CASE STUDY: Expectile (XVaR) Estimation by Three Variants of Regression (xvar, xvar_dev, pm_pen, pm2_pen, avg)

**Background**

Rockafellar and Uryasev (2013) developed a paradigm called the Fundamental Risk Quadrangle, which links Risk Management, Reliability, Statistics, and Stochastic Optimization. Risk Quadrangles unite risk functions in groups (Quadrangles) consisting of four stochastic functions and a statistic:

- Risk \( R(X) \), providing a numerical surrogate for the overall hazard in \( X \);
- Deviation \( D(X) \), measuring nonconstancy in \( X \);
- Error \( E(X) \), measuring nonzeroness in \( X \);
- Regret \( V(X) \), which quantifying the regret in outcomes \( X > 0 \) versus \( X < 0 \);
- Statistic \( S(X) \) associated with \( X \) through \( E \) and \( V \).

The VaR Quadrangle (see, Rockafellar and Uryasev (2013)) and CVaR Quadrangle (Rockafellar et al (2014)) are popular quadrangles. VaR and CVaR functions are statistics in the corresponding quadrangles.

(Kuzmenko, Malandii, Uryasev (2023)) considered the Expectile (XVaR) Quadrangle (with piecewise linear error) where expectile is both Risk and Statistic. All functions are piece-wise linear in this quadrangle.

(Kuzmenko, Malandii, Uryasev (2023)) also considered two another Expectile Quadrangles where Expectile is a statistic only. One of these two quadrangles, Expectile Quadrangle (with asymmetric variance error), is motivated by the definition of Expectile by Newey and Powell (1987).

The considered Case Study estimates Expectile by linear regression, minimizing the Error function from Expectile Quadrangles with piecewise linear and asymmetric variance errors. The first Error is piece-wise linear, while the second one is quadratic w.r.t. decision variables.

Alternatively, Expectile regression can be done in two steps (Rockafellar et al (2013)):
1. Minimize Expectile Deviation from Expectile Quadrangle with the residual depending only on loading factors.
2. Calculate intercept = Expectile for the optimal value of variables from Step1.

The Case Study implements the following variants of Expectile regression:
1. Minimization of the Expectile piece-wise linear error expressed through the PSG functions \( \text{avg}, \text{pm}_\text{pen}, \text{variable} \).
2. Alternative regression – two-step procedure including minimization of the Expectile Deviation (from the quadrangle piece-wise linear error) with PSG function \( \text{xvar}_\text{dev} \) and calculation Expectile at optimal point using PSG functions, \( \text{xvar} \).
3. Minimization of the Expectile asymmetric quadratic error expressed through PSG functions \( \text{pm}_2\text{pen}, \text{pm2}_\text{pen}_\text{g} \).

See description of PSG function in [PSG Help](#).

The first two variants of regression are equivalent. However, the third variant may give different regression coefficients, compared to the first/second variant (the third variant is equivalent to the first/second only asymptotically).

The first and second variants of regression (Problems 1 and 2) are run with the dataset from the Case Study: [Estimation of CVaR through Explanatory Factors with CVaR (Superquantile) Regression](#).

The first and third variants (problems 3a, 4a, 3b, 4b) are done with the dataset from the Case Study: [Logistic Regression and Regularized Logistics Regression Applied to Estimating the Probability of Cesarean Section](#).
When all variables except the "intercept" (column containing "1" in all rows) are set to zero in problems 3a, 4a, the first and the third variants give the same optimal value for variable "intercept", which equals to Expectile (problems 3b, 4b).

References


Notations

\( I \) = number of factors in regression, \( i = 1, ..., I \) index of factor;
\( \theta_i \) = independent random variable (value of factor \( i \)), \( i = 1, ..., I \);
\( \theta_0 \) = dependent random variable;
\( \theta = (\theta_0, \theta_1, ..., \theta_I) \) = random vector;
\( J \) = number of scenarios (observations); \( j = \) index of scenario, \( j = 1, ..., J \);
\( p_j \) = probability of scenario, \( j = 1, ..., J \);
\( \{\theta_{i1}, ..., \theta_{ij}\} \) = scenarios of random variable \( \theta_i \), \( i = 0, 1, ..., I \);
\( x = (x_1, ..., x_I) \) = vector of regression coefficients (loading factors);
\( x_0 \) = intercept;
\( L(x_0, x, \theta) = \theta_0 - \left( x_0 + \sum_{i=1}^{I} \theta_i x_i \right) \) = residual of the regression (Loss function);
\( L_0(x, \theta) = \theta_0 - \sum_{i=1}^{I} \theta_i x_i \) = residual without intercept;
\( q \in (0.5, 1) \) = parameter of Expectile function in Quadrangle with asymmetric variance error;
\( K \in (0, \infty) \) = parameter of Expectile function in the Quadrangle with piecewise linear error;
\( K = \frac{1-q}{2q-1} \) = correspondence between parameters of two quadrangles;
\((X)_+ = \{X, \text{for } X > 0\}, (X)_- = \{-X, \text{for } X < 0\} = \{0, \text{otherwise}\}.

**Error, Deviation and Statistic: Expectile Quadrangle (with piecewise linear error) (Kuzmenko (2020))**
\[ \mathcal{E}_q(X) = \max \left\{ -E[X]; \frac{1}{K} E[(X)_+] \right\} \] = piecewise linear Error;
\[ \mathcal{D}_q(X) = XVaR_{dev}_K(X) \] = piecewise linear Deviation;
\[ \mathcal{S}_q(X) = XVaR_{q}(X) \] = Expectile (Statistic).

**Error, Deviation and Statistic: Expectile Quadrangle (with asymmetric variance error) (Kuzmenko, Malandii, Uryasev (2023))**
\[ \mathcal{E}_q(X) = qE[(X)_+^2] + (1-q)E[(X)_-^2] \] = asymmetric variance Error;
\[ \mathcal{D}_q(X) = qE\left[\left((X - e_q(X))_+\right)^2\right] + (1-q)E\left[\left((X - e_q(X))_-\right)^2\right] \] = asymmetric Deviation;
\[ \mathcal{S}_q(X) = XVaR_{q}(X) = e_q(X) = \text{Expectile (Statistic)}. \]

**PSG functions** used in Case Study (see PSG Help):
\( xvar_q(L(\cdot)) \) = Expectile with parameter \( q \) of the random Loss function;
\( xvar_{dev}_q(L(\cdot)) \) = Expectile Deviation with parameter \( q \) of the random Loss function;
\( avg(L(\cdot)) \) = Average of the random Loss function;
\( pm\_pen_0(L(\cdot)) \) = Partial Moment with threshold 0 of the random Loss function;
\( pm2\_pen_0(L(\cdot)) \) = Partial Moment Quadratic with threshold 0 of the random Loss function;
\( pm2\_pen\_g_0(L(\cdot)) \) = Partial Moment Quadratic with threshold 0 of the random Gain function (\( Gain = -Loss \)).

**Optimization problems**

**Optimization Problem 1**
Expectile regression by minimizing piecewise linear Error.

**Minimizing piecewise linear Error**
\[ \min_{x_0 \in \mathbb{R}, \ x \in \mathbb{R}^j} \mathcal{E}_q(L(x_0, x, \theta)) \]

Optimization with PSG functions:
\[ \min_{err, \ x_0 \in \mathbb{R}, \ x \in \mathbb{R}^j} err \]
\[ \text{s.t. constraints} \]
\[ err \geq -avg(L(x_0, x, \theta)) \]
\[
\text{where } K = \frac{1-q}{2q-1}.
\]

**Optimization Problem 2**

Expectile regression problem in two steps.

- **Step 1.** Find an optimal vector \( \mathbf{x}^* \) by minimizing deviation from Expectile Quadrangle (with piecewise linear error):

  \[
  \min_{\mathbf{x} \in \mathbb{R}^I} D_q(L_0(\mathbf{x}, \theta))
  \]

- **Step 2.** Calculate intercept \( x_0^* \)

  \[
  x_0^* = S_q(L_0(x^*, \theta))
  \]

Optimization with PSG functions:

\[
\min_{\mathbf{x} \in \mathbb{R}^I} \text{xvar}_{dev_q}(L_0(\mathbf{x}, \theta))
\]

calculate

\[
xvar_q(L_0(\mathbf{x}, \theta))
\]

**Optimization Problem 3**

Expectile regression by minimizing asymmetric variance Error.

Minimizing asymmetric variance error

\[
\min_{x_0 \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^I} E_q(L(x_0, \mathbf{x}, \theta))
\]

Optimization with PSG functions:

\[
\min_{x_0 \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^I} \left( q \cdot \text{pm2}_{pen_0}(L(x_0, \mathbf{x}, \theta)) + (1-q) \cdot \text{pm2}_{pen_g_0}(L(x_0, \mathbf{x}, \theta)) \right)
\]