Chapter 9: Fat-Tailed Regression Models

Peng Shi

Northern Illinois University

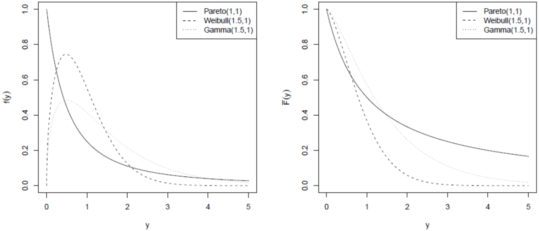
2013 Ratemaking and Product Management Seminar
March 11-13, 2013

Outline

- Introduction
- Regression Models
  - Transformation
  - Exponential Family
  - Parametric Regression
  - Quantile Regression

Introduction

- Claims data often presents fat-tails: auto, home, medical costs ...
- Fat-tailed Distribution
  - The frequency of extreme events is higher than that implied by the normal distribution
  - Alternative term: heavy-tailed or long-tailed



Introduction

- Tail measure

$$\lim_{y\to+\infty}\frac{\Pr(Y_1 > y)}{\Pr(Y_2 > y)} = \lim_{y\to+\infty}\frac{\bar{F}_{Y_1}(y)}{\bar{F}_{Y_2}(y)} = \lim_{y\to+\infty}\frac{f_{Y_1}(y)}{f_{Y_2}(y)}$$

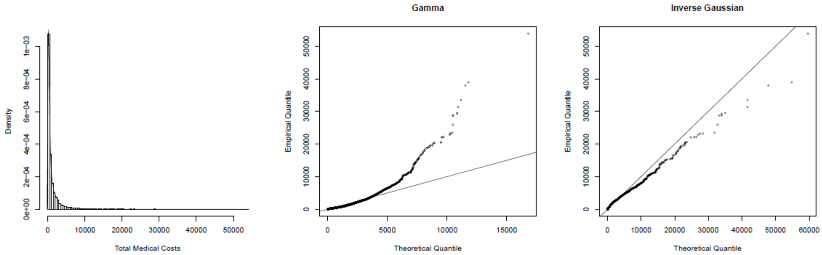
  - A limiting value of zero of the ratio indicates that the distribution of Y_2 has a heavier tail than that of Y_1

- Example

$$\lim_{y\to+\infty}\frac{\bar{F}_{Weibull}(y)}{\bar{F}_{Pareto}(y)} = \lim_{y\to+\infty}\frac{\exp(-(y/\lambda)^{\tau})}{\theta^{\alpha}(y+\theta)^{-\alpha}} = 0$$

Parametric Regression

- Two special cases:
  - Transformation: e.g. log
  - Exponential family: gamma
- Motivating example: Medical care expenditure
  - Look at individual medical care expenditure for office-based visit
  - Covariates include social economic charateristics, health status, employment etc.

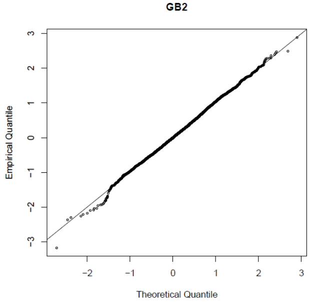


Parametric Regression

- General distributions
  - Consider transformation

$$\ln Y = \mu + \sigma \ln \frac{B(\phi_1, \phi_2)}{1 - B(\phi_1, \phi_2)}$$

  - Y is known to follow a GB2 distribution denoted by $GB2(\mu, \sigma, \phi_1, \phi_2)$
  - Many special cases, including GG and Burr XII



Multivariate Regression

- Examples
  - Claims of different type in auto insurance
  - Claims in multi-perils in homeowner insurance
  - Medical costs by physician and non-physician
- Copula regression
  - A copula is a multivariate joint distribution with all marginal follows uniform distribution on interval $[0, 1]$

$$C(u_1, \cdots, u_d) = \Pr(U_1 \leq u_1, \cdots, U_d \leq u_d)$$

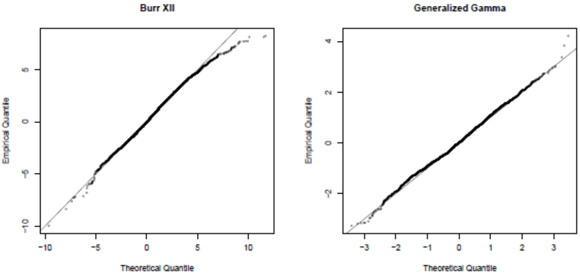
  - According to Sklar's theorem, for any multivariate distribution F with continuous marginals $F_1, \cdots, F_d$, there exists a copula C such that

$$F(y_1, \cdots, y_d) = C(F_1(y_1), \cdots, F_d(y_d))$$

  - Allows for flexible distributions for the marginal models while accommodates association

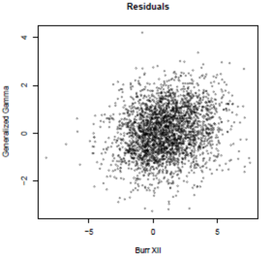
Medical Costs Example

- Again consider individual medical care costs of office-based visits
- Break down into two categories: doctor v.s. non-doctor
- Consider a bivariate copula regression
  - Burr XII for doctor and generalized gamma for non-doctor



Medical Costs Example

- Residuals



- Use Gaussian copula to join two marginals

$$f(y_1, y_2) = c(F_1(y_1), F_2(y_2))f_1(y_1)f_2(y_2)$$

- The estimated association parameter ρ is about 0.25

Quantile Regression

- Linear model

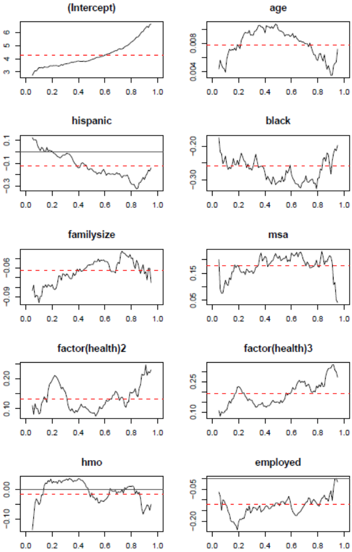
$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \sim F$$

- Consider the quantile function $Q_\tau(y|\mathbf{x}) = \mathbf{x}'\beta(\tau)$ with $0 < \tau < 1$, the regression coefficient $\beta(\tau)$ can be found by solving

$$\hat{\beta}(\tau) = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i'\boldsymbol{\beta})$$

  - $\rho_\tau(u) = u(\tau - I(u \leq 0))$ is known as the check function
- Use for skewed data and data with heteroscedasticity

Medical Costs Example: Revisit



Concluding Remarks

- See Volume I of the predictive modeling book for details
- More case studies in the Volume II