A Bayesian Semiparametric Quantile Regression Model for Longitudinal Data with Application to Insurance Company Costs

Karthik Sriram, Peng Shi and Pulak Ghosh

December 7, 2011

Karthik Sriram is a Phd student in the department of Quantitative Methods and Information Systems at Indian Institute of Management, Bangalore,Karnataka, India(Email: karthik.sriram09@iimb.ernet.in). Peng Shi is an Assistant Professor in Division of Statistics, Northern Illinois University, DeKalb, IL, USA(Email: pshi@niu.edu). Pulak Ghosh is a Professor in the department of Quantitative Methods and Information Systems at Indian Institute of Management, Bangalore, Karnataka, India(Email: pulak.ghosh@iimb.ernet.in).

Abstract

This article examines the average cost function for property and casualty insurers. Cost function describes the relationship between a firm's minimum production cost and outputs. The comparison of cost functions could shed light on the relative cost efficiency of individual firms, which is of interest to many market participants and has been given extensive attention in the insurance industry. To identify and compare the cost function of insurers, common practice is to assume an identical functional form between cost and outputs and to rank insurers according to the center (usually mean) of the cost distribution. Such assumption could be misleading because insurers tend to adopt different technologies that are reflected by the cost function in their production process. We find that the cost distribution is skewed with a heavy tail. Therefore, the center-based comparison could lead to biased inference regarding an insurer's efficiency in operation. To address these issues, we propose Bayesian semiparametric quantile regression approach to model the longitudinal data on production cost of insurance companies. Particularly we formulate the quantile regression using the asymmetric Laplace distribution; the effects of various firm characteristics on the quantiles of cost are specified as a linear function; the firm-specific relation between cost and multiple outputs is captured by a single-index function based on a spline basis where the coefficients are assumed to have a Dirichlet process prior. The single-index formulation with splines renders flexibility in modeling the nonlinear cost-output relation and the use of Dirichlet process leads to a natural clustering of insurers with similar cost efficiency. The method is applied to data on US property casualty insurers from the National Association of Insurance Commissioners (NAIC). The analysis of average cost at different quantiles indicates that better insights on efficiency are gained by comparing the whole cost distribution. A comparison of the model results with an external financial strength ratings for property casualty insurers provides interesting insights on which part of the cost distribution are perhaps weighted more by the rating agency.

Key words: Bayesian quantile regression; Asymmetric Laplace distribution; Single-index ; Dirichlet process; Spline; Clustering; Longitudinal data

1 Introduction

The insurance industry is one of the more important economic sectors in well-developed nations. The global insurance premiums totaled $4.27 trillion US dollars in 2008 with 58% from life insurance and 42% from property-casualty (nonlife) insurance. As the world's largest insurance market, the US insurance industry underwrites $1.24 trillion US dollars, among which, life and nonlife insurance contribute $0.58 and $0.66 trillion US dollars, respectively (I.I.I. Insurance Fact Book 2009). A well-functioning insurance industry is thus critical to the stable growth of any developed economy.

In a competitive insurance market, a profitable and financially healthy insurer is built on a cost-efficient operation. A fair indicator of an insurer's financial status is a letter-based rating issued by independent financial rating agencies. Such rating reflects to a great extent an insurer's cost efficiency. For example, the financial strength rating of A.M. Best Company (from A++ to D) offers the largest coverage of insurers and reinsurers in the United States.

In this paper, we wish to analyze the cost efficiency and thus determine the financial strength of companies in the US property casualty insurance industry. Our goals are three fold. First, to identify the cost curve of each insurer (i.e. cost as a function of produced output) and show that insurers choose to operate at different positions of their own cost curves. Second, to compare individual insurers according to the level of efficiency in operation and validate results with independent financial strength ratings. Third, to identify groups of companies with similar efficiency level. To achieve these goals, we examine the average cost curve for property and casualty insurers. We define a more cost-efficient operation as one with lower average cost, i.e. the cost per unit of output produced.

In microeconomics, cost function describes the relationship between a firm's minimum production cost and outputs. A comparison of cost functions of individual firms is very important due to several reasons. First of all, a better understanding of cost-output relationship helps achieve economies of scale, improve profitability, and align a company's financial and operational plans. Secondly, insurance is a competitive busi-

ness, thus a cost-efficient operation is the key to sustainable growth. Therefore, such a comparison would help investors in the evaluation of operation efficiency, to identify more profitable insurers, and to make better-informed investment decisions. Another important aspect of efficiency analysis is to provide guidance to regulators and policy makers regarding the problems and development in the industry or the economy. For example, in the past two decades, the US property casualty insurance industry has experienced a wave of mergers and acquisitions, partly motivated by the adoption of regulatory risk-based capital standards. As insurers enjoy the risk diversification through consolidations, it is interesting to regulatory authorities whether consolidation is beneficial or detrimental to the scale efficiency of insurers. Not surprisingly, Berger and Humphrey (1997) identified more than 130 articles on cost efficiency studies for banking and insurance industries. In less than ten years after Cummins and Weiss (2000), where 21 studies for the insurance industry are surveyed, anther 95 studies are identified for the insurance industry alone (Eling and Luhnen (2010)).

To identify and compare the cost function of insurers, usual practice is to assume an identical functional form between cost and outputs and to rank insurers according to the center(usually mean) of the cost distribution. Such assumption could be inappropriate because insurers tend to adopt different technologies in their production process, suggesting a variable cost-output relation across firms. Usually cost distributions are skewed with long tails. For example, in our motivating data that is analyzed later, we find that insurer's cost distribution is highly skewed and heavy tailed (see Figure 1). Therefore, a center-based comparison can lead to biased inference regarding an insurer's efficiency in operation. Contrary to the usual practice, in this work we look into insurers' average cost rather than total production costs. According to microeconomic theory, a company's average cost, as a function of its output, is a U-shaped curve. Each firm chooses the optimal output level to maximize its profit. As a result, one insurer might end up with an output level with economies of scale (the decreasing section of the U-shaped curve), while another insurer might end up with an output level with diseconomies of scale (the increasing section of the U-shaped curve). Thus, by investigating average cost, one obtains additional insights regarding returns to scale

in its production process without sacrificing the capability of ranking an insurer's cost efficiency. Comparisons of the average cost are not straightforward because of the complicated production decision process. To identify the unique average cost curve of each insurer, one needs to observe an insurer repeatedly over time. To further compare two distinct cost curves, ideally one should examine the whole cost distributions rather than focusing on the center. Motivated by such observations, we propose Bayesian semiparametric quantile regression to model the data on production cost of insurance companies over time. In our application, the method is used to derive the distributions of average cost and to compare the cost efficiency among insurers. As a highlight of our model, the quantile regression is formulated using the asymmetric Laplace distribution(ALD); the effects of various firm characteristics on the quantiles of cost are specified as a linear function; the firm-specific relation between cost and multiple outputs is captured by a single-index function (e.g. Hardle et al. (1993),Ichimura (1993), Yu and Ruppert (2002)) with a Dirichlet process prior (Ferguson (1973)). In the following, we briefly introduce the two key concepts, namely single-index formulation and longitudinal quantile regression method, that are necessary to develop our proposed model.

1.1 Single-index Formulation and Economies of Scale

One challenge in our study is the analysis of economies (diseconomies) of scale for insurers, i.e., to determine whether an insurer operates on the increasing or decreasing side of it's average cost curve. In microeconomics, the economies of scale for a single-output firm could be simply measured by the elasticity of average cost with respect to that output (in a linear regression model, this would be the regression coefficient of output). However, insurance companies produce multiple intangible outputs. According to Cummins and Weiss (2000), property-casualty insurers provide two principal services: risk transfer and financial intermediation. Hence, two outputs related to these types of services are typically used in efficiency studies. What has been ignored in the multiple-output case is the combined effect of the two outputs and their possible nonlinear effects on the production cost. In our application, we use a single-index

formulation in order to arrive at a single measure for output based on both types of services. In doing so, we measure their combined effect on average cost and capture the potential nonlinear relationship.

In a single-index formulation, the parameter of interest (θ) is modeled as a function of multiple covariates (X) as $\theta = g(h_\beta(X))$, where the function $h_{\boldsymbol{\beta}}(\cdot)$ is known up to the finite-dimensional parameters $\boldsymbol{\beta}$ and the "link" function g is unknown. On one hand, this formulation provides much desired flexibility over a known parametric formulation. On the other hand, it does not suffer from the curse of dimensionality that arises from a complete nonparametric formulation, since the argument of function $g(\cdot)$ is still univariate. See Hardle et al. (1993), Ichimura (1993), Yu and Ruppert (2002) for some key developments in the analysis of single-index models and Wu et al. (2010) for it's application in quantile regression. Bayesian analysis of single-index models has also received much attention recently (e.g. Antoniadis et al. (2004) and Wang (2009)).

In our model, the parameter of interest θ is the τ^{th} quantile of the distribution of average cost of an insurer, which is modeled as follows.

$$\begin{aligned} &h_\beta(\mathbf{X}, \text{risk transfer, financial intermediation}) \\ &= \mathbf{X}^T\boldsymbol{\beta} + g(\text{a} \times \text{risk transfer} + \text{b} \times \text{financial intermediation}) \end{aligned}$$

The covariates $\mathbf{X}$ represent various company characteristics such as size, product mix, and investment portfolio etc, that are modeled linearly in the first term. The dependence on outputs namely, risk transfer and financial intermediation is modeled by the second term. The function $g(\cdot)$ can be interpreted as the τ^{th} quantile of the average cost after controlling for firm characteristics. Since we are interested in the combined effects of multiple outputs, we define insurer's output as a linear combination of the individual outputs ($i.e., \kappa_1 \times$ risk transfer $+ \kappa_2 \times$ financial intermediation) and its effect on average cost will be captured by function $g(\cdot)$.

1.2 Longitudinal quantile regression and clustering

Our cost analysis resorts to quantile regression because conditional quantiles provide a more complete picture of the cost distribution. Quantile regression was introduced

by Koenker and Bassett (1978), and has been investigated for cross sectional and time series data, for example, see Koenker and Xiao (2002), Koenker and Xiao (2004), and Koenker and Xiao (2006). Although there have been recent studies on the use of quantile regression in longitudinal data (see Jung (1996), Lipsitz et al. (1997), Koenker (2004) among others), research into the Bayesian approach for the same is relatively new and limited. There are however a small yet growing number of studies (see Geraci and Bottai (2007), Yuan and Yin (2010) and Yue and Rue (2011)) where Bayesian quantile regression has been extended in a few directions including nonparametric setting.

In this article, we adopt a new Bayesian hierarchical modeling approach to analyze cost data from property casualty insurers that address the above mentioned issues. The key contributions of our work are:

1. We perform a quantile regression analysis for the longitudinal data on costs of property casualty insurers, which allows for the comparison of the whole cost distribution. In fact, we show that better insights on efficiency are gained by examining different (conditional) quantiles of average cost.

2. We identify firm-specific cost function via a single-index formulation on the multiple outputs of insurers. Such a formulation renders flexibility in modeling the nonlinear relationship of cost with multiple outputs and enables us to determine on which part of the average cost curve an insurer operates

3. We allow the single-index function to vary by firm with a Dirichlet process prior on its coefficients. While the longitudinal data helps capture the relationship between average cost and output, the prior on the coefficients provides the necessary shrinkage on the large number of parameters needed to model the relationship at firm level.

4. Moreover, as a consequence of using Dirichlet process, there is a borrowing of strength within the coefficients across firms, since the Dirichlet process model results in automatic clustering of companies with similar cost curves. The clustering of insurers is interesting in quantile regression context, because potentially

different clustering results can be obtained at various quantiles.

5. On the one hand, we use the letter-based financial strength ratings assigned by an independent rating company to validate our results. On the other hand, our analysis sheds light on which part of the cost distribution (left tail, middle, or right tail) is weighted more heavily by rating agencies.

It is also important to note that given the objective of estimating firm-level cost function and the fact that the data is limited to a few years, a purely classical formulation of the model may have estimation problems and thus a Bayesian approach may be more useful.

The rest of the article is structured as follows: Section 2 introduces the data that motivated the work. In Section 3, we propose the Bayesian quantile regression model with a subject-specific single-index formulation and in Section 4, the Bayesian inference. In Section 5, we analyze data and show the benefit of the proposed approach, develop how to cluster insurers and compare cost efficiency. We conclude in Section 6. Technical notes on simulating from posterior distribution of the various model parameters are provided in the appendix.

2 Motivating Data

We consider a dataset of US property-casualty insurers from the National Association of Insurance Commissioners (NAIC) database. The NAIC, an organization of insurance regulators, maintains one of the largest insurance regulatory database in the world. The database contains financial statements of thousands of insurers writing business in US. Further details of this database is available in Shi and Frees (2010).

The definition and selection of variables are all followed from the insurance literature. The average cost of insurers represents the cost per unit of output. As pointed out earlier, a property casualty insurer produces two types of output: "risk transfer", which protects a personor businesses from property losses and legal liabilities, and "financial intermediation", which compensates for the opportunity cost of the funds held by the insurer through a discount in premiums. The production process is accom-

plished by property-casualty insurers through two principal functions, underwriting and investment, associated with three types of cost: underwriting, investment, and loss adjustment expenses. Underwriting expenses aggregate policy acquisition and maintenance costs; Investment expenses are associated with the portfolio management of an insurer's invested assets; Loss adjustment expenses result from loss investigation and claim settlement. Using gross premiums written and invested assets to measure the quantities of "risk transfer" and "financial intermediation", respectively, we define the average cost for a property- casualty insurer as:

$$\text{Average cost} = \frac{Underwriting\ expenses + Loss\ adjustment\ expenses}{Gross\ premiums\ written} + \frac{Investment\ expenses}{Invested\ assets}$$

In addition to average cost and output variables, we control important firm characteristics in the quantile regression: business mix is captured by the distribution in different lines of business; Investment portfolio is described by the allocation of premiums funds into various types of financial assets; Ownership structure is categorized as stock verses mutual insurer; Firm affiliation is differentiated by whether an insurer is a single entity or belongs to a group. Table 1 summarizes these variables along with their descriptions.

We use firm-level observations for active primary insurers from years 2001-2006. To construct our modeling dataset, we exclude from the data: 1) Companies with non-positive net written premiums in all observation years. These companies are not underwriting business, either because they have stopped issuing new policies or because they are doing other types of business (for example investment) under a shell. 2) Records with an inactive company status. The NAIC records an inactive status when the company is merged with or acquired by another company, the company is voluntarily out of business, or the charter is inactive. 3) Certain specialty insurers, such as financial guaranty and title insurers, insurers that do not file statements with the NAIC, as well as professional reinsures classified by the NAIC.

We analyze data for individual companies, as opposed to groups of affiliated insurers. The above criterion produces a dataset of 9,362 company-year observations for

1,741 property casualty insurers over a six-year time horizon. In the following analysis, all variables expressed in monetary values are deflated to 2001 dollars using the consumer price index. The descriptive statistics are presented in Table 2.

In this work, our dependent variable is avg_cost (i.e. the cost per unit of output produced). The histogram of this dependent variable is displayed in Figure 1. It is evident that the distribution is right skewed and presents long tails, suggesting that focusing on the center is not sufficient for a comprehensive description of an insurer's cost distribution. Such observation motivates the use of quantile regression, where a more complete picture of cost distribution is captured by conditional quantiles. Also, quantile regression analysis helps gain much insights into the relationship of average cost with different firm characteristics. To account for the variation across the data, we control for different firm characteristics as well as an overall macro effect of time that impacts all companies. As will be seen later in Figure 3, the strength of relationship can be different for different variables, and moreover for the same variable, can vary across quantiles. A mean regression approach in such a situation could give an incomplete picture and some times mislead us into concluding that no relationship exists.

3 Methodology

3.1 Bayesian quantile regression model

Given dependent variables Y_i $(i = 1, 2, ..., N)$ and explanatory variables $\mathbf{X}_i$, the quantile regression problem (see Koenker and Bassett (1978), Koenker (2005)) involves solving for $\boldsymbol{\beta}$ in the following problem.

$$\min_{\beta} \sum_{i=1}^{N} \rho_\tau(Y_i - \mathbf{X}_i^T \boldsymbol{\beta})$$

where $\rho_\tau(u) = u(\tau - I_{(u \leq 0)})$ with $I_{(\cdot)}$ being the indicator function and $0 < \tau < 1$. This is equivalent to the maximum likelihood estimation problem by assuming asymmetric Laplace distribution(ALD) for the response, i.e. $Y_i \sim ALD(., \mu_i^\tau, \sigma, \tau)$, where $\mu_i^\tau =$

$X_i^T\boldsymbol{\beta}$ and

$$ALD(y;\mu^\tau,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma}\exp\left\{-(y-\mu^\tau)/\sigma \times (\tau - I_{(y\leq\mu^\tau)})\right\},\ -\infty < y < \infty$$

The τ^{th} quantile of this distribution happens to be μ^τ. In light of this fact, Yu and Moyeed (2001) introduced Bayesian methods in quantile regression by formulating the problem as a generalized linear model using ALD for the response. Based on empirical findings, they observe that using ALD for response is robust to the true underlying likelihood. Sriram et al. (2011)(unpublished manuscript) explored a theoretical justification for using ALD in Bayesian quantile regression. An alternative to ALD is using Bayesian nonparametric methods which allow for relaxing the distributional assumption (see e.g. Reich et al. (2010), Reich et al. (2011)). However, given the robustness of ALD to the true underlying distribution and ease of implementation, we formulate our problem using ALD for the response.

Let Y_{it} be the dependent variable for subject i at time t and $\mathbf{X}_{it}$ be the vector of firm specific factors. Further, let $\mathbf{V}_{it}$ is a vector of factors for which the interest is to derive a more precise functional relationship with the dependent variable. For a fixed $\tau \in (0,1)$, we use the following formulation to model the τ^{th} quantile of the dependent variable.

$$Y_{it} \quad \sim \quad ALD(\cdot, \mu_{it}^\tau, \sigma, \tau) \tag{1}$$

where μ_{it}^τ is specified semi-parametrically as,

$$\mu_{it}^\tau = \mathbf{X}_{it}^T\boldsymbol{\beta} \ + \ \sum_{l=2}^{T} \alpha_l I_{[t=l]} \ + \ g_i(\mathbf{V}_{it}^T\boldsymbol{\kappa}) \tag{2}$$

Here, $\boldsymbol{\beta}$ is a vector of fixed effects for the firm specific factors, I_x is the indicator function for condition x, α_l allows for the yearly effect. The single-index part $g_i(\mathbf{V}_{it}^T\boldsymbol{\kappa})$ models for subject specific functional relationship between the τ^{th} quantile of Y_{it} and a linear combination of the vector of covariates $\mathbf{V}_{it}$. Note that this linear combination is given by $\mathbf{V}_{it}^T\boldsymbol{\kappa}$ where the coefficients $\boldsymbol{\kappa}$ are unknown but common across subjects. We vary the nonparametric function $g_i(\cdot)$ by firm. The proposed semiparametric quantile regression model offers great flexibility in modeling the response. In particular, the

model reduces to a linear model when $g_i(\cdot)$ is the identity function. Following, Ruppert et al. (2003), we specify $g_i(\cdot)$ by the following spline formulation based on a piecewise truncated polynomial of degree L:

$$g_i(V_{it}^T\kappa) \quad = \quad \gamma_{i1} + \gamma_{i2}(\mathbf{V}_{it}^T\boldsymbol{\kappa}) + \cdots + \gamma_{iL}(\mathbf{V}_{it}^T\boldsymbol{\kappa})^L + \sum_{d=1}^{D} \gamma_{i,L+d+1}(\mathbf{V}_{it}^T\boldsymbol{\kappa} - \eta_d)_+^L \quad (3)$$

where $(x)_+ = x$ if $x > 0$, and 0 otherwise, and $\eta_1 < \eta_2 < \cdots < \eta_D$ are the fixed knots, which are typically placed at quantiles of the distribution of unique values of $\mathbf{V}_{it}^T\boldsymbol{\kappa}$. The vector of random effects $\boldsymbol{\gamma}_i = (\gamma_{i1}, \cdots, \gamma_{i,L+D+1})$ is modeled nonparametrically by using a Dirichlet process prior. The above spline model of order L represents adequate fits for most of the data situations. However, the number of parameters may not be practical for smaller data sets. In those situations, simpler spline models such as linear splines may be used, or subject specific splines may be dropped. Typically, linear ($L = 1$), quadratic ($L = 2$) or a cubic ($L = 3$) splines are common choices in practice (Ruppert et al. (2003)) as they ensure a certain degree of smoothness in the fitted curve .

3.2 Dirichlet Process

The spline coefficient vectors $\boldsymbol{\gamma}_i$, for $i = 1, 2, .., N$, are usually assumed to be i.i.d with a parametric multivariate normal distribution. Recently, Ghosh et al. (2009) have shown that this assumption can be misleading if the actual distribution is misspecified. Thus, to make the model robust to misspecified distributions, we assume a Dirichlet process (Ferguson (1973)) prior for the spline coefficients. A Dirichlet process would assume an unknown probability measure G for $\boldsymbol{\gamma}_i$ and thus incorporate infinitely-many parameters in order to more flexibly model the uncertainty in G. In order to allow G to be an unknown distribution on the Euclidean space, let $G \sim DP(\nu, H_0)$, where $\nu > 0$ is called the concentration parameter, which characterizes the prior precision and H_0 a base distribution on the Euclidean space. By choosing a Dirichlet process prior for G, one allows G to be an unknown distribution, with H_0 corresponding to the best guess for G and ν expressing confidence in this guess. In particular, H_0 can be chosen to be a normal distribution with some mean and variance parameters. In addition, ν

is commonly assigned a gamma hyperprior to allow the data to inform more strongly about the extent to which G is close to H_0. Based on the above discussion we assume the following :

$$\begin{aligned} &\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_N | G \overset{\text{iid}}{\sim} G, \quad G|\nu, H_0 \sim DP(\nu, H_0), \\ &H_0|\boldsymbol{\mu}_{\mathbf{h}}, \boldsymbol{\Sigma} \ \sim N_{L+D+1}(\boldsymbol{\mu}_{\mathbf{h}}, \boldsymbol{\Sigma}), \quad \nu|a, b \ \sim Gamma(a, b) \end{aligned} \tag{4}$$

There are several ways to implement a Dirichlet process prior. Recent research has focused on using the following constructive definition to produce Markov Chain Monte Carlo (MCMC) algorithms (Sethuraman and Tiwari (1981) and Sethuraman (1994)), where the unknown random probability measure G can be written as:

$$\begin{aligned} G(\cdot) \quad &= \quad \sum_{r=1}^{\infty} p_r \delta_{\boldsymbol{\phi}_r}(\cdot) \qquad (5) \\ \text{where} \qquad &\boldsymbol{\phi}_r \overset{\text{iid}}{\sim} H_0 \text{ and } \delta_{\boldsymbol{\phi}_r}(\cdot) \text{ is the degenerate probability at } \boldsymbol{\phi}_r, \\ &p_1 = q_1, \ p_r = q_r \prod_{j=1}^{r-1}(1 - q_j) \text{ and } q_r \overset{\text{iid}}{\sim} \text{Beta}(1, \nu), \ r \geq 1 \end{aligned}$$

Recently, a common approach has been to truncate the above sum at a chosen large integer to obtain a finite approximation (Ishwaran and Zarepour (2002), Ishwaran and James (2002)). However, we prefer to use an approach due to Walker (2007) that circumvents the requirement of taking a finite approximation. The idea of this method is as follows. Let $\mathbf{p} = \{p_1, p_2, ...\}$ and $\boldsymbol{\phi} = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ...\}$. For each fixed i, a latent variable U_i is introduced such that the joint density of (γ_i, U_i) conditional on $(\mathbf{p}, \boldsymbol{\phi})$ is given by,

$$g_{(\mathbf{p},\boldsymbol{\phi})}(\boldsymbol{\gamma_i}, u_i) = \sum_{r=1}^{\infty} I_{u_i \leq p_r} \times \delta_{\boldsymbol{\phi}_r}(\boldsymbol{\gamma_i})$$

Then the marginal density of $\boldsymbol{\gamma}_i$ obtained by integrating out u_i w.r.t the Lebesgue measure, will be as in equation (5) and the marginal density of U_i will be:

$$g_{\mathbf{p}}(u_i) = \sum_{r \in A_p(u_i)} I_{u_i \leq p_r}, \text{ where } A_{\mathbf{p}}(u_i) = \{r : \ u_i \leq p_r\}$$

Note that for a given u_i and $\mathbf{p}$, the set $A_{\mathbf{p}}(u_i)$ is finite (because $\sum_{r=1}^{\infty} p_r = 1$ implying $p_r \to 0 \ as \ r \to \infty$). The conditional density of $\boldsymbol{\gamma}_i$ given U_i is then given by

$$g_{\mathbf{p},\boldsymbol{\phi}}(\boldsymbol{\gamma}_i | U_i = u_i) = \frac{1}{g_{\mathbf{p}}(u_i)} \sum_{r \in A_{\mathbf{p}}(u_i)} \delta_{\boldsymbol{\phi}_r}(\boldsymbol{\gamma}_i)$$

Further, another latent variable δ_i is introduced to indicate the component $k \in A_{\mathbf{p}}(u_i)$ for which $\boldsymbol{\gamma}_i = \boldsymbol{\phi}_k$. The joint density now becomes

$$g_{(\mathbf{p},\boldsymbol{\phi})}(\boldsymbol{\gamma}_i, U_i = u_i, \delta_i = k) = \sum_{k=1}^{\infty} I_{k \in A_{\mathbf{p}}(u_i)} \times \delta_{\boldsymbol{\phi}_k}(\boldsymbol{\gamma})$$

The idea is to then design the MCMC sampling scheme to iteratively generate $\boldsymbol{\gamma}_i$, U_i and δ_i for every $i \in \{1, 2, ..., N\}$.

3.3 Clustering of functions

A key feature of the Dirichlet process is the almost sure discreteness of the random measure G that assigns positive probability to common values among all $\boldsymbol{\gamma}_i$'s. The pattern of ties among the members of $\{\boldsymbol{\gamma}_i,\ i = 1, 2, ..., N\}$ determines a partition of the set of distinct subjects of $C = \{1, 2, \cdots, N\}$. In other words, C is divided into m unordered sets of disjoint nonempty subsets $C_1, C_2, \cdots, C_m$ whose union is C. If we associate a latent variable s_i with each $\boldsymbol{\gamma}_i$, then we can write $s_i = l$ if $i \in C_l$ and define $\boldsymbol{\gamma}_i = \boldsymbol{\phi}_{s_i}$. The s_i acts as a classification variable to identify $\boldsymbol{\phi}_l$ corresponding to a specific $\boldsymbol{\gamma}_i$. Thus, given the classification vector $\mathbf{s} = (s_1, \cdots, s_N)^T$ one can describe the clustering behavior of $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_N)^T$, which in turn is equivalent to clustering the functions $g_i(\cdot)$.

However, it is important to note that the resulting clustering depends on the particular realization of the random distribution G. Therefore, each simulation of G from an MCMC scheme will result in a possibly different clustering configuration. Understandably, one needs to arrive at a final clustering based on all the simulations of G (post the burn-in period). Our approach is similar to that of Medvedovic and Sivaganesan (2002) and can be briefly described as follows.

1. We compute the percentage of simulations in which subjects i and j are classified into the same cluster and denote this by P_{ij}.

2. We then compute the pair-wise distance matrix $\mathbf{D}$ with $(i, j)^{th}$ entry as $D_{ij} = 1 - P_{ij}$

3. Finally, we cluster the subjects using standard clustering techniques (more specifically "Complete-Linkage method") with $\mathbf{D}$ as the distance matrix. (This can be carried out using the hclust function in R or Proc clust procedure in SAS).

Since we carry out quantile regression at different quantiles, we get a membership matrix $[P_{ij}]$ at each quantile. One way to apply the above method would be to choose a particular quantile on which we would like to base the clutering. In our case, instead of preferring any one quantile model for clustering, we use all the quantiles by computing P_{ij} as the average pairwise membership probability across all the quantile models.

4 Bayesian Inference

4.1 Likelihood

The model specification for the data $\{(Y_{it}, \mathbf{X}_{it}, \mathbf{V}_{it})\ i = 1, 2, ..., N,\ t \in \{1, 2, ..., T\}\ \}$, is given by equations (1), (2) and (3). Let $S_i \subseteq \{1, 2, .., T\}$ be the time points at which data for subject i is observed. In order to subsequently derive an effective MCMC sampling scheme, we find it advantageous to use the representation of ALD as a scale mixture of normals (see Tsionas (2003), Yue and Rue (2011)). If $W_{it} \overset{\text{iid}}{\sim}$ Exponential (with mean=σ) and $Z_{it} \sim N(0, 1)$, then

$$Y_{it} = \left(\mu_{it}^{\tau} + \xi W_{it} + \epsilon Z_{it}\sqrt{\sigma W_{it}}\right) \sim\ ALD(\cdot, \mu_{it}^{\tau}, \sigma, \tau)$$

where $\xi = (1-2\tau) \times (\tau(1-\tau))^{-1}$ and $\epsilon^2 = 2 \times (\tau(1-\tau))^{-1}$. Accordingly, the likelihood based on the model specifications for a fixed $\tau \in (0, 1)$ can be written as follows.

$$L(\mathbf{Y}|\mathbf{X}, \mathbf{W}, \boldsymbol{\beta}, \boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \sigma) = \prod_{i=1}^{N} L_i(X_{i1}, .., X_{iT}, W_{i1}, .., W_{iT}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{\gamma_i}, \sigma) \quad (6)$$

where,

$$\begin{aligned} & L_i(X_{i1}, .., X_{iT}, W_{i1}, .., W_{iT}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \boldsymbol{\gamma_i}, \sigma) \\ = & \prod_{t \in S_i} \left\{ (2\pi\sigma\epsilon^2 W_{it})^{-1/2} e^{-\left(Y_{it} - \xi W_{it} - \mathbf{X}_{it}^T\boldsymbol{\beta} - \sum_{l=2}^{T} \alpha_l I_{[t=l]} - g_i(\mathbf{V}_{it}^T\boldsymbol{\kappa})\right)^2 / \left(2\sigma\epsilon^2 W_{it}\right)} \right\} \quad (7) \end{aligned}$$

where $\mathbf{Y}, \mathbf{X}$ and $\mathbf{W}$ respectively denote the collections of Y_{it}, X_{it} and W_{it} over all possible i and t. Similarly, $\boldsymbol{\alpha} = (\alpha_2, .., \alpha_T)$ and $\boldsymbol{\gamma}$ is the collection of $\boldsymbol{\gamma_i}$ over all i. For

identifiability, the intercept term is added to the spline function $g_i(\cdot)$ and excluded from $\mathbf{X}_{it}$. Also, note that the parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ are common for all subjects whereas the parameter coefficients for the spline function $\boldsymbol{\gamma}_i$ vary by subject.

4.2 Prior specification

The Bayesian specification is completed by specifying prior distributions for the set of parameters $\Omega = \{\boldsymbol{\beta}, \alpha_2, .., \alpha_T, \sigma, \boldsymbol{\kappa}, \boldsymbol{\gamma_1}, ..., \boldsymbol{\gamma_N}\}$. Recall that $W_{it} \overset{i.i.d}{\sim} exp(1/\sigma)$. We assume elements of Ω are independent. We assign conjugate priors for the regression coeffcients of the linear part (namely, $\boldsymbol{\beta}, \alpha_2, .., \alpha_T$) by assuming a normal density for each and for the scale parameter σ by assuming a inverse-gamma(IG) density. More specifically, $\boldsymbol{\beta} \sim N(\boldsymbol{\mu_\beta}, \boldsymbol{\Sigma_\beta})$, $\alpha_l \overset{\text{iid}}{\sim} N(\mu_l, \sigma_l), l = 2, .., T$ and $\sigma \sim IG(a, b)$, which is the distribution of $1/S$, where S follows a gamma distribution wth density function $f(s) \propto s^{a-1}\exp\{-b \times s\}$.

We assume a Dirichlet process prior for $\boldsymbol{\gamma}_i$ as described in equation (4). The functions $g_i(\cdot)$ and parameters $\boldsymbol{\kappa}$ are identifiable only up to a scalar multiple and hence we restrict the parameter space of $\boldsymbol{\kappa}$ to vectors with unit norm. Therefore, we assume a uniform distribution on the set $\{\boldsymbol{\kappa} : \boldsymbol{\kappa}^T\boldsymbol{\kappa} = 1\}$. It is worth noting that a careful choice of prior on σ drastically improves the accuracy of Bayesian estimation. Especially, in our problem the support of the dependent variable (average cost)is mostly between 0 and 1, with more than 99% values falling within this range. Therefore, a prior on σ which ensures that a large probability under the specified ALD density lies between 0 and 1, helps the estimation. We also impose a prior on the hyper parameter $\nu \sim Gamma(a_\nu, b_\nu)$.

4.3 MCMC algorithm

MCMC techniques have been extensively used in Bayesian modeling for the computation of posterior distributions (see Hastings (1970), Gelfand and Smith (1990)). Simulation for all the parameters except $\boldsymbol{\kappa}$ is based on a Gibbs sampling. Following Karabatsos (2009), we use an Adaptive Random-Walk Metropolis (ARWM) algorithm for simulat-

ing $\boldsymbol{\kappa}$. This uses a Metropolis-Hastings procedure with a symmetric distribution on the unit sphere as the proposal distribution and modifies the proposal distribution by introducing a scalar parameters λ so as to optimize the acceptance rate of the algorithm. Hence we introduce the parameter λ in our sampling. Also, we use the representation of Walker (2007) as described in Section 3.2 to simulate from the Dirichlet process. Let $\mathbf{U} = (U_1, U_2, ..., U_N), \boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_N)$, where the elements within the vectors are as in Section 3.2. The MCMC procedure is designed to generate all the required parameters, namely $\boldsymbol{\beta}, \sigma, \boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\kappa}, \lambda, \boldsymbol{\gamma_1}, ..., \boldsymbol{\gamma_N}, \mathbf{U}, \boldsymbol{\delta}, \mathbf{p}, \boldsymbol{\phi}$. The detailed procedure is provided in the appendix.

5 Data Analysis

5.1 Model specification

We analyze the insurance cost data described in Section 2 using the proposed modeling methodology. The data consists of information on the variables shown in Table 1 for $N = 1741$ insurance companies, collected over the years 2001 to 2006 (indexed using $t = 1, \ldots, 6$). In our context, the dependent variable is $Y_{it} =$ avg_cost for company i in year t. Our interest is to analyze the τ^{th} quantile of the distribution of avg_cost for each company, for various values of τ. Specifically, we consider $\tau \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$. This is accomplished by using the ALD formulation as described in Section 3. An important part of our model is the subject specific single-index curve which is formulated as a spline function that models the relationship between avg_cost and output. We try to obtain output as a linear combination of risk transfer and financial intermediation which are operationally measured by gross written premiums and total invested assets respectively. A key feature of a Bayesian approach is that it accommodates dependence across observations through random effects. In our case, the dependence within subject is modeled via the random effects in

the firm specific spline coefficients. Accordingly, the model formulation is as follows.

$$\begin{aligned}
\text{avg_cost}_{it} &\sim ALD(., \mu^{\tau}_{it}, \sigma, \tau),\ i = 1, 2, .., N,\ t \in \{1, 2, .., 6\} \\
\mu^{\tau}_{it} &= \beta_1\ \text{pct_property}_{it} + \beta_2\ \text{pct_liability}_{it} + \beta_3\ \text{pct_combined}_{it} \\
&\quad + \beta_4\ \text{pct_equity}_{it} + \beta_5\ \text{pct_cash}_{it} + \beta_6\ \text{group}_{it} + \beta_7\ \text{stock}_{it} \\
&\quad + \sum_{l=2}^{6} \alpha_l I_{(t=l)} + g_i\left(\kappa_1\ \text{prem_tot}_{it} + \kappa_2\ \text{assets_inv}_{it}\right)
\end{aligned}$$

Note that a company specific intercept enters the model through the spline function. We use a linear spline basis (degree $L = 1$) with knots chosen at the ten different deciles of the variable ($\kappa_1\ \text{prem\_tot}_{it} + \kappa_2\ \text{assets\_inv}_{it}$) in the data. This specification results in a basis with $11(= L + D + 1)$ degrees of freedom. The advantage of a linear spline over those of higher degrees is its simplicity, which also translates into a better convergence of the MCMC scheme. In line with our objective of deriving a firm-specific dependence of avg_cost on output, the coefficients of the spline function ($\gamma_{il}(\cdot)$) are taken to be firm-specific. Coefficients $(\alpha_2, .., \alpha_6)$ are the fixed effects for the years 2002 to 2006 respectively. The fixed effect for the year 2001 is excluded to ensure identifiability. For $\beta_1, .., \beta_7$ and $\alpha_2, .., \alpha_6$ we take independent $N(0, 1)$ priors. For $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{i(L+D+1)})$, we take a Dirichlet process prior with base probability measure $H_0 = N_{L+D+1}(0, I)$ and precision parameter $\nu \sim\ Gamma(1, 1)$. For σ we take $IG(10, 0.1)$. This particular choice of prior on σ helps ensure that the essential support of ALD density resembles that of the variable avg_cost in the data and to a great extent helps the convergence of MCMC scheme. Also, we impose a uniform prior on the set $\{(\kappa_1, \kappa_2) : \kappa_1^2 + \kappa_2^2 = 1\}$

5.2 Model Results

Implementing the MCMC scheme described in Section 4, the quantile regression model is estimated for five different quantile values, i.e. $10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}$ quantiles. To achieve convergence, the first 100,000 simulations are discarded as burn-in samples and further 10,000 simulations are generated for the analysis. Simulation of $\boldsymbol{\kappa}$ involved a Metropolis-Hastings step which resulted in approximately 20% acceptance rate.

Figure 1 presents a comparison of the empirical density curves for the modeled values at different quantile regressions with the histogram of actual data. As expected we see that the density curves for the lower quantiles are to the left of data and higher to the right, with the median model density curve mostly matching with the data. The relation between densities with modeled values and actual data suggests that the range of values resulting from the models are reasonable compared to that in the data.

To provide a little more insight into the cost distribution of individual firms, we show in Figure 2 the comparison of the actual average cost with modeled values for two different insurers. The insurer in the left panel presents higher volatility in that the average cost jumps back and forth across various quantiles. In contrast, the average cost of the insurer in the right panel is more stable, staying close to the 50^{th} quantile.

Figure 3 summarizes the parameter estimates at various quantile regressions, with the plus sign indicating the posterior mean of the parameter and the dotted lines marking the 95% credible interval. The horizontal solid line denotes the constant line at zero. Though one would conclude with an insignificant effect if zero lies within the credible interval, we emphasize that in a Bayesian context, we are more interested in the likelihood that a covariate would affect an insurer's cost. In this sense, if most of the interval is above (below) the zero line, the chance is high that the variable has a positive (negative) effect. Firstly, we observe that most of firm characteristics affect insurers' cost significantly. It is interesting to see the mixed effects of underwriting mix variables which are significant at lower and higher quantiles but not at the middle quantiles. Furthermore, the sign of coefficients of these variables is positive at lower quantiles and negative at higher quantiles, suggesting a positive effect for thrifty insurers and a negative effect for ones with higher costs. This result also emphasizes the importance of carrying out quantile regression at different values of τ. The traditional regression or a simple median regression would lead us to the biased conclusion that underwriting mix variables do not affect the insurer's production cost. As for investment mix variables, though zero is mostly within the credible interval, cash equivalent tends to have a negative effect on cost and equity tends to have a positive effect on cost. This could be explained by the lower transaction cost involved in the cash equivalent investment.

Another noticeable result is the effect of organization form on the average cost. Figure 3 suggests that stock insurer has lower cost at higher quantiles but higher cost at lower quantiles. This is consistent with the coexistence of stock and mutual property-casualty insurers in the market, though the agency theoretical hypothesis predicts that mutual insurers are less successful than stocks in minimizing costs (Cummins et al. (1999)). Finally, as anticipated, the fixed effect of years are less pronounced at all quantiles, because the year effects presumably have been captured by the individual effects in the single-index function.

5.2.1 Average Cost Curve

One application of our approach is to identify the unique average cost curve for each insurer. As discussed earlier, microeconomic theory implies a U-shaped average cost curve as a function of output for individual firms and different firms might produce at different positions of its own cost curve. However, in practice, curves need not exactly conform to this theory. For example, there could be certain ranges of output for which the curve is flat. Therefore, a method that allows for more flexible curve shapes is desirable. The single-index formulation based on splines provides this flexibility and allows us to capture the aggregated effect of multiple outputs. Note that in reality one does not observe the entire cost curve for an insurer, thus the estimated curve has more credibility in the range of observed output. To illustrate, we choose to present the average cost curve for two individual insurers, as shown in Figure 4. The cost curve for insurer X is estimated from the 50% quantile regression and the cost curve for insurer Y is estimated from the 75% quantile regression. The actual average costs in different years are marked by solid triangles. Figure 4 suggests decreasing return to scale for company X and constant return to scale for company Y. Presuming that the insurance market is competitive, company X is expected to be more profitable than company Y.

5.2.2 Comparison with Independent Financial Ratings

Additional insights could be gained by comparing our result with an external measure. In doing so, we look into the A.M. Best financial strength rating for the US property

casualty insurers. A letter-based score is assigned for each insurer by A.M. Best Company based on an independent evaluation of the insurer's ability to meet its ongoing contract obligations. In general, each insurer will be categorized into one of the four rating scales: superior, excellent, good, and vulnerable. Figure 5 shows the mean cost curves at different quantiles by groups of A.M. Best ratings. Presumably, a more cost-efficient insurer tends to be more profitable and thus to have a better financial status. The better alignment of the 75% quantile model with the A.M. Best rating suggests that the 75th percentile of the insurers' cost is more weighted in the evaluation process. This is not surprising as the cost distribution is right skewed as seen in Figure 1. Note that the A.M. Best rating incorporates many other factors apart from cost, thus we do not expect an exact match.

5.2.3 Clustering of Insurers

The utilization of the Dirichlet process prior allows us to cluster insurers according to the similarity in the level of cost and shape of the curve. Specifically, we group insurers based on the pairwise membership matrices obtained for the coefficient vector $\boldsymbol{\gamma}_i, i = 1, \ldots, N$ of the single-index spline function $g_i(\cdot)$. As described in Section 3.3, a specification of the number of target clusters is needed for the hierarchical clustering procedure. Since the number of clusters from the Dirichlet process varies across quantiles, we chose the average number (=9) of clusters that is implied by the MCMC simulation of the Dirichlet process for different quantile regressions. For illustration purposes, we show the plot of individual cost curves within each cluster from the 75% quantile model in Figure 6. To help discern the typical shape of the curves within each cluster, we also exhibit the mean curve (black solid line). As we expected, the shapes of the curves do not exactly conform to theory of being U-shaped. However, they are not totally inconsistent either. The initial part of the curve corresponding to approximately half the range of output is clearly part of a U-shape. The rest of the curve would have been if not for a portion that is flat or has a slight dip. The latter behavior can perhaps be attributed to significant changes in the production process beyond certain levels of output. In general, clustering points us to companies with

similar returns to scale. In particular, cluster 1 and cluster 9 show steeper returns to scale than clusters 2 or 5, which are relatively flat.

5.2.4 Comparison of cost efficiency

Another application of the proposed method is to compare the cost efficiency across insurers. The quantile regression specification enables us to derive the cumulative distribution function of average cost for individual insurers. Insights of relative efficiency among insurers could be gained by comparing the distribution of their average costs. For demonstration purposes, we display in Figure 7 the implied cumulative distribution function of average cost for two pairs of insurers. Note that the distribution of average cost is derived after controlling for firm characteristics. In the left panel, we see that insurer A has first-order stochastic dominance over insurer B, indicating a higher cost efficiency of insurer A than insurer B. In contrast, the right panel exhibits a case that ranking the cost efficiency among insurers is not straightforward. While the medians for the two companies are close, the average cost of insurer C is higher at low quantiles but lower at high quantiles than insurer D. This example again emphasizes the advantages of quantile regression analysis: more information could be revealed by looking into different quantiles of the cost distribution rather than focusing on the center, where two insurers show similar level of cost efficiency.

6 Concluding Remarks

We examined average cost of property-casualty insurers using a Bayesian semiparametric quantile regression approach. Although much of the motivation for this article was developed in the context of insurance industry, in principle, our results are useful for any market where each firm has multiple outputs and cost function varies across individual firms. Additional potential applications are easy to imagine, for example, banking industry. Thus the size of the economic sector such as insurance industry provides sufficient motivation for this work.

The average cost curve by definition is unobservable. For a given firm, it measures

the hidden relationship between average production cost and output level given the technology adopted by the company in the production process. To compare the average cost among insurers, we proposed a Bayesian quantile regression based on the asymmetric Laplace distribution. The quantile regression examines the whole distribution of average cost instead of just focusing on the center. We show that such analysis is in particular useful for skewed and heavy-tailed distributions such as insurers' production costs in our application. We also show that more insights on relative cost efficiency could be gained by comparing the cumulative distribution function of average cost derived from the regression quantiles. Our formulation of the quantile regression employed a single-index formulation to capture the nonlinear dependence of average cost on multiple outputs. Thus, we arrive at a single measure of output that helps identify returns to scale for individual insurers. Motivated by the observation that insurers adopt different technologies in their production, we also showed that the unique cost function of each insurer could be identified by allowing the single-index formulation to be firm-specific within a longitudinal context. By using a Dirichlet process prior, we show that insurers could be grouped according to similar level of cost efficiency. A comparison of our results with an independent financial strength rating leads to an interesting finding that the right tail of the cost distribution is perhaps weighted more by rating agencies.

While the semiparametric quantile regression using single-index turns out insightful, few limitations of our method must be underlined. Firstly, the quantile regressions at different quantiles are estimated individually by running the model separately for each desired quantile. One may be interested to look at more than one quantile simultaneously. Although there have been some recent developments on simultaneous modeling of quantiles, their adaptation to the current problem is not straight forward and requires further investigation. Secondly, the parameters $\boldsymbol{\kappa}$ can vary for different quantile regressions, which is a problem if one wishes to fix it across quantiles and only vary the function g_i. A good way to address this would again involve simultaneous modeling of quantiles. We are currently exploring these issues pf modeling. Notwithstanding these limitations, this research has pointed to a new road map for the modeling of insurance

cost data and addresses some of the key challenges.

Appendix

Details of the MCMC algorithm

By way of notation, $\mathbf{X}$ will include the dummy columns indicating the year variable. Let $\mathbf{V}$ be the matrix with elements of the set $\{\mathbf{V}_{it}, i = 1, .., N, t \in \{1, .., T\}\}$ as it's rows. Let v= number of columns of $\mathbf{V}$. Similarly let $\mathbf{W}$ be the column containing the elements W_{it}. For a matrix $\mathbf{Z}_1$ and column vector $\mathbf{z}_2$ with same number of rows, $\mathbf{Z}_1/\mathbf{z_2}$ will mean the matrix, whose columns are obtained by taking term-wise ratios of each column of $\mathbf{Z}_1$ with the vector $\mathbf{z}_2$. Similarly $\sqrt{\mathbf{Z}_1}$ will be the matrix obtained by taking square-root of each element and matrix multiplication $\mathbf{Z}_1.\mathbf{Z}_2$ for two equally dimensioned matricies $\mathbf{Z}_1$ and $\mathbf{Z}_2$, will be a matrix of the same dimension whose entries will be the scalar product of the entries of the two matricies in the same position. Also recall that $\xi = (1 - 2\tau) \times (\tau(1 - \tau))^{-1}$ *and* $\epsilon^2 = 2 \times (\tau(1 - \tau))^{-1}$. After starting with some initial values, let the simulated values of the parameters of interest at step s be given by $\boldsymbol{\beta}^{(s)}, \sigma^{(s)}, \boldsymbol{\nu}^{(s)}, \boldsymbol{\alpha}^{(s)}, \boldsymbol{\kappa}^{(s)}, \lambda^{(s)}, \boldsymbol{\gamma}_i^{(s)}, \mathbf{U}_i^{(s)}, \boldsymbol{\delta}^{(s)}, \mathbf{p}^{(s)}, \boldsymbol{\phi}^{(s)}$.

Step 1: Simulating $\boldsymbol{\kappa}$ and λ

We generate these parameters in the lines of Karabatsos (2009). Using their method, $\boldsymbol{\kappa}^{(s+1)}$ is obtained after generating a proposal $\boldsymbol{\kappa}^*$ as follows

- $\boldsymbol{\kappa} \sim N_v\left(\boldsymbol{\kappa}^{(s)} \times 2 \times {\lambda^{(s)}}^2, \mathbf{I}_v\right)$, where $\mathbf{I}_v$ is the identity matrix of order v.
- $\boldsymbol{\kappa}^* = \boldsymbol{\kappa}/\boldsymbol{\kappa}'\boldsymbol{\kappa}$
- Compute $\rho = \min\left(1, L(\mathbf{Y}|\mathbf{X}, ..., \boldsymbol{\kappa}^*, ...)/L(\mathbf{Y}|\mathbf{X}, ..., \boldsymbol{\kappa}^{(s)}, ...)\right)$, where parameters other than $\boldsymbol{\kappa}$ in computing likelihood $L(\cdot)$ (as in equation 6) are held fixed at the values obtained at the s^{th} step.
- $\boldsymbol{\kappa}^{(s+1)} = \boldsymbol{\kappa}^*$ with probability ρ and $\boldsymbol{\kappa}^{(s+1)} = \boldsymbol{\kappa}^{(s)}$ with probability $(1 - \rho)$.
- $\lambda^{(s+1)} = max\left(0, \lambda^{(s)} + (s + 1)^{-1/2}(.234 - \rho)\right)$. This step is to approximately ensure a 23% acceptance rate in the sampling.

Step 2: Simulating $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$

Although $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ represent two types of variables (viz. company characteristics and year dummies), we find it easier to simulate them as a combined block, denoted by $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}, \alpha_2, ..., \alpha_l)^T$. The corresponding covariate matrix $\tilde{\boldsymbol{X}}$ is formed by stacking the rows X_{it} for $i \in \{1, 2, ..., N\},\ t \in \{1, 2..., T\}$ and by appending dummy columns for the year variables. Accordingly the prior for $\tilde{\boldsymbol{\beta}} \sim Normal(\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}})$, which is obtained by combining the independent priors specified for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ in Section 4.2. So, $\tilde{\boldsymbol{\beta}}^{(s+1)}$ is simulated as follows.

- $\mathbf{X}^* = \tilde{\mathbf{X}}/\sqrt{\sigma\epsilon^2\mathbf{W}^{(s)}}$
- $\mathbf{Y}^* = (\mathbf{Y} - \xi\mathbf{W}^{(s)} - \tilde{\mathbf{X}}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}} - \mathbf{g}^{(s)}(\mathbf{V}\boldsymbol{\kappa}^{(s+1)}))/\sqrt{\sigma\epsilon^2\mathbf{W}^{(s)}}$, where $\mathbf{g}^{(s)}(\mathbf{V}\boldsymbol{\kappa}^{(s+1)})$ is the column vector with entries $\left\{g_i^{(s)}(\mathbf{V}_{it}\boldsymbol{\kappa}^{(s+1)}), i = 1, 2, ..., N,\ t \in \{1, 2, .., T\}\right\}$ and

$$g_i^{(s)}(x) \quad = \quad \gamma_{i1}^{(s)} + \gamma_{i2}^{(s)}x + \cdots + \gamma_{iL}^{(s)}x^L + \sum_{d=1}^{D}\gamma_{i,L+d+1}^{(s)}(x - \eta_d)_+^L$$

  with $(\eta_1, .., \eta_D)$ being equally spaced quantiles of $\{\mathbf{V}_{it}\boldsymbol{\kappa}^{(s+1)}$ for $i = 1, .., N$, $t \in \{1, 2, ..T\}\}$. Note that the superscripts on $\boldsymbol{\kappa}$ and $\boldsymbol{\gamma}$ are differnt since an $(s+1)^{th}$ updated value for $\boldsymbol{\kappa}$ is available from the previous step but not yet for $\boldsymbol{\gamma}$.
- $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^* = \left(\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}{}^{-1} + \mathbf{X}^{*T}\mathbf{X}^*\right)^{-1}$
- $\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}^* = \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^* \times \left(\mathbf{X}^{*T}\mathbf{Y}^* + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^{-1}\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}\right)$
- Simulate $\tilde{\boldsymbol{\beta}}^{(\boldsymbol{s+1})}$ from $N_b(\boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}^*, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}^*)$, where b is the dimension of $\tilde{\boldsymbol{\beta}}$

Step 3: Simulating W

It can be seen (as in Yue and Rue (2011)) that the distribution of W_{it}^{-1} for each (i, t), conditioned on other parameters and the data is Inverse Gaussian. As mentioned in their paper, the inverse gaussian density with parameters $(\lambda', \mu')$is given by

$$f(x) = \sqrt{\frac{\lambda'}{2\pi}}x^{-3/2}exp\left(-\frac{\lambda'(x-\mu')^2}{2(\mu')^2x}\right); x > 0$$

We simulate $\mathbf{W}^{(s+1)}$ as follows.

- Compute the vector $\boldsymbol{\mu}^{'} = (\xi^2 + 2\epsilon^2)/\left(\mathbf{Y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}^{(s+1)} - \mathbf{g}^{(s)}(\mathbf{V}\boldsymbol{\kappa}^{(s+1)})\right)^2$, where $\tilde{\mathbf{X}}, \tilde{\boldsymbol{\beta}}^{(s+1)}, \mathbf{g}^{(s)}(\mathbf{V}\boldsymbol{\kappa}^{(s+1)})$ are as in the previous step.
- Compute $\lambda^{'} = (\xi^2 + 2\epsilon^2)/\sigma\epsilon$.
- Simulate ${W_{it}^{(s+1)}}^{-1}$ from $Inverse\ Guassian(\mu_{it}^{'}, \lambda^{'})$, where $\mu_{it}^{'}$ is an individual element of the vector $\boldsymbol{\mu}^{'}$.

Step 4: Simulating σ

- Compute $\mathbf{Y}^* = \left(\mathbf{Y} - \xi\mathbf{W}^{(s+1)} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}^{(s+1)} - \ \mathbf{g}^{(s)}(\mathbf{V}\boldsymbol{\kappa}^{(s+1)})\right) / \left(2\epsilon^2\mathbf{W}^{(s+1)}\right)$.
- Compute $b^* = b + \mathbf{Y}^{*\prime}\mathbf{Y}^* + \sum_{i=1}^{N}\sum_{t\in S_i} W_{it}$.
- Compute $a^* = a + N + N/2$.
- Simulate ${\sigma^{(s+1)}}^{-1}$ from $Gamma(a^*, b^*)$.

Step 5: Simulating $\boldsymbol{\gamma}$

Recall that $\boldsymbol{\delta}$ is a vector of dimension N, whose i^{th} entry indicates the position of subject i in the infinite series of equation (5).

- Simulate $\mathbf{U}_i^{(s+1)}$ from $Uniform\left[0, p_i^{(s)}\right]$, for each $i = 1, 2, ..., N$.
- For $k \leq \max\left\{\delta_i^{(s)}, i = 1, 2, ..., N\right\}$, such that $k \neq \delta_i^{(s)}$, for any $i \in \{1, 2.., N\}$, simulate $\boldsymbol{\phi}_k^{(s+1)}$ from $N_{L+D+1}(\mathbf{0}, \boldsymbol{\Sigma})$.
- For $k \leq \max\left\{\delta_i^{(s)}, i = 1, 2, ..., N\right\}$, such that $k = \delta_i^{(s)}$, for some $i \in \{1, 2, .., N\}$, simulate $\boldsymbol{\phi}_k^{(s+1)}$ as follows.

  (i) Let $\mathbf{Y}_\gamma = \left(\mathbf{Y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}^{(s+1))} - \xi\mathbf{W}^{(s+1)}\right) / \sqrt{\epsilon^2\sigma\mathbf{W}^{(s+1)}}$ and $\mathbf{Y}_{\gamma i}$ be the sub-vector of the vector $\mathbf{Y}_\gamma$ containing only rows that corresponding to subject i.

  (ii) Let $\mathbf{G}_\gamma$ be the matrix of the following columns.

$$\left[1, (\mathbf{V}\boldsymbol{\kappa}^{(s+1)})^1, \cdots, (\mathbf{V}\boldsymbol{\kappa}^{(s+1)})^L, (\mathbf{V}\boldsymbol{\kappa}^{(s+1)} - \eta_d^{(s+1)})_+^1, \cdots, (\mathbf{V}\boldsymbol{\kappa}^{(s+1)} - \eta_d^{(s+1)})_+^L\right]$$

  and let $\mathbf{G}_{\gamma i}$ be the submatrix of the matrix $\mathbf{G}_\gamma$ containing only rows that corresponding to subject i.

(iii) Compute $\boldsymbol{\Sigma}_{\boldsymbol{\phi}_k} = \left(\boldsymbol{\Sigma}^{-1} + \mathbf{G}_{\gamma i}^T \mathbf{G}_{\gamma i}\right)^{-1}$.

(iv) Compute $\boldsymbol{\mu}_{\boldsymbol{\phi}_k} = \boldsymbol{\Sigma}_{\boldsymbol{\phi}_k}\left(\mathbf{G}_{\gamma i}^T \mathbf{Y}_{\gamma i} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_h\right)$

(v) Simulate $\boldsymbol{\phi}_k^{(s+1)}$ from $N_{L+D+1}(\boldsymbol{\mu}_{\boldsymbol{\phi}_k}, \boldsymbol{\Sigma}_{\boldsymbol{\phi}_k})$

- For $k \leq \max\left\{\delta_i^{(s)}, i = 1, 2, ..., N\right\}$, obtain $p_k^{(s+1)}$ as follows

  (i) Let $a_{(p)} = max_{\{r:\delta_r^{(s)}=k\}}\left\{U_r^{(s+1)} / \left(\prod_{l<k}\left(1 - q_l^{(s)}\right)\right)\right\}$

  (ii) Let $b_{(p)} = 1 - max_{\{r:\delta_r^{(s)}>k\}}\left\{U_r^{(s+1)} / \left(q_{\delta_r}\prod_{l<\delta_r, l\neq k}\left(1 - q_l^{(s)}\right)\right)\right\}$

  (iii) Simulate $q_k^{(s+1)}$ from truncated beta $\propto Beta(1, \nu^{(s)}) I_{(a_{(p)}<q<b_{(p)})}(\cdot)$

  (iv) Compute $p_k^{(s+1)}$ from $q_k^{(s+1)}$ using the relation under equation (5)

- In order to simulate the $(s+1)^{th}$ step of $\delta^{(s+1)}$, it may appear as though we need $p_k^{(s+1)}$ and $\phi_k^{(s+1)}$ for all k. Note that so far the simulation has been described only for $k \leq \max(\delta_i^{(s)}, i = 1, 2, ..., N)$. The key feature of Walker's method is that the introduction of latent variables U_i helps put an upper bound on k, which is the smallest k^* such that $\sum_{k=1}^{k^*} p_k^{(s+1)} > 1 - \min\left(U_1^{(s+1)}, .., U_N^{(s+1)}\right)$. Now, if $k^* \leq \max(\delta_i^{(s)}, i = 1, 2, ..., N)$ then there is no need to simulate additional terms and we can discard the terms $p_k^{(s+1)}$, $\boldsymbol{\phi}_k^{(s+1)}$ after k^*. If $k^* > \max(\delta_i^{(s)}, i = 1, 2, ..., N)$ then for $k > \max(\delta_i^{(s)}, i = 1, 2, ..., N)$ up to k^*, simulate $\boldsymbol{\phi}_k^{(s+1)}$ from $N(\mu_h, \Sigma)$, $q_k^{(s+1)} \sim Beta(1, \nu^s)$ and update $p_k^{(s+1)}$.

- For each $i \in \{1, 2, .., N\}$, simulate $\delta_i^{(s+1)}$ from a discrete distribution taking values in $\{1, 2, ..., k^*\}$ with probabilities $\{\pi_1, ..., \pi_{k^*}\}$ such that $\pi_r \propto L_r\left(\beta^{(s+1)}, \alpha^{(s+1)}, W_{i1}^{(s+1)}, .., W_{iT}^{(s+1)}, \gamma_i = \phi_r^{(s+1)}\right)$, where the function $L_r(.)$ is as in equation (7).

- Then for $i \in \{1, 2, .., N\}$, compute $\gamma_i^{(s+1)} = \phi_{\delta_i^{(s+1)}}^{(s+1)}$.

Simulating hyper parameter ν

Here we follow the methodology in Escobar and West (1995).

- Let $C_\gamma^{(s+1)}$ = number of distinct elements in the set $\left\{\boldsymbol{\gamma}_1^{(s+1)}, .., \boldsymbol{\gamma}_N^{(s+1)}\right\}$. Note that this is indeed the number of clusters obtained from the $(s+1)^{th}$ simulation.

- Let η_γ be a simulated value from $Beta(\nu^{(s)}+1, N)$
- Compute $O_\eta = (a_\nu + C_\gamma^{(s+1)} - 1)/(N \times (b_\nu - log(\eta_\gamma)))$ and $\pi_\eta = O_\eta/(1+O_\eta)$
- Simulate $\nu^{(s+1)}$ from the mixture distribution

$$\pi_\eta Gamma\left(a_\nu + C_\gamma^{(s+1)}, b_\nu - log\left(\eta_\gamma\right)\right)$$
$$+(1-\eta_\gamma) \times Gamma\left(a_\nu + C_\gamma^{(s+1)} - 1, b_\nu - log(\eta_\gamma)\right)$$

References

Antoniadis, A., Grégoire, G., and Mckeague, I. W. (2004), "Bayesian Estimation in Single-index Models," *Statistica Sinica*, 14, 1147–1164.

Berger, A. and Humphrey, D. (1997), "Efficiency of Financial Institutions: International Survey and Directions for Future Research," *European Journal of Operational Research*, 98, 175–212.

Cummins, J. and Weiss, M. (2000), "Analyzing Firm Performance in the Insurance Industry using Frontier Efficiency Methods," in *Handbook of Insurance*, ed. Dionne, G., pp. 767–829.

Cummins, J., Weiss, M., and Zi, H. (1999), "Organizational Form and Efficiency: the Coexistence of Stock and Mutual Property-Liability Insurers," *Management Science*, 45, 1254–1269.

Eling, M. and Luhnen, M. (2010), "Frontier Efficiency Methodologies to Measure Performance in the Insurance Industry: Overview, Systematization, and Recent Developments," *The Geneva Papers on Risk and Insurance-Issues and Practice*, 35, 217–265.

Escobar, M. D. and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Ferguson, T. S. (1973), "A Bayesian Analysis of some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Geraci, M. and Bottai, M. (2007), "Quantile Regression for Longitudinal Data using the Asymmetric Laplace Distribution," *Biostatistics*, 8, 140–154.

Ghosh, P., Basu, S., and Tiwari, R. C. (2009), "Bayesian Analysis of Cancer Rates from SEER Program using Parametric and Semiparametric Joinpoint Regression Models," *Journal of the American Statistical Association*, 104, 439–452.

Hardle, W., Hall, P., and Ichimura, H. (1993), "Optimal Smoothing in Single-Index Models," *The Annals of Statistics*, 21, 157–178.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71–120.

Ishwaran, H. and James, L. F. (2002), "Approximate Dirichlet Process Computing in Finite Normal Mixtures: Smoothing and Prior Information," *Journal of Computational and Graphical Statistics*, 11, 508–532.

Ishwaran, H. and Zarepour, M. (2002), "Dirichlet Prior Sieves in Finite Normal Mixtures," *Statistica Sinica*, 12, 941–963.

Jung, S.-H. (1996), "Quasi-Likelihood for Median Regression Models," *Journal of the American Statistical Association*, 91, 251–257.

Karabatsos, G. (2009), "Modeling Heteroscedasticity in the Single-Index Model with the Dirichlet Process," *Advances and Applications in Statistical Sciences*, 1, 83–104.

Koenker, R. (2004), "Quantile regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89, special Issue on Semiparametric and Nonparametric Mixed Models.

— (2005), *Quantile Regression (Econometric Society Monographs)*, Cambridge University Press.

Koenker, R. and Bassett, Gilbert, J. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R. and Xiao, Z. (2002), "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

— (2004), "Unit Root Quantile Autoregression Inference," *Journal of the American Statistical Association*, 99, 775–787.

— (2006), "Quantile Autoregression," *Journal of the American Statistical Association*, 101, 980–990.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G., and Zhao, L. P. (1997), "Quantile Regression Methods For Longitudinal Data with Drop-Outs: Application to CD4 Cell Counts of Patients Infected with the Human Immunodeficiency Virus," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 46, 463–476.

Medvedovic, M. and Sivaganesan, S. (2002), "Bayesian Infinite Mixture Model Based Clustering of Gene Expression Profiles," *Bioinformatics*, 18, 1194–1206.

Reich, B. J., Bondell, H. D., and Wang, H. J. (2010), "Flexible Bayesian Quantile Regression for Independent and Clustered data," *Biostatistics*, 11, 337–352.

Reich, B. J., Fuentes, M., and Dunson, D. B. (2011), "Bayesian Spatial Quantile Regression," *Journal of the American Statistical Association*, 106, 6–20.

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge University Press.

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 4, 639–650.

Sethuraman, J. and Tiwari, R. C. (1981), "Convergence of Dirichlet Measures and the Interpretation of Their Parameter," Tech. Rep. M583, Department of Statistics, Florida State University.

Shi, P. and Frees, E. (2010), "Long-tail Longitudinal Modeling of Insurance Company Expenses," *Insurance: Mathematics and Economics*, 47, 303–314.

Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2011), "Posterior Consistency of Bayesian Quantile Regression Based on the Mis-specified Asymmetric Laplace Density," *Unpublished Manuscript*.

Tsionas, E. G. (2003), "Bayesian Quantile Inference," *Journal of Statistical Computation and Simulation*, 73, 659–674.

Walker, S. G. (2007), "Sampling the Dirichlet Mixture Model with Slices," *Communications in Statistics - Simulation and Computation*, 36, 45–54.

Wang, H.-B. (2009), "Bayesian Estimation and Variable Selection for Single-Index Models," *Compututational Statistics and Data Analysis*, 53, 2617–2627.

Wu, T. Z., Yu, K., and Yu, Y. (2010), "Single-Index Quantile Regression," *Journal of Multivariate Analysis*, 101, 1607–1621.

Yu, K. and Moyeed, R. A. (2001), "Bayesian Quantile Regression," *Statistics and Probability Letters*, 54, 437–447.

Yu, Y. and Ruppert, D. (2002), "Penalized Spline Estimation for Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 97, 1042–1054.

Yuan, Y. and Yin, G. (2010), "Bayesian Quantile Regression for Longitudinal Studies with Nonignorable Missing Data," *Biometrics*, 66, 105–114.

Yue, Y. R. and Rue, H. (2011), "Bayesian Inference for Additive Mixed Quantile Regression Models," *Compututational Statistics and Data Analysis*, 55, 84–96.

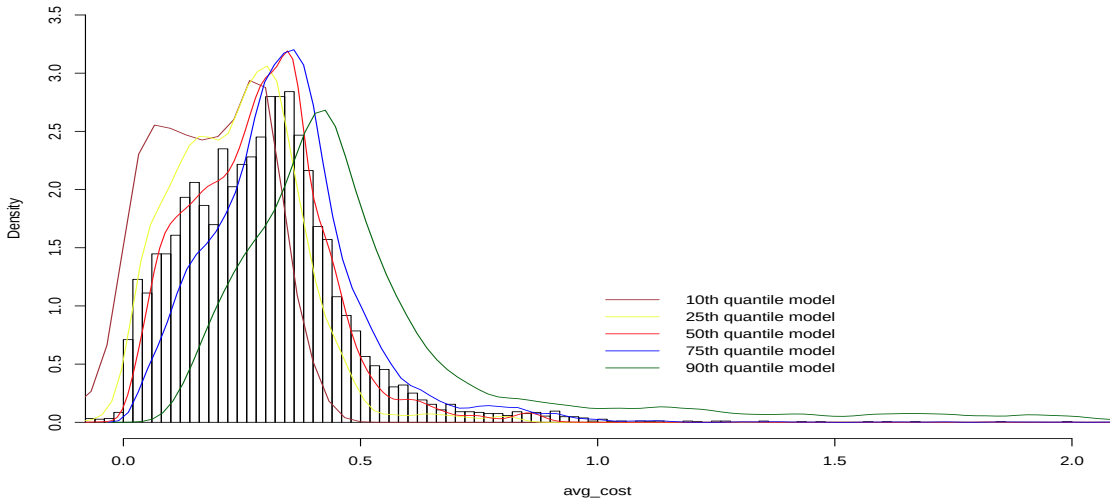

Figure 1: Histogram of avg_cost - The empirical density curves for the modeled values at different quantile regressions along with the histogram (bar chart) for actual data.

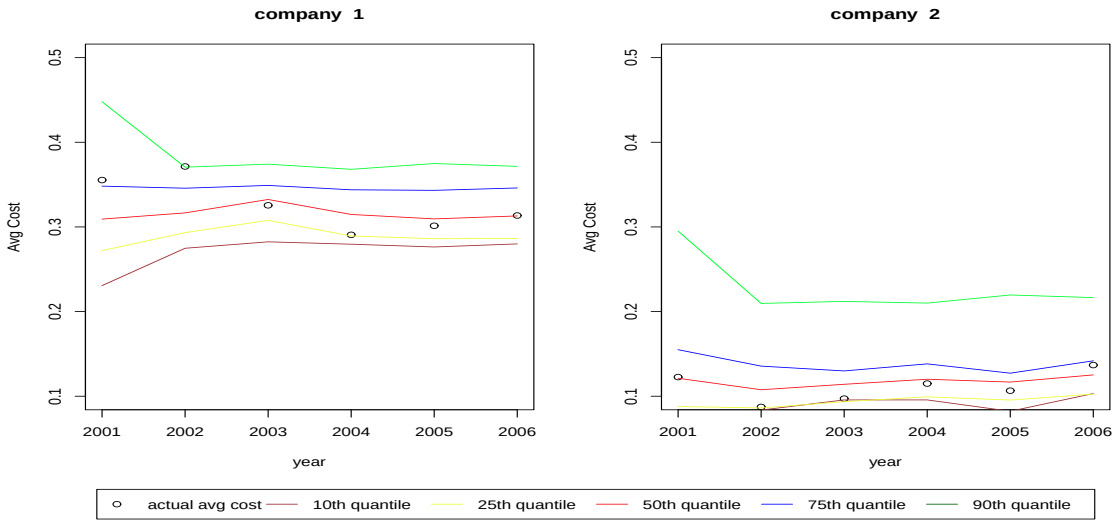

Figure 2: Actual vs modeled by year- A comparison of the actual average cost with modeled values over time for two different insurers. The small circles indicate the actual average cost in the data. Position of the circle is suggestive of the quantile of the average cost distribution at which the company is operating in a given year.

Table 1: Description of variables

Variable	Description
avg_cost	Average cost
pct_property	Percentage of premium written in property lines
pct_liability	Percentage of premium written in liability lines
pct_combined	Percentage of premium written in combined lines
pct_equity	Percentage of assets invested in equity
pct_cash	Percentage of assets invested in cash equivalent
prem_tot	Gross premiums written
assets_inv	Total invested assets
group	equals 1 if the insurer is affiliated to a group, and 0 otherwise
stock	equals 1 if the insurer is a stock company, and 0 otherwise

Table 2: Descriptive statistics of variables†

Variable	Mean	SD	Q1	Median	Q3
avg_cost	0.298	0.264	0.168	0.287	0.380
pct_property	0.217	0.231	0.003	0.178	0.322
pct_liability	0.505	0.369	0.168	0.494	0.773
pct_combined	0.230	0.297	0.000	0.085	0.385
pct_equity	0.137	0.167	0.000	0.079	0.21
pct_cash	0.162	0.218	0.035	0.084	0.192
prem_tot	396.300	1583.551	14.718	57.164	212.620
assets_inv	590.606	2834.126	20.347	68.42	250.904
group	0.662				
stock	0.702				

† Monetary variables are in million dollars.

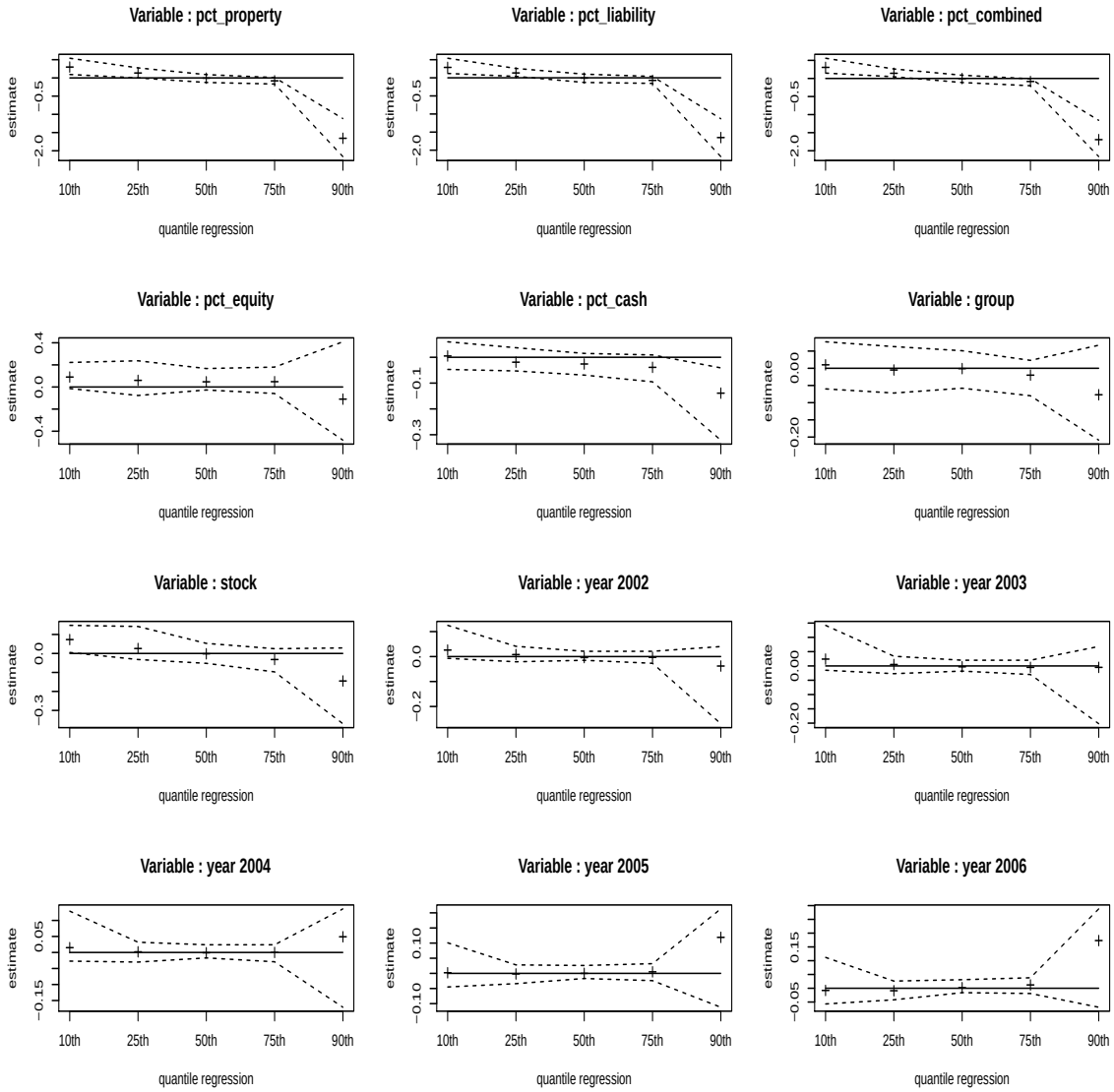

Figure 3: Parameter estimates- + indicates the estimated posterior mean of the parameter. The dotted lines mark the 95 percent credible interval. The solid line marks the constant line at 0. The x-axis shows for which quantile of avg_cost the regression has been run (viz; 10th, 25th, 50th, 75th and 90th)

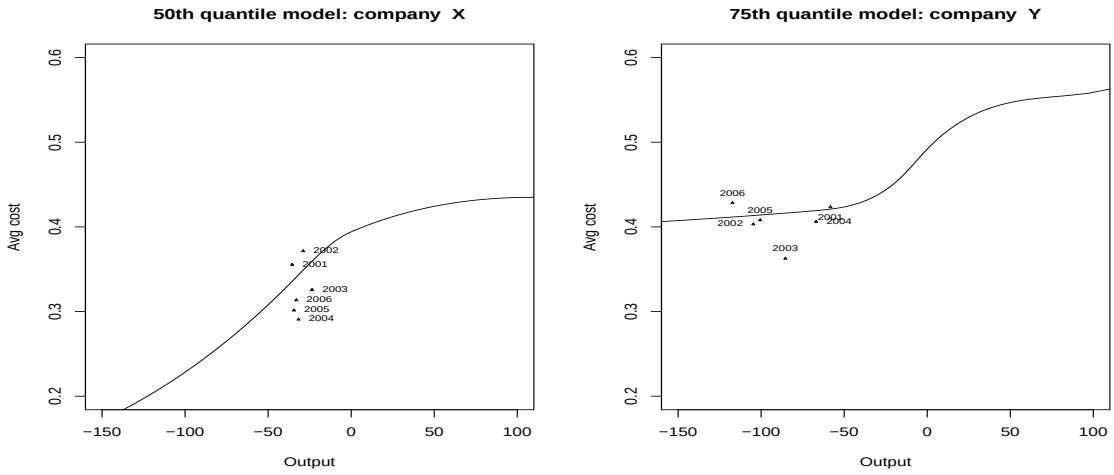

Figure 4: Individual company cost curves- The average cost curve for two individual insurers. The cost curve for insurer X is estimated from the 50% quantile regression and for insurer Y is estimated from the 75% quantile regression. The actual average costs in different years are marked by solid triangles.

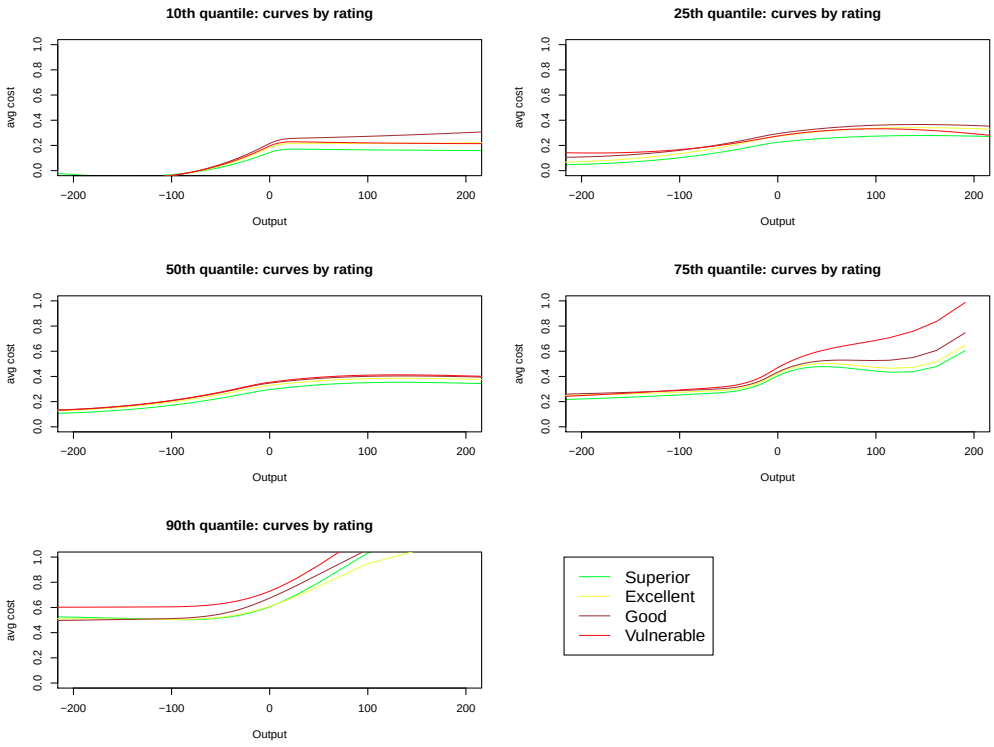

Figure 5: Mean Cost curves by AM Best rating - Mean cost curves at different quantiles by groups of A.M. Best ratings. Each insurer is categorized into one of the four rating scales: superior, excellent, good, and vulnerable. The curves shown are the average of all the curves within a given class.

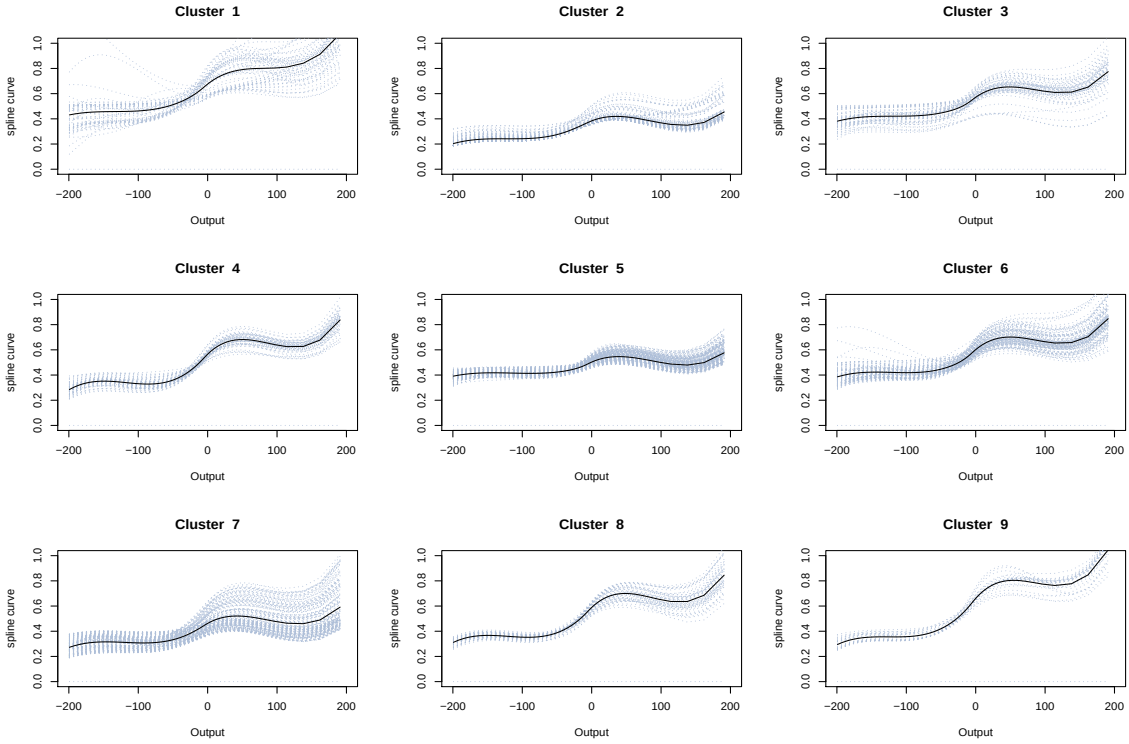

Figure 6: Individual Cost curves within clusters for 75th quantile regression- Plot of individual cost curves within each cluster from the 75% quantile model. The black solid line indicates the mean curve within the cluster.

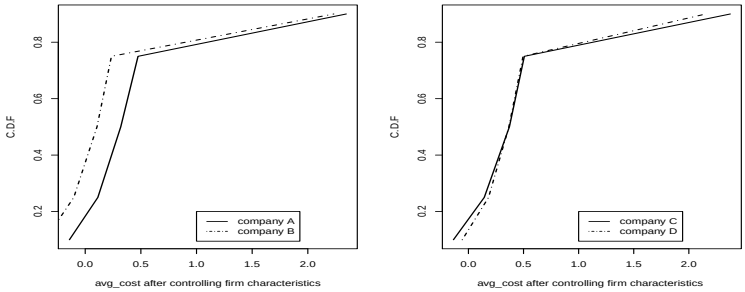

Figure 7: Comparing cost distribution for different companies- The implied cumulative distribution function of average cost for two pairs of insurers. The distribution of average cost is derived after controlling for firm characteristics