Sparse and robust normal and t-portfolios by penalized $L_q$-likelihood minimization

Davide Ferrari$^1$, Margherita Giuzio$^2$, Sandra Paterlini$^2$

$^1$University of Melbourne, AU
$^2$EBS Universität für Wirtschaft und Recht, GER

Università degli Studi di Bergamo

Seminar 07.04.2014
Portfolio Selection

Task: Select the "optimal" assets and their "optimal" weights

- **Ideal Characteristics**
  - Good Out-of-sample Performance
  - Stable weights
  - Robustness
  - Cheap to implement and maintain

- **Main Issues**
  - Normality assumption
  - Highly correlated returns
  - Sensitivity of Weights to Estimation Errors and Model Misspecification
  - High-dimensionality
Main Issue

S&P500 Log-Return

Sample Quantiles

Theoretical Quantiles
• **Robust and Shrinkage Estimators** $\implies$ Estimation Errors
  Huber (1981)
  Ledoit and Wolf (2004)

• **Penalized Least Squares** $\implies$ Model Misspecifications and Sparsity
  DeMiguel, Garlappi, Nogales, Uppal (2007)
  Fan, Zhang, Yu (2012)
  Fastrich, Paterlini, Winker (2012)

• **Minimum Divergence Methods** $\implies$ Model Misspecifications
  Ferrari, Yang (2010)
  Ferrari, Paterlini (2010)
  Ferrari, La Vecchia (2012)
Our Proposal

Targets:
- **Robustness** to both Estimation Errors and Model Misspecification
- **Sparsity**: automatically select a small number of relevant active positions

**How can we achieve that?**

**Generalized Description Length Approach**
- based on $q$-Entropy
- using priors as sparsifying operator
Maximum Likelihood

• If observations are i.i.d. from a probability distribution \( f(x; \theta_0) \), maximizing

\[
\log L(\theta_0) = \log \prod_{i=1}^{n} f(X_i \mid \theta) = \sum_{i=1}^{n} \log f(X_i; \theta)
\]

we obtain the maximum log-likelihood estimator of \( \theta_0 \).

\[
\sum_{i=1}^{n} \log f(X_i; \theta) \to E[\log f(X_i; \theta)],
\]

as \( n \to \infty \)

• Properties
  • Consistency, Efficiency
  • Asymptotic normality
  • Easy to implement
Shannon Entropy

- Information Theory (1940s): how to measure uncertainty with respect to a probability distribution $f(x)$

\[ H(X) = -E[\log f(x)] \]

- $- \log f(x)$ is the information included in the observation $x$
- $H(X)$ represents the average uncertainty removed after observing the outcome of the variable $X$.

**Relationship**

Given $n$ i.i.d. observations, the Maximum Likelihood Estimator can be seen as a minimization of the Shannon entropy

\[ \max_{\theta} \sum_{i=1}^{n} \log f(X_i; \theta) = \min -E[\log f(X_i; \theta)] \]
Havrda and Charvát (1967) proposed a generalized measure of entropy of order $q$, then used in physics, biology and finance

$$H(X) = \frac{E[1 - f(x)^{q-1}]}{1 - q}, \ q > 0$$

Tsallis (1988) arrived to the following specification:

$$H(X) = -E[L_q f(x)],$$

where

$$L_q = \begin{cases} 
(u^{1-q} - 1)/(1 - q), & q \neq 1, \\
\log(u), & q = 1, 
\end{cases}$$
Maximum $L_q$ Likelihood

- Given $n$ i.i.d. observations from a probability distribution $f(x; \theta_0)$, maximizing

$$\sum_{i=1}^{n} L_q[f(X_i; \theta)]$$

we obtain the Maximum $L_q$-Likelihood Estimator of $\theta_0$.

- Properties
  - Consistency and Fisher-consistency (Ferrari, La Vecchia, 2009)
  - Asymptotic normality
  - Changing $q$ we balance the trade-off between bias and variance, robustness and efficiency
  - When $q \to 1$, ML$q$E $\to$ MLE
Let:

- $\mathbf{X} = (X_1, \ldots, X_p)^T$ a $p$-dimensional random vector such that $E(\mathbf{X}) = \mu$ and $\text{Var}(\mathbf{X}) = \Sigma$

- $\mathbf{Y} = \beta^T \mathbf{X}$ a portfolio where $\beta^T$ is the weights vector

- the portfolio mean, $\mu^*$ is a fixed target
Generalized Description Length criterion (GDL)

Assumptions

Distribution of the Data: \( f(y; \mu, \sigma^2) = f(x_i^T \beta; \mu, \sigma^2) \)

Prior Distribution of Portfolio Weights: \( \pi(\beta_j; \lambda) \) for the \( j \)th weight \( \beta_j \)

\[
\hat{\beta} = \arg\min_{\beta} \left\{ -\sum_{i=1}^{n} L_q \left\{ f \left( \frac{x_i^T \beta - \mu^*}{\sigma} \right) \right\} - \sum_{j=1}^{p} L_q \left\{ \pi(\beta_j; \lambda) \right\} \right\}, \quad (1)
\]

with \( q \)-entropy function

\[
L_q = \begin{cases} 
(u^{1-q} - 1)/(1 - q), & q \neq 1, \\
\log(u), & q = 1
\end{cases}
\]

- Information from the data given the model
- Information from the model itself
Penalized $q$-entropy minimization

Computing the first derivative

$$0 = \sum_{i=1}^{n} w_i(x_i, \beta, \sigma) \nabla \log f(\sigma^{-1}(x_i^T \beta - \mu^*)) + \sum_{j=1}^{p} v_j(\beta_j, \lambda) \nabla \log \pi(\beta_j; \lambda).$$

(2)

Define $w_i$ and $v_j$ as

- Weights on obs $y$
  $$w_i(x_i, \beta, \sigma) = f(\sigma^{-1}(x_i^T \beta - \mu^*))^{1-q}$$

- Weights on $\beta$
  $$v_j(\beta_j, \lambda) = \pi(\beta_j; \lambda)^{1-q}$$

Role of $q$

- Changing $q$, we modify the role of unusual observations $x_i$ and sparsify the parameters $\beta_j$
- When $q < 1$ we gain robustness to models misspecification
GDL Penalty Function

\[- \sum_{j=1}^{p} L_q \{ \pi(\beta_j; \lambda) \}\]
GDL vs Other Penalties

- GDL Penalty
- Lasso Penalty
- SCAD Penalty
- Logarithm Penalty

Ferrari, Giuzio, Paterlini (AU & GER)
Model 1: Normal portfolios with Laplace prior

Model Specification

\( Y \sim N_1(\mu, \sigma^2) \)

\( \pi(\beta_j; \lambda) = \lambda \exp \{-\lambda |\beta_j|\}/2 \)

- **GDL N**

\[
\hat{\beta}^{(s)} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \hat{w}_i^{(s-1)} \frac{1}{2} \left( \frac{\mu^* - x_i^T \beta}{\hat{\sigma}(s-1)} \right)^2 + \lambda \sum_{j=1}^{p} \hat{v}_j^{(s-1)} |\beta_j| \right\} \tag{3}
\]

- **Lasso**

\[
\hat{\beta}^{(s)} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (\mu^* - x_i^T \beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \tag{4}
\]
Model 2: \(t\)-portfolios with Laplace prior

Model Specification

\[ Y \sim t\text{-Student} \quad f_{\nu}(y; \mu, \sigma^2) = N(\mu, \sigma^2 Z^{-1}_i), \quad Z_i \sim \Gamma(\nu/2, \nu/2) \]

\[ \pi(\beta_j; \lambda) = \lambda \exp \{-\lambda |\beta_j|\}/2 \]

- **GDL \(t\)**

\[
\text{argmin}_{\beta, \sigma} \left\{ - \left( \frac{\nu + 1}{2} \right) \sum_{i=1}^{n} \hat{w}_i^{(s-1)} \log \left\{ 1 + \frac{(x_i \beta^T - \mu^*)^2}{\nu \sigma^2} \right\} + \lambda \sum_{j=1}^{p} \hat{v}_j^{(s-1)} |\beta_j| \right\}
\]

(5)

where \(\nu_q = q\nu + (q - 1)\).
Behaviour of $\hat{w}_i$ – Observations Weights

$$\hat{w}_i^{(s)} = f((x_i \hat{\beta}^{(s-1)} - \mu^*)/\hat{\sigma}^{(s-1)})^{1-q} \quad (6)$$

$q = 0.5$

$q = 0.9$
Behaviour of $\hat{\nu}_j$ – Parameters Weights

$$\hat{\nu}_j^{(s)} = \pi(\hat{\beta}_j^{(s-1)}; \lambda)^{1-q}.$$  \hspace{1cm} (7)

\begin{align*}
q &= 0.5 \\
q &= 0.9
\end{align*}

Minimizing the GDL criterion

- Modify the role of extreme observations and parameters by varying $q$
- Minimize the divergence between the hypothetical and the transformed version of the density and prior distributions
Selection of $\lambda$

Select $\lambda$ (for a given $q$) by minimizing

- "Standard" AIC and BIC
- Robust AIC and BIC (Ronchetti, 1997 and Machado, 1993)

\[
AIC_q = -2 \sum_{i=1}^{n} L_q \left\{ f \left( \frac{x_i^T \hat{\beta}_{q,\lambda} - \mu^*}{\hat{\sigma}_{q,\lambda}} \right) \right\} + 2k, \quad (8)
\]

\[
BIC_q = -2 \sum_{i=1}^{n} L_q \left\{ f \left( \frac{x_i^T \hat{\beta}_{q,\lambda} - \mu^*}{\hat{\sigma}_{q,\lambda}} \right) \right\} + \log(n)k, \quad (9)
\]

where $k$ is the number of active portfolio positions
Geometric Interpretation of GDL

Example

Let's assume two asset returns have bivariate normal distribution

\[
\begin{pmatrix}
X_1 \\
X_2
\end{pmatrix}
\sim \mathcal{N}_2 \left( \begin{pmatrix}
0 \\
1
\end{pmatrix}, \begin{pmatrix}
2 & 1/\sqrt{2} \\
1/\sqrt{2} & 1
\end{pmatrix} \right),
\]

When comparing Sharpe Ratios, the second asset is preferable.

- Compare the optimal portfolios obtained with GDL N and LASSO
- Contaminated vs not contaminated data
Geometric Interpretation of GDL

Ferrari, Giuzio, Paterlini (AU & GER)

Penalized Lq-likelihood minimization
Geometric Interpretation of GDL
Simulation Study

Empirical Setup

- \( n = 200 \) observations, \( p = 50 \) assets, \( k = 10 \) optimal positions
- Four settings: \( \rho = 0.2, 0.4, 0.6, 0.8 \), 50 simulations
- \( q = 0.9 \), target return \( \mu^* = k = 10 \)

Model 1: \( \mathbf{X} \sim N_p(\mu, \Sigma) \)

\[
\mu_j = \begin{cases} 
1, & j \leq k, \\
0, & j > k, 
\end{cases}
\]

\[
\Sigma_{ij} = \begin{cases} 
1, & i = j, \\
\rho, & i \neq j, 
\end{cases}
\]

\( \nu = 6 \).
Simulation Study

Goals:

- Portfolio with low risk and high return, compared to a target
- Sparsity

In each simulation:

- Given a grid of $\lambda$, compute estimates of $\hat{\beta}$ with GDL N, GDL t, Lasso, Zhang, Scad, Log, Lq
- Choose the model with the lowest BIC
- Compare $\hat{\beta}$ to the optimal vector of $\beta^*$
Simulation Study

- Assess the whole performance by the Monte Carlo mean squared error

\[
\hat{MSE} = \frac{1}{50} \sum_{b=1}^{50} \left( \frac{\mu^T \hat{\beta}_b - \mu^*}{\sqrt{\hat{\beta}_b^T \Sigma \hat{\beta}_b}} \right)^2,
\]

where:

- \( \mu^*/\sigma^* \rightarrow \) target Sharpe ratio with \( \sigma^* = \sqrt{\hat{\beta}_b^T \Sigma \hat{\beta}_b} \)

- \( \mu^T \hat{\beta}_b / \sqrt{\hat{\beta}_b^T \Sigma \hat{\beta}_b} \rightarrow \) portfolio Sharpe ratio

- Verify the selection of the \( k \) active assets by the F-measure:

\[
F\text{-measure} = 2 \frac{|supp(\beta^*)| \cap |supp(\hat{\beta})|}{|supp(\beta^*)| + |supp(\hat{\beta})|},
\]

where \( supp(\beta) = \{j : |\beta_j| \geq \tau\} \).
Simulation Results – Sparsity

Model 1

Model 2
Simulation Results – MSE

Model 1

Model 2

Mean Squared Error

GDL N
GDL t
Lasso
Zhang
Log
Lq

Ferrari, Giuzio, Paterlini (AU & GER)  
Penalized Lq-likelihood minimization  
Seminar 07.04.2014
Index Tracking

\( \mathbf{y} \) returns vector of a Financial Index

\( \mathbf{X} \) returns matrix of the Index Components

\( \mathbf{\beta} \) vector of asset weights to be estimated

<table>
<thead>
<tr>
<th>Data</th>
<th>( n )</th>
<th>( p )</th>
<th>( \bar{r} )</th>
<th>( \hat{\sigma} )</th>
<th>( \hat{S} )</th>
<th>( \hat{K} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>F&amp;F 100</td>
<td>1401</td>
<td>100</td>
<td>13.97</td>
<td>18.24</td>
<td>-0.03</td>
<td>5.27</td>
</tr>
<tr>
<td>S&amp;P 200</td>
<td>1401</td>
<td>200</td>
<td>10.70</td>
<td>26.94</td>
<td>-0.23</td>
<td>10.57</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>1401</td>
<td>500</td>
<td>9.80</td>
<td>28.33</td>
<td>-0.30</td>
<td>13.09</td>
</tr>
</tbody>
</table>

Period from 23.08.2002 to 27.03.2008

Goal: build sparse portfolios able to track as closely as possible the index
Financial Data

![F&F 100 Log-returns Abs. Frequency](image1.png)

![S&P 200 Log-returns Abs. Frequency](image2.png)

![S&P 500 Log-returns Abs. Frequency](image3.png)
Performance Evaluation

- Rolling Window of 250 obs, stepsize = 1 obs
- Compare GDL N and GDL t with $q = 0.5, 0.9$
- Lasso and Eq. Weighted (1/N) Portfolios

Performance measures

- Risk-return performance:
  \[ IR = \frac{ER}{TEV} \]

- Sparsity and stability:
  \[ \hat{k}, \text{ average turnover (TO)} \]

- Tracking ability:
  OOS correlation and beta wrt Index
## Performance Evaluation

<table>
<thead>
<tr>
<th>Index</th>
<th>GDL N (q = 0.9)</th>
<th>GDL N (q = 0.5)</th>
<th>GDL t (q = 0.9)</th>
<th>GDL t (q = 0.5)</th>
<th>Lasso</th>
<th>1/N</th>
</tr>
</thead>
<tbody>
<tr>
<td>F&amp;F 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER (%)</td>
<td>0.338</td>
<td>0.302</td>
<td>0.170</td>
<td>0.186</td>
<td>1.030</td>
<td>0.929</td>
</tr>
<tr>
<td>TEV (%)</td>
<td>0.624</td>
<td>0.659</td>
<td>0.492</td>
<td>0.527</td>
<td>2.117</td>
<td>3.854</td>
</tr>
<tr>
<td>IR</td>
<td><strong>0.542</strong></td>
<td>0.458</td>
<td>0.346</td>
<td>0.352</td>
<td>0.486</td>
<td>0.241</td>
</tr>
<tr>
<td>S&amp;P 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER (%)</td>
<td>0.319</td>
<td>0.639</td>
<td>-2.421</td>
<td>-0.196</td>
<td>4.760</td>
<td>5.937</td>
</tr>
<tr>
<td>TEV (%)</td>
<td>4.500</td>
<td>4.961</td>
<td>4.897</td>
<td>4.614</td>
<td>7.267</td>
<td>2.518</td>
</tr>
<tr>
<td>IR</td>
<td>0.071</td>
<td><strong>0.129</strong></td>
<td>-0.494</td>
<td>-0.042</td>
<td>0.655</td>
<td><strong>2.357</strong></td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER (%)</td>
<td>2.906</td>
<td>-1.989</td>
<td>1.192</td>
<td>2.287</td>
<td>2.986</td>
<td>5.113</td>
</tr>
<tr>
<td>TEV (%)</td>
<td>6.966</td>
<td>8.100</td>
<td>9.018</td>
<td>8.859</td>
<td>10.315</td>
<td>3.107</td>
</tr>
<tr>
<td>IR</td>
<td><strong>0.417</strong></td>
<td>-0.245</td>
<td>0.132</td>
<td>0.258</td>
<td>0.289</td>
<td><strong>1.646</strong></td>
</tr>
</tbody>
</table>
### Performance Evaluation

<table>
<thead>
<tr>
<th>Index</th>
<th>GDL N</th>
<th>GDL N</th>
<th>GDL t</th>
<th>GDL t</th>
<th>Lasso</th>
<th>1/N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$q = 0.9$</td>
<td>$q = 0.5$</td>
<td>$q = 0.9$</td>
<td>$q = 0.5$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F&amp;F 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{k}$</td>
<td>37.749</td>
<td>38.244</td>
<td>32.241</td>
<td>32.121</td>
<td>65.939</td>
<td>98</td>
</tr>
<tr>
<td>TO</td>
<td>0.068</td>
<td>0.060</td>
<td>0.066</td>
<td>0.064</td>
<td>0.017</td>
<td>-</td>
</tr>
<tr>
<td>S&amp;P 200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{k}$</td>
<td>36.950</td>
<td>43.627</td>
<td>28.431</td>
<td>28.011</td>
<td>66.532</td>
<td>200</td>
</tr>
<tr>
<td>TO</td>
<td>0.399</td>
<td>0.346</td>
<td>0.520</td>
<td>0.481</td>
<td>0.037</td>
<td>-</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{k}$</td>
<td>44.564</td>
<td>45.736</td>
<td>27.770</td>
<td>26.944</td>
<td>66.407</td>
<td>500</td>
</tr>
<tr>
<td>TO</td>
<td>0.605</td>
<td>0.322</td>
<td>0.811</td>
<td>0.801</td>
<td>0.053</td>
<td>-</td>
</tr>
</tbody>
</table>
Tracking ability - Normalized trends

- F&F 100
- 1/N
- Lasso
- GDL t q=0.5

- S&P 200
- 1/N
- Lasso
- GDL N q=0.5

- S&P 500
- 1/N
- Lasso
- GDL N q=0.9

Ferrari, Giuzio, Paterlini (AU & GER)  
Penalized L_q-likelihood minimization  
Seminar 07.04.2014
**Conclusion**

**Penalized $q$-entropy minimization helps to**

- select a sparse portfolio
- handle highly-correlated variables
- handle outliers and control for estimation errors

**On-going work...**

- test on other financial data
- extend to mean-variance framework
- develop a method to optimally choose parameters $q$ and $\lambda$
Appendix
Other Penalties

- **LASSO**

\[
\lambda \sum_{i=1}^{p} \rho(\beta_i) = \lambda \| \beta \|_1.
\]

- **SCAD**

\[
\lambda \sum_{i=1}^{p} \rho(\beta_i) = \sum_{i=1}^{p} \left\{ \begin{array}{ll}
\frac{\lambda |\beta_i|}{\beta_i^2 + 2a \lambda |\beta_i| - \lambda^2}, & |\beta_i| \leq \lambda, \\
\frac{\beta_i^2 + 2a \lambda |\beta_i| - \lambda^2}{2(a-1)} - \frac{\lambda^2}{(a+1)\lambda^2}, & \lambda < |\beta_i| \leq a\lambda, \\
\frac{(a+1)\lambda^2 - 2a \lambda |\beta_i| + \beta_i^2}{2}, & a\lambda < |\beta_i| \end{array} \right.
\]
Other Penalties

- **Zhang**

\[
\lambda \sum_{i=1}^{p} \rho(\beta_i) = \sum_{i=1}^{p} \left\{ \begin{array}{ll}
\lambda |\beta_i|, & |\beta_i| \leq \eta, \\
\lambda \eta, & \eta < |\beta_i|
\end{array} \right.
\]

- **Lq**

\[
\lambda \sum_{i=1}^{p} \rho(\beta_i) = \lambda \|\beta\|_q^q, \quad 0 < q < 1.
\]

- **Log**

\[
\lambda \sum_{i=1}^{p} \rho(\beta_i) = \lambda \sum_{i=1}^{p} (\log(|\beta_i| + \Phi) - \log(\Phi)).
\]
General Re-weighting Algorithm

Given $q$, $\lambda$ and $\mu^*$

0. At Step $s = 0$, compute initial parameter values $\widehat{\beta}^{(s)}$ and $\widehat{\sigma}^{(s)}$.

1. Set $s = s + 1$, and update the data weights

$$\widehat{w}_i^{(s)} = f\left(\frac{(x_i^T \beta^{(s-1)} - \mu^*)}{\widehat{\sigma}^{(s-1)}}\right)^{1-q}, \quad i = 1, \ldots, n,$$

and the penalty weights

$$\widehat{v}_j^{(s)} = \pi(\beta_j^{(s-1)}; \lambda)^{1-q}, \quad j = 1, \ldots, p.$$

2. Find the parameter values $\widetilde{\beta}$ and $\widetilde{\sigma}$ by minimizing

$$\sum_{i=1}^{n} \widehat{w}_i \log f\left(\frac{(x_i^T \beta - \mu^*)}{\sigma}\right) + \sum_{j=1}^{p} \widehat{v}_j \log \pi(\beta_j; \lambda), \quad (12)$$

3. Compute $\widehat{\beta}^{(s)}$ and $\widehat{\sigma}^{(s)}$ by solving

$$f\left(\frac{(x_i^T \beta - \mu^*)}{\sigma}\right) \propto f\left(\frac{x_i^T \widetilde{\beta} - \mu^*}{\widetilde{\sigma}}\right)^q \text{ for } \beta \text{ and } \sigma.$$

4. Repeat Steps 1 and 2 until a stopping criterion is satisfied.
Model 1: Normal portfolios with Laplace prior

The weights $\hat{w}_i$ and $\hat{v}_j$ are updated using estimates obtained in Step $s-1$ as follows:

$$\hat{w}_i^{(s-1)} = \frac{1}{\sqrt{2\pi\hat{\sigma}^2(s-1)}} \exp \left\{ -\frac{\left( \mu^* - x_i^T \hat{\beta}^{(s-1)} \right)^2}{2\hat{\sigma}^2(s-1)} \right\}^{1-q},$$

(13)

$$\hat{v}_j^{(s-1)} = \left[ \frac{\lambda}{2} \exp \left\{ -\lambda |\hat{\beta}_j^{(s-1)}| \right\} \right]^{1-q}.$$ 

(14)

The portfolio variance is also updated using estimates from Step $s-1$ as

$$\hat{\sigma}^2(s) = \frac{\sum_{i=1}^n \hat{w}_i^{(s-1)} \left( \mu^* - x_i^T \hat{\beta}^{(s-1)} \right)^2}{q \sum_{i=1}^n \hat{w}_i^{(s-1)}}.$$

(15)
Model 2: t-portfolios with Laplace prior

Given $\hat{w}_i^{(s-1)}$ and $\hat{v}_i^{(s-1)}$, set the initial mixture weights $\hat{z}_i = 1/n$. Then, obtain the updated estimates $\hat{\beta}^{(s)}$ and $\hat{\sigma}^{(s)}$ by iterating the following expectation-maximization steps

- **M-Step:**

\[
\beta' = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} \hat{w}_i^{(s-1)} \hat{z}_i^{(s-1)} \frac{1}{2} \left( \frac{x_i^T \beta - \mu}{\hat{\sigma}^{(s-1)}} \right)^2 + \lambda \sum_{j=1}^{p} \hat{v}_j^{(s-1)} |\beta_j| \right\}, \quad (16)
\]

\[
\sigma'^2 = \frac{\sum_{i=1}^{n} \hat{w}_i^{(s-1)} \hat{z}_i^{(s-1)} \left( x_i^T \hat{\beta}' - \mu \right)^2}{\sum_{i=1}^{n} \hat{w}_i^{(s-1)} \hat{z}_i^{(s-1)}} \times \frac{\nu}{(\nu + 1)q - 1}. \quad (17)
\]

- **E-Step:**

\[
\hat{z}_i = \frac{(\nu_q + 1)\sigma'^2}{\nu_q \sigma'^2 + \hat{w}_i^{(s-1)}(x_i^T \beta' - \mu)^2}, \quad i = 1, \ldots, n, \quad (18)
\]

where $\nu_q = (\nu + 1)q - 1$. 

Ferrari, Giuzio, Paterlini (AU & GER) 
Penalized L$q$-likelihood minimization 
Seminar 07.04.2014


