CASE STUDY: Style Classification with Quantile Regression (kb_err, pm_pen, pm_pen_g)

Background

This case study applies percentile regression to the return-based style classification of a mutual fund. The procedure regresses fund return by several indices as explanatory variables. The estimated coefficients represent the fund's style with respect to each of the indices. This problem was considered by Carhart (1997) and Sharpe (1992). They estimated conditional expectation of a fund return distribution (under the condition that a realization of explanatory variables is observed).

Basset and Chen (2001) extended this approach and conducted style analyses of quantiles of the return distribution. This extension is based on the quantile regression approach suggested by Koenker and Basset (1978). The quantile regression model is more flexible compared to the standard least squares regression because it can identify dependence of various parts of the distribution from explanatory variables. A portfolio style depends on how a factor influences the entire return distribution, and this influence cannot be described by a single number. The single number given by the least squares regression may obscure the tail behavior (which could be of a prime interest to a manager). With the quantile regression we can estimate, for instance, the impact of explanatory variables on the 99-th percentile of the loss distribution. Portfolios having exposures to derivatives may have very different regression coefficients of the mean value and tail quantiles. For instance, let us consider a strategy in investing into naked deep out-of-the-money options. This strategy in most cases behaves like a bond paying some interest, however, in rare cases the strategy loses some amount of money (may be quite significant). Therefore, the mean value and 99-th percentile may have very different regression coefficients for the explanatory variables.

We regresses quantile of the return distribution of the Fidelity Magellan Fund on the Russell Value Index (RUJ), RUSSELL 1000 VALUE INDEX (RLV), Russell 2000 Growth Index (RUO) and Russell 1000 Growth Index (RLG). We want to calculate coefficients for the explanatory variables of the tail of the distribution of residuals (these coefficients may differ from the regression coefficients for the mean and the median of the distribution). The confidence level in quantile regression is 0.1.

Quantile regression methodology

Let $Y_1, Y_2, \ldots, Y_J$ be J observations of a random value Y and $\boldsymbol{\zeta}_j = (\zeta_{j1}, \zeta_{j2}, \ldots, \zeta_{jI})$ be J corresponding observations of a covariate (explanatory) vector $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_I)$. The estimate of α-percentile of Y (i.e., VaR of Y with confidence level α) under the condition that a new realization $(\hat{\zeta}_1, \hat{\zeta}_2, \ldots, \hat{\zeta}_I)$ is observed is equal to $\sum_{i=1}^{I} \hat{\zeta}_i x_i^a$ where the vector $\boldsymbol{x}^a = (x_1^a, x_2^a, \ldots x_I^a)$ is obtained by solving the following quantile regression minimization problem (see, Koenker and Basset (1978)):

$$\boldsymbol{x}^\alpha = \underset{\boldsymbol{x}}{\arg\min} \left\{ \sum_{j: Y_j \geq \sum_{i=1}^{I} \zeta_{ji} x_i} \alpha \left(Y_j - \sum_{i=1}^{I} \zeta_{ji} x_i \right) + \sum_{j: Y_j < \sum_{i=1}^{I} \zeta_{ji} x_i} (1-\alpha) \left(\sum_{i=1}^{I} \zeta_{ji} x_i - Y_j \right) \right\}. \tag{CS.1}$$

The minimization problem (CS.1) can be reformulated in terms of the Partial Moment Penalty for Loss (pm_pen) and Partial Moment Penalty for Gain (pm_pen_g).

Let $L(\boldsymbol{x}, \boldsymbol{\theta}_j) = \theta_{j0} - \sum_{i=1}^{I} \theta_{ji} x_i$ be a loss function associated with the decision vector $\boldsymbol{x}$. pm_pen is defined as follows:

$$pm\_pen(L(\boldsymbol{x}, \boldsymbol{\theta}), w) = \sum_{j=1}^{J} p_j \max\{0, L(\boldsymbol{x}, \boldsymbol{\theta}_j) - w\},$$

and pm_pen_g equals

$$pm\_pen\_g(L(\boldsymbol{x},\boldsymbol{\theta}),w)=\sum_{j=1}^{J}p_j\max\{0,-L(\boldsymbol{x},\boldsymbol{\theta}_j)-w\},$$

where $p_j=\frac{1}{J}$, $j=1,\ldots,J$; w is some target value. Denote $\theta_{j0}=Y_j$; $\theta_{ji}=\zeta_{ji}$; and let w = 0. Then, the minimization problem (CS.1) can be reformulated as follows:

$$\hat{\boldsymbol{x}}_\alpha=\underset{\boldsymbol{x}}{\operatorname{argmin}}\{\alpha\cdot pm\_pen(L(\boldsymbol{x},\boldsymbol{\theta}),0)+(1-\alpha)\cdot pm\_pen\_g(L(\boldsymbol{x},\boldsymbol{\theta}),0)\} \qquad \text{(CS.2)}$$

The function

$$kb\_err_\alpha(L(\boldsymbol{x},\boldsymbol{\theta}))=\alpha\cdot pm\_pen(L(\boldsymbol{x},\boldsymbol{\theta}),0)+(1-\alpha)\cdot pm\_pen\_g(L(\boldsymbol{x},\boldsymbol{\theta}),0)$$

is called the Koenker and Basset error function. In terms of this function the minimization problem (CS.1) can be also reformulated as follows:

$$\hat{\boldsymbol{x}}_\alpha=\underset{\boldsymbol{x}}{\operatorname{argmin}}\{kb\_err_\alpha(L(\boldsymbol{x},\boldsymbol{\theta})\}. \qquad \text{(CS.3)}$$

The Koenker and Basset error function is an element of Quntile-based Risk Quadrangle defined in Rockafellar and Uryasev (2011).

References

- Basset G.W., Chen H-L. (2001): Portfolio Style: Return-based Attribution Using Quantile Regression. Empirical Economics 26, 293-305.
- Carhart M.M. (1997): On Persistence in Mutual Fund Performance. Journal of Finance 52, 57-82.
- Koenker R, Basset G. (1978): Regression Quantiles. Econometrica 46, 33-50.
- Sharpe W.F. (1992): Asset Allocation: Management Style and Performance Measurement. Journal of Portfolio Management (Winter), 7-19.
- Rockafellar, R.T. and S. Uryasev (2013): The Fundamental Risk Quadrangle in Risk Management, Optimization, and Statistical Estimation. Surveys in Operations Research and Management Science, 18 (to appear).

Notations

I= number of style indices used for classification. We consider four indices: Russell 1000 value index (optimization variable, *i* = 1), Russell 1000 growth index (optimization variable, *i* = 2), Russell 2000 value index (optimization variable, *i* = 3), and Russell 2000 growth index (optimization variable, *i* = 4);

J = number of scenarios (time periods) *j*={1,…,*J*} index of scenarios;

θ_{j0} = monthly rate of return of the fund, for which the classification is conducted, under scenario j; scenarios are equally probable (in the current case study θ_{j0} is the monthly historical returns of the Fidelity Magellan Fund);

θ_{ji} = monthly rate of return of *i*-th style index (*i* = 1,2,…,*I*) under scenario j, scenarios are equally probable;

θ_i = random value having *J* equally probable scenarios, $\{\theta_{1i}\ldots,\theta_{Ji}\}$, *i* = 1,2,…,*I*;

$\boldsymbol{\theta}=(\theta_0,\theta_1,\ldots,\theta_I)$ = random scenario vector;

$\boldsymbol{x}=(x_1,x_2,\ldots x_I)$ = vector of regression coefficients (loading factors);

$L(\boldsymbol{x},\boldsymbol{\theta}) = \theta_0 - \sum_{i=1}^{I} \theta_i x_i$ = loss function.

Optimization Problem 1

minimizing Koenker and Basset error function

$$\hat{\boldsymbol{x}}_\alpha = \min_{\boldsymbol{x}} \{kb\_err_\alpha(L(\boldsymbol{x},\boldsymbol{\theta})\} \quad . \tag{CS.4}$$

Optimization Problem 2

minimizing Koenker and Basset error function presented with Partial Moments

$$\min_{\boldsymbol{x}} \{\alpha \cdot pm\_pen(L(\boldsymbol{x},\boldsymbol{\theta}),0) + (1-\alpha) \cdot pm\_pen\_g(L(\boldsymbol{x},\boldsymbol{\theta}),0)\}. \tag{CS.5}$$