Financial Data Analysis

Regime–switching Models

June 30, 2010
Mixture Models for Financial Time Series

- The idea of the mixture approach to modeling asset returns is that the distribution of returns depends on an unobserved state (or regime) of the market.

- For example, expected returns as well as variances and correlations may differ in bull and bear markets.

- Assume that there are \( k \) different states of the market and that, given that the market is in state \( j \) at time \( t \), the \( N \times 1 \) vector of returns under consideration is normal with mean \( \mu_j \) and covariance matrix \( \Sigma_j \), i.e., its density is

\[
f(r_t|s_t = j) = \phi(r_t; \mu_j, \Sigma_j) = \frac{1}{(2\pi)^{N/2} \sqrt{|\Sigma_j|}} \exp \left\{ -\frac{1}{2} (r_t - \mu_j)' \Sigma_j^{-1} (r_t - \mu_j) \right\},
\]

where \( s_t \in \{1, \ldots, k\} \) is a variable indicating the market regime at time \( t \).
• At time $t$, the market is in state $j$ with (conditional) probability $\pi_{jt}$, i.e.,

$$\Pr_{t-1}(s_t = j) = \pi_{jt}, \quad j = 1, \ldots, k.$$  

• Then the (conditional) distribution of $r_t$ at time $t$ is a $k$–component finite normal mixture distribution, with (conditional) density

$$f_{t-1}(r_t) = \sum_{j=1}^{k} \pi_{jt} \phi(r_t; \mu_j, \Sigma_j), \quad (2)$$

where $\phi(\cdot; \mu_j, \Sigma_j)$ is the multivariate normal density given in (1).

• In (2), the $\pi_{jt}$ are the (conditional) **mixing weights**, and the $\phi(r_t; \mu_j, \Sigma_j)$ are the **component densities**, or **mixture components**, with **component means** $\mu_j$, and **component covariance matrices** $\Sigma_j$, $j = 1, \ldots, k$. 
• The normal mixture has finite moments of all orders, which can be calculated by making use of the properties of the normal distribution.

• For example, the mean and the covariance matrix are given by

\[ \mu := \mathbb{E}_{t-1}(r_t) = \sum_{j=1}^{k} \pi_{jt} \mu_j \]

and

\[ \operatorname{Var}_{t-1}(r_t) = \sum_{j=1}^{k} \pi_{jt} \Sigma_j + \sum_{j=1}^{k} \pi_{jt} (\mu_j - \mu)(\mu_j - \mu)', \]

respectively.
• We have already considered independent normal mixture distributions, where the mixing weights are constant over time, i.e.,

\[ \pi_{jt} = \lambda_j \quad \text{for all } t. \]

• A mixture of a few normals, say two or three, is capable of capturing the skewness and excess kurtosis detected in empirical return distributions.

• In the multivariate framework, the mixture approach is able to account for regime-specific dependence structures (correlation matrices) in a natural way, while still appealing to correlation matrices in the context of (conditionally) normally distributed returns.

• For example, it is often argued that stock returns are more highly correlated during high-volatility periods, which are often associated with market downturns, i.e., bear markets.

• However, it is in times of adverse market conditions that the benefits from diversification are most urgently needed.
Reminder: Portfolio Diversification

• Suppose we have $N$ risky assets with returns $r_i$, $i = 1, \ldots, n$.

• Let the mean and the covariance matrix of the returns be denoted by $\mu$ and $\Sigma$, i.e.,

$$
\mu = \begin{bmatrix}
\text{E}(r_1) \\
\text{E}(r_2) \\
\vdots \\
\text{E}(r_N)
\end{bmatrix}, \quad \Sigma = \begin{bmatrix}
\text{Var}(r_1) & \text{Cov}(r_1, r_2) & \cdots & \text{Cov}(r_1, r_N) \\
\text{Cov}(r_1, r_2) & \text{Var}(r_2) & \cdots & \text{Cov}(r_2, r_N) \\
\vdots & \vdots & \ddots & \vdots \\
\text{Cov}(r_1, r_N) & \text{Cov}(r_2, r_N) & \cdots & \text{Var}(r_N)
\end{bmatrix}.
$$

• Then, for a vector of portfolio weights, $w$, the mean and the variance of the portfolio return, $r_p$, are given by

$$
\mu_p = w'\mu = \sum_{i=1}^{N} w_i \text{E}(r_i),
$$

$$
\sigma^2_p = w'\Sigma w = \sum_{i=1}^{N} w_i^2 \text{Var}(r_i) + 2 \sum_{j=1}^{N} \sum_{i<j} w_i w_j \text{Cov}(r_i, r_j).
$$
Now suppose (common correlation model)

\[ \text{Var}(r_i) = \sigma^2, \quad i = 1, \ldots, N, \]
\[ \text{Corr}(r_i, r_j) = \rho \quad i, j = 1, \ldots, N, \quad i \neq j, \]

and consider equally weighted portfolio, i.e.,

\[ w_i = \frac{1}{N}, \quad i = 1, \ldots, N. \]

Then the portfolio variance

\[
\sigma_p^2 = \sum_{i=1}^{N} w_i^2 \text{Var}(r_i) + 2 \sum_{j=1}^{N} \sum_{i<j} w_i w_j \text{Cov}(r_i, r_j)
= \sigma^2 \rho
= \frac{\sigma^2}{N} + \frac{N(N-1)}{N^2} \sigma^2 \rho
= \sigma^2 \left( \frac{1 - \rho}{N} + \rho \right)
\approx \rho \sigma^2 \quad \text{for large } N \text{ (many assets)}. \tag{3}
\]
• Question: What if $\rho$ in (4) is negative?

• Answer: This cannot be. It can be shown that the common correlation coefficient of $N$ assets has to satisfy

$$\frac{-1}{N-1} < \rho < 1, \quad (5)$$

so that the expression in brackets in (3) will always be positive.

• As noted by Paul A. Samuelson,\(^1\) this is rather plausible intuitively, since it shows that

“although there is no limit on the degree to which all investments can be positively intercorrelated, it is impossible for all to be strongly negatively correlated.”

---

Modeling the (conditional) mixing weights

- The independent mixture model fits the unconditional return distribution well.

- It does not capture the dynamic properties of asset returns, however, such as volatility clustering.

- Economically, this means that we expect the regimes to be persistent.

- That is, if we are in a bull market currently, the probability of being in a bull market in the next period will be larger than if the current regime were a bear market.

- In this framework, volatility clustering is generating by the tendency of high–volatility regimes being followed by high–volatility regimes and low–volatility–regimes being followed by low–volatility regimes.
Markov–switching Models


Markov–switching Models

- Markov–switching models have become very popular in economics and finance since Hamilton (1989).²

- In this model, it is assumed that the probability of being in regime $j$ at time $t$ depends on the regime at time $t - 1$ via the time–invariant transition probabilities $p_{ij}$, defined by

$$p_{ij} := \Pr(s_t = j|s_{t-1} = i), \quad j = 1, \ldots, k,$$

where

$$p_{ik} = 1 - \sum_{j=1}^{k-1} p_{ij}, \quad i = 1, \ldots, k.$$

• It will be useful to collect the transition probabilities in the $k \times k$ transition matrix $P$, 

$$
P = \begin{pmatrix}
p_{11} & p_{21} & \cdots & p_{k1} \\
p_{12} & p_{22} & \cdots & p_{k2} \\
\vdots & \vdots & \ddots & \vdots \\
p_{1k} & p_{2k} & \cdots & p_{kk}
\end{pmatrix}.
$$

• If we are in regime $j$ at time $t$, we anticipate that regime $j$ will continue with probability $p_{jj}$.

• Thus, if regimes are persistent, this will be reflected in rather large diagonal elements of the transition matrix $P$, which can also be characterized as the “staying probabilities”.
Basic Properties of the Mixing Process

• Assume that we are given a vector of regime probabilities at time $t$,

$$
\pi_t := [\pi_{1t}, \pi_{2t}, \ldots, \pi_{kt}]',
$$

where, as before,

$$
\pi_{jt} = \Pr_t(s_t = j), \quad j = 1, \ldots, k.
$$

• Recall the law of total probability: Let $A_1, \ldots, A_n$ be a partition of the sample space, i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$, and $\Pr(\bigcup_i A_i) = 1$, then

$$
\Pr(B) = \sum_i \Pr(A_i) \Pr(B|A_i).
$$
• Using this, and given the information in (6), the probability of being in state $j$ at time $t + 1$ is

$$
\Pr_{t-1}(s_{t+1} = j) = \pi_{j,t+1}
$$

$$
= \sum_{i=1}^{k} \Pr(s_t = i) \times \Pr(s_{t+1} = j | s_t = i)
$$

$$
= \sum_{i=1}^{k} \pi_{it} p_{ij}, \quad j = 1, \ldots, k.
$$

In terms of the transition matrix,

$$
\pi_{t+1} = P \pi_t.
$$

Iterating this, we get multi–step–ahead forecasts,

$$
\pi_{t+\tau} = P^{\tau} \pi_t.
$$

• Thus, the $\tau$–step transition probabilities are given by the respective elements of $P^{\tau}$.
That is, the probability that an observation from Regime $i$ will be followed $\tau$ periods later by an observation from Regime $j$ is given by the element in the $j$th row and $i$th column of the matrix $P^\tau$. 
• What happens if the forecast horizon becomes large, i.e., $\tau \to \infty$?

• Under rather general conditions (usually satisfied in practice),

$$\pi_\infty := \lim_{\tau \to \infty} P^\tau \pi_t$$

(7)

exists and is independent of the initial probability vector $\pi_t$.

• Then the values of $\pi_\infty = [\pi_{1,\infty}, \pi_{2,\infty}, \ldots, \pi_{k,\infty}]'$ are called the limiting, or unconditional, or log–run regime probabilities.

• These probabilities reflect the frequency of the regimes over longer time horizons.

• The convergence in (7) is due to the fact that

$$P_\infty := \lim_{\tau \to \infty} P^\tau = [\pi_\infty, \ldots, \pi_\infty] = \pi_\infty 1_k',$$

where

$$1_k = \left[\underbrace{1, 1, \ldots, 1}_{k \text{ times}}\right]'$$
• This shows that we also have

\[ P\pi_\infty = \pi_\infty, \]  

which shows that \( \pi_\infty \) is the stationary distribution of the chain.

• (8) also shows that \( \pi_\infty \) is a right eigenvector of the transition matrix corresponding to the eigenvalue 1.

• Moreover, all the other eigenvalues of \( P \) are smaller than 1 in absolute value for \( P^\tau \) to converge and for \( \pi_\infty \) to be unique.

• The largest subdominant eigenvalue (i.e., the largest eigenvalue with absolute value smaller than unity) may be viewed as a measure for the degree of overall regime persistence, i.e., for how rapidly \( P^\tau \) approaches \( P_\infty \) with \( \tau \).
• **Example**: Consider the two–state Markov chain with transition matrix

\[
P = \begin{bmatrix}
p_{11} & 1 - p_{22} \\
1 - p_{11} & p_{22}
\end{bmatrix} = \begin{bmatrix}
p & 1 - q \\
1 - p & q
\end{bmatrix}.
\]  

(9)

• The eigenvalues of this matrix are 1 and \(\delta = p + q - 1\), which measures the strength of persistence in the regime process.

• For example,

\[
\delta = 0 \iff p = 1 - q \iff 1 - p = q.
\]

• That is, the probability of staying in Regime 1 is identical to the probability of moving from Regime 2 to Regime 1.

• That is, the probability for the occurrence of Regime 1 does not depend on the current regime \(\Rightarrow\) no persistence; this is the independent mixture situation.

• The same applies to regime 2.
• The unconditional regime probabilities are given by

\[
\pi_{1,\infty} = \frac{1 - q}{2 - p - q}, \quad \pi_{2,\infty} = 1 - \pi_{1,\infty} = \frac{1 - p}{2 - p - q},
\]  

(10)

and it can be shown that

\[
P^\tau = \begin{bmatrix}
\pi_{1,\infty} + \delta^\tau \pi_{2,\infty} & (1 - \delta^\tau)\pi_{1,\infty} \\
(1 - \delta^\tau)\pi_{2,\infty} & \pi_{2,\infty} + \delta^\tau \pi_{1,\infty}
\end{bmatrix} 
= \begin{bmatrix}
\pi_{1,\infty} & \pi_{1,\infty} \\
\pi_{2,\infty} & \pi_{2,\infty}
\end{bmatrix} + \delta^\tau \begin{bmatrix}
\pi_{2,\infty} & -\pi_{1,\infty} \\
-\pi_{2,\infty} & \pi_{1,\infty}
\end{bmatrix},
\]  

(11)

so

\[
\lim_{\tau \to \infty} P^\tau = P_\infty = \begin{bmatrix}
\pi_{1,\infty} & \pi_{1,\infty} \\
\pi_{2,\infty} & \pi_{2,\infty}
\end{bmatrix} = [\pi_\infty, \pi_\infty],
\]  

(12)

and the speed of convergence is determined by the magnitude of \(\delta\).
Digression I

• Note (12) requires $|\delta| < 1$.

• For example, if $\delta = 1$,

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

(13)

which is a specific from of a reducible chain with two absorbing states.

• Another example of a reducible chain is

$$P = \begin{bmatrix} 1 & 0.5 \\ 0 & 0.5 \end{bmatrix}, \quad P^\tau = \begin{bmatrix} 1 & 1 - 0.5^\tau \\ 0 & 0.5^\tau \end{bmatrix} \xrightarrow{\tau \to \infty} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},$$

so $\pi_\infty = [1, 0]'$.

• This generalizes to higher–dimensional chains.

• We only consider irreducible chains.
Digression II

- If $\delta = -1$, the chain is *periodic* with

\[
P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}
\]

not converging, and the chain switches back and forth between the states.

- We only consider *irreducible* and *aperiodic* chains.
Expected Regime Durations

• The expected duration is also often of interest.

• That is, how many periods, on average, will Regime stay in Regime $j$?

• Once we are in Regime $j$, the duration $D_j \geq 1$ is geometrically distributed with probability $p_{jj}$, i.e.,

$$\Pr(D_j = d) = p_{jj}^{d-1}(1 - p_{jj}),$$

and so

$$E(D_j) = \frac{1}{1 - p_{jj}}.$$
Properties of the Return Process

• We focus on the univariate situation.

• The unconditional distribution of a return generated by a $k$–regime Markov–switching process of the form discussed so far, i.e.,

$$r_t = \mu_{s_t} + \sigma_{s_t} \eta_t, \quad \eta_t \overset{iid}{\sim} \mathcal{N}(0, 1),$$

is a $k$–component finite normal mixture with mixing weights $\pi_{1,\infty}, \pi_{2,\infty}, \ldots, \pi_{k,\infty}$.

• However, due to the regime–persistence, the process also captures the volatility clustering in the series.

• The dependence properties of the Markov chain $\{s_t\}$ are transferred to those of the returns.
• For example, the covariance function of the returns and squared returns are given by

\[
\text{Cov}(r_t, r_{t-\tau}) = \pi_{1,\infty} \pi_{2,\infty} \delta^\tau (\mu_1 - \mu_2)^2,
\]

and

\[
\text{Cov}(r^2_t, r^2_{t-\tau}) = \pi_{1,\infty} \pi_{2,\infty} \delta^\tau (\sigma^2_1 + \mu_1^2 - \sigma^2_2 - \mu_2^2)^2,
\]

respectively.

• These formulas can be extended to the situation of \( k \) regimes (and more general processes).\(^3\)

• Intuitively, if regimes are persistent, then high-return and low-volatility regimes (bull markets) tend to be followed by high-return and low-volatility regimes, respectively.

Inference about Market Regimes and the Likelihood Function

• As the market regimes are not observable, we can only use observed returns to make probability statements about the market’s past, current, or future regimes.

• To this end, we introduce, for each point of time, \( t \), a new (unobserved) \( k \)-dimensional random vector \( z_t = (z_{1t}, \ldots, z_{kt})' \), \( t = 1, \ldots, T \), with elements \( z_{jt} \) such that

\[
  z_{jt} = \begin{cases} 
    1 & \text{if } s_t = j \\
    0 & \text{if } s_t \neq j 
  \end{cases} \quad j = 1, \ldots, k.
\]

• That is, \( z_{jt} \) is one or zero according as to whether the return vector at time \( t \), \( r_t \), has been generated by the \( j \)th component of the mixture.

• Moreover, let \( R_\tau \) be the return history up to time \( \tau \), i.e., \( R_\tau = (r_1, \ldots, r_\tau) \), \( \tau = 1, \ldots, T \), and let \( \theta \) be the vector of model parameters.
• Then our probability inference of being in state \( j \) at time \( t \), based on the return history up to time \( \tau \), \( R_\tau \), and the parameter vector, \( \theta \), will be denoted by

\[
\Pr(z_{jt} = 1|R_\tau, \theta) = z_{jt|\tau},
\]

and

\[
z_{t|\tau} = (z_{1t|\tau}, \ldots, z_{kt|\tau})'.
\]

• Several cases are usually distinguished terminologically:

  – If \( \tau = t \), we have so-called **filtered regime inferences**. Then

\[
z_{t|t} = \begin{bmatrix}
\Pr(z_{1t} = 1|R_t, \theta) \\
\vdots \\
\Pr(z_{kt} = 1|R_t, \theta)
\end{bmatrix}
\]

  has our probability assessments that we are in a certain regime at time \( t \), given the observed time series up to time \( t \).

  – For \( t > \tau \), we have **predictive** regime inferences.

  – Finally, for \( \tau > t \), we have so-called **smoothed** regime inferences.
• Now suppose that

\[ z_{t|t-1} = \begin{bmatrix}
\Pr(z_{1t} = 1|R_{t-1}, \theta) \\
\vdots \\
\Pr(z_{kt} = 1|R_{t-1}, \theta)
\end{bmatrix} \]

is given.

• Then the conditional density of \( r_t \), given the information up to time \( t-1 \), can be calculated:

\[
f(r_t|R_{t-1}, \theta) = \sum_{j=1}^{k} z_{jt|t-1} f(r_t|s_t = j)
\]

\[ = 1_k'(z_{t|t-1} \odot \eta_t),\]

where

\[
\eta_t = \begin{bmatrix}
f(r_t|s_t = 1) \\
\vdots \\
f(r_t|s_t = k)
\end{bmatrix},
\]
and the Hadamard product denotes the elementwise multiplication of conformable matrices, i.e.,

\[
\begin{bmatrix}
  a_1 \\ a_2 \\ \vdots \\ a_n
\end{bmatrix} \odot \begin{bmatrix}
  b_1 \\ b_2 \\ \vdots \\ b_n
\end{bmatrix} = \begin{bmatrix}
  a_1 b_1 \\ a_2 b_2 \\ \vdots \\ a_n b_n
\end{bmatrix}.
\]
• Now define the joint density

\[ p(s_t = j, r_t | R_{t-1}, \theta) = z_{jt|t-1} f(r_t | s_t = j), \]

\[ j = 1, \ldots, k. \]

• This is to be interpreted in the sense that if we integrate over an interval \( A \), then

\[ \Pr(s_t = j, r_t \in A | R_{t-1}) = z_{jt|t-1} \int_A f(r_t | s_t = j) dr_t. \]

• If we divide this by \( f(r_t | R_{t-1}, \theta) \) given above, we get

\[
\frac{p(s_t = j, r_t | R_{t-1}, \theta)}{f(r_t | R_{t-1}, \theta)} = \Pr(s_t = j | r_t, R_{t-1}, \theta) \\
= \Pr(s_t = j | R_t, \theta) \\
= z_{jt|t} \\
\]

\[ j = 1, \ldots, k. \]
• Thus we have the filtered regime inferences

\[ z_{t|t} = \begin{bmatrix}
\Pr(z_{1t} = 1|R_t, \theta) \\
\vdots \\
\Pr(z_{kt} = 1|R_t, \theta)
\end{bmatrix} = \frac{z_{t|t-1} \odot \eta_t}{1_k'(z_{t|t-1} \odot \eta_t)}. \]

• From the filtered inferences, the one-step-ahead predictive regime inferences follow by multiplying by the transition matrix, i.e.,

\[ z_{t+1|t} = Pz_{t|t}. \]

Summarizing:

\[ z_{t|t} = \frac{z_{t|t-1} \odot \eta_t}{1_k'(z_{t|t-1} \odot \eta_t)} \quad (14) \]

\[ z_{t+1|t} = Pz_{t|t} \quad (15) \]
To initialize the recursion given by (14) and (15), a vector of initial probabilities, $\varrho$, with elements

$$\varrho_j := z_{1|0} = \Pr(s_1 = j), \quad j = 1, \ldots, k,$$

needs to be either fixed or estimated.

An obvious choice is to put $\varrho = \pi_\infty$, the unconditional regime probabilities.

Another variant is to treat the $\varrho_j$'s as parameters to be estimated.

If the EM algorithm is used for parameter estimation, then a natural estimate for the $\varrho_j$'s follows. See Hamilton (1994), Chapter 22, for discussion of the EM algorithm for Markov–switching models, which is a generalization of (and very similar to) the EM algorithm for independent mixture models.

For the EM algorithm, we need the *smoothed* probabilities, the derivation of which is somewhat tedious, see the appendix of Ch. 22 of Hamilton (1994).
Having specified an initial probability, the algorithm (14) and (15) can be used to compute the likelihood function of the sample of observed data, $R_T$, at a value of the parameter vector, $\theta$, as

$$\log L(\theta|R_T) = \sum_{t=1}^{T} \log f(r_t|R_{t-1}, \theta)$$

$$= \sum_{t=1}^{T} \log [1_k'(z_{t|t-1} \odot \eta_{t})],$$

which can be maximized numerically by standard algorithms.

However, in the context of MS–models, the EM algorithm provides a very attractive tool for doing the estimation, in particular if we allow for complex regime–specific dynamics and exogenous variables.

Also, the one–step–ahead predictive regime inferences are easily generalized to $\tau$–step–ahead inferences for calculating multi–step–ahead forecasts, i.e.,

$$\hat{z}_{t+\tau|t} = P^\tau \hat{z}_{t|t}.$$
Application to stock returns

- Consider three index return series of major European stock markets, namely, daily returns of the MSCI indices for France, Germany and the United Kingdom (UK) from January 1990 to April 2009 ($T = 4993$ observations).

- All returns are measured in Euro.

- Continuously compounded percentage returns are used, i.e., the return of index $i$ at time $t$ is

  \[ r_{it} = 100 \times \log \left( \frac{I_{it}}{I_{i,t-1}} \right), \]

  where $I_{it}$ denotes the level of index $i$ at time $t$

- The return vector at time $t$ is defined as $r_t = (r_{1t}, r_{2t}, r_{3t})'$, where $r_{1t}$, $r_{2t}$, and $r_{3t}$ are the time–$t$ returns of the indices for France, Germany, and the UK, respectively.
• In addition to the three country indices, the returns of an equally weighted portfolio of the three indices are considered, calculated as the simple average of the individual indices’ returns.

• In this regard, recall that, due to our use of continuously compounded returns, the linear relation between the individual returns and the portfolio return is only an approximation.

• In case of a (conditional) normal mixture distribution, the return on a portfolio, $r^p_t$, i.e., $r^p_t = w'r_t$, where $w$ is an $N \times 1$ vector of portfolio weights, has a $k$–component univariate normal mixture distribution, i.e.,

$$f(r^p_t) = \sum_{j=1}^{k} \frac{\pi_{jt}}{\sqrt{2\pi}\sigma^*_j} \exp \left\{ -\frac{1}{2} \left( \frac{r^p_t - \mu^*_j}{\sigma^*_j} \right)^2 \right\},$$

(16)

where $\mu^*_j = w'\mu_j$, and $\sigma^*_j = \sqrt{w'\Sigma_j w}$. 

33
• Gaussian Markov–switching models have often been successfully applied to monthly return data.

• However, in applications to daily returns, use of Gaussian components may not be appropriate.

• Namely, although a normal mixture can accommodate a considerable degree of excess kurtosis, its tails will eventually decay in a Gaussian manner, i.e., squared exponentially, whereas the tails of financial returns are often found to be more realistically described by a power–law behavior due to the regular occurrence of rather large realizations.
• This feature, which is of particular importance for risk management, can be captured by replacing the Gaussian with a Student’s $t$ distribution, i.e., by replacing the multivariate Gaussian with

$$f_j(x; \theta_j) = f(x; \mu_j, \Sigma_j, \nu_j)$$

$$= \frac{\Gamma \left( \frac{\nu_j+p}{2} \right)}{\Gamma(\nu_j/2)(\pi\nu_j)^{p/2}\sqrt{\mid\Sigma_j\mid}} \left\{ 1 + \frac{(x - \mu_j)\Sigma_j^{-1}(x - \mu_j)}{\nu_j} \right\}^{-(\nu_j+p)/2},$$

$j = 1, \ldots, k$, where $\nu_j > 0$ is the degrees of freedom parameter characterizing the tail of the $j$th component.

• The interpretation of the other parameters is similar to the Gaussian with the exception that, for $\nu_j > 2$, the covariance matrix of component $j$ is $\frac{\nu_j}{\nu_j - 2}\Sigma_j$ rather than just $\Sigma_j$.

• McLachlan and Peel (2000, Ch. 7) and Peel and McLachlan (2000) show how mixtures of $t$ distributions can be estimated via the EM algorithm.\(^4\)

---

• Portfolios then follow univariate Student’s $t$ distributions, similar to Equation (16).

\footnote{For a detailed discussion of the properties of the multivariate $t$ distribution, see, e.g., Appendix B in Zellner (1971): An Introduction to Bayesian Inference in Econometrics, Wiley.}
Model specification

• We first estimate the models over the (approximately) first ten years of data, i.e., the first 2500 observations (ranging from January 1990 to the end of August 1999), while the remaining observations are retained for the computation and backtesting of out–of–sample Value–at–Risk measures.

• To keep the analysis manageable, we have to select an upper bound for the number of regimes.

• The BIC is often used in the context of mixture models, given by

\[ \text{BIC} = -2 \times \log L + K \log T, \]

where \( \log L \) is the value of the maximized log–likelihood achieved by a given model, \( K \) is the number of parameters and \( T \) is the sample size.

• Tables 1 and 2 report, for the in–sample period from January 1990 to August 1999, the values of the maximized log–likelihood and the BIC for models with \( k \leq 4 \) for univariate and multivariate models, respectively.
• Considering the results for univariate models in Table 1, we observe that, according to BIC, a Gaussian mixture with $k = 3$ is preferred for France, whereas Student’s $t$ mixtures with two and three regimes are preferred for the UK and Germany, respectively, with $\nu_1 = \nu_2$ imposed in the former case.

• Among the multivariate models, as reported in Table 2, a Student’s $t$ mixture with three regimes performs best.

• Moreover, in no case is a model with $k = 4$ preferred over a model with $k = 3$, which leads us to restrict the analysis to specifications with $k \leq 3$. 
Table 1: Likelihood–based goodness–of–fit measures for the in–sample period (1990-1999): Univariate models

<table>
<thead>
<tr>
<th></th>
<th>France</th>
<th></th>
<th>Germany</th>
<th></th>
<th>United Kingdom</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K$</td>
<td>$\log L$</td>
<td>BIC</td>
<td>$\log L$</td>
<td>BIC</td>
<td>$\log L$</td>
</tr>
<tr>
<td>Gaussian models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 1$</td>
<td>2</td>
<td>-3880.7</td>
<td>7777.0</td>
<td>-3998.2</td>
<td>8012.1</td>
<td>-3665.1</td>
</tr>
<tr>
<td>$k = 2$</td>
<td>6</td>
<td>-3734.2</td>
<td>7515.3</td>
<td>-3694.3</td>
<td>7435.5</td>
<td>-3531.8</td>
</tr>
<tr>
<td>$k = 3$</td>
<td>12</td>
<td>-3706.1</td>
<td>7506.2</td>
<td>-3626.6</td>
<td>7347.1</td>
<td></td>
</tr>
<tr>
<td>$k = 4$</td>
<td>20</td>
<td>-3699.7</td>
<td>7556.0</td>
<td>-3617.8</td>
<td>7392.1</td>
<td>-3500.7</td>
</tr>
<tr>
<td>Student’s $t$ models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 1$</td>
<td>3</td>
<td>-3797.4</td>
<td>7618.3</td>
<td>-3816.8</td>
<td>7657.1</td>
<td>-3592.1</td>
</tr>
<tr>
<td>$k = 2$, $\nu_1 = \nu_2$</td>
<td>7</td>
<td>-3729.9</td>
<td>7514.6</td>
<td>-3656.4</td>
<td>7367.6</td>
<td>-3523.8</td>
</tr>
<tr>
<td>$k = 2$, $\nu_1 \neq \nu_2$</td>
<td>8</td>
<td>-3729.9</td>
<td>7522.3</td>
<td>-3656.1</td>
<td>7374.9</td>
<td>-3522.0</td>
</tr>
<tr>
<td>$k = 3$</td>
<td>13</td>
<td>-3705.1</td>
<td>7511.9</td>
<td>-3617.7</td>
<td><strong>7337.0</strong></td>
<td>-3503.7</td>
</tr>
<tr>
<td>$k = 4$</td>
<td>21</td>
<td>-3699.0</td>
<td>7562.3</td>
<td>-3607.2</td>
<td>7378.7</td>
<td>-3498.8</td>
</tr>
</tbody>
</table>

Smaller values of BIC are preferred, and boldface entries indicate the best model according to BIC.

<table>
<thead>
<tr>
<th>Gaussian</th>
<th>( k = 1 )</th>
<th>( k = 2 )</th>
<th>( k = 3 )</th>
<th>( k = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K )</td>
<td>9</td>
<td>20</td>
<td>33</td>
<td>48</td>
</tr>
<tr>
<td>( \log L )</td>
<td>(-10229)</td>
<td>(-9788.6)</td>
<td>(-9666.7)</td>
<td>(-9619.6)</td>
</tr>
<tr>
<td>BIC</td>
<td>20530</td>
<td>19734</td>
<td>19592</td>
<td>19615</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student's t</th>
<th>( k = 1 )</th>
<th>( k = 2 ), ( \nu_1 = \nu_2 )</th>
<th>( k = 3 )</th>
<th>( k = 4 )</th>
<th>( \nu_1 \neq \nu_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K )</td>
<td>10</td>
<td>21</td>
<td>22</td>
<td>34</td>
<td>49</td>
</tr>
<tr>
<td>( \log L )</td>
<td>(-9879.1)</td>
<td>(-9689.2)</td>
<td>(-9686.9)</td>
<td>(-9629.7)</td>
<td>(-9594.9)</td>
</tr>
<tr>
<td>BIC</td>
<td>19836</td>
<td>19543</td>
<td>19546</td>
<td><strong>19525</strong></td>
<td>19573</td>
</tr>
</tbody>
</table>

Smaller values of BIC are preferred, and boldface entries indicate the best model according to BIC.
Parameter estimates for multivariate models

• Estimation results are reported in Tables 3 and 4 for models with Gaussian and Student’s $t$ distributions, respectively, where the regimes have been ordered with respect to a declining stationary regime probability, i.e., $\pi_1 > \pi_2 > \pi_3$.

• For purpose of comparison, we also report the results for single–regime models, where $k = 1$.

• For the models with $k = 2$ regimes, the first regime is, for Gaussian as well as Student’s $t$ components, characterized by a higher expected return, lower variances and lower correlations, implying that it can be interpreted as the bull market regime, whereas the second regime represents the bear market.

• However, there are also some differences between the results for Gaussian and Student’s $t$ components.
• Namely, in the Gaussian model, the bear market regime is less persistent (the estimated “staying probability” $p_{22}$ is lower) and has a lower stationary probability $\pi_2$ than in the Student’s $t$ model.

• This can be related to the observation that, in the former model, the bear market is more “extreme” than in the latter.

• In particular, with the exception of the UK, expected bear market returns are negative in the Gaussian model, whereas they are positive in the $t$ model, and the ratios of bear market variances to bull market variances are 3.6, 5.4, and 3.3 in the Gaussian model for the French, German, and UK market, respectively, whereas they are only 2.1, 4.2, and 2.1 in the $t$ model.

• On the other hand, the differences between the correlations are more distinctive in the model based on Student’s $t$ regimes.

• These observations can be explained by the pronounced nonnormality of the regime densities implied by the estimated degrees of freedom parameter $\hat{\nu} = 7.9$. 
• In particular, a Gaussian regime–switching model will detect regime switches more frequently due to an untypical observation in an otherwise low– or high–volatility period, whereas Student’s t components, as a result of their higher peaks and fatter tails, can better accommodate such realizations within a given regime, leading to regimes that are more persistent.

• Moreover, in normal mixture models, excess kurtosis is exclusively generated by the ratio of the component variances.

• On the other hand, in Student’s t mixtures, there is an additional source of leptokurtosis, namely conditional (within–regime) kurtosis, so that, to achieve the same degree of overall kurtosis, the differences in volatility across regimes can be more moderate.

• Similarly, by using a (more robust) t distribution, extreme observations are given less weight in calculating the parameter estimates, resulting in the aforementioned differences in the bear market expected returns.
Table 3: Gaussian Markov mixture parameter estimates.

<table>
<thead>
<tr>
<th></th>
<th>$k = 1$</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\pi}_1$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}'_1$</td>
<td>$\begin{pmatrix} 0.042 &amp; 0.040 &amp; 0.045 \ 0.023 &amp; 0.024 &amp; 0.021 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.069 &amp; 0.094 &amp; 0.048 \ 0.022 &amp; 0.021 &amp; 0.021 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.047 &amp; 0.049 &amp; 0.044 \ 0.036 &amp; 0.038 &amp; 0.033 \end{pmatrix}$</td>
</tr>
<tr>
<td>$\hat{\Sigma}_{1/\hat{R}_1}$</td>
<td>$\begin{pmatrix} 0.616 &amp; 1.434 &amp; 0.638 \ 0.041 &amp; 0.028 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.515 &amp; 0.737 &amp; 0.346 \ 0.037 &amp; 0.024 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.617 &amp; 1.447 &amp; 0.643 \ 0.098 &amp; 0.058 \end{pmatrix}$</td>
</tr>
<tr>
<td>$\hat{\pi}_2$</td>
<td>0</td>
<td>0.208</td>
<td>0.410</td>
</tr>
<tr>
<td>$\hat{\mu}'_2$</td>
<td>$\begin{pmatrix} -0.057 &amp; -0.162 &amp; 0.036 \ 0.082 &amp; 0.094 &amp; 0.073 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.076 &amp; 0.088 &amp; 0.058 \ 0.031 &amp; 0.024 &amp; 0.028 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.428 &amp; 0.402 &amp; 0.205 \ 0.031 &amp; 0.020 \end{pmatrix}$</td>
</tr>
<tr>
<td>$\hat{\Sigma}_{2/\hat{R}_2}$</td>
<td>$\begin{pmatrix} 3.027 &amp; 2.450 &amp; 1.827 \ 0.256 &amp; 0.239 &amp; 0.178 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.705 &amp; 3.992 &amp; 1.730 \ 0.332 &amp; 0.186 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.428 &amp; 0.402 &amp; 0.205 \ 0.031 &amp; 0.020 \end{pmatrix}$</td>
</tr>
<tr>
<td>$\hat{\pi}_3$</td>
<td>0</td>
<td>0</td>
<td>0.080</td>
</tr>
<tr>
<td>$\hat{\mu}'_3$</td>
<td>$\begin{pmatrix} -0.152 &amp; -0.254 &amp; -0.008 \ 0.175 &amp; 0.192 &amp; 0.154 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 5.136 &amp; 4.113 &amp; 3.009 \ 0.624 &amp; 0.660 &amp; 0.436 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.716 &amp; 6.427 &amp; 2.758 \ 0.800 &amp; 0.463 \end{pmatrix}$</td>
</tr>
<tr>
<td>$\hat{\Sigma}_{3/\hat{R}_3}$</td>
<td>$\begin{pmatrix} 0.716 &amp; 6.427 &amp; 2.758 \ 0.800 &amp; 0.463 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.671 &amp; 0.550 &amp; 3.917 \ 0.465 \end{pmatrix}$</td>
<td>$\begin{pmatrix} 0.960 &amp; 0.010 &amp; 0.204 \ 0.011 &amp; 0.007 &amp; 0.056 \end{pmatrix}$</td>
</tr>
<tr>
<td>$\hat{P}$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\delta}$</td>
<td></td>
<td>0.865</td>
<td>0.965</td>
</tr>
</tbody>
</table>
**Table 4: Student's t Markov mixture parameter estimates.**

<table>
<thead>
<tr>
<th></th>
<th>$k = 1$</th>
<th>$k = 2$</th>
<th>$k = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\pi}_1$</td>
<td>1</td>
<td>0.581</td>
<td>0.505</td>
</tr>
<tr>
<td>$\hat{\mu}_1'$</td>
<td>(0.058 0.068 0.053)</td>
<td>(0.059 0.083 0.060)</td>
<td>(0.074 0.089 0.054)</td>
</tr>
<tr>
<td></td>
<td>(0.020 0.020 0.019)</td>
<td>(0.024 0.020 0.022)</td>
<td>(0.031 0.033 0.029)</td>
</tr>
<tr>
<td>$\hat{\Sigma}_1/\hat{R}_1$</td>
<td>(0.845 0.490 0.504)</td>
<td>(0.662 0.214 0.356)</td>
<td>(0.962 0.542 0.586)</td>
</tr>
<tr>
<td></td>
<td>(0.031 0.024 0.023)</td>
<td>(0.031 0.020 0.023)</td>
<td>(0.049 0.041 0.037)</td>
</tr>
<tr>
<td>$\hat{\pi}_2$</td>
<td>0</td>
<td>0.419</td>
<td>0.364</td>
</tr>
<tr>
<td>$\hat{\mu}_2'$</td>
<td>(0.039 0.020 0.033)</td>
<td>(0.039 0.020 0.033)</td>
<td>(0.050 0.066 0.054)</td>
</tr>
<tr>
<td></td>
<td>(0.041 0.047 0.037)</td>
<td>(0.041 0.047 0.037)</td>
<td>(0.029 0.023 0.028)</td>
</tr>
<tr>
<td>$\hat{\Sigma}_2/\hat{R}_2$</td>
<td>(1.375 1.147 0.893)</td>
<td>(0.717 1.858 0.868)</td>
<td>(0.717 1.858 0.868)</td>
</tr>
<tr>
<td></td>
<td>(0.081 0.081 0.060)</td>
<td>(0.114 0.066 0.066)</td>
<td>(0.114 0.066 0.066)</td>
</tr>
<tr>
<td>$\hat{\pi}_3$</td>
<td>0</td>
<td>0</td>
<td>0.131</td>
</tr>
<tr>
<td>$\hat{\mu}_3'$</td>
<td>(0.704 0.589 1.170)</td>
<td>(0.704 0.589 1.170)</td>
<td>(0.704 0.589 1.170)</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.067)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>$\hat{\Sigma}_3/\hat{R}_3$</td>
<td>(0.704 0.589 1.170)</td>
<td>(0.704 0.589 1.170)</td>
<td>(0.704 0.589 1.170)</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.067)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>$\hat{\nu}$</td>
<td>5.813 (0.369)</td>
<td>7.919 (1.010)</td>
<td>10.55 (1.272)</td>
</tr>
<tr>
<td>$\hat{\lambda}$</td>
<td>(0.992 0.010)</td>
<td>(0.992 0.010)</td>
<td>(0.992 0.010)</td>
</tr>
<tr>
<td></td>
<td>(0.003 0.004)</td>
<td>(0.003 0.004)</td>
<td>(0.003 0.004)</td>
</tr>
<tr>
<td>$\hat{\delta}$</td>
<td>0.982</td>
<td>0.982</td>
<td>47</td>
</tr>
</tbody>
</table>

(All parameter estimates are rounded to two decimal places.)
Backtesting Value–at–Risk (VaR) measures

- Both in industry and in academia, Value–at–Risk (VaR) is a widely employed measure to characterize the downside risk of a financial position.

- VaR can be defined in terms of the conditional quantile of the portfolio return distribution for a given horizon (typically a day or a week) and a given (typically small) shortfall probability.

- To be more precise, consider a time series of portfolio returns, $r_t$, and an associated series of ex–ante VaR measures with shortfall probability $\xi$, $\text{VaR}_t(\xi)$.

- The $\text{VaR}_t(\xi)$ implied by a model $\mathcal{M}$ is defined by

$$ F_{t-1}^\mathcal{M}(\text{VaR}_t(\xi)) = \xi, \quad (17) $$

where $F_{t-1}^\mathcal{M}$ is the (conditional) cumulative distribution function (cdf) derived from model $\mathcal{M}$ using the information up to time $t - 1$. 
• Thus, economically, $\text{VaR}_t(\xi)$ is defined so that, over the next period, the probability that the portfolio suffers a loss larger than the VaR is $100 \times \xi\%$.

• For a nominal VaR shortfall probability $\xi$ and a correctly specified VaR model, we expect $100 \times \xi\%$ of the observed return values not to exceed the respective VaR forecast.

• To test the models’ suitability for calculating accurate ex-ante VaR measures, we say that a violation or hit occurs at time $t$ if $r_t < \text{VaR}_t$ and define the binary sequence

$$I_t = \begin{cases} 1, & \text{if } r_t < \text{VaR}_t, \\ 0, & \text{if } r_t \geq \text{VaR}_t. \end{cases}$$

(18)

• Then the empirical shortfall probability is $\hat{\xi} = x/T$, where $x = \sum_{t=1}^{T} I_t$ is the number of observed violations, and $T$ is the number of forecasts evaluated.
• If \( \hat{\xi} \) is significantly higher (less) than \( \xi \), then the model under study tends to underestimate (overestimate) the risk of the financial position.

• If the model is correctly specified, the hit sequence is a sample of size \( T \) from the Bernoulli distribution with parameter \( \xi \), with pdf

\[
p(I_t; \xi) = \xi^I_t(1 - \xi)^{1-I_t},
\]

and the likelihood of the sample is

\[
L(\xi) = \xi^{\sum_{t=1}^{T} I_t}(1 - \xi)^{T - \sum_{t=1}^{T} I_t} = \xi^x(1 - \xi)^{T-x},
\]

with log–likelihood

\[
\log L(\xi) = x \log \xi + (T - x) \log(1 - \xi).
\]

• The maximum likelihood estimator is obtained via

\[
\frac{\partial \log L(\xi)}{\partial \xi} = \frac{x}{\xi} - \frac{T - x}{1 - \xi} = 0 \Rightarrow \hat{\xi} = \frac{x}{T}.
\]
• The likelihood ratio test statistic is two times the unrestricted log-likelihood \( \log L(\hat{\xi}) = x \log(x/T) + (T - x) \log\{(T - x)/x\} \) minus the log-likelihood under the null that the actual shortfall probability is equal to the nominal shortfall probability \( \xi \).

• To formally test whether a model correctly estimates the VaR, that is, whether the empirical shortfall probability, \( \hat{\xi} \), is statistically indistinguishable from the nominal shortfall probability, \( \xi \), we use the likelihood ratio test statistic,

\[
L_{RT_1} = -2\{x \log(\xi/\hat{\xi}) + (T - x) \log[(1 - \xi)/(1 - \hat{\xi})]\} \overset{asy}{\sim} \chi^2(1). \tag{23}
\]
• We calculate one–step–ahead out–of–sample VaR measures and consider the VaR levels $\xi = 0.0025, 0.005, 0.01, 0.025, 0.05,$ and $0.1$, as defined in (17).

• For each model, the parameter estimates are updated every week (5 trading days) employing a moving window of data, i.e., using the most recent 2500 observations in the sample.

• In this manner, we obtain, for each model and return series, 2493 one–step–ahead out–of–sample VaR measures.

• Note that the conditional density of $r_t$ implied by a Gaussian mixture model at time $t - 1$ is a $k$–component normal mixture,

$$f(r_t|R_{t-1}, \theta) = \sum_{j=1}^{k} z_{jt|t-1} f(r_t|s_t = j),$$

(24)

where the conditional mixing weights $z_{jt|t-1}$ are determined via the recursion (14)–(15).

• The component densities in (24) may be either normal or Student’s $t$. 

52
• Our backtesting sample spans the period from September 1999 to April 2009, and thus covers the recent financial crisis which seriously affected the stock markets since the beginning of 2008.

• The sharp decline of the markets in 2008 was accompanied by a burst of volatility that appears rather extreme even in view of the former high–volatility periods covered by the sample under study.
• In view of these events, we first perform the backtesting for the time span September 1999 to December 2007 (corresponding to the first 2159 observations of the backtesting sample).

• Subsequently, we do the tests for the entire out–of–sample period.

• By following this procedure, it is possible to assess whether a model’s (potential) failure is chronic or can be attributed mainly to a clustering of violations due to the outbreak of unprecedented volatility during the turmoil.
Results for the period September 1999 to December 2007

• We first observe that the single–regime models, both Gaussian and Student’s $t$, are not adequate for all markets.

• Considering the multi–regime specifications, the most striking result is the superiority of VaR measures calculated on the basis of univariate as compared to multivariate models.

• Comparing the univariate multi–regime specifications, we find that the two–regime Student’s $t$ processes perform best overall, both with $\nu_1 = \nu_2$ and $\nu_1 \neq \nu_2$.

• The hypothesis of correct coverage of VaR measures implied by these models cannot be rejected with the exception of the higher VaR levels ($\xi \geq 0.025$) for the German market.

• The performance of the Student’s $t$ models with $k = 3$ regimes and the Gaussian models is less consistent.
Table 5: Evaluation of Value–at–Risk measures for the French market, September 1999 to December 2007: Gaussian models

<table>
<thead>
<tr>
<th>100 × ξ</th>
<th>0.25</th>
<th>0.5</th>
<th>1</th>
<th>2.5</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model Coverage (100 × ˆξ)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gaussian</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k = 1</td>
<td>1.575***</td>
<td>1.899***</td>
<td>2.409***</td>
<td>3.983***</td>
<td>6.346***</td>
<td>9.680</td>
</tr>
<tr>
<td>uni., k = 2</td>
<td>0.324</td>
<td>0.602</td>
<td>1.297</td>
<td>3.057</td>
<td>5.697</td>
<td>9.680</td>
</tr>
<tr>
<td>uni., k = 3</td>
<td>0.463*</td>
<td>0.880**</td>
<td>1.621***</td>
<td>3.566***</td>
<td>6.716***</td>
<td>11.58**</td>
</tr>
<tr>
<td>multi., k = 2</td>
<td>0.602***</td>
<td>1.019***</td>
<td>1.945***</td>
<td>3.891***</td>
<td>6.531***</td>
<td>11.35**</td>
</tr>
<tr>
<td>multi., k = 3</td>
<td>0.648***</td>
<td>1.112***</td>
<td>1.806***</td>
<td>4.122***</td>
<td>7.179***</td>
<td>11.90***</td>
</tr>
<tr>
<td><strong>Student’s t</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uni., k = 1</td>
<td>0.463*</td>
<td>1.065***</td>
<td>1.853***</td>
<td>3.520***</td>
<td>7.133***</td>
<td>11.63**</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ = ν₂</td>
<td>0.278</td>
<td>0.463</td>
<td>0.973</td>
<td>2.779</td>
<td>5.697</td>
<td>10.05</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ ≠ ν₂</td>
<td>0.278</td>
<td>0.509</td>
<td>0.926</td>
<td>2.733</td>
<td>5.604</td>
<td>9.912</td>
</tr>
<tr>
<td>uni., k = 3</td>
<td>0.371</td>
<td>0.695</td>
<td>1.390*</td>
<td>3.520***</td>
<td>6.577***</td>
<td>11.90***</td>
</tr>
<tr>
<td>multi., k = 1</td>
<td>0.695***</td>
<td>1.297***</td>
<td>1.945***</td>
<td>3.891***</td>
<td>7.133***</td>
<td>11.72***</td>
</tr>
<tr>
<td>multi., k = 2, ν₁ = ν₂</td>
<td>0.509**</td>
<td>0.926**</td>
<td>1.806***</td>
<td>3.844***</td>
<td>6.577***</td>
<td>11.53**</td>
</tr>
<tr>
<td>multi., k = 2, ν₁ ≠ ν₂</td>
<td>0.463*</td>
<td>0.834**</td>
<td>1.760***</td>
<td>3.844***</td>
<td>6.531***</td>
<td>11.58**</td>
</tr>
<tr>
<td>multi., k = 3</td>
<td>0.463*</td>
<td>0.741</td>
<td>1.390*</td>
<td>3.335**</td>
<td>6.207**</td>
<td>11.26*</td>
</tr>
</tbody>
</table>

Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.
Table 6: Evaluation of Value–at–Risk measures for the German market, September 1999 to December 2007: Gaussian models

<table>
<thead>
<tr>
<th>$100 \times \xi$</th>
<th>0.25</th>
<th>0.5</th>
<th>1</th>
<th>2.5</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Coverage $(100 \times \hat{\xi})$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 1$</td>
<td>1.621***</td>
<td>2.362***</td>
<td>3.011***</td>
<td>4.539***</td>
<td>6.531***</td>
<td>9.958</td>
</tr>
<tr>
<td>uni., $k = 2$</td>
<td>0.371</td>
<td>0.741</td>
<td>1.343</td>
<td>3.150*</td>
<td>6.160**</td>
<td>10.89</td>
</tr>
<tr>
<td>uni., $k = 3$</td>
<td>0.185</td>
<td>0.509</td>
<td>1.158</td>
<td>3.103*</td>
<td>6.623***</td>
<td>11.53**</td>
</tr>
<tr>
<td>multi., $k = 2$</td>
<td>0.417</td>
<td>0.880**</td>
<td>1.853***</td>
<td>4.261***</td>
<td>7.133***</td>
<td>11.86***</td>
</tr>
<tr>
<td>multi., $k = 3$</td>
<td>0.648***</td>
<td>0.834**</td>
<td>1.806***</td>
<td>4.122***</td>
<td>7.272***</td>
<td>12.32***</td>
</tr>
<tr>
<td>Student’s $t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uni., $k = 1$</td>
<td>0.278</td>
<td>0.787*</td>
<td>1.945***</td>
<td>4.308***</td>
<td>7.596***</td>
<td>13.25***</td>
</tr>
<tr>
<td>uni., $k = 2$, $\nu_1 = \nu_2$</td>
<td>0.139</td>
<td>0.556</td>
<td>1.158</td>
<td>3.103*</td>
<td>6.438***</td>
<td>11.35**</td>
</tr>
<tr>
<td>uni., $k = 2$, $\nu_1 \neq \nu_2$</td>
<td>0.139</td>
<td>0.556</td>
<td>1.112</td>
<td>3.057</td>
<td>6.207**</td>
<td>11.30**</td>
</tr>
<tr>
<td>uni., $k = 3$</td>
<td>0.139</td>
<td>0.417</td>
<td>0.973</td>
<td>2.918</td>
<td>6.531***</td>
<td>11.90***</td>
</tr>
<tr>
<td>multi., $k = 1$</td>
<td>0.787***</td>
<td>1.528***</td>
<td>2.686***</td>
<td>4.863***</td>
<td>7.596***</td>
<td>12.51***</td>
</tr>
<tr>
<td>multi., $k = 2$, $\nu_1 = \nu_2$</td>
<td>0.324</td>
<td>0.695</td>
<td>1.436*</td>
<td>3.705***</td>
<td>6.716***</td>
<td>12.23***</td>
</tr>
<tr>
<td>multi., $k = 2$, $\nu_1 \neq \nu_2$</td>
<td>0.278</td>
<td>0.741</td>
<td>1.436*</td>
<td>3.659***</td>
<td>6.670***</td>
<td>12.23***</td>
</tr>
<tr>
<td>multi., $k = 3$</td>
<td>0.278</td>
<td>0.556</td>
<td>1.065</td>
<td>3.474***</td>
<td>6.346***</td>
<td>11.35**</td>
</tr>
</tbody>
</table>

Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.
<table>
<thead>
<tr>
<th>$100 \times \xi$</th>
<th>0.25</th>
<th>0.5</th>
<th>1</th>
<th>2.5</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 1$</td>
<td>1.158***</td>
<td>1.806***</td>
<td>2.177***</td>
<td>3.613***</td>
<td>5.651</td>
<td>9.495</td>
</tr>
<tr>
<td>uni., $k = 2$</td>
<td>0.463*</td>
<td>0.787*</td>
<td>1.390*</td>
<td>2.779</td>
<td>5.095</td>
<td>10.01</td>
</tr>
<tr>
<td>uni., $k = 3$</td>
<td>0.371</td>
<td>0.787*</td>
<td>1.390*</td>
<td>3.150*</td>
<td>5.049</td>
<td>10.51</td>
</tr>
<tr>
<td>multi., $k = 2$</td>
<td>0.648***</td>
<td>1.065***</td>
<td>1.528**</td>
<td>3.150*</td>
<td>5.558</td>
<td>10.61</td>
</tr>
<tr>
<td>multi., $k = 3$</td>
<td>0.602***</td>
<td>0.880**</td>
<td>1.436*</td>
<td>3.150*</td>
<td>6.068**</td>
<td>11.72***</td>
</tr>
<tr>
<td>Student’s $t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uni., $k = 1$</td>
<td>0.417</td>
<td>0.834**</td>
<td>1.714***</td>
<td>3.566***</td>
<td>6.021**</td>
<td>10.93</td>
</tr>
<tr>
<td>uni., $k = 2$, $\nu_1 = \nu_2$</td>
<td>0.371</td>
<td>0.695</td>
<td>1.343</td>
<td>2.872</td>
<td>4.771</td>
<td>10.47</td>
</tr>
<tr>
<td>uni., $k = 2$, $\nu_1 \neq \nu_2$</td>
<td>0.371</td>
<td>0.741</td>
<td>1.297</td>
<td>2.918</td>
<td>4.817</td>
<td>10.47</td>
</tr>
<tr>
<td>uni., $k = 3$</td>
<td>0.324</td>
<td>0.695</td>
<td>1.297</td>
<td>3.103*</td>
<td>5.280</td>
<td>11.07</td>
</tr>
<tr>
<td>multi., $k = 1$</td>
<td>0.463*</td>
<td>0.787*</td>
<td>1.760***</td>
<td>3.520***</td>
<td>6.160**</td>
<td>11.02</td>
</tr>
<tr>
<td>multi., $k = 2$, $\nu_1 = \nu_2$</td>
<td>0.463*</td>
<td>0.787*</td>
<td>1.343</td>
<td>2.964</td>
<td>5.604</td>
<td>10.70</td>
</tr>
<tr>
<td>multi., $k = 2$, $\nu_1 \neq \nu_2$</td>
<td>0.463*</td>
<td>0.787*</td>
<td>1.390*</td>
<td>2.918</td>
<td>5.604</td>
<td>10.75</td>
</tr>
<tr>
<td>multi., $k = 3$</td>
<td>0.509**</td>
<td>0.787*</td>
<td>1.297</td>
<td>2.594</td>
<td>5.373</td>
<td>10.65</td>
</tr>
</tbody>
</table>

Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.
Table 8: Evaluation of Value–at–Risk measures for equally weighted portfolio, September 1999 to December 2007: Gaussian models

<table>
<thead>
<tr>
<th>$100 \times \xi$</th>
<th>0.25</th>
<th>0.5</th>
<th>1</th>
<th>2.5</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model Coverage (100 \times \hat{\xi})</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gaussian</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$k = 1$</td>
<td>1.575***</td>
<td>1.945***</td>
<td>3.103***</td>
<td>4.771***</td>
<td>7.226***</td>
<td>10.42</td>
</tr>
<tr>
<td>uni., $k = 2$</td>
<td>0.509**</td>
<td>0.880**</td>
<td>1.343</td>
<td>3.242**</td>
<td>6.114**</td>
<td>10.05</td>
</tr>
<tr>
<td>uni., $k = 3$</td>
<td>0.417</td>
<td>0.695</td>
<td>1.019</td>
<td>3.103*</td>
<td>6.392***</td>
<td>11.39**</td>
</tr>
<tr>
<td>multi., $k = 2$</td>
<td>0.648***</td>
<td>1.204***</td>
<td>2.270***</td>
<td>4.400***</td>
<td>7.318***</td>
<td>12.18***</td>
</tr>
<tr>
<td>multi., $k = 3$</td>
<td>0.787***</td>
<td>1.112***</td>
<td>2.084***</td>
<td>4.215***</td>
<td>7.596***</td>
<td>12.69***</td>
</tr>
<tr>
<td><strong>Student’s $t$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uni., $k = 1$</td>
<td>0.463*</td>
<td>1.158***</td>
<td>1.806***</td>
<td>4.585***</td>
<td>8.059***</td>
<td>13.48***</td>
</tr>
<tr>
<td>uni., $k = 2$, $\nu_1 = \nu_2$</td>
<td>0.371</td>
<td>0.648</td>
<td>1.343</td>
<td>2.872</td>
<td>5.651</td>
<td>10.51</td>
</tr>
<tr>
<td>uni., $k = 2$, $\nu_1 \neq \nu_2$</td>
<td>0.324</td>
<td>0.695</td>
<td>1.251</td>
<td>2.964</td>
<td>5.604</td>
<td>10.65</td>
</tr>
<tr>
<td>uni., $k = 3$</td>
<td>0.371</td>
<td>0.695</td>
<td>1.158</td>
<td>2.964</td>
<td>6.346***</td>
<td>12.04***</td>
</tr>
<tr>
<td>multi., $k = 1$</td>
<td>0.926***</td>
<td>1.482***</td>
<td>2.362***</td>
<td>4.910***</td>
<td>7.967***</td>
<td>13.20***</td>
</tr>
<tr>
<td>multi., $k = 2$, $\nu_1 = \nu_2$</td>
<td>0.602***</td>
<td>0.880**</td>
<td>1.853***</td>
<td>3.659***</td>
<td>7.133***</td>
<td>12.32***</td>
</tr>
<tr>
<td>multi., $k = 2$, $\nu_1 \neq \nu_2$</td>
<td>0.509*</td>
<td>0.973***</td>
<td>1.760***</td>
<td>3.520***</td>
<td>7.179***</td>
<td>12.23***</td>
</tr>
<tr>
<td>multi., $k = 3$</td>
<td>0.556**</td>
<td>0.741</td>
<td>1.343</td>
<td>3.196**</td>
<td>6.577***</td>
<td>12.14***</td>
</tr>
</tbody>
</table>

Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.
Results for the entire out–of–sample period
Table 9: Evaluation of Value–at–Risk measures, September 1999 to April 2009:
Univariate Student’s t models

<table>
<thead>
<tr>
<th>100 × ξ</th>
<th>0.25</th>
<th>0.5</th>
<th>1</th>
<th>2.5</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uni., k = 1</td>
<td>0.682***</td>
<td>1.364***</td>
<td>2.286***</td>
<td>4.332***</td>
<td>8.022***</td>
<td>13.08***</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ = ν₂</td>
<td>0.401</td>
<td>0.722</td>
<td>1.203</td>
<td>3.329**</td>
<td>6.458***</td>
<td>10.83</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ ≠ ν₂</td>
<td>0.401</td>
<td>0.762*</td>
<td>1.163</td>
<td>3.289**</td>
<td>6.378***</td>
<td>10.67</td>
</tr>
<tr>
<td>uni., k = 3</td>
<td>0.401</td>
<td>0.762*</td>
<td>1.524**</td>
<td>3.771***</td>
<td>6.980***</td>
<td>12.23***</td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uni., k = 1</td>
<td>0.401</td>
<td>0.963***</td>
<td>2.447***</td>
<td>4.813***</td>
<td>8.022***</td>
<td>14.08***</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ = ν₂</td>
<td>0.201</td>
<td>0.602</td>
<td>1.324</td>
<td>3.490**</td>
<td>6.980***</td>
<td>11.91***</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ ≠ ν₂</td>
<td>0.201</td>
<td>0.602</td>
<td>1.284</td>
<td>3.450**</td>
<td>6.779***</td>
<td>11.87***</td>
</tr>
<tr>
<td>uni., k = 3</td>
<td>0.201</td>
<td>0.562</td>
<td>1.083</td>
<td>3.169**</td>
<td>6.899***</td>
<td>12.23***</td>
</tr>
<tr>
<td>United Kingdom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uni., k = 1</td>
<td>0.842***</td>
<td>1.284***</td>
<td>2.246***</td>
<td>4.412***</td>
<td>7.702***</td>
<td>13.04***</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ = ν₂</td>
<td>0.802***</td>
<td>1.123***</td>
<td>1.765***</td>
<td>3.369***</td>
<td>5.576</td>
<td>11.55**</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ ≠ ν₂</td>
<td>0.762***</td>
<td>1.123***</td>
<td>1.725***</td>
<td>3.369***</td>
<td>5.616</td>
<td>11.47**</td>
</tr>
<tr>
<td>uni., k = 3</td>
<td>0.481**</td>
<td>0.882**</td>
<td>1.484**</td>
<td>3.329**</td>
<td>5.856*</td>
<td>11.55**</td>
</tr>
<tr>
<td>Equally weighted portfolio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uni., k = 1</td>
<td>0.722***</td>
<td>1.484***</td>
<td>2.367***</td>
<td>5.174***</td>
<td>9.105***</td>
<td>14.80***</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ = ν₂</td>
<td>0.521**</td>
<td>0.842**</td>
<td>1.645***</td>
<td>3.450**</td>
<td>6.498***</td>
<td>11.11*</td>
</tr>
<tr>
<td>uni., k = 2, ν₁ ≠ ν₂</td>
<td>0.481**</td>
<td>0.882**</td>
<td>1.564***</td>
<td>3.530**</td>
<td>6.458***</td>
<td>11.19*</td>
</tr>
<tr>
<td>uni., k = 3</td>
<td>0.481**</td>
<td>0.842**</td>
<td>1.284</td>
<td>3.410***</td>
<td>6.699***</td>
<td>12.27***</td>
</tr>
</tbody>
</table>
• Now we turn to the VaR measures for the entire backtesting sample, including the crisis period.

• For the univariate Student’s $t$ models, results are reported in Table 9.

• As may have been expected, the performance of the models worsens dramatically.

• The reason for this failure can be seen in the fact that the conditional variance of these processes is essentially bounded by the volatility parameter of their high–volatility component, which incorporates the information of previous high–volatility periods.

• Thus, as rolling parameter estimates evolve slowly, the models will have difficulties to react promptly if volatility rises to a historically new level.

• This is illustrated in the next figure by means of the example of the equally weighted portfolio of the three stock markets.
• For the two–component process with $\nu_1 = \nu_2$ and moving estimation window, the upper panel of the figure depicts the out–of–sample return series along with the ex ante VaR measures at the 1% level.

• The figure shows that there is a clustering of violations due to the outbreak of extraordinarily high volatility at the end of September 2008.

• For example, in the period from September 29 to October 31, there are 7 violations of the 1% VaR, corresponding to a 28% hit rate in this period, which clearly deteriorates the overall performance.

• In view of the fact that the three–component models tend to identify a “very high–volatility regime” (regime 3 in Tables 3 and 4), we may expect them to do better in such turbulent periods.

• The Figure shows that this is indeed the case.

• For this model, we only register 4 violations in the period in question, which is still a hit rate of 16%, however.
Returns and 1% Value-at-Risk (VaR) implied by a two-regime model

Returns and 1% Value-at-Risk (VaR) implied by a three-regime model
Summary of the Example

• Most of the time, a Markov mixture of two Student’s $t$ distributions provides a reliable model for calculating Value–at–Risk measures for national stock market indices.

• However, during the period of the recent financial crisis, all models failed, where most of the violations leading to this failure were concentrated in a single month, namely October 2008.

• The reason for the breakdown of the models can be seen in the fact that the conditional volatility of the class of Markov–switching models studied in this paper is essentially bounded by that of the highest–volatility regime.

• Since rolling parameter estimates adjust only sluggishly, this makes these models unable to cope with a burst of previously unseen volatility.

• During such periods, three–regime specifications turned out to be much more capable of detecting sudden jumps in the risk level, but their performance was less satisfying in “normal times”.
• Models reacting immediately and more flexibly to shocks, such as GARCH, can thus be expected be more appropriate to meet such previously unseen bursts of volatility.
Comparison with GARCH

- To investigate the last issue, we repeat the exercise with a skewed and fat-tailed GARCH process, i.e., a GARCH(1,1) with a skewed Student’s $t$ error distribution,

$$
\begin{align*}
    r_t &= \mu + \sigma_t \eta_t, \quad \eta_t \overset{iid}{\sim} \text{Skewt}(\nu, p, \theta) \\
    \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2,
\end{align*}
$$

where the skewed $t$ is defined via the density

$$
f(z; \nu, p, \theta) = \frac{\theta}{1 + \theta^2 \nu^{1/p} B(\nu, 1/p)} \left\{ \begin{array}{ll}
    \left(1 + \frac{|z| \theta^p}{\nu}\right)^{-\nu-1/p} & \text{if } z < 0 \\
    \left(1 + \frac{z \theta^p}{\nu}\right)^{-\nu-1/p} & \text{if } z \geq 0,
\end{array} \right. \quad (25)
$$

where $\nu, p, \theta > 0$. 
Table 10: Evaluation of Value–at–Risk measures, September 1999 to April 2009: skewed $t$–GARCH(1,1)

<table>
<thead>
<tr>
<th></th>
<th>0.25</th>
<th>0.5</th>
<th>1</th>
<th>2.5</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage ($100 \times \hat{\xi}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>0.2810</td>
<td>0.4020</td>
<td>0.9240</td>
<td>2.7710</td>
<td>5.7030</td>
<td>10.9640</td>
</tr>
<tr>
<td>Germany</td>
<td>0.2010</td>
<td>0.4420</td>
<td>0.8030</td>
<td>2.4900</td>
<td>5.6630</td>
<td>11.0840*</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.4020</td>
<td>0.5220</td>
<td>1.2450</td>
<td>2.7710</td>
<td>5.1000</td>
<td>10.6430</td>
</tr>
<tr>
<td>Equally weighted portfolio</td>
<td>0.2810</td>
<td>0.5620</td>
<td>0.7230</td>
<td>2.6510</td>
<td>5.9840**</td>
<td>10.9640</td>
</tr>
</tbody>
</table>
• The figures are again for the equally weighted portfolio.
Returns and 1% VaR implied by Skewed-t GARCH(1,1)
Returns and 1% VaR implied by Skewed-t GARCH(1,1) (red) and two-regime Markov-switching (green)
Returns and 1% VaR implied by Skewed-t GARCH(1,1) (red) and three-regime Markov-switching (green)
Exceedance correlations

• In basic portfolio theory, we are interested in the first two moments of the (portfolio) return distribution, i.e., mean and variance.

• In this framework, correlations between assets are of predominant interest, because the strength of the correlations determines the degree of risk (variance) reduction that can be achieved by efficient portfolio diversification.

• Simple correlation estimates may be misleading, however, due to asymmetric dependence structures.

• This refers to the observation that, for example, stock returns are more dependent in bear markets (market downturns) than in bull markets.

• Therefore, diversification might fail when the benefits from diversification are most urgently needed.
Exceedance correlations

• A popular tool to describe this asymmetric dependence structure are the exceedance correlations of Longin and Solnik (2001).\(^6\)

• For a given threshold \(\theta\), the exceedance correlation between (demeaned) returns \(r_1\) and \(r_2\) is given by

\[
\rho(\theta) = \begin{cases} 
\text{Corr}(x, y|x > \theta, y > \theta) & \text{for } \theta \geq 0 \\
\text{Corr}(x, y|x < \theta, y < \theta) & \text{for } \theta \leq 0
\end{cases}
\]  

(26)

• Let us consider monthly returns of MSCI stock market indices for the US and Germany from January 1970 to June 2008.

exceedance threshold $\theta = -5$
exceedance threshold $\theta = 5$
Exceedance correlations

\[ r_t = (r_{t,US}, r_{t,Ger}) \]
Table 11: Parameter Estimates for three-regime Gaussian MS-GARCH

<table>
<thead>
<tr>
<th></th>
<th>Regime 1</th>
<th>Regime 2</th>
<th>Regime 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean return</td>
<td>[0.874, 1.024]</td>
<td>[3.479, 2.593]</td>
<td>[−1.553, −1.528]</td>
</tr>
<tr>
<td>std. deviation</td>
<td>[3.091, 3.693]</td>
<td>[4.482, 5.810]</td>
<td>[5.593, 8.338]</td>
</tr>
<tr>
<td>correlation</td>
<td>0.466</td>
<td>−0.068</td>
<td>0.798</td>
</tr>
<tr>
<td>stationary prob.</td>
<td>0.583</td>
<td>0.209</td>
<td>0.208</td>
</tr>
</tbody>
</table>

- transition matrix

\[ P = \begin{pmatrix}
0.940 & 0.168 & 0.000 \\
0.038 & 0.729 & 0.165 \\
0.022 & 0.103 & 0.835
\end{pmatrix} \]  

- Regime 1: “business as usual”
- Regime 2: bull market
- Regime 3: bear market
Überschreitungskorrelationen für die MSCI Indizes USA/Deutschland

Daten
Normalverteilung
Markov-Switching Modell