Financial Data Analysis

Regime–switching Models

July 31, 2011

Mixture Models for Financial Time Series

- A possible interpretation of mixture distributions for modeling asset returns is that the distribution of returns depends on an (typically 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.

- If the market is in state j at time t, the $N \times 1$ vector of returns under consideration is multivariate normal with mean μ_j and covariance matrix Σ_j, i.e., its density is

$$
\begin{aligned}
f(r_t | s_t = j) &= \phi(r_t; \mu_j, \Sigma_j) \qquad (1) \\
&= \frac{1}{(2\pi)^{N/2}\sqrt{|\Sigma_j|}} \exp\left\{ -\frac{1}{2}(r_t - \mu_j)' \Sigma_j^{-1} (r_t - \mu_j) \right\}.
\end{aligned}
$$

- In (1), $s_t \in \{1, \ldots, k\}$ is the regime variable indicating the regime at time t.

- At time t, the market is in state j with (conditional) probability π_{jt}, i.e.,

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

- Thus the model can be written

$$r_t | I_{t-1} \sim \begin{cases} \text{Normal}(\mu_1, \Sigma_1) & \text{with probability } \pi_{1t} \\ \text{Normal}(\mu_2, \Sigma_2) & \text{with probability } \pi_{2t} \\ \ldots & \\ \text{Normal}(\mu_k, \Sigma_k) & \text{with probability } \pi_{kt}, \end{cases} \tag{2}$$

where I_{t-1} is the information set available up to time $t-1$.

- The (conditional) distribution of r_t at time t is a k–component finite normal mixture distribution, with density

$$f_{t-1}(r_t) = \sum_{j=1}^{k} \pi_{jt} \phi(r_t; \mu_j, \Sigma_j), \tag{3}$$

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

- In (3), the π_{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** μ_j, and **component covariance matrices** Σ_j, $j = 1, \ldots, k$.

- Since this is a linear combination of normal densities, the *raw moments*[1] can be calculated as linear combinations of normal moments.

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

$$\mu := \mathsf{E}_{t-1}(r_t) = \sum_{j=1}^{k} \pi_{jt}\mu_j \tag{4}$$

and

$$\mathsf{Var}_{t-1}(r_t) = \sum_{j=1}^{k} \pi_{jt}\Sigma_j + \sum_{j=1}^{k} \pi_{jt}(\mu_j - \mu)(\mu_j - \mu)', \tag{5}$$

respectively.

- The mean (4) is just the probability–weighted average of the component means.

[1] As opposed to centered moments, which are calculated around the mean. The variance is a centered moment, for example. However, centered moments can be obtained from the raw moments. For example, the variance (the second centered moment) can be obtained from the first and second raw moments, since $E(X-\mu)^2 = E(X^2) - E^2(X)$, where $\mu = E(X)$.

- The variance formula (5) is a bit more complex.

- It can be interpreted as

"expectation of the variance (first term) $+$ variance of the expectation (second term)".

- For univariate mixtures (for ease of notation only), the variance formula can be derived as follows.

- For univariate mixtures, the variance formula (5) becomes (dropping the time subscript t for simplicity)

$$\begin{aligned}\sum_j \pi_j \sigma_j^2 + \sum_j \pi_j (\mu_j - \mu)^2 &= \sum_j \pi_j \sigma_j^2 + \sum_j \pi_j (\mu_j^2 - 2\mu_j \mu + \mu^2) \\ &= \sum_j \pi_j (\sigma_j^2 + \mu_j^2) - \mu^2 \\ &= \mathsf{E}(r^2) - \mathsf{E}^2(r) = \mathsf{Var}(r),\end{aligned}$$

since, from the properties of the normal distribution,

$$E(r) = \sum_j \pi_j \mu_j, \quad \mathsf{E}(r^2) = \sum_j \pi_j (\sigma_j^2 + \mu_j^2). \tag{6}$$

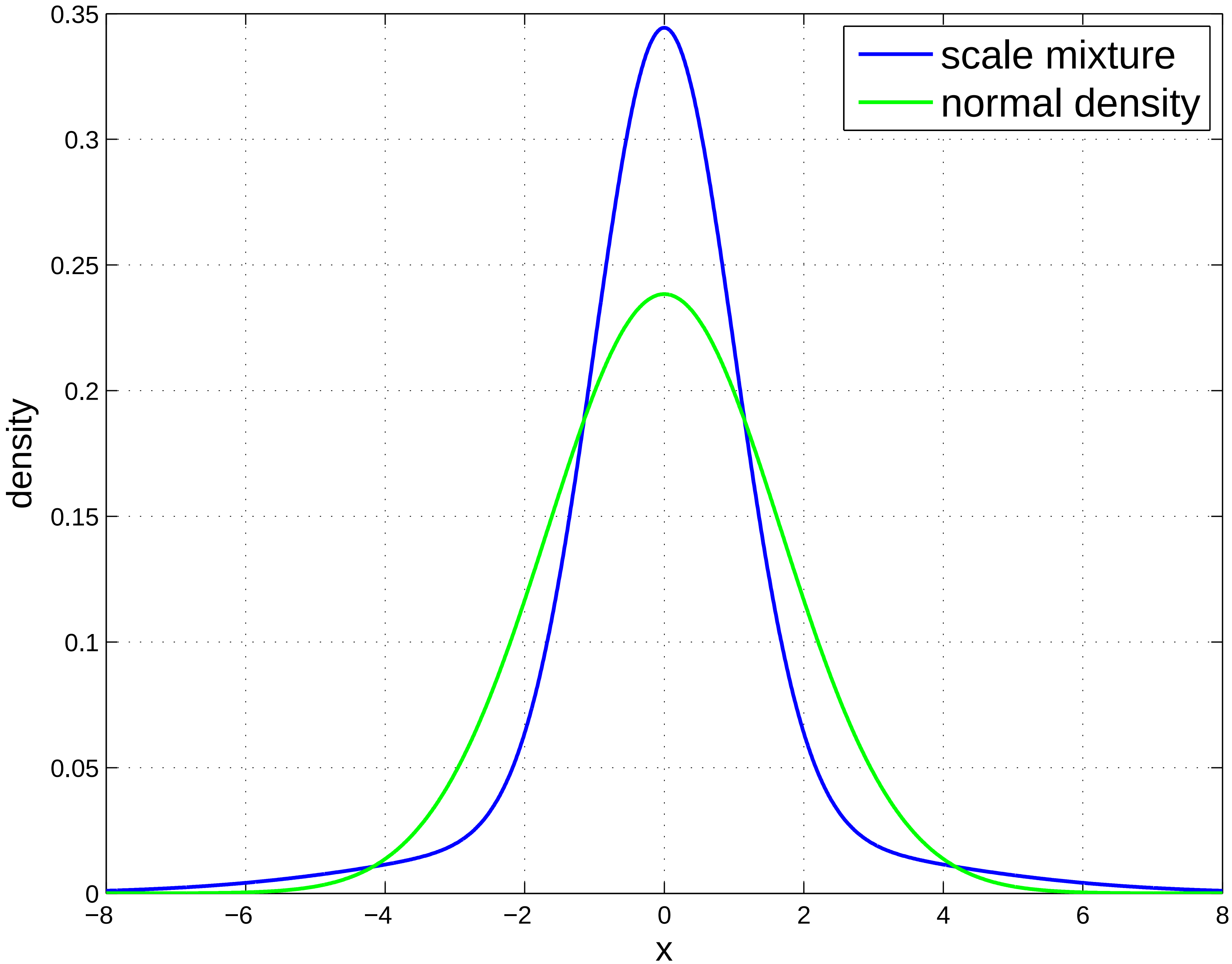
- Normal mixture distributions are known to exhibit great flexibility, and they can capture the skewness and excess kurtosis which characterizes many financial variables.

- For example, consider the *scale normal mixture*, where only the variances are component–specific, whereas the component means are all equal to μ. This gives rise to a leptokurtic density.

- Consider example

$$\pi_1 = 0.8, \quad \mu_1 = \mu_2 = 0, \quad \sigma_1^2 = 1, \quad \sigma_2^2 = 10,$$

and the normal distribution with the same mean $\mu = 0$ and variance $\sigma^2 = 0.8 \cdot 1 + 0.2 \cdot 10 = 2.8$.


scale mixture
normal density
density
x
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
−8
−6
−4
−2
0
2
4
6
8

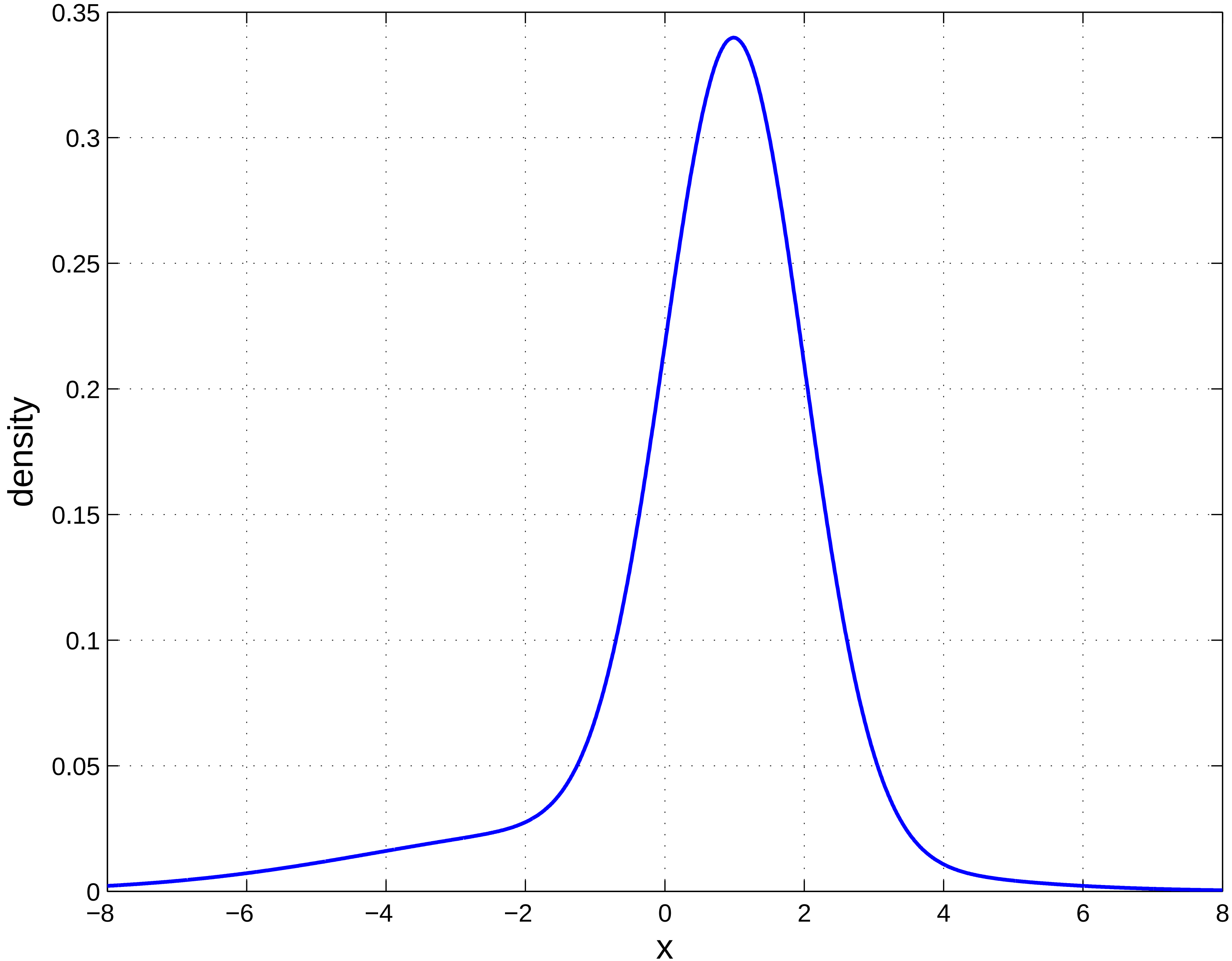
- A skewed density can be generated by allowing the component means to differ.

- The typical situation where left–skewness emerges is where the component with the smaller probability (mixing weight) has the greater variance and the smaller mean.

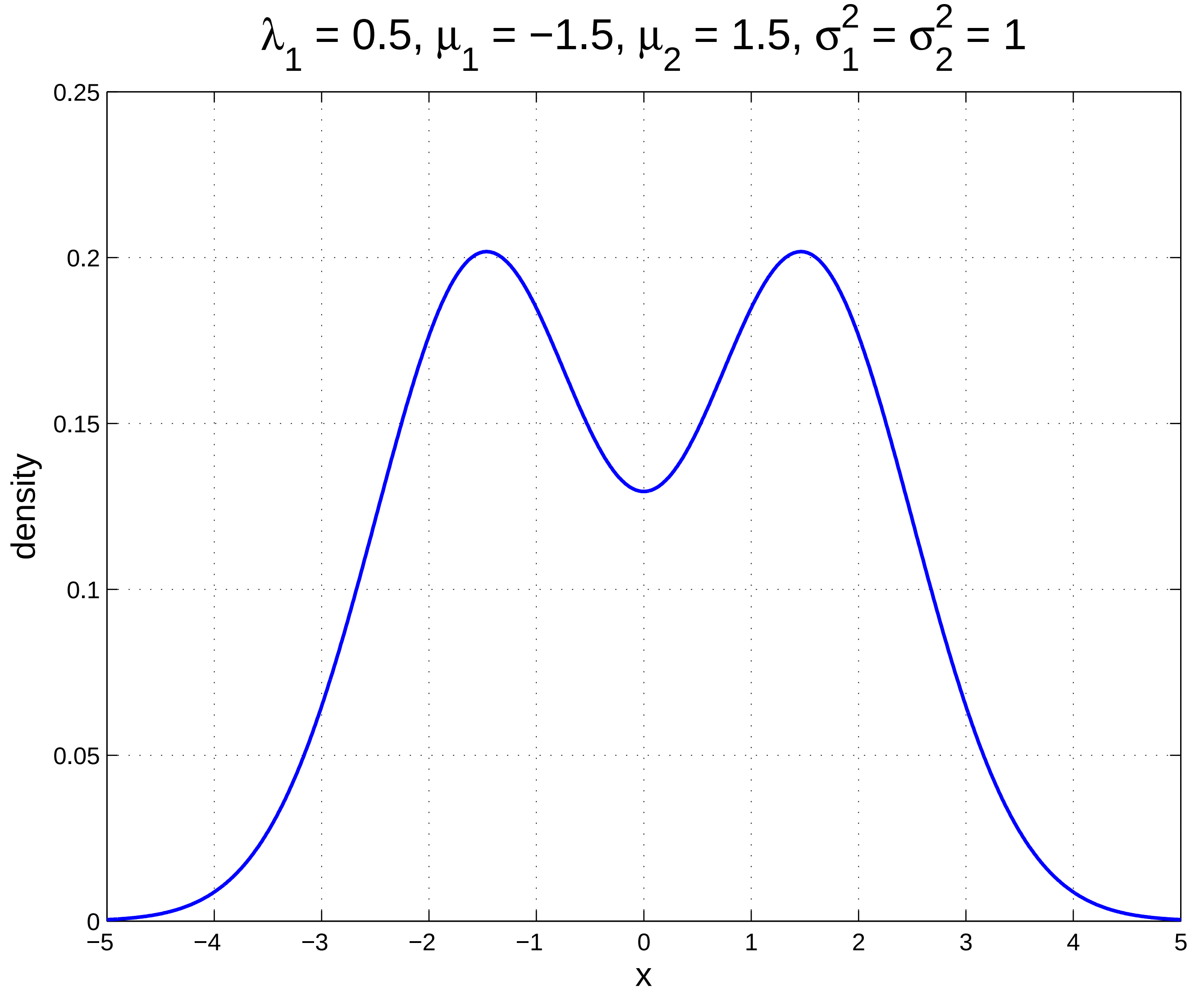
- In terms of the bull/bear market interpretation of the two–component model, this means that the bear market component (smaller mean return) has smaller probability and a higher volatility

- Consider example

$$\pi_1 = 0.8, \quad \mu_1 = 1, \quad \mu_2 = -1, \quad \sigma_1^2 = 1, \quad \sigma_2^2 = 10.$$


0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
density
−8
−6
−4
−2
0
2
4
6
8
x

- Bimodality is the result of the component means being sufficiently far apart, relative to the magnitude of the variances.



Example

- For illustration, let us consider the simplest case.

- This is a univariate iid mixture model, where the conditional mixing weights (regime probabilities) are constant over time:

$$\pi_{jt} = \lambda_j, \quad j = 1, \ldots, k, \quad \text{for all } t. \tag{7}$$

- The parameter vector to be estimated is $\theta = (\lambda_1, \ldots, \lambda_{k-1}, \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2)$.

- Mixture models are typically estimated by either maximum likelihood or Bayesian approaches.

- The log–likelihood function of the iid mixture model is

$$\log L = \sum_{t=1}^{T} \log \left\{ \sum_{j=1}^{k} \lambda_j \phi(r_t; \mu_j, \sigma_j^2) \right\}. \tag{8}$$

- The log–likelihood (8) can conveniently be maximized by means of the Expectation–Maximization (EM) algorithm, which can also be constructed for more complex mixture models.[2]

- However, mixture likelihoods often have more than a single local maximum

- There are further subtleties of likelihood inference in mixture models, which are discussed in the literature.[3]

---

[2]E.g., for Markov–switching mixtures, as birefly discussed below.

[3]E.g., G. J. McLachlan and T. Krishnan (2008): *The EM algorithm and extensions*, Wiley; McLachlan and Peel (2000): *Finite Mixture Models*, Wiley.

Example: European Stock Market Returns

- Consider mixture models for the major European stock markets.

Table 1: Likelihood–based goodness–of–fit

	$k=2$		$k=3$		$k=4$	
	$\log L$	BIC	$\log L$	BIC	$\log L$	BIC
CAC 40	-8451.7	**16946**	-8439.9	16948	-8438.5	16971
DAX 30	-8628.4	17299	-8602.4	**17273**	-8599.8	17293
FTSE 100	-7313.7	14670	-7295.3	**14659**	-7294.7	14683

Table 2: Two–component normal mixture parameter estimates for European stock markets

	$\widehat{\lambda}_1$	$\widehat{\mu}_1$	$\widehat{\sigma}_1^2$	$\widehat{\lambda}_2$	$\widehat{\mu}_2$	$\widehat{\sigma}_2^2$
CAC 40	0.8396 (0.0199)	0.0679 (0.0190)	1.0310 (0.0462)	0.1604 (0.0199)	-0.2023 (0.1102)	7.0192 (0.6372)
DAX 30	0.8221 (0.0205)	0.0898 (0.0195)	1.0253 (0.0500)	0.1779 (0.0205)	-0.2851 (0.1085)	7.5843 (0.6527)
FTSE 100	0.8357 (0.0174)	0.0542 (0.0145)	0.6120 (0.0268)	0.1643 (0.0174)	-0.0944 (0.0880)	4.9196 (0.4118)

Table 3: Three–component normal mixture parameter estimates for European stock markets

	$\widehat{\lambda}_1$	$\widehat{\mu}_1$	$\widehat{\sigma}_1^2$	$\widehat{\lambda}_2$	$\widehat{\mu}_2$	$\widehat{\sigma}_2^2$	$\widehat{\lambda}_3$	$\widehat{\mu}_3$	$\widehat{\sigma}_3^2$
CAC 40	0.672 (0.0842)	0.098 (0.0257)	0.829 (0.0897)	0.290 (0.0726)	−0.144 (0.1035)	3.133 (0.7571)	0.037	0.003 (0.4205)	13.931 (3.8774)
DAX 30	0.703 (0.0495)	0.046 (0.0333)	1.509 (0.1383)	0.187 (0.0563)	0.154 (0.0446)	0.239 (0.0848)	0.109	−0.354 (0.1608)	9.913 (1.1378)
FTSE 100	0.662 (0.0596)	0.060 (0.0182)	0.480 (0.0407)	0.300 (0.0528)	−0.008 (0.0579)	2.034 (0.3675)	0.036	−0.213 (0.3065)	10.580 (2.4531)

- BIC $= -2 \times \log L + (3k-1) \times \log T$, where $3k-1$ is the number of parameters of a mixture model with k components. Bold entries indicate the best model according to BIC.

- The cdf of the normal mixture is calculated as

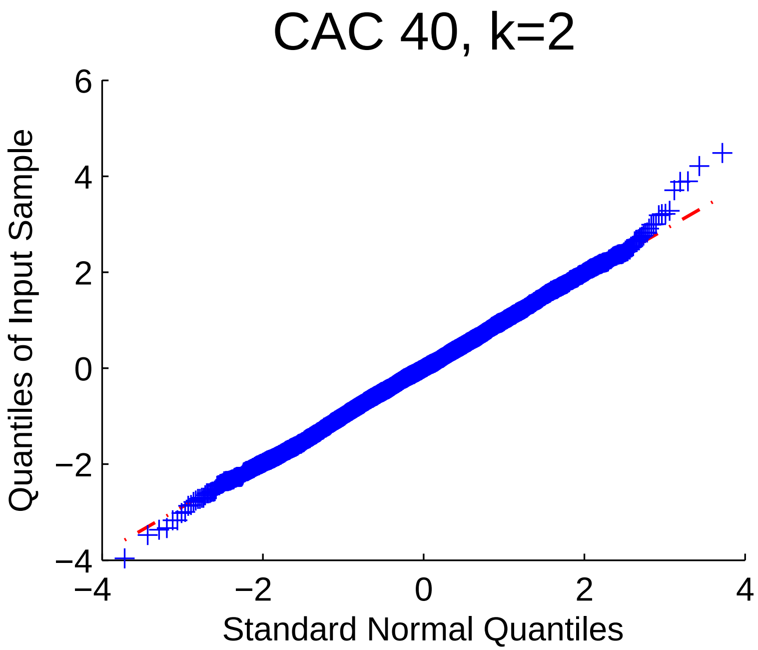
$$F(r_t;\theta) = \sum_{j=1}^{k} \lambda_j \Phi\left(\frac{r_t-\mu_j}{\sigma_j}\right), \tag{9}$$

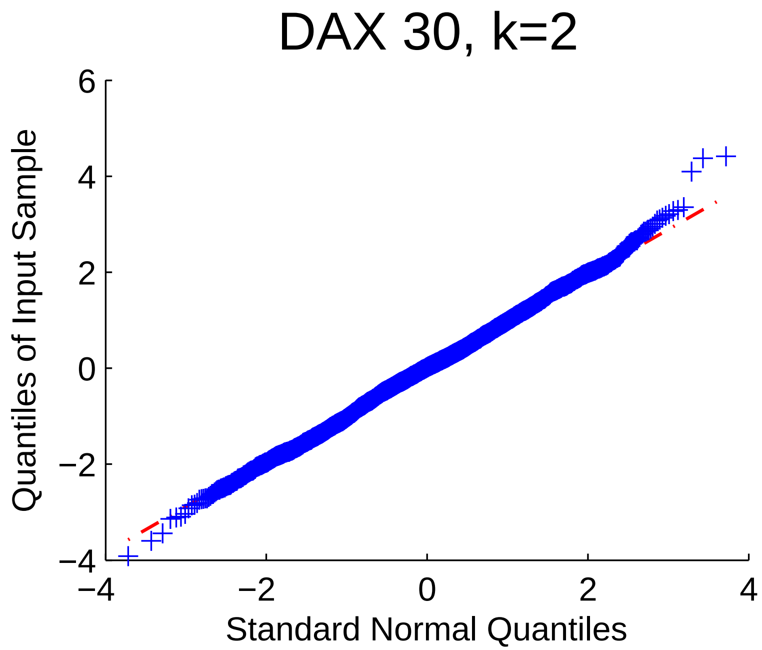
where Φ is the standard normal cdf.

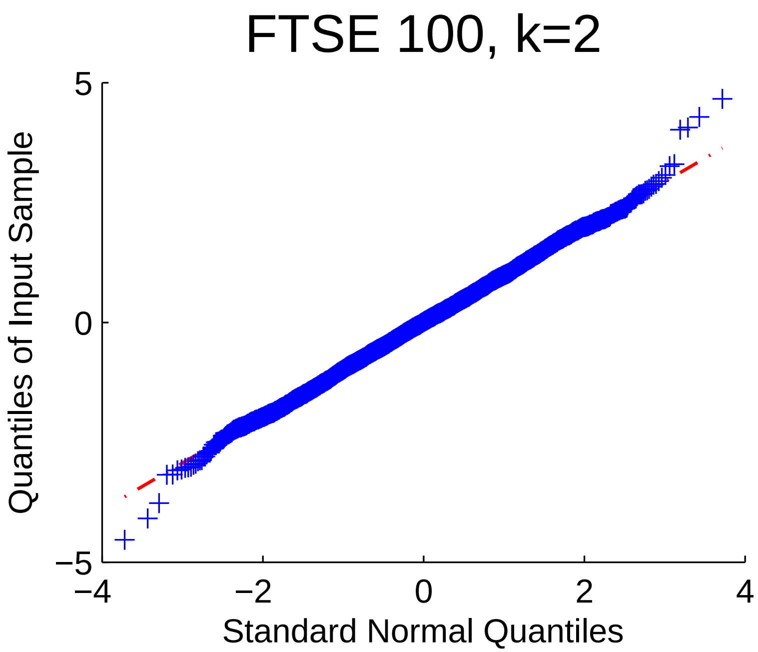
- The sequence (9) should be uniformly distributed over the unit interval, and then

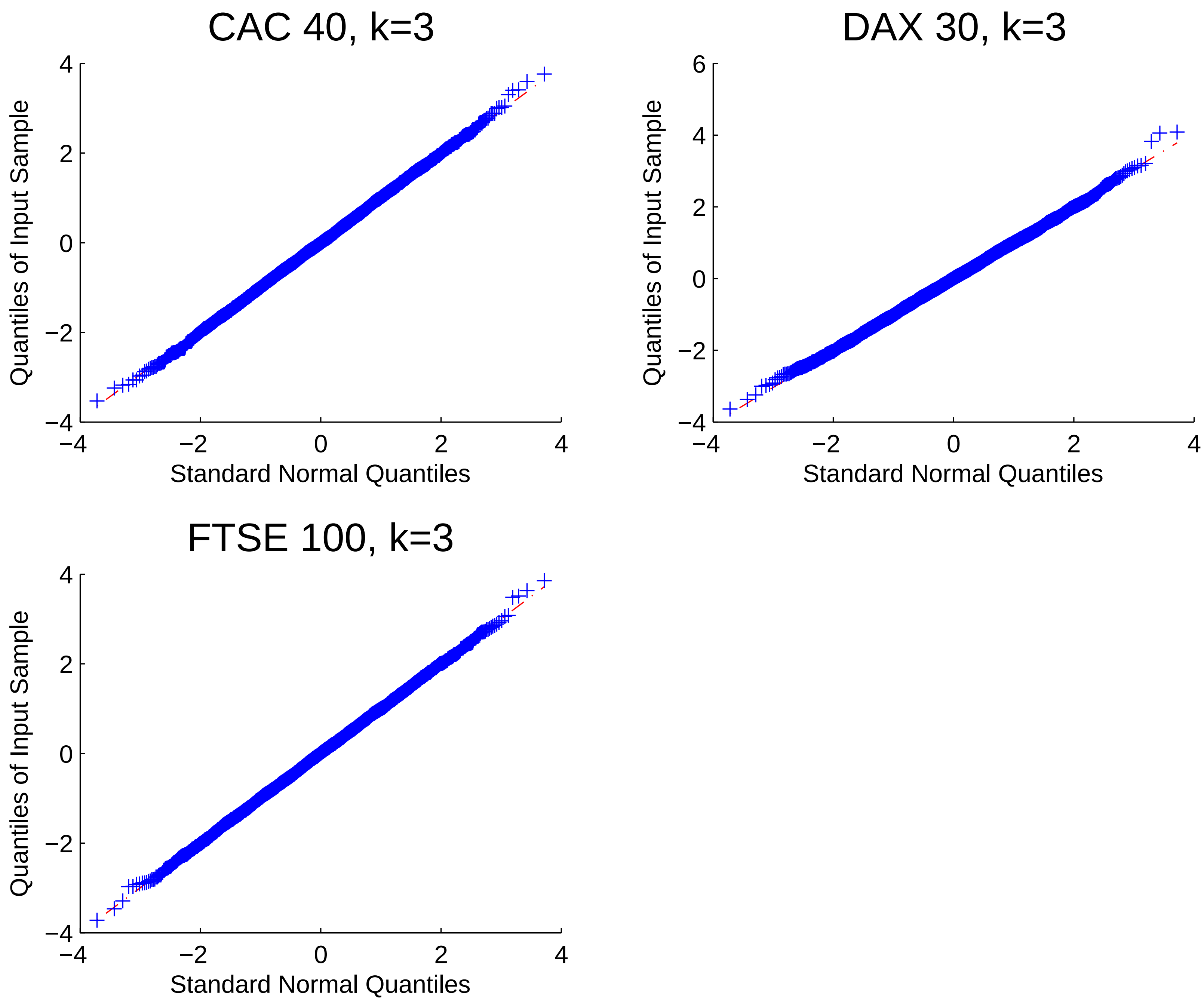
$$z = \Phi^{-1}\left(\sum_{j=1}^{k} \lambda_j \Phi\left(\frac{r_t - \mu_j}{\sigma_j}\right)\right) \tag{10}$$

should have a standard normal distribution, where Φ^{-1} denotes the inverse standard normal cdf.


CAC 40, k=2
Quantiles of Input Sample
Standard Normal Quantiles
6
4
2
0
−2
−4
−4
−2
0
2
4


DAX 30, k=2
Quantiles of Input Sample
Standard Normal Quantiles
6
4
2
0
−2
−4
−4
−2
0
2
4


FTSE 100, k=2
Quantiles of Input Sample
Standard Normal Quantiles
5
0
−5
−4
−2
0
2
4


CAC 40, k=3
DAX 30, k=3
FTSE 100, k=3
Quantiles of Input Sample
Standard Normal Quantiles
−4
−2
0
2
4
6

Multivariate Mixture models

- In a multivariate framework, the mixture approach is also 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 μ and Σ, i.e.,

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

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

$$\begin{aligned} \mu_p &= w'\mu = \sum_{i=1}^{N} w_i \mathsf{E}(r_i), \\ \sigma_p^2 &= w'\Sigma w = \sum_{i=1}^{N} w_i^2 \mathsf{Var}(r_i) + 2 \sum_{j=1}^{N} \sum_{i<j} w_i w_j \mathsf{Cov}(r_i, r_j). \end{aligned}$$

- Now suppose (common correlation model)

$$\begin{aligned} \text{Var}(r_i) &= \sigma^2, \quad i = 1, \ldots, N, \\ \text{Corr}(r_i, r_j) &= \rho \quad i, j = 1, \ldots, N, \quad i \neq j, \end{aligned}$$

and consider equally weighted portfolio, i.e.,

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

- Then the portfolio variance

$$\begin{aligned} \sigma_p^2 &= \sum_{i=1}^{N} w_i^2 \text{Var}(r_i) + 2 \sum_{j=1}^{N} \sum_{i<j} w_i w_j \underbrace{\text{Cov}(r_i, r_j)}_{=\sigma^2 \rho} \\ &= \frac{\sigma^2}{N} + \frac{N(N-1)\sigma^2 \rho}{N^2} \\ &= \sigma^2 \left(\frac{1-\rho}{N} + \rho \right) && (11) \\ &\approx \rho \sigma^2 \quad \text{for large } N \text{ (many assets).} && (12) \end{aligned}$$

Digression

- Question: What if ρ in (12) is negative?

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

$$-\frac{1}{N-1} < \rho < 1, \quad (13)$$

so that the expression in brackets in (11) will always be positive.

- As noted by Paul A. Samuelson,[4] 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."

[4] Samuelson (1967): General Proof that Diversification Pays, *Journal of Financial and Quantitative Analysis*, 2, 1–13.

Modeling the (conditional) mixing weights

- The independent mixture model often 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 (and hence predictable).
- 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 generated 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 have become very popular in economics and finance since Hamilton (1989).[5]

- 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

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

where

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

[5]A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," Econometrica , March 1989.

- Collect the transition probabilities in the $k \times k$ **transition matrix** P,

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

- 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,

$$\pi_t := [\pi_{1t}, \pi_{2t}, \ldots, \pi_{kt}]', \tag{14}$$

where, as before,

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

- 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 (14), the probability of being in state j at time $t+1$ is

$$\begin{aligned} \text{Pr}_{t-1}(s_{t+1} = j) &= \pi_{j,t+1} \quad (15) \\ &= \sum_{i=1}^{k} \text{Pr}_{t-1}(s_t = i) \times \text{Pr}(s_{t+1} = j | s_t = i) \quad (16) \\ &= \sum_{i=1}^{k} \pi_{it} p_{ij}, \quad j = 1, \ldots, k. \quad (17) \end{aligned}$$

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 τ–**step transition probabilities** are given by the respective elements of P^τ.

- That is, the probability that an observation from Regime i will be followed τ periods later by an observation from Regime j is given by the element in the jth row and ith column of the matrix P^{τ}.

- What happens if the forecast horizon becomes large, i.e., $\tau \to \infty$?

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

$$\pi_\infty := \lim_{\tau \to \infty} P^\tau \pi_t \tag{18}$$

exists and is independent of the initial probability vector π_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 relative frequency of the regimes over longer time horizons.

- The convergence in (18) is due to the fact that

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

where

$$\mathbf{1}_k = [\underbrace{1, 1, \ldots, 1}_{k \text{ times}}]'.$$

- This shows that we also have

$$P\pi_\infty = \pi_\infty, \tag{19}$$

which shows that π_∞ is the stationary distribution of the chain.

- The *unconditional distribution* of r_t is then a k–component normal mixture distribution with weights $\pi_{j,\infty}$, $j = 1, \ldots, k$, i.e.,

$$f(r_r) = \sum_{j=1}^{k} \frac{\pi_{j,\infty}}{(2\pi)^{N/2}\sqrt{|\Sigma_j|}} \exp\left\{-\frac{1}{2}(r_t - \mu_j)'\Sigma_j^{-1}(r_t - \mu_j)\right\}. \tag{20}$$

- **Example**: Consider the two–state Markov chain with transition matrix

$$P = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix}. \tag{21}$$

Then (15) becomes

$$\begin{aligned}
\pi_{1,t+\tau} &= p_{11}\pi_{1,t+\tau-1} + p_{21}\pi_{2,t+\tau-1} \\
&= p_{11}\pi_{1,t+\tau-1} + (1 - p_{22})(1 - \pi_{1t}) \\
&= (p_{11} + p_{22} - 1)\pi_{1,t+\tau-1} + (1 - p_{22}) \\
&= (p_{11} + p_{22} - 1)^2\pi_{1,t+\tau-2} + (p_{11} + p_{22} - 1)(1 - p_{22}) + (1 - p_{22}) \\
&\vdots \\
&= (p_{11} + p_{22} - 1)^\tau\pi_{1t} + (1 - p_{22})\sum_{i=1}^{\tau-1}(p_{11} + p_{22} - 1)^i \\
&= \frac{1 - p_{22}}{2 - p_{11} - p_{22}} + \delta^\tau\left(\pi_{1t} - \frac{1 - p_{22}}{2 - p_{11} - p_{22}}\right) \\
&= \pi_{1,\infty} + \delta^\tau(\pi_{1t} - \pi_{1,\infty}),
\end{aligned}$$

where

$$\delta := p_{11} + p_{22} - 1 \tag{22}$$

measures the persistence of the regimes, and

$$\pi_{1,\infty} = \lim_{\tau \to \infty} \pi_{1,t+\tau} = \frac{1 - p_{22}}{2 - p_{11} - p_{22}}. \tag{23}$$

Similarly,

$$\pi_{2,t+\tau} = \pi_{2,\infty} + \delta^\tau \left(\pi_{2t} - \pi_{2,\infty} \right), \tag{24}$$

where

$$\pi_{2,\infty} = 1 - \pi_{1,\infty} = \frac{1 - p_{11}}{2 - p_{11} - p_{22}}. \tag{25}$$

- The same argument can be made directly for the τ–step transition probabilities in P^τ.

- Let $p_{ij}^{(\tau)} = \Pr(\Delta_{t+\tau} = j | \Delta_t = i)$. Then $PP^{\tau-1} = P^\tau$, i.e.,

$$\left[\begin{array}{cc} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{array} \right] \left[\begin{array}{cc} p_{11}^{(\tau-1)} & 1 - p_{22}^{(\tau-1)} \\ 1 - p_{11}^{(\tau-1)} & p_{22}^{(\tau-1)} \end{array} \right] = \left[\begin{array}{cc} p_{11}^{(\tau)} & 1 - p_{22}^{(\tau)} \\ 1 - p_{11}^{(\tau)} & p_{22}^{(\tau)} \end{array} \right],$$

or, written explicitly for the first element,

$$\begin{aligned} p_{11} \cdot p_{11}^{(\tau-1)} + (1-p_{22})(1-p_{11}^{(\tau-1)}) &= p_{11}^{(\tau)} \\ (1-p_{22}) + \delta p_{11}^{(\tau-1)} &= p_{11}^{(\tau)}. \end{aligned} \tag{26}$$

- Solving (26) recursively as before, we get, with $p_{11}^{(1)} = p_{11}$

$$\begin{aligned} p_{11}^{(\tau)} &= (1-p_{22}) + \delta p_{11}^{(\tau-1)} = (1-p_{22}) + \delta(1-p_{22}) + \delta^2 p_{11}^{(\tau-2)} \\ &\cdots \\ &= (1-p_{22}) \sum_{i=0}^{\tau-2} \delta^i + \delta^{\tau-i} p_{11} = (1-p_{22}) \frac{1-\delta^{\tau-1}}{1-\delta} + \delta^{\tau-i} p_{11} \\ &= \pi_{1,\infty} + \delta^{\tau-1} \left(p_{11} - \frac{1-p_{22}}{1-\delta} \right) = \pi_{1,\infty} + \delta^{\tau-1} \frac{\delta(1-p_{11})}{1-\delta} \\ &= \pi_{1,\infty} + \delta^{\tau} \frac{1-p_{11}}{2-p_{11}-p_{22}} = \pi_{1,\infty} + \delta^{\tau} \pi_{2,\infty}. \end{aligned}$$

- Similar calculations can be done for $p_{22}^{(\tau)}$, and it follows that P^τ containing the τ–step regime probabilities is given by

$$\begin{aligned} P^\tau &= \begin{bmatrix} \pi_{1,\infty} + \delta^\tau \pi_{2,\infty} & (1-\delta^\tau)\pi_{1,\infty} \\ (1-\delta^\tau)\pi_{2,\infty} & \pi_{2,\infty} + \delta^\tau \pi_{1,\infty} \end{bmatrix} \\ &= \begin{bmatrix} \pi_{1,\infty} & \pi_{1,\infty} \\ \pi_{2,\infty} & \pi_{2,\infty} \end{bmatrix} + \delta^\tau \begin{bmatrix} \pi_{2,\infty} & -\pi_{1,\infty} \\ -\pi_{2,\infty} & \pi_{1,\infty} \end{bmatrix}, \end{aligned} \tag{27}$$

so

$$\lim_{\tau\to\infty} P^\tau = P_\infty = \begin{bmatrix} \pi_{1,\infty} & \pi_{1,\infty} \\ \pi_{2,\infty} & \pi_{2,\infty} \end{bmatrix} = [\pi_\infty, \pi_\infty], \tag{28}$$

and the speed of convergence is determined by the magnitude of δ.

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.,

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

and so, once we are in regime j, we expect it to last for

$$\mathsf{E}(D_j) = \frac{1}{1 - p_{jj}} \quad \text{periods,}$$

since

$$
\begin{aligned}
\mathsf{E}(D_j) &= \sum_{d=1}^{\infty} d p_{jj}^{d-1}(1-p_{jj}) = \sum_{d=0}^{\infty}(d+1)p_{jj}^{d}(1-p_{jj}) \\
&= p_{jj}\sum_{d=0}^{\infty} d p_{jj}^{d-1}(1-p_{jj}) + (1-p_{jj})\sum_{d=0}^{\infty} p_{jj}^{d} \\
&= p_{jj}\sum_{d=1}^{\infty} d p_{jj}^{d-1}(1-p_{jj}) + 1 \\
&= p_{jj}\mathsf{E}(D_j) + 1.
\end{aligned}
$$

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.,

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

  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.

- To illustrate, consider a generalization of the **law of total probability**.

- The law of total probability is as follows: 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

$$\Pr(B) = \sum_i \Pr(A_i) \Pr(B|A_i).$$

- The generalization to conditional expectations: For random variable X,

$$\mathsf{E}(X) = \sum_{i=1}^{n} \mathsf{Pr}(A_i)\mathsf{E}(X|A_i). \tag{29}$$

- Hence

$$\begin{aligned} \mathsf{E}(r_t r_{t-\tau}) &= \sum_{i=1}^{k}\sum_{j=1}^{k} \mathsf{Pr}(s_{t-\tau} = i \cap s_t = j)\mathsf{E}(r_t r_{t-\tau}|s_{t-\tau} = i \cap s_t = j) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} \pi_{i,\infty} p_{ij}^{(\tau)} \mu_i \mu_j, \end{aligned} \tag{30}$$

since $\text{Pr}(s_{t-\tau} = i \cap s_t = j) = \text{Pr}(s_{t-\tau} = i)\text{Pr}(s_t = j | s_{t-\tau} = i) = \pi_{i,\infty} p_{ij}^{(\tau)}$.

- For the two–regime model, inserting the expressions in (27) for the τ–step probabilities, (30) yields after a few calculations

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

- For the squares, similar calculations lead to

$$\text{Cov}(r_t^2, r_{t-\tau}^2), = \pi_{1,\infty}\pi_{2,\infty}\delta^\tau(\sigma_1^2 + \mu_1^2 - \sigma_2^2 - \mu_2^2)^2,$$

respectively.

- 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.

- Such probabilities are also required for calculation of the likelihood function.

- Algorithms for calculating such probabilities have been developed.[6]

- Let $z_{jt|t-1}$ be our probability inference that we are in regime j at time t, given information up to time $t-1$.

[6] See Hamilton (1994): *Time Series Analysis*, Chapter 22.

- Then the conditional density of r_t, given the information up to time $t-1$, is

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

and the log–likelihood for a sample of size T

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

which can be maximized by means of the EM algorithm.

Illustration: Asymmetric Correlations in Bull and Bear Markets

- In classical 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

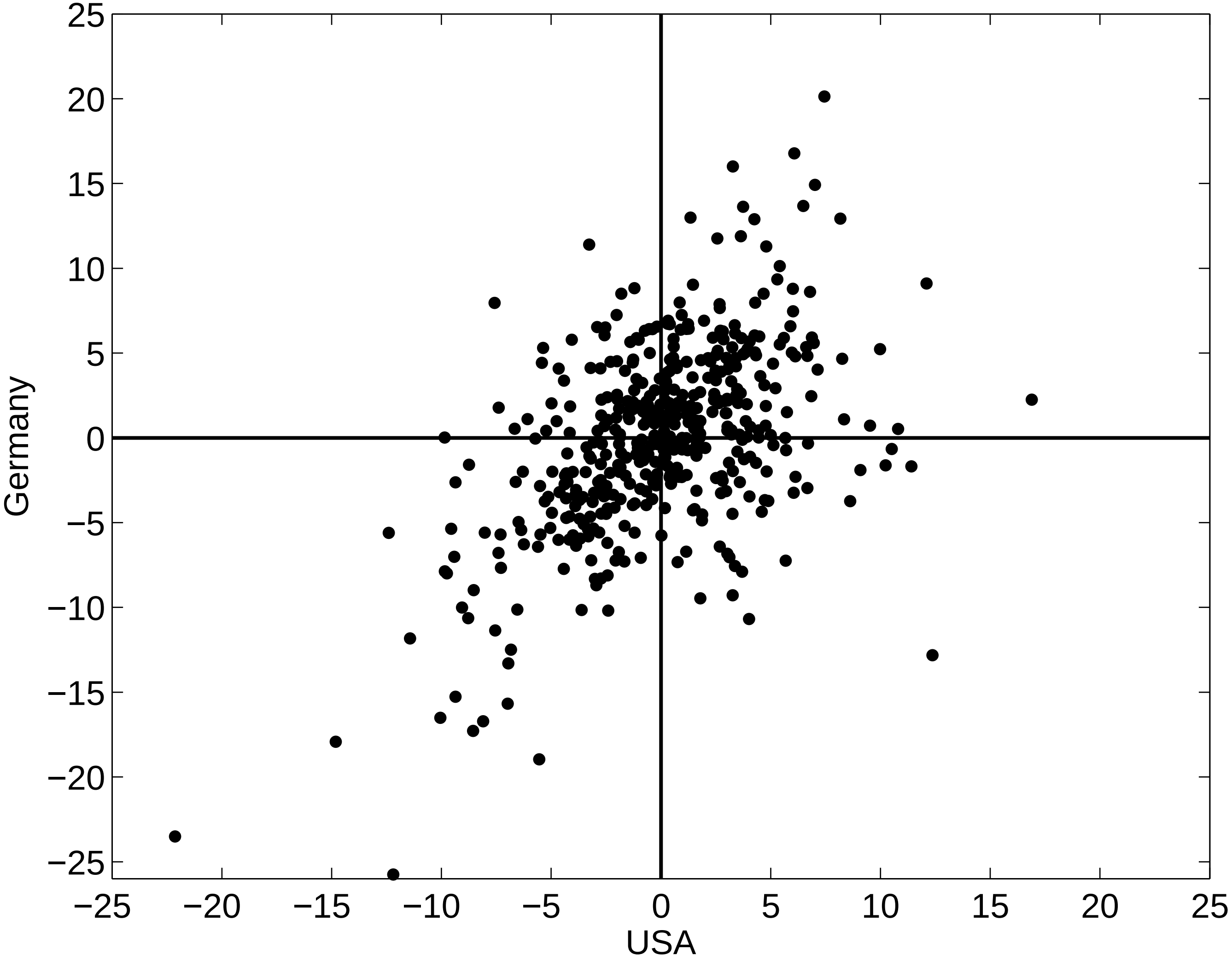
- A popular tool to describe this asymmetric dependence structure are the **exceedance correlations** of Longin and Solnik (2001).[7]

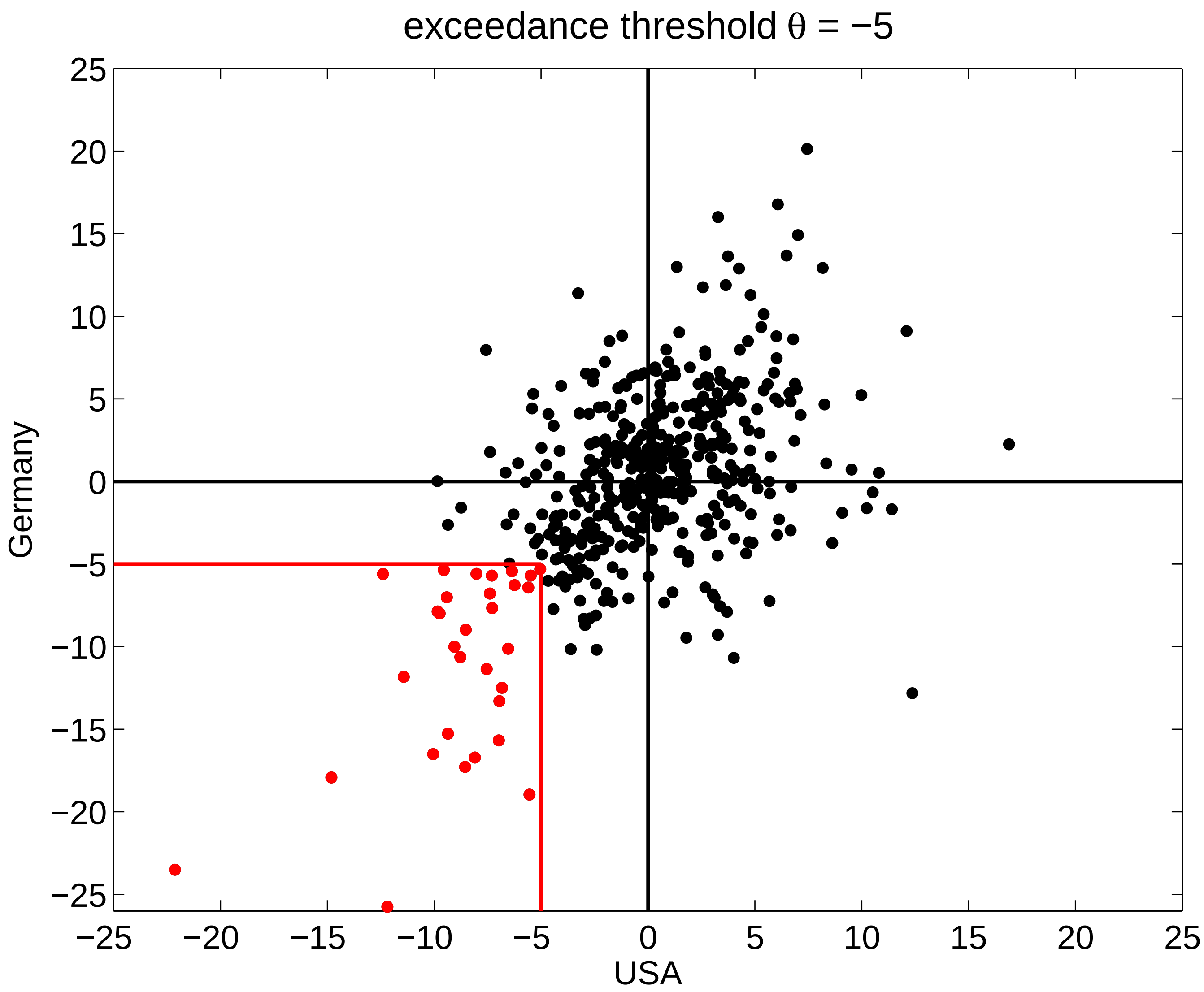
- For a given threshold θ, the exceedance correlation between (demeaned) returns r_1 and r_2 is given by

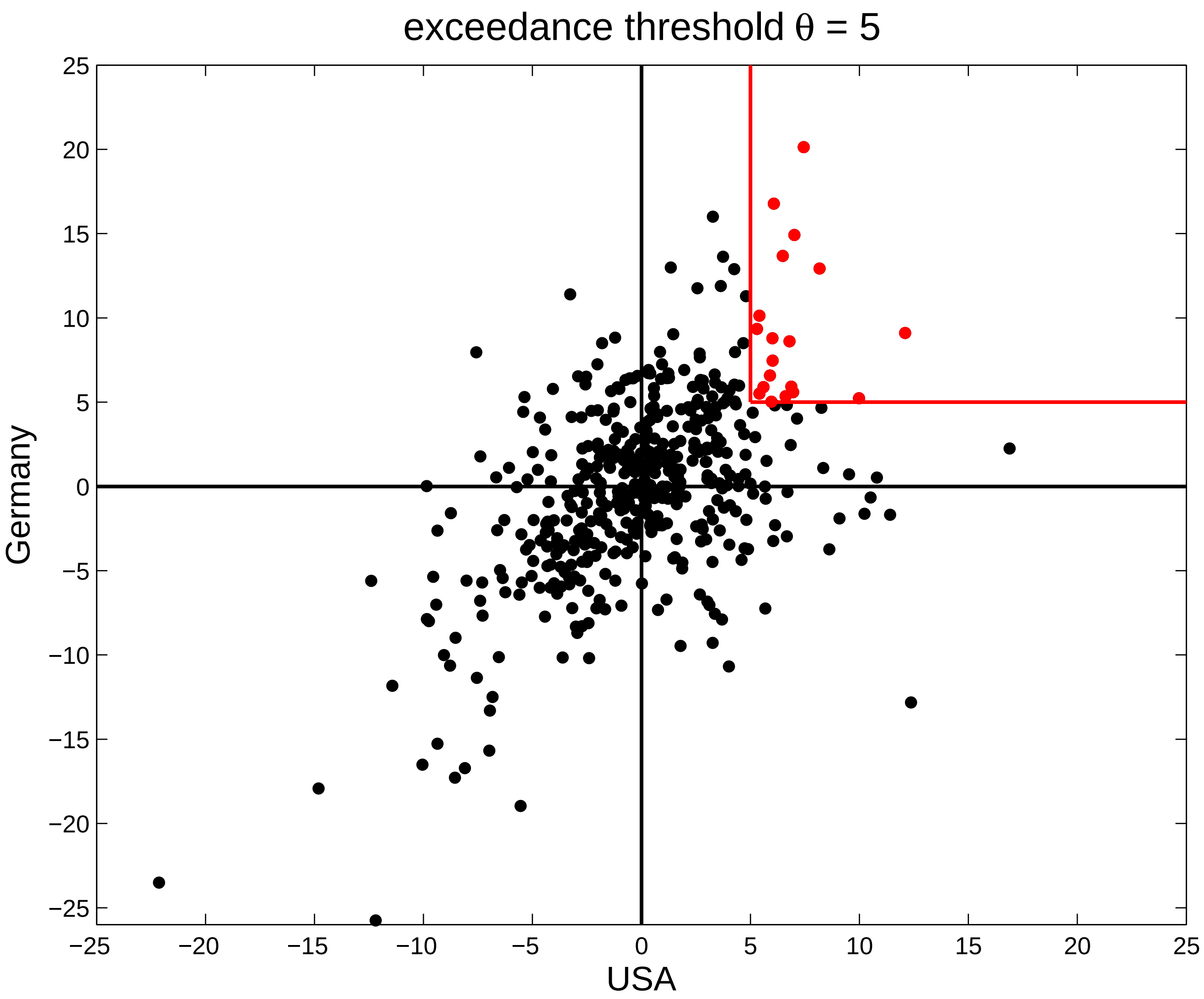
$$\rho(\theta) = \begin{cases} \text{Corr}(x, y | x > \theta, y > \theta) & \text{for } \theta \geq 0 \\ \text{Corr}(x, y | x < \theta, y < \theta) & \text{for } \theta \leq 0 \end{cases} \tag{31}$$

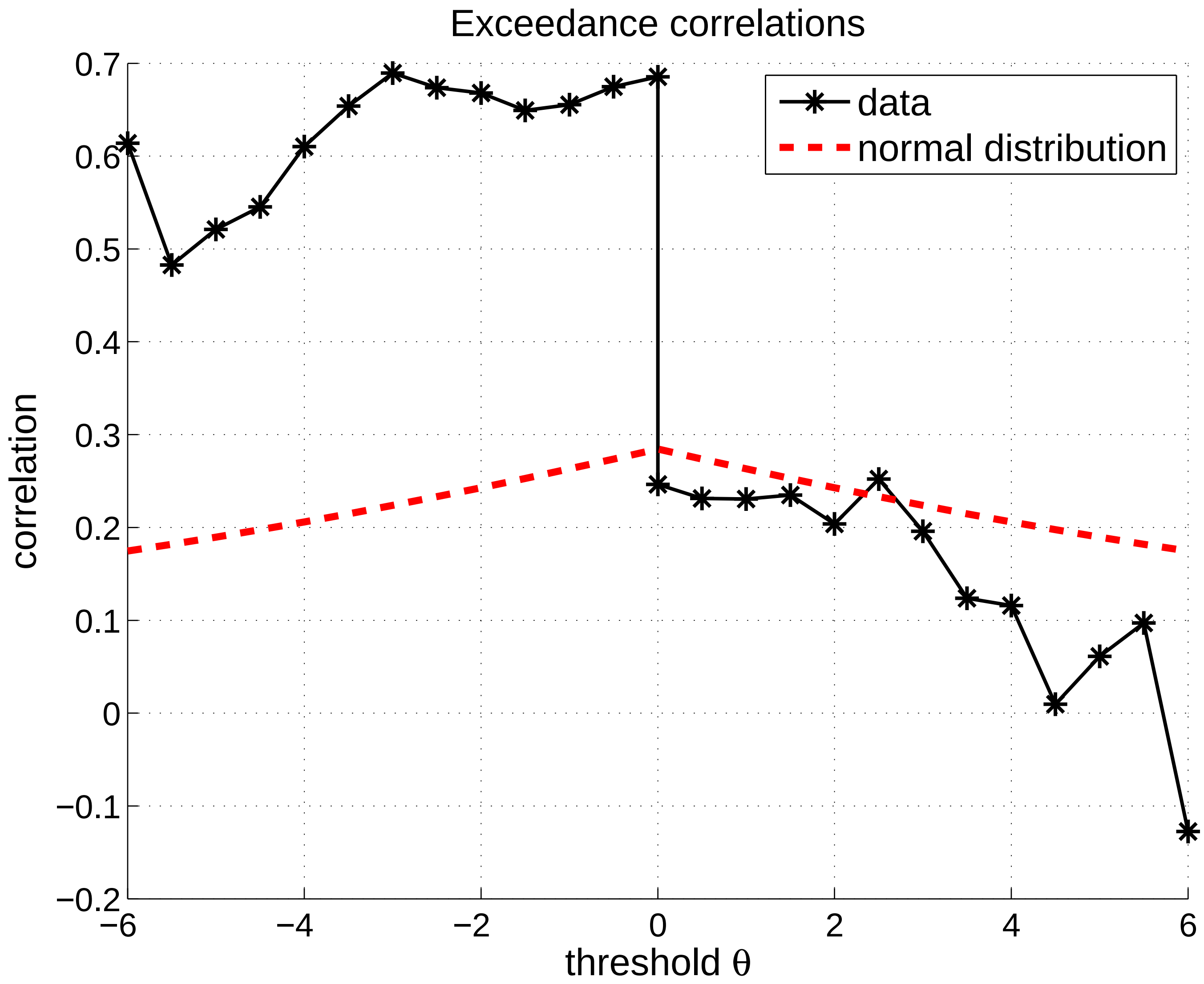
- Let us consider monthly returns of MSCI stock market indices for the US and Germany from January 1970 to June 2008.

[7] Extreme Correlation of International Equity Markets. *Journal of Finance 56, 649-676.*


25
20
15
10
5
0
−5
−10
−15
−20
−25
Germany
−25
−20
−15
−10
−5
0
5
10
15
20
25
USA


exceedance threshold θ = −5
Germany
USA
25
20
15
10
5
0
−5
−10
−15
−20
−25
−25
−20
−15
−10
−5
0
5
10
15
20
25


exceedance threshold θ = 5
Germany
USA
25
20
15
10
5
0
−5
−10
−15
−20
−25
−25
−20
−15
−10
−5
0
5
10
15
20
25



- $r_t = (r_{t,US}, r_{t,Ger})$

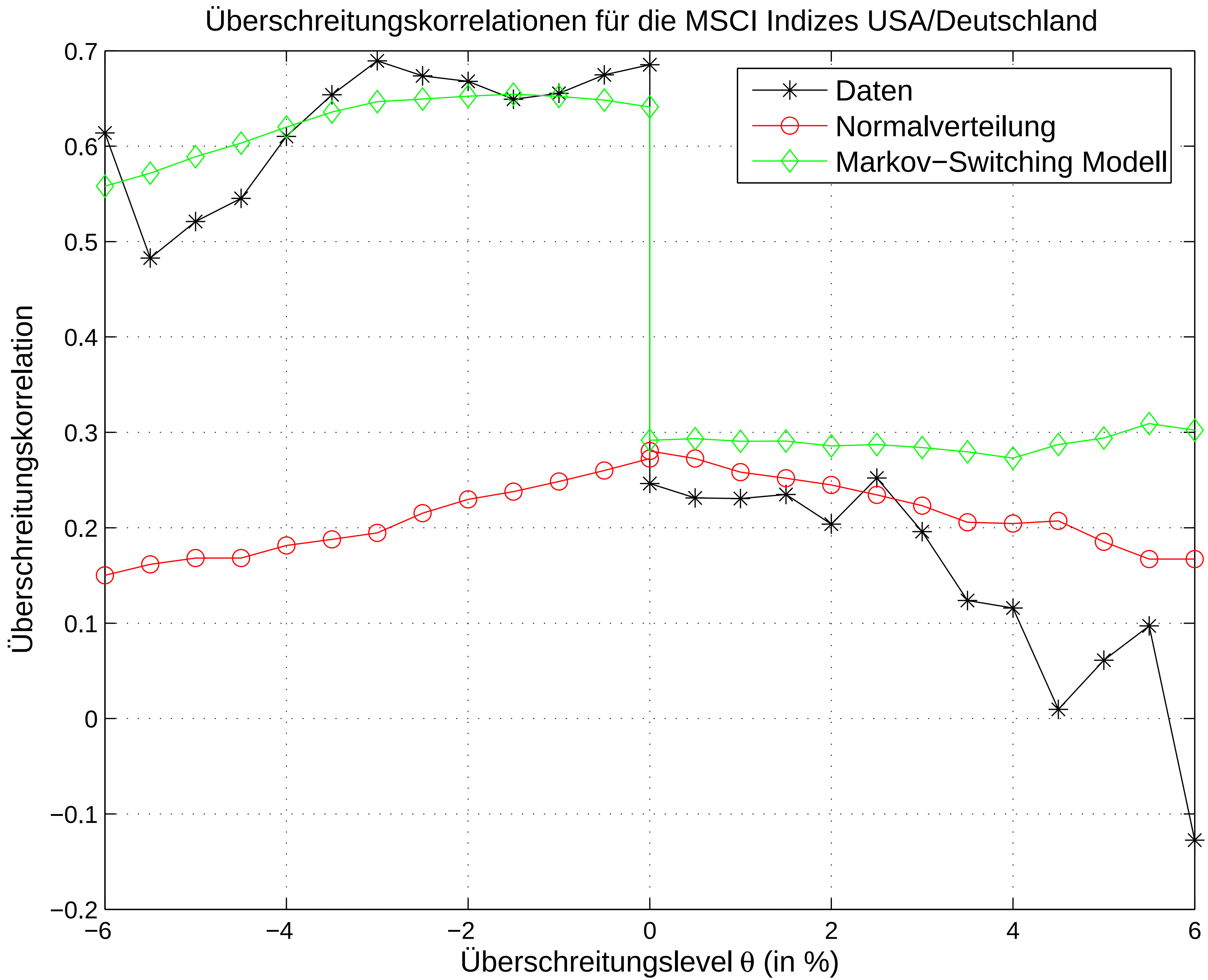
Table 4: Parameter Estimates for three–regime Markov–switching model ($r_t = (r_{t,US}, r_{t,Ger})$)

	Regime 1	Regime 2	Regime 3
mean return	$[0.874, 1.024]$	$[3.479, 2.593]$	$[-1.553, -1.528]$
std. deviation	$[3.091, 3.693]$	$[4.482, 5.810]$	$[5.593, 8.338]$
correlation	0.466	-0.068	0.798
stationary prob.	0.583	0.209	0.208

- transition matrix

$$\widehat{P} = \begin{pmatrix} 0.940 & 0.168 & 0.000 \\ 0.038 & 0.729 & 0.165 \\ 0.022 & 0.103 & 0.835 \end{pmatrix} \qquad (32)$$

- 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
Überschreitungskorrelation
Überschreitungslevel θ (in %)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−0.1
−0.2
−6
−4
−2
0
2
4
6

Observation switching

- Use predetermined variables $x_{1t}, \ldots, x_{pt}$ to model the mixing weights.
- For example, in a two–component mixture, model the weight of the first component in a logistic fashion via

$$\pi_{1t} = \frac{\exp\{\gamma_0 + \gamma_1 x_{1t} + \cdots + \gamma_p x_{pt}\}}{1 + \exp\{\gamma_0 + \gamma_1 x_{1t} + \cdots + \gamma_p x_{pt}\}}. \tag{33}$$

- The choice of the predetermined variables can be based on economic arguments or model (forecasting) performance, depending on the specific application.
- For example, Bauwens et al. (2006) consider a specification where

$$\pi_{1t} = \frac{\exp\{\gamma_0 + \gamma_1 \epsilon_{t-1}^2\}}{1 + \exp\{\gamma_0 + \gamma_1 \epsilon_{t-1}^2\}}, \tag{34}$$

where $\epsilon_t = r_t - \mathsf{E}_{t-1}(r_t)$ is the unexpected shock at time t.

- Then $\gamma_1 > 0$ implies that

$$\pi_{1t} \longrightarrow 1 \quad \text{as } \epsilon_{t-1}^2 \text{ becomes large.} \tag{35}$$

- If the first component has lower volatility, this means that "large shocks have the effect of 'relieving pressure' by reducing the probability of a large shock in the next period."