Lasso Quantile Trading Strategy

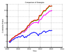
Sergey Nasekin

Wolfgang Karl Härdle

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

Portfolio Management

- portfolio diversification
- portfolio construction
- asset allocation.



Preliminary Comparison - S&P 500 Stocks

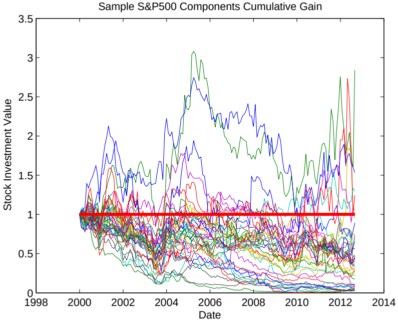
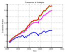


Figure 1: 38 random S&P 500 Sample Components' Cumulative Return: **87%** of stocks lost the value of the initial investment (thick red line)



Preliminary Comparison - Hedge Funds

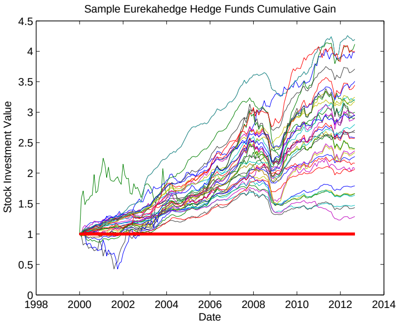
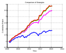


Figure 2: 38 Eurekahedge Hedge Funds Indices' Cumulative Return: **0%** of funds lost the value of the initial investment (thick red line)

How Hedge Funds Can Help?

A **hedge fund** is an "aggressively managed portfolio of investments that uses advanced investment strategies such as leveraged, long, short and derivative positions in both domestic and international markets with the goal of generating high returns" (Investopedia).

- diversification - reduction of the portfolio risk
- construction - a more diverse universe of assets
- allocation - a higher risk-adjusted return.



Hedge Funds and Diversification

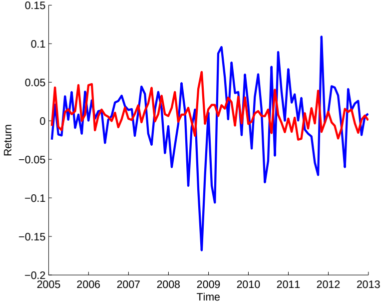


Figure 3: S&P 500 (in blue) and Eurekahedge North America Macro Hedge Fund Index (in red) returns in 31.01.2005-31.12.2012

Outline

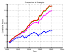
1. Motivation ✓
2. Hedge Funds' Potential For Diversification
3. Asset Allocation with Hedge Funds
4. Portfolio Trading Strategy
5. Results

Mean-Variance Optimization

Markowitz diversification rule:

$$\begin{aligned} \underset{w \in \mathbb{R}^d}{\text{minimize}} \quad & w^\top \Sigma w \\ \text{subject to} \quad & w^\top \bar{r} = r_T, \\ & \sum_{i=1}^{d} w_i = 1, \\ & w_i \geq 0 \end{aligned}$$

where w_i, $i = 1, \ldots, d$ are weights, $\Sigma \in \mathbb{R}^{d \times d}$ is the covariance matrix for d portfolio asset returns $\bar{r}_i$, r_T is the "target" return for the portfolio.



Diversification Concept

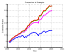
Portfolio diversification is a tool to reduce specific risk and remain only with market risk. Mean-variance theory implies that

- diversification is efficient when portfolio asset returns are uncorrelated or negatively correlated;
- increasing diversification increases certainty when returns are uncorrelated and variances are identical;
- it is necessary to avoid investing in securities with high covariances among themselves.

Correlation Examples - Traditional Assets

MSCI Indices	US	UK	SW	GER	JAP
US (US)	1.00				
UK (UK)	0.69	1.00			
SW (Switzerland)	0.51	0.58	1.00		
GER (Germany)	0.60	0.59	0.50	1.00	
JAP (Japan)	0.47	0.45	0.38	0.24	1.00

Table 1: Correlation statistics for traditional asset class indices; based on monthly data Jan. 1994 - Aug. 2001; table from Lhabitant (2002, p.158)

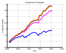


Correlation Examples - Traditional Assets and Hedge Fund Indices

Hedge Funds	US	UK	SW	GER	JAP
Conv. arb.	0.10	0.08	0.07	0.10	−0.02
Dedic. sh. bias	−0.77	−0.53	−0.33	−0.46	−0.48
Fix. inc. arb.	0.10	0.13	0.01	0.08	−0.10
Glob. macro	0.30	0.19	0.10	0.27	−0.11
Man. fut.	−0.10	0.02	−0.09	−0.03	0.03

Table 2: Correlation statistics for traditional asset class and hedge funds' indices; based on monthly data Jan. 1994 - Aug. 2001; table from Lhabitant (2002, p.164)

Details for Hedge Funds Strategies



Asset Allocation and Hedge Funds

Asset allocation: combine several assets optimally to maximize risk-adjusted performance consistently with the investor's preferences.

Hedge funds superior in asset allocation:

- may in reality be a conservative investment
- offer superior risk-adjusted returns
- better diversification
- dynamic.

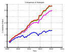
Alternative Allocation Approach

Motivation:

- hedge funds' returns often have negative skewness and/or positive excess kurtosis - mean-variance approach tends to underestimate portfolio risk
- financial returns' covariance structure is often time-changing

Possible remedies:

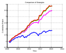
- use VaR as the objective optimization function
- adjust VaR for skewness and kurtosis, e.g, via Cornish-Fisher (CF) expansion
- use a multivariate GARCH framework to model variance-covariance structure



Modelling Variance-Covariance Structure

Problem: model the covariance matrix Σ_t of financial returns r_t, as in $r_t | \mathcal{F}_{t-1} \sim \mathrm{N}(0, \Sigma_t)$

1. Orthogonal GARCH framework: modelling $\Sigma_t = B_t \Delta_t B_t^\top + \Omega_t$ ▸ Details
2. Dynamic Conditional Correlation (DCC) framework: modelling $\Sigma_t = D_t R_t D_t$ ▸ Details



Cornish-Fisher VaR Optimization

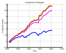
The modified optimization problem becomes

$$\begin{aligned} \underset{w\in\mathbb{R}^d}{\text{minimize}} \quad & W(\bar{r}_p(w) - \omega(w)\cdot\sigma_p(w)) \\ \text{subject to} \quad & w^\top \bar{r} = r_T, \\ & \sum_{i=1}^{d} w_i = 1, \;\; w_i \geq 0 \end{aligned}$$

where W is portfolio value, $\bar{r}_p(w) \stackrel{\text{def}}{=} w^\top \bar{r}$, $\sigma_p^2(w) \stackrel{\text{def}}{=} w^\top \Sigma w$,

$$\omega(w) \stackrel{\text{def}}{=} z_\alpha + (z_\alpha^2 - 1)\frac{S_p(w)}{6} + (z_\alpha^3 - 3z_\alpha)\frac{K_p(w)}{24} - (2z_\alpha^3 - 5z_\alpha)\frac{S_p(w)^2}{36},$$

where $S_p(w)$ and $K_p(w)$ are, respectively, portfolio skewness and kurtosis, z_α is standard normal α-quantile



Efficient Frontier

Efficient frontier is a set of optimal portfolios that offers the highest expected return for a defined level of risk or the lowest risk for a given level of expected return.

Portfolios that lie below the efficient frontier are *sub-optimal*: they do not provide enough return for the level of risk.

Portfolios made solely of stocks are sub-optimal to those which include hedge funds.

Efficient Frontier Example

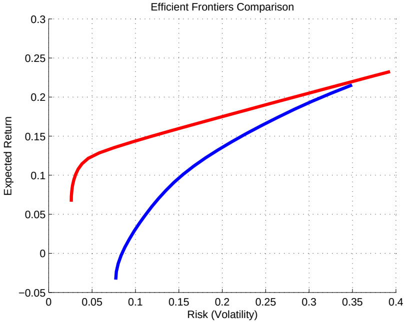
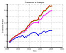


Figure 4: Efficient frontiers built by using all S&P 500 components only (in blue) and by mixing them with Eurekahedge hedge funds indices (in red)



Risk-Adjusted Return

Hedge funds offer superior risk-adjusted returns

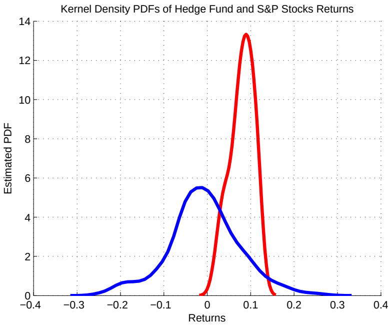


Figure 5: Estimated kernel densities for S&P 500 components (in blue) and for Eurekahedge hedge funds indices (in red) returns

Hedge Funds-Based Tail Events Strategy

How to choose hedge funds which are negatively related to S&P 500 in the lower tail?

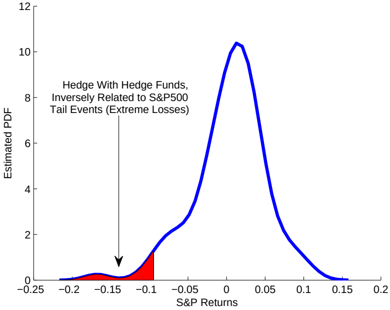
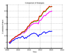


Figure 6: A strategy is needed to estimate negative dependence between S&P 500 and hedge funds in the lower tail

Hedge Funds-Based Tail Events Strategy

The method of Adaptive Non-Positive Lasso Quantile Regression (ALQR)

- deals with the high dimensionality problem (how to choose from today's over 5,200 hedge funds?)
- consistently estimates true nonzero coefficients measuring negative relationship between hedge funds (X, hedge funds' log-returns) and the benchmark asset (Y, S& P 500 log-returns) in the tails



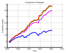
ALQR Estimator

Given $Y = X\beta + \varepsilon$; $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^p$, $\varepsilon \in \mathbb{R}^n$; $\varepsilon_1, ..., \varepsilon_n$ i.i.d., independent of $\{X_i; i = 1, ..., n\}$, $\mathsf{E}(\varepsilon_i) = 0$; $\rho_\tau(u) = u\{\tau - \mathrm{I}(u < 0)\}$, $\tau \in (0, 1)$, $\hat{\omega}$ are weights, λ is regularization parameter, the ALQR estimator is ▸ Details

$$\hat{\beta}_{\tau,\hat{\lambda}} = \arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta) + \lambda \|\hat{\omega}^\top \beta\|_1$$

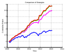
$$\text{such that} \quad \beta \leq 0$$

$\hat{\beta}_{\tau,\hat{\lambda};j}$ measures the "strength" of the linear relationship between Y and $\{X_j; j = 1, \ldots, p\}$ on quantile level



ALQR Monte-Carlo Analysis

- $\lambda_n = 0.25\sqrt{\|\hat{\beta}^{\text{init}}\|_0 \log(n \vee p)}(\log n)^{0.1/2}$, $\hat{\omega}_j = 1/|\hat{\beta}_j^{\text{init}}| \wedge \sqrt{n}$; $\hat{\beta}_j^{\text{init}}$ are from (3) (for oracle ▸ Details results to hold);
- linear model (1) with $X_i \sim \mathrm{N}(0, \Omega)$, $n = 50$, $p = 300$, $\beta = (-5, -5, -5, -5, -5, -5, 0, ..., 0)$, $q = 6$, $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$;
- $\Omega_{i,j} = 0.5^{|i-j|}$, $\sigma = 0.1, 0.5, 1$ (three levels of noise);
- for $\hat{\beta}^{\text{init}}$ estimator $\hat{\beta}_{\tau,\hat{\lambda}}$ from the model (2) is used, where $\hat{\lambda}$ is chosen according to the GACV criterion (Li, Zhu, 2008);
- number of replications is 100



Accuracy Criteria

1. Standardized L_2-norm

$$\text{Dev} \stackrel{\text{def}}{=} \frac{\|\beta - \hat{\beta}\|_2}{\|\beta\|_2}$$

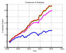
2. Sign consistency

$$\text{Acc} \stackrel{\text{def}}{=} \sum_{j=1}^{p} |\operatorname{sign}(\beta_j) - \operatorname{sign}(\hat{\beta}_j)|$$

3. Least angle

$$\text{Angle} \stackrel{\text{def}}{=} \frac{< \beta, \hat{\beta} >}{\|\beta\|_2 \|\hat{\beta}\|_2}$$

4. Estimate of q: Est $\stackrel{\text{def}}{=} \hat{q}$



Monte-Carlo Analysis Results

Table 3: Criteria Results under Different Models and Quantiles

Accuracy Crit. and Model		Noise Levels and Quantile Indices								
		σ=0.1			σ=0.5			σ=1		
		τ=0.1	**τ=0.5**	**τ=0.9**	**τ=0.1**	**τ=0.5**	**τ=0.9**	**τ=0.1**	**τ=0.5**	**τ=0.9**
Dev	**ALQR**	*0.1(0.23)*	*0.03(0.11)*	*0.06(0.2)*	*0.1(0.19)*	*0.04(0.1)*	*0.1(0.2)*	*0.17(0.24)*	*0.05(0.06)*	*0.16(0.2)*
	LQR	0.12(0.23)	0.06(0.13)	0.09(0.2)	0.18(0.19)	0.14(0.13)	0.19(0.2)	0.25(0.21)	0.21(0.13)	0.26(0.21)
Acc	**ALQR**	*0.44(1.2)*	*0.11(0.53)*	*0.32(1.25)*	*0.3(0.93)*	*0.06(0.31)*	*0.42(1.22)*	*0.71(1.72)*	*0.14(0.49)*	*0.5(1.21)*
	LQR	5.33(2)	0.95(1.1)	5.66(2.24)	5.59(1.96)	1.07(1.39)	5.51(2.08)	6.15(2.3)	1.11(1.36)	6.06(1.95)
Angle	**ALQR**	*0.07(0.18)*	*0.01(0.07)*	*0.05(0.2)*	*0.05(0.14)*	*0.01(0.08)*	*0.06(0.18)*	*0.1(0.25)*	*0.01(0.04)*	*0.08(0.19)*
	LQR	0.11(0.3)	0.03(0.14)	0.07(0.25)	0.09(0.19)	0.05(0.15)	0.11(0.22)	0.16(0.3)	0.07(0.11)	0.18(0.36)
Est	**ALQR**	*5.81(0.58)*	*5.97(0.30)*	*5.89(0.42)*	*5.87(0.42)*	*5.97(0.22)*	*5.92(0.37)*	*5.81(0.61)*	*6.03(0.33)*	*5.84(0.58)*
	LQR	8.51(1.60)	6.35(0.72)	8.54(1.72)	8.69(1.63)	6.42(0.84)	8.60(1.91)	8.85(1.68)	6.43(0.74)	8.83(1.52)

Model notation: ALQR - Adaptive Lasso-penalized quantile regression; LQR - simple Lasso-penalized quantile regression

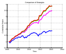
Standard deviations are given in brackets

Data for Analysis

- the covariate matrix $X \in \mathbb{R}^{96\times 170}$, data on 96 monthly returns of 170 Eurekahedge hedge funds indices in the period of 31.01.2005 - 31.12.2012;
- the response vector $Y \in \mathbb{R}^{96}$ consists of data on 96 monthly returns of S&P 500 in the period of 31.01.2005 - 31.12.2012;
- the idea is to hedge the benchmark asset (e.g.,S&P 500) with a security (e.g., hedge fund) moving in opposite direction at different quantiles;

Tail Events Trading Strategy

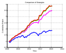
- moving window, width $l = 50$
- $\tau_{1,2,3,4,5} = (0.05, 0.15, 0.25, 0.35, 0.50)$
- F_n is the **edf** of S&P 500 log-returns
- $\hat{q}_\tau \stackrel{\text{def}}{=} F_n^{-1}(\tau)$ is the S&P 500 log-returns empirical quantile function
- $\hat{\beta}_{\tau,\hat{\lambda}}$ are the estimated non-zero ALQR coefficients.



Tail Events Trading Strategy

At each time moment t; $t = l, \ldots, n$

1. determine the S& P 500 return r_t
2. choose $\tau_{j,t}, j = 1, \ldots, 5$ corresponding to the right-hand side $\hat{q}_{\tau_t}$ in one of the conditions which holds simultaneously:
   $r_t \leq \hat{q}_{\tau_{1,t}}$, $\hat{q}_{\tau_{1,t}} < r_t \leq \hat{q}_{\tau_{2,t}}$, $\hat{q}_{\tau_{2,t}} < r_t \leq \hat{q}_{\tau_{3,t}}$,
   $\hat{q}_{\tau_{3,t}} < r_t \leq \hat{q}_{\tau_{4,t}}$, $\hat{q}_{\tau_{4,t}} < r_t \leq \hat{q}_{\tau_{5,t}}$
3. solve the ALQR problem for $\hat{\beta}_{\tau_{j,t},\lambda_n}$ on the moving window using the observations $X \in \mathbb{R}^{t-l+1,\ldots,t \times p}$, $Y \in \mathbb{R}^{t-l+1,\ldots,t}$, buy the hedge funds with $\hat{\beta}_{\tau_{j,t},\lambda_n} \neq 0$ taken with optimal weights
4. if none of the inequalities from Step 2 holds, invest into the benchmark asset (S&P 500) at r_t.



Tail Events Trading Strategy - Example

1. Suppose $t = 58$, accumulated wealth $W_{58} = 1.429$, $r_{58} = -1.85\%$.

2. It occurs that $\hat{q}_{0.15,58} = -4.18\%$, $\hat{q}_{0.25,58} = -1.85\%$ and so $\hat{q}_{0.15,58} < r_{58} \leq \hat{q}_{0.25,58}$.

3. Solving the ALQR problem using $X \in \mathbb{R}^{9,\dots,58 \times 170}$, $Y \in \mathbb{R}^{9,\dots,58}$ yields $\hat{\beta}_{0.25,58} = (-0.77, -1.12, -0.41)$, which correspond to three hedge funds, namely, *Latin American Arbitrage*, *North America Macro*, *Emerging Markets CTA/Managed Futures*.

4. CF-VaR optimization problem yields $w = (0.22, 0.16, 0.62)$ as solution, this portfolio yields a return of 0.62% ($W_{59} = 1.438$), while the benchmark asset return has been -1.85%.

Alternative Strategies

1. Strategy 2 - base case S&P 500 "buy-and-hold"
2. Strategy 3 - ALQR-based "naive" diversification (always equal weights, no optimization)
3. Strategy 4 - based on Orthogonal GARCH model and simple variance-covariance VaR optimization

Strategies' Comparison

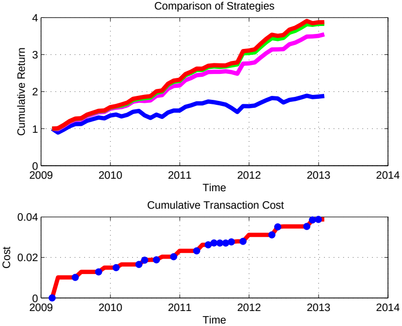
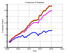
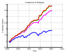


Figure 7: Upper graph: Strategies' cumulative returns' comparison: Strategy 1 (in red), Strategy 2 (in blue), Strategy 3 (in green), Strategy 4 (in magenta); lower graph: cumulative transaction cost for Strategies 1,3,4 (each time 1% of trade value); blue dots denote portfolio rebalancing



-$\widehat{\beta}$ in each window, $\tau = 0.05$

Figure 8: Different -$\widehat{\beta}$ in application; $\tau = 0.05$



-$\widehat{\beta}$ in each window, $\tau = 0.35$

Figure 9: Different $-\widehat{\beta}$ in application; $\tau = 0.35$

Histograms of $\hat{q}$

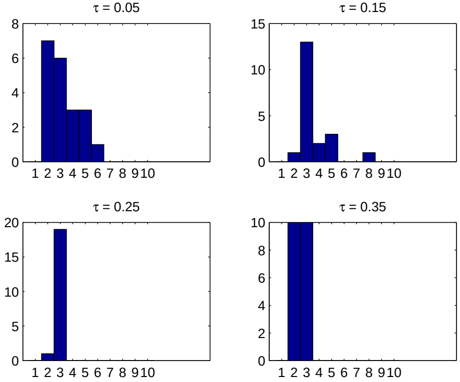


Figure 10: Frequency of the number of selected variables for 4 different τ

The estimated value of λ

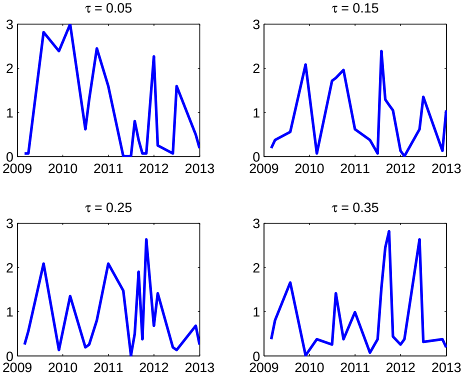


Figure 11: The estimated λ for 4 different τ

The Influential Variables

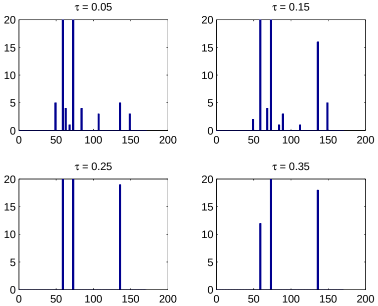
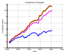
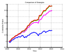


Figure 12: The frequency of the hedge funds. Most frequently selected hedge funds: *"Latin American Arbitrage Hedge Fund Index"* and *"North America Macro Hedge Fund Index"* with frequencies 37 at $\tau = 0.05, 0.15$



Conclusions

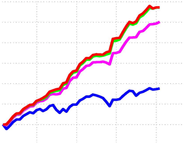
- hedge funds can successfully replace conventional assets in portfolios;
- diversification with hedge funds is superior to traditional financial instruments;
- portfolio trading strategy using the Adaptive Lasso Quantile Regression Method performs significantly better than other strategies.



Lasso Quantile Trading Strategy

Sergey Nasekin

Wolfgang Karl Härdle



Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de

Lasso Shrinkage

Linear model: $Y = X\beta + \varepsilon$; $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^p$, $\varepsilon \in \mathbb{R}^n$; $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d., independent of $\{X_i; i = 1, \ldots, n\}$, $\mathrm{E}(\varepsilon_i) = 0$

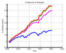
The optimization problem for the lasso estimator:

$$\begin{aligned} \hat{\beta}^{\text{lasso}} = & \arg \min_{\beta \in \mathbb{R}^p} f(\beta) \\ & \text{subject to} \quad g(\beta) \geq 0 \end{aligned} \tag{1}$$

where

$$f(\beta) = \frac{1}{2}(y - X\beta)^\top (y - X\beta)$$
$$g(\beta) = t - \|\beta\|_1$$

where t is the size constraint on $\|\beta\|_1$ Back to "ALQR Estimator"



Lasso Duality

If (1) is convex programming problem, then the Lagrangian is

$$L(\beta,\lambda) = f(\beta) - \lambda g(\beta).$$

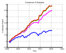
and the primal-dual relationship is

$$\underbrace{\underset{\beta}{\text{minimize}} \ \sup_{\lambda \geq 0} \ L(\beta,\lambda)}_{\text{primal}} \geq \underbrace{\underset{\lambda \geq 0}{\text{maximize}} \ \inf_{\beta} \ L(\beta,\lambda)}_{\text{dual}}$$

Then the dual function $L^*(\lambda) = \inf_{\beta} \ L(\beta,\lambda)$ is

$$L^*(\lambda) = \frac{1}{2} y^\top y - \frac{1}{2} \hat{\beta}^\top X^\top X \hat{\beta} - t \frac{(y - X\hat{\beta})^\top X \hat{\beta}}{\|\hat{\beta}\|_1}$$

with $(y - X\hat{\beta})^\top X \hat{\beta} / \|\hat{\beta}\|_1 = \lambda$



Paths of Lasso Coefficients

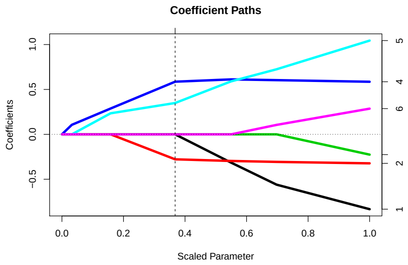
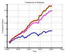


Figure 13: Lasso shrinkage of coefficients in the hedge funds dataset example (6 covariates were chosen for illustration); each curve represents a coeffcient as a function of the scaled parameter $\hat{s} = t/\|\beta\|_1$; the dashed line represents the model selected by the BIC information criterion ($\hat{s} = 3.7$)



Quantile Regression

The loss $\rho_\tau(u) = u\{\tau - \mathbf{I}(u < 0)\}$ gives the (conditional) quantiles $F^{-1}_{y|x}(\tau) \stackrel{\text{def}}{=} q_\tau(x)$.

Minimize

$$\hat{\beta}_\tau = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta).$$

Re-write:

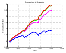
$$\underset{(\xi,\zeta)\in\mathbb{R}^{2n}_{+}}{\text{minimize}} \quad \left\{ \tau 1_n^\top \xi + (1-\tau) 1_n^\top \zeta | X\beta + \xi - \zeta = Y \right\}$$

with ξ, ζ are vectors of "slack" variables

Non-Positive (NP) Lasso-Penalized QR

The lasso-penalized QR problem with an additional non-positivity constraint takes the following form:

$$
\begin{aligned}
\underset{(\xi,\zeta,\eta,\tilde{\beta})\in\mathbb{R}_{+}^{2n+p}\times\mathbb{R}^{p}}{\text{minimize}} \quad & \tau 1_n^\top \xi + (1-\tau) 1_n^\top \zeta + \lambda 1_n^\top \eta \\
\text{subject to} \quad & \xi - \zeta + \eta = Y + X\tilde{\beta}, \\
& \xi \geq 0, \\
& \zeta \geq 0, \\
& \eta \geq \tilde{\beta}, \\
& \eta \geq -\tilde{\beta}, \\
& \tilde{\beta} \geq 0, \qquad \tilde{\beta} \stackrel{\text{def}}{=} -\beta
\end{aligned}
\tag{2}
$$



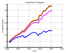
Solution

Transform into matrix (I_n is $n \times n$ identity matrix; $E_{n\times k} = \begin{pmatrix} I_k \\ 0 \end{pmatrix}$):

$$\begin{aligned} &\text{minimize} && c^\top x \\ &\text{subject to} && Ax = b, \;\; Bx \le 0 \end{aligned}$$

where $A = \begin{pmatrix} I_n & -I_n & I_n & X \end{pmatrix}$, $b = Y$, $x = \begin{pmatrix} \xi & \zeta & \eta & \beta \end{pmatrix}^\top$,

$$c = \begin{pmatrix} \tau 1_n \\ (1-\tau)1_n \\ \lambda 1_n \\ 0 1_n \end{pmatrix}, \quad B = \begin{pmatrix} -I_n & 0 & 0 & 0 \\ 0 & -I_n & 0 & 0 \\ 0 & 0 & -E_{n\times k} & E_{n\times k} \\ 0 & 0 & -E_{n\times k} & -E_{n\times k} \\ 0 & 0 & 0 & E_{n\times k} \end{pmatrix}$$



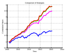
Solution - Continued

The previous problem may be reformulated as follows:

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Cx = d, \\ & x + s = u, \;\; x \geq 0, s \geq 0 \end{aligned}$$

and the dual problem is:

$$\begin{aligned} \text{maximize} \quad & d^\top y - u^\top w \\ \text{subject to} \quad & C^\top y - w + z = c, \;\; z \geq 0, w \geq 0 \end{aligned}$$



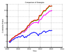
Solution - Continued

The KKT conditions for this linear program are

$$F(x, y, z, s, w) = \left\{ \begin{array}{c} Cx - d \\ x + s - u \\ C^\top y - w + z - c \\ x \circ z \\ s \circ w \end{array} \right\} = 0,$$

with $y \geq 0$, $z \geq 0$ dual slacks, $s \geq 0$ primal slacks, $w \geq 0$ dual variables.

This can be solved by a primal-dual path following algorithm based on the *Newton method*



Adaptive Lasso Procedure

Lasso can be inconsistent (Zou, 2006) in some scenarios.

Lasso soft-threshold function gives biased results

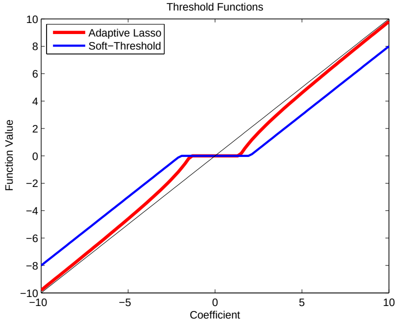
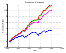


Figure 14: Threshold functions for simple and adaptive Lasso



Adaptive Lasso Procedure

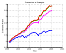
The adaptive Lasso (Zou, 2006) yields a sparser solution and is less biased.

L_1 - penalty replaced by a re-weighted version; $\hat{\omega} = 1/|\hat{\beta}^{\text{init}}|^{\gamma}$, $\gamma = 1$, $\hat{\beta}^{\text{init}}$ is from (1)

The adaptive lasso estimates are given by:

$$\hat{\beta}_{\hat{\lambda}}^{\text{adapt}} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (Y_i - X_i^{\top}\beta)^2 + \lambda \|\hat{\omega}^{\top}\beta\|_1$$

(Bühlmann, van de Geer, 2011): $\hat{\beta}_j^{\text{init}} = 0$, then $\hat{\beta}_j^{\text{adapt}} = 0$



Simple and Adaptive Lasso Penalized QR

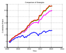
Simple lasso-penalized QR optimization problem is:

$$\hat{\beta}_{\tau,\hat{\lambda}} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta) + \lambda \|\beta\|_1 \tag{3}$$

Adaptive lasso-penalized QR model uses the re-weighted penalty:

$$\hat{\beta}^{\text{adapt}}_{\tau,\hat{\lambda}} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i^\top \beta) + \lambda \|\hat{\omega}^\top \beta\|_1 \tag{4}$$

Adaptive lasso-penalized QR procedure can ensure oracle properties for the estimator ▸ Details



Algorithm for Adaptive Lasso Penalized QR

The optimization for the adaptive lasso can be re-formulated as a lasso problem:

- the covariates are rescaled: $\tilde{X} = (X_1 \circ \hat{\beta}_1^{\text{init}}, \dots, X_p \circ \hat{\beta}_p^{\text{init}})$;
- the lasso problem (3) is solved:

$$\hat{\tilde{\beta}}_{\tau,\hat{\lambda}} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - \tilde{X}_i^\top \beta) + \lambda \|\beta\|_1$$

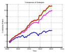
- the coefficients are re-weighted as $\hat{\beta}^{\text{adapt}} = \hat{\tilde{\beta}}_{\tau,\hat{\lambda}} \circ \hat{\beta}^{\text{init}}$

Oracle Properties of an Estimator

An estimator has oracle properties if (Zheng et al., 2013):

- it selects the correct model with probability converging to 1;
- the model estimates are consistent with an appropriate convergence rate (He, Shao, 2000);
- estimates are asymptotically normal with the same asymptotic variance as that knowing the true model

▸ Back to "Simple and Adaptive Lasso Penalized QR"



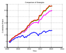
Oracle Properties for Adaptive Lasso QR

In the linear model, let $Y = X\beta + \varepsilon = X^1\beta^1 + X^2\beta^2 + \varepsilon$, where $X = (X^1, X^2)$, $X^1 \in \mathbb{R}^{n\times q}$, $X^2 \in \mathbb{R}^{n\times(p-q)}$; β^1_q are true nonzero coefficients, $\beta^2_{p-q} = 0$ are noise coefficients; $q = \|\beta\|_0$;
$\lim_{n\to\infty} X^\top X/n = \Sigma$, where

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where Σ_{11} is the $q \times q$ covariance matrix knowing the true subset model $Y = X^1\beta^1 + \varepsilon$;

$\lambda q/\sqrt{n} \to 0$ and $\lambda/\{\sqrt{q}\log(n \vee p)\} \to \infty$ Back



Oracle Properties for Adaptive Lasso QR

Then under certain regularity conditions ▸ Details the adaptive L_1 QR estimator has the oracle properties (Zheng et al., 2013):

1. Variable selection consistency:

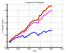
$$\mathsf{P}(\beta^2 = 0) \geq 1 - 6\exp\left\{-\frac{\log(n \vee p)}{4}\right\}.$$

2. Estimation consistency: $\|\beta - \hat{\beta}\| = \mathcal{O}_p(\sqrt{q/n})$

3. Asymptotic normality: $u_q^2 \stackrel{\text{def}}{=} \alpha^\top \Sigma_{11} \alpha$, $\forall \alpha \in \mathbb{R}^q$, $\|\alpha\| < \infty$,

$$n^{1/2} u_q^{-1} \alpha^\top (\beta^1 - \hat{\beta}^1) \xrightarrow{\mathcal{L}} \mathsf{N}\left\{0, \frac{(1-\tau)\tau}{f^2(\gamma^*)}\right\}$$

where γ^* is the τth quantile and f is the pdf of ε ▸ Back



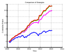
Selected Hedge Funds' Strategies

1. *Convertible arbitrage* funds - attempt to profit from mispricing of converible securities, and/or expected trends in factors influencing their prices.

2. *Dedicated short bias* funds - attempt to capture profits when the market declines, by holding investments that are overall biased to the short side.

3. *Fixed income arbitrage* funds - attempt to profit from observed relative pricing inefficiencies between related fixed income securities and/or expected changes in inter-market spreads.

▸ Return



Selected Hedge Funds' Strategies

4. *Global macro* funds - attempt to profit by making very large directional bets that reflect their forecasts of market directions, as influenced by major economic trends and/or particular events.

5. *Managed futures* funds - use their own proprietary trading methods and money management techniques to establish positions on behalf of their clients on a discretionary basis.

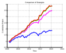
The Orthogonal GARCH Model

- Y_t is a time-dependent matrix of asset returns, $\Gamma_t = B_t \in \mathbb{R}^{p \times p}$ is the matrix of standardized eigenvectors of $\frac{1}{n} Y_t^\top Y_t$ ordered according to decreasing magnitude of eigenvalues

- $F_t = P_t \stackrel{\text{def}}{=} Y_t \Gamma_t$ is the matrix of principal components of Y_t

- retaining only the first k most important factors f and introducing noise terms u_i gives $y_j = b_{j1} f_1 + b_{j2} f_2 + \ldots + b_{jk} f_k + u_i$ or $Y_t = F_t B_t^\top + U_t$

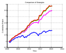
- then $\Sigma_t = \text{Var}(Y_t) = \text{Var}(F_t B_t^\top) + \text{Var}(U_t) = B_t \Delta_t B_t^\top + \Omega_t$, where $\Delta_t = \text{Var}(F_t)$ is a diagonal matrix of principal component variances at t and B_t is assumed to be known at time t; Ω_t assumed to be constant and diagonal Back



The DCC Model

$$
\begin{aligned}
r_t|\mathcal{F}_{t-1} &\sim \mathrm{N}(0, D_t R_t D_t),\\
D_t^2 &= \mathrm{diag}(\omega_i) + \mathrm{diag}(\alpha_i) \odot r_{t-1} r_{t-1}^\top + \mathrm{diag}(\beta_i) \odot D_{t-1}^2,\\
\varepsilon_t &= D_t^{-1} r_t,\\
Q_t &= S \odot (\imath\imath^\top - A - B) + A \odot \varepsilon_{t-1}\varepsilon_{t-1}^\top + B \odot Q_{t-1},\\
R_t &= \{\mathrm{diag}(Q_t)\}^{-1} Q_t \{\mathrm{diag}(Q_t)\}^{-1}
\end{aligned}
$$

where r_t is an $d \times 1$ vector of returns t, D_t is an $d \times d$ diagonal matrix of time-varying standard deviations σ_{it}, $i = 1, \ldots, d$, modeled by univariate GARCH, ε_t is an $d \times 1$ vector of standardized returns with $\varepsilon_{it} \overset{\mathrm{def}}{=} r_{it}\sigma_{it}^{-1}$, $\imath$ is a vector of ones, A and B are positive semidefinite coefficient matrices, $S = (1/T)\sum_{t=1}^{T} \varepsilon_t \varepsilon_t^\top$ Back



Regularity Conditions for Adaptive Lasso QR

A1 Sampling and smoothness: $\forall x$ in the support of X_i, $\forall y \in \mathbb{R}$, $f_{Y_i|X_i}(y|x)$, $f \in \mathcal{C}^k(\mathbb{R})$, $|f_{Y_i|X_i}(y|x)| < \overline{f}$, $|f'_{Y_i|X_i}(y|x)| < \overline{f'}$; $\exists \underline{f}$, such that $f_{Y_i|X_i}(x^\top \beta_\tau | x) > \underline{f} > 0$

A2 Restricted identifiability and nonlinearity: let $\delta \in \mathbb{R}^p$, $T \subset \{0, 1, ..., p\}$, δ_T such that $\delta_{Tj} = \delta_j$ if $j \in T$, $\delta_{Tj} = 0$ if $j \notin T$; $T = \{0, 1, ..., s\}$, $\overline{T}(\delta, m) \subset \{0, 1, ..., p\} \backslash T$, then $\exists m \geq 0$, $c \geq 0$ such that

$$\inf_{\delta \in A, \delta \neq 0} \frac{\delta^{\top} \mathrm{E}(X_i X_i^\top)\delta}{\|\delta_{T \cup \overline{T}(\delta, m)}\|^2} > 0, \qquad \frac{3\underline{f}^{3/2}}{8\overline{f}'} \inf_{\delta \in A, \delta \neq 0} \frac{\mathrm{E}[|X_i^\top \delta|^2]^{3/2}}{\mathrm{E}[|X_i^\top \delta|^3]} > 0,$$

where $A \stackrel{\text{def}}{=} \{\delta \in \mathbb{R}^p : \|\delta_{T^c}\|_1 \leq c\|\delta_T\|_1, \|\delta_{T^c}\|_0 \leq n\}$

▸ Back

Regularity Conditions - Continued

A3 Growth rate of covariates:

$$\frac{q^3\{\log(n \vee p)\}^{2+\eta}}{n} \rightarrow 0, \eta > 0$$

A4 Moments of covariates: Cramér condition

$$\mathsf{E}[|x_{ij}|^k] \leq 0.5\,C_m M^{k-2} k!$$

for some constants C_m, M, $\forall k \geq 2$, $j = 1, ..., p$

A5 Well-separated regression coefficients: $\exists b_0 > 0$, such that $\forall j \leq q$, $|\hat{\beta}_j| > b_0$

References

Peter Bühlmann, Sara van de Geer

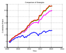
Statistics for High-Dimensional Data: Methods, Theory and Applications

Springer, 2011

Lhabitant, François-Serge

Hedge Funds. Myths and Limits

Wiley, 2002



References

Roger Koenker, Gilbert Bassett, Jr

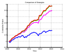
Regression Quantiles

Econometrica, Vol. 46, No. 1 (January, 1978)

Youjuan Li, Ji Zhu

L_1-norm Quantile Regression

Journal of Computational and Graphical Statistics, Vol. 17, No. 1, pp. 1-23



References

Markowitz, Harry

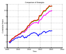
Portfolio Selection

Journal of Finance, Vol. 7, No. 1 (Mar., 1952), pp. 77-91

Robert Tibshirani

Regression Shrinkage and Selection via the Lasso

J. R. Statist. Soc (1996), No.1 pp. 267-288



References

Qi Zheng, Colin Gallagher, K.B. Kulasekera

Adaptive Penalized Quantile Regression for High-Dimensional Data

Journal of Statistical Planning and Inference, 143 (2013) 1029-1038

Hui Zou

The Adaptive Lasso and Its Oracle Properties

Journal of the American Statistical Association, Dec., 2006, Vol. 101, No. 476