Time Series Models for Measuring Market Risk

José Miguel Hernández Lobato

Universidad Autónoma de Madrid, Computer Science Department

June 28, 2007
Outline

1. Introduction
2. Competitive and collaborative mixtures of experts
3. GARCH processes with non-parametric innovations
4. Conclusions and future work
Introduction

1. Introduction
2. Competitive and collaborative mixtures of experts
3. GARCH processes with non-parametric innovations
4. Conclusions and future work
**Risk, market risk and risk measures**

**Risk** has two components:
1. Uncertainty, and
2. Exposure.

**Market risk** is caused by exposure to uncertainty in the market value of a portfolio.

The risk involved by a situation is captured by the **probability distribution** $\mathcal{P}$ of its possible outcomes. A **risk measure** $\rho$ maps $\mathcal{P}$ to a **risk measurement**:

$$\rho : \mathcal{D} \rightarrow \mathbb{R} \cup \pm \infty,$$

(1)
Risk measures: Value at Risk and Expected Shortfall

Value at Risk (VaR)

It is the worst result within the $\alpha$ fraction of best results:

$$\rho_{\text{VaR}}(\mathcal{P}) = -\mathcal{P}^{-1}(1 - \alpha). \quad (2)$$

Expected Shortfall (ES)

It is the average result when the result is worse than the VaR:

$$\rho_{\text{ES}}(\mathcal{P}) = -\frac{1}{1 - \alpha} \int_{-\infty}^{\mathcal{P}^{-1}(1 - \alpha)} x \, d\mathcal{P}(x). \quad (3)$$

$\alpha$ is usually high, e.g. 0.95 or 0.99.
A risk measuring example

Figure: Density for the loss after holding 100 euros of an imaginary asset for one year. 0.95 VaR is 29.5 euros and the 0.95 ES is 35.3 euros.

However, we do not know the distribution of future price changes in real financial assets!!! We have to estimate it from historical data.
Price variations

- We focus on logarithmic returns: \( r_t = \log(P_t) - \log(P_{t-1}) \).

Figure: Left, daily prices of Microsoft stocks from January 4, 2000 to April 18, 2007. Right, corresponding daily percentage logarithmic returns of Microsoft stocks.
Statistical properties of returns

Linear autocorrelation, time-dependent volatility and heavy tails.

Figure: Left, histogram for the returns of Microsoft stocks and Gaussian density. Right, normal q-q plot for returns of Microsoft stocks.
Volatility models: GARCH processes

We say \( \{r_t\}_{t=1}^{T} \) follows a GARCH(1,1) process if:

\[
\begin{align*}
    r_t &= \sigma_t \varepsilon_t, \\
    \sigma_t^2 &= \gamma + \alpha |r_t|^2 + \beta \sigma_{t-1}^2,
\end{align*}
\]

where \( \varepsilon_t \sim \mathcal{N}(0, 1), \gamma > 0, \alpha \geq 0, \beta \geq 0, \alpha + \beta < 1. \)

- The volatility process is correctly modeled (more or less).
- The heavy-tails are not.
Example

Figure: Left, volatility of Microsoft returns estimated by a GARCH process. Right, normal q-q plot for the GARCH residuals.
Validating market risk models: **Backtesting**

**Step 1**
- Normalized training window
- Test point

**Step 2**
- Normalized training window
- Test point

**Step 3**
- Normalized training window
- Test point

- Normal distribution sample under the null hypothesis
- Test based on the functional delta method
- p-value

Functional

---

11 / 32
Outline

1. Introduction

2. Competitive and collaborative mixtures of experts

3. GARCH processes with non-parametric innovations

4. Conclusions and future work
Mixture of experts paradigm

- Mixture of experts paradigm
- Competitive and collaborative mixtures of experts
The model

Volatility is modeled by a GARCH process with mean $\mu$:

$$ r_t = \mu + \sigma_t \varepsilon_t $$

$$ \sigma^2_t = \gamma + \alpha |r_t - \mu|^2 + \beta \sigma^2_{t-1}, $$

We eliminate heteroskedasticity with the normalization:

$$ z_t = r_t - \mu \sigma_t. $$

Linear autocorrelations and heavy-tails in $\{z_t\}_{t=1}^T$ are modeled by competitive or collaborative mixtures of experts.
Example of a normalization process

Log-Return vs. Day

Normalized Log-Return vs. Day
The experts

- First order autoregressive processes:
  \[ z_t = \phi_0 + \phi_1 z_{t-1} + \sigma \varepsilon_t , \]  
  where \(|\phi_1| < 1, \varepsilon_t \sim \mathcal{N}(0, 1)\) and \(\sigma > 0\).
- Mixtures include up to 3 experts on a **single level**.
The gating network

Time Series Models for Measuring Market Risk
Competitive and collaborative mixtures of experts
### The strategies

**Collaboration**

The output of the mixture is a **weighted average** of the output of the experts. The gating network determines the weights.

**Hard Competition**

Only **one expert** is **active at a given time**. The gating network outputs 1 for that expert and 0 for the others.

**Soft Competition**

The output of the network determines the **probability** of each expert to be the single one that generates the output of the system.
Fitting the mixtures

- Models are fitted by the maximum likelihood method.

**Optimization problems**

Sometimes an expert gets stuck to a single data point close to its mean. This makes the variance of the expert tend to 0.

**Solution based on quasi-bayesian techniques**

We include a prior for the variances of the experts as if we had observed 0.1 points with sample variance 1.5 known to have been generated by each expert. This penalizes small variances.
Sliding window experiment

We take the time series of log-returns of the Spanish stock index IBEX-35 from 12/29/1989 to 1/31/2006. 4034 points.

Each mixture is trained on a normalized window of 1000 points and tested on the first out-of-sample point.

The window moves forward and the process repeats obtaining 3034 points that should be standard normal distributed. We apply statistical tests to those points.
## Test results

<table>
<thead>
<tr>
<th>#AR(1)</th>
<th>Strategy</th>
<th>VaR</th>
<th>Exc</th>
<th>Expt. Shortfall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>99%</td>
<td>95%</td>
<td>99%</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>0.01</td>
<td>0.93</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99%</td>
<td>95%</td>
<td>3 $\cdot 10^{-8}$</td>
</tr>
<tr>
<td>2</td>
<td>CL</td>
<td>0.01</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>SC</td>
<td>0.26</td>
<td>0.56</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.05</td>
<td>0.48</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99%</td>
<td>95%</td>
<td>2 $\cdot 10^{-9}$</td>
</tr>
<tr>
<td>3</td>
<td>CL</td>
<td>0.02</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>SC</td>
<td>0.44</td>
<td>0.31</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>HC</td>
<td>0.02</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99%</td>
<td>95%</td>
<td>8 $\cdot 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99%</td>
<td>95%</td>
<td>4 $\cdot 10^{-6}$</td>
</tr>
</tbody>
</table>
Analysis

- A single Gaussian distribution with time dependent mean and variance (Hard Competition and Collaboration) fail to model the heavy tails of the conditional distribution of log-returns.

- Soft Competitive models predict the conditional distribution as a mixture of Gaussians, a much more expressive model.
Outline

1. Introduction
2. Competitive and collaborative mixtures of experts
3. GARCH processes with non-parametric innovations
4. Conclusions and future work
The model

A power autoregressive GARCH(1,1,1) with non-parametric innovations:

\[
\begin{align*}
    r_t &= \phi_0 + \phi_1 r_{t-1} + \sigma_t \varepsilon_t \\
    \sigma_t &= \gamma + \alpha |r_t - \phi_0 - \phi_1 r_{t-2}| + \beta \sigma_{t-1},
\end{align*}
\]

where \( \varepsilon_t \sim \hat{f}(x; \{c_i\}_{i=1}^N) \) and

\[
\hat{f}(x; \{c_i\}_{i=1}^N) = \frac{1}{Nh} \sum_{i=1}^N K \left( \frac{x - c_i}{h} \right),
\]

where \( K \) is the standard Gaussian density. \( \theta = \{\phi_0, \phi_1, \gamma, \alpha, \beta\} \).
Training procedure: maximum likelihood

If $u_t = r_t - \phi_0 - \phi_1 r_{t-1}$ is the empirical autoregressive residual of the process ($u_0 = 0$, $u_1 = r_1 - \hat{\mu}$ and $\sigma_0 = \hat{\sigma}$) we maximize:

$$
L(\theta, \{c_i\}_{i=1}^N | \{r_t\}_{t=1}^T) = \prod_{t=1}^{T} \hat{f} \left( \frac{u_t}{\sigma_t}; \{c_i\}_{i=1}^N \right) \frac{1}{\sigma_t}, \quad (13)
$$

If $N = T$ we just have to repeat

1. We standardize $\{u_t/\sigma_t\}_{t=1}^T$.
2. We hold $\theta$ fixed and maximize (13) with respect to $\{c_i\}_{i=1}^T$ by setting $c_t = u_t/\sigma_t$, $t = 1, \ldots, T$.
3. We fix $h$ by means of $\{c_i\}_{i=1}^T$ and Silverman’s rule.
4. We hold $\{c_i\}_{i=1}^T$ fixed and maximize (13) with respect to $\theta$ by means of a non-linear optimization routine.
Problem

Kernel density estimates have difficulty modeling leptokurtic densities. Solution: perform the estimation in a transformed space where the data looks more Gaussian.

![Kernel Density Estimate for IBM Returns](image1)

![Density Estimate in a Transformed Space](image2)
Choosing a good transformation

We choose \( g_\lambda(x) = \Phi^{-1}(\hat{F}_\lambda(x)) \). Where \( \Phi \) is the standard Gaussian distribution and \( \hat{F} \) is a stable distribution. \( \lambda \) is found by maximum likelihood to \( \{u_t/\sigma_t\}_{t=1}^T \).
We perform another sliding window experiment with the returns of IBM stocks from 1962/07/03 to 1998/12/31. 9190 points.
## Test results

<table>
<thead>
<tr>
<th>Model</th>
<th>VaR</th>
<th>Exc</th>
<th>Expt. Shortfall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99%</td>
<td>95%</td>
<td>99%</td>
</tr>
<tr>
<td>stable</td>
<td>1.5e-3</td>
<td>0.01</td>
<td>3.7e-5</td>
</tr>
<tr>
<td>non-parametric</td>
<td>0.51</td>
<td>0.66</td>
<td>0.49</td>
</tr>
<tr>
<td>mix. of 2 exp.</td>
<td>5.7e-3</td>
<td>0.30</td>
<td>0.025</td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. Competitive and collaborative mixtures of experts
3. GARCH processes with non-parametric innovations
4. Conclusions and future work
Conclusions

Asset returns show time-dependent volatility and heavy tails.

Soft competitive mixtures predict a distribution which is a mixture of a few Gaussians and can account for the heavy tails.

GARCH processes with non-parametric innovations outperform soft competitive mixtures because they employ constrained mixtures of thousands of Gaussians without causing overfitting.
Future work

- Risk measuring in multi-asset portfolios.
- Use high frequency data to build improved models.
- Measure risk with a higher horizon in the future.
- Employ different risk measures.