CASE STUDY: Style Classification with Quantile Regression (kb\_err, pm\_pen, pm\_pen\_g, cvar\_dev, var\_risk)

## 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.

Stile classification problem was considered by Sharpe (1992) and Carhart (1997) with the standard regression approach based on the mean square error. 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 suggested by Koenker and Bassett (1978). The quantile regression 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 estimate. The single estimate given by a least squared regression may obscure performance in the tail of the distribution (which could be of a prime interest to a manager). Quantile regression can estimate, for instance, the impact of explanatory variables on the 99-th percentile of the loss distribution. Portfolios including derivatives may have quite different regression coefficients of mean value and tail quantiles. For instance, let us consider a strategy investing in 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 may have loses (potentially quite significant). Therefore, the mean value and 99-th percentile may have different regression coefficients of explanatory variables.

We considered 3 equivalent formulations of the regression problem:

- 1. Minimization of Koenker-Bassett error, see, Koenker and Bassett (1978);
- 2. Minimization of Koenker-Bassett error with representation through partial moments;
- 3. Decomposition Theorem using CVaR deviation from Quantile Quadrangle, see Rockafellar and Uryasev (2013) and Rockafellar, et al. (2008).

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). The confidence level in quantile regression is 0.1.

### References

- Bassett 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 and G. Bassett (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).
- Rockafellar, R. T., Uryasev, S., and M. Zabarankin (2008): Risk Tuning with Generalized Linear Regression. Mathematics of Operations Research, 33(3), 712–729.

### 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;

 $\theta_0$  = dependent random variable modeling monthly return of the fund;

 $\theta_1, \ldots, \theta_I$  = independent random variables modeling monthly return of style indexes;

 $\theta_{j0}$  = monthly return of the fund, for which the classification is conducted, under scenario j; scenarios are

equally probable (in the current case study  $\theta_{j0}$  = monthly historical returns of the Fidelity Magellan Fund);

 $\theta_{ji}$  = monthly return of *i*-th style index (*i* = 1,2,...,*I*) under scenario *j*, scenarios are equally probable;  $\boldsymbol{\theta}_{i} = (\theta_{i0}, \theta_{i1}, \dots, \theta_{il}) = j$  -th scenario vector;  $p_j = \frac{1}{I}$  = probability of *j*-th scenario, *j* = 1,...,*J*;  $\theta_i$  = random value having *J* equally probable scenarios,  $\{\theta_{1i}, \dots, \theta_{Ji}\}, i = 1, 2, \dots, I$ ;  $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_I)$  = random scenario vector;  $\tilde{\boldsymbol{\theta}} = (\theta_1, \dots, \theta_L) = \text{vector } \boldsymbol{\theta} \text{ without component } \theta_0;$  $\boldsymbol{x} = (x_0, x_1, \dots, x_I)$  = vector of regression coefficients (loading factors);  $\widetilde{\boldsymbol{x}} = (x_1, \dots, x_l) = \text{vector } \boldsymbol{x} \text{ without intercept;}$  $x_0^a + \sum_{i=1}^{I} \hat{\theta}_i x_i^a$  = estimate of  $\alpha$  -percentile of  $\theta_0$  (i.e., VaR of  $\theta_0$  with confidence level  $\alpha$ ) under the condition that a new realization  $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_r)$  is observed;  $L(\boldsymbol{x},\boldsymbol{\theta}) = \theta_0 - x_0 - \sum_{i=1}^{I} \theta_i x_i = \text{loss function, having scenarios } L(\boldsymbol{x},\boldsymbol{\theta}_1), \dots, L(\boldsymbol{x},\boldsymbol{\theta}_J);$  $L(\boldsymbol{x},\boldsymbol{\theta}_j) = \theta_{j0} - x_0 - \sum_{i=1}^{I} \theta_{ji} x_i = j \text{ -th scenario of loss function } L(\boldsymbol{x},\boldsymbol{\theta}), \quad j = 1,...,J;$ w = threshold in a Partial Moment function;  $pm_pen(L(\boldsymbol{x},\boldsymbol{\theta}),w) = \sum_{j=1}^{n} p_j \max\{0, L(\boldsymbol{x},\boldsymbol{\theta}_j) - w\} = \text{expected excess of loss } L(\boldsymbol{x},\boldsymbol{\theta}) \text{ over } w \text{ (Partial Moment})$ Penalty for Loss);  $pm_pen_g(L(\boldsymbol{x},\boldsymbol{\theta}),w) = \sum_{i=1}^{J} p_j \max\{0, -L(\boldsymbol{x},\boldsymbol{\theta}_j) - w\} = \text{expected excess of } -L(\boldsymbol{x},\boldsymbol{\theta}) \text{ over } w \text{ (Partial)}$ Moment Penalty for Gain);  $\alpha$  = confidence level,  $0 < \alpha < 1$ ; ٦

$$kb\_err_{\alpha}(L(\boldsymbol{x},\boldsymbol{\theta})) = \left[\sum_{j: \theta_{j0} \ge x_{0} + \sum_{i=1}^{I} \theta_{ji}x_{i}} \alpha \left(\theta_{j0} - x_{0} - \sum_{i=1}^{I} \theta_{ji}x_{i}\right) + \sum_{j: \theta_{j0} < x_{0} + \sum_{i=1}^{I} \theta_{ji}x_{i}} (1 - \alpha) \left(x_{0} + \sum_{i=1}^{I} \theta_{ji}x_{i} - \theta_{j0}\right)\right] = \alpha \cdot pm\_pen(L(\boldsymbol{x},\boldsymbol{\theta}), 0) + (1 - \alpha) \cdot pm\_pen\_g(L(\boldsymbol{x},\boldsymbol{\theta}), 0) = \text{Koenker-Bassett error};$$

#### Quantile regression methodology

Expression  $x_0^a + \sum_{i=1}^{I} \hat{\theta}_i x_i^a$  estimates  $\alpha$ -percentile of  $\theta_0$ . Vector of regression coefficients  $\mathbf{x}^a = (x_0^a, x_1^a, \dots, x_I^a)^T$  can be obtained by using one of the following approaches:

- 1. Minimization of the Koenker-Bassett Error (see, Koenker and Bassett (1978));
- 2. Decomposition Theorem using CVaR deviation from Quantile Quadrangle, see Rockafellar and Uryasev (2013) and Rockafellar, et al. (2008).

The first method solves the following quantile regression minimization problem:

$$\boldsymbol{x}^{\alpha} = \arg\min_{\boldsymbol{x}} kb\_err_{\alpha}(L(\boldsymbol{x},\boldsymbol{\theta})).$$
(CS.1)

The Koenker-Bassett error is an element of Quantile-Based Risk Quadrangle defined in Rockafellar and Uryasev (2013). The Risk Quadrangle methodology combines risk functions for a random value Y in groups (Quadrangles) consisting of five elements.

Further we provide list functions of the Quantile Quadrangle for  $\alpha \in (0,1)$ . See, Rockafellar and Uryasev (2013)):

Statistic: 
$$S_{\alpha}(X) = VaR_{\alpha}(X) = VaR$$
 (quantile) Statistic;

<u>Risk</u>:  $R_{\alpha}(X) = CVaR_{\alpha}(X) = \min_{C} \{C - V_{\alpha}(X - C)\} = CVaR$  Risk;

Deviation: 
$$D_{\alpha}(X) = CVaR_{\alpha}^{\Delta}(X) = CVaR_{\alpha}(X) - E[X] = \min_{C} \{\mathcal{E}_{\alpha}(X - C)\} = CVaR$$
 Deviation

<u>Regret</u>:  $V_{\alpha}(X) = \frac{1}{1-\alpha} E[X^+]$  = average absolute loss, scaled; <u>Error</u>:  $\mathcal{E}_{\alpha}(X) = E\left[\frac{\alpha}{1-\alpha}X^+ + X^-\right] = V_{\alpha}(X) - E[X]$  = normalized Koenker-Bassett Error.

The quantile regression based on Quantile Quadrangle methodology and decomposition theorem, solves the problem in two steps.

Step 1. Find an optimal vector 
$$\widetilde{\boldsymbol{x}}^{a} = (x_{1}^{a}, \dots, x_{I}^{a})^{T}$$
 by minimizing CVaR deviation:  

$$\min_{\boldsymbol{x}} D_{\alpha}(Z_{0}(\widetilde{\boldsymbol{x}})) = \min_{\boldsymbol{x}} CVaR_{\alpha}^{\Delta}(Z_{0}(\widetilde{\boldsymbol{x}})), \qquad (CS.2)$$

where  $Z_0(\tilde{\boldsymbol{x}}) = \theta_0 - \tilde{\boldsymbol{x}}^T \tilde{\boldsymbol{\theta}}$ .

**Step 2**. Calculate intercept,  $x_0^{\alpha}$ ,

$$x_0^{\alpha} \in S_{\alpha}(Z_0(\widetilde{\boldsymbol{x}}^{\alpha})) = VaR_{\alpha}(Z_0(\widetilde{\boldsymbol{x}}^{\alpha})).$$
(CS.3)

#### **Optimization Problem 1**

minimizing Koenker-Bassett error

$$\hat{\boldsymbol{x}}_{\alpha} = \min_{\boldsymbol{x}} \left\{ kb_{err_{\alpha}}(L(\boldsymbol{x},\boldsymbol{\theta})) \right\}$$
(CS.4)

The minimization problem (CS.4) is reformulated with PSG Partial Moment Functions, pm\_pen and pm\_pen\_g.

### **Optimization Problem 2**

minimizing Koenker-Bassett error presented with Partial Moments

$$\min_{\mathbf{x}} \left\{ \alpha \cdot pm_{pen}(L(\mathbf{x}, \boldsymbol{\theta}), 0) + (1 - \alpha) \cdot pm_{pen} g(L(\mathbf{x}, \boldsymbol{\theta}), 0) \right\}.$$
(CS.5)

Two step procedure is implemented with PSG functions, CVaR Deviation (cvar\_dev) and VaR Risk (var\_risk).

# **Optimization Problem 3**

Step 1. minimizing CVaR deviation

$$\min_{\mathbf{x}} \left\{ cvar\_dev_{\alpha}(Z_0(\tilde{\mathbf{x}})) \right\}.$$
(CS.6)

Step 2.

$$x_0^{\alpha} = var\_risk_{\alpha}(Z_0(\widetilde{\boldsymbol{x}}^*)).$$
(CS.7)