Models based on the mixture of normal distributions

Motivation

- Propose stress scenarios of simultaneous FX rates movements for CZK / EUR and PLN / EUR based on historical data
- Idea: Use the statistical properties of FX rates changes in turbulent periods
- Problem: In turmoil periods, both the volatility and the correlation increase significantly

Basic idea

- Assumptions:
  - Data are generated from two distinct distributions („quiet“ and „hectic“)
  - Both these distributions are normal
  - Data from each distribution are drawn with some probability

Model

- Model is based on the combination of two multivariate distributions, e.g.

$$(x, y) \sim \begin{cases} \mathrm{MVN}\left(\begin{bmatrix} \mu_{x1} \\ \mu_{y1} \end{bmatrix}, \begin{bmatrix} \sigma_{x1}^2 & \sigma_{x1}\rho_1\sigma_{y1} \\ \sigma_{x1}\rho_1\sigma_{y1} & \sigma_{y1}^2 \end{bmatrix}\right) & \text{with probability } 1-\omega \quad \text{(quiet days),} \\ \mathrm{MVN}\left(\begin{bmatrix} \mu_{x2} \\ \mu_{y2} \end{bmatrix}, \begin{bmatrix} \sigma_{x2}^2 & \sigma_{x2}\rho_2\sigma_{y2} \\ \sigma_{x2}\rho_2\sigma_{y2} & \sigma_{y2}^2 \end{bmatrix}\right) & \text{with probability } \omega \quad \text{(hectic days),} \end{cases}$$

(x and y denotes logaritmic changes in the FX rates)

- Problem: 11 parameters!

Alternative approach

- Compute the conditional correlation between time series assuming that one of them has been generated from distribution corresponding to the hectic period
  - We will use that for each individual observation, we can calculate the probability that it was drawn from the hectic distribution
  - Hectic distributions are not defined apriori!

Model formulation

- Assume that one of the exchange rates is the „core“ one
  - Denote the relative changes of this echange rate by x_t
- Formulation of the model for this exchange rate:

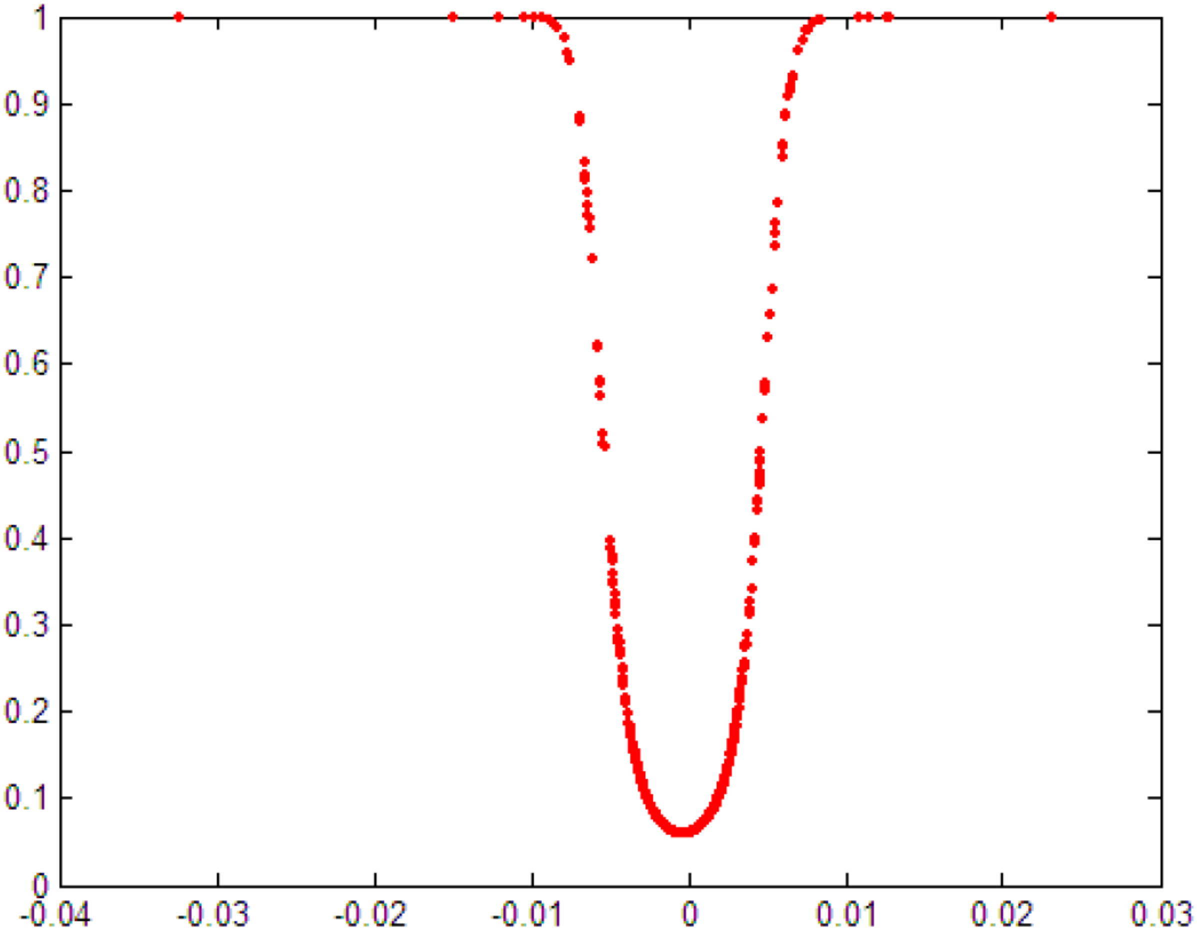
$$x \sim \begin{cases} \mathrm{UVN}(\mu_{x1}, \sigma_{x1}^2) & \text{with probability } 1-\omega \quad \text{(quiet days)}, \\ \mathrm{UVN}(\mu_{x2}, \sigma_{x2}^2) & \text{with probability } \omega \quad \text{(hectic days)}, \end{cases}$$

- Probability that a particular relative change was drawn from the hectic distribution:
  - Conditional probability of the hectic period (Bayes formula)

$$H(x_t) = \frac{\omega \; f(x_t | \, \mu_2, \sigma_2^2)}{(1-\omega) \; f(x_t | \, \mu_1, \sigma_1^2) \; + \; \omega \; f(x_t | \, \mu_2, \sigma_2^2)}$$

Model formulation

- Probability (denoted by $H(x_t)$) that a particular observation is drawn from hectic distribution



Application of the model

- Assume that we have determined the scenario for the change in the „core“ exchange rate. How the corresponding changes in other (=peripheral) exchange rate (denoted by y_t) can be calculated?

$$\mu_{y_2} = \frac{\sum_{t=1}^{n} H(x_t)\, y_t}{\sum_{t=1}^{n} H(x_t)}.$$

$$\sigma_{y_2} = \sqrt{\frac{\sum_{t=1}^{n} H(x_t)\,(y_t - \mu_{y_2})^2}{\sum_{t=1}^{n} H(x_t)}}.$$

$$\rho_2 = \frac{\sum_{t=1}^{n} H(x_t)\,(x_t - \mu_{x_2})\,(y_t - \mu_{y_2})}{\sigma_{x_2}\,\sigma_{y_2}\sum_{t=1}^{n} H(x_t)}$$

- **Exercise:** How the conditional values of the parameter corresponding to the quiet period for the y_t can be calculated?

Application of the model

- Given that we know the mean and volatility for the core as well as for all peripheral FX rates together with correlation between them, the values for the stress scenarios can be calculated using the following formula:

$$\left(\frac{y_t - \mu_y}{\sigma_y}\right) = \rho \left(\frac{x_t - \mu_x}{\sigma_x}\right) + \sqrt{(1 - \rho^2)}\, \varepsilon_t$$

Parameter estimation

- Maximum likelihood method
- The probability density function of the model based on mixture of two normal distributions has the following form:

$$f(x_i,(\mu_1,\sigma_1,\mu_2,\sigma_2,\omega)) = (1-\omega)\,\frac{1}{\sqrt{2\pi\sigma_1^2}}\,e^{-\frac{(x_i-\mu_1)^2}{2\sigma_1^2}} + \omega\,\frac{1}{\sqrt{2\pi\sigma_2^2}}\,e^{-\frac{(x_i-\mu_2)^2}{2\sigma_2^2}}$$

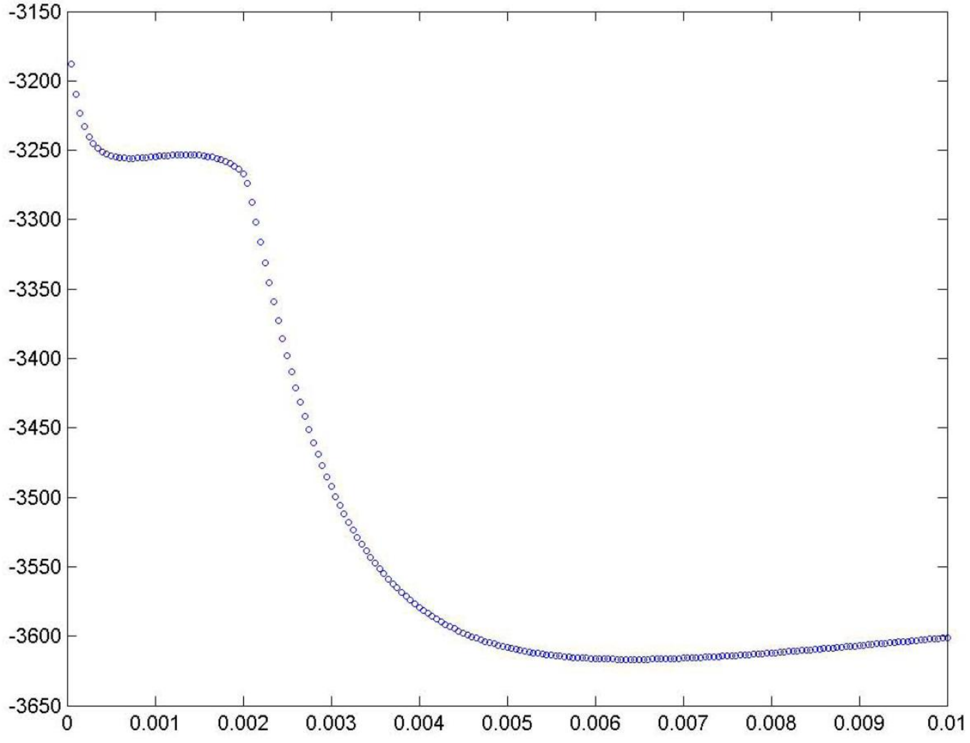
- The logarithmic likelihood function is a function of 5 parameters:

$$\ln L(\mathbf{x},(\mu_1,\sigma_1,\mu_2,\sigma_2,\omega)) = \ln\prod_{i=1}^{n} f(x_i,(\mu_1,\sigma_1,\mu_2,\sigma_2,\omega)) = \sum_{i=1}^{n}\ln f(x_i,(\mu_1,\sigma_1,\mu_2,\sigma_2,\omega)).$$

- Exercise: Propose a statistical test to the test that the model based on the mixture of two normal distribution is significantly better than the model based on one normal distribution

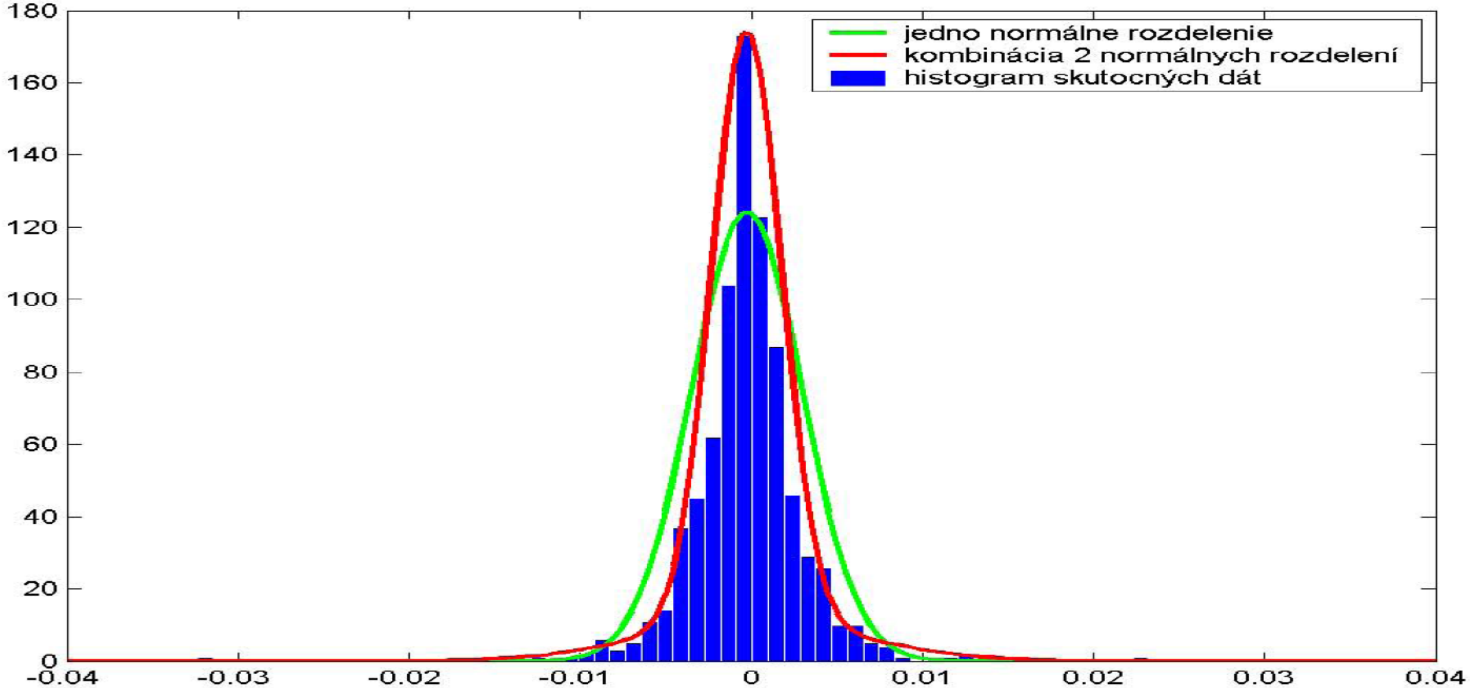
Parameter estimation

- In the bachelor thesis of Potisková (2007), the method of simulated annealing has been used (stochastic optimalization method)
- Problem with non-convexity of the likelihood function in the parameter σ_2



Quality of the model

Comparison of the probability density functions corresponding to the emprical distribution and the distribution implied by the model



Source: Potisková, L. (2007)

Statistical significance of the model based on the mixture of normal distributions can be tested using the likelihood ratio test

References

- Kim, J. – Finger, Ch. C. (2000): A Stress Test to Incorporate Correlation Breakdown, RiskMetrics Journal, Volume 1

- Application to the stress testing of FX risk in the NBS:

  Jurča, P. – Rychtárik, Š. (2006): Stress testting the Slovak banking sector, Banking journal BIATEC, Vol. XIV, No. 4, National Bank of Slovakia

- Lucia Potisková (2007): Stresové testovanie s využitím modelu založenom na kombinácii normálnych rozdelení (Bachelor thesis, in Slovak)