Longitudinal Modeling of Insurance Company Expenses

Peng Shi

University of Wisconsin-Madison

joint work with

Edward W. (Jed) Frees - University of Wisconsin-Madison

July 31, 2009

Outline

I. Introduction: Motivation and Objective

II. Data Description

III. Longitudinal Quantile Regression Model

IV. Copula Model Inference: Rescaling and Transformation

V. Model Validation

VI. Concluding Remarks

- Expenses by Type
  - Underwriting expenses: policy acquisition cost, administrative expenses
  - Investment expenses: trading activities, portfolio management
  - Loss adjustment expenses: investigation cost, legal fees
- Benefits of Expense Analysis
  - Insurers: rate making, cost control, strategic decision
  - Investors: cost efficiency and profitability analysis
  - Regulators: expense factor, industry benchmark, economic hypothesis
- Limitations of Current Practice
  - Ignored three features of insurance company expenses: skewness, negative values and intertemporal dependence

- Expenses by Type
  - Underwriting expenses: policy acquisition cost, administrative expenses
  - Investment expenses: trading activities, portfolio management
  - Loss adjustment expenses: investigation cost, legal fees
- Benefits of Expense Analysis
  - Insurers: rate making, cost control, strategic decision
  - Investors: cost efficiency and profitability analysis
  - Regulators: expense factor, industry benchmark, economic hypothesis
- Limitations of Current Practice
  - Ignored three features of insurance company expenses: skewness, negative values and intertemporal dependence

- Expenses by Type
  - Underwriting expenses: policy acquisition cost, administrative expenses
  - Investment expenses: trading activities, portfolio management
  - Loss adjustment expenses: investigation cost, legal fees
- Benefits of Expense Analysis
  - Insurers: rate making, cost control, strategic decision
  - Investors: cost efficiency and profitability analysis
  - Regulators: expense factor, industry benchmark, economic hypothesis
- Limitations of Current Practice
  - Ignored three features of insurance company expenses: skewness, negative values and intertemporal dependence

Objective

GOAL:

To develop longitudinal models that can be used for prediction, to identify unusual behavior, and to eventually measure firm inefficiency, by addressing above three features.

- Statistical Viewpoint
  - Basic regression set-up - almost every analyst is familiar with
    - It is part of the basic actuarial education curriculum
  - Incorporating cross-sectional and time patterns is the subject of longitudinal data analysis - a widely available statistical methodology
  - Quantile regression focuses on the quantiles of response variable - a relatively new regression technique

GOAL:

To develop longitudinal models that can be used for prediction, to identify unusual behavior, and to eventually measure firm inefficiency, by addressing above three features.

- Statistical Viewpoint
  - Basic regression set-up - almost every analyst is familiar with
    - It is part of the basic actuarial education curriculum
  - Incorporating cross-sectional and time patterns is the subject of longitudinal data analysis - a widely available statistical methodology
  - Quantile regression focuses on the quantiles of response variable - a relatively new regression technique

Sampling Procedure

- Firm level data of property-casualty insurers from NAIC
- Observe from 2001 to 2006
- Two types of observations are removed:
  (1) Companies with non-positive net premiums written in all years
  (2) Records with inactive company status in the last observation year
- Final sample consists of 2,660 companies and 13,925 observations
- Variables in money values are deflated to 2001 US dollars

Table 1. Descriptive Statistics of Total Expenses ($1,000,000)

	2001	2002	2003	2004	2005	2006
Number	2,286	2,269	2,303	2,320	2,354	2,393
Mean	57.01	61.25	64.47	65.37	64.06	63.66
Median	6.00	6.22	6.01	5.99	5.83	6.12
StdDev	332.03	353.63	364.09	359.50	354.74	352.29
Minimum	-190.46	-38.16	-32.63	-26.28	-111.08	-20.42
Maximum	10,410.17	11,307.32	10,966.07	10,397.76	9,809.33	10,051.16
Percentage of Negative Obs	5.86%	6.30%	6.34%	5.56%	6.07%	5.56%

- Total expenses are skewed and heavy-tailed distributed
- Lack of balance
- Negative expenses: (1) Adjustment for prior reporting year (2) Reinsurance arrangement
- Strong serial correlation and individual effects

Literatures on Long-tail Longitudinal Models

- Two techniques to handle skewed and long-tailed data
  - Transformation, see Carroll and Ruppert (1988)
  - Parametric regression
    - Generalized linear model (GLM), see Haberman and Renshaw (1996), Parametric survival model, see Lawless (2003) and GB2 regression, see Sun et al. (2008), Frees and Valdez (2008), Frees et al. (2008)
    - Random effects are use to account for heterogeneity and serial correlation
- Quantile Regression
  - First introduced by Koenker and Bassett Jr (1978)
  - Advantages in long-tail regression modeling include easier interpretation, higher efficiency and robustness to outliers
  - Longitudinal Quantile Regression
    - Jung (1996): quasi-likelihood method
    - Lipsitz et al. (1997): weighted generalized estimating equations
    - Koenker (2004): regularization method
    - Geraci and Bottai (2007): asymmetric Laplace density

Literatures on Long-tail Longitudinal Models

- Two techniques to handle skewed and long-tailed data
  - Transformation, see Carroll and Ruppert (1988)
  - Parametric regression
    - Generalized linear model (GLM), see Haberman and Renshaw (1996), Parametric survival model, see Lawless (2003) and GB2 regression, see Sun et al. (2008), Frees and Valdez (2008), Frees et al. (2008)
    - Random effects are use to account for heterogeneity and serial correlation
- Quantile Regression
  - First introduced by Koenker and Bassett Jr (1978)
  - Advantages in long-tail regression modeling include easier interpretation, higher efficiency and robustness to outliers
  - Longitudinal Quantile Regression
    - Jung (1996): quasi-likelihood method
    - Lipsitz et al. (1997): weighted generalized estimating equations
    - Koenker (2004): regularization method
    - Geraci and Bottai (2007): asymmetric Laplace density

Quantile Regression

The regression quantiles $\beta(\tau)$ in the τth conditional quantile function $Q_\tau(y|\mathbf{x}) = \mathbf{x}'\beta(\tau)$ can be estimated by solving

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

Also, $\rho_\tau(u) = u(\tau - \mathrm{I}(u \le 0))$ is check function and $\mathrm{I}(\cdot)$ is the indicator.

Asymmetric Laplace Distribution

$$f(y;\mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma}\exp(-\frac{y-\mu}{\sigma}[\tau - \mathrm{I}(y \le \mu)])$$

- Defined on $(-\infty, +\infty)$
- Location μ, scale σ, skewness τ (Yu and Zhang (2005))
- Under $\mu = \mathrm{x}^{'}\beta$, the MLE with $\mathrm{ALD}(\mu,\sigma,\tau)$ assumption results in regression quantiles (Yu et al. (2003))
- $\mathrm{E}(y|x) = \mu(x) + \frac{\sigma(1-2\tau)}{\tau(1-\tau)}$

Quantile Regression

The regression quantiles $\beta(\tau)$ in the τth conditional quantile function $Q_\tau(y|\mathbf{x}) = \mathbf{x}'\beta(\tau)$ can be estimated by solving

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

Also, $\rho_\tau(u) = u(\tau - \mathrm{I}(u \le 0))$ is check function and $\mathrm{I}(\cdot)$ is the indicator.

Asymmetric Laplace Distribution

$$f(y;\mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left(-\frac{y-\mu}{\sigma}[\tau - \mathrm{I}(y \le \mu)]\right)$$

- Defined on $(-\infty, +\infty)$
- Location μ, scale σ, skewness τ (Yu and Zhang (2005))
- Under $\mu = \mathbf{x}'\beta$, the MLE with $\mathrm{ALD}(\mu,\sigma,\tau)$ assumption results in regression quantiles (Yu et al. (2003))
- $\mathrm{E}(y|\boldsymbol{x}) = \mu(\boldsymbol{x}) + \frac{\sigma(1-2\tau)}{\tau(1-\tau)}$

- Use ALD(μ, σ, τ) for marginals
- Use copula function to model intertemporal dependence

$$f_i(y_{i1}, \ldots, y_{iT_i}) = c(F_{i1}, \ldots, F_{iT_i}; \phi) \prod_{t=1}^{T_i} f_{it}$$

- Parameterize $\mu_{it} = \mathbf{x}_{it}^{'}\beta$ in $\mathrm{ALD}(\mu, \sigma, \tau)$, then the log-likelihood function for ith insurer is shown as

$$l_i = \ln \frac{\tau(1-\tau)}{\sigma} - \frac{1}{\sigma} \sum_{t=1}^{T_i} \rho_\tau(y_{it} - \mathbf{x}_{it}^{'}\beta) + \ln c(F_{i1}, \ldots, F_{iT_i}; \phi)$$

- Quantile regression are preserved for marginals and we are interested in the τ of best fit

Approach I: Rescaling

$$Y_{it} = \frac{\text{Total Expenses}_{it}}{\text{Total Assets}_{it}}.$$

- allows one to compare different sized firm
- requires prediction of total assets

Approach II: Transformation

- Idea: transform data to ALD
- Normality-improving and variance-stabilizing (Pierce and Shafer (1986))
- To create new distributions (Bali (2003), Bali and Theodossiou (2008))
- We consider modulus transformation (John and Draper (1980)), IHS (Burbidge and Magee (1988)), modified modulus transformation (Yeo and Johnson (2000))

Model Extension

Approach I: Rescaling

$$Y_{it} = \frac{\text{Total Expenses}_{it}}{\text{Total Assets}_{it}}.$$

- allows one to compare different sized firm
- requires prediction of total assets

Approach II: Transformation

- Idea: transform data to ALD
- Normality-improving and variance-stabilizing (Pierce and Shafer (1986))
- To create new distributions (Bali (2003), Bali and Theodossiou (2008))
- We consider modulus transformation (John and Draper (1980)), IHS (Burbidge and Magee (1988)), modified modulus transformation (Yeo and Johnson (2000))

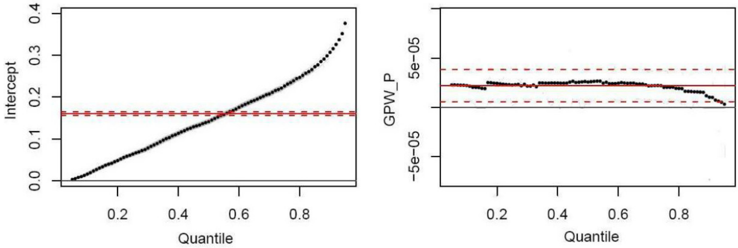
Table 3. Description of Covariates

Covariate	Description
GPW_P	Gross premium written of personal lines
GPW_C	Gross premium written of commercial lines
IRATIO	Cash and invested assets (net admitted)
LOSS_L	Losses incurred for long tail line of business
LOSS_S	Losses incurred for short tail lines of business
ASSET_CURR	Net admitted assets in current year
STOCK	Indicates if the company is a stock company
MUTUAL	Indicates if the company is a mutual company
GROUP	Indicates if the company is affiliated or unaffiliated company

Table 3. Description of Covariates

Covariate	Description
GPW_P	Gross premium written of personal lines
GPW_C	Gross premium written of commercial lines
IRATIO	Cash and invested assets (net admitted)
LOSS_L	Losses incurred for long tail line of business
LOSS_S	Losses incurred for short tail lines of business
ASSET_CURR	Net admitted assets in current year
STOCK	Indicates if the company is a stock company
MUTUAL	Indicates if the company is a mutual company
GROUP	Indicates if the company is affiliated or unaffiliated company

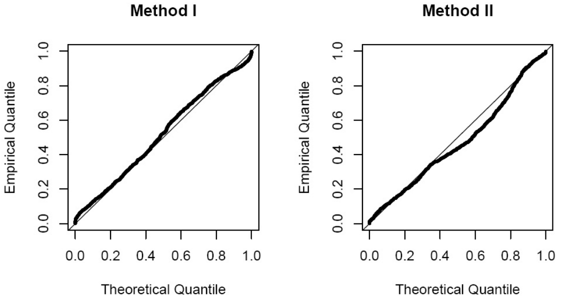
- Regression quantiles for intercept and GPW_P



- Out-of-sample validation is based on the predictive density:

$$f(y_{i,T+1}|y_{i1},\ldots,y_{iT}) = \frac{c(F_{i1}(y_{i1}),\ldots,F_{i,T+1}(y_{i,T+1}))}{c(F_{i1}(y_{i1}),\ldots,F_{i,T}(y_{i,T}))} f_{i,T+1}(y_{i,T+1})$$

- Calculate the percentile of y_{i2006} by $p_i = F(y_{i2006})$ for $i = 1,\ldots,n_h$, where $F(\cdot)$ is the cdf of the predictive distribution
- p_i should be uniform if the model is well specified



Cost Efficiency Validation

Idea

A residual is the company expense, controlled for company characteristics. A small residual means an inexpensive company. We look into residuals to identify cost efficient companies.

- We have no external measures to validate our notions of an "inexpensive" company but can look to A. M. Best Ratings
  - Ratings indicate the financial strength of an insurer
  - Not the same as the expense situation for a company
  - Still, a less expensive insurer tends to be more profitable, and thus has a healthier financial status and higher rating

Idea

A residual is the company expense, controlled for company characteristics. A small residual means an inexpensive company. We look into residuals to identify cost efficient companies.

- We have no external measures to validate our notions of an “inexpensive" company but can look to A. M. Best Ratings
  - Ratings indicate the financial strength of an insurer
  - Not the same as the expense situation for a company
  - Still, a less expensive insurer tends to be more profitable, and thus has a healthier financial status and higher rating

Cost Efficiency Validation

- The average residuals over 2001-2005 are employed in the analysis
- Define the residual percentile as the ratio of the rank of an residual to the number of insurers
- A financially strong company will have low expenses, meaning that the percentiles of the distribution of expenses are small
- Counts of Insurers with Secure Rating

Residual	Superior		Excellent		Good	
Percentile	Copula	RE	Copula	RE	Copula	RE
0-10	51	42	83	56	20	2
10-25	78	69	126	117	51	26
25-50	119	125	197	210	88	98
$>$50	96	108	459	482	87	120
Totals	344	344	865	865	246	246

- The copula model outperforms the random effects model in classifying more insurers into higher efficiency range (top 50th percentile) for all categories of secure rating

Conclusion

- Model features:
  - Introduces a quantile regression model for longitudinal data
  - Captures heavy tailed nature of insurance company expenses
  - Allows for negative values of expenses
  - Captures intertemporal dependence of expenses through a copula function
  - Allows for covariates for expenses
  - Provides a predictive distribution for insurer's expenses
- Future work:
  - Will look at each type of expenses
  - Will examine the efficiency of insurers using more formal “stochastic frontier" models

Transformation Method

- Generic form $Y^{(\lambda)} = \psi(Y, \lambda)$
- Three transformations
  - Modulus

$$y^{(\lambda)} = \begin{cases} \mathrm{sign(y)}\left\{(|y|+1)^{\lambda} - 1\right\}/\lambda, & \lambda \neq 0 \\ \mathrm{sign(y)}\log(|y|+1), & \lambda = 0 \end{cases}$$

  - IHS

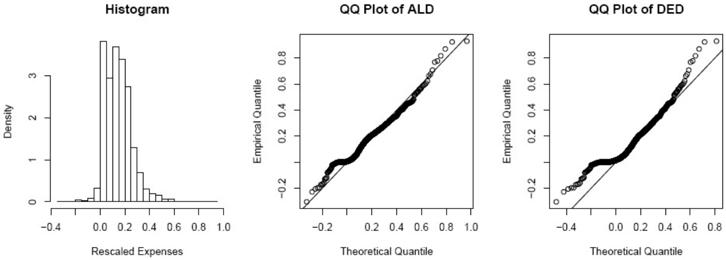
$$\begin{aligned} y^{(\lambda)} \quad &= \sinh^{-1}(\lambda y)/\lambda \\ &= \ln(\lambda y + (\lambda^2 y^2 + 1)^{1/2})^{(1/\lambda)} \end{aligned}$$

  - Modified Modulus

$$y^{(\lambda)} = \begin{cases} \{(y+1)^{\lambda} - 1\}/\lambda, & y \geq 0, \lambda \neq 0 \\ \log(y+1) & y \geq 0, \lambda = 0 \\ -\{(-y+1)^{2-\lambda} - 1\}/(2-\lambda), & y < 0, \lambda \neq 2 \\ -\log(-y+1) & y < 0, \lambda = 2 \end{cases}$$

Analysis of Rescaled Expenses

- Marginal distribution



- Intertemporal dependence (Define $\hat{\varepsilon}_{it} = (y_{it} - \hat{\mu}_{it})/\hat{\sigma}$, $\hat{\mu}_{it} = \mathbf{x}_{it}^{'}\hat{\beta}$)

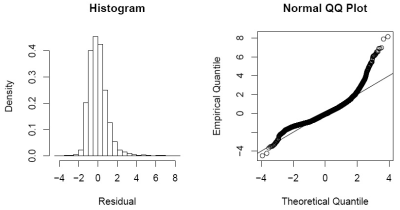
	2001	2002	2003	2004	2005	2006
2001	1.000					
2002	0.857	1.000				
2003	0.774	0.852	1.000			
2004	0.686	0.754	0.823	1.000		
2005	0.642	0.692	0.740	0.824	1.000	
2006	0.625	0.667	0.691	0.759	0.844	1.000

Analysis of Rescaled Expenses

- Estimates for the Longitudinal Quantile Regression Model with Different Copulas

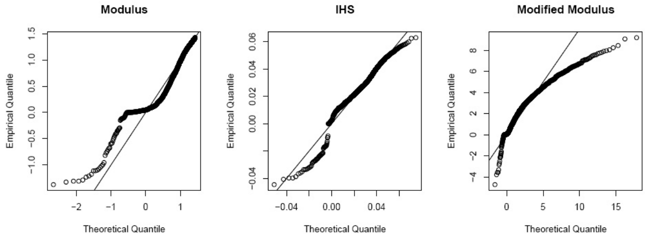
	Gaussian		Student		RE	
	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat
SIGMA	0.0282	54.62	0.0281	222.50		
TAU	0.2130	54.87	0.2129	63.47		
BINT	0.0983	245.22	0.0984	242.41	0.1843	23.42
BLOSS_L	0.0287	5.99	0.0228	5.63	0.0389	3.51
BLOSS_S	0.0255	11.23	0.0169	7.29	0.0201	3.31
BPREM_P	0.0122	5.98	0.0168	5.57	0.0158	2.74
BPREM_C	0.0092	8.11	0.0095	7.15	0.0061	1.86
BASSET_CURR	-0.0057	-7.77	-0.0043	-6.22	-0.0108	-7.31
BIRATIO	-0.0639	-83.12	-0.0638	-162.86	-0.0423	-6.05
BGROUP	-0.0226	-7.54	-0.0225	-60.65	-0.0369	-10.21
BSTOCK	0.0200	6.91	0.0201	20.19	0.0276	4.54
BMUTUAL	0.0706	171.84	0.0705	24.39	0.0487	6.95
RHO1	0.8371	205.71	0.8390	134.13	0.7807	
RHO2	0.7405	106.76	0.7561	152.60	0.6685	
RHO3	0.6643	66.42	0.6893	87.63	0.5723	
RHO4	0.5979	42.09	0.6416	47.14	0.4953	
TDF			6.3453	31.34		
LogLikelihood	14,090.68		15,741.96		13,150.55	
AIC	-28,149.37		-31,449.92		-26,269.10	

- Histogram and QQ plot of residuals of random effect model

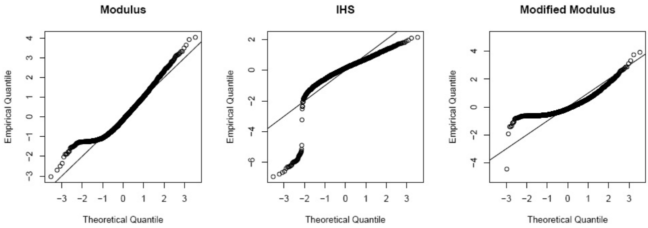


Analysis of Transformed Expenses

- QQ plots of transformed asymmetric Laplace distributions



- QQ plots of transformed normal distributions



- Estimates for t-Copula Models with Different Dependence Structure

	AR1		Exchangable		Toeplitz		Unstructured	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
SIGMA	0.0028	78.03	0.0019	123.08	0.0029	50.31	0.0029	78.77
TAU	0.6777	117.20	0.8268	400.44	0.6776	83.27	0.6796	161.63
LAMBDA	245.4257	704.66	244.0306	675.03	246.5235	268.81	244.7425	662.24
BINT	0.0299	332.27	0.0312	307.94	0.0302	170.85	0.0304	391.70
BLOSS_L	0.0006	0.91	0.0007	1.25	0.0010	1.55	0.0009	1.38
BLOSS_S	0.0028	5.84	0.0078	13.60	0.0025	5.55	0.0027	5.89
BPREM_P	0.0036	7.33	0.0023	5.72	0.0034	7.01	0.0033	7.08
BPREM_C	0.0039	14.38	0.0042	17.97	0.0040	14.30	0.0041	15.20
BASSET_CURR	0.0009	7.42	0.0017	16.66	0.0008	6.61	0.0007	6.66
BIRATIO	-0.0011	-14.79	-0.0012	-8.24	-0.0013	-8.99	-0.0015	-17.05
BGROUP	0.0018	9.75	0.0013	7.46	0.0017	26.29	0.0016	21.46
BSTOCK	0.0020	12.67	0.0035	35.18	0.0021	18.07	0.0019	27.19
BMUTUAL	0.0022	15.66	0.0032	11.13	0.0020	4.82	0.0021	18.46
RHO12	0.8629	188.23	0.7220	60.08	0.8624	306.83	0.8876	106.01
RHO13					0.7715	31.51	0.7869	76.74
RHO14					0.6952	58.12	0.7019	58.46
RHO15					0.6336	36.79	0.6313	50.80
RHO23							0.8743	131.31
RHO24							0.7769	39.15
RHO25							0.6901	57.13
RHO34							0.8565	66.49
RHO35							0.7543	82.22
RHO45							0.8314	61.12
TDF	1.3890	292.91	1.3814	433.19	1.3796	290.90	1.3889	229.88
Loglikelihood	47,625.04		47,256.25		47,648.23		47,662.17	
AIC	-95,220.07		-94,482.50		-95,260.46		-95,276.33	