Some Recent Developments in Expected Shortfall Regression

Xuming He

Washington University in St. Louis

July 3, 2023; Warsaw

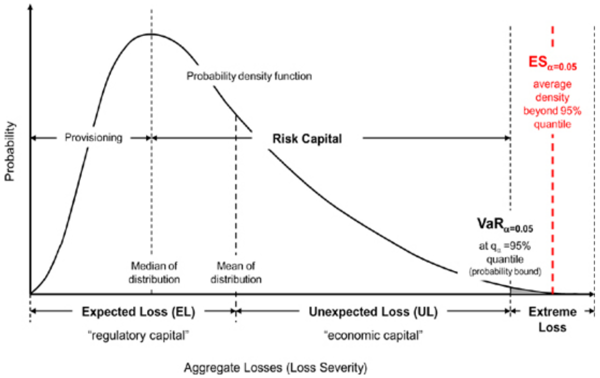
Outline

1. Introduction to expected shortfall
2. Conditional expected shortfall
3. Estimation methods
4. Example
5. Final words

Summary statistics are used everywhere for communication and understanding.

- University of Warsaw is ranked #2 in Poland by Times Higher Education Rankings.
- Poland passport is ranked #9 in the World by the Guide Passport Ranking Index.
- Poland is 25-th safest country ranked by Global Peace Index.

Investment Risk (IMF eLibrary)



Expected Shortfall (upper tail):

For a continuous random variable Y, the expected shortfall of level τ is

$$\boxed{v_Y(\tau) = E[Y|Y \geq q(\tau)]},$$

where $q(\tau)$ is the τ-quantile of Y.

An equivalent expression: $\boxed{v(\tau) = \frac{1}{1-\tau}\int_\tau^1 q(u)du}.$

Other names:

- Conditional Value-at-Rick (CVaR)
- Superquantile
- Expected tail loss; Tail average

ES as measure of treatment effect

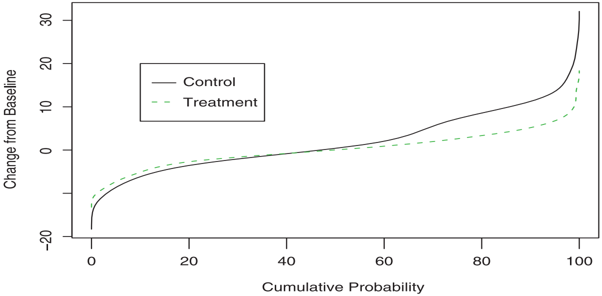
- Treatment $T \in \{0, 1\}$
- Potential outcome $Y(0), Y(1)$
- Average treatment effect $(ATE) = E[Y(1)] - E[Y(0)]$
- Tail average treatment effect

$$TATE = v_{Y(1)}(\tau) - v_{Y(0)}(\tau).$$

A Clinical Example:

- Rheumatoid arthritis is a chronic inflammatory disorder (of joints).
- Most treatments work for some but not for all.
- Total Sharp Score, a measure of joint space narrowing and erosion.

X. HE, Y.-H. HSU AND M. HU

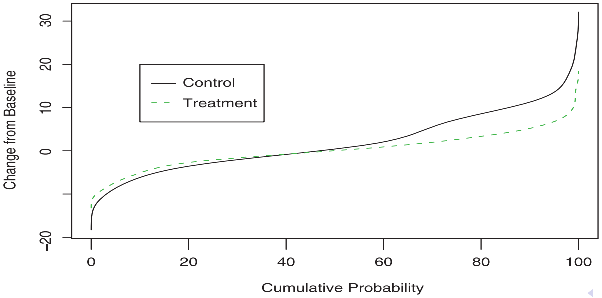


Quantile function of the TSS change shows that the groups differ mostly in the upper tails.

Treatment effect in one tail

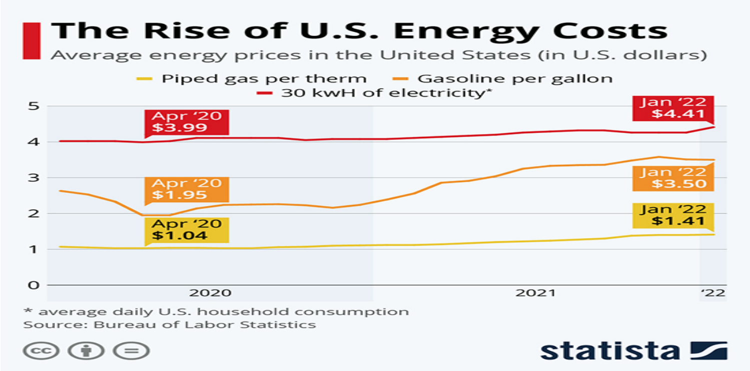
- Median treatment effect is nearly zero.
- Mean treatment effect is harder to detect.
- Expected shortfall at $\tau = 0.75$ is a useful target. Substantial sample size reduction can be achieved.

X. HE, Y.-H. HSU AND M. HU



Other Applications

- Impact of higher energy price on household consumption for the heavier users.
- Effect of job training program on income for the lower-paid or higher paid workers.



Why conditional expected shortfall?

- Losses of different portfolios can be highly correlated, and depend on some common factors. Lesson from the 2007-2009 financial crisis!

$$\boxed{v_{Y|X}(\tau) = v(\tau, x)}$$

- Covariate-adjustments can improve statistical power in treatment effect estimation and detection.

$$v(\tau, x) = \alpha(\tau) + x^T \beta(\tau).$$

Challenges in Estimation

- Insufficient data for any given $X = x$;
- Tail quantile estimates tend to be unstable;
- Expected shortfall is not elicitable (Gneiting, 2011). What?

Mean: $\mu = \arg\min_{\boldsymbol{u}} E[(Y - \boldsymbol{u})^2].$

Quantile: $\boldsymbol{\beta}(\tau) = \arg\min_{\boldsymbol{u}} E[\rho_\tau(Y - \boldsymbol{u})],$

where $\rho_\tau(u) = u[\tau - \mathbf{1}(u < 0)].$

Challenges in Estimation

- Insufficient data for any given X;
- Tail quantile estimates tend to be unstable;
- Expected shortfall is not elicitable (Gneiting, 2011).

Quantile: $\boldsymbol{\beta}(\tau) = \arg\min_{\boldsymbol{u}} E[\rho_\tau(Y - \boldsymbol{X}_i^T \boldsymbol{u})]$,

where $\rho_\tau(u) = u[\tau - \mathbf{1}(u < 0)]$, but

no such loss function defines the expected shortfall.

Solutions: two starting points

1. Start from τ-quantile $q(\tau)$.

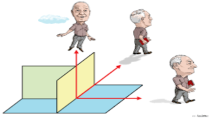
   -  (Quantile, Expected shortfall) is jointly elicitable (Fissler and Ziegel, 2016) $\hookrightarrow$ joint estimation.

     The loss funciton is highly nonconves and non-smooth.

2. Start from a preliminary estimate of $v(\tau)$.

Solutions: two starting points

1. Start from τ-quantile $q(\tau)$.
   -  (Quantile, Expected shortfall) is jointly elicitable (Fissler and Ziegel, 2016) $\hookrightarrow$ joint estimation.

     The loss funciton is highly nonconves and non-smooth.
   -  Find a Neyman orthogonalized score for estimating the expected shortfall $\hookrightarrow$ Two-step estimation. (Barendse, 2020)
2. Start from a preliminary estimate of $v(\tau)$.

Solutions: two starting points

1. Start from τ-quantile $q(\tau)$.
   - Joint estimation.
   - Two-step estimation.
2. Start from a preliminary estimate of $v(\tau, x)$ at every or some x, and find a way to linearize over x.

Two-step Estimation: with data (X_i, Y_i)

Assume $q(\tau)$ and $v(\tau)$ (lower tail) are

$$q(\tau) = x^T \beta(\tau), \quad v(\tau) = x^T \gamma(\tau).$$

1. Find quantile regression $\hat{\beta}$. (Koenker and Bassett Jr, 1978)
2. Among possible joint loss functions, find one with Neyman orthogonal score for estimation of γ. (Barendse, 2020)

$$Z_i := (Y_i - X_i^T \hat{\beta}) I(Y_i \leq X_i^T \hat{\beta}) + \tau X_i^T \hat{\beta}$$

$$\boxed{\hat{\gamma} = arg\,min_\gamma \sum_i (Z_i - \tau X_i^T \gamma)^2.}$$

Two-step Estimation (continued)

$$v(\tau, x) = (1-\tau)^{-1} \int_{\tau}^{1} q(u, x) du = h(F_{Y|X}) = x^T \gamma,$$

where h is a functional on the conditional distribution of $Y|X$.

1. Use nonparametric or machine learning methods for $F_{Y|X}$
2. Use influence function adjustment to construct an estimating equation for γ (Chetverikov et al., 2022)

R_i : some pseudo response

$$\boxed{\hat{\gamma} = arg\, min_{\gamma} \sum_i (R_i - X_i^T \gamma)^2.}$$

Two-step Estimation (continued)

For possibly heavier tailed outcomes, robust estimation can be done by replacing

$$\boxed{\hat{\gamma} = arg\,min_{\gamma} \sum_i (R_i - X_i^T \gamma)^2.}$$

by

$$\hat{\gamma} = arg\,min_{\gamma} \sum_i h_c(R_i - X_i^T \gamma),$$

where h_c is the Huber loss.
Reference: He et al. (2023)

Solutions: two starting points (re-visited)

1. Start from τ-quantile $q(\tau)$.
    - Joint estimation.
    - Two-step estimation. 
2. Start from a preliminary estimate of $v(\tau, x)$ at every or some x, and find a way to linearize over x. More to come.

Optimization Approach

Key fact: (Rockafellar et al., 2014)

$$v(\tau) = \text{argmin}_c \Big[c + \frac{1}{1-\tau} \int_0^1 (v(u) - c)^+ du \Big].$$

The loss function is $convex$ in c if $v(u)$ is given.

Useful for the computation of expected shortfall?

Optimization Approach

Key fact: (Rockafellar et al., 2014)

$$v(\tau) = argmin_c \Big[c + \frac{1}{1-\tau} \int_0^1 (v(u) - c)^+ du \Big].$$

Generalization to regression:

$$\gamma(\tau) = argmin_\gamma \Big[E\{X^T\gamma\} + \frac{1}{1-\tau} \int_0^1 \{v(u, Y - X^T\gamma)\}^+ du \Big],$$

where $v(u, V)$ is the (upper) u-expected shortfall of V.

Optimization Approach: bad news?

$$\gamma(\tau) = argmin_{\gamma}\left[E\{X^T\gamma\} + \frac{1}{1-\tau}\int_0^1 \{v(u, Y - X^T\gamma)\}^+ du\right],$$

can be solved iteratively but is not Fisher consistent to the conditional expected shortfall regression coefficient in general. Yet, this approach has been used quite regularly.

A New Characterization:

$$\gamma(\tau) = arg\,min_{\gamma} E\Big[\rho_{\tau}(v(U, X) - X^{T}\gamma)\Big],$$

where $v(u, x)$ is the u-expected shortfall of $Y|X = x$, and

$$U \sim Unif(0, 1),$$

and the expectation is taken over (X, U).

Dissertation work of

Yuanzhi Li, Ph.D. (2022)
Just joined Five Rings Capital, New York

New Approach: Practically Speaking

$$\gamma(\tau) = argmin_{\gamma} E\Big[\rho_\tau(v(U, X) - X^T\gamma)\Big],$$

$$\hat{\gamma}(\tau) = argmin_{\gamma} \sum_i \sum_j \Big[\rho_\tau(\hat{v}(U_j, X_i) - X_i^T\gamma)\Big],$$

where $\{u_j\}$ is a uniform grid of the interval

$$[\tau - \delta\tau, \tau + \delta(1 - \tau)]$$

for any constant $\delta \in (0, 1]$.

Input: initial expected shortfall estimates $\hat{v}(u_j, X_i)$.

New Approach: a “minor” detail

$$\hat{\gamma}(\tau) = argmin_{\gamma} \sum_i \sum_j \left[\rho_\tau(\hat{v}(u_j, X_i) - X_i^T \gamma) \right],$$

where $\{u_j\} \in [\tau - \delta\tau, \tau + \delta(1-\tau)]$ for any constant $\delta \in (0, 1]$.

- $\delta = 1 \longrightarrow u_j \in [0, 1]$
- $\delta < 1$ avoids the extreme quantile calculations.

New Approach: Practically Speaking

$$\hat{\gamma}(\tau) = argmin_{\gamma} \sum_i \sum_j \left[\rho_\tau(\hat{v}(u_j, X_i) - X_i^T \gamma) \right],$$

How about iRock?

New Approach: Theoretically Speaking

$$\hat{\gamma}(\tau) \approx argmin_{\gamma} \sum_{Y_i > q_i} \left[w_i (Y_i - X_i^T \gamma)^2 \right],$$

where $q_i = q(\tau, X_i)$, τ-quantile of $Y|(X = X_i)$,
and w_i is the weight that depends on the model.

Our findings:

- *iRock* adapts to data heterogeneity nicely.
- The initial estimates do not need to be root-n consistent.

Statistical Property of $iRock$

$$\sqrt{n}(\hat{\gamma}(\tau) - \gamma(\tau)) \xrightarrow{d} N(0, D^{-1}\Omega D^{-1}),$$

where

$$D = E\left(\frac{XX^T}{v(\tau, X) - q(\tau, X)}\right),$$

$$\Omega = E\left(\frac{\sigma_\tau^2(X)XX^T}{[v(\tau, X) - q(\tau, X)]^2}\right),$$

$$(1-\tau)\sigma_\tau^2(x) = \underbrace{\tau[v(\tau, x) - q(\tau, x)]^2}_{} + \underline{\underline{Var(Y|X = x, Y \geq q(\tau, x))}}.$$

Statistical Property of $iRock$

$$\sqrt{n}(\hat{\gamma}(\tau) - \gamma(\tau)) \xrightarrow{d} N(0, D^{-1}\Omega D^{-1}),$$

Requirements on initial estimators:

- $\hat{v}(u, x)$ are sufficiently smooth in u and converges at the rate of $o_p(n^{-1/4})$.
- Some aggregated average of $\hat{v}(\tau, x) - v(\tau, x)$ over $x = X_i$ converges to 0 at the rate of $O_p(n^{-1/2})$.

Example: Binning of the covariates.

A Comparison

- Location-scale model (all quantile functions are linear)

$$Y = (X^T\beta) + (X^T\gamma)e.$$

- Comparison with
  - N_2 = Neyman-orthogonalized estimator
  - J_1, J_2 = two specific joint estimators advocated by (Dimitriadis et al., 2019) and (Patton et al., 2019), respectively

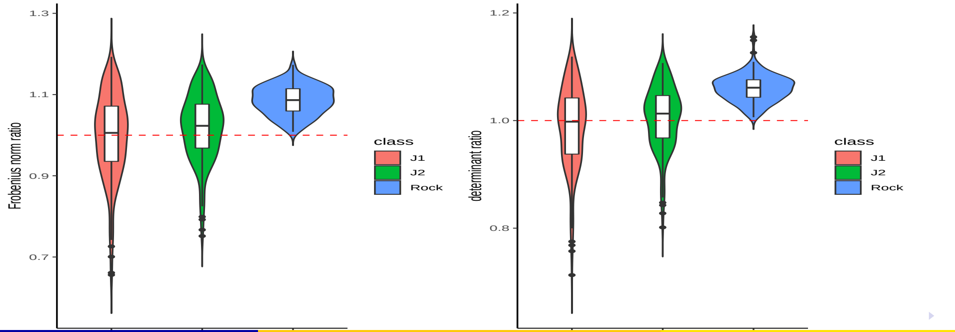
- If $X^T\gamma = 1$, but $X^T\beta$ is not constant, then

$$\boxed{Var_{iRock} = Var_{N_2} < min\{V_{J_1}, V_{J_2}\}}$$

A comparison in case of heterogeneity

$$Y = (X^T\beta) + (X^T\gamma)e.$$

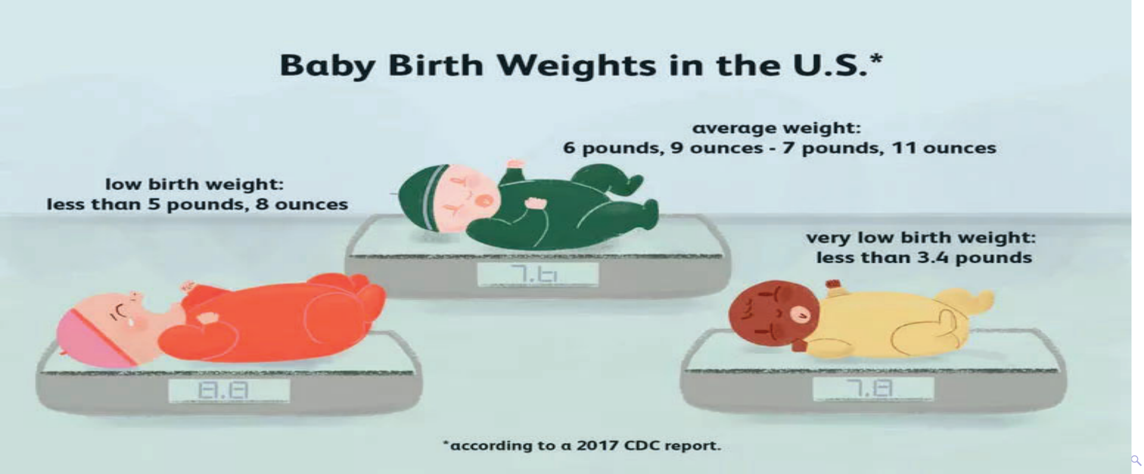
Figure: Efficiency comparison relative to the 2-step estimator N_2; With different values of γ and β; $dim(X) = 3$, $e \sim N(0,1)$.



Conclusions about $iRock$

1. Rockafeller's optimization-based characterization of superquantile motivated us to find a valid approach to expected shortfall regression.
2. It requires no joint modeling of quantile regression.
3. It is automatically adaptive to data heterogeneity in a wide variety of models (often outperforming other methods).
4. Its implementation requires initial expected shortfall estimation, which is challenging if $\dim(X)$ is high.

Example: low birth weight



Example: Nulliparity as a risk factor for low birth weight

- Nulliparity is known to be a risk factor of preterm birth and low birth weight (Shah, 2010)
- The effect has been observed across countries and age groups, but the effect sizes vary.
- Our focus: male singleton birth to generally healthy older mothers (age 36+)
- Data: 2020 U.S. birth-weight dataset, available at the National Center for Health Statistics. (Black or white mothers; married, college-educated, non-smokers; $n = 79,336$)

Example: Summary Statistics

Table: Average and interquartile range in the sample

Variable	Parity=1	Parity > 1
Birth weight (g)	3301 (3005–3657)	3482 (3185–3820)
Maternal age	38.0 (36.0 – 39.0)	38.1 (36.0 – 39.0)
Gestational weight gain (lb)	30.4 (21.0 – 38.0)	30.2 (22.0 – 38.0)
% of Black mothers	10.2	10.7

Example: Analysis of (lower tail) Expected Shortfall

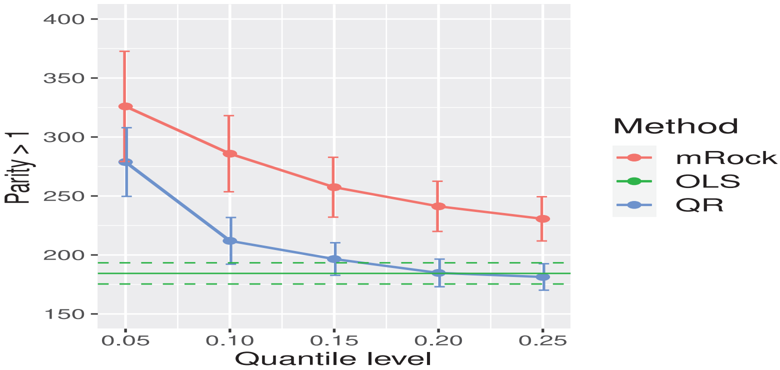
$$\text{Birth weight} = (\text{Parity} > 1) + (\text{Race} = \text{black}) + \text{Mom age} + \text{Mom weight gain} + (\text{Mom weight gain})^2$$

Table: Coefficients in ES Regression

Covariates	$\tau = 0.05$	$\tau = 0.20$
Intercept	2006.83 (21.53)	2580.32 (9.62)
Parity	325.93 (23.35)	241.22 (10.64)
Race	-505.92 (39.09)	-297.43 (16.62)
Mom age	-34.03 (5.33)	-22.20 (2.48)
MWG	21.39 (0.92)	12.67 (0.43)
MWG^2	-0.41 (0.04)	-0.22 (0.02)

Example: Analysis of (lower tail) Expected Shortfall

For the lowest 5% of the birth weight distribution (given race, mother's age and weight gain), the average "loss" for nulliparous women in this sub-population was 326g.



Not yet covered in this presentation

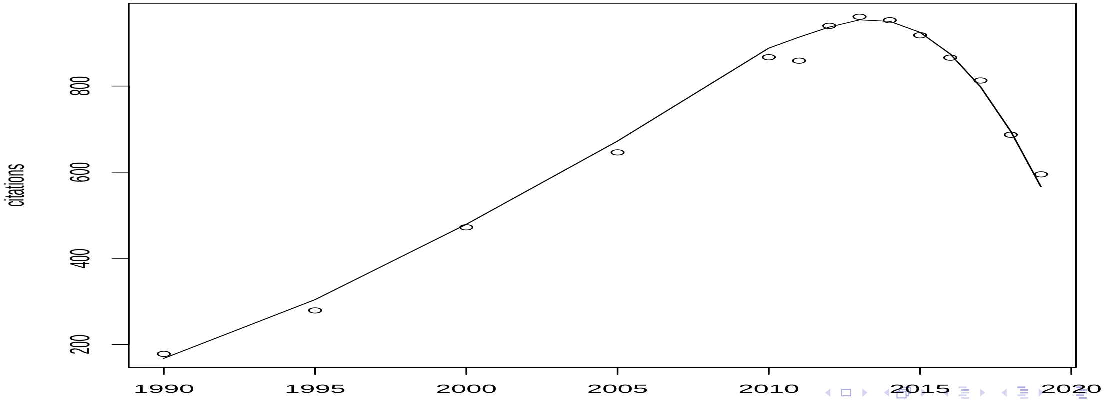
Regression Inference

Causal Inference

Computation/Implementation

A warning about “regression”?

Regression may need an age regression therapy …



Thank you!

Barendse, S. (2020). Efficiently weighted estimation of tail and interquantile expectations. *Tinbergen Institute Discussion Paper 2017-034/IIIs*.

Chetverikov, D., Liu, Y., and Tsyvinski, A. (2022). Weighted-average quantile regression. *https://doi.org/10.48550/arXiv.2203.03032*.

Dimitriadis, T., Bayer, S., et al. (2019). A joint quantile and expected shortfall regression framework. *Electronic Journal of Statistics*, 13(1):1823–1871.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and osband's principle. *The Annals of Statistics*, 44(4):1680 – 1707.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 58(494):746–762.

He, X., Tan, K., and Zhou, W. (2023). Robust estimation and inference for expected shortfall regression with many regressors. *J. Royal Statistical Society, Series B*.

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413.

Rockafellar, R. T., Royset, J. O., and Miranda, S. I. (2014). Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. *European Journal of Operational Research*, 234(1):140–154.

Shah, P. S. (2010). Parity and low birth weight and preterm birth: a systematic review and meta-analyses. *Acta Obstet Gynecol Scand.*, 89(7):862–75.