Supplement for
Multivariate Dynamic Copula Models:
Parameter Estimation and Forecast Evaluation

July 5, 2015
A  Static Copulas

This section presents the static multivariate copulas employed in the paper including the corresponding simulation algorithms.

A.1 Elliptical Copulas

Sklar’s theorem indicates that typical multivariate distributions describe central dependence structures which is why elliptical copulas are also known as implicit copulas: A multivariate normal distribution entails a Gaussian copula $C_{G\text{a}}$, whereas a multivariate t-distribution entails a Student-t copula $C_t$. The advantage of implicit copulas is their ability to easily extend to multiple dimensions and their parameter plurality (Nelsen, 2006).

Gaussian Copula

The $d$-dimensional Gaussian copula for a correlation matrix $\Sigma$ is given by

$$C^{G\text{a}}_{\Sigma}(u_1, \ldots, u_d) = \Phi_{\Sigma}(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d))$$

$$= \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_d)} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left( -\frac{1}{2} y' \Sigma^{-1} y \right) dy_1 \cdots dy_d,$$  \hspace{1cm} (1)

where $\Phi$ is the cumulative distribution function of a standard normal distribution, whereas $\Phi_{\Sigma}$ is the cumulative distribution function of the multivariate normal distribution having a mean of zero and a covariance matrix $\Sigma$. The density of any copula which proves to be adequately differentiable may be computed with

$$c(u_1, \ldots u_d) = \frac{f(F^{-1}_1(u_1), \ldots, F^{-1}_d(u_d))}{f_1(F^{-1}_1(u_1)), \ldots, f_d(F^{-1}_d(u_d))},$$

where $f$ is the joint density and $f_1, \ldots, f_d$ are the marginal densities (McNeil et al., 2005; Schmidt, 2007).

Simulation from the Gaussian copula is straightforward. Step number one is to generate random variables according to the underlying multivariate distribution. Step two then consists of transforming them to uniform marginal distributions by quantile transformation. The resulting algorithm is the following (McNeil et al., 2005; Schmidt, 2007):

1. Obtain the correlation matrix $\Sigma$ from any covariance matrix $\Sigma$ by scaling each component to variance 1.

2. Compute the Cholesky-decomposition $\Sigma = A' A$

3. Generate independent and identically distributed standard normal random variables $\tilde{X}_1, \ldots, \tilde{X}_d$
4. From \( \tilde{X} = (\tilde{X}_1, ..., \tilde{X}_d)' \) calculate \( (X_1, ..., X_d)' = X = A\tilde{X} \).

5. Return \( U = (\Phi(X_1), ..., \Phi(X_d))' \) where \( \Phi \) equals the cumulative standard normal distribution function.

**Student-\( t \) Copula**

Identical to the multivariate normal distribution one may obtain an implicit copula from any other distribution with continuous marginal distribution functions. For \( t \)-distributions the \( d \)-dimensional \( t \)-copula with \( \nu \) degrees of freedom is given by

\[
C^t_{\nu, \Sigma}(u_1, ..., u_d) = t_{\nu, \Sigma}(t^{-1}_\nu(u_1),...,t^{-1}_\nu(u_d))
\]

\[
= \int_{-\infty}^{t^{-1}_\nu(u_1)} \cdots \int_{-\infty}^{t^{-1}_\nu(u_d)} \frac{\Gamma((\nu+d)/2) |\Sigma|^{-\frac{1}{2}}}{\Gamma(\frac{\nu}{2}) (\nu\pi)^{\frac{d}{2}}} \left( 1 + \frac{1}{\nu} y' \Sigma^{-1} y \right)^{-\nu+d/2} dy_1 \cdots dy_d, \tag{3}
\]

where \( \Sigma \) is a correlation matrix, \( t_\nu \) is the cumulative distribution function of the one dimensional \( t_\nu \)-distribution and \( t_{(\nu, \Sigma)} \) is the cumulative distribution function of the multivariate \( t_{(\nu, \Sigma)} \)-distribution.

The density of the multivariate Student-\( t \) copula has the form

\[
f^t_{\nu, \Sigma}(u) = \frac{f_{\nu, \Sigma}(t^{-1}_\nu(u_1), ..., t^{-1}_\nu(u_d))}{\prod_{i=1}^d f_{\nu}(t^{-1}_\nu(u_i))}, \tag{4}
\]

where \( u \in (0, 1)^d \), \( f_{\nu, \Sigma} \) is the joint density of a \( t_{\nu}(v, 0, \Sigma) \)-distributed random vector and \( f_{\nu} \) is the density of the univariate standard Student-\( t \)-distribution with \( \nu \) degrees of freedom (Demarta and McNeil, 2005). Simulation from the Student-\( t \) copula can be done using the following algorithm [McNeil et al. (2005)]:

1. Use steps 1 to 4 of the algorithm for the Gaussian copula to generate multivariate normal \( X \) with covariance \( \Sigma \).
2. Generate independent \( \xi \chi^2_\nu \) by e.g. using \( \xi = \sum_{i=1}^\nu Y_i^2 \), where \( Y_i \) are independent and identically \( N(0, 1) \) distributed.
3. Return \( U = (t_\nu(X_1/\sqrt{\xi/\nu}), ..., t_\nu(X_d/\sqrt{\xi/\nu}))' \) where \( t_\nu \) denotes the cumulative distribution function of a univariate \( t \)-distribution with \( \nu \) degrees of freedom.

While both the Gaussian and the Student-\( t \) copula are dependence structures implied by elliptical distributions, they differ with regards to tail dependence. Lower (upper) tail dependence refers to the density in the lower (upper) tail of the copula function and represents the probability of observing joint negative (positive) extremes. The Student-\( t \) copula is capable of capturing equal lower and upper tail dependence, whereas the Gaussian copula has no tail dependence implying independence of the extreme realizations. The level of tail dependence of the Student-\( t \) copula is governed by \( \nu \): the greater the degrees of freedom, the lower the level of tail dependence, converging in the limit \( \nu \to \infty \) to the Gaussian copula.
A.2 Archimedean Copulas

The name Archimedean refers to the copulas’ algebraic property which resembles the Archimedean axiom for real numbers. In contrast to elliptical copulas, Archimedean copulas are given explicitly and capture all information about the dependence structure in the univariate generator function \( \phi \). There exists a wide selection of different bivariate Archimedean copulas, however, choices on the multivariate level are limited. While Schweizer and Sklar (1983) prove that the generator \( \phi \) creates a bivariate copula if and only if it is convex, McNeil and Něslehová (2009) show that in case \( \phi : [0, 1] \to [0, \infty] \) is a strict Archimedean copula generator, then

\[
C(u_1, \ldots, u_d) = \phi^{-1}(\phi(u_1) + \ldots + \phi(u_d))
\]  
(5)

induces a copula in any dimension \( d \) if and only if the generator inverse \( \phi^{-1} : [0, \infty] \to [0, 1] \) is \( d \)-monotone. Accordingly, the strictly decreasing function \( \phi \) has to be continuous on \([0, \infty]\), admit derivatives up to the order \( d - 2 \) and satisfy

\[
(-1)^k \frac{d^k}{dt^k} \phi(t) \geq 0, \quad k \in \{0, \ldots, d - 2\}, t \in (0, \infty).
\]  
(6)

Mostly, it is assumed, that \( \phi \) is completely monotonic, meaning that \( k \in \mathbb{N}_0 \) (Hofert et al., 2013).

With the Clayton and the Frank copula, this paper includes two copulas of the Archimedean family. Both have completely monotonic generators and are thus well-defined for multiple dimensions. While the Frank copula is symmetric exhibiting no tail dependence, the asymmetric Clayton copula is particularly interesting for risk forecasting purposes, since it is capable of modeling lower tail dependence.

**Clayton Copula**

The generator function of the Clayton copula is defined as

\[
\phi_{Cl}(u) = \frac{1}{\theta} (u^{-\theta} - 1),
\]  
(7)

where the permissible parameter range is \( \theta \in (0, \infty) \). A \( d \)-dimensional Clayton copula is given by

\[
C(u_1, \ldots, u_d) = \left( \sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right)^{-\frac{1}{\theta}}.
\]  
(8)

As the copula parameter \( \theta \) tends to infinity, the dependence becomes maximal while the limiting case \( \theta = 0 \) should be interpreted as the \( d \)-dimensional independence copula (McNeil et al., 2005).

The density of the multivariate Clayton copula is

\[
\frac{\partial^d C}{\partial u_1 \ldots \partial u_d} = \theta^d \frac{\Gamma \left( \frac{1}{\theta} + d \right)}{\Gamma \left( \frac{1}{\theta} \right)} \left( \sum_{i=1}^{d} u_i^{-\theta} - d + 1 \right)^{-\frac{1}{\theta} - d} \left( \prod_{i=1}^{d} u_i^{-\theta - 1} \right),
\]  
(9)
where $\Gamma$ denotes the usual Euler $\Gamma$ function.

The contribution of [Marshall and Olkin 1988] can be used to elegantly simulate from multivariate Archimedean copulas. The simulation algorithm exploits the fact that every completely monotonic function mapping from $[0, \infty]$ to $[0, 1]$ can be expressed in terms of Laplace-Stieltjes transforms of distribution functions on $\mathbb{R}^+$ and therewith provide a way of describing multivariate Archimedean copulas.

Let $G$ be a distribution function on $\mathbb{R}^+$ which satisfies $G(0) = 0$ with Laplace-Stieltjes transform

$$ \hat{G}(t) = \int_0^{+\infty} e^{-tx} dG_\gamma(x). \quad (10) $$

For $\hat{G}(\infty) := 0$ it is clear that $\hat{G} : [0, \infty] \to [0, 1]$ is a continuous, strictly decreasing function with the property of complete monotonicity [McNeil et al. 2005]. It can thus serve as a candidate of a multivariate Archimedean copula generator inverse which leads to the following simulation procedure for the Clayton copula [Hofert 2008, McNeil et al. 2005]:

1. Generate a gamma variable $\gamma \sim \text{Gamma}(1/\theta, 1)$ with $\theta \in (0, \infty)$. The distribution function of $\gamma$ thus has Laplace transform $\hat{G}(t) = (1 + t)^{-\frac{1}{\theta}}$.
2. Generate independent uniform variates $(X_1, ..., X_d) \sim U[0, 1]$.
3. Return $(U_1, ..., U_d) = (\hat{G}(-\log(X_1)/\gamma), ..., \hat{G}(-\log(X_d)/\gamma))'$.

**Frank Copula**

The Frank copula generator is given by

$$ \phi_{Fr}(u) = \log \left( \frac{\exp(-\theta u) - 1}{\exp(-\theta) - 1} \right), \quad (11) $$

hence

$$ \phi_{Fr}^{-1}(u) = \frac{1}{\theta} \log \left( 1 + e^u (e^{-\theta} - 1) \right) \quad (12) $$

is completely monotonic if $\theta \in (0, \infty)$. The multivariate Frank copula is

$$ C(u_1, ..., u_d) = -\frac{1}{\theta} \log \left( 1 + \prod_{i=1}^d \frac{(e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{d-1}} \right). \quad (13) $$

The independence copula is attained for $\theta = 0$, whereas with $\theta \rightarrow \infty$ maximal dependence is achieved. The density of the multivariate Frank copula is given by:

1. Generate a discrete variable $V$ with probability mass function $p(k) = P(V = k) = (1 - \exp(-\theta))^k/(k\theta)$ for $k = 1, 2, ..., n$ and $\theta \in (0, \infty)$.
2. Generate independent uniform variates $(X_1, ..., X_d) \sim U[0, 1]$. 4
3. Return \((U_1, ..., U_d) = (\hat{G}(-\log(X_1)/V), ..., \hat{G}(-\log(X_d)/V))'\).

As for the Clayton copula simulation, the algorithm is especially efficient in large dimensions, as only \(d+1\) random numbers are required for the generation of a \(d\)-dimensional observation.

### A.3 Mixture Copulas

Nelsen (2006) demonstrates that a convex combination of different copulas is yet again a copula. Mixing copulas with different dependence features provides an appealing flexibility in constructing dependence characterizations and broadens the range of copula structures applicable on the multivariate level. Mixture copulas were applied, e.g., by Hu (2006), Dias and Embrechts (2010), Weiss (2011) and Ruenzi and Weigert (2013) in a bivariate setting and by Braun (2011) in a multivariate context.

This paper constructs dependence structures which are capable of capturing asymmetries by amalgamating the lower tail dependence feature of the Clayton copula with the symmetric dependence structure of the other copulas under consideration by combining them into mixture structures. The applied static mixture copulas consist of two different copulas but of \(d\) dimensions and have distribution functions of the form

\[
C (u_1, ..., u_d; w, \theta) = wC_1(u_1, ..., u_d, \theta_1) + (1 - w)C_2(u_1, ..., u_d, \theta_2),
\]

where \(\theta_1\) and \(\theta_2\) are the parameter sets of the different copulas. The density of the static mixture construct is simply the convex combination of the copula densities involved in the mixture:

\[
c (u_1, ..., u_d; w, \theta) = wc_1(u_1, ..., u_d, \theta_1) + (1 - w)c_2(u_1, ..., u_d, \theta_2).
\]

While the degree of dependence is carried by the parameters of the copulas within the mixture, the shape of the dependence is summarized by the weight of each individual copula. Compared to the stand-alone static copulas, the mixture copula provides increased flexibility to adapt to different dependence structures, since it unites the diverse features of the enclosed copulas. Combining an elliptical copula with an Archimedean copula means that two different copulas describe the dependence structure of the data according to their proportion in the mixture construct. As a consequence, each copula is adjusted solely to the share of the data set it represents. Such a partition of the dependence structure should allow a more precise calibration of the copula parameters, as they only have to accommodate their fraction of the data set. In contrast, the parameters of stand-alone copulas are required to cover the entire data set which may entail larger compromises and inaccuracies in the calibration.

Figure [1] gives a visual illustration of the mixture concept on the bivariate level. The density of the Gaussian copula is depicted on the left, the Clayton copula density on the right and the plot in the center shows the density of the combination of these two copulas in a mixture copula with \(w = 0.5\). This mixture copula has the advantage of accommodating the parameter plurality of the Gaussian copula (with increasing number of dimensions).
and the asymmetric dependence feature of the Clayton copula. While static mixture copulas aim at a more precise representation of the dependence structure compared to stand-alone copulas, they are not capable of capturing dependence shifts over time, since their shapes, parameters and mixture weights remain stable over time. Therefore, in the paper we derive regime-switching and dynamic copula models and analyze their in-sample fit and out-of-sample forecast performance.

Figure 1: Gaussian-Clayton Mixture Copula Density

The figure shows the Gaussian (left), Gauss-Clayton mixture with equal weights (center) and Clayton copula density (right).

B Multi-Stage Maximum Likelihood Estimation

B.1 Estimation Procedure

The model estimation takes advantage of the fact that the copula is independent of the marginal distributions and separates the procedure into different stages (see, e.g., [Joe (2001); McNeil et al. (2005); Chollete et al. (2009)]). The overall log likelihood depends on all the data $Y = (Y_1^T, ..., Y_T^T)$, and is given by

$$
\log L(Y; \theta_m, \theta_c) = \sum_{t=1}^{T} \log f(Y_t|Y_{t-1}; \theta_m, \theta_c),
$$

(16)

where $Y_{t-1} = (Y_1, ..., Y_t)$ represents the history of the entire process. The likelihood can thus be decomposed into one part $\log L_m$ containing the marginal densities and a second part $\log L_c$ which contains the copula densities

$$
\log L(Y; \theta_m, \theta_c) = \log L_m(Y; \theta_m) + \log L_c(Y; \theta_m, \theta_c),
$$

(17)

$$
\log L_m(Y; \theta_m) = \sum_{t=1}^{T} \sum_{i=1}^{d} \log f_i(y_{i,t}|y_{i,t-1}; \theta_m),
$$

(18)
\[ \log \mathcal{L}_c(Y; \theta_m, \theta_c) = \sum_{t=1}^{T} \log c(F_1(y_{1,t}|y_{1}^{t-1}; \theta_{m,1}), ..., F_d(y_{d,t}|y_{d}^{t-1}; \theta_{m,d}); \theta_c), \tag{19} \]

where \( y_{i}^{t-1} = (y_{i,1}, ..., y_{i,t}) \) is the entire history of variable \( i \).

The marginal models’ likelihood \( \log \mathcal{L}_m \) is a function of the parameter vector \( \theta_m = (\theta_{m,1}, ..., \theta_{m,d}) \) which collects the parameters for each of the \( d \) marginal density functions \( f_i \). The likelihood of the copula \( \log \mathcal{L}_c \) directly depends on vector \( \theta_c \). For the estimation of singular static copulas, \( \theta_c \) contains the copula parameters. In case of the static mixtures, \( \theta_c = (\theta^{(1)}_c, \theta^{(2)}_c, w) \) comprises the parameters of both copulas in the mixture copula and the mixture weight. For the estimation of regime-switching copula containing singular copulas for regimes, the vector \( \theta_c = (\theta^{(1)}_c, ..., \theta^{(k)}_c, P) \) collects the copula parameters over all regimes plus the parameters of the transition probability matrix \( P \), while for regime-switching copulas comprising mixtures, the mixture weight parameter \( w \) is also collected \( \theta_c = (\theta^{(1a)}_c, \theta^{(1b)}_C, \theta^{(2)}_c, P, w_{1a}) \). Through the distribution function \( F_i \) this parameter vector \( \theta_c \) also indirectly depends on the parameters of the marginal densities, since \( F_i \) transforms the observations into uniform [0, 1] variables based on which the copula is estimated.

The models in this paper accommodate a number of parameters such that a full single-step likelihood maximization is numerically rather intricate and time consuming. Maximizing the parameters separately for the margins and the copula is also referred to as “inference functions for margins” or IFM (see, e.g., Joe (2001); McNeil et al. (2005)), though more generally it is known as multi-stage maximum likelihood estimation (MSMLE) (Patton (2013)). Compared to a one-stage estimation it represents a much more tractable procedure whose properties have been studied by Chen and Fan (2006), Joe (2005), and Patton (2006), and which has repeatedly been applied in the context of copulas (see, e.g., Patton (2006); Chollete et al. (2009); Dias and Embrechts (2010); Garcia and Tsafack (2011)). While the multi-stage estimation generally entails some loss of efficiency in comparison to estimating the entire joint distribution in one single step, it substantially simplifies the computational burden and comes with a low loss of efficiency (Joe (2005) and Patton (2006)).

In the first step, it is assumed that the marginals are independent from each other and depend only on their own history:

\[ \hat{\theta}_m = \theta_m \mathcal{L}_m(Y; \theta_m). \tag{20} \]

This estimation can further be simplified since the univariate model parameters of each time series can be calibrated separately:

\[ \hat{\theta}_{m,i} = \theta_{m,i} \sum_{t=1}^{T} \log f_i(y_{i,t}|y_{i}^{t-1}; \theta_{m,i}). \tag{21} \]

For the second step, the marginal parameters are bundled in a vector \( \hat{\theta}_m = (\hat{\theta}_{m,i}, ..., \hat{\theta}_{m,n}) \) and are considered as given in order to calibrate the copula:

\[ \hat{\theta}_c = \theta_c \mathcal{L}_c(Y; \hat{\theta}_m, \theta_c). \tag{22} \]

This procedure tremendously reduces the computational effort and time, firstly because a numerical optimization with lots of parameters is much more time consuming compared
to several numerical optimizations, each with fewer parameters. Secondly, the multi-stage method reduces the overall computational burden since different copulas are estimated given the same univariate models, which means that the parameter vector \( \hat{\theta}_m \) can be reused.

To guarantee positive semi-definiteness of the correlation matrices \( \Sigma \) contained in the elliptical copula models, the numerical optimization is effectively carried out for the lower triangular matrix \( A \) of Cholesky factors, such that \( \Sigma = AA' \). All maximum likelihood estimation procedures require starting values for the optimization to initialize. With the plurality of the parameters to be estimated in the more elaborated dependence models, the choice of initial values can have a direct impact on the overall estimation time. Naturally the closer the starting values are to the final parameters, the faster the optimization algorithm converges. For the correlation parameters of the Gaussian copulas, the Cholesky factors of the quantile transformed uniform values’ correlation matrix are used as initial values. For the Student-\( t \) copula’s correlation parameters, the fact that sample rank correlations can be used to partially calibrate Student-\( t \) copulas is exploited (see, e.g., [McNeil et al. (2005)]). The relationship between Kendall’s tau and the \( t \)-copula’s correlation is

\[
\rho_{\tau}(y_i, y_j) = \frac{2}{\pi} \arcsin \rho_{i,j}^c,
\]

such that a possible estimator of the Student-\( t \) copula correlation matrix is given by a matrix with the components

\[
\rho_{i,j}^c = \sin\left(\frac{1}{2} \frac{\pi \rho_{\tau}^T}{2}\right).
\]

For the Archimedean copulas, in a first step the \( (d\choose2) \) pairwise Kendall’s tau estimators are computed. Using the arithmetic average over all these estimators, the functional relationship between Kendall’s tau \( \rho_{\tau} \) and the Archimedean copula parameter \( \theta \), presented in Equation (19) of the paper, is exploited. The usage of these initial values was found to decrease estimation time supporting the findings of [Hofert et al. (2013)], who in contrast to some statements in literature (see, e.g., [Berg and Aas (2009) or Weiss (2010)]) find that maximum-likelihood estimation is feasible in higher dimensions and performs well.

### B.2 Standard Error Computation

To calculate the standard errors within the MSMLE framework, [Joe (2001)] and [Durrleman et al. (2000)] show that the vector of parameter estimates, \( \hat{\theta}_{MSMLE} = [\hat{\theta}_{m,1}, ..., \hat{\theta}_{m,d}, \hat{\theta}_c] \) verifies the property of asymptotic normality such that

\[
\sqrt{T}(\hat{\theta}_{MSMLE} - \theta_0) \xrightarrow{d} N(0, G^{-1}(\theta_0)) \quad \text{as} \quad T \to \infty,
\]

where \( G(\theta_0) \) is the asymptotic variance-covariance matrix known as the information matrix of Godambe. It is based on the theory of inference functions, which imposes optimality criteria on the score functions of the estimating equations rather than on the estimators received from them [Godambe (1960, 1976, 1991)].

Following [Joe (2001)] in defining a score function \( g \):

\[
g(\theta) = (\partial_{\theta_1} \log \mathcal{L}_{m,1}, ..., \partial_{\theta_d} \log \mathcal{L}_{m,d}, \partial_{\theta_c} \log \mathcal{L}_c),
\]

where \( \mathcal{L} \) is the likelihood function.
the Godambe information matrix takes the form

$$G(\theta_0) = D^{-1}M(D^{-1})',$$  \hspace{1cm} (27)

where

$$D = E\left[\frac{\partial g(\theta)'}{\partial \theta}\right], \quad M = E[g(\theta)'g(\theta)].$$  \hspace{1cm} (28)

For the models at hand, the estimation of the Godambe information matrix requires the computation of a number of derivatives (i.e., score functions for likelihood functions) which can be computationally demanding. Moreover, the main issue in implementing the multi-stage estimation method compared to a one-step maximum likelihood method is its loss of performance in the estimation of the parameters since the marginal estimation procedure in the first step neglects the possible dependence between $\theta_{m,1}, ..., \theta_{m,d}$ when estimating $\theta_c$ (Liu and Luger (2009)).

In order to adequately compute standard errors for the models, we therefore follow Silva Filho et al. (2012) and De Lira Salvatierra and Patton (2013) in employing a bootstrap approach to calculate the covariance matrix of the parameters. To asymptotically preserve the cross-sectional dependence in the time series data, the block bootstrap method of Politis and Romano (1994) is applied. The optimal block size is determined based on the automatic block-length selection of Politis and White (2004) and Patton et al. (2009). The standard error computation procedure is as follows:

1. Estimate parameters $\hat{\theta}$ as described in section B.1.
2. Generate a block bootstrap sample of the same size as the data set.
3. Estimate the parameters $\hat{\theta}^{(r)}$ on the simulated data.
4. Repeat steps (2) and (3) $R$ times.
5. Compute standard errors using the covariance matrix

$$R^{-1} \sum_{r=1}^{R} (\hat{\theta}^{(r)} - \hat{\theta})(\hat{\theta}^{(r)} - \hat{\theta})',$$

where $\hat{\theta}^{(r)}$ is the column vector of the estimated parameters for every replication $r$.

\footnote{See Gonçalves and White (2004) for theoretical justification.}
C In-Sample Analysis

C.1 Univariate Models

The marginal models play an important role for dependence modeling since they filter the univariate risk factor evolutions from autocorrelation, heteroscedasticity and leverage and yield the input data for the copula estimation. Misspecification of the marginals can thus result in biased copula parameter estimates which in turn yields inadequate portfolio risk forecasts. Models which are traditionally utilized in time series analysis to capture (linear) serial dependence in the returns are for example ARMA($p,q$) models that read:

\[ y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \xi_j \epsilon_{t-j} + \epsilon_t. \]  

(29)

To model the volatility clusters within Equation (29), the residuals $\epsilon_t$ are decomposed such that:

\[ \epsilon_t = z_t \sigma_t. \]  

(30)

In order to choose the most appropriate model specification for each of the return series, the likelihood ratio test is not applicable since the different models are not nested. Instead, the information criterion of Akaike (1973) (AIC) and the Bayesian Information Criterion (BIC) of Schwarz (1978) can be used to find a suitable univariate model. Given the $m$ models $M_1, \ldots, M_m$ for the $n$-dimensional return vector $Y$, where model $j$ has $k_j$ parameters denoted $\theta_j = (\theta_{j1}, \ldots, \theta_{jk})'$ with the likelihood function $L_j$, the criteria are defined as

\begin{align*}
AIC(M_j) &= -2 \log L_j(\hat{\theta}_j; Y) + 2k_j, \\
BIC(M_j) &= -2 \log L_j(\hat{\theta}_j; Y) + k_j \log(n),
\end{align*}

(31) (32)

where $\hat{\theta}_j$ denotes the maximum likelihood estimation of $\theta_j$. Each of these criteria has two terms; the first term measures the goodness-of-fit and the second term penalizes model complexity. The model favored is the one for which the respective criterion is minimized. Since the penalty of the BIC is a function of the sample size of $Y$, it is more severe for the applications at hand. The BIC thus favors more parsimonious models which is why it is the model selection criterion of choice in case the two information criteria favor different model specifications. The univariate model for each index return series was found by selecting the AIC and BIC optimal model considering ARMA($p,q$) specifications for the conditional mean up to order ($p=3,q=3$) and GARCH($P,Q$), EGARCH($P,Q$), and GJR-GARCH($P,Q$) volatility models up to order ($P=3,Q=3$). The EGARCH($P,Q$) model specification reads:

\[ \log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{P} \alpha_i |z_{t-i}| + \sum_{i=1}^{P} \gamma_i z_{t-i} + \sum_{j=1}^{Q} \beta_j \log(\sigma_{t-j}^2). \]  

(33)

The parameters of the accordingly selected univariate models (ARMA(1,0) EGARCH(1,1) for the equity index portfolio are listed in Table 1. Note that the automatic model selection by AIC / BIC yields parsimonious specifications, as no model is of an order higher
than one. The results in Table 1 show that a moving average parameter is not part of any optimal model for the return series. With the DAX 30, the CAC 40 and the S&P 500, three out of the eight series contain an autoregressive part for the conditional mean, while the mean of the other series is simply characterized by a constant. Leverage appears to be a feature in all of the return series as the information criteria favor a leveraged conditional volatility model over the standard GARCH model for all indices. The EGARCH specification (as shown in Table 1) is selected by the information criteria to be more suitable than the GJR model for every equity return series (see Equation (33)). All leverage parameters are negative which is in line with the economic interpretation, as in the EGARCH a negative parameter captures the phenomenon of negative innovations to returns having a more significant impact on volatility than positive return innovations.
<table>
<thead>
<tr>
<th>Index</th>
<th>SMI</th>
<th>DAX</th>
<th>CAC</th>
<th>UKX</th>
<th>SPX</th>
<th>TSX</th>
<th>HS</th>
<th>Nik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>μ</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>φ</td>
<td>-0.073</td>
<td>-0.088</td>
<td>-0.104</td>
<td>-0.073</td>
<td>-0.088</td>
<td>-0.104</td>
<td>-0.073</td>
<td>-0.088</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.034)</td>
<td>(0.037)</td>
<td>(0.034)</td>
<td>(0.034)</td>
<td>(0.037)</td>
<td>(0.034)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>α₀</td>
<td>-0.873</td>
<td>-0.576</td>
<td>-0.538</td>
<td>-0.511</td>
<td>-0.519</td>
<td>-0.311</td>
<td>-0.306</td>
<td>-0.583</td>
</tr>
<tr>
<td></td>
<td>(0.382)</td>
<td>(0.332)</td>
<td>(0.316)</td>
<td>(0.451)</td>
<td>(0.328)</td>
<td>(0.326)</td>
<td>(0.234)</td>
<td>(0.379)</td>
</tr>
<tr>
<td>α₁</td>
<td>0.257</td>
<td>0.243</td>
<td>0.224</td>
<td>0.176</td>
<td>0.174</td>
<td>0.194</td>
<td>0.266</td>
<td>0.212</td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.078)</td>
<td>(0.079)</td>
<td>(0.080)</td>
<td>(0.075)</td>
<td>(0.082)</td>
<td>(0.065)</td>
<td>(0.085)</td>
</tr>
<tr>
<td>γ₁</td>
<td>-0.186</td>
<td>-0.133</td>
<td>-0.144</td>
<td>-0.174</td>
<td>-0.162</td>
<td>-0.067</td>
<td>-0.041</td>
<td>-0.109</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.051)</td>
<td>(0.049)</td>
<td>(0.052)</td>
<td>(0.054)</td>
<td>(0.046)</td>
<td>(0.047)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>β₁</td>
<td>0.884</td>
<td>0.919</td>
<td>0.919</td>
<td>0.934</td>
<td>0.933</td>
<td>0.960</td>
<td>0.955</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(0.122)</td>
<td>(0.138)</td>
<td>(0.123)</td>
<td>(0.102)</td>
<td>(0.129)</td>
<td>(0.081)</td>
<td>(0.137)</td>
</tr>
<tr>
<td></td>
<td>(0.901)</td>
<td>(0.912)</td>
<td>(1.463)</td>
<td>(1.263)</td>
<td>(0.696)</td>
<td>(1.194)</td>
<td>(1.081)</td>
<td>(1.257)</td>
</tr>
<tr>
<td>λ</td>
<td>-0.218</td>
<td>-0.269</td>
<td>-0.272</td>
<td>-0.204</td>
<td>-0.275</td>
<td>-0.186</td>
<td>-0.215</td>
<td>-0.137</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.036)</td>
<td>(0.037)</td>
<td>(0.038)</td>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.043)</td>
<td>(0.043)</td>
</tr>
</tbody>
</table>

This table shows parameter estimates for the models of the weekly returns over the full sample period from June 30, 1988 to June 5, 2013 for the SMI, DAX 30, CAC 40, FTSE 100 (UKX), S&P 500 (SPX), S&P/TSX 60, Hang Seng (HSI) and Nikkei 225 index. The Model line indicates the volatility model selected by the AIC / BIC, where $E$ stands for the EGARCH specification. KS, CvM and AD report the $p$-values of the parameter estimation error adjusted results (based on 1,000 simulations) of the Kolmogorov-Smirnov (KS), Cramer-von Mises (CvM) and Anderson-Darling (AD) goodness-of-fit tests for the models of the conditional marginal distributions. LBQ($k$) reports the $p$-value of the Ljung-Box Q-statistics (adjusted for the number of estimated parameters) assessing the null hypothesis of no autocorrelation of the squared standardized residuals for $k$ lags.

To complete the univariate model specifications we still need to specify the marginal distributions of the standardized residuals, $z_t$. Returns exhibit asymmetric fat tails (as shown in the section Descriptive Statistics in the paper). To further account for this feature in our modeling approach, we modeled the uniform variates $z_t$ (Equation (30)) by a skewed-$t$ distribution. As shown in Table 1, the estimated skewness parameters $\lambda$ are all negative and the degrees of freedom parameters range from 7.731 to 15.069 both with significantly low standard errors. Recalling that the skewed-$t$ distribution converges to a Student-$t$ distribution when $\lambda = 0$ and to a normal distribution when $\lambda = 0$ and $\nu \to \infty$, one may conclude that the skewed-$t$ is a more suitable model for the residual distribution.

$^2$The usual $t$-statistic $\hat{\alpha}/\hat{\sigma}_\alpha$ is used to gauge the significance of a parameter $\alpha$. 

12
than either the Gaussian or the standard Student-t. The negative skewness parameters indicate that the residual distributions exhibit longer left tails compared to the right tails emphasizing the importance of accounting for asymmetric return distributions. The DAX 30 reveals the most prominent left tail incorporating the lowest degrees of freedom and the third largest skewness. In case the fat tails are neglected or underestimated, the respective quantiles are inadequate which in turn distorts the risk forecasts.

Figure 2: QQ plots of Empirical vs. Fitted skewed-t Quantiles

This figure shows the quartile-quartile the fitted skewed-t quantiles of the models’ standardized innovations. If the standardized univariate model residuals adhere to the skewed-t distribution (with parameters for each series listed in Table 1) then the data markers will fall on the dashed 45° line.

A first inspection of the quality of the fittings is done graphically through quantile-quartile plots of standardized innovations. Figure 2 shows the plots of the empirical versus the fitted skewed-t quantiles and points to a fairly good fit since the pairs of quantiles lay almost all very close to the main diagonal. The skewed-t distribution seems to be particularly capable of adequately capturing the lower tails of the residual distributions. To formally test whether the estimated model is appropriate for all return series, three goodness-of-fit tests are applied: The Kolmogorov-Smirnov (KS), the Cramer-von Mises (CvM), and the Anderson-Darling (AD) test with the test statistics defined as:

\[
KS_i = \max_t \left| \hat{U}_{i,t}(t) - \frac{t}{T} \right|, \tag{34}
\]

\[
CvM_i = \sum_{t=1}^{T} \left( \hat{U}_{i,t}(t) - \frac{t}{T} \right)^2, \tag{35}
\]

\[
AD = -T - \frac{1}{T} \sum_{t=1}^{T} \left( (2t - 1) \log(\hat{U}_{i,t}) + (2T + 1 - 2t) \log(1 - \hat{U}_{i,t}) \right), \tag{36}
\]
where \( \hat{U}_{i,t} \) is the \( t^{th} \) order statistic of \( \left\{ \hat{U}_{i,j} \right\}_{j=1}^{T} \), i.e. the \( t^{th} \) largest value of \( \left\{ \hat{U}_{i,j} \right\}_{j=1}^{T} \).

The test statistics are based on the estimated probability integral transformations \( \hat{U}_{i,t} \equiv F_{\text{skewed-}t}(\hat{\epsilon}_{it}; \hat{\nu}_{i}, \hat{\lambda}_{i}) \) and have asymptotic distributions that are known in the absence of parameter estimation errors. However, these asymptotic distributions cannot be applied here as the univariate models are based on a number of estimated parameters. Since the parametric models for the mean, variance and error distribution completely characterize the conditional distribution, this can be overcome by obtaining critical values using the following simulation-based method (Patton (2013) and Genest and Rémillard (2008)):

1. Simulate \( T \) returns from a univariate model using the estimated parameters.
2. Estimate the model on the simulated returns.
3. Compute the test statistics on the estimated probability integral transforms of the simulated returns.
4. Repeat steps (1) to (3) \( R = 1'000 \) times.
5. Use the upper \( 1-\alpha \) quantile of \( \left\{ (KS_{(r)}, CvM_{(r)}, AD_{(r)}) \right\}_{r=1}^{R} \) as the critical value for the tests.

The implementation of this procedure yields the \( p \)-values of the three test statistics listed in the Table 1. The lowest \( p \)-values for all three tests are found for the model of the DAX 30 which yields 0.377 (KS), 0.382 (CvM) and 0.441 (AD). The three tests therewith cannot reject the null hypothesis that the skewed-\( t \) models are well-suited for the equity index return series. The magnitude of the \( p \)-values further confirms the conclusion suggested by the quantile-quantile plots, that the skewed-\( t \) distribution adapts very well to the empirical data.
This figure shows the autocorrelation functions of the squared standardized residuals from the univariate models for the equity returns.

The standardized residuals obtained by filtering each of the returns series with its AIC / BIC-optimal model should be independent and identically distributed. This means that the series of squared standardized residuals must also be independent and identically distributed. Figure 3 shows a graphical display for the estimates of serial correlation at different lags together with 95% confidence bounds. The visual inspection of these correlograms indicates that an independent and identical distribution seems to be given for all series, as the autocorrelation remains within the confidence bounds for all lags $k = 1, \ldots, 20$. In order to formally test the null hypothesis of zero autocorrelation, the portmanteau test of Ljung and Box (1978) is performed on the squared standardized residuals using lags up to the twentieth order. To account for the estimated parameters, the degrees of freedom of the statistic’s limiting $\chi^2$ distribution is adjusted for the number of estimated parameters $p$ (excluding constants): $LBQ(k) \overset{d}{\rightarrow} \chi^2_{k-p}$. Table 1 lists the $p$-values of the tests at lags 7, 10, 15, and 20. The test results for all tested orders indicate that the null hypothesis of zero correlation in the squared standardized residuals of any of the univariate models cannot be rejected. The Ljung-Box test therewith supports the findings of the goodness-of-fit tests and contributes to the conclusion that the univariate models are well fitted.

The results in Table 1 show that a moving average parameter is not part of any optimal model for the return series. With the DAX 30, the CAC 40 and the S&P 500 three out of the eight series contain an autoregressive part for the conditional mean, while the mean of the other series is simply characterized by a constant. Leverage appears to be a feature in all of the return series as the information criteria favor a leveraged condi-

\footnote{Note that with the adjustment for the number of estimated parameters, the test is valid only for $k > p$. \cite{Tsay2005} finds evidence of increased test power by setting $k \approx \log(T)$, i.e. $k = \log(1300) \approx 7$.}
tional volatility model over the standard GARCH model for all indices. The EGARCH specification is selected by the information criteria to be more suitable than the GJR model for every equity return series. All leverage parameters are negative which is in line with the economic interpretation, as in the EGARCH a negative parameter captures the phenomenon of negative innovations to returns having a more significant impact on volatility than positive return innovations.

C.2 Static Dependence

To calibrate the static copulas, the filtered standardized residuals from the univariate models are transformed to uniform variates by inversion using the corresponding cumulative skewed-t distribution function. Three static mixture copulas are constructed by combining the asymmetric Clayton copula with the other three (symmetric) copulas. The combination of the two Archimedean copulas into a mixture construct yields a parsimonious model which is able of capturing lower tail dependence. Mixing the Clayton with the Gaussian copula combines the parameter plurality of the elliptical copula with the lower tail dependence feature of the Clayton copula and creates a flexible model which is capable of capturing asymmetries in the dependence structure. Adding the Clayton to the Student-\(t\) copula finally results in an adaptive model capable of modeling different degrees of upper and lower tail dependence. Results are shown in Table 2.

Using the information criteria to rank the fit of the static models, the Student-\(t\)-Clayton mixture fits best according to both AIC and BIC, followed by the Student-\(t\) copula and the Gauss-Clayton mixture. Overall, the in-sample analysis shows that the best fitting static dependence model is the Student-\(t\)-Clayton mixture copula. The two models with the second-best fit are the Gaussian-Clayton mixture copula and the Student-\(t\) copula. All Archimedean copulas rank far behind with the stand-alone Archimedean copulas having the largest AIC/BIC values and the lowest likelihood value, respectively.

To gauge how well the different copula models fit the data, standard goodness-of-fit tests based on the comparison with the empirical copula employing the test statistics in Equations (34), (35) and (36) are not applicable. Patton (2013) emphasizes that such tests rely on the empirical copula to serve as a non-parametric estimate of the true conditional copula, but when the true conditional copula is not constant, the empirical copula cannot be used for such tests. Therefore, following Dias and Embrechts (2010), Guégan and Zhang (2010), the information criteria are used to rank the fit of the different models.

The higher log likelihood value and therewith - as both copulas have a single parameter - lower AIC and BIC indicate that the Clayton copula provides a better fit to the data among the two standalone Archimedean copulas. The Clayton’s superiority is due to its ability to capture the returns’ lower tail dependence. The mixture of the two Archimedean copulas have a better fit than the stand-alone Archimedean copulas as shown by both information criteria. Intuitively, the Frank-Clayton mixture model improves the adaptability of the Frank dependence structure by creating a possibly asymmetric structure allowing for lower tail dependence. With 58.7% and a low standard error of 6.5%, the Clayton copula has the larger weight in the mixture structure. Clayton’s \(\theta_C\) is further substantially higher in the mixture than in the stand-alone Clayton copula, which
<table>
<thead>
<tr>
<th>Copula</th>
<th>Clayton</th>
<th>Frank</th>
<th>Gaussian</th>
<th>Student-t</th>
<th>FC Mix</th>
<th>GC Mix</th>
<th>TC Mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMI:DAX</td>
<td>0.697 (0.018)</td>
<td>0.720 (0.016)</td>
<td>0.807 (0.023)</td>
<td>0.781 (0.019)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI:CAC</td>
<td>0.683 (0.019)</td>
<td>0.716 (0.018)</td>
<td>0.824 (0.028)</td>
<td>0.796 (0.026)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI:FTS</td>
<td>0.660 (0.019)</td>
<td>0.690 (0.017)</td>
<td>0.789 (0.030)</td>
<td>0.757 (0.025)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI:SPX</td>
<td>0.591 (0.021)</td>
<td>0.607 (0.019)</td>
<td>0.690 (0.026)</td>
<td>0.659 (0.023)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI:TSX</td>
<td>0.499 (0.025)</td>
<td>0.505 (0.024)</td>
<td>0.575 (0.030)</td>
<td>0.544 (0.027)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI:HSE</td>
<td>0.418 (0.027)</td>
<td>0.447 (0.025)</td>
<td>0.523 (0.041)</td>
<td>0.493 (0.033)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMI:NIK</td>
<td>0.381 (0.026)</td>
<td>0.407 (0.026)</td>
<td>0.456 (0.040)</td>
<td>0.448 (0.035)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAX:CAC</td>
<td>0.793 (0.015)</td>
<td>0.818 (0.014)</td>
<td>0.895 (0.018)</td>
<td>0.880 (0.015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAX:FTS</td>
<td>0.670 (0.022)</td>
<td>0.698 (0.020)</td>
<td>0.794 (0.027)</td>
<td>0.770 (0.024)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAX:SPX</td>
<td>0.637 (0.021)</td>
<td>0.658 (0.019)</td>
<td>0.740 (0.025)</td>
<td>0.715 (0.022)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAX:TSX</td>
<td>0.547 (0.021)</td>
<td>0.551 (0.020)</td>
<td>0.618 (0.028)</td>
<td>0.591 (0.024)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAX:HSE</td>
<td>0.489 (0.025)</td>
<td>0.505 (0.024)</td>
<td>0.561 (0.036)</td>
<td>0.544 (0.029)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAX:NIK</td>
<td>0.439 (0.025)</td>
<td>0.463 (0.024)</td>
<td>0.509 (0.037)</td>
<td>0.506 (0.033)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAC:FTS</td>
<td>0.725 (0.017)</td>
<td>0.754 (0.015)</td>
<td>0.843 (0.023)</td>
<td>0.822 (0.021)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAC:SPX</td>
<td>0.641 (0.021)</td>
<td>0.661 (0.020)</td>
<td>0.750 (0.027)</td>
<td>0.724 (0.024)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAC:TSX</td>
<td>0.560 (0.023)</td>
<td>0.566 (0.021)</td>
<td>0.640 (0.030)</td>
<td>0.612 (0.025)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAC:HSE</td>
<td>0.490 (0.023)</td>
<td>0.501 (0.024)</td>
<td>0.559 (0.033)</td>
<td>0.539 (0.027)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAC:NIK</td>
<td>0.459 (0.026)</td>
<td>0.482 (0.025)</td>
<td>0.530 (0.037)</td>
<td>0.523 (0.032)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTS:SPX</td>
<td>0.662 (0.020)</td>
<td>0.667 (0.019)</td>
<td>0.741 (0.023)</td>
<td>0.720 (0.022)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTS:TSX</td>
<td>0.566 (0.022)</td>
<td>0.573 (0.020)</td>
<td>0.645 (0.027)</td>
<td>0.620 (0.024)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTS:HSE</td>
<td>0.532 (0.026)</td>
<td>0.535 (0.025)</td>
<td>0.600 (0.032)</td>
<td>0.576 (0.027)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTS:NIK</td>
<td>0.432 (0.024)</td>
<td>0.448 (0.023)</td>
<td>0.509 (0.037)</td>
<td>0.493 (0.032)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPX:TSX</td>
<td>0.721 (0.015)</td>
<td>0.714 (0.014)</td>
<td>0.757 (0.016)</td>
<td>0.740 (0.015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPX:HSE</td>
<td>0.483 (0.027)</td>
<td>0.490 (0.026)</td>
<td>0.561 (0.035)</td>
<td>0.530 (0.028)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPX:NIK</td>
<td>0.431 (0.024)</td>
<td>0.438 (0.026)</td>
<td>0.485 (0.034)</td>
<td>0.471 (0.031)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSX:HSE</td>
<td>0.455 (0.025)</td>
<td>0.443 (0.025)</td>
<td>0.488 (0.033)</td>
<td>0.465 (0.027)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSX:NIK</td>
<td>0.392 (0.027)</td>
<td>0.389 (0.027)</td>
<td>0.440 (0.032)</td>
<td>0.419 (0.030)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSE:NIK</td>
<td>0.416 (0.030)</td>
<td>0.430 (0.029)</td>
<td>0.469 (0.039)</td>
<td>0.461 (0.032)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \nu )</td>
<td></td>
<td></td>
<td>10.655 (0.803)</td>
<td>13.355 (1.324)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| \( \theta_F \) |         | 3.453 (0.141) | 2.311 (1.282) |        |        |        |
| \( \theta_C \) | 0.714 (0.038) | 1.166 (0.194) | 0.280 (0.306) | 0.330 (0.411) |        |        |        |
| \( w_C \) |         | 0.587 (0.065) | 0.587 (0.065) | 0.199 (0.029) | 0.149 (0.029) |        |        |
| \( \log L \) |         | 2269 | 2234 | 3172 | 3345 | 2522 | 3345 | 3399 |
| AIC | -4536 | -467 | -6289 | -6632 | -5039 | -6690 | -6735 |
| BIC | -4531 | -461 | -6143 | -6482 | -5023 | -6475 | -6576 |

This table shows the estimates of the static copulas with standard errors in parentheses. The mixture copulas are abbreviated: Frank-Clayton mixture (FC Mix), Gaussian-Clayton mixture (GC Mix) and Student-t-Clayton mixture (TC Mix). \( w_C \) denotes the weight of the Clayton copula in the mixture.
reflects the increased flexibility of the mixture copula to adapt to the data. The Clayton proportion of the mixture copula captures the asymmetric dependence, while the Frank proportion captures symmetric dependence. The higher standard errors of the Clayton and particularly of the Frank copula parameter shows that the increase in flexibility comes at a price of less reliable parameters.

With their correlation matrices, the elliptical copulas contain a multiple of the number of parameters in the Archimedean copulas. Their increased likelihood value thus reflects the additional number of parameters. However, both information criteria account for the number of parameters, and both indicate a better fit of the stand-alone elliptical copulas compared to both the stand-alone and the mixture Archimedean copulas. The degrees of freedom parameter \( \nu = 10.655 \) of the Student-t copula is to be considered as rather low which confirms tail dependence in the return data. However, the Student-t copula implies the same degree of lower and upper tail dependence, even if only the former is present.

This drawback is addressed by mixing the asymmetric Clayton with one of the symmetric elliptical copulas. The weight of the Clayton in the Gaussian-Clayton mixture amounts to 19.9%, however \( \theta_C \) is rather low. Furthermore, the uncertainty in the Clayton copula parameter indicated by the according standard error is high. As one might expect, the Student-t-Clayton mixture displays a higher \( \nu \), since the Clayton copula fraction of the mixture accounts for lower tail dependence. However, for both the Gaussian-Clayton and the Student-t-Clayton mixture copula, the standard errors do not allow to conclude with certainty that \( \theta_C \) is different from zero. Using the information criteria to rank the fit of the static models, the Student-t-Clayton mixture fits best according to both AIC and BIC, followed by the Student-t copula and the Gauss-Clayton mixture.

C.3 Regime-Switching Modeling Results

In the last two columns of Table 3 we show the three-regime-switching copula parameters, in addition to the two-regime-switching models (already shown in the paper). In the G/T/C model, as shown in Figure 4, besides a high and a low dependence state, the third regime seems to model some middle or “normal” state of the economy with a dependence level midway between the ones of the extreme states. In contrast to what one might expect, this “normal” state is not modeled by the Gaussian copula but by the Clayton copula. The low probability of staying in the third regime conditional on being in regime three suggests that this regime is more of a pass-through between the other regimes which both display much higher probabilities \( p_{1|1}, p_{2|2} \). Following Hamilton (1989), the transition probabilities can be used to compute the probability to stay \( k \) weeks in regime \( j \) as \( P(D_j = k) = p_{jj}^{k-1}(1 - p_{jj}) \), where \( D_j \) denotes the number of periods the Markov chain is in state \( j \). This implies that the expected duration of this state is

\[
E(D_j) = \sum_{k=1}^{\infty} kp_{jj}^{k-1}(1 - p_{jj}) = \frac{1}{1 - p_{jj}}. \tag{37}
\]
Table 3: Regime Switching Copula Parameters

<table>
<thead>
<tr>
<th>Copula</th>
<th>G/G</th>
<th>G/T</th>
<th>T/T</th>
<th>T/TC</th>
<th>G/T/C</th>
<th>G/C/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMI:DAX</td>
<td>0.838 (0.154)</td>
<td>0.496 (0.108)</td>
<td>0.822 (0.125)</td>
<td>0.826 (0.090)</td>
<td>0.452 (0.104)</td>
<td>0.518 (0.119)</td>
</tr>
<tr>
<td>SMI:CAC</td>
<td>0.858 (0.188)</td>
<td>0.409 (0.141)</td>
<td>0.847 (0.166)</td>
<td>0.853 (0.130)</td>
<td>0.340 (0.139)</td>
<td>0.452 (0.169)</td>
</tr>
<tr>
<td>SMI:FTS</td>
<td>0.838 (0.179)</td>
<td>0.414 (0.129)</td>
<td>0.820 (0.157)</td>
<td>0.830 (0.139)</td>
<td>0.327 (0.139)</td>
<td>0.386 (0.179)</td>
</tr>
<tr>
<td>SMI:SPX</td>
<td>0.717 (0.160)</td>
<td>0.408 (0.094)</td>
<td>0.712 (0.102)</td>
<td>0.690 (0.105)</td>
<td>0.381 (0.100)</td>
<td>0.441 (0.125)</td>
</tr>
<tr>
<td>SMI:FTS</td>
<td>0.612 (0.220)</td>
<td>0.183 (0.136)</td>
<td>0.561 (0.152)</td>
<td>0.605 (0.160)</td>
<td>0.133 (0.156)</td>
<td>0.160 (0.192)</td>
</tr>
<tr>
<td>SMI:HSE</td>
<td>0.573 (0.249)</td>
<td>0.086 (0.142)</td>
<td>0.546 (0.170)</td>
<td>0.518 (0.195)</td>
<td>0.093 (0.185)</td>
<td>0.072 (0.228)</td>
</tr>
<tr>
<td>DAX:CAC</td>
<td>0.932 (0.138)</td>
<td>0.570 (0.100)</td>
<td>0.922 (0.127)</td>
<td>0.919 (0.095)</td>
<td>0.501 (0.100)</td>
<td>0.600 (0.118)</td>
</tr>
<tr>
<td>DAX:FTS</td>
<td>0.849 (0.185)</td>
<td>0.429 (0.129)</td>
<td>0.821 (0.128)</td>
<td>0.835 (0.136)</td>
<td>0.435 (0.166)</td>
<td>0.457 (0.180)</td>
</tr>
<tr>
<td>DAX:SPX</td>
<td>0.798 (0.173)</td>
<td>0.406 (0.108)</td>
<td>0.791 (0.140)</td>
<td>0.768 (0.121)</td>
<td>0.386 (0.120)</td>
<td>0.436 (0.146)</td>
</tr>
<tr>
<td>DAX:FTS</td>
<td>0.657 (0.150)</td>
<td>0.392 (0.081)</td>
<td>0.621 (0.076)</td>
<td>0.644 (0.088)</td>
<td>0.394 (0.099)</td>
<td>0.356 (0.108)</td>
</tr>
<tr>
<td>DAX:HSE</td>
<td>0.641 (0.193)</td>
<td>0.290 (0.107)</td>
<td>0.602 (0.123)</td>
<td>0.657 (0.135)</td>
<td>0.225 (0.126)</td>
<td>0.326 (0.156)</td>
</tr>
<tr>
<td>DAX:NK</td>
<td>0.619 (0.234)</td>
<td>0.170 (0.135)</td>
<td>0.597 (0.154)</td>
<td>0.563 (0.195)</td>
<td>0.193 (0.172)</td>
<td>0.139 (0.217)</td>
</tr>
<tr>
<td>CAC:FTS</td>
<td>0.883 (0.158)</td>
<td>0.495 (0.115)</td>
<td>0.861 (0.144)</td>
<td>0.873 (0.114)</td>
<td>0.418 (0.116)</td>
<td>0.536 (0.133)</td>
</tr>
<tr>
<td>CAC:SPX</td>
<td>0.801 (0.185)</td>
<td>0.404 (0.121)</td>
<td>0.804 (0.152)</td>
<td>0.791 (0.138)</td>
<td>0.383 (0.133)</td>
<td>0.432 (0.165)</td>
</tr>
<tr>
<td>CAC:TSX</td>
<td>0.683 (0.171)</td>
<td>0.380 (0.103)</td>
<td>0.649 (0.097)</td>
<td>0.673 (0.108)</td>
<td>0.338 (0.105)</td>
<td>0.366 (0.130)</td>
</tr>
<tr>
<td>CAC:HSE</td>
<td>0.646 (0.192)</td>
<td>0.284 (0.105)</td>
<td>0.604 (0.126)</td>
<td>0.650 (0.139)</td>
<td>0.207 (0.132)</td>
<td>0.333 (0.157)</td>
</tr>
<tr>
<td>CAC:NK</td>
<td>0.632 (0.236)</td>
<td>0.220 (0.133)</td>
<td>0.613 (0.152)</td>
<td>0.581 (0.188)</td>
<td>0.256 (0.169)</td>
<td>0.168 (0.216)</td>
</tr>
<tr>
<td>FTS:SPX</td>
<td>0.775 (0.145)</td>
<td>0.517 (0.094)</td>
<td>0.763 (0.113)</td>
<td>0.776 (0.102)</td>
<td>0.460 (0.095)</td>
<td>0.543 (0.115)</td>
</tr>
<tr>
<td>FTS:TSX</td>
<td>0.708 (0.174)</td>
<td>0.370 (0.103)</td>
<td>0.664 (0.102)</td>
<td>0.687 (0.107)</td>
<td>0.355 (0.104)</td>
<td>0.361 (0.131)</td>
</tr>
<tr>
<td>FTS:HSE</td>
<td>0.664 (0.191)</td>
<td>0.353 (0.112)</td>
<td>0.623 (0.139)</td>
<td>0.673 (0.138)</td>
<td>0.251 (0.127)</td>
<td>0.429 (0.161)</td>
</tr>
<tr>
<td>FTS:NK</td>
<td>0.594 (0.230)</td>
<td>0.209 (0.126)</td>
<td>0.559 (0.147)</td>
<td>0.540 (0.178)</td>
<td>0.244 (0.160)</td>
<td>0.188 (0.198)</td>
</tr>
<tr>
<td>FTS:NIK</td>
<td>0.751 (0.054)</td>
<td>0.669 (0.032)</td>
<td>0.732 (0.034)</td>
<td>0.744 (0.028)</td>
<td>0.640 (0.026)</td>
<td>0.691 (0.037)</td>
</tr>
<tr>
<td>SPX:TSX</td>
<td>0.638 (0.186)</td>
<td>0.281 (0.116)</td>
<td>0.593 (0.130)</td>
<td>0.643 (0.132)</td>
<td>0.214 (0.127)</td>
<td>0.325 (0.156)</td>
</tr>
<tr>
<td>SPX:HSE</td>
<td>0.561 (0.194)</td>
<td>0.249 (0.116)</td>
<td>0.550 (0.131)</td>
<td>0.521 (0.162)</td>
<td>0.296 (0.152)</td>
<td>0.229 (0.176)</td>
</tr>
<tr>
<td>SPX:NK</td>
<td>0.548 (0.129)</td>
<td>0.303 (0.082)</td>
<td>0.496 (0.067)</td>
<td>0.554 (0.091)</td>
<td>0.239 (0.090)</td>
<td>0.347 (0.109)</td>
</tr>
<tr>
<td>TSX:NIK</td>
<td>0.511 (0.166)</td>
<td>0.220 (0.097)</td>
<td>0.472 (0.097)</td>
<td>0.452 (0.123)</td>
<td>0.246 (0.118)</td>
<td>0.219 (0.139)</td>
</tr>
<tr>
<td>HSE:NK</td>
<td>0.590 (0.209)</td>
<td>0.186 (0.123)</td>
<td>0.558 (0.135)</td>
<td>0.560 (0.177)</td>
<td>0.196 (0.172)</td>
<td>0.157 (0.207)</td>
</tr>
</tbody>
</table>

Continued on next page

This table shows regime switching copula parameters with standard errors in parentheses. The copulas are abbreviated: Frank (F), Clayton (C), Gaussian (G), Student-t (T), Gaussian-Clayton mixture (GC) and Student-t-Clayton mixture (TC). \( \omega_C \) denotes the weight of the Clayton copula in the mixture. The forward slash indicates the separate regimes i.e. G/T/C and G/C/F are three-state models. \( p_{ij} \) denotes the probability of staying in regime \( i \).
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.513 (0.154)</td>
<td>0.815 (0.066)</td>
<td>0.511 (0.126)</td>
<td>0.477 (0.134)</td>
<td>0.826 (0.090)</td>
<td>0.811 (0.095)</td>
<td>0.450 (0.185)</td>
<td>0.846 (0.088)</td>
<td>0.433 (0.168)</td>
<td>0.415 (0.190)</td>
<td>0.853 (0.130)</td>
<td>0.838 (0.149)</td>
<td>0.360 (0.142)</td>
<td>0.563 (0.050)</td>
<td>0.385 (0.068)</td>
<td>0.340 (0.106)</td>
<td>0.583 (0.087)</td>
<td>0.552 (0.089)</td>
<td>0.117 (0.250)</td>
<td>0.543 (0.113)</td>
<td>0.144 (0.169)</td>
<td>0.053 (0.222)</td>
<td>0.518 (0.195)</td>
<td>0.540 (0.209)</td>
<td>0.608 (0.139)</td>
<td>0.917 (0.072)</td>
<td>0.503 (0.131)</td>
<td>0.573 (0.133)</td>
<td>0.919 (0.095)</td>
<td>0.918 (0.102)</td>
<td>0.438 (0.193)</td>
<td>0.815 (0.092)</td>
<td>0.431 (0.172)</td>
<td>0.422 (0.176)</td>
<td>0.825 (0.128)</td>
</tr>
</tbody>
</table>
The figure shows Kim filtered regime probabilities of the three-regime-switching copula models over the entire sample period. The copulas are abbreviated: Gaussian (G), Student-t (T), Frank (F), Clayton (C), Gaussian-Clayton mixture (GC) and Student-t-Clayton mixture (TC). The solid black line marks the high dependence regime.

Table 4 lists the expected durations of the regimes in the different models. It shows that for the two-regime models, the high dependence regime (marked in bold) has a much longer expected duration compared to the other regime(s) in every copula model. The table illustrates that although the transition probabilities $p_{11}$ and $p_{22}$ are high probabilities (exceeding 0.82 for every two-state model), they reflect a very different regime persistence. The additional length in expected duration in the two-regime setups range from one third in the T/T model to more than two thirds in the T/GC structure. The differences in the three-regime models are even larger. However, the minimal expected duration of the third regime in the G/T/C model suggests that this regime may be irrelevant. In the G/C/F model, the regime modeled by the Frank copula has a parameter $\theta_F$ of virtually zero, which means that in this setup there are times when the returns are independent. Even though the transition probability $p_{33}$ is not as low as in the other three regime model, it has the second lowest expected duration (three weeks).

Table 4: Expected Regime Durations $E(D_{R_j})$

<table>
<thead>
<tr>
<th></th>
<th>G/G</th>
<th>G/T</th>
<th>T/T</th>
<th>T/GC</th>
<th>G/TC</th>
<th>G/T/C</th>
<th>G/C/F</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(D_{R_1})$</td>
<td>13</td>
<td>8</td>
<td>8</td>
<td>22</td>
<td>11</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>$E(D_{R_2})$</td>
<td>9</td>
<td>13</td>
<td>6</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>$E(D_{R_3})$</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table presents the expected regime durations (in weeks) under the regime switching models. The expected duration of the high dependence regime is marked in bold. The copulas are abbreviated: Gaussian (G), Student-t (T), Clayton (C), Frank (F), Gaussian-Clayton mixture (GC) and Student-t-Clayton mixture (TC). The forward slash indicates the separation of the regimes.
D Forecast Evaluation

D.1 Backtesting Risk Measures

The most common reported risk measure is the Value-at-Risk (VaR) as it represents the industry and regulatory standard for the calculation of risk capital in banking and insurance (Embrechts et al., 2013). The VaR for a given confidence level $\alpha$ is defined as (Christoffersen 2012):

$$ Pr(-y_t > VaR_t(\alpha)) = 1 - \alpha \iff Pr(y_t < -VaR_t(\alpha)) = 1 - \alpha. $$

Hence, VaR($\alpha$) is the negative return that will not be exceeded with probability $\alpha$. Formally, the VaR($\alpha$) corresponds to the $(1 - \alpha)$ percentile of the portfolio return distribution. The most widely used alternative to VaR is expected shortfall (ES), also known as Conditional VaR (CVaR), which is defined as

$$ ES(\alpha) = \mathbb{E}[y_t \mid y_t < -VaR_t(\alpha)], $$

and possesses some advantages over VaR as it is a coherent risk measure representing the expected return conditional on VaR being violated. Despite theoretical advantages, the vast majority of financial institutions uses VaR and not ES. This may be due to problems with the backtesting procedure of ES, as it requires estimates of the tail expectation to be compared with the forecast of ES. Danielsson (2011) points out that in a backtest, forecasted ES can only be compared to a model output while VaR can be compared with real observations. The resulting complete profit and loss functions allow to calculate VaR forecasts at arbitrary significance levels $\alpha$. The most common confidence levels 90%, 95%, and 99% are deployed. If the actual return of the portfolio over the forecast period falls short of the forecast, then the VaR limit is said to have been violated. In a backtesting procedure, violations over time are a sequence of zeros and ones, also called “hit sequence”, denoted $\eta_t$. Formally, a violation of the VaR($\alpha$) forecast is an event such that

$$ \eta_{t,\alpha} = \begin{cases} 1, & \text{if } y_t \leq -VaR_t(\alpha) \\ 0, & \text{if } y_t > -VaR_t(\alpha) \end{cases}, $$

where $y_t$ is the portfolio return in time $t$. A judgment on the quality of the models forecast can then be made by calculating the hit ratio, which reflects the percentage of times when the portfolio return exceeds the forecasted VaR($\alpha$) in a sample with size $T$:

$$ \text{Hit Ratio}(\alpha) = \frac{\sum_{t=1}^{T} \eta_{t,\alpha}}{T}. $$

Backtesting for example the VaR($\alpha = 99\%$) means that one expects to observe a VaR violation in 1% of the time. Naturally, the closer the hit ratio is to the expected value $(1 - \alpha)$, the better the forecasts of the risk model. If the hit ratio is greater than the expectation, then the model underforecasts the portfolio risk; with a hit ratio smaller than $(1 - \alpha)$, the model overforecasts risk.

---

4See Artzner et al. (1999) or McNeil et al. (2005) for a detailed discussion on the property of coherence.
To formally test whether the difference between observed and expected hit ratio is significant, statistical tests are called for. According to Christoffersen (1998), the measurement of the accuracy of a VaR forecast model can be reduced to determining whether the hit sequence $\eta$ satisfies two properties: the unconditional coverage property and the independence property.

**Unconditional Coverage Test**

The unconditional coverage test was introduced by Kupiec (1995). The property of unconditional coverage ensures that the empirical hit ratio corresponds with the theoretical significance level $p = (1 - \alpha)$. If a violation on day $t$ occurred, then variable $\eta_t$ takes the value 1 and 0 otherwise (see Equation (40)). $\eta_t$ is thus a sequence of Bernoulli-distributed random variables. The unconditional coverage test is used to determine the violations’ proportion. For VaR violations, the null hypothesis is

$$H_0 : \eta \sim \text{i.i.d.} B(p),$$

(42)

where $B$ stands for the Bernoulli distribution with density

$$f(\eta_t; p) = (1 - p)^{1-\eta_t} p^{\eta_t}, \quad \eta_t = 0, 1.$$  

(43)

To test whether a hit ratio $\pi$ obtained by a risk model is equal to the expected fraction $p$, the likelihood of an i.i.d. Bernoulli($\pi$) hit sequence is used:

$$L(\pi) = \prod_{t=1}^{T} (1 - \pi)^{1-\eta_t} \pi^{\eta_t} = \prod_{t=1}^{T} (1 - \pi)^{T_0} \pi^{T_1},$$

(44)

where $T_1 = \sum_{t=1}^{T} \eta_t$ and $T_0 = T - T_1$ are the number of zeros and ones in the hit sequence $\eta$ and $\pi = T_1 / T$. The test statistic for the null hypothesis $H_0 : \pi = p$ is given by the likelihood ratio

$$LR_{uc} = -2 \log \left( \frac{(1 - p)^{T_0} p^{T_1}}{(1 - \pi)^{T_0} \pi^{T_1}} \right),$$

(45)

which is asymptotically $\chi^2$ distributed with one degree of freedom. Since the unconditional coverage test does not assume a distribution for the returns, it is a nonparametric test which commonly provides good benchmarks for the assessment of the accuracy of VaR models (Danielsson, 2011).

**Independence Test**

Theoretically, violations should spread out over time such that an adequate risk model would not yield VaR violation clusters. Based on this idea, Christoffersen (1998) generalized the approach of Kupiec (1995) to include a test of independence. To fulfill the property of independence, any two observations in the hit sequence must be independent of one another. The fact that a violation has been observed in time $t$ should thus not yield any information about the likelihood of observing a violation in $t + 1$. In case previous VaR violations predict a future VaR violation, this indicates a general inadequacy
of the reported risk measure. In order to establish a test of the independence of the VaR violations, Christoffersen (1998, 2012) assumes that the hit sequence is dependent over time and that it can be described by a first-order Markov chain with the transition probability matrix. The probability of a violation in \( t + 1 \), given a violation in \( t \) is defined by

\[ p_{11} = \Pr(\eta_{t+1} = 1|\eta_t = 1), \]

while the probability of a violation in \( t + 1 \), given there is no violation in \( t \) is \( p_{01} = \Pr(\eta_{t+1} = 1|\eta_t = 0) \). The probability of no violation in \( t + 1 \) following no violation in \( t \) is \( 1 - p_{01} \). With a sample of \( T \) observations, the likelihood function of the first-order Markov process is given by

\[
L = (1 - \pi_{01})^{T_{00}} \pi_{01}^{T_{01}} (1 - \pi_{11})^{T_{10}} \pi_{11}^{T_{11}},
\]

(46)

where \( T_{ij} \) equals the number of observations with a \( j \) following an \( i \). By taking the first derivatives with respect to \( \pi_{01} \) and \( \pi_{11} \) and setting them equal to zero, one may obtain the maximum likelihood estimates \( \hat{\pi}_{01} = T_{01}/(T_{00} + T_{01}) \) and \( \hat{\pi}_{11} = T_{11}/(T_{10} + T_{11}) \). The matrix of the estimated transition probabilities is therewith:

\[
\hat{\Pi}_1 = \begin{bmatrix} 1 - \hat{\pi}_{01} & \hat{\pi}_{01} \\ 1 - \hat{\pi}_{11} & \hat{\pi}_{11} \end{bmatrix} = \begin{bmatrix} \frac{T_{00}}{(T_{00} + T_{01})} & \frac{T_{01}}{(T_{00} + T_{01})} \\ \frac{T_{10}}{(T_{10} + T_{11})} & \frac{T_{11}}{(T_{10} + T_{11})} \end{bmatrix}.
\]

(47)

If the violations are independent over time, then \( \hat{\pi}_{01} = \hat{\pi}_{11} = \hat{\pi} \) such that the transition probability matrix is

\[
\hat{\Pi}_0 = \begin{bmatrix} 1 - \hat{\pi} & \hat{\pi} \\ 1 - \hat{\pi} & \hat{\pi} \end{bmatrix}.
\]

(48)

To test the null hypothesis of independence of the violations \( H_0 : \hat{\pi}_{01} = \hat{\pi}_{11} \) the likelihood ratio test is used:

\[
LR_{\text{ind}} = -2 \log \frac{L(\hat{\Pi}_0)}{L(\hat{\Pi}_1)}.
\]

(49)

The statistic is asymptotically \( \chi^2 \) distributed with one degree of freedom (Christoffersen, 2012).

**Joint Test**

An accurate VaR forecast has to feature both the independence and unconditional coverage property. A test that jointly examines the unconditional coverage and independence properties thus gives an opportunity to identify VaR measures which are defective in one way or the other. One may conclude, that the joint test should generally be preferred to the individual tests of coverage or independence. However, this is usually not the case since the joint test entails less power to reject a model which only satisfies either one of the properties (Danielsson, 2011). To jointly test if the observed violations are significantly different from the expected ones and if the violations are independent over time, we apply the joint test, also known as conditional coverage test. The test statistic of the joint test is the sum of the test statistics for the individual properties

\[
LR_{\text{joint}} = LR_{\text{unc}} + LR_{\text{ind}},
\]

(50)

which is asymptotically \( \chi^2 \) distributed with two degrees of freedom.
Basel Three-Zone Approach

The Basel regulatory framework provides a categorization of risk models based on their performance in the backtesting procedure. Accordingly, a model is categorized into one of the three traffic light zones “green”, “yellow” and “red” based on the probability of obtaining up to the observed number of $x$ exceedances of the 1% VaR when the true coverage level is 99%. Formally, this corresponds to the binomial cumulative distribution function:

$$Pr(X \leq x; T, p) = \sum_{i=0}^{x} \binom{T}{i} p^i (1-p)^{T-i},$$  \hspace{1cm} (51)$$

where $x$ is the number of observed exceedances of the model’s 1% VaR in the backtesting period, $p = 1\%$ and $T$ equals the size of the backtesting sample. Models, whose backtesting performance is such that $Pr(X \leq x) < 95\%$ are categorized as green, which means that the backtesting results do not suggest a problem with the quality or accuracy of the model. Models with $x$ exceedances such that $95\% \leq Pr(X \leq x) \leq 99.99\%$ are categorized as yellow, which is to be interpreted as a model that raises questions but no definitive conclusion is possible. Finally, a model with a backtesting performance such that the probability of obtaining up to the backtested number of $x$ exceptions equals or exceeds 99.99%, i.e. $Pr(X \leq x) > 99.99\%$ is classified as red, which indicates that there is almost certainly a problem with the risk model (Basel Committee on Banking Supervision, 2013).

Expected Shortfall Evaluation

A backtest of the expected shortfall is a rather intricate task, because the aim is to test an expectation rather than a single quantile. As the expected shortfall is the expected loss, given a violation of VaR, Danielsson (2011) proposes to compare for all points in time $t$ when the VaR is violated, to which extent on average the ES corresponds to the realized return. This indicates how much of the realized return, given a VaR violation, is on average forecasted by the model:

$$\text{ES Ratio}(\alpha) = T_1^{-1} \left( \sum_{t=1}^{T} \eta_{t,\alpha} \times \frac{ES_{t,\alpha}}{y_t} \right),$$  \hspace{1cm} (52)$$

where $T_1$ is the number of ones in the hit sequence $\eta$. Naturally, the closer the ES ratio is to one, the more accurate the forecasts. In case the ES ratio is larger than one, the model overforecasts the expected shortfall and if the ES ratio is smaller than one, the model underestimates the loss given a violation of the VaR. Since ES ratio takes the average over all those times when the VaR forecast of a model is violated, this measure hinges on the model’s hit ratio. While it gives a further insight into the predictive power of a single model, it only allows for a direct comparison of competing forecast models if their hit ratios are identical.

Ziegel (2014) points out that other procedures for evaluating ES forecasts such as the one proposed by McNeil and Frey (2000) do not permit a direct comparison of the predictive performance of the competing forecasting methods either. Gneiting (2011)
introduces the notion of elicitability, where elicitable broadly means that a measure is “properly” backtestable. He shows that in general, VaR is elicitable, whereas ES is not. Specifically, [Gneiting (2011)] proves that the existence of convex level sets is a necessary condition for the elicitability of a risk measure and disproves the existence of convex level sets for the ES. This provides a potential explanation for the lack of literature on backtesting ES. As a simple option, the [Basel Committee on Banking Supervision (2011)] suggests to backtest the VaR instead of the ES - if VaR is inaccurate, the corresponding ES can hardly be correct.

Gneiting and Ranjan (2011), however, propose to compare density forecasts with emphasis on different regions of interest, such as the tails of the distributions. Given that with the ES the goal is to accurately predict a functional focused on the tails, this approach seems promising (Ziegel, 2014). Testing density forecasts means that the object of interest is shifted from a particular risk measure to the entire profit and loss distribution, respectively its entire tail.

### D.2 Backtesting the Entire Distribution

Instead of focusing on particular risk measures from the profit and loss distribution such as the Value-at-Risk or the Expected Shortfall, one might instead decide to backtest the entire forecasted profit and loss distribution from the risk model. This has the benefit of potentially further increasing the power to reject bad risk models. The idea of testing the entire predictive distribution goes back to [Berkowitz (2001)] and has the appeal that it makes use of the entire predictive distribution, compared to the approaches in Subsection D.1 which effectively throw away information. The attractiveness of the approach of [Berkowitz] is also based on the use of the probability integral transformation, which allows conducting much more powerful tests than otherwise possible.

In time $t$ the portfolio risk forecast models produce a cumulative profit and loss distribution forecast $F_t(\cdot)$ for $t + 1$. At the end of $t + 1$, with the realized portfolio return $y_{t+1}$, the risk model’s probability of observing a return below the actual one can be computed with

$$\hat{p}_{t+1} = F_t(y_{t+1}).$$

Under the null hypothesis that the forecast model is accurate, the time series of observed probabilities $\hat{p}_{t+1}$ should then be independently and identically uniform $(0,1)$ distributed:

$$H_0 : \hat{p}_{t+1} \sim U(0,1).$$

A visual evaluation of the distribution can be made by constructing a histogram of $\hat{p}_{t+1}$ to determine if the distribution is reasonably flat (Christoffersen, 2012). While visual evaluations are less precise in comparison to statistical tests, they are constructive in the sense that they can provide guidance as to why and where a statistical test is rejected. The model adequacy can be statistically examined by testing whether $\hat{p}_{t+1}$ is uniform as predicted. This uniformity is tested by conducting Pearson’s $\chi^2$-test, which compares the observed with the expected number of values in different sub-intervals of the unit interval.
The test statistic for \( k \) sub-intervals is given by

\[
Q = \sum_{i=1}^{k} \frac{(N(l_i, u_i) - N(u_i - l_i))^2}{N(u_i - l_i)},
\]

where \( N(l_i, u_i) \) refers to the number of observations in the \( i^{th} \) sub-interval and \( N \) refers to the total number of observations being used to construct the test. Also, \( l_i \) and \( u_i \) refer to the lower and upper bound of each sub-interval. The test statistic is approximately distributed according to the \( \chi^2 \)-distribution with \( k - 1 \) degrees of freedom.

In order to apply powerful normality tests, Berkowitz (2001) suggests to further transform the \( i.i.d. \) uniform \( \hat{p}_{t+1} \) into standard normally distributed variables under the null by using the inverse cumulative normal distribution function \( \Phi^{-1} \):

\[
\begin{align*}
H_0 & : \hat{p}_{t+1} \sim U(0, 1) \\
H_0 & : \hat{z}_{t+1} = \Phi^{-1}(\hat{p}_{t+1}) = \Phi^{-1}(F_t(y_{t+1})) \sim \mathcal{N}(0, 1).
\end{align*}
\]

To investigate whether the \( \hat{z}_{t+1} \) are \( i.i.d. \) normally distributed, the Kolmogorov-Smirnov (KS) test and the Anderson-Darling (AD) test are employed.

### D.3 Backtesting the Entire Lower Tail

While it is useful to backtest the estimate of the whole predictive distribution to obtain additional information, the primary interest of financial risk management lies in the lower tail of the profit and loss distribution (McNeil et al., 2005). Testing the entire density as in Sub-section D.2 might lead to a rejection of risk models which capture the lower tail of the profit and loss distribution well, but are not accurate for the rest of the distribution. Instead, Christoffersen (2012) proposes to construct a test that directly focuses on assessing the risk capability of the model to capture the lower tail of the distribution, which contains the largest losses. Christoffersen (2012) restricts the attention to the tail of the distribution below the 10\% quantile. His test consists of examining whether the probability integral transformations below this threshold are themselves uniform, constructing a rescaled probability integral transformed variable as

\[
\hat{p}^{*}_{t+1} = \begin{cases} 
10 \times F_t(y_{t+1}) & \text{if } y_{t+1} \leq -VaR_t \\
\text{Else not defined.} & 
\end{cases}
\]

The null hypothesis that the risk model provides the correct tail distribution is given by

\[
H_0 : \hat{p}^{*}_{t+1} \sim U(0, 1)
\]

or equivalently

\[
H_0 : \hat{z}^{*}_{t+1} = \Phi^{-1}(\hat{p}^{*}_{t+1}) \sim \mathcal{N}(0, 1).
\]

To do formal statistical testing, the alternative hypothesis is constructed as

\[
\hat{z}^{*}_{t+1} = b_0 + b_1 \hat{z}^{*}_t + \sigma_z z_{t+1}, \quad \text{with} \quad z_{t+1} \sim \mathcal{N}(0, 1).
\]
Then the log-likelihood of a sample of $T$ observations of $z_{t+1}^*$ under the alternative hypothesis conditioned on an initial observation is

$$
\log L = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \sum_{t=1}^{T} \left( \frac{(z_{t+1}^* - b_0 - b_1 z_t^*)^2}{2\sigma^2} \right).
$$

With the parameter estimates $\hat{b}_0, \hat{b}_1, \hat{\sigma}^2$ obtained by the maximum likelihood estimation, the likelihood ratio test of a correct lower tail distribution is given by

$$
LR_{lt} = -2 \left( \log L(0, 0, 1) - \log L(\hat{b}_0, \hat{b}_1, \hat{\sigma}^2) \right),
$$

which is asymptotically $\chi^2$-distributed with three degrees of freedom (Christoffersen, 2012).

References


