

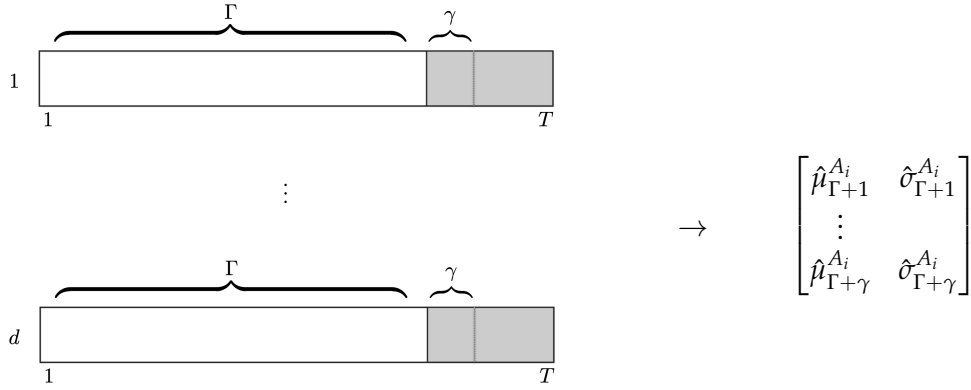


Figure 16: Fitting the marginal models and forecasting the conditional means and variances.

2. Fitting a vine copula model

In this step we use the fitted marginal models from the previous step to fit a vine copula model. As we did in Steps 3 of Section 3.1, we first calculate the standardized residuals of the fitted ARMA-GARCH models. In particular we calculate $\hat{z}_{\Gamma-\Psi+1}^{A_i}, \dots, \hat{z}_{\Gamma}^{A_i}, i = 1, \dots, d$ which are used to fit a regular vine copula model.

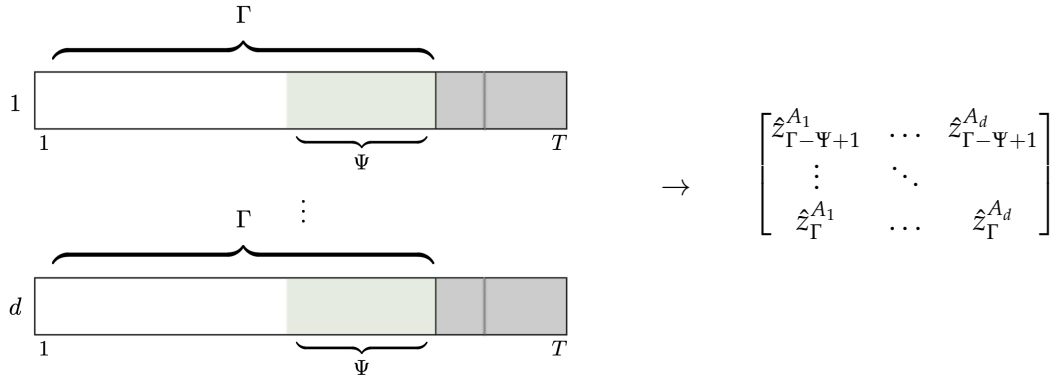


Figure 17: Calculating the standardized residuals from the fitted marginal models.

Now similar to what we did in Step 4 of Section 3.1, we use the standardized residuals to fit a regular vine copula model V . Once, again we first transform the residuals to the copula scale using Equation (28).

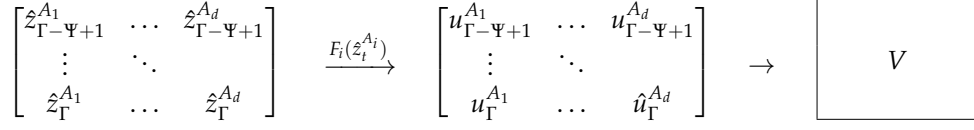


Figure 18: Shows the workflow of fitting the regular vine copula model V using the standardized residuals of the fitted marginal models.

3. Simulating returns

Next we simulate a sample of the returns following a similar procedure to that of Step 5 of Section 3.1; however, here we are interested in simulating a sample for each day in the set $\{\Gamma + 1, \dots, \Gamma + \kappa\}$ rather than just one day which is done as follows:

1. Simulate a random sample $\hat{\mathbf{u}}_i = (\hat{u}_i^{A_1}, \dots, \hat{u}_i^{A_d})^t, i = 1, \dots, k$ for some constant k large enough (e.g. $k \geq 1000$).
2. Transform the simulated data to the original scale using Equation (29):

$$z_i'^{A_j} = F_j^{-1}(\hat{u}_i^{A_j}),$$

for $i = 1, \dots, k$, and $j = 1, \dots, d$.

3. Use $z_i'^{A_j}$ to get a sample of returns on days $\{\Gamma + 1, \dots, \Gamma + \kappa\}$ using Equation (30):

$$\hat{r}_t^{A_j, i} = \hat{\mu}_t^{A_j} + \hat{\sigma}_t^{A_j} z_i'^{A_j},$$

for $i = 1, \dots, k, j = 1, \dots, d, t = \Gamma + 1, \dots, \Gamma + \kappa$.

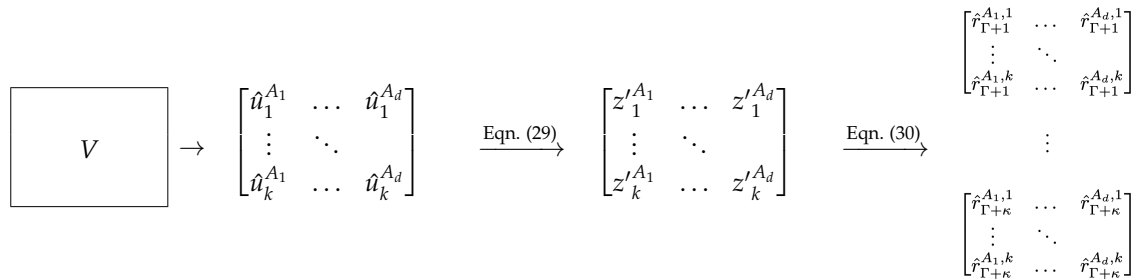


Figure 19: Shows the workflow of simulating a sample of returns for each day in the set $\{\Gamma + 1, \dots, \Gamma + \kappa\}$ using the simulated data from the vine copula model V

4. Estimating the VaR

Now that we have an approximate sample of returns for each day in the set $\{\Gamma + 1, \dots, \Gamma + \kappa\}$, we can get an approximate sample of the portfolio returns on those days which can be used to estimate the VaR similar to how we did it in Step 6 of Section 3.1 as shown in Figure 20.

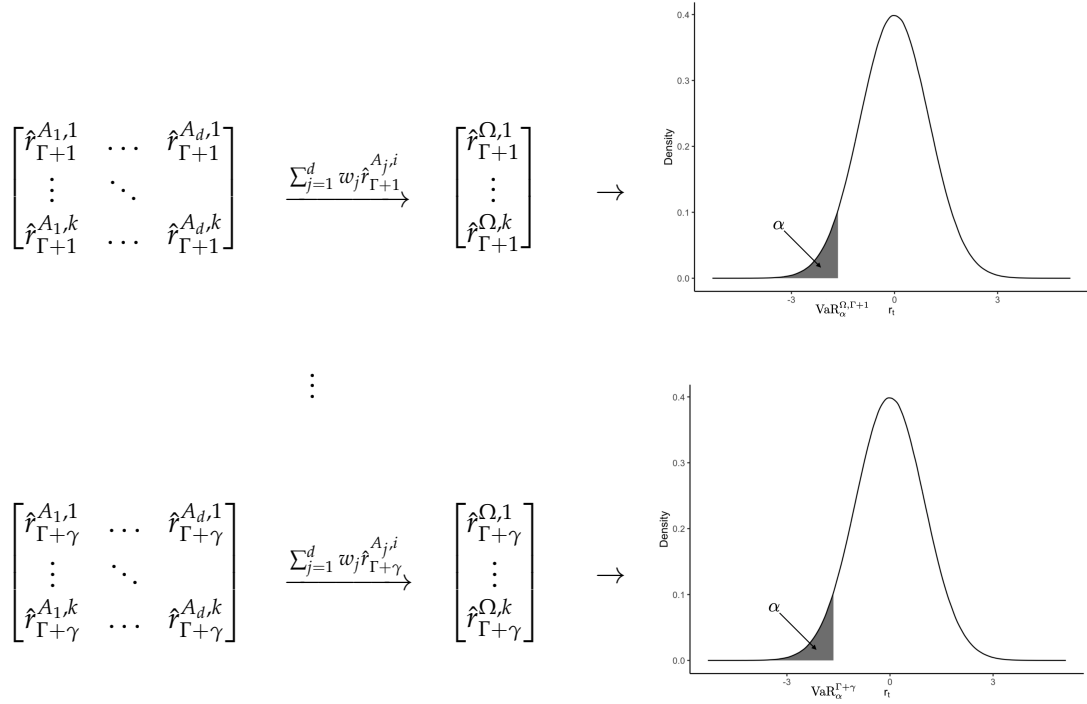


Figure 20: Shows the process of estimating $\text{VaR}_{\alpha}^{\Omega, \Gamma+1}, \dots, \text{VaR}_{\alpha}^{\Omega, \Gamma+\gamma}$ using the approximate sample of returns of the d assets.

5. Rolling the vine window

In Step 4, we estimated the $\text{VaR}_{\alpha}^{\Omega, \Gamma+1}, \dots, \text{VaR}_{\alpha}^{\Omega, \Gamma+\kappa}$. Now we are interested in refitting the vine copula model to forecast the next κ values. To do that we roll the *vine window* by κ as shown in Figure 21.

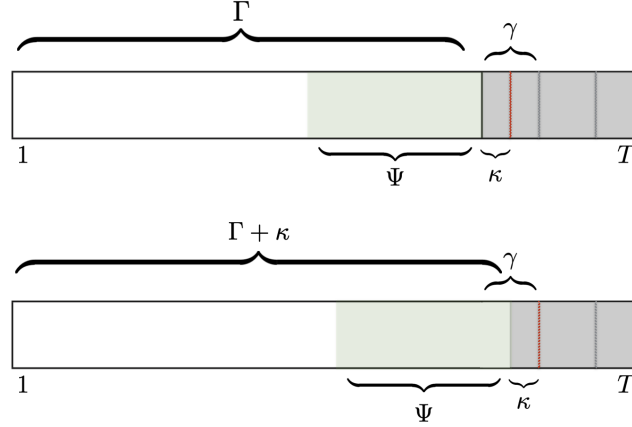


Figure 21: Rolling the vine window.

Therefore, to forecast $\text{VaR}_\alpha^{\Omega, \Gamma + \kappa + 1}, \dots, \text{VaR}_\alpha^{\Omega, \Gamma + 2\kappa}$, we repeat Steps 2, 3, and 4 using the shifted standardized residuals $\hat{z}_{\Gamma - \Psi + \kappa + 1}^{A_i}, \dots, \hat{z}_{\Gamma + \kappa}^{A_i}, i = 1, \dots, d$. However, it is worth mentioning that we distinguish 2 cases while calculating the standardized residuals:

- Case 1: $t \leq \Gamma$, then $\hat{z}_t^{A_i}$ is calculated as before using Equation (27).
- Case 2: $t > \gamma$, then $\hat{z}_t^{A_i}$ is in fact a forecasted residuals which is calculated using:

$$\hat{z}_t^{A_i} = \frac{r_t^{A_i} - \hat{\mu}_t^{A_i}}{\hat{\sigma}_t^{A_i}},$$

for $i = 1, \dots, d$. Where $\hat{\mu}_t^{A_i}, \hat{\sigma}_t^{A_i}$ refer to the forecasted mean and the forecasted volatility which are calculated in Step 1 (a) and (b).

Next, we keep repeating this step by rolling the vine window κ steps at a time until we have γ VaR forecasts. Namely, $\text{VaR}_\alpha^{\Omega, \Gamma + 1}, \dots, \text{VaR}_\alpha^{\Omega, \Gamma + \gamma}$ which conclude the first *GARCH window*.

6. Rolling the GARCH window

Now that we have forecasted the first γ values, namely, $\text{VaR}_\alpha^{\Omega, \Gamma + 1}, \dots, \text{VaR}_\alpha^{\Omega, \Gamma + \gamma}$, we are interested in repeating the steps above to forecast the next γ values. In particular, we are interested in repeating Steps 1-5 by rolling the GARCH window γ observations. More precisely, we use the observations $r_{\gamma+1}^{A_i}, \dots, r_{\gamma+\Gamma}^{A_i}, i = 1, \dots, d$ to fit the marginal models, and then we proceed following the steps described above to forecast $\text{VaR}_\alpha^{\Omega, \Gamma + \gamma + 1}, \dots, \text{VaR}_\alpha^{\Omega, \Gamma + 2\gamma}$.

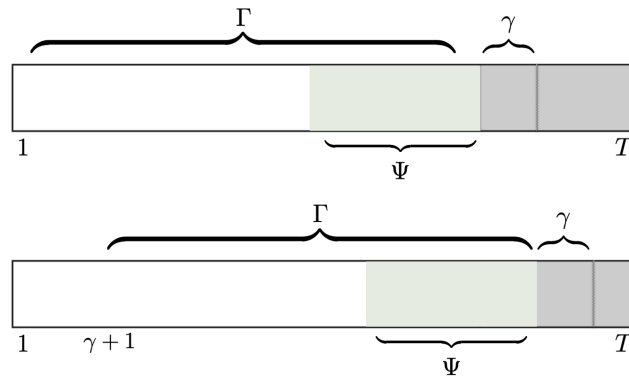


Figure 22: Rolling the GARCH window.

Next, we keep repeating this step by rolling the GARCH window γ steps at a time until we have forecasted all the values of interest. Namely, until we have estimated $\text{VaR}_\alpha^{\Omega, \Gamma+1}, \dots, \text{VaR}_\alpha^{\Omega, T}$ which then can be used to backtest the model using the different methods described in Section 2.14.

4 R-package

In the following, we will show an example of the rolling forecast method described above using the R package `VaRVine` which was developed for this purpose.

Data

Our portfolio consists of 30 assets which are equally weighted from different sectors as shown in Table 1. The return series are between 2012-01-11 and 2016-01-01. More precisely, we have exactly 1000 observations, and we are interested in estimating the VaR for the last 250 days for levels $\alpha = (0.01, 0.05)$.

Asset	Sector	Asset	Sector	Asset	Sector
AAPL	IT	APD	Materials	GS	Financials
ABT	Health Care	ATVI	Communication Services	J	Industrials
ADBE	IT	BA	Industrials	LNT	Utilities
AES	Utilities	BLK	Financials	LVS	Consumer Discretionary
AFL	Financials	CE	Materials	MMM	Industrials
AGCO	Industrials	CMI	Industrials	MOS	Materials
AKAM	IT	CVS	Health Care	NOV	Energy
ALL	Financials	GE	Industrials	NVDA	IT
AMD	IT	GME	Retail	OMI	Health Care
AMZN	Consumer Discretionary	GOOG	IT	TEX	Manufacturing

Table 1: List of the stock tickers used in the example with their corresponding sectors.

Model settings for testing period Jan 2015-Jan 2016

We use the following settings for our model:

- $\Gamma = 750$ (marginal models training period)
- $\Psi = 250$ (vine model training period)
- $\gamma = 50$ (marginal models refitting frequency)
- $\kappa = 25$ (vine model refitting frequency)

This implies that we have 5 GARCH windows in total with 2 vine windows in each GARCH window ($\frac{50}{25} = 2$) and therefore, we have 10 vine windows in total.

Code

```
library(VaRVine)
head(data)
```

.	AAPL	ABT	ADBE	...
2012-01-11	-0.00163	-0.01237	0.00410	...
2012-01-12	-0.00275	-0.00470	-0.00307	...
2012-01-13	-0.00376	0.00488	-0.00893	...
2012-01-17	0.01158	0.00504	0.01303	...
2012-01-18	0.01033	-0.00630	0.01823	...
2012-01-19	-0.00317	0.00126	0.01164	...

```
# Here  $\Gamma$  is 'train.size', and  $\gamma$  is 'refit.every'.
garch.setting <- garch_settings(train.size = 750, refit.every = 50)
```

```
# Here  $\Psi$  is 'train.size', and  $\kappa$  is 'refit.every'.
vine.setting <- vine_settings(train.size = 250, refit.every = 25,
```

```

        family.set = 'all')

# This method performs the rolling window forecast.
roll <- garchvineroll(data, garch.setting, vine.setting)

Fitting garch models: AAPL ABT ADBE AES AFL AGCO AKAM
ALL AMD AMZN APD ATVI BA BLK CE CMI CVS GE GME GOOG
GS J LNT LVS MMM MOS NOV NVDA OMI TEX

Window 1 :...
Window 2 :...
Window 3 :...
Window 4 :...
Window 5 :...

.

# Computation time using MacBook pro with
# Intel Core i7 2.6 GHz processor and 16 GB of RAM.
roll@time.taken
>> 29.33719 mins

# Accessing the VaR forecasts.
roll@VaR.forecast

```

.	date	alpha_0.01	alpha_0.05	realized
1	01/06/2015 00:00:00	-0.02382	-0.01486	-0.01386
2	01/07/2015 00:00:00	-0.02430	-0.01513	0.00519
3	01/08/2015 00:00:00	-0.02395	-0.01498	0.01811
4	01/09/2015 00:00:00	-0.02451	-0.01541	-0.01370
5	01/12/2015 00:00:00	-0.02419	-0.01501	-0.00715
6	01/13/2015 00:00:00	-0.02420	-0.01519	-0.00337

```

# Plotting the VaR forecasts and highlighting
# the violations.
VaRVine::plot(roll, alpha=0.01)
VaRVine::plot(roll, alpha=0.05)

# Unconditional and conditional backtests for  $\alpha = 0.05$ 
VaRVine::backtest(roll, alpha = 0.05)

```

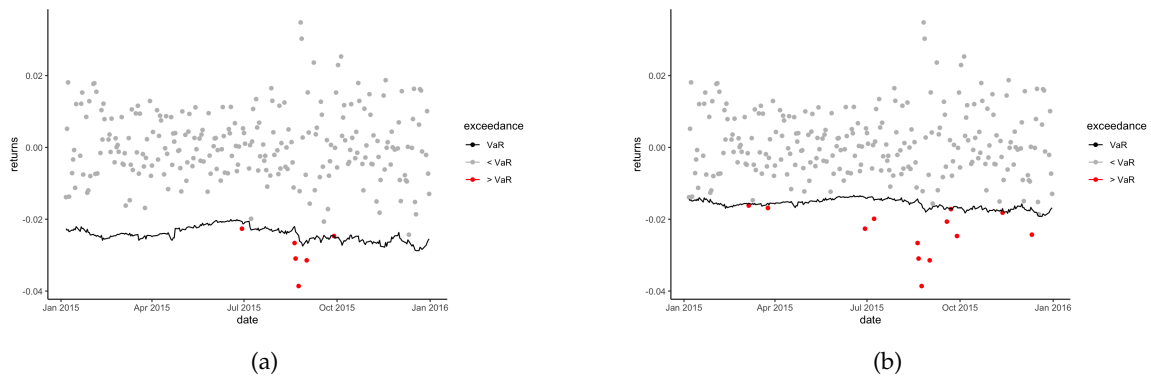
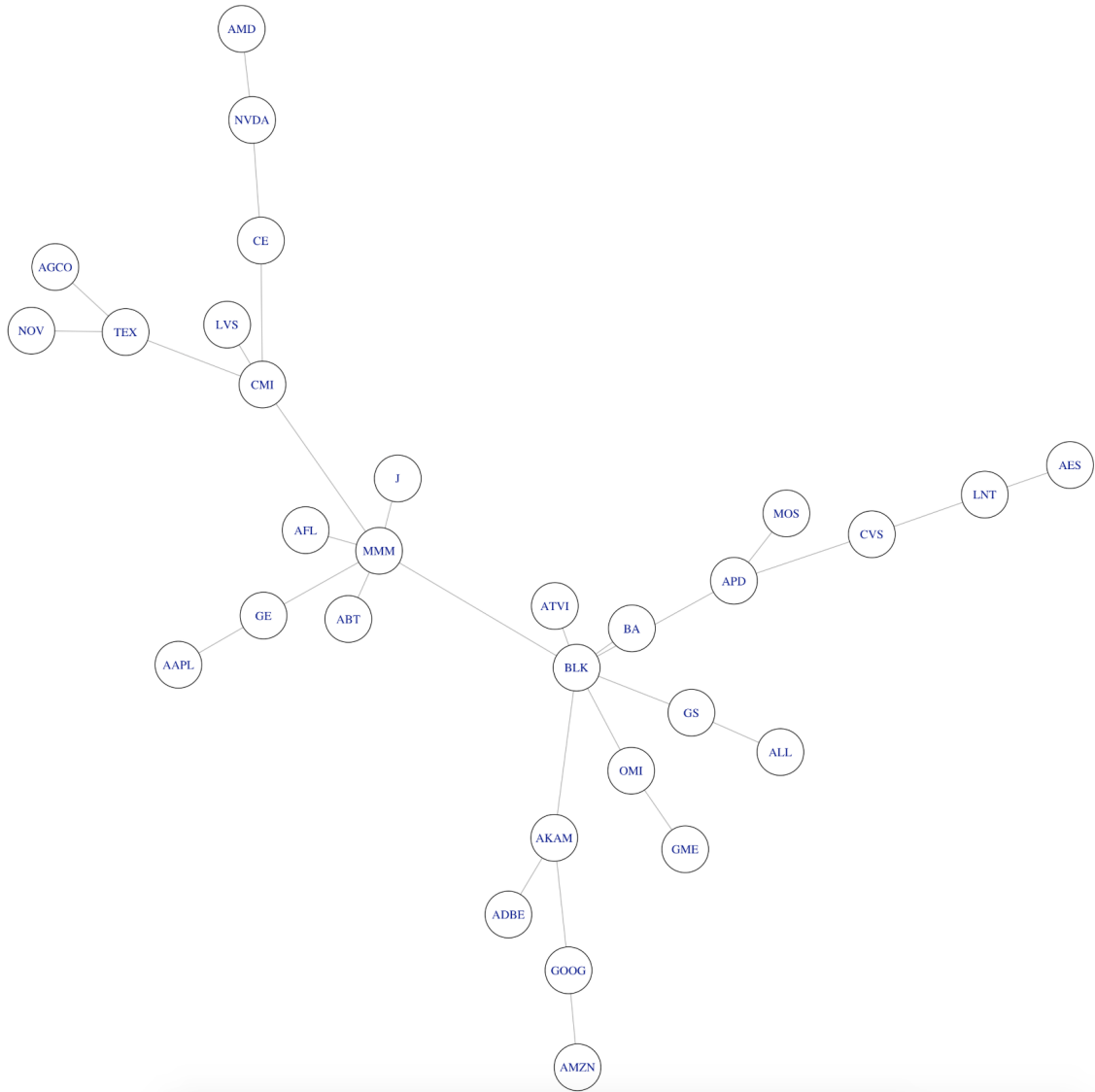


Figure 23: VaR plots for the testing period Jan 2015-Jan 2016. Left plot corresponds to $\alpha = 0.01$, while right plot corresponds to $\alpha = 0.05$. The violations are highlighted in red.

Unconditional Test	Conditional Test
- expected.exceed: 12.	- expected.exceed: 12.
- actual.exceed: 13.	- actual.exceed: 13.
- uc.H0: Correct Exceedances.	- cc.H0: Correct Exceedances & Independent.
- uc.LRstat: 0.02079.	- cc.LRstat: 2.0029.
- uc.critical: 3.84146.	- cc.critical: 5.9914.
- uc.LRp: 0.88535.	- cc.LRp: 0.3673.
- uc.Decision: Fail to Reject H0.	- cc.Decision: Fail to Reject H0.

Evolution of the vine model over time

As mentioned above, we have 10 vine windows in total where each vine window has its own vine copula model. As described in Section 2.12.1, the structure of a vine copula model is represented by a sequence of trees where each edge corresponding to a bivariate copula. Figures 24, 25, and 26 show how the vine structure changes with time. In particular it shows plots of the vine structure of the first tree for the first, fifth, and tenth vine windows.

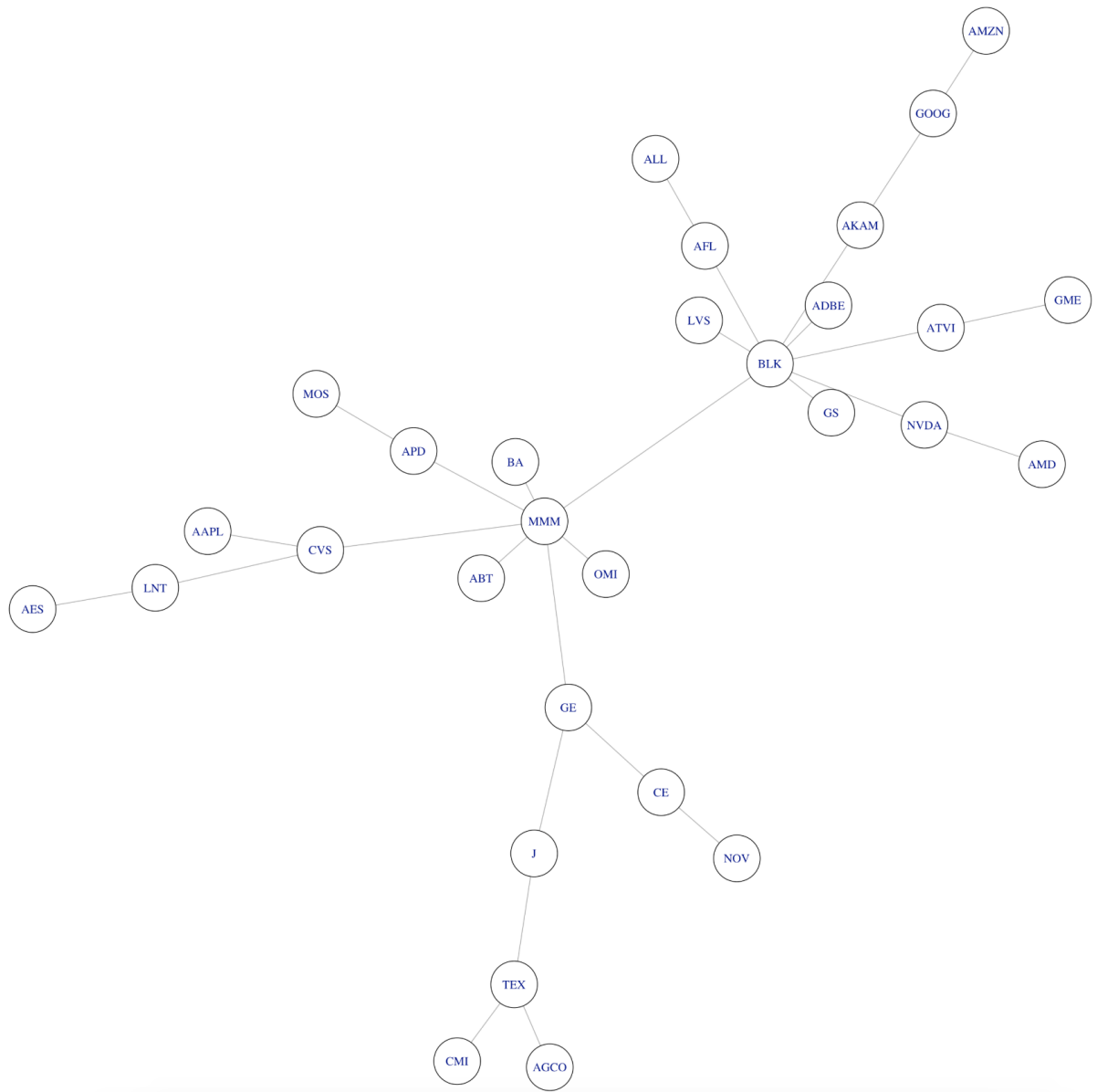


(a)

Edge	Copula family ($ \tau $)	Edge	Copula family ($ \tau $)
{AMD, NVDA}	gumbel(0.3)	{NVDA, CE}	frank(0.36)
{CE, CMI}	gaussian(0.43)	{CMI, LVS }	t(0.32)
{CMI, TEX}	bb1(0.48)	{TEX, AGCO }	frank(0.39)
{TEX, NOV}	gumbel(0.37)	{CMI, MMM }	gumbel(0.46)
{MMM, GE}	gaussian(0.47)	{GE, AAPL }	bb8(0.27)
{MMM, AFL}	bb1(0.42)	{MMM, ABT }	t(0.36)
{MMM, J}	t(0.4)	{MMM, BLK }	gumbel(0.47)
{BLK, ATVI}	bb8(0.34)	{BLK, APD }	gumbel(0.43)
{APD, MOS}	gumbel(0.34)	{CVS, APD }	gumbel(0.34)
{CVS, LNT}	gumbel(0.29)	{LNT, AES }	bb1(0.37)
{BLK, BA}	gaussian(0.38)	{BLK, OMI }	bb8(0.34)
{OMI, GME}	bb8(0.2)	{BLK, AKAM }	t(0.41)
{AKAM, ADBE}	bb8(0.41)	{AKAM, GOOG }	gumbel(0.36)
{GOOG, AMZN}	t(0.43)	{BLK, GS }	t(0.48)
{GS, ALL}	gumbel(0.38)		

(b)

Figure 24: (a) Shows the vine structure first tree of the first vine window for testing period Jan 2015-Jan 2016. (b) Shows the copula family and the magnitude of Kendall's τ for the bivariate copulas which correspond to the edges of the vine tree structure.

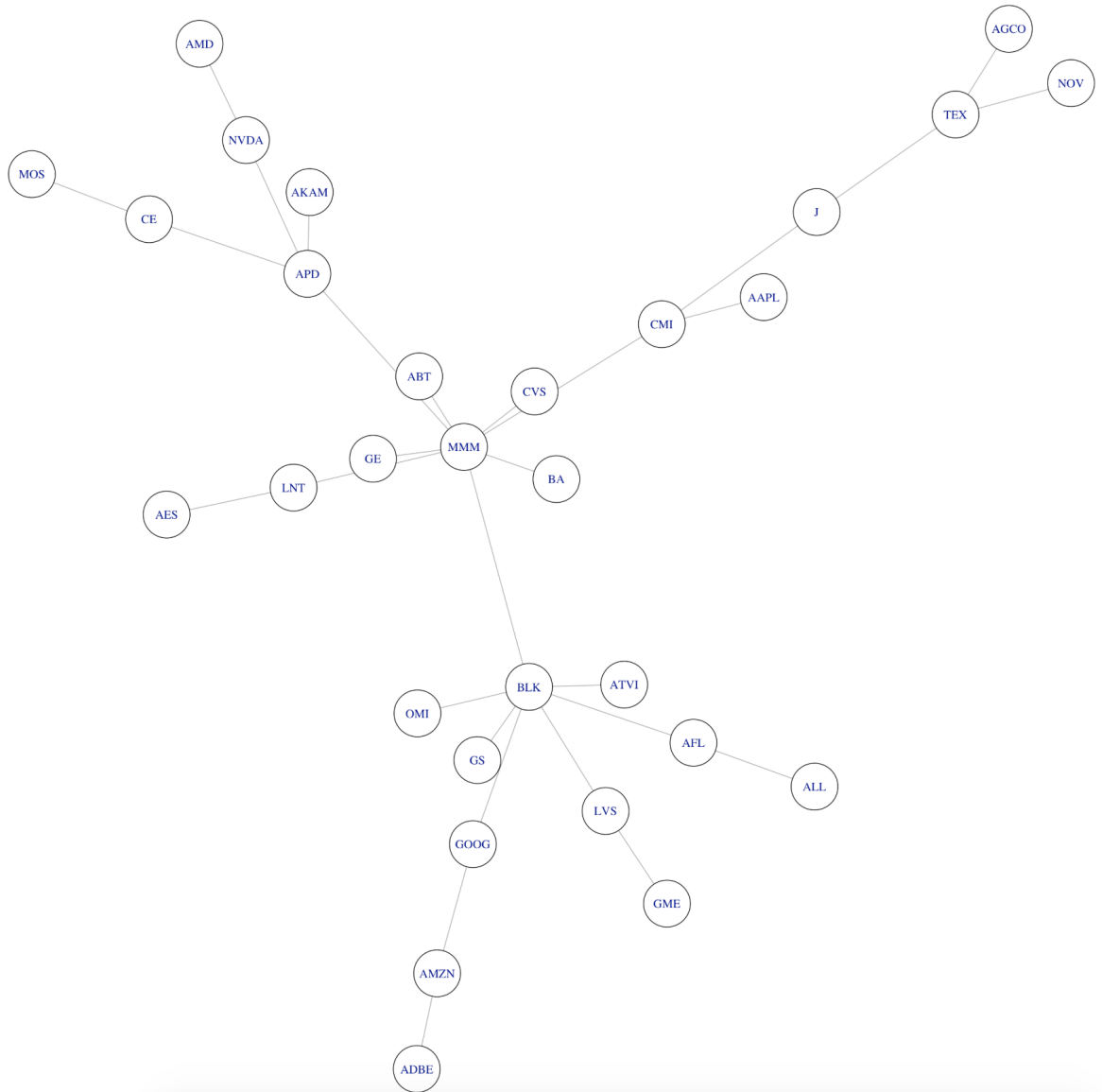


(a)

Edge	Copula family ($ \tau $)	Edge	Copula family ($ \tau $)
{AGCO, TEX}	frank(0.42)	{TEX, CMI}	t(0.45)
{J, TEX}	frank(0.44)	{J, GE}	frank(0.44)
{GE, CE}	gumbel(0.35)	{CE, NOV}	t(0.36)
{GE, MMM}	gumbel(0.41)	{MMM, APD}	bb8(0.46)
{APD, MOS}	gumbel(0.34)	{MMM, ABT}	frank(0.46)
{MMM, BA}	frank(0.46)	{MMM, OMI}	frank(0.39)
{MMM, CVS}	frank(0.43)	{CVS, LNT}	gumbel(0.31)
{LNT, AES}	t(0.41)	{CVS, AAPL}	t(0.32)
{MMM, BKJ}	bb8(0.51)	{BLK, LVS}	frank(0.3)
{BLK, AFL}	t(0.49)	{AFL, ALL}	t(0.45)
{BLK, AKAM}	bb8(0.44)	{AKAM, GOOG}	t(0.39)
{GOOG, AMZN}	gaussian(0.45)	{BLK, ADBE}	t(0.39)
{BLK, NVDA}	bb8(0.38)	{NVDA, AMD}	t(0.26)
{BLK, ATVI}	bb8(0.4)	{ATVI, GME}	gumbel(0.24)
{BLK, GS}	gumbel(0.51)		

(b)

Figure 25: (a) Shows the vine structure first tree of the fifth vine window for testing period Jan 2015-Jan 2016. (b) Shows the copula family and the magnitude of Kendall's τ for the bivariate copulas which correspond to the edges of the vine tree structure.



(a)

Edge	Copula family ($ \tau $)	Edge	Copula family ($ \tau $)
{AGCO, TEX}	frank(0.42)	{TEX, NOV}	frank(0.39)
{TEX, J}	bb8(0.46)	{J, CMI}	t(0.42)
{CMI, AAPL}	gaussian(0.37)	{MMM, CMI}	t(0.41)
{MMM, APD}	bb8(0.46)	{APD, CE}	bb8(0.38)
{CE, MOS}	bb8(0.38)	{APD, AKAM}	t(0.41)
{APD, NVDA}	t(0.35)	{NVDA, AMD}	t(0.27)
{MMM, GE}	gumbel(0.42)	{MMM, BA}	frank(0.5)
{MMM, ABT}	frank(0.49)	{MMM, BLK}	bb8(0.51)
{MMM, CVS}	frank(0.46)	{MMM, LNT}	frank(0.34)
{LNT, AES}	frank(0.34)	{BLK, OMI}	t(0.4)
{BLK, AFL}	t(0.51)	{AFL, ALL}	frank(0.52)
{BLK, GS}	gumbel(0.5)	{BLK, LVS}	frank(0.36)
{LVS, GME}	frank(0.31)	{BLK, GOOG}	gumbel(0.37)
{GOOG, AMZN}	t(0.5)	{AMZN, ADBE}	gumbel(0.34)
{BLK, ATVI}	bb8(0.39)		

(b)

Figure 26: (a) Shows the vine structure first tree of the tenth (final) vine window for testing period Jan 2015-Jan 2016. (b) Shows the copula family and the magnitude of Kendall's τ for the bivariate copulas which correspond to the edges of the vine tree structure.

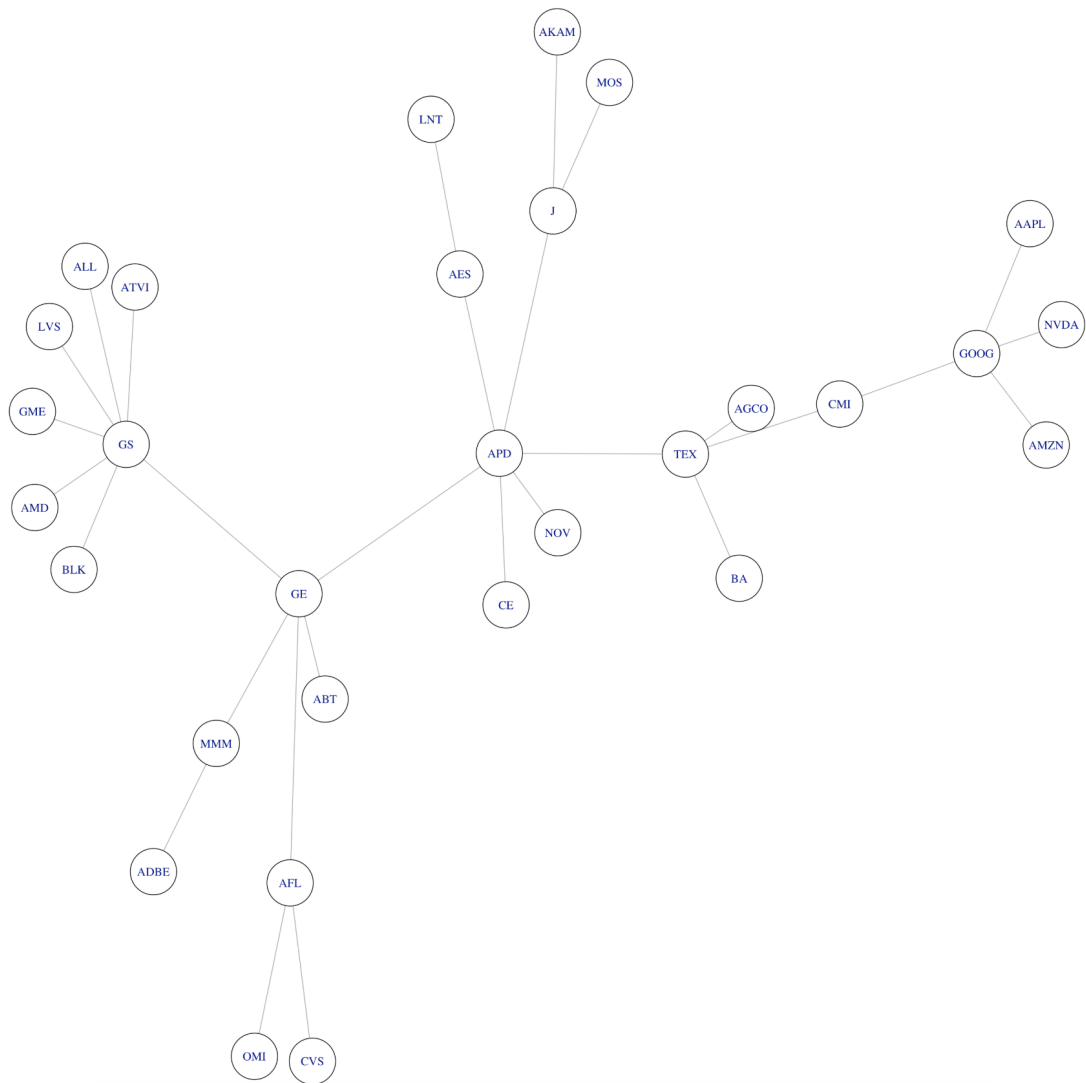
Comparing two models in the crises period July 2008-July 2009

During normal market conditions, changing the model settings has a little effect on the accuracy of the model. However, during crises times, we have found that changing some of the parameters has a considerable effect on the model performance. As an example, consider the the same list of stocks used above, but this time, we forecast the VaR for the financial crises of 2008. Namely, we forecast the VaR between July of 2008 and July of 2009. First, we fit a base model with settings similar to the ones used in the previous example. Namely:

- $\Gamma = 750$ (marginal models training period)
- $\Psi = 250$ (vine model training period)
- $\gamma = 50$ (marginal models refitting frequency)
- $\kappa = 25$ (vine model refitting frequency)

Once again, we have 5 GARCH windows, and 10 vine windows where each GARCH window contains 2 vine windows.

Figures 27, 28, 29 show the evolution of the vine structure with time of our base model similar to what we did in the previous example.

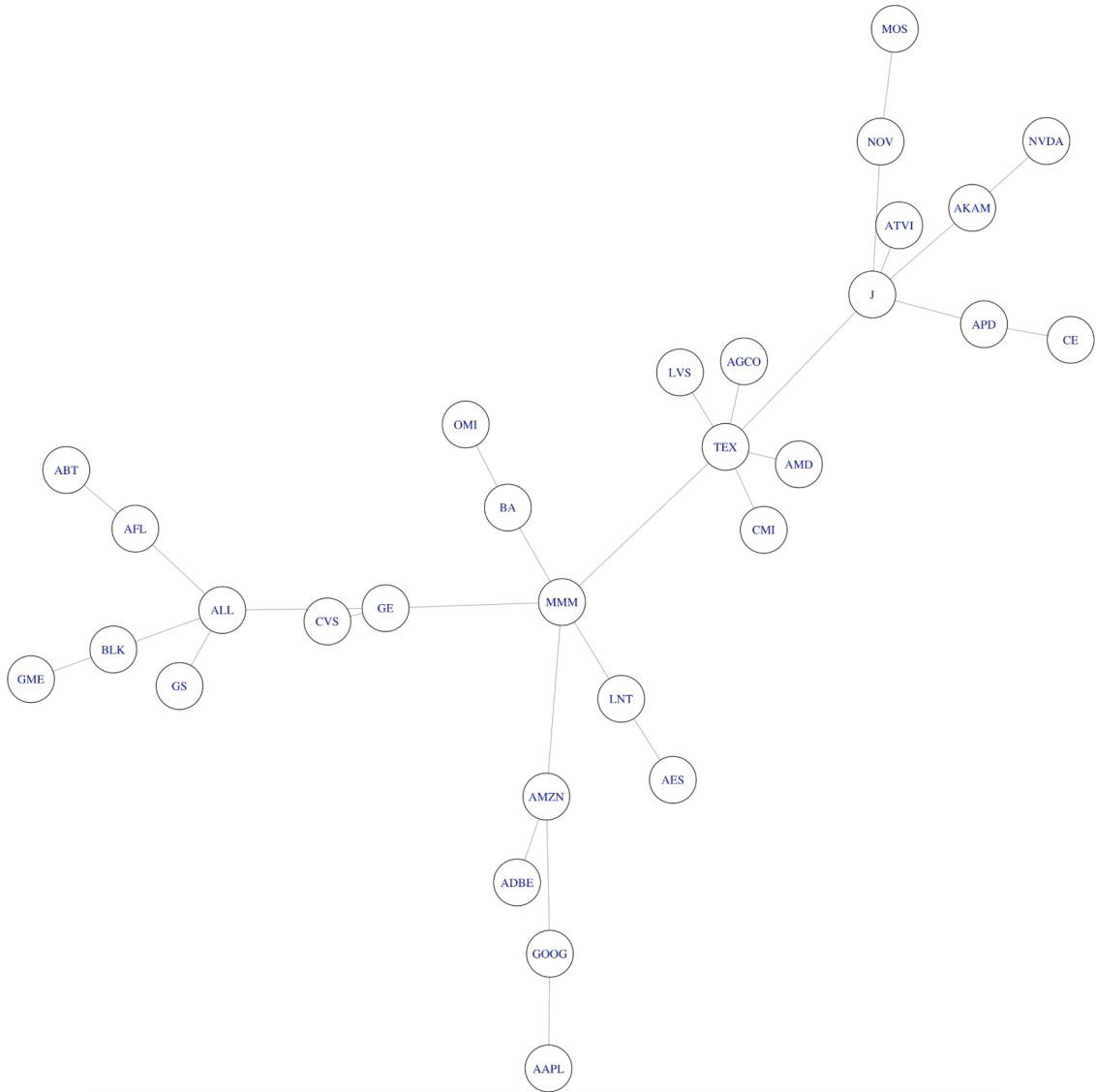


(a)

Edge	Copula family ($ \tau $)	Edge	Copula family ($ \tau $)
{NVDA, GOOG}	gumbel(0.31)	{GOOG, AMZN}	bb1(0.38)
{GOOG, AAPL}	t(0.43)	{GOOG, CMI}	t(0.33)
{CMI, TEX}	bb1(0.45)	{TEX, AGCO}	gumbel(0.41)
{TEX, BA}	bb8(0.3)	{TEX, APD}	t(0.43)
{APD, NOV}	gaussian(0.38)	{APD, J}	t(0.44)
{J, MOS}	gaussian(0.36)	{J, AKAM}	bb8(0.35)
{APD, CE}	t(0.36)	{APD, AES}	bb8(0.37)
{AES, LNT}	gumbel(0.35)	{APD, GE}	gaussian(0.43)
{GE, MMM}	t(0.48)	{MMM, ADBE}	gumbel(0.28)
{GE, AFL}	t(0.43)	{AFL, CVS}	frank(0.32)
{AFL, OMI}	gumbel(0.27)	{GE, ABT}	gaussian(0.33)
{GE, GS}	bb1(0.4)	{GS, LVS}	frank(0.28)
{GS, ATVI}	frank(0.26)	{GS, AMD}	gumbel(0.27)
{GS, GME}	gumbel(0.3)	{GS, ALL}	bb1(0.41)
{GS, BLK}	bb1(0.38)		

(b)

Figure 27: (a) Shows the vine structure first tree of the first vine window for testing period July 2008-July 2009. (b) Shows the copula family and the magnitude of Kendall's τ for the bivariate copulas which correspond to the edges of the vine tree structure.

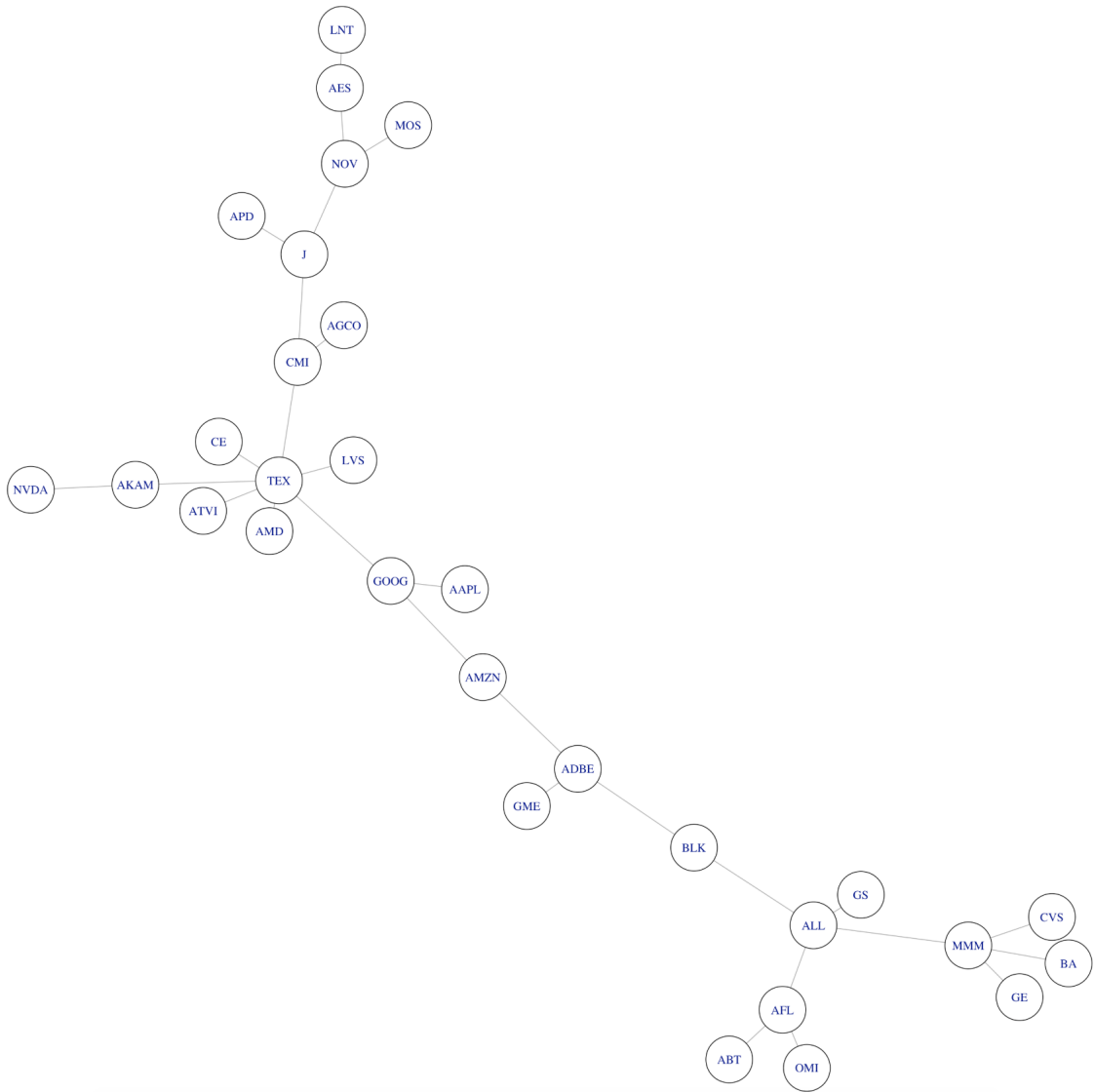


(a)

Edge	Copula family ($ \tau $)	Edge	Copula family ($ \tau $)
{AAPL, GOOG}	t(0.45)	{GOOG, AMZN}	t(0.42)
{AMZN, ADBE}	bb7(0.36)	{AMZN, MMM}	t(0.36)
{MMM, BA}	bb1(0.35)	{BA, OMI}	frank(0.29)
{MMM, GE}	t(0.53)	{GE, CVS}	gaussian(0.35)
{GE, ALL}	frank(0.45)	{ALL, GS}	bb1(0.43)
{ALL, BLK}	bb1(0.4)	{BLK, GME}	bb1(0.31)
{ALL, AFL}	gaussian(0.46)	{AFL, ABT}	bb8(0.32)
{MMM, LNT}	bb1(0.33)	{LNT, AES}	gumbel(0.37)
{MMM, TEX}	bb1(0.39)	{TEX, LVS}	bb8(0.3)
{TEX, CMI}	t(0.47)	{TEX, AMD}	bb1(0.28)
{TEX, AGCO}	bb8(0.45)	{TEX, J}	gumbel(0.44)
{J, ATVI}	bb1(0.27)	{J, AKAM}	t(0.41)
{AKAM, NVDA}	gumbel(0.38)	{J, NOV}	gumbel(0.43)
{NOV, MOS}	t(0.42)	{J, APD}	t(0.47)
{APD, CE}	bb1(0.4)		

(b)

Figure 28: (a) Shows the vine structure first tree of the fifth vine window for testing period July 2008-July 2009. (b) Shows the copula family and the magnitude of Kendall's τ for the bivariate copulas which correspond to the edges of the vine tree structure.



(a)

Edge	Copula family ($ \tau $)	Edge	Copula family ($ \tau $)
{OMI, AFL}	gumbel(0.28)	{AFL, ABT}	frank(0.32)
{AFL, ALL}	t(0.48)	{ALL, GS}	t(0.36)
{ALL, MMM}	t(0.44)	{MMM, GE}	t(0.48)
{MMM, BA}	t(0.45)	{MMM, CVS}	bb1(0.36)
{ALL, BLK}	t(0.47)	{BLK, ADBE}	t(0.44)
{ADBE, GME}	t(0.38)	{ADBE, AMZN}	frank(0.45)
{AMZN, GOOG}	t(0.47)	{GOOG, AAPL}	t(0.5)
{GOOG, TEX}	frank(0.49)	{TEX, LVS}	bb8(0.38)
{TEX, AMD}	bb8(0.36)	{TEX, CMI}	bb8(0.57)
{CMI, AGCO}	t(0.51)	{CMI, J}	t(0.51)
{J, APD}	gaussian(0.48)	{J, NOV}	gaussian(0.51)
{NOV, MOS}	bb8(0.48)	{NOV, AES}	gumbel(0.37)
{AES, LNT}	gumbel(0.37)	{TEX, CE}	t(0.46)
{TEX, ATVI}	frank(0.39)	{TEX, AKAM}	bb8(0.48)
{AKAM, NVDA}	frank(0.5)		

(b)

Figure 29: (a) Shows the vine structure first tree of the tenth (final) vine window for testing period July 2008-July 2009. (b) Shows the copula family and the magnitude of Kendall's τ for the bivariate copulas which correspond to the edges of the vine tree structure.

To study the effect of the vine training period Ψ , we have fitted another model which has the same settings as our base model described above; however, in this model we used a vine training period $\Psi = 750$ instead of 250. Plotting the pinball loss shows that the base model outperforms this model as displayed in Figure 30.

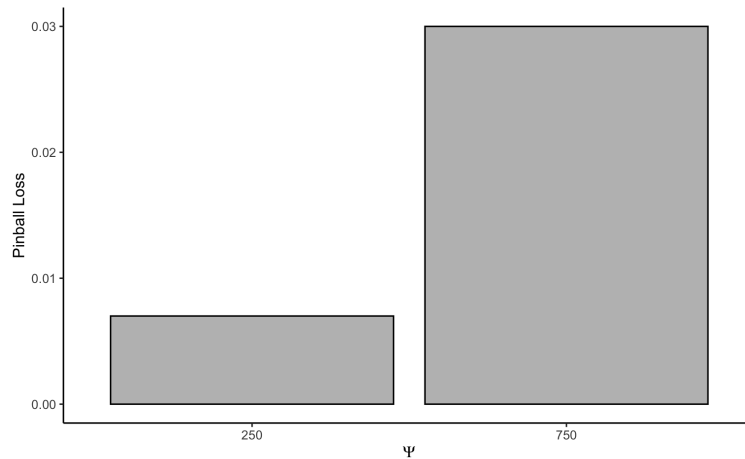


Figure 30: Shows the pinball for two models with different vine training period Ψ for testing period July 2008-July 2009.

To study the effect of the marginal training period Γ , this time we have fitted another model with identical model settings to our base model except for $\Gamma = 500$. Figure 31 shows that once again the base model outperforms this alternative model.

Computation time

The running time of the rolling window method depends primarily on the running time of fitting the vine copula model. Therefore, it exhibit varying computation times based on the size of the portfolio, and the different model settings which determine the number of times a vine copula model is fitted. To get a better sense of the computation time, we executed the rolling window method using different portfolio sizes and different model settings on the period of Jan 2012-Jan 2016 which are shown in Table 2.

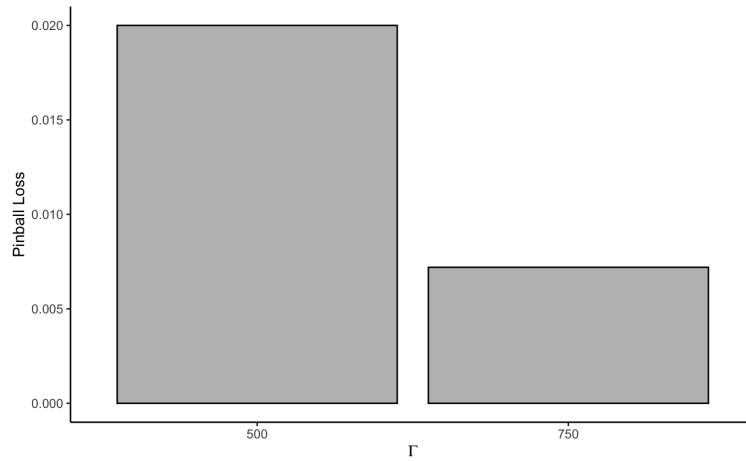


Figure 31: Shows the pinball for two models with different marginal training period Γ for testing period July 2008-July 2009.

Portfolio Size	Γ	γ	Ψ	κ	Allowed Copula Families	Computation Time
3	750	50	250	25	all	1.5 mins
10	750	50	250	25	all	4.4 mins
10	500	50	250	50	all	5.5 mins
10	500	50	250	25	all	8.5 mins
10	750	50	250	50	all	3 mins
10	750	50	250	50	Student t & Gaussian	1.8 mins
20	750	50	250	25	all	15.6 mins
30	750	50	250	25	all	29.3 mins

Table 2: Shows the computation time of the rolling window model corresponding to different portfolio sizes and model settings.

5 Conclusion

In this thesis, we have tackled the problem of assessing the market risk of a portfolio of investments. In particular, large portfolios that are too complicated to work with using traditional methods. We took advantage of vine copula models which are powerful tools that allowed us to capture the dependence among the portfolio components in an efficient way, and by that simplifying our problem considerably.

In addition to giving the theoretical background behind the methods and models used, we have described a model that uses a vine copula model alongside a group of ARMA-GARCH models for forecasting the VaR for a single day. In addition, we have explained how to backtest this model by describing a method for multi-day VaR forecast, which was implemented in an efficient rolling window fashion, that avoids refitting the models used too frequently.

Finally, we gave two examples of using the rolling window method for forecasting the VaR of a portfolio consisting of 30 assets. Furthermore, we show cased the R package VaRVine, which we have developed to tackle the problem addressed in this thesis.

References

- C. Czado. Analyzing dependent data with vine copulas. *Lecture Notes in Statistics*, Springer, 2019.
- J.-Á. Jiménez-Martín, M. McAleer, and T. Pérez-Amaral. The ten commandments for managing value at risk under the basel ii accord. *Journal of Economic Surveys*, 23(5): 850–855, 2009.
- H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of multivariate Analysis*, 94(2):401–419, 2005.
- R. Koenker and K. F. Hallock. Quantile regression. *Journal of Economic Perspectives*, 15 (4):143–156, December 2001. doi: 10.1257/jep.15.4.143. URL <https://www.aeaweb.org/articles?id=10.1257/jep.15.4.143>.
- M. McAleer. The ten commandments for optimizing value-at-risk and daily capital charges. *Journal of economic surveys*, 23(5):831–849, 2009.
- S. Roccioletti. *Backtesting value at risk and expected shortfall*. Springer, 2015.
- R. S. Tsay. *Analysis of financial time series*. John wiley & sons, 3 edition, 2010.