Risk-averse inference using higher moment coherent risk measures

Spiridon I. Penev [1] and But-Elle Hatooka
The University of New South Wales, Sydney, Australia

[1]Corresponding author: Spiridon Penev, e-mail: s.penev@unsw.edu.au

Abstract

The higher moment coherent risk measures are parameterized by a parameter $p \geq 1$ which denotes the norm used in their definition. The particular case $p = 1$ corresponds to the average value at risk, but our interest is in the cases where $p > 1$. Since financial returns are recognised to be skewed and leptokurtic, it is desirable for a risk functional to incorporate the information of higher moments more directly in its definition, hence using the higher moment coherent risk measures is believed to more accurately reflect the risk aversity to large negative outliers. We consider the conditional analogues of these higher order measures whereby the risk evaluation takes into account a set of exogenous random variables. We focus on inference using such conditional risk measures. Finding point estimators of the conditional risk can be done using methods from constrained convex optimization theory. Furthermore, we discuss the issue about confidence interval construction where jackknife-type methods can be utilised. We compare the performance of models that use higher moment coherent risk measures as a criterion with the models that use average value at risk. To this end, the residuals of the resulting regression models are examined carefully. Both simulated and empirical data illustrate the benefits of using the higher moment coherent risk measures by a risk-averse investor.

Key Words: constrained convex optimization, jackknife, regression, risk measure.

1 Introduction

There has been increased interest in applying coherent risk measures in Finance. After the pioneering paper [1] introduced the axiomatic definition of a *coherent risk measure*, the coherency conditions have been slightly modified or extended by several authors. We formulate the convex coherent risk measure axioms below. For random variables X, Y, the coherency of the risk ρ (with values on the real axis) requires:

- convexity $\rho(\lambda X + (1-\lambda)Y) \leq \lambda\rho(X) + (1-\lambda)\rho(Y), \lambda \in [0, 1]$.
- monotonicity: if $X \leq Y$ then $\rho(X) \geq \rho(Y)$
- positive homogeneity: if $\lambda \geq 0$ then $\rho(\lambda X) = \lambda\rho(X)$
- translation invariance: if $m \in \mathbb{R}$ then $\rho(Y + m) = \rho(Y) - m$.

The convexity requirement reflects the view that diversification should not increase risk. As the popular Value at Risk (VaR) is not a coherent risk measure, alternatives to VaR have been suggested and investigated. This has not only been an academic activity. Recently, the Basel Committee on Banking Supervision was considering scrapping the use of VaR as a mean tool for modelling market risk. As alternative, the Average Value at Risk (AVaR)([9]) has been suggested.

The AVaR plays a central role in the description of every coherent risk measure via the *Kusuoka representation* (see [8], [3] and the next Section). AVaR is a coherent risk measure hence it is preferred in stochastic optimization. However, there are other coherent risk measures, generated from AVaR via the Kusuoka representation, namely the Higher Order Tail Risk measures, that may have advantages in comparison to AVaR for the purpose of risk management, as discussed next.

2 Higher Moment Coherent Risk Measures (HMCRM)

The HMCRM were investigated in [5] (who gave them the name), and in [3]. The latter paper studies the dual representation of these measures (that is, their Kusuoka representation). To explain Kusuoka's representation, we need the definition of a Kusuoka measure of risk. We define $\rho(Z)$ to be a *Kusuoka measure of risk* if there exists a convex set $\mathcal{M}$ in the set $\mathcal{P}((0,1])$ of probability measures on $(0,1]$ such that for all such Z we have

$$\rho(Z) = \sup_{m\in\mathcal{M}} \int_0^1 \mathrm{AVaR}_\alpha(Z)\, m(d\alpha). \tag{1}$$

Kusuoka's fundamental result [6] states that under general conditions, a coherent measure of risk on $\mathcal{L}^\infty(\Omega,\mathcal{F},P)$ is a Kusuoka measure. It was extended for $\mathcal{L}^p$ spaces ($p \geq 1$) (see [10]). It can be used in two ways: to show a Kusuoka representation for a coherent measure or, by choosing the set $\mathcal{M}$, to generate new coherent risk measures. In [3], the explicit Kusuoka representation for the HMCRM (5) was found. The advantage of HMCRM in comparison to other preceding coherent risk measures is that their tail cutoff point is adjustable to the chosen level $\alpha \in (0,1)$.

For a random variable $X \in \mathcal{L}^1(\Omega,\mathcal{F},P)$, we denote the cumulative distribution function (cdf) and the higher order cdf as

$$F_X(\eta) = \mathbb{P}[X \leq \eta] = F_X^{(1)}(\eta),\ F_X^{(k)}(\eta) = \int_{-\infty}^{\eta} F_X^{(k-1)}(\alpha)\, d\alpha \quad \text{for ,} \quad k \geq 2. \tag{2}$$

Denote by $F_X^{(-1)}(\alpha) = \inf\,\{\eta : F_X(\eta) \geq \alpha\}$ for $0 < \alpha < 1$ the left-continuous inverse of the cumulative distribution function. Then $\mathrm{VaR}_\alpha(X) = -F_X^{(-1)}(\alpha)$. Also, $F_X^{(-2)}(\alpha) = \int_0^\alpha F_X^{(-1)}(t)dt, \alpha \in (0,1]$ is the *absolute Lorenz function.* Using the introduced notation, the *Average Value at Risk* of X at level α is defined as

$$\mathrm{AVaR}_\alpha(X) = -\frac{1}{\alpha} F_X^{(-2)}(\alpha) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_t(X)\, dt. \tag{3}$$

The following extremal representation of $\mathrm{AVaR}_\alpha(X)$ is valid ([3]):

$$\mathrm{AVaR}_\alpha(X) = -\frac{1}{\alpha}\sup_{\eta\in\mathbb{R}}\{\eta\alpha - F^{(2)}(\eta)\} = \inf_{\eta\in\mathbb{R}}\left\{\frac{1}{\alpha}\mathbb{E}[(\eta - X)_+] - \eta\right\}. \tag{4}$$

This extremal representation was generalized in [5] to suggest the HMCRM:

$$\inf_{\eta\in\mathbb{R}}\left\{\frac{1}{\alpha}\|(\eta - X)_+\|_p - \eta\right\}, \quad p > 1. \tag{5}$$

Examples of stock portfolio optimization models using (5) were shown in [5], and advantages of these measures in comparison to the mean-variance model of Markowitz were demonstrated. Our work ([7]) also indicates that if such type of risk measures (specifically, with values of $p = 2$ or $p = 3$) is used as a risk criterion in European option portfolio optimization, the time evolution of the portfolio is superior to the evolution of a portfolio optimized with respect to AVaR or to the Markowitz model.

3 Detailed treatment

Of course, amenability to easy calculation of plug-in estimators for any risk measure using data is also an important factor determining its widespread use. Let

$$\varrho_{c,p}(X) = \inf_{\eta\in\mathbb{R}}\{c\|(\eta - X)_+\|_p - \eta\} \tag{6}$$

where $c > 1$ is a constant and $p > 1$ is fixed by the choice of the norm. Given i.i.d. $X_i, i = 1, 2, \ldots, n$ and replacing by empirical we have:

$$\hat{\varrho}_{c,p} = \inf_{\eta\in\mathbb{R}}\{c\frac{1}{n^{1/p}}(\sum_{i=1}^{n} max(\eta - X_i, 0)^p)^{1/p} - \eta\} \tag{7}$$

Denoting $d_i = max(\eta - X_i, 0), i = 1, 2, \ldots, n$ and putting all $d_i, i = 1, 2, \ldots, n$ in a vector d we can rewrite (7) as follows:

$$\begin{gathered}\min_{\eta,d}\{c\frac{1}{n^{1/p}}(\sum_{i=1}^{n} d_i^p)^{1/p} - \eta\} \\ \text{subject to } -X_i + \eta \le d_i,\ i = 1, 2, \ldots, n; \quad d_i \ge 0\end{gathered} \tag{8}$$

Note that for $p = 1$, (8) is a *linear* optimization problem that can also be solved by the simplex method. Incidentally, this solution coincides with an explicit plug-in formula for estimating AVaR via a linear combination of order statistics. For HMCRM (i.e., for $p > 1$) no explicit solution exists, yet a unique numerical solution can always be obtained since the optimization problem and the constraints are convex.

3.1 Conditional HMCRM in regression problems

In linear regression setting we need to deal with a conditional variant of the risk measure since the risk of the output variable is considered conditionally on a realisation of the regressors. Consider $\mathcal{L}_p(\Omega, \mathcal{F}_1, \mathbb{P})$ and $\mathcal{L}_p(\Omega, \mathcal{F}_2, \mathbb{P})$ with $\mathcal{F}_1 \subset \mathcal{F}_2$. Define for $X \in \mathcal{L}_p(\Omega, \mathcal{F}_2, \mathbb{P})$ a conditional mapping

$$\varrho_{p,\alpha}(X|\mathcal{F}_1) = argmin_{\eta(\omega)\in\mathcal{L}_p(\Omega,\mathcal{F}_1,\mathbb{P})} \left\{ -\eta(\omega) + \frac{1}{\alpha}\mathbb{E}[(\eta - X)_+^p|\mathcal{F}_1]^{1/p}(\omega) \right\}.$$

When $p > 1$ we call it conditional HMCRM, and for $p = 1$ this is the conditional AVaR. The conditioning argument leads to an easy treatment of conditional risk measures in linear regression setting as discussed in [2] for the case of AVaR. However our goal here is the application for HMCRM. Consider a linear regression model

$$Y_i = \beta_0 + \boldsymbol{\beta}^T \boldsymbol{X}_i + \epsilon_i, i = 1, 2, \dots, n$$

with $\boldsymbol{X}_i, i = 1, 2, \dots, n$ being $K \times 1$ i.i.d. random vectors, and errors $\epsilon_i, i = 1, 2, \dots, n$ homoscedastic random variables, with finite variance σ^2. We solve:

$$\min_{\eta,\beta_0,\boldsymbol{\beta}} \left\{ \frac{1}{\alpha}\frac{1}{n^{1/p}}\big(\sum_{i=1}^n d_i^p\big)^{1/p} - \eta \right\} \tag{9}$$

subject to

$$-Y_i + \beta_0 + \boldsymbol{\beta}^T \boldsymbol{X}_i - \eta \le d_i, \qquad d_i \ge 0, i = 1, 2, \dots, n$$

with the added constraint

$$\sum_{i=1}^n Y_i = n\beta_0 + \sum_{i=1}^n \boldsymbol{\beta}^T \boldsymbol{X}_i$$

Denote the solution by $\tilde{\beta}$ and $\tilde{\beta}_0$ and the residual by $\tilde{\epsilon}_i = Y_i - \tilde{\beta}^T \mathbf{X}_i - \tilde{\beta}_0$. The coherency implies: $\tilde{\varrho}_{p,\alpha|\boldsymbol{X}_i}(Y_i) = \tilde{\varrho}_{p,\alpha}(\tilde{\epsilon}) - \tilde{\boldsymbol{\beta}}^T \boldsymbol{X}_i - \tilde{\beta}_0$. Since no explicit form exists for the variance of the conditional risk estimator, we approximate it using the conventional jackknife estimator or by approximating the influence function in the non-parametric delta method.

3.2 Numerical Implementation and discussion of the results

Our objective is to demonstrate that the residuals obtained from using the β-coefficients in (9) with $p > 1$ are more conducive to "risk-aversity" than those yielded when $p = 1$ or when estimating the coefficients using the more conventional OLS approach . All the residuals for $p \ge 1$ were calculated using Matlab with accessibility of the convex optimisation package `cvx` ([4]) and we call them *the CVX residuals.*

We performed Monte Carlo simulations whereby the error terms are generated from the following distributions: (i) $N(0, 4)$, (ii) t distribution with 5 degrees of

Table 1: Results for $t(5)$ (left) and for CN(1,25) (right)

n	p	Min.$\hat{\epsilon}$	No$<\hat{w}_L$	Min.$\tilde{\epsilon}$	No$<\tilde{w}_L$	Min.$\hat{\epsilon}$	No$<\hat{w}_L$	Min.$\tilde{\epsilon}$	No$<\tilde{w}_L$
500	1	**-8.065**	**10**	-8.069	**10**	-6.997	3	-6.846	2
	2			-6.743	**0**			-6.717	2
	3			**-6.743**	**0**			-6.717	2
1000	1	**-8.759**	**16**	-8.823	**15**	-9.723	19	-9.642	18
	2			-6.787	**4**			-9.031	12
	3			**-6.616**	**4**			-8.972	12

freedom (denoted as $t(5)$), (iii) Contaminated Normal, where $N(0,4)$ is contaminated with 20% $N(1,25)$ (denoted as $CN(1,25)$), and (iv) Log-Normal,with parameters 0 and 1 (denoted as $LN(0,1)$). The heavy-tailed distributions (ii) - (iv) present appropriate tests for our claims. We generate $X_i \sim U[-2,2]$ and construct linear models

$$Y_i = 1.5 + X_i + \epsilon_i, i = 1, 2, \dots, n$$

with $n = 500, 1000$. Due to a lack of space, we summarise our findings in a Table 1 although graphs are more telling. We also choose $\alpha = 0.05$ to discuss, however maintain that the results are reflective for the other $\alpha = 0.01$, 0.1 that were also simulated. For both $n = 500$ and $n = 1000$, the most prominent effect occurs on the $t(5)$ and $CN(1,25)$ distributions. We denote the CVX estimates by $\tilde{\boldsymbol{\beta}}, \tilde{\beta}_0$, as opposed to the OLS estimates $\hat{\boldsymbol{\beta}}, \hat{\beta}_0$. The residuals are denoted by $\tilde{\epsilon}$ and $\hat{\epsilon}$, accordingly. We compare the OLS and CVX residuals of these distributions in our example. The contrast in the distribution of these residuals is clearer from the boxplots of the two methods, with lower OLS whisker $\hat{w}_L > \tilde{w}_L$, and $|\hat{w}_L - \hat{q}_{0.25}| < |\tilde{w}_L - \tilde{q}_{0.25}|$, where q_α denotes the α-quantile. We calculated each whisker using Matlab's default, that is $w = q_{0.25} - 1.5(q_{0.75} - q_{0.25})$. The results suggest that the CVX residuals are more negatively skewed than their OLS counterparts, with large negatives occurring with lower probability. As we can see, the CVX residuals for $p = 2, 3$ reflect the notion of risk-aversity, consistently outperforming the benchmark in AVaR, as well as their OLS counterparts. By implementing the risk-averse conditional HMCRM, the largest negative residual is reduced in magnitude, and there are countable less negative outliers. We justify measuring outliers relative to the lower whisker of each distribution, even though these values differ, since $\hat{w}_L$ and $\tilde{w}_L$ are functions of the quantiles and capture the shape of the residual distribution.

We apply our methods to real data consisting of 1010 daily closing prices in cents from 01/07/2008 – 02/07/2012. We let the dependent variable be the daily log return of Apple Inc. (NASDAQ:AAPL) stock price, so that $Y_i = \log(A_{i+1}/A_i)$, where $\{A_i\}$ denote the raw daily return for AAPL stock at time i. Similarly, we let the explanatory variables X_1 and X_2 be the daily log return of Google Inc. (NASDAQ:GOOG) and the index NASDAQ-100 (INDEX:IUXX) respectively.

We construct an autoregressive model on the raw data of the model, so that $Y_i, X_{1,i}$ and $X_{2,i}$ are now the raw daily returns of AAPL, GOOG and NASDAQ-100. The third predictor $X_{3,i}$ is taken to be the closing price of AAPL at time $i-1$, so

Table 2: Results for autoregression model

α	p	Min. $\hat{\epsilon}$ (USD)	Number $< \hat{w}_L$	Min. $\tilde{\epsilon}$ (USD)	Number $< \tilde{w}_L$
0.01	1	**-24.721**	**27**	-25.340	**5**
	2			**-19.866**	**0**
0.05	1	-24.721	**27**	-25.614	**8**
	2			**-19.887**	**0**
0.1	1	-24.721	**27**	-25.025	**14**
	2			**-23.625**	**4**

that the model is given by

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 Y_{i-1} + \epsilon_i, i = 2, 3, \ldots, 1010.$$

The results of the residual analysis are presented in Table 2 whose interpretation is the same as Table 1. A complete discussion of the adequacy of this model and of the inference will be given in an extended version of this paper. However, Table 2 succinctly illustrates the effects of prime interest in this note.

References

[1] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999) "Coherent measures of risk," *Mathematical Finance*, 9, 203–228.

[2] Chun, S., Shapiro, A., Uryasev, S. (2012) "Conditional Value-at-Risk and Average Value-at-Risk: Estimation and Asymptotics," *Operations Research*, 60, 739–756.

[3] Dentcheva, D., Penev, S., Ruszczyński, A. (2010) "Kusuoka representation of higher order dual risk measures," *Annals of Operations Research*, 181, 325–335.

[4] Grant, M., Boyd, S. (2011) *CVX: Matlab software for disciplined convex programming, version 1.21*, `http://cvxr.com/cvx/`

[5] Krokhmal, P. (2007) "Higher moment coherent risk measures," *Quantitative Finance*, 7, 373–387.

[6] Kusuoka, S. (2001) "On law invariant coherent risk measures," *Adv.Math. Econ.*, 3, 83–95.

[7] Matmoura, Y. , Penev, S. (2013) "Multistage Optimization of Option Portfolio using Higher Order Coherent Risk Measures," *European Journal of Operational Research*, 227, 190–198.

[8] Pflug, G., Römisch, W. (2007) *Modeling, measuring and managing risk,* World Scientific.

[9] Rockafellar, R., Uryasev, S. (2002), "Conditional value-at-risk for general loss distributions," *J. Bank. Financ.,* 26, 1443–1471.

[10] Shapiro, A., Dentcheva, D., Ruszczyński, A., 2009. *Lectures on Stochastic programming: Modelling and Theory,* SIAM.