RISK MARGIN QUANTILE FUNCTION VIA PARAMETRIC AND NON-PARAMETRIC BAYESIAN APPROACHES

BY

ALICE X.D. DONG, JENNIFER S.K. CHAN AND GARETH W. PETERS

ABSTRACT

We develop quantile functions from regression models in order to derive risk margin and to evaluate capital in non-life insurance applications. By utilizing the entire range of conditional quantile functions, especially higher quantile levels, we detail how quantile regression is capable of providing an accurate estimation of risk margin and an overview of implied capital based on the historical volatility of a general insurers loss portfolio. Two modeling frameworks are considered based around parametric and non-parametric regression models which we develop specifically in this insurance setting. In the parametric framework, quantile functions are derived using several distributions including the flexible generalized beta (GB2) distribution family, asymmetric Laplace (AL) distribution and power-Pareto (PP) distribution. In these parametric model based quantile regressions, we detail two basic formulations. The first involves embedding the quantile regression loss function from the nonparameteric setting into the argument of the kernel of a parametric data likelihood model, this is well known to naturally lead to the AL parametric model case. The second formulation we utilize in the parametric setting adopts an alternative quantile regression formulation in which we assume a structural expression for the regression trend and volatility functions which act to modify a base quantile function in order to produce the conditional data quantile function. This second approach allows a range of flexible parametric models to be considered with different tail behaviors. We demonstrate how to perform estimation of the resulting parametric models under a Bayesian regression framework. To achieve this, we design Markov chain Monte Carlo (MCMC) sampling strategies for the resulting Bayesian posterior quantile regression models. In the non-parametric framework, we construct quantile functions by minimizing an asymmetrically weighted loss function and estimate the parameters under the AL proxy distribution to resemble the minimization process. This quantile regression model is contrasted to the parametric AL mean regression model and both are expressed as a scale mixture of uniform distributions to facilitate efficient implementation. The models are extended to adopt dynamic mean, variance and skewness and applied to analyze two real loss reserve data sets to perform inference and discuss interesting features of quantile regression for risk margin calculations.

Astin Bulletin 45(3), 503-550. doi: 10.1017/asb.2015.8

KEYWORDS

Asymmetric Laplace distribution, Bayesian inference, Markov chain Monte Carlo methods, Quantile regression, loss reserve, risk margin, central estimate.

1. BACKGROUND ON RISK MARGIN CALCULATION

A core component of the work performed by general insurance actuaries involves the assessment, analysis and evaluation of the uncertainty involved in the claim process with a view to assessing appropriate risk margins for inclusion in insurance liabilities. An appropriate valuation of insurance liabilities including risk margin is one of the most important issues for a general insurer. Risk margin is the component of the value of claims liability that relates to the inherent uncertainty.

The significance of this task is well understood by the actuarial profession and has been debated by both practitioners and academic actuaries alike. Much of the attention involves the non-prescriptive nature of risk margin requirements discussed in regulatory guidelines such as Article 77 and Article 101 of the Solvency II Directives. In Australia, a general task force was established, developing a report on risk margin evaluation methodologies presented to the Australian actuarial profession at the Institute of Actuaries of Australia during the 16th General Insurance Seminar in 2008. This report aimed to highlight approaches to risk margin calculations that are often considered. Before briefly discussing these aspects, we first note the following Solvency II items which relate to the Solvency Capital Requirement (SCR) and the risk margin.
Article 101 of the Solvency II Directive states,

> "*The SCR shall correspond to the Value-at-Risk (VaR) of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of 99.5% over a one-year period*".

Essentially, the basic own funds are defined as the excess of assets over liabilities, under specific valuation rules. In this regard, a core challenge is the capital market-consistent value of insurance liabilities, which requires a best estimate typically defined as the expected present value of future cash flows under Solvency II plus a risk margin calculated using a cost of capital approach.
Furthermore, under Article 77 of the 2009 Solvency II Directive, it states that the risk margin calculation is described as

> "*The risk margin shall be such as to ensure that the value of the technical provisions is equivalent to the amount insurance undertakings would be expected to require in order to take over and meet the insurance obligations... it shall be calculated by determining the cost of providing an amount of eligible own funds equal to the SCR necessary to support the insurance obligations over the lifetime thereof...*".

As can be seen from such specifications, the recommendations to be adopted are not prescriptive in the required model approaches. Therefore, as discussed in the white paper produced by the Risk Margins Taskforce, 1998 (RMT), there have been several approaches considered which range from those that involve little analysis of the underlying claim portfolio to those that involve significant analysis of the uncertainty using a wide range of information and techniques, including stochastic modeling. The RMT highlighted several approaches adopted in practice for the assessment of risk margins and pointed to percentile or quantile methods as being most prevalent in practice, this provides a good foundation for the methods we consider.

Traditionally, actuaries that adopt a stochastic framework would evaluate claims liability using a central estimate which is typically defined as the expected value over the entire range of outcomes. However, with the inherent uncertainty that may arise from such an estimator which is not statistically robust and therefore sensitive to outlier claims, claims liability measures often differ from their central estimates. In practice, the approach adopted is typically to then set an insurance provision so that, to a specified probability, the provision will eventually be sufficient to cover the run-off claims. For instance, in order to satisfy the requirement of the Australian Prudential Regulation Authority (APRA) to provide sufficient provision at a 75% probability level, the risk margin should be modeled statistically so that it can capture the inherent uncertainty of the mean estimate. When this margin is then added to the central estimate, it should provide a reasonable valuation of claims liability and therefore increases the likelihood of providing sufficient provision to meet the level required in GPS 320. In this regard, it is worth noting that the more volatile a portfolios run-offs or those that display heavy tailed features may require a higher risk margin, since the potential for large swings in reserves is greater than that of a more stable portfolio.

To accommodate these ideas, two common methods for risk margin estimation have been proposed in practice. These are the cost of capital and the percentile methods. Under the cost of capital method the actuary determines the risk margin by measuring the return on the capital required to protect against adverse development of those unpaid claim liabilities. It is evident that application of the cost of capital method requires an estimate of the initial capital to support the unpaid claim liabilities and also the estimate of return on that capital. Alternatively, under the percentile or quantile method that we consider in this paper, which is currently used in Australia, the actuary takes the perspective that the insurer must be able to meet its liability with some probability under some assumptions on the distribution of liabilities. Risk margin is then calculated by subtracting the central estimate from a predefined critical percentile value.

Since the percentile-based method involves the estimation of quantiles, it is therefore natural to consider quantile functions in risk margin calculation. Quantile regression is a statistical technique to estimate conditional quantile functions. It provides the ability to incorporate in a rigorous statistical manner, regression factors that may be driven by both exogenous features directly related to the insurance claims run-off stochastic process as well as endogenous factors that are related to the current micro or macro-economic conditions and the regulatory environment. Moreover, the model allows one to explain the proportion of variation in the risk margin allowing for accurate estimation and prediction of loss reserve.

Just as classical linear regression models based on minimizing sums of squared residuals enable one to estimate parameters for conditional mean functions in non-parametric models, quantile regression models offer a mechanism for estimating parameters for the conditional median function as well as the full range of other conditional quantile functions. Specifically, this mechanism refers to minimizing an asymmetrically weighted loss function without the need of specifying any distributional assumption. Hence, this quantile regression model is *non-parametric* in nature and robust to distributional assumption. By supplementing the estimation of conditional mean functions with techniques for estimating an entire family of conditional quantile functions, quantile regression model reveals the effect of explanatory variables on the entire conditional distribution of the response variable and not only on its center. Hence, it provides a more complete statistical analysis of the stochastic relationships between response and explanatory variables.

Quantile regression has a wide range of applications in economics and finance. In quantitative investment, least square regression-based analysis is extensively used in analyzing factor performance, assessing the relative attractiveness of different firms, and monitoring the risks in their portfolios. Engle and Manganelli (2004) consider the quantile regression for calculating VaR and constructed a conditional autoregressive VaR (CAVaR) model. VaR is a popular risk measure defined as a quantile of the loss distribution of a portfolio within a given time period at a certain confidence level. Accurate VaR estimation can help financial institutions maintain appropriate capital levels to cover the risk from the corresponding portfolio. Despite its ample application in finance, quantile regression still has limited application in a claim reserving context for risk margin estimation. We propose the use of quantile regression model to estimate the risk margin and demonstrate its advantages through real application. We highlight its features that have been popularized in finance and explain how they can be adopted in insurance for risk margin calculation. Through the proposed quantile regression framework, we develop factors and covariates that explain the risk margin variation directly. It provides a richer characterization of the data, especially when the data is heavy tailed, allowing us to consider the impact of a covariate on the entire distribution, not merely its conditional mean.

This proposed methodology of applying quantile functions to estimate risk margin differs from the traditional loss reserving approaches of developing statistical models to capture all features of the claims run-off stochastic structure. Instead, we propose to target explicitly the conditional quantile functions in a regression structure. From a statistical perspective, this is a fundamentally different approach to those traditional loss reserving models. However, we will illustrate that we can borrow from such models in developing our risk margin quantile regression framework. In fact, the associate parameter estimation via loss functions, parameter estimator properties and the resulting quantile in sample and out of sample forecasts will significantly differ to those achieved when trying to develop a model for the entire process rather than targeting a particular quantile level. This is clear from the perspective that only under a Gaussian distributional assumption for such reserve models (on log scale) would a standard least squares approach be optimal under the viewpoint of Gauss–Markov theory. In situations where returns are heavy tailed and skewed alternative models will prove more appropriate as we will discuss. On the other hand, quantile regression while focusing on just a particular quantile level is free from such sensitivity, a property in coherence with its non-parametric nature.

However, non-parametric quantile function does have its context in *parametric* regression model because the inverse cumulative distribution function (CDF) of any data distribution is itself a quantile function. A significant literature on parametric loss reserving models has been devoted to the study of the choice of appropriate distributions. For instance, Taylor (2006) estimates percentile-based risk margins via a parametric model based on the assumption of a log normal distribution of liability. Other sophisticated distributions to capture flexible shapes and tail behaviors are also proposed to model severity distribution on aggregated claim data. These distributions include the generalized-t (GT) (McDonald and Newey, 1988), Pareto (Embrechts *et al.*, 1997), the Stable family (Paulson and Faris, 1985; Peters *et al.*, 2011a, b), the Pearson family (Aiuppa, 1988), the loggamma and lognormal (Ramlau-Hansen, 1988), the lognormal and Burr 12 (Cummins *et al.*, 1999) and type II GB2 distributions (Cummins *et al.*, 1990, 1999, 2007).

Some of these distributions are defined on a real support, for instance the GT, and are flexible to model both leptokurtic and platykurtic data. However, they require log-transformation for claims data and the resulting log-linear model may be more sensitive to low values than large values (Chan *et al.*, 2008). On the other hand, the GB2 distribution family with a positive support avoids such transformation and is very flexible as it includes both heavy-tailed and light-tailed severity distributions, such as gamma, Weibull, Pareto, Burr12, lognormal and the Pearson family, hence providing convenient functional forms to model claims liability. Recently, Dong and Chan (2013) consider an alternative class of flexible skew and heavy tail models involving the GB2 distribution with dynamic mean functions and mixture model representation to model long tail loss reserving data and show that GB2 distribution outperforms some conventional distributions such as Gamma and generalized Gamma.

New distributions and models are also derived to facilitate accurate loss reserving. Peters *et al.* (2009) adopt a Poisson–Tweedie family of models which incorporates families such as normal, compound Poisson–Gamma, positive stable and extreme stable distributions into a family of models. It is shown how such a generalized regression structure could be used in a claims reserving setting to model the claims process whilst incorporating covariate structures from the loss reserving structure. In this instance, a multiplicative structure for the mean and variance functions is considered and quantiles are derived from modeling the entire distribution, rather than specifically targeting a model at the conditional quantile function. From the perspective of quantile specific regression models, Cai (2010) proposed a PP model which combines the quantile functions for both power and Pareto distributions. These combinations provide flexible quantile functions as they enable the modeling of both the main body and tails of a distribution.

For model inference, both frequentist and Bayesian approaches can be applied to estimate the traditional parametric regression model. For (nonparametric) quantile regression, Koenker and Machado (1999) show that the AL distributional family provides a useful model structure which naturally fits into a quantile regression framework. Yu and Moyeed (2001) further propose Bayesian approach via the proxy AL distribution expressed as scale mixtures of uniforms (SMU) to simplify the MCMC simulation. The benefit of using a Bayesian procedure lies in the adoption of available prior information and the provision of a complete predictive distribution for the required reserves (de Alba, 2002). For quantile regression in general, Hu *et al.* (2012) develop a fully Bayesian approach for fitting single-index models and in the context of loss reserving, Zhang *et al.* (2012) propose a Bayesian nonlinear hierarchical model with growth curves to model the loss development process, using data from individual companies forming various cohorts of claims. Ntzoufras and Dellaportas (2002) investigate various models for outstanding claims problems and show that the computational flexibility of a Bayesian approach facilitated the implementation of complex models. For non-Bayesian experts, the models can be implemented using the user friendly Bayesian software `WinBUGS`.

1.1. Contributions

The contribution of this paper is three-fold. First, we develop a range of non-parametric and parametric quantile regression models in Bayesian approach for loss reserving. These proposed models allow direct modeling of risk margin, and hence loss provision, instead of having to estimate the mean then apply a risk margin. For parametric quantile regression models, we explore different distributional assumptions with both real and positive supports, and distinct distributional features. Secondly, we generalize these distributions to incorporate dynamic structures to the mean and, where possible, the variance parameters and compare their performances. Finally, we show that the estimation of dynamic shape parameter by accident year provides an analytical framework to estimate risk margin. This allows us to capture the feature that the cohort of claims in different accident year may be heterogeneous, and hence applying different risk margin to different accident year gives us a more appropriate provision in loss reserving.

The rest of the paper is organized as follows. Section 2 explains the parametric and non-parametric models proposed. Section 3 presents the posterior quantile regression models in a Bayesian framework. Section 4 details the way to calculate risk measures and risk margin using our models. Then, we apply the methodology to two real loss reserve data sets in Sections 5 and 6. Section 7 concludes.

2. QUANTILE REGRESSION FOR CLAIMS RESERVING

In this section, we present quantile regression models and explain their relevance to loss reserving, this will be undertaken in a non-parametric and a parametric modeling framework respectively under the Bayesian paradigm. In the process, we propose a novel analytical approach to perform estimation of the risk margin under various quantile regression model structures. Of particular focus in this paper is the class of models based on the AL distributional family. In the special case of the AL distribution, we demonstrate that risk margin estimation is achieved naturally through modeling the shape parameters of the AL distribution and hence the inference on the model parameters directly informs the inference of the risk margin.

In developing a quantile function framework for general insurance claims development triangles we assume that there is a run-off triangle containing claims development data in which Y_{ij} will denote the claims with indices $i \in \{0, \ldots, I\}$ and $j \in \{0, \ldots, J\}$, where i denotes the accident year and j denotes the development year (claims can refer to payments, claims incurred, etc.). Furthermore, without loss of generality, we make the simplifying assumption that the number of accident years is equal to the number of observed development years, that is, $I = J$ with $N = \frac{1}{2}I(I+1)$ observations. At time I, the index set in the *upper* triangular is

$$\mathcal{D}_o = \{(i, j) : \ i + j \leq I + 1\}, \tag{1}$$

and for claims reserving at time I the index set to predict the future claims in the *lower* triangle is

$$\mathcal{D}_l = \{(i, j) : \ i + j > I + 1, \ i \leq I, \ j \leq I\}. \tag{2}$$

Therefore, the vector of observed Y_{ij} in the upper triangle is given by $\boldsymbol{Y}_o = \left\{Y_{ij} : (i, j) \in \mathcal{D}_o\right\}$ and the corresponding vector of covariates is denoted by $\boldsymbol{x}_o = \left\{\boldsymbol{x}_{ij} : (i, j) \in \mathcal{D}_o\right\}$. Similarly, $\boldsymbol{Y}_l = \left\{Y_{ij} : (i, j) \in \mathcal{D}_l\right\}$ and $\boldsymbol{x}_l = \left\{\boldsymbol{x}_{ij} : (i, j) \in \mathcal{D}_l\right\}$ are the vectors of claims and covariates in the lower triangle.

In the regression structures, we aim to make inference on the quantile function of the data within sample, in each cell of $\boldsymbol{Y}_o$ as well as the predictive out-off sample quantile function based on the claim cells in $\boldsymbol{Y}_l$ in lower triangle. The estimation of the quantile function has three main components:

- The conditional distribution of the regression model which defines a conditional quantile function of the dependent variables, the claims data, given the explanatory variables.
- The structural component of the regression model based on the link functions and imposed model structures linking the regression structures with the covariates to the location and scale of the conditional distribution which defines a conditional quantile functions of the response.
- The actual choice of independent variables, that is, the covariates in the regression model as well as some basis function regression structures in some of the models proposed.

In the following subsections, we discuss each of these components in turn, starting with the distributional aspects of the quantile regression models we consider.

2.1. Non-parametric quantile regression models

In a non-parametric quantile regression approach, we estimate regression coefficients without the need to make any assumptions on the distribution of the response, or equivalently the residuals. If $Y_{ij} > 0$ is a set of observed losses and $\boldsymbol{x}_{ij} = (1, x_{ij1}, \ldots, x_{ijm})$ is a vector of covariates that describe Y_{ij}, the quantile function for the log transformed data $Y^*_{ij} = \ln Y_{ij} \in \Re$ is

$$Q_{Y^*}(u|\boldsymbol{x}_{ij}) = \alpha_{0,u} + \sum_{k=1}^{m} \alpha_{k,u}\, x_{ijk}, \tag{3}$$

where $u \in (0, 1)$ is the quantile level, $\boldsymbol{\alpha}_u = (\alpha_{0,u}, \ldots, \alpha_{k,u})$ are the linear model coefficients for quantile level u which are estimated by solving the following loss function

$$\min_{\alpha_{0,u},\ldots,\alpha_{m,u}} \sum_{i,j\leq I} \rho_u(\epsilon_{ij}) = \sum_{i,j\leq I} \epsilon_{ij}[u - I(\epsilon_{ij} < 0)], \tag{4}$$

and $\epsilon_{ij} = y^*_{ij} - \alpha_{0,u} - \sum_{k=1}^{m} \alpha_{k,u}\, x_{ijk}$. Then the quantile function for the original data is $Q_Y(u|\boldsymbol{x}_{ij}) = \exp(Q_{Y^*}(u|\boldsymbol{x}_{ij}))$. Koenker and Hallock (2001) illustrate the loss function ρ_u for quantile regression as we represent in Figure 1.

Koenker and Machado (1999) and Yu and Moyeed (2001) show that the parameter estimates of $\boldsymbol{\alpha}_u$ by minimizing the loss function in (4) is equivalent to the maximum likelihood estimates of $\boldsymbol{\alpha}_u$ when Y^*_{ij} follow the AL proxy distribution with pdf

$$f(y^*_{ij}|\mu_{ij}, \sigma^2_{ij}, p) = \frac{p(1-p)}{\sigma_{ij}} \exp\left(-\frac{(y^*_{ij} - \mu^*_{ij})}{\sigma_{ij}}[p - I(y^*_{ij} \leq \mu_{ij})]\right), \tag{5}$$

where the location parameter or mode μ^*_{ij} equals to $Q_{Y^*}(u|\boldsymbol{x}_{ij})$ in (3), the scale parameter $\sigma_{ij} > 0$ and the skewness parameter $p \in (0, 1)$ equals to the quantile

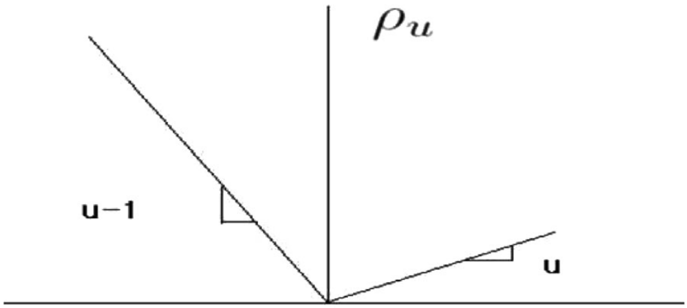

FIGURE 1: Loss function.

level u. Since the pdf (5) contains the loss function (4), it is clear that parameter estimates which maximize (5) will minimize (4).

In this formulation, the AL distribution represents the conditional distribution of the observed dependent variables (responses) given the covariates. More precisely, the location parameter μ_{ij} of the AL distribution links the coefficient vector $\boldsymbol{\alpha}_u$ and associated independent variable covariates in the linear regression model to the location of the AL distribution. It is also worth noting that under this representation it is straightforward to extend the quantile regression model to allow for heteroscedasticity in the response which may vary as a function of the quantile level u under study. To achieve this, one can simply add a regression structure linked to the scale parameter σ_{ij} in the same manner as was done for the location parameter.

Equivalently, we assume Y_{ij}^* conditionally follows an AL distribution denoted by $Y_{ij}^* \sim AL(\mu_{ij}^*, \sigma_{ij}^2, u)$. Then

$$Y_{ij}^* = \mu_{ij}^* + \epsilon_{ij}^* \sigma_{ij}, \tag{6}$$

where $\epsilon_{ij}^* \sim AL(0, 1, u)$, $\mu_{ij}^* = \alpha_{0,u} + \sum_{k=1}^{m} \alpha_{k,u}\, x_{ijk}$,
$\sigma_{ij}^2 = \exp(\beta_{0,u} + \sum_{k=1}^{v} \beta_{k,u}\, s_{ijk})$ and s_{ijk} are covariates in the variance function. Discussion on the parametric regression model, in particular, the choice of link function and structure of regression terms will be undertaken in later sections.

2.2. Parametric quantile regression models

In the parametric model based quantile regressions we detail two basic formulations. The first involves embedding the quantile regression loss function from the non-parametric setting into the argument of the kernel of a parametric data likelihood model, this is well known to naturally lead to the AL parametric model case, see Yu and Zhang (2005) and further details in Section 2.2.1.

The second formulation we utilize in the parametric setting adopts an alternative quantile regression formulation in which we specify an expression for the regression general linear model incorporating a structural trend and variance

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

function which each act to modify a base quantile function in order to produce the conditional data quantile function, see discussion on such approaches for instance in chapter 12 of Gilchrist (2002). This allows us to adopt a parametric approach to study the structure of quantile function.

Two types of distributions, on real support $\Re$ or positive support $\Re^+$ can be considered and we begin with distributions on $\Re$. In this case, we assume that $Y^*_{ij} \sim F(y^*|\boldsymbol{\theta})$ where $F(y^*|\boldsymbol{\theta})$ is the conditional CDF and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a vector of model parameters including all unknown coefficient parameters and distributional parameters. The quantile function for the conditional distribution of Y^*_{ij} given $\boldsymbol{x}_{ij}$ at a quantile level $u \in (0, 1)$ is given by

$$Q_{Y^*}(u|\boldsymbol{x}_{ij}) \equiv \inf\{y^* : \ F(y^*|\boldsymbol{\theta}) \geq u\}. \tag{7}$$

Under this formulation, the conditional quantile function in (7) can be written as

$$Q_{Y^*}(u|\boldsymbol{x}_{ij}) = \mu^*_{ij} + Q_{\epsilon^*}(u)\sigma_{ij}, \tag{8}$$

where $Q_{\epsilon^*}(u) = F^{-1}_{z^*}(u)$ is the inverse cdf for the standardized variable $Z^*_{ij} = \frac{Y^*_{ij}-\mu^*_{ij}}{\sigma_{ij}}$ and again one may incorporate regression structures given as follows for location and scale functions:

$$\textbf{location:} \quad \mu^*_{ij} = \alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ijk}, \tag{9}$$

$$\textbf{scale:} \quad \sigma^2_{ij} = \exp(\beta_0 + \sum_{k=1}^{\nu} \beta_k s_{ijk}). \tag{10}$$

We note that the parameters in (9) for parametric models are not quantile dependent as the parameters in (3). *Hence, we remark that the difference between the non-parametric and the parametric quantile functions in this paper is that in the parametric structure we make explicit the quantile function of the "residual" denoted by $Q_\epsilon(u)$ when the regression structure is still estimated at the mean (GB2) or mode (AL) of the distribution.* To transform the quantile function $Q_{Y^*}(u|\boldsymbol{x}_{ij})$ back to the original scale of the data $Y_{ij} = \exp(Y^*_{ij})$, we suggest $Q_Y(u|\boldsymbol{x}_{ij}) = \exp(Q_{Y^*}(u|\boldsymbol{x}_{ij}))$. It can be shown that the proposed transformation $Q_Y(u|\boldsymbol{x}_{ij}) = \exp(Q_{Y^*}(u|\boldsymbol{x}_{ij}))$ equals to the quantile function for the log-AL distribution.

Lemma: If Y follows a log asymmetric Laplace distribution such that $Y^* = \ln Y$ follows a asymmetric Laplace distribution with quantile function $Q_{Y^*}(u)$, the quantile function for Y is $Q_Y(u) = \exp(Q_{Y^*}(u))$.

For distributions on $\Re^+$, we assume that $Y_{ij} \sim F(y|\boldsymbol{\theta})$ with mean $\exp(\mu^*_{ij})$ where μ^*_{ij} is given in (9). Next, we make explicit several possible parametric models one may consider in quantile regressions for risk margin. Each model has different associated properties with regard to the relationship of the skewness, kurtosis and heaviness of the tail that it imposes on the quantile function of the response given the covariates.

2.2.1. *Asymmetric Laplace distribution.* As discussed above, the AL distributional family is a useful model structure which naturally fits into a quantile regression framework. As made explicit above, the AL distribution is a three parameter distribution which has been shown to be directly linked to the estimation of quantiles in a quantile regression framework, see further details in Yu and Zhang (2005).

Since this realization, the AL family has been utilized in several financial risk and econometric settings such as Guermat and Harris (2001) who use the symmetric laplace distribution with GARCH volatility to model short-horizon asset returns. Chen *et al.* (2012) extend this to allow skewness via AL distribution. Yu and Moyeed (2001) apply AL distribution for quantile regression purposes, though as yet, no such developments have been made in the insurance and particularly the risk margin context. Here, we propose such a model for risk margin estimation.

If we model the residuals ϵ_{ij} by an AL distribution, the quantile function for observed data Y^*_{ij} is given by (8) where $F^{-1}_{z^*}(u)$ is the inverse cdf (quantile function)

$$F^{-1}_{\mathrm{AL}}(u|\mu,\sigma^2,p) = \begin{cases} \mu + \frac{\sigma}{1-p}\log(\frac{u}{p}), & \text{if } 0 \le u \le p, \\ \mu - \frac{\sigma}{p}\log(\frac{1-u}{1-p}), & \text{if } p < u \le 1. \end{cases} \tag{11}$$

To understand how the three location, shape and scale parameters of the AL distribution affect the shape and tails of the distribution it is also useful to note the following relationship between the parameters and the mean, variance, skewness S and kurtosis K of AL distribution:

$$E(Y^*) = \mu + \frac{\sigma(1-2p)}{p(1-p)}, \qquad \mathrm{Var}(Y^*) = \frac{\sigma^2(1-2p+2p^2)}{(1-p)^2p^2}, \tag{12}$$

$$S(Y^*) = \frac{2[(1-p)^3 - p^3]}{((1-p)^2+p^2)^{3/2}}, \qquad K(Y^*) = \frac{9p^4 + 6p^2(1-p)^2 + 9(1-p)^4}{(1-2p+2p^2)^2}. \tag{13}$$

Note that the shape parameter p of the AL distribution gives the magnitude and direction of skewness. AL distribution is skewed to left when $p > 0.5$ and skewed to right when $p < 0.5$ and hence it can model the left skewness of most log transformed loss data directly through this shape parameter p. Moreover, as the risk margin adopted in insurance industry is mostly greater than 50%, AL distribution allows the calculation of quantiles rather than mean estimates fairly easily. Figures 2(a) and (b) show a variety of pdf for AL distribution and its skewness and kurtosis respectively.

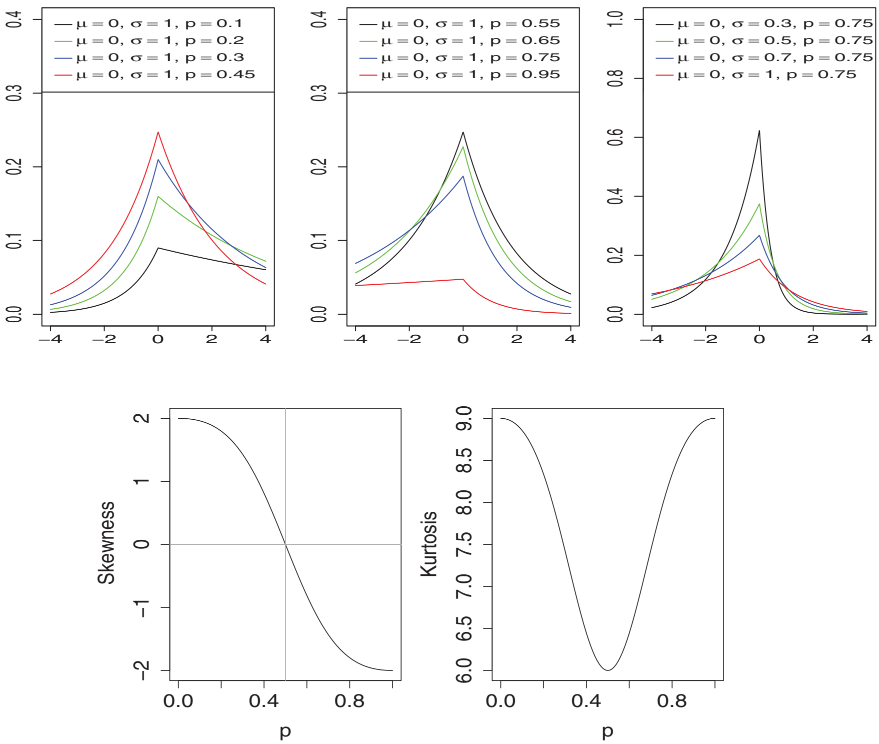

FIGURE 2: (a) The pdf of asymmetric Laplace distribution. (b) The skewness and kurtosis of asymmetric Laplace distribution. (Color online)

2.2.2. *Power-Pareto model.* As the second choice of parametric quantile regression model we consider the framework of Cai (2010). In this approach, a polynomial PP quantile function model is developed. This model combines a power distribution with a Pareto distribution, which enables us to model both the main body and the tails of a distribution. In considering the PP model, the conditional quantile function of the response (reserve in each cell) are comprised of two components:

- component 1: a power distribution $F_1(y) = y^{\frac{1}{\gamma_1}}$ where $y \in [0, 1]$ and $\gamma_1 > 0$ with a corresponding quantile function then given by $Q_1(u; \gamma_1) = u^{\gamma_1}$ for $u \in [0, 1]$; and
- component 2: a Pareto distribution function $F_2(y) = 1 - y^{-\frac{1}{\gamma_2}}$ where $y \geq 1$ and $\gamma_2 > 0$ with a corresponding quantile function then given by $Q_2(u; \gamma_2) = (1 - u)^{-\gamma_2}$.

One may use the fact that the product of the two quantile functions will remain a strictly valid quantile function producing the new quantile function family known as the polynomial-PP model. The resulting structural form given by the

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

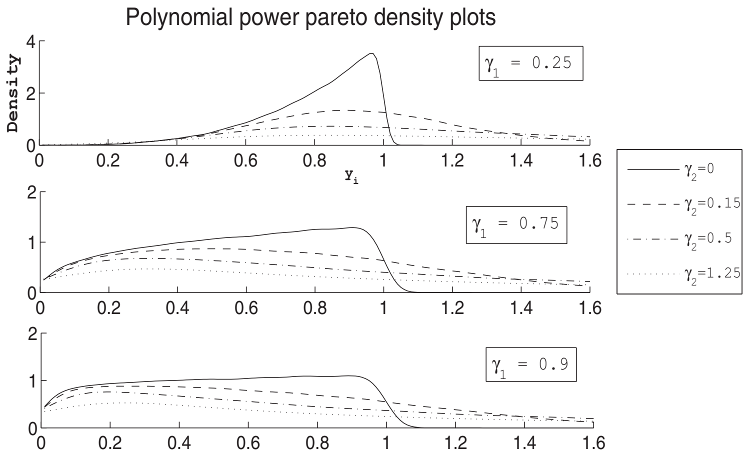

FIGURE 3: The pdf of power-Pareto distribution.

inverse cdf of the Pareto distribution with an additional polynomial power term:

$$F_{PP}^{-1}(u|\gamma_1, \gamma_2) = u^{\gamma_1}(1-u)^{-\gamma_2}. \tag{14}$$

Hence, the quantile function is again given by (8) where $Q_{\epsilon^*}(u) = F_{PP}^{-1}(u)$ and $Q_Y(u) = \exp(Q_{Y^*}(u))$.

From the specification of this quantile function, one may then derive the resulting pdf of the PP model for $Y_{ij}^* = \ln Y_{ij}$ which is given by

$$f_{PP}(y_{ij}^*|\gamma_1, \gamma_2) = \frac{u_{ij}^{1-\gamma_1}(1-u_{ij})^{\gamma_2+1}}{\sigma_{ij}[\gamma_2 u_{ij} + \gamma_1(1-u_{ij})]},$$

where u_{ij} is an implicit function of the regression structure which can be obtained by solving the system of equations defined for each observation

$$y_{ij}^* = \mu_{ij}^* + u_{ij}^{\gamma_1}\left(1-u_{ij}\right)^{-\gamma_2}\sigma_{ij}, \tag{15}$$

where again we treat the location $\mu_{ij}^* = \mu_{ij}^*(\boldsymbol{\alpha})$ in (9) and scale $\sigma_{ij} = \sigma_{ij}(\boldsymbol{\beta})$ in (10) as functions of the regression coefficients and associated covariates.

To complete the specification of the polynomial PP model, we plot the shape of the density that can be obtained for a range of different power parameters for the power and pareto components with a unit scale factor $\sigma = 1$. These plots in Figure 3 demonstrate the flexible skewness, kurtosis and tail features that can be obtained from such a model by varying the parameters γ_1 and γ_2.

2.2.3. *Generalized Beta distribution of the second type family.* The models discussed so far, that is, the AL and PP families of regression models, require a log transformation of the data before the modeling to ensure the data has real support $\Re$ that these distributions are defined upon. In performing this transformation, one must analyze carefully the effect of the transformation on the ability to fit such models and the resulting model interpretability must be interpreted with regard to the transformation. This is particularly the case if zero counts are present in the data for some accident and development years. Moreover, in the context of claims reserving, loss data often exhibits heavy-tailed behavior, particularly for long tail business classes. To account for such features and to avoid the need of pre-transformation of the data one may consider the family of GB2 distributions of the second kind.

The type two GB2 distributionGB2 has attractive features for modeling loss reserve data, as it has a positive support $\Re^+$ and nests a number of important distributions as its special cases. The GB2 distribution has four parameters, which allows it to be expressed in various flexible densities. See Dong and Chan (2013) for a more detailed description of GB2 distribution including its pdf and distribution family.

If $Y_{ij} \in \Re^+$ conditionally follows a GB2 distribution, then it can be characterized by the density given by

$$f_{\mathrm{GB2}}(y_{ij}|a, b_{ij}, p, q) = \frac{\frac{a}{b_{ij}}(\frac{y_{ij}}{b_{ij}})^{ap-1}}{B(p,q)[1+(\frac{y_{ij}}{b_{ij}})^a]^{p+q}}, \quad \text{for } y_{ij} \geq 0, \tag{16}$$

where a, p and q are shape parameters and b_{ij} is the scale parameter.

In *mean regression*, b_{ij} can be linked to the mean μ_{ij} of the distribution as follows:

$$b_{ij} = \frac{\mu_{ij} B(p,q)}{B(p+1/a, q-1/a)}, \tag{17}$$

where μ_{ij} is log-linked to a linear function of covariates μ^*_{ij} in (9) according to the relationship:

$$E(Y_{ij}) = \mu_{ij} = \exp\left(\alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ijk}\right). \tag{18}$$

Then the variance is given by

$$\mathrm{Var}(Y_{ij}) = \mu^2_{ij}\left\{\frac{B(p,q)B(p+2/a, q-2/a)}{[B(p+1/a, q-1/a)]^2} - 1\right\}. \tag{19}$$

The GB2 distribution is a generalization from the beta distribution with pdf:

$$f_B(z_{ij}|p,q) = \frac{1}{B(p,q)} z_{ij}^{p-1}(1-z_{ij})^{p+q}, \tag{20}$$

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

via the transformation $z_{ij} = \frac{(\frac{y_{ij}}{b_{ij}})^a}{1+(\frac{y_{ij}}{b_{ij}})^a}$. Hence, the cdf of GB2 distribution is given by

$$F_{GB2}(y_{ij}|a, b_{ij}, p, q) = \int_0^{z_{ij}} \frac{t^{p-1}(1-t)^{(q-1)}}{B(p,q)} dt = \frac{B(z_{ij}|p,q)}{B(p,q)} = F_B(z_{ij}|p,q) \tag{21}$$

where $B(z_{ij}|p, q)$ is the incomplete beta function.

The GB2 is directly relevant for quantile regression models since one may also find its quantile function in closed form according to the following expression:

$$Q_Y(u) = \frac{\exp\left(\alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ijk}\right) B(p,q)}{B(p+1/a, q-1/a)} \left(\frac{F_B^{-1}(u|p,q)}{1-F_B^{-1}(u|p,q)}\right)^{\frac{1}{a}}. \tag{22}$$

There are many widely known and utilized sub-families of the GB2 family, we present two examples of relevance to the context of risk margin estimation that we will explore, corresponding to the GG and the gamma distribution sub-families.

2.2.4. *Two special cases of GB2.* To understand the flexibility of the GB2 family of models, we consider the case when $q = \infty$, then the resulting GB2 distribution sub-family becomes the GG distribution, see discussion in McDonald (1984). The GG family of models was independently introduced by Stacy (1962), as a three parameter distribution with pdf given by

$$\begin{aligned} f_{GG}(y_{ij}|a, b_{ij}, p) &= \lim_{q\to\infty} \frac{\frac{a}{b_{ij}}(\frac{y_{ij}}{b_{ij}})^{ap-1}}{B(p,q)[1+(\frac{y_{ij}}{b_{ij}})^a]^{p+q}} \\ &= \frac{a(\frac{y_{ij}}{b_{ij}})^{ap}\exp[-(\frac{y_{ij}}{b_{ij}})^a]}{y_{ij}\Gamma(p)}, \quad \text{for } y_{ij} > 0 \end{aligned} \tag{23}$$

where a and p are shape parameters and b_{ij} is scale parameter linked to the mean of the distribution as

$$b_{ij} = \frac{\mu_{ij}\Gamma(p)}{\Gamma(p+1/a)}, \tag{24}$$

and the mean is again log-linked to a linear function of covariates in (18). The cdf is

$$F_{GG}(y_{ij}|a, b_{ij}, p) = \int_0^{z_{ij}} \frac{t^{p-1}e^{-t}}{\Gamma(p)} dt = \frac{\gamma_1(z_{ij}|p)}{\Gamma(p)} = F_G(z_{ij}|1, p),$$

where $\gamma_1(z_{ij}|p)$ is the lower incomplete gamma function and $z_{ij} = (\frac{y_{ij}}{b_{ij}})^a$. Hence, the quantile function is given by

$$Q_Y(u) = \frac{\exp\left(\alpha_0 + \sum_{k=1}^{m} \alpha_k x_{ijk}\right) \Gamma(p)}{\Gamma(p + 1/a)} \left(F_G^{-1}(u|1, p)\right)^{\frac{1}{a}}. \tag{25}$$

The second case is nested within the GG family and corresponds to the two parameter Gamma distribution which is obtained by further restricting $a = 1$. Its pdf and quantile function are well known and can be expressed using Equations (23) and (25) by replacing a with 1. Having defined clearly the three different quantile regression distributional families that will be considered in the parametric quantile regression framework, we will introduce the different regression structures we consider in the quantile function under each distributional assumption in the next section. Meanwhile, we provide a summary of distinct features in the quantile functions using non-parametric and parametric approaches, giving insights to actuaries to choose between the two approaches.

In summary, regression under non-parametric approach is conducted at different quantile level by minimizing the loss function in (4) condition on quantile level u. The quantile level specific regression coefficients reveal relationships between responses across quantiles, which is of significant interest in estimating risk margin and VaR in insurance and finance applications. Moreover, as there is no distribution assumption, parameter estimates are more robust to heavy tailed data. To estimate these parameters, AL distribution is adopted as a proxy distribution converting the minimization of loss function to the maximization of likelihood function for AL distribution. Theoretically, the set of parameters for non-parametric quantile regression at quantile level u are the same as those for parametric mode regression with AL distribution and shape parameter $p = u$ but their quantile functions, as given by (3) and (8) respectively, are different. An important drawback of non-parametric quantile function is its possible crossover which may occur particularly at extreme quantiles when the data are rare and are heavily weighted by the asymmetric loss function. As there is no simple solution to this problem, we believe it is important to be aware of this limitation when using this framework. On the other hand, parametric quantile functions are inverses of distribution functions. Hence, there is a clear mathematical framework for their orders across quantile levels. Moreover, parametric quantile functions can capture more dynamic features in the data through the modeling of mean/mode (mean/mode regression), variance and skewness when adopting distributions on real support. However, transformation of data is required under these distributions.

2.3. Structural components of the regression framework

In the following subsections, we explain how under each different distributional assumptions for the regression structure, one may introduce a link function to relate regression models using independent explanatory variables to the response quantiles in order to model trend behaviors in the location and scale of the quantile function. To simplify all different model considerations we consider only log link functions for distributions on positive support such as GB2 and GG, avoiding the positive constraints on regression parameters to ensure a positive mean ($\mu_{ij} > 0$). Given the set of covariates for the dependent variable quantity of interest, in this case the conditional quantile function, we assume the observations are independent.

The possible regression structures we consider will be classified as: location based explanatory factors, that is, trends in accident and development years; and scale (heteroscedasticity or variance) based explanatory factors for accident and development years. We note that when it comes to different distributional choices since we may transform the observations by taking log to fit distributions on real support such as AL and PP or consider a log link function in the mean of distributions on positive support such as GB2 and GG, we are actually considering multiplicative instead of additive terms in our regressions for the observations. As such, we explore aspects of ANOVA as well as ANCOVA regression structures in the regression setting. A summary of the model structures we consider for the location (additive in log scale for μ^*_{ij}) and scale components of each model is provided in Table 7 in Appendix B. In the context of non-parametric quantile regression, we will allow the influence of covariates to affect different quantile levels by adopting different weights in the loss function (4), making for an interesting analysis on the effect of model structure on quantile level.

We note that since the focus of this manuscript differs to that undertaken in the Poisson–Tweedie regression context of Peters *et al.* (2009), in that the focus of the regression model comparison will be primarily concerned with the model choice for the distributional form of the conditional quantile function, not so much on the model structure uncertainty related to all possible covariate model sub-space structures and nested models, therefore we limit the analysis to the ANOVA and ANCOVA structures. If one is interested in specialized techniques to explore and compare all possible models sub-spaces within each distributional model, we suggest the approach adopted in Peters *et al.* (2009) or recently in Verrall and Wuthrich (2013).

2.3.1. *Location: Development and accident year trend model structures.* The primary sets of covariates we consider correspond to the accident year and the development year in the claims reserving structure, as well as transformations of these through basis functions. From Table 7 in Appendix B, one may observe that we label models using two subscripts according to their location (mean/mode/quantile) and variance functions respectively. Models 0• (denoted by $M_{0\bullet}$) and 1• (denoted by $M_{1\bullet}$) are parsimonious location structure specifications for the general model in (9) with $m = 2$, that is, the additive structure is

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

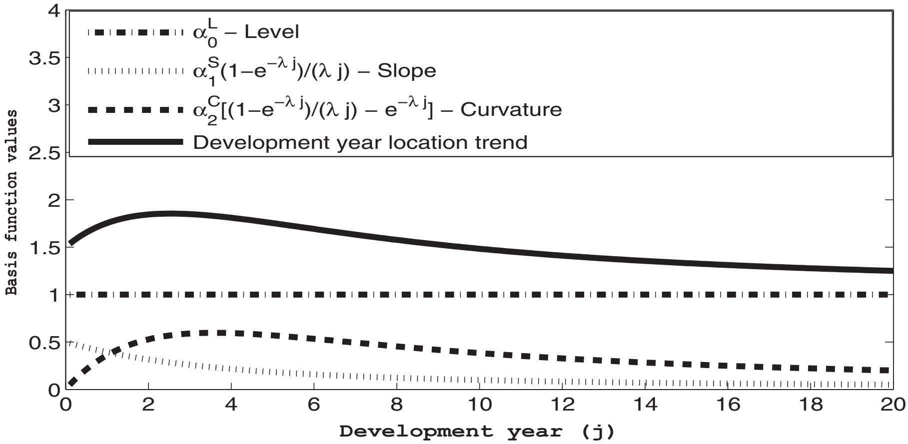

FIGURE 4: Basis function regression structure for development years in location parameter in the AL model ($M_{1\bullet}$). Decomposition of the role the level, slope and curvature basis functions play in the regression with example coefficients: $\alpha_0 = 1$, $\alpha_1^S = 0.5$, $\alpha_2^C = 2$ and $\lambda = 0.5$ with $j \in \{1, 2 \ldots, J\}$ in years.

given by:

$$\text{Model } 0\bullet: \quad \mu_{ij}^* = \alpha_0 + \alpha_1 \times i + \alpha_2 \times j, \tag{26}$$

$$\text{Model } 1\bullet: \quad \mu_{ij}^* = \alpha_0 + \alpha_1^S F_1(j) + \alpha_2^C F_2(j), \tag{27}$$

where $F_1(\cdot)$ and $F_2(\cdot)$ are basis functions defined in Table 7. Under $M_{0\bullet}$ one assumes a linear trend across accident and development years. If a nonlinear trend across development years is considered with an assumption of common behavior down the accident years, one may consider $M_{1\bullet}$ which is a basis regression model popular in term structure models and known as the Nelson–Siegel model (Nelson and Siegel, 1987). Examples of typical basis functions we considered under this choice for the location are given in Figure 4, where we show the "level", "slope" and "curvature" structure of the location trend from such a model.

In the context of an ANOVA model specification for the location, one can assume a form given by

$$\text{Model } 2\bullet: \quad \mu_{ij}^* = \alpha_0 + \alpha_{1i} + \alpha_{2j}. \tag{28}$$

This location (trend) function corresponds to the general model in (9) with $m = 2$,

$$\alpha_1 x_{ij1} = \alpha_{1i} \quad \text{and} \quad \alpha_2 x_{ij2} = \alpha_{2j}.$$

The parameters α_{1i} and α_{2j} denote the accident year and development year effects respectively and they satisfy the following constraints:

$$\alpha_{11} = \alpha_{21} = 0. \tag{29}$$

This parametrization is set up in the context of loss reserving so that all parameters are relative to the first accident year which has the most information. These location functions (26) to (28) apply to both AL and PP distributions in general. For Gamma, GG and GB2 distributions with positive support, a log link function is considered and the location functions become $\mu_{ij} = \exp(\mu^*_{ij})$. When the AL distribution, with the shape parameter $p = u$ is applied, Model 3• ($M_{3\bullet}$) corresponds to a non-parametric model given by

$$\text{Model } 3\bullet\text{:} \quad \mu^*_{ij,u} = \alpha_{0,u} + \alpha_{1i,u} + \alpha_{2j,u}, \tag{30}$$

where $\alpha_{\bullet,u}$ are parameters at quantile level u. Then, the quantile functions are given by $Q_Y(u|\boldsymbol{x}_{ij}) = \exp(Q_{Y^*}(u|\boldsymbol{x}_{ij})) = \exp(\mu^*_{ij,u})$ from (3).

2.3.2. *Scale: Development and accident year variance model structures.* There are different choices for the structure of the variance function for the AL and PP distributions but Gamma, GG and GB2 distributions do not have a component to model σ^2 directly. Model • 0 ($M_{\bullet 0}$) assumes homoscedastic variance $\sigma^2_{ij} = \sigma^2$. Models •0 ($M_{\bullet 0}$) to •3 ($M_{\bullet 3}$) are specified as below:

$$\text{Model } \bullet 0: \quad \sigma^2_{ij} = \sigma^2, \tag{31}$$

$$\text{Model } \bullet 1: \quad \sigma^2_{ij} = \exp(\beta_0 + \beta_{1i}), \tag{32}$$

$$\text{Model } \bullet 2: \quad \sigma^2_{ij} = \exp(\beta_0 + \beta_{2j}), \tag{33}$$

$$\text{Model } \bullet 3: \quad \sigma^2_{ij} = \exp(\beta_0 + \beta_{1i} + \beta_{2j}), \tag{34}$$

where the parameters β_{1i} and β_{2j}, which denote the accident year and development year effects respectively, satisfy the following constraints:

$$\beta_{11} = \beta_{21} = 0. \tag{35}$$

Again, Models •1 to •3 correspond to (10) with $\beta_1 s_{ij1} = \beta_{1i}$ and $\beta_2 s_{ij2} = \beta_{2j}$. Furthermore, for Model 23′, the shape parameter in the AL distribution is further modeled by the accident year effect, which is specified as follows:

$$\text{Model } 23': \quad p_i = \phi_0 + \phi_{1i}. \tag{36}$$

where the parameters ϕ_{1i} denote the accident year effect and satisfy the following constraints:

$$\phi_{11} = 0. \tag{37}$$

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

3. Bayesian framework: Posterior quantile regression

The estimation of quantile regression models is straightforward to adopt under a Bayesian formulation. One of the key advantage of using Bayesian procedures for practical models such as those we develop above lies in the adoption of available prior information and the provision of a complete predictive distribution for the required reserves (de Alba, 2002).

To complete the posterior distribution specification in each model it suffices to consider the representation of two components: the likelihood of the data for the regression structure (that is, the density not the quantile function); and the prior specifications for the model parameters. In the above sections, the quantile function of the likelihood is presented, along with the associated density for the observations conditional upon the parameters and covariates, that is, the likelihood for each model. Therefore, to formulate the Bayesian structure we simply need to present the prior structures we consider for the parameters in each model. This will be relatively straightforward for models formed from the AL distribution structure and the GB2 structures, but not so trivial for the case of the PP model.

In the real data examples we consider, we adopt an objective Bayesian perspective in which we consider relatively uninformative priors. This reflects our lack of prior knowledge for the model parameters likely ranges or magnitudes. For instance, the priors for parameters (coefficients) in mean, variance and skewness quantile regression functions prior to the link transformations are all selected as Gaussian

$$\alpha_0,\ \alpha_1,\ \alpha_1^S,\ \alpha_{1i},\ \alpha_2,\ \alpha_2^C,\ \alpha_{2j},\ \beta_{1i},\ \beta_{2j},\ \phi_0,\ \phi_{1i} \sim N(0, 100), \tag{38}$$

and for the shape parameters of the GB2 distribution are

$$a \sim N(0, 100), \qquad p \sim Ga(0.001, 0.001), \qquad q \sim Ga(0.001, 0.001). \tag{39}$$

Normal and gamma distributions are standard choices of priors for parameters with a real and positive support respectively, see discussions on possible choices in Denison *et al.* (2002). In the case of the AL and GB2 models, these priors combined with the resulting likelihoods produce in each case standard and well-defined posterior distributions.

In the case of the PP model, one has to be careful to define the posterior support to ensure the resulting distribution is normalized and therefore a proper posterior density. To ensure this is the case, one must impose constraints on the prior support which can be uniquely characterized by the three sets of parameter space constraints Ω_1, Ω_2 and Ω_3, for coefficient vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and (γ_1, γ_2)

respectively, given by:

$$
\begin{aligned}
\Omega_1 &= \left\{ \left(\alpha_{0,u}, \ldots, \alpha_{m,u}\right) \;:\; \alpha_{0,u} + \sum_{k=1}^{m} \alpha_{k,u} x_{ijk} < y_{ij}, \quad \forall i, j \in \{1, 2, \ldots, I\} \right\}, \\
\Omega_2 &= \left\{ \left(\beta_{0,u}, \ldots, \beta_{\nu,u}\right) \;:\; \beta_{0,u} + \sum_{k=1}^{\nu} \beta_{k,u} s_{ijk} > \epsilon > 0, \quad \forall i, j \in \{1, 2, \ldots, I\} \right\}, \\
\Omega_3 &= (0, M] \times (0, \infty), \quad M \in \Re^{+}.
\end{aligned} \tag{40}
$$

Under these parameter space restrictions the resulting posterior for the PP model can be shown to be well defined as a proper density, see a derivation and proof in Theorem 1 of Cai (2010).

In Cai (2010), they consider an MCMC scheme for the resulting posteriors based on standard Metropolis–Hastings steps with rejection when the proposed parameter values fail to satisfy the posterior support constraints. In general, this results in a very slowly mixing MCMC chain which will have very poor properties. We replace this idea with simple block Metropolis within Gibbs updates which allow for smaller moves in each component of the constrained posterior support making it more likely to satisfy the constraints and also simpler to design and tune the proposal for the MCMC scheme. This was a significant improvement compared to the approach proposed in Cai (2010). We implement this sampler in `R`. For the other Bayesian models from the AL and GB2 models, sampling from the intractable posterior distributions involved the Gibbs sampling algorithm (Smith and Roberts, 1993; Gilks *et al.*, 1996) and Metropolis–Hastings algorithm (Hastings, 1970; Metropolis *et al.*, 1953) which are the most popular MCMC techniques. For readers who are less familiar with Bayesian computation techniques, we suggest the `WinBUGS` (Bayesian analysis Using Gibbs Sampling) package, see Spiegelhalter *et al.* (2002). The MCMC algorithms that are implemented for each model in `WinBugs` and `R` are available upon request.

In the Gibbs sampling scheme, a single Markov chain is run for 60,000 to 1,10,000 iterations, discarding the initial 10,000 iterations as the burn-in period to ensure convergence of parameter estimates. Convergence is also carefully checked by the history and autocorrelation function (ACF) plots. Then, every 10th simulated value from the Gibbs sampler after the burn-in period was selected to produce a down-sampled set of samples that will mimic a random sample of size 10,000 from the joint posterior distribution for inference purposes. Parameter estimates are given by the sample means, which will be the posterior mean under a quadratic loss function.

4. Quantile prediction for risk measures and risk margin

There are two different places in claims reserving where knowledge of the predicted quantile function obtained from our regression framework can be directly of use. The first is related to a risk margin adjustment and the second is related to risk measure calculations such as those that may be required for SCR under Article 101 of the Solvency II Directive.

It is common in practice to perform estimation of the predicted reserves by predicting the mean reserve in each cell in the lower triangle $\mathcal{D}_l$ and then obtaining a total mean reserve. In such settings, it is typical to then make an adjustment to the estimated reserve to allow for some amount of uncertainty in this prediction that could come from a range of different sources. One way to achieve this is to adopt a risk margin adjustment. In such a setting, one takes the estimated reserve and makes a risk margin adjustment based on knowledge of the quantile function of the distribution of the predicted reserves at a regulatory approved quantile level of significance.

It is important to note that such forms of risk margin adjustment have been the focus of regulatory discussions in recent months. For instance, it was noted in September 2014 by the UK's Bank of England Prudential Regulation Authority that:

"*The risk margin is designed to ensure that the overall value of a firm's technical provisions is equivalent to the amount that would be expected to be required in order that a third party can take over and meet the insurance and reinsurance obligations of the firm*".

As such the risk margin can be seen as being designed not as a substitute for, insurers' normal capital requirements, instead it relates only to the non-hedgeable risks of cash flows, such as operational, underwriting and certain credit risks. Its main purpose is to protect against worse than expected outcomes. The risk margin should ensure that insurers have sufficient assets to safely wind up and transfer obligations to a third party in the event of insolvency

A second place where knowledge of the predictive quantile function of the reserves is relevant is for evaluation of a risk measure. A common approach to reserving based on the mean reserve that is complementary is to also report, along with the mean reserve, a risk measure. This would involve evaluating an alternative estimator for the reserve based on for instance the quantification of a risk measures formed from the quantile function of the distribution of the predicted reserves. Examples of such alternatives that include specifically information on the quantile function include the VaR at some specified quantile level, which is known in Solvency II as the SCR, often then calculated over a predefined time frame, see discussions in Article 101 of the Solvency II Directive discussed in the introduction. Other risk measures that could also be considered and can be obtained from knowledge of the predictive quantile function include the expected shortfall (ES) and spectral risk measures (SRM), see discussions on properties of such quantile based risk measures for capital and reserving in for instance Embrechts *et al.* (1997), Artzner (1999), Dowd *et al.* (2006), Delbaen (2002) and Peters *et al.* (2013) and the references therein. It is also worth noting that the Solvency II Directive clearly defines the risk margin and the SCR as separate concepts.

When calculating any of these required measures for the resulting total outstanding reserves (OR) one requires to first obtain a quantity such as the predictive density. The pure Bayesian approach to achieve this is obtained from the following for each $Y_{ij} \in \mathcal{D}_l$:

- **Full predictive posterior distribution:**

$$F_{Y_{ij}}\left(y_{ij}|\mathcal{D}_0\right) = \int_0^{y_{ij}} f_{Y_{ij}}\left(y|\mathcal{D}_0\right)\, dy = \int_0^{y_{ij}} \int f_{Y_{ij}}\left(y|\boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}|\mathcal{D}_0\right)\, d\boldsymbol{\theta}\, dy.$$

  Here, all posterior parameter uncertainty is integrated out of the predictive distribution.

In practice, typically performing this integration for the predictive distribution cannot be done in closed form and is typically approximated using MCMC sampling procedures. There are also alternative methods that can be considered to obtain the predictive distribution. One such alternative approach is often referred to as Empirical Bayes (Carlin and Thomas, 2000). Under this approach, one would first obtain a posterior point estimator for model parameters from the posterior conditional on the upper triangular data, denoted generically by $\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)$. This could be the posterior mean, mode, median or some other desired point estimator from the posterior $\pi(\boldsymbol{\theta}|\mathcal{D}_0)$. This would then be substituted to obtain the approximate predictive distribution. We note that some authors prefer to refer to empirical Bayes as maximum marginal likelihood.

A frequentist approach that is often considered in practice, would involve first obtaining a point estimator for model parameters based on the data in the upper triangular claims triangle, denoted by $\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)$. In general, under a frequentist approach, this estimator may be formed from minimization of a desired loss function. We note that this differs from Empirical Bayes in that the estimator obtained is not based on a prior on the parameters.

In either case, the generically denoted estimator $\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)$, would then be substituted to the model to obtain future predictions. This would produce the expression given as follows, with different interpretations for how one obtained $\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)$.

- **Conditional predictive distribution:**

$$F_{Y_{ij}}\left(y_{ij}|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right) = \int_0^{y_{ij}} f_{Y_{ij}}\left(y|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right)\, dy.$$

Using one of these predictive distributions, one may also be interested in quantities such as the distribution of the total outstanding claim given by the sum of the losses in the lower triangle according to the random variable $Y_T := \sum_{(i,j)\in\mathcal{D}_l} Y_{ij}$ which has distribution given under the full predictive distribution according to convolution given by

$$\begin{aligned} F_{Y_T}\left(y_t|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right) &:= *_{(i,j)\in\mathcal{D}_l} F_{Y_{ij}}\left(y|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right) \\ &= \left(F_{Y_{I,2}} * F_{Y_{I-1,3}} * F_{Y_{I-2,4}} * \cdots * F_{Y_{I,I}}\right)\left(y|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right). \end{aligned} \tag{41}$$

where, one convolves the distributions for the loss elements in the lower triangle with $*$ the convolution operator. One can then state several features about the tail behavior of the total loss distribution and also therefore of the high quantiles as $y \to \infty$, depending on the properties of the individual loss random variables in the sum.

At this stage, we observe that there is a rich literature on quantile approximations and asymptotics that have been studied in the actuarial context. In particular, there is an actively growing literature on the highly related topic of quantile and conditional quantile asymptotics, see for instance detailed discussions of relevance to the actuarial context in Beirlant *et al.* (2006), Goegebeur *et al.* (2014), Borovkov and Borovkov (2008), Kluppelberg and Mikosch (1998), Daouia *et al.* (2012) and Ori and Cohen (1996)

Here, we consider a simplified discussion on such ideas applied to the quantile regression context considered in this paper. This is not aimed to be a major component of the study as the general aim of this paper is to discuss the context of Bayesian quantile regression, instead we give a brief glimpse of how these methods can complement the ideas we present and can be developed further in future studies dedicated to studying further this relationship.

For instance, if one has loss distributions on $\Re^+$ then one can obtain the lower bound given by

$$\begin{aligned} \overline{F_{Y_T}}\left(y_t|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right) &:= \overline{\left(F_{Y_{I,2}} * F_{Y_{I-1,3}} * F_{Y_{I-2,4}} * \cdots * F_{Y_{I,I}}\right)}\left(y|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right) \\ &\sim c \sum_{(i,j)\in\mathcal{D}_l} \overline{F_{Y_{ij}}}\left(y|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right), \quad \text{as } y \to \infty, \end{aligned} \tag{42}$$

for some $c \geq 1$. Note, if at least one of the lower triangle losses Y_{ij} is distributed according to a heavy tailed loss distribution, such as sub-exponential, regularly varying or long tailed loss distributions then one can find the precise value for c. For instance, if the total loss is max-sum equivalent, then $c = 1$, see definitions for regular variation, sub-exponential, long tailed and max-sum equivalence in Bingham *et al.* (1989) and in the context of insurance and quantile function approximations as discussed here, see the recent tutorial and references therein from Peters *et al.* (2013).

These conditional predictive distributions can be obtained for any model approximately by solving the integrals using the MCMC samples obtained from the posterior $\pi\left(\boldsymbol{\theta}|\mathcal{D}_0\right)$. Then, given a predictive distribution, one can then find quantile functions according to the following approaches:

- **Full predictive posterior quantile function:** is given by $Q_{Y_{ij}|\mathcal{D}_0}(u) := F_{Y_{ij}}^{-1}\left(y_{ij}|\mathcal{D}_0\right)$ which is the solution to the second order ordinary differential equation

$$\frac{d}{dQ_{Y_{ij}|\mathcal{D}_0}} f_{Y_{ij}}\left(Q_{Y_{ij}|\mathcal{D}_0}(u)\,|\mathcal{D}_0\right)\left(\frac{dQ_{Y_{ij}|\mathcal{D}_0}}{du}\right)^2 + f_{Y_{ij}}\left(Q_{Y_{ij}|\mathcal{D}_0}(u)\,|\mathcal{D}_0\right)\frac{d^2Q_{Y_{ij}|\mathcal{D}_0}}{du^2} = 0,$$

which is obtained by twice differentiating the following identity:

$$F_{Y_{ij}}\left(Q_{Y_{ij}|\mathcal{D}_0}(u)\,|\mathcal{D}_0\right) = \int_0^{Q_{Y_{ij}|\mathcal{D}_0}(u)} f_{Y_{ij}}\left(y|\mathcal{D}_0\right)dy = u. \tag{43}$$

The solution to this second order ordinary differential equation can often be found in the form of a power series, see discussions in Gyorgy and Shaw (2008).

- **Conditional predictive quantile function:**

$$Q_{Y_{ij}|\widehat{\theta}(\mathcal{D}_0)}(u) := F_{Y_{ij}}^{-1}\left(u|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right), \tag{44}$$

which is the most convenient choice that we recommend since the inverse of the predictive distribution in this case takes the closed form expressions for the particular model considered as detailed in Section 2.2.

- **Conditional total reserve quantile function:** In many cases, one is also interested in finding the quantile function of the distribution corresponding to the total reserve, which under conditional independence is given by $F_{Y_T}^{-1}\left(y_t|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right)$ where this is given by the quantile function of the distribution in equation (41). In general, finding the convolution and inverse of this convolved distribution must be done numerically. There are many basic results known about these quantities such as asymptotic results and bounds for different properties of light and heavy tailed random variables, independent or dependent, see a discussion in Kaas *et al.* (2000).

4.1. Light tailed run-off for claims process

In the case in which no loss cells in the claims triangle are heavy tailed, then in general, one would need to approximate the tail quantile for the partial sum of all losses. In Kaas *et al.* (2000), they study partial sums of random variables with no assumption of independence or of identical marginal distributions. The only assumption is that the tails are not so heavy for each marginal, such that each marginal has finite mean. It will be useful to recall that for two random variables X and Y, X proceeds Y under convex ordering $X \leq_{CX} Y$ iff for all convex real functions $g(\cdot)$ with finite expectations one has

$$\mathbb{E}\left[g(X)\right] \leq \mathbb{E}\left[g(Y)\right]. \tag{45}$$

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

Thus, two random variables X and Y with equal mean are convex ordered if their cdfs cross once.

Then, one can show that in such cases for any sequence of loss distributions $\{F_{Y_{ij}}\}_{(i,j)\in\mathcal{D}_l}$ the following convex order relationship holds

$$\sum_{(i,j)\in\mathcal{D}_l} Y_{ij} \leq_{CX} \sum_{(i,j)\in\mathcal{D}_l} F^{-1}_{Y_{ij}}(U), \tag{46}$$

for $U \sim$ uniform[0, 1], see derivations in Goovaerts *et al.* (2000). This result means that the total loss Y_T in the convex order sense, comprised of the most risky joint vector of losses with given marginals, has the comonotonous joint distribution. The components of which are maximally dependent since all components are non-decreasing functions of a common random variable U. Hence, we consider the following quantile function approximation for the total loss based on the most conservative estimate using the above bound, given by

$$F^{-1}_{Y_T}(u) = \sum_{(i,j)\in\mathcal{D}_l} F^{-1}_{Y_{ij}}(u). \tag{47}$$

4.2. Heavy tailed run-off for claims process

Alternatively, if additional features of the loss distributions in the lower triangle are known, such as the loss models contain at least one heavy tailed loss distribution, then one can bound the total quantile function result. This can be done conservatively by instead considering the $\mathcal{T}$-fold convolution of the distribution, say $F^{(*\mathcal{T})}_{Y_{i*j*}}$ which correspond to the loss distribution amongst all the lower triangular loss elements with the dominant index of regular variation (that is, with the heaviest tails). In such cases, it would be popular to utilize an asymptotic result for the quantile function of the sum, as the quantile level becomes large $u \to 1$. For instance, one could use the first order or second order asymptotic results, see discussions in Peters *et al.* (2013) and Cruz *et al.* (2014). As an example, if the quantile regression was structured such that the distribution of the partial sum $Y_T = \sum_{(i,j)\in\mathcal{D}_l} Y_{ij} \sim F_{Y_T}$ is regularly varying with index $\rho \geq 0$ with conditionally independent and identically distributed (i.i.d.) Y_{ij} with each Y_{ij} taking positive support, then one can write the first order tail approximation which is asymptotically equivalent to the following

$$\overline{F}_{Y_T}(y) \sim \mathcal{T}\overline{F}_{Y_{i*j*}}(y), \quad y \to \infty, \tag{48}$$

see detailed tutorial in Peters *et al.* (2013). This would lead to the approximation of the required quantile asymptotically by the expression

$$\begin{aligned} Q_{Y_{\mathcal{T}}|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)}(u) &:= \inf\left\{y \in \mathbb{R}^+ : \ F_{Y_{\mathcal{T}}}(y) > u\right\} \\ &\approx \inf\left\{y \in \mathbb{R}^+ : \ \mathcal{T}\overline{F}_{Y_{i*j*}}(y) < 1-u\right\} \\ &\approx Q_{Y_{i*j*}|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)}\left(1-\frac{1-u}{\mathcal{T}}\right) := F^{-1}_{Y_{i*j*}}\left(1-\frac{1-u}{\mathcal{T}}|\widehat{\boldsymbol{\theta}}(\mathcal{D}_0)\right). \end{aligned} \tag{49}$$

5. Model structure analysis for Israel data

In this section, we perform two core studies: The first involves isolating the structural components for the quantile regressions, in order to perform a study on the mean function and variance functions that are most suitable for an example of a representative claims reserving data set. This is therefore performed using Bayesian formulations of the AL model with different assumptions on the mean and variance functions in both parametric and non-parametric approaches. The second involves isolating the distributional choices of the regression model, where we take the best fitting parametric model mean and variance function structures and use these to study distributional properties under the different quantile function choices.

The data set used throughout this section is interesting for such a benchmark exercise as it has been previously studied and its features are reasonably well known, see Chan *et al.* (2008) for more details on the Israel data set. The data are available in Figure 17 in Appendix A and represent the paid out claim amounts y_{ij} for an Israel insurance company, covering periods from 1978 to 1995, containing 171 observations. For mathematical convenience, two zero claim amounts have been replaced with 0.01. Some general trends are observed in this data. Given an accident year, the claim development amounts generally increase till the first 4 to 6 development years then this increase is followed by a generally decreasing trend thereafter. The mean, median, variance and kurtosis of this data are 4459.7, 3,871, 12,059,232.6 and -0.4 respectively. The overall skewness is 0.58 and on a log scale is -2.67.

This data has been studied in Chan *et al.* (2008) using the GT distribution expressed as scale mixtures of uniform which facilitates the Bayesian implementation. They adopt the ANOVA and ANCOVA mean structures to study the accident year and development year effects on the conditional mean functions but not on any quantile level. Moreover, they also remark that the log transformed data become negatively skewed which the symmetric GT distribution fails to accommodate. Hence, they suggest to adopt some skewed error distributions to improve the model performance.

Our primary point of departure for these previous studies on this data is the conjecture that using a measure of average effects may not be appropriate for understanding loss reserves at higher quantiles. Higher quantile projection is critical in loss reserving, for reinsurance premium calculations and also in deriving the risk margin. In this section, we use all the models in Section 2 for quantile projection with an aim to provide a more comprehensive study on model performance with a wide range of distributions having different tails behavior and model structures for the quantile trends and heteroscedasticity in the accident and development years.

5.1. Analysis of quantile regression models: Location and scale

To investigate the model structures for location (mean/mode/quantile) and scale (variance) functions, we consider two settings: the first class of models involves the *parametric* models using the AL distribution with p left to be estimated (denoted by est), the *mode* functions given by (26) to (28) and variance being constant (31) (Models 00–20) or given by (34) (Models 03–23); the second class of models involves a set of *non-parametric* models which are also studied with *quantile* function (30) and variance being constant (31) or given by (34) (Models 30 and 33) using AL as a proxy distribution with p fixed (denoted by fix) at different quantile levels.

For model comparison, deviance information criterion (DIC) is adopted, as detailed in for instance Claeskens and Hjort (2008) or Spiegelhalter *et al.* (2002). In general, one can consider the DIC as a hierarchical model generalization of other popular information criterions typically used for model selection such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). DIC is widely used in Bayesian model selection problems especially in cases when the posterior distributions of the models are obtained by MCMC sampling, as adopted in this paper.

Since, models with smaller DIC are preferred to those with larger DIC, then the results of the model comparisons provided in Table 1 show that among the parametric models in which p is estimated, M_{23} with an ANOVA model for both accident and development years in modeling both the mode and variance functions is the best fitting model according to DIC. This result is further supported by the two measures, namely the posterior mean deviance $\bar{D}$ and the deviance evaluated at the posterior mean of parameters $\hat{D}$ as defined under Table 1 which indicate model fit without model complexity penalty as in DIC. This shows that the accident year and development year effects are both important in describing the dynamics of the location and variance. Hence, these ANOVA-type location and variance functions are applied to most of the subsequent analyses whenever possible. For the non-parametric models in which p is fixed at various quantile levels, M_{33} with ANOVA variance provides better fit than M_{30} with constant variance.

Between parametric model M_{23} and non-parametric models M_{33}, the non-parametric models provide better model performance according to DIC. These

TABLE 1

ESTIMATES OF p AND MODEL FIT MEASURES FOR AL PARAMETRIC AND NON-PARAMETRIC MODELS. (MODELS DETAILED IN APPENDIX B, TABLE 7).

Models	DIC	$\bar{D}^{\dagger}$	$\hat{D}^{\ddagger}$	p	Models	DIC	$\bar{D}^{\dagger}$	$\hat{D}^{\ddagger}$	p
	Variance Constant					Variance Function			
M_{00}	195.41	255.21	315.02	0.85 (est)	M_{03}	272.82	334.74	396.66	0.93 (est)
M_{10}	223.30	284.10	344.91	0.88 (est)	M_{13}	199.14	247.49	295.85	0.95 (est)
M_{20}	50.94	120.17	189.40	0.81 (est)	M_{23}	–20.81	24.91	70.63	0.75 (est)
M_{30}	55.94	125.61	195.28	0.30 (fix)	M_{33}	–37.06	38.34	113.74	0.30 (fix)
M_{30}	73.10	152.26	231.43	0.50 (fix)	M_{33}	–38.80	35.51	109.82	0.50 (fix)
M_{30}	55.26	132.56	209.87	0.75 (fix)	M_{33}	–17.33	53.40	124.12	0.75 (fix)
M_{30}	44.86	116.38	187.91	0.95 (fix)	M_{33}	–64.26	3.68	71.62	0.95 (fix)

† $\bar{D}$ is the posterior mean deviance $E_{\theta}[-2\log f(\boldsymbol{y}|\boldsymbol{\theta})]$; ‡ $\hat{D} = -2\log f(\boldsymbol{y}|\bar{\boldsymbol{\theta}})$ where $\bar{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$.

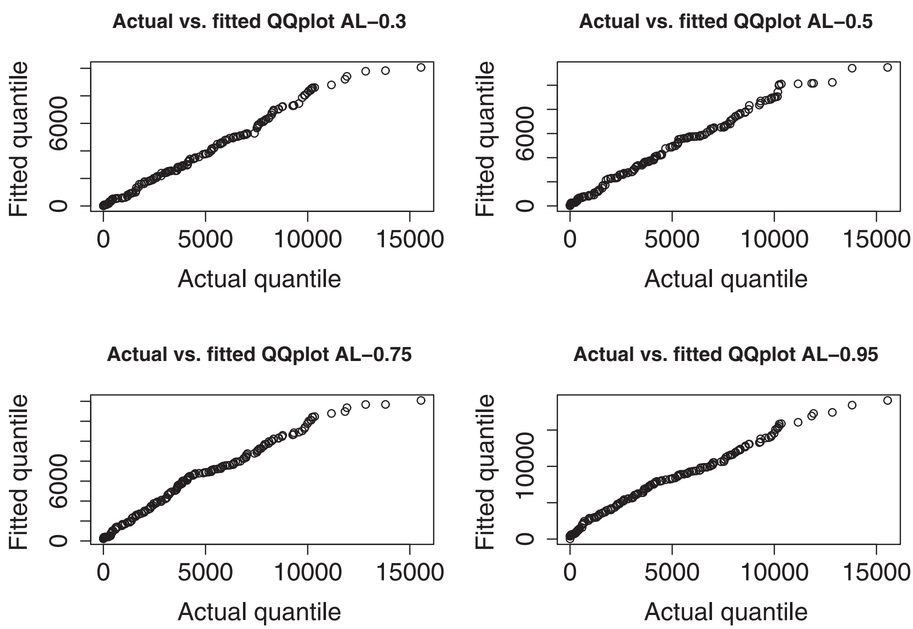

FIGURE 5: QQ plot for non-parametric models M_{33} at different quantile levels.

M_{33} models correspond to the AL models with location and variance -functions for a range of fixed quantile levels $p \in \{0.3, 0.5, 0.75, 0.95\}$. Then, we study their performances using quantile–quantile (QQ) plot as shown in Figure 5. The QQ plot shows linearity except for a few very extreme quantiles and indicates appropriate fits from the specified model structures for this range of quantile levels.

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

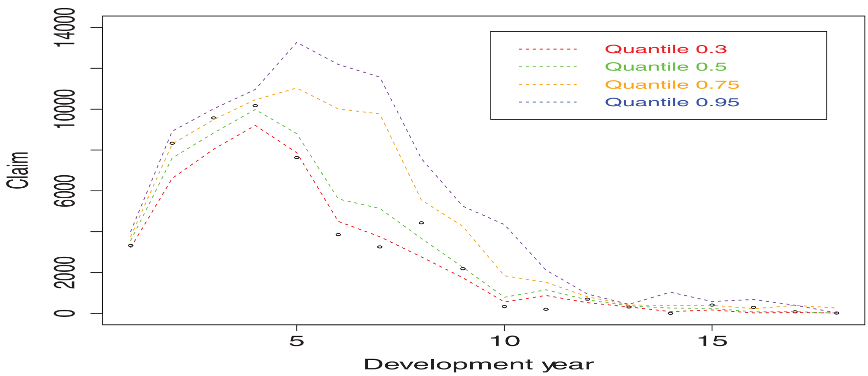

FIGURE 6: Fitted loss of the first accident year across quantiles using M_{33} with AL distribution. (Color online)

In addition, we investigate the trends of development year effects as depicted in Figure 6 which reports the fitted loss $\widehat{Y}_{1j} = \exp(\mu^*_{1j})$ where μ^*_{1j} is given by (30) and calculated using the conditional predictive quantile function in (44) for the first accident year ($i = 1$). The quantile levels u correspond to the shape parameter p set to 0.3, 0.5, 0.75 and 0.95 respectively in AL distribution. The figure demonstrates that there is a clear requirement for a nonlinear trend in the development year covariate at all quantile levels. Hence, the trend of fitted losses increases uniformly until $j = 4$ and subsequently decreases thereafter at all quantile levels and they agree closely with the observed trend particularly at lower quantile levels.

To conclude the benchmark analysis on model structure, we also present for the best model M_{33} with quantile and variance functions the estimated model trends for all accident years, depicted in Figure 7 as four triangular heat maps. The heat maps each depicts the fitted loss by accident and development years in the upper triangle at all four quantile levels, where the first row corresponds to that which was studied in Figure 6. All heat maps show a consistent trend across development years for all accident years and quantile levels with high levels of loss as indicated by light colours being around the fourth development year, particularly for lower accident years. With increasing quantile levels, the width of light colours for each accident year increases showing higher levels of fitted losses around the peak.

Although non-parametric models have lower DIC values in general, Table 1 shows that parametric model M_{23} actually provides comparable model fit according to $\bar{D}$s before model complexity penalty was applied. This is because parametric models with additional shape parameters are subject to heavier model complexity penalty in DIC. However, it should be noted that parametric models provide better model fit in general over a range of models and quantile levels. For example, $\bar{D}$ for M_{23} is 24.91 which is less than those of M_{33} except when $p = 0.95$. In addition, the parametric models have additional advantages that they will be more readily interpretable as well as directly

TABLE 2

PARAMETER ESTIMATES AND MODEL FIT MEASURES FOR AL MODELS WITH ANOVA MEAN AND VARIOUS VARIANCE FUNCTIONS.

Models	DIC	$\bar{D}$	$\hat{D}$	MSE	p	σ^2
$\mathcal{M}_{20}$	50.94	120.17	189.40	1015.71	0.80	0.02
$\mathcal{M}_{21}$	−4.32	56.66	117.64	849.91	0.74	0.04
$\mathcal{M}_{22}$	6.63	54.29	101.95	755.66	0.68	0.19
$\mathcal{M}_{23}$	−20.81	24.91	70.63	850.10	0.75	0.17

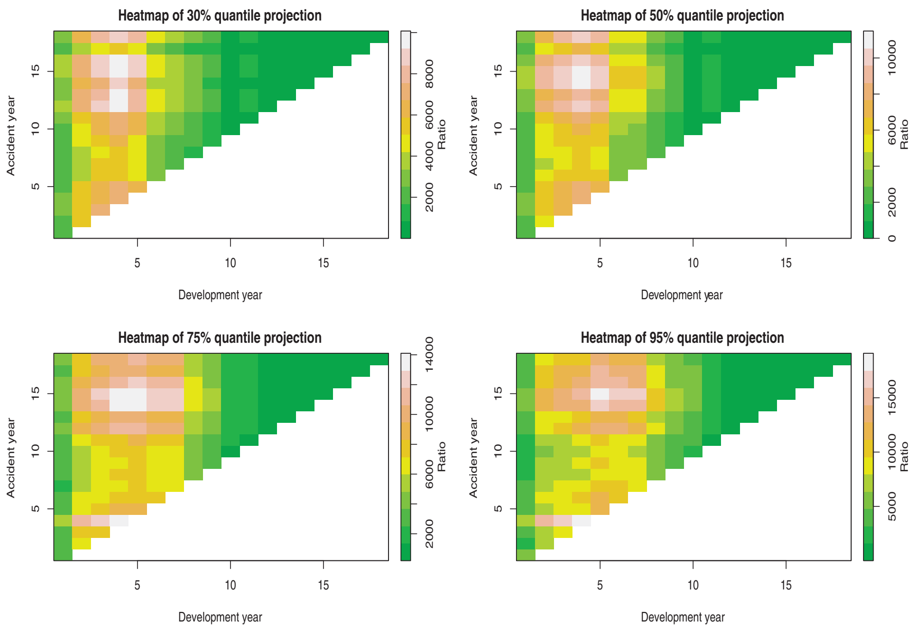

FIGURE 7: Fitted loss of the upper triangle across quantiles using $\mathcal{M}_{33}$ with AL distribution. (Color online)

usable when calculating risk margins and quantile based risk measures with no crossover of quantile functions, as was discussed in Section 2.2. For the mode structure corresponding to model choice $\mathcal{M}_{2\bullet}$ under parametric model, we also studied different variance structures, in order to explore the different choices of variance functions under the AL distribution.

Again, we confirm that amongst all models with AL distribution, $\mathcal{M}_{23}$ which incorporates both accident and development year effects for the mode and variance demonstrates the best model fit according to all DIC, $\bar{D}$ and $\hat{D}$. On the other hand, MSE favors $\mathcal{M}_{22}$ which adopts only development year effect for the variance. One possible reason might be that the payments made in different accident years are relatively stable compared to those across development years, and hence the development year effect dominates in the variance estimation, a specific phenomenon in this example.

5.2. Analysis of quantile regression models: Quantile distribution

In this section, we analyze the different model choices from the distributional perspective. This is not directly trivial to achieve, since each model has different features that must be taken into consideration in the comparison. It is clear from previous studies that one should always utilize an ANOVA-type location function with accident and development years effect ($M_{2\bullet}$), or at a minimum incorporate a quadratic or basis function form for the development year effects such as $M_{1\bullet}$. In the case of the GB2 and AL models, we will therefore consider mean structures in $M_{2\bullet}$. However, in the case of the PP model we will consider $M_{1\bullet}$, since purely from a computational perspective it will be easier to implement an efficient MCMC sampler for $M_{1\bullet}$ compared to $M_{2\bullet}$. The reason for this is due to the rejection stage in the Metropolis–Hastings acceptance probability where under the PP model the prior constraint regions will be easier to satisfy with less model complexity. In terms of the variance functions, when working with the GB2 models, we will consider $M_{2\bullet}$ in which we do not specify variance functions as there is no variance parameter in the distribution to model the variance directly. The variance of the models are given by (19). Then, in the case of the AL model we consider M_{20} as well as M_{23} and for the PP model we consider M_{10} and M_{13}.

Table 3 reports the results for the Israel data set which are split according to models with constant, unspecified and dynamic variance functions. In the case of constant or unspecified variance, the best performing model is again the AL model, followed by the GG model. Among distributions in the GB2 family with positive support, GG provides the best model fit according to DIC with model complexity penalty while GB2 model provides the best model prediction according to MSE. Comparing $\bar{D}$s without model complexity penalty, GG and GB2 provide very similar model fit. Besides, it is clear that the PP model with only the basis function regression structure for the mean, given by a quadratic polynomial for the trend in the development year covariate, and a constant variance was not sufficient to capture all the features required. We believe that this is largely due to the fact that such a model is more suitable for heavy tailed run-off in the claims development and the Israel data clearly does not display such a feature. It is therefore expected that such a heavy tailed quantile regression model will not perform as well for this data. When the variance is also modeled, the AL model is clearly significantly better than all the other models considered, again making M_{23} with AL model optimal compared to all choices. Since, the PP model is shown to be not suitable for this data, we will consider analyses going forward with only the GB2 and AL models.

To further assess the in-sample model fits, we display in Figure 8 the observed losses Y_{ij} and fitted losses $\widehat{Y}_{ij} = \exp(\mu^*_{ij})$ for AL model and $\widehat{Y}_{ij} = \mu_{ij}$ for the GB2 family models in the upper triangle arranged in ascending order against

TABLE 3

PARAMETER ESTIMATES AND MODEL FIT MEASURES FOR MODELS WITH VARIOUS DISTRIBUTIONS.

Models	*DIC*	$\bar{D}$	$\hat{D}$	*MSE*	a	p	q	σ^2
			Quantile Regression: Unspecified Variance Function					
M_2. Gamma	3064.50	3028.93	2993.36	537.82	1	1.87	∞	–
M_2. GG	2707.42	2932.97	3158.52	582.78	33.22	0.08	∞	–
M_2. GB2	3002.82	2964.60	2926.37	526.65	−7.94	1.78	0.17	–
			Quantile Regression: Constant Variance Function					
M_{10} PP	3272.14	1021.71	1230.01	1132.12	–	–	–	14.15
M_{20} AL	50.94	120.17	189.40	1015.71	–	0.80	–	0.02
			Quantile Regression: Non-Constant Variance Function					
M_{13} PP	1502.19	1906.49	2310.98	923.00	–	–	–	9.10
M_{23} AL	-20.81	24.91	70.63	850.10	–	0.75	–	0.17

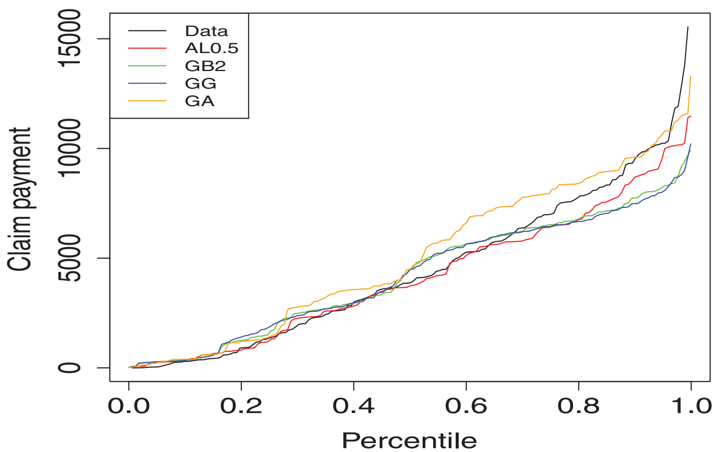

FIGURE 8: Percentiles of fitted losses in the upper triangle using GB2 family and AL distributions. (Color online)

the order (percentile) p in the data. In the AL model, we set $p = 0.5$ that corresponds to non-parametric median regression to facilitate comparison with the GB2, GG and gamma models which have mean regression. We can see that the fitted losses using AL model with M_{33} are closest to the observed losses, GG and GB2 models provide very similar fitted losses and gamma model provides the poorest fit. Table 4 reports the observed and fitted losses in Figure 8 for some specified percentile $p = 0.3, 0.5, 0.75$ and 0.95 in the data using the four models.

As the model assessments show adequate model fits, we apply the quantile functions to calculate the losses for each cell (i, j) in the upper triangle at different quantile levels u. The quantile function is $Q_{Y^*}(u|\boldsymbol{x}_{ij})$ in (8) where

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

TABLE 4

SELECTED PERCENTILES OF FITTED LOSSES IN THE UPPER TRIANGLE USING GB2 AND AL MODELS.

Models	0.30	0.50	0.75	0.95
Observed	1,985	3,871	6,990	10,200
M_2. Gamma	2,760	4,496	8,036	10,700
M_2. GG	2,378	4,498	6,451	8,040
M_2. GB2	2,480	4,463	6,526	8,247
M_{23} AL ($p = 0.5$)	2,255	3,734	6,422	9,715

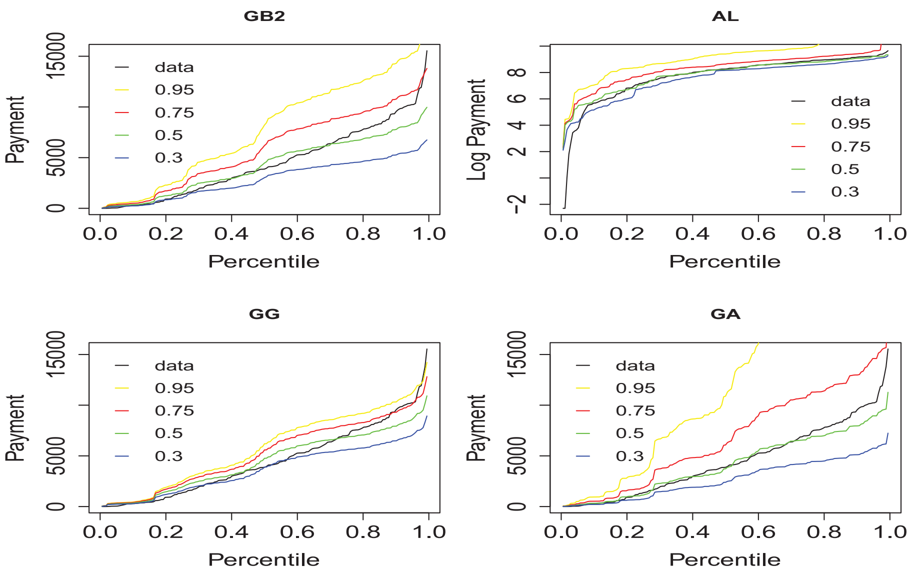

FIGURE 9: Percentiles of predicted quantiles in the upper triangle using GB2 and AL models. (Color online)

$Q_{\epsilon^*}(u) = F_{AL}^{-1}(u)$ is given by (11) for the AL model and it is $Q_Y(u|\boldsymbol{x}_{ij})$ in (22) for the GB2 model and (25) for the GG and Gamma models. Figure 9 plots quantile estimates $\widehat{Q}_{Y^*}(u|\boldsymbol{x}_{ij})$ or $\widehat{Q}_Y(u|\boldsymbol{x}_{ij})$ in ascending order against the percentile p. This is similar to Figure 8, for example, the fitted quantile $\widehat{Q}_Y(u|\boldsymbol{x}_{ij})$ instead of fitted mean $\widehat{Y}_{ij}$ for the GB2 family is plotted against the percentile. Each line in Figure 9 corresponds to a quantile level $u = 0.3, 0.5, 0.75$ and 0.95. These so-called empirical quantile lines are dense for GG model, sparse for gamma model and moderate for GB2 model indicating that GB2 distribution provides quantile estimates which can reasonably cover the observed losses across percentile when the quantile level u gradually increases. We also remark that the empirical quantiles for AL model in the log scale are convex rather than concave and are more dense because of the log transformation.

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

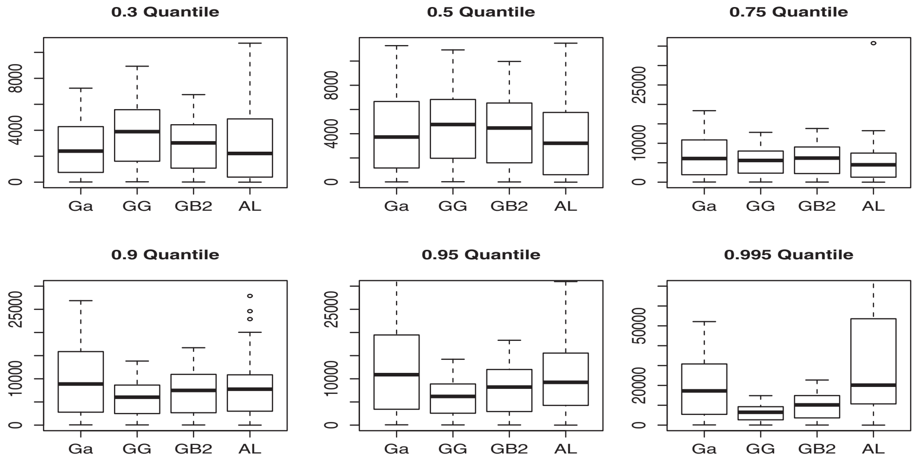

FIGURE 10: Boxplots of predicted quantile in the upper triangle using GB2 family and AL distributions.

Then Figure 10 reveals the shape of distribution using boxplot for the estimated losses at each quantile level (line) in Figure 9. The quantile estimates are $\exp(\widehat{Q}_{Y^*}(u|\boldsymbol{x}_{ij}))$ for the AL model and $\widehat{Q}_Y(u|\boldsymbol{x}_{ij})$ for the GB2 family models. Comparing boxplots across models and quantile levels, the boxplots for AL model have the heaviest right tails and the ranges of boxplots differ more at higher quantile levels. In particular, the ranges for gamma and AL models increase much faster across quantile levels than the GG and GB2 models.

Then Figure 11 plots the quantile functions $\exp(Q_{Y^*}(u|\boldsymbol{x}_o))$ for the AL model and $Q_Y(u|\boldsymbol{x}_o)$ for the GB2 family models against the quantile $u \in (0, 1)$ as compared to the empirical quantiles plotted against data percentile p in Figure 9. Since, the quantile functions do not refer to any particular cell in the upper triangle, the location parameters μ^* and μ involved in calculating $\exp(Q_{Y^*}(u|\boldsymbol{x}_o))$ and $Q_Y(u|\boldsymbol{x}_o)$ respectively are taken to be the average of μ^*_{ij} or μ_{ij} over $\boldsymbol{x}_0$ in the upper triangle. Hence, this graph presents the average reserve level for each risk cell across quantile level u for the four models. The AL model is the most conservative at higher quantiles and has the heaviest right tail. This is partially due to the log transformation.

We further adopt these models to calculate the OR as reported in Table 5 using the conditional predictive approach where the conditional total reserve quantile function in (47) is adopted for the case of light tailed run-off in the claim process because the claim distribution was shown to be light tailed in the previous analyses. Under the Solvency II framework, insurers will have to establish technical provisions to cover future claims expected from policyholders. Insurers must also have available financial resources sufficient to cover both a minimum capital requirement and a SCR. The SCR is based on a VaR measure

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

TABLE 5

OUTSTANDING RESERVES AT DIFFERENT QUANTILE LEVELS USING GB2 FAMILY AND AL DISTRIBUTIONS.

Models	0.30	0.50	0.75	0.95
M_2. Gamma	127,816	198,907	324,515	581,302
M_2. GG	203,207	248,409	291,457	323,346
M_2. GB2	152,315	225,017	311,625	413,525
M_{23} AL ($p = 0.5$)	145,031	176,926	314,454	462,980

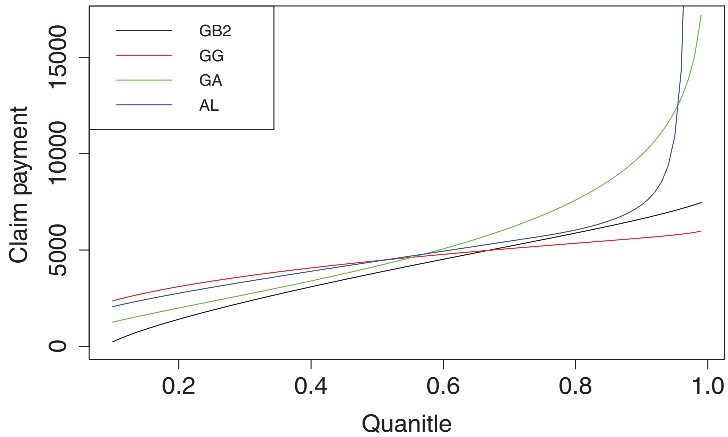

FIGURE 11: Quantile functions using GB2 family and AL distributions. (Color online)

calibrated to a 99.5% confidence level over a one-year time horizon. Results in Table 5 show that the OR projection increases gradually up to 0.95 quantile level. To fulfill the requirement of SCR, estimates of OR at 0.995 quantile level are also calculated and they are 5,12,731 and 5,60,430 using GB2 and AL models respectively. It seems that all the predictions at 0.75 quantile level are quite close in Table 5. The APRA requires licensed Australian insurers to have a minimum probability of adequacy of 75% . Hence, all estimations from 75% quantiles onwards can be considered.

6. RISK MARGIN: AUSTRALIAN CASE STUDY

In general, the guidance on calculation of risk margin by regulators leaves flexibility in the practical modeling approach adopted by practitioners. There are a few popular approaches considered in practice, some of which involve a degree of expert opinion. In this section, we aim to consider only approaches based on statistical models and in particular quantile based methods. In this context, the standard practice is to consider the reserve estimate and then try to quantify the uncertainty associated with the reserve estimator. This uncertainty is typically measured via a standard error, which is utilized to adjust the reserve. Traditionally, if a loss distribution produces an estimator for the reserve which admits a normal distribution (approximately under a central limit theorem result), then setting the risk margin to equal the sample estimator for the reserve plus 0.675 times the sample estimator of standard deviation would result in risk margins calibrated to approximately the 75th percentile. Note, whilst the total loss distribution may not have finite second moment if a heavy tailed run-off is present, the variance of the sample estimator for the distribution of the reserve will always be well defined. It should be noted that this method suffers from drawbacks as there is both an influential judgment in determining the appropriate multiple, especially when the normality assumption is not present due to sample estimators distribution being skewed. The first plot in Figure 12 illustrates graphically this traditional method where the blue line represents the assumed normal liability (loss) distribution, the red arrow indicates the risk margin calculated using traditional method and the red line denotes the underlying loss distribution which differ considerably from the assumed normal distribution.

Alternatively, if the traditionally utilized estimate of reserve based on the *mean* of the loss distribution is considered but normality is dropped. Then, two scenarios may arise if one uses the risk margin adjustment based on the tail quantile of the total loss distribution at say 75%. In this case, the estimated mean reserve could be *above* the desired risk margin quantile level of the total loss distribution, in which case it may be reasonable to make no further adjustment if the risk margin is already at a tail quantile such as 75%. Alternatively, if the estimated mean reserve is *below* the desired risk margin quantile level of the total loss distribution, then the difference would be the resulting risk margin.

Alternatively, one may utilize the quantile regression model obtained for the total loss distribution. There are two basic ways to achieve this, for instance one could take instead of a mean reserve, a quantile based reserve. The first way considers the *median* of the total loss distribution as a central measure and makes a risk margin adjustment based on the tail quantile of the total loss distribution at say 75% (as is considered in practice). The second way applies a risk measure such as VaR which represents a tail quantile of the total loss distribution at say 99.95%, in which case one may judge that a conservative measure of reserve is obtained from such a tail measure and so no additional risk margin is required. This is standard in banking regulations such as Basel II/III and being considered in insurance regulations.

In this section, we adopt the quantile approach for risk margin estimation. We first extend the best model, model M_{23} with AL distribution, in the previous sections to model the risk margin statistically. To achieve this, we generalize the AL distribution to model the shape parameter p via the following regression $p_i = \phi_0 + \phi_i$ where ϕ_0 is the intercept and ϕ_i denotes accident year effect. Accident year effect is chosen because risk capital allocation is by accident

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

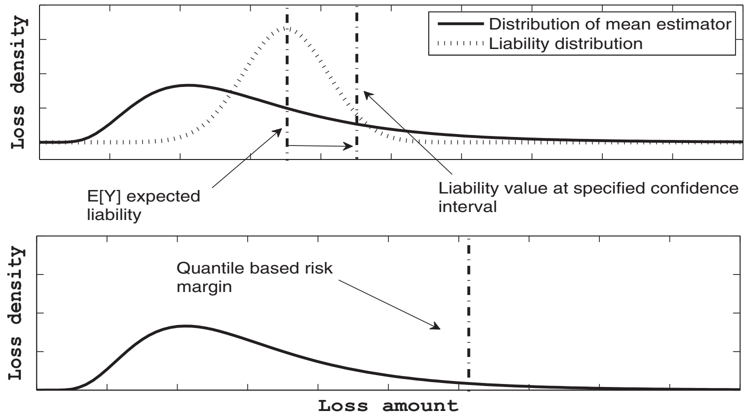

FIGURE 12: Traditional method (upper) versus proposed method (lower).

years. It is worth noting an important assumption which is stated as underlying this method: actual outstanding claim payments are assumed to be uncorrelated between accident years. Therefore, the estimated shape parameter p, which presents quantile in AL distribution, and also infers risk margin in the percentile method, is an applicable risk margin estimate for outstanding claims payments. The difference between our proposed quantile based method in the second plot and the traditional method is clearly illustrated in Figure 12.

The data that we used to demonstrate our model is the amount of payments for all the compulsory third party (CTP) policies in Queensland (QLD) as of June 2008. CTP insurance policy covers risk that would be referred to as Auto Bodily Injury in the U.S. and Motor Bodily Injury in the U.K. The data are in the units of millions summarized by accident and development quarters covering periods from December 2002 to June 2008. It contains 276 observations over 23 accident quarters. In order to remove the influence of inflation for reserving purposes, we utilize the average weekly earning index from the Australian Bureau of Statistics (ABS) to inflate all the values to December 2008 dollars. Hence, the data used in this analysis represents the inflated cumulative payment for QLD CTP portfolio as reported in Figure 18 in Appendix A.

To review features of the data, Figure 13 plots the observed variance across accident year on original and log scale. It shows that the variance fluctuates a lot across accident year on the original scale but displays a sharp drop on the log scale. Figure 14 shows that the skewness are mostly negative on the original and log scales. The overall skewness of the data is 0.61 and that on the log scale is -1.08. Trend of skewness reveals a sharp drop at the start and then it fluctuates across accident years for data on the original scale but increases monotony for data on the log scale. These changes confirm the necessity of adopting dynamic variance and skewness in modeling the data.

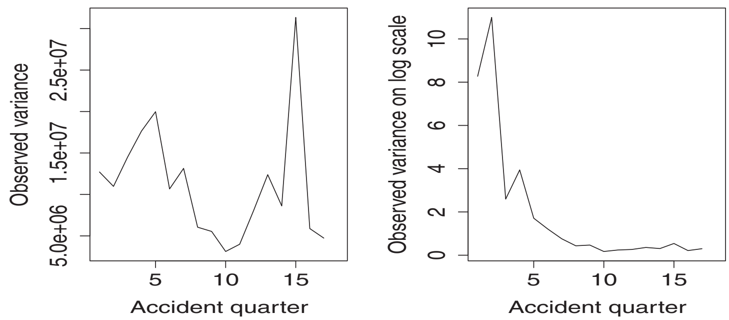

FIGURE 13: Observed variance of QLD CTP payment data by accident year.

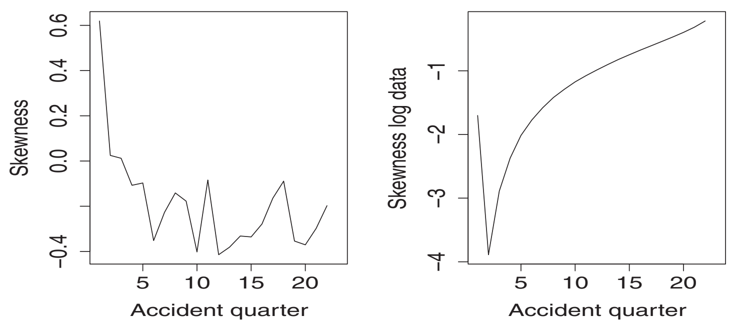

FIGURE 14: Observed skewness of QLD CTP payment data by accident year.

Among choices of distributions, the AL distribution allows flexibility in modeling variance and skewness through modeling directly the scale and shape parameters σ^2 and p respectively. Furthermore, in the context of non-parametric regression using AL as a proxy distribution for model implementation, p indicates the quantile level of a model which corresponds to risk margin in loss reserving. In the analysis of QLD CTP data, we adopt model M_{23} with ANOVA type location and variance functions as preliminary study shows that M_{23} also provides the best model performance, similar to Israel data. We further propose modeling the risk margin p as a linear function of accident quarters. One reason is that as accident quarter increases, there are more uncertainty involved in estimating the reserves; hence, it is an important factor in risk margin estimation. This model is called $M_{23'}$ in Table 7, Appendix B.

Then $M_{23'}$ with dynamic variance and skewness is compared to two models, M_{20} with constant variance and skewness and M_{23} with just dynamic variance in Table 6. Although M_{20} outperform $M_{23'}$ according to DIC, $M_{23'}$ provides the best model fit according to $\bar{D}$ which measures solely the model fit without model complexity penalty. As our aim is to provide the most accurate risk margin estimates, we adopt $M_{23'}$ in the subsequent risk margin analysis. From a modeling perspective, it reconciles with our risk margin estimation approach.

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

TABLE 6

PARAMETER ESTIMATES AND MODEL FIT MEASURES FOR ANOVA MODELS USING QLD CTP PAYMENT DATA.

Models	DIC	$\bar{D}$	$\hat{D}$	$E(Y)$	$Var(Y)$	$S(Y)$
M_{20} Constant variance & skewness	−322.55	−215.65	−108.75	4.33	0.008	−0.28
M_{23} Dynamic variance	−311.36	−197.71	−84.06	7.67	0.22	−0.57
$M_{23'}$ Dynamic variance & skewness	−255.03	−229.46	−203.90	4.77	0.10	−0.18

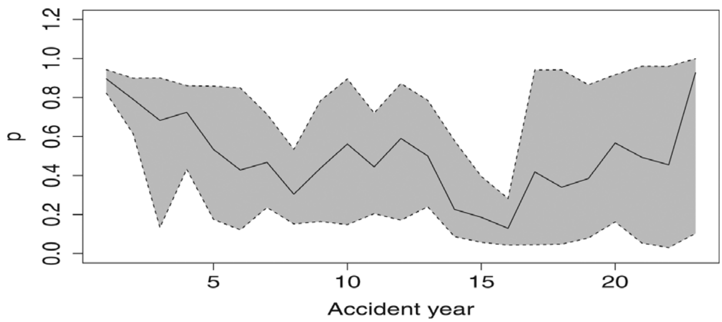

FIGURE 15: Change of p across accident year using $M_{23'}$ for risk margin analysis.

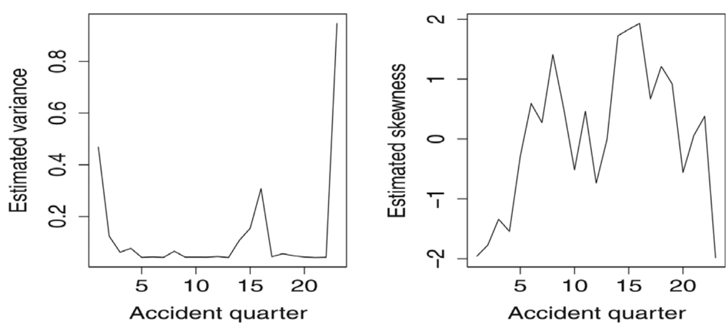

FIGURE 16: Estimated variance and skewness in $M_{23'}$ for risk margin analysis.

Figure 15 demonstrates how the estimated risk margin $\hat{p}_i$ changes across accident quarters i, superimposed with its credibility interval. Figure 16 displays the corresponding changes in estimated variance and skewness using the variance and skewness equations in (12) and (13) respectively. The risk margin $\hat{p}_i$ starts at 0.895 at accident quarter 1 when the variance is quite high. Afterwards, it decreases gradually to 0.439 in accident quarter 8 when the variance is much smaller. From accident quarter 17 onwards, the risk margin increases again when the variance is large when the data become rare and there are more

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

development quarters ahead. In actuarial practice, the calculation of the risk margin is often not based on a sound model but various simplified methods are used. This approach enables us to calculate a risk margin for non-life insurance run-off liabilities in a mathematically consistent way, and provides reasonable risk margin estimates.

7. Conclusion

We apply regression models to estimate loss reserve and risk margin and compare the performance of quantile functions from two approaches. In the non-parametric approach, we adopt AL as a proxy distribution for parameter inference and find models with ANOVA type location and variance and higher skewness (higher shape parameter p) which corresponds to higher quantile level, provide better model-fit. In the parametric framework, we built five models, namely AL, PP, GB2, GG and gamma on real and positive supports. The AL model provides the best fit because it has specific location and variance parameters to allow dynamic modeling of both components. We also investigate three different regression structures, namely ANCOVA, ANOVA and Poisson–Tweedie regression. The ANOVA model performs the best in our empirical data study.

Furthermore, we adopt the best performed model, which is the AL model in a parametric approach with ANOVA location and variance functions, to estimate risk margin. This AL model is further generalized to adopt a dynamic shape parameter p and the resultant model provides us a mathematically consistent way of estimating risk margin because the estimated p corresponds to the quantile level in a non-parametric model and hence is the best quantile level for risk margin according to the data. Overall, the results of our studies indicate that this new risk margins framework offers considerable potential benefits for reserving purpose.

References

AIUPPA, A. (1988) Evaluation of Pearson curves as an approximation of the maximum probable annual aggregate loss. *Journal of Risk and Insurance*, **55**(3), 425–441.

ARTZNER, P. (1999) Application of coherent risk measures to capital requirements in insurance. *North American Actuarial Journal*, **3**(2), 11–25.

Australian Prudential Regulatory Authority, Prudential Standard GPS 320, Actuarial and Related Matters. (May 2012) http://www.apra.gov.au/CrossIndustry/Consultations/Documents/Draft-GPS-320-Actuarial-and-Related-Matters-May-2012.pdf.

Bank of England Prodential Regulation Authority, Solvency II: An update on implementation. (August 2014) http://www.bankofengland.co.uk/pra/Documents/solvency2/solvency2updateaugust2014.pdf.

BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J. and TEUGELS, J. (2006) *Statistics of Extremes: Theory and Applications.* John Wiley & Sons.

BINGHAM, N.H., GOLDIE, C.M. and TEUGELS, J.L. (1989) *Regular Variation*. Cambridge University Press.

BOROVKOV, A.A. and BOROVKOV K.A. (2008) *Asymptotic Analysis of Random Walks: Heavy-Tailed Distributions.* Cambridge University Press.

CAI, Y. (2010) Polynomial power-Pareto quantile function models. *Extremes*, **13**(3), 291–314.

CARLIN B.P. and THOMAS A.L. (2000) *Bayes and Empirical Bayes Methods for Data Analysis.* CRC Press.

CHAN, J.S.K., CHOY, S.T.B. and MAKOV, U.E. (2008) Robust Bayesian analysis of loss reserves data using the generalized-t distribution. *Astin Bulletin*, **38**(1), 207–230.

CHEN, Q., GERLACH, R. and LU, Z. (2012) Bayesian value-at-risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics and Data Analysis*, **56**(11), 3498–3516.

CLAESKENS, G. and HJORT, N.L. (2008) *Model Selection and Model Averaging*, vol. 330. Cambridge: Cambridge University Press.

CRUZ, M.G., PETERS, G.W. and SHEVCHENKO, P.V. (2014) *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk*. John Wiley and Sons.

CUMMINS, J.D., MCDONALD, J.B. and CRAIG, M. (2007) Risk loss distributions and modelling the loss reserve pay-out tail. *Review of Applied Economics*, **3**(1–2), 1–23.

DAOUIA, A., GARDES, L. and GIRARD, S. (2012) On kernel smoothing for extremal quantile regression. *Bernoulli*, **19**(5B), 2557–2589.

DENISON, D.G.T., HOLMES, C.C., MALLICK, B.K. and SMITH, A.F.M. (2002) *Bayesian Methods for Nonlinear Classification and Regression*. John Wiley and Sons.

DE ALBA, E. (2002) Bayesian estimation of outstanding Claim Reserves. *North American Actuarial Journal*, **6**(4), 1–20.

DELBAEN, F. (2002) Coherent risk measures on general probability spaces. In *Advances in Finance and Stochastics*, pp. 1–37. Berlin Heidelberg: Springer.

DONG, X. and CHAN, J. (2013) Bayesian analysis of loss reserving using dynamic models with generalized beta distribution. *Insurance: Mathematics and Economics*, **53**(2), 355–365.

DOWD, K. and BLAKE D. (2006) After VaR: The theory, estimation, and insurance applications of quantile-based risk measures. *Journal of Risk and Insurance*, **73**(2), 193–229.

EMBRECHTS, P., KLUPPELBERG, C. and MIKOSCH, T. (1997) *Modeling Extreme Events for Insurance and Finance*. Berlin: Springer.

ENGLE, R. and MANGANELLI, S. (2004) CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, **22**(4), 367–381.

GILCHRIST, W. (2002) *Statistical Modelling with Quantile Functions*. CRC Press.

GILKS, W.R., RICHARDSON, S. and SPIEGELHALTER, D.J. (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

GOEGEBEUR, Y., GUILLOU, A. and SCHORGEN, A. (2014) Nonparametric regression estimation of conditional tails: The random covariate case. *Statistics*, **48**(4), 732–755.

GOOVAERTS, M.J., DHAENE, J. and DE SCHEPPER, A. (2000) Stochastic upper bounds for present value functions. *Journal of Risk and Insurance Theory*, **67**(1), 1–14.

GUERMAT, C. and HARRIS, R.D.F. (2002) Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns. *International Journal of Forecasting*, **18**(3), 409–419.

GYORGY, S. and SHAW, W.T. (2008) Quantile mechanics. *European Journal of Applied Mathematics*, **19**(2), 87–112.

HASTINGS, W.K. (1970) Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, **57**(1), 97–109.

HU, Y., GRIMACY, R.B. and LIAN, H. (2012) Bayesian quantile regression for single-index models. *Statistics and Computing*, **23**(4), 437–454.

KAAS, R., DHAENE, J. and GOOVAERTS M. (2000) Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, **27**(2), 151–168.

KLUPPELBERG, C. and MIKOSCH, T. (1998) Large deviations of heavy-tailed random sums with applications in insurance and finance. *Extremes*, **1**(1), 81–110.

KOENKER, R. and HALLOCK, K. (2001) Quantile regression: An introduction. *Journal of Economic Perspectives*, **15**(1), 143–156.

KOENKER, R. and MACHADO, A. F. (1999) Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, **94**(448), 1296–310.

MARSHALL, K., COLLINGS, S., HODSON, M. and O'DOWD, C. (2008) *A framework for assessing risk margins. Prepared by the Risk Margins Task Force for Institute of Actuaries of Australia, 16th General Insurance Seminar, 9–12*** November 2008*, Coolum, Australia.

MCDONALD, J.B. (1984) Some generalized functions for the size distribution of income. *Econometrica*, **52**(3), 647–663.

MCDONALD, J.B. and NEWEY, W.K. (1988) Partially adaptive estimation of regression models via the Generalized t distribution. *Econometric Theory*, **4**(3), 428–457.

METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N. and TELLER, A.H. (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**(6), 1087–1091.

NELSON, C.R. and SIEGEL, A.F. (1987) Parsimonious modeling of yield curves. *Journal of Business*, **60**(4), 473–489.

NTZOUFRAS, I. and DELLAPORTAS, P. (2002) Bayesian modeling of outstanding claim reservesliabilites incorporating claim count uncertainty. *North American Actuarial Journal*, **6**(1), 113–128.

ORI, R. and COHEN, A. (1996) Extreme percentile regression. In *Statistical Theory and Computational Aspects of Smoothing*, pp. 200–2014. Physica-Verlag HD.

PAULSON, A.S. and FARIS, N.J. (1985) A practical approach to measuring the distribution of total annual claims. In *Strategic Planning and Modeling in Property-Liability Insurance* (ed. J.D. Cumins), Norwell, MA: Kluwer Academic Publishers.

PETERS, G.W., BYRNES, A.D. and SHEVCHENKO, P.V. (2011a) Impact of insurance for operational risk: Is it worthwhile to insure or be insured for severe losses? *Insurance: Mathematics and Economics*, **48**(2), 287–303.

PETERS, G.W., SHEVCHENKO, P.V. and WUTHRICH, M.V. (2009) Model uncertainty in claims reserving within Tweedie compound Poisson models. *ASTIN Bulletin*, **39**(1), 1–33.

PETERS, G.W., SHEVCHENKO, P.V., YOUNG, M. and YIP, W. (2011b) Analytic loss distributional approach models for operational risk from the α-stable doubly stochastic compound processes and implications for capital allocation. *Insurance: Mathematics and Economics*, **49**(3), 565–579.

PETERS, G.W., TARGINO, R.S. and SHEVCHENKO, P.V. (2013) *Understanding operational risk capital approximations: First and second orders*. Governance and Regulation (Invited Special Issue 8th International Conference "International Competition in Banking: Theory and Practice", Sumy, Ukraine), **2**(3), 58–79.

RAMLAU-HANSEN, H. (1988) A solvency study in non-life insurance. Part 1. analysis of fire, Windstorm, and glass claims. *Scandinavian Actuarial Journal*, 3–34.

SMITH, A.F.M. and ROBERTS, G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, **55**, 3–23.

SPIEGELHALTER, D., BEST, N.G., CARLIN, B.P. and VAN DER LINDE, A. (2002) Bayesian measures of model complexity and fit. (with Discussion). *Journal of the Royal Statistical Society, Series B*, **64**(4), 583–616.

STACY, E.W. (1962) A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, **33**(3), 1187–92.

TAYLOR, G. APRA (2006) GENERAL INSURANCE RISK MARGINS. (February 2006). `http://fbe.unimelb.edu.au/__data/assets/pdf_file/0011/806267/136.pdf`.

VERRALL, R.J. and WUTHRICH, M. (2013) Reversible jump Markov chain Monte Carlo method for parameter reduction in claims reserving. To appear in *North American Actuarial Journal*.

YU, K. and MOYEED, R.A. (2001) Bayesian quantile regression. *Statists and Probability Letters*, **54**(4), 437–447.

YU, K. and ZHANG, J. (2005) A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics Theory and Methods*, **34**(9), 1867–1879.

ZHANG, Y., DUKIC, V. and GUSZCZA, J. (2012) A Bayesian nonlinear model for forecasting insurance loss payments. *Journal of the Royal Statistical Society, Series A*, **175**(2), 1–20.

ALICE X.D. DONG (Corresponding author)
School of Mathematics and Statistics,
The University of Sydney,
NSW 2006, Australia
E-Mail: alice.dong10@gmail.com

JENNIFER S.K. CHAN
School of Mathematics and Statistics,
The University of Sydney,
NSW 2006, Australia
E-Mail: jenniferskchan@gmail.com

GARETH W. PETERS
Department of Statistical Science,
University College London UCL,
London, UK
E-Mail: gareth.peters@ucl.ac.uk

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

APPENDIX A

	Development quarters																	
Accident quarters	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1	3,323	8,332	9,572	10,172	7,631	3,855	3,252	4,433	2,188	333	199	692	311	0	405	293	76	14
2	3,785	10,342	8,330	7,849	2,839	3,577	1,404	1,721	1,065	156	35	259	250	420	6	1	0	
3	4,677	9,989	8,746	10,228	8,572	5,787	3,855	1,445	1,612	626	1,172	589	438	473	370	31		
4	5,288	8,089	12,839	11,829	7,560	6,383	4,118	3,016	1,575	1,985	2,645	266	38	45	115			
5	2,294	9,869	10,242	13,808	8,775	5,419	2,424	1,597	4,149	1,296	917	295	428	359				
6	3,600	7,514	8,247	9,327	8,584	4,245	4,096	3,216	2,014	593	1,188	691	368					
7	3,642	7,394	9,838	9,733	6,377	4,884	11,920	4,188	4,492	1,760	944	921						
8	2,463	5,033	6,980	7,722	6,702	7,834	5,579	3,622	1,300	3,069	1,370							
9	2,267	5,959	6,175	7,051	8,102	6,339	6,978	4,396	3,107	903								
10	2,009	3,700	5,298	6,885	6,477	7,570	5,855	5,751	3,871									
11	1,860	5,282	3,640	7,538	5,157	5,766	6,862	2,572										
12	2,331	3,517	5,310	6,066	10,149	9,265	5,262											
13	2,314	4,487	4,112	7,000	11,163	10,057												
14	2,607	3,952	8,228	7,895	9,317													
15	2,595	5,403	6,579	15,546														
16	3,155	4,974	7,961															
17	2,626	5,704																
18	2,827																	

FIGURE 17: Israel payment data.

https://doi.org/10.1017/asb.2015.8 Published online by Cambridge University Press

	Development quarter																							
Accident quarter	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**Exposure**
Dec-02	0.1	0.6	1.3	2.1	3.0	4.3	5.8	5.6	4.9	6.7	6.1	14.3	13.0	9.8	7.9	8.9	8.0	4.0	9.4	3.4	3.3	2.2	5.4	2.6
Mar-03	0.1	0.6	0.9	0.9	1.3	1.8	3.0	2.6	2.7	5.9	6.1	7.4	7.9	8.5	9.1	9.3	5.4	6.6	8.4	5.3	1.4	2.5		2.6
Jun-03	0.1	0.6	0.7	0.8	1.2	1.7	2.6	3.0	4.3	6.0	6.9	5.8	7.1	10.2	11.0	5.5	8.1	6.2	4.8	1.3	7.3			2.6
Sep-03	0.1	0.6	0.8	1.0	1.3	2.0	2.1	3.6	4.5	8.0	5.8	5.4	11.0	9.6	6.5	10.5	7.4	6.2	5.4	4.0				2.6
Dec-03	0.0	0.6	0.9	1.1	1.0	1.1	2.5	3.8	5.2	4.5	7.4	6.0	12.1	5.3	9.7	8.0	6.7	4.4	8.2					2.7
Mar-04	0.1	0.5	1.0	0.9	0.9	1.7	2.4	4.2	5.0	8.1	7.5	8.7	7.9	9.9	8.6	11.8	10.0	5.6						2.7
Jun-04	0.1	0.5	0.9	0.7	1.0	1.8	2.6	4.1	5.8	5.6	7.2	6.2	10.0	6.5	11.5	6.8	5.4							2.7
Sep-04	0.1	0.6	0.7	0.9	1.1	2.1	3.2	4.8	5.8	7.5	8.0	14.2	9.4	15.0	8.4	11.7								2.8
Dec-04	0.1	0.5	0.8	0.9	1.1	1.8	4.1	5.9	6.5	5.9	12.5	9.2	13.9	6.2	10.0									2.8
Mar-05	0.0	0.5	1.0	0.8	1.1	1.5	4.1	6.7	5.5	10.7	7.9	8.2	8.8	8.7										2.8
Jun-05	0.1	0.7	0.8	0.9	1.3	2.3	5.3	4.4	7.4	8.3	9.1	7.3	15.2											2.9
Sep-05	0.1	0.6	1.1	0.9	2.3	2.9	3.8	8.2	8.9	9.6	6.9	8.6												2.9
Dec-05	0.1	0.7	1.0	1.2	1.5	3.0	5.7	7.2	8.7	7.8	9.8													2.9
Mar-06	0.0	0.4	0.6	0.8	0.9	3.2	4.7	7.2	6.8	7.0														3.0
Jun-06	0.1	0.6	0.8	0.7	1.9	4.4	7.5	6.5	7.7															3.0
Sep-06	0.0	0.5	0.7	0.8	1.9	6.2	6.9	7.7																3.0
Dec-06	0.1	0.5	1.0	1.5	2.0	4.4	7.6																	3.1
Mar-07	0.0	0.7	1.0	1.0	1.6	5.6																		3.1
Jun-07	0.1	0.6	0.8	0.9	2.0																			3.2
Sep-07	0.1	0.7	0.9	1.1																				3.2
Dec-07	0.1	0.5	0.7																					3.2
Mar-08	0.1	0.6																						3.3
Jun-08	0.1																							3.3

FIGURE 18: QLD CTP payment data.

APPENDIX B

The following table shows the model structures considered for each regression analysis.

TABLE 7

MODEL STRUCTURES IN THE QUANTILE REGRESSIONS. NOTE: BASIS FUNCTION CHOICES $F_1(j) = \left(\frac{1-e^{-\lambda\times j}}{\lambda\times j}\right)$, $F_2(j) = \left(\frac{1-e^{-\lambda\times j}}{\lambda\times j} - e^{-\lambda\times j}\right)$.

Model Index	Model Location Structure	Model Scale Structure	Distribution Types	Model Description
M_{00}	$\mu^*_{ij} = \alpha_0 + \alpha_1 \times i + \alpha_2 \times j$	$\sigma_{ij} = \sigma$	AL	**Location:** Simple Additive Model (parsimonious) common trend in accident years and development years. **Scale:** homoscedasticity in development years scale parameter (common across accident years).
M_{10}	$\mu^*_{ij} = \alpha_0 + \alpha_1^S F_1(j) + \alpha_2^C F_2(j)$	$\sigma_{ij} = \sigma$	AL	**Location:** Basis function regression model with trend component for development years given by Level, Slope and Curvature components (common across accident years). **Scale:** homoscedasticity in development years scale parameter (common across accident years).
M_{20}	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \sigma$	AL, PP	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** homoscedasticity in development years scale parameter (common across accident years).
$M_{2\bullet}$	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	Eqn 19.	GB2	**Location:** Fully parameterized model with individual trend components in accident and development years.

TABLE 8

MODEL STRUCTURES IN THE QUANTILE REGRESSIONS.

Model Index	Model Location Structure	Model Scale Structure	Distribution Types	Model Description
M_{21}	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \beta_0 + \beta_{1i}$	AL	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** heteroscedasticity in accident years with common variance over development years scale parameter.
M_{22}	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \beta_0 + \beta_{2j}$	AL	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** heteroscedasticity in development years with common variance over accident years scale parameter.
M_{23}	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \beta_0 + \beta_{1i} + \beta_{2j}$	AL	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** heteroscedasticity in development and accident years scale parameter.
$M_{23'}$	$\mu^*_{ij} = \alpha_0 + \alpha_{1i} + \alpha_{2j}$	$\sigma_{ij} = \beta_0 + \beta_{1i} + \beta_{2j}$ $p = \phi_0 + \phi_{1i}$	AL	**Location:** Fully parameterized model with individual trend components in accident and development years. **Scale:** homoscedasticity in scale parameter and shape parameter p (quantile level) has trend in the accident years (common across all development years).
M_{30}	$\mu^*_{ij} = \alpha_{0,u} + \alpha_{1i,u} + \alpha_{2j,u}$	$\sigma_{ij} = \sigma$	AL as proxy	**Location:** Non-parameterized model with individual trend components in accident and development years. **Scale:** not defined in the model.