

Article

# An EM Algorithm for Double-Pareto-Lognormal Generalized Linear Model Applied to Heavy-Tailed Insurance Claims

Enrique Calderín-Ojeda \*, Kevin Fergusson and Xueyuan Wu 

Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Melbourne, VIC 3010, Australia; kevin.fergusson@unimelb.edu.au (K.F.); xueyuanw@unimelb.edu.au (X.W.)

\* Correspondence: enrique.calderin@unimelb.edu.au, Tel.: +61-39-035-8176

Received: 27 September 2017; Accepted: 3 November 2017; Published: 7 November 2017

**Abstract:** Generalized linear models might not be appropriate when the probability of extreme events is higher than that implied by the normal distribution. Extending the method for estimating the parameters of a double Pareto lognormal distribution (DPLN) in Reed and Jorgensen (2004), we develop an EM algorithm for the heavy-tailed Double-Pareto-lognormal generalized linear model. The DPLN distribution is obtained as a mixture of a lognormal distribution with a double Pareto distribution. In this paper the associated generalized linear model has the location parameter equal to a linear predictor which is used to model insurance claim amounts for various data sets. The performance is compared with those of the generalized beta (of the second kind) and lognormal distributions.

**Keywords:** insurance claim; double Pareto lognormal distribution; heavy-tailed; generalized beta distribution of the second kind; EM algorithm

## 1. Introduction

Heavy-tailed distributions are an important tool for actuaries working in insurance where many insurable events have low likelihoods and high severities and the associated insurance policies require adequate pricing and reserving. In such cases the four-parameter generalized beta distribution of the second kind (GB2) and the three-parameter generalized gamma distribution fulfil this purpose, as demonstrated in McDonald (1990), Wills et al. (2006), Frees and Valdez (2008), Wills et al. (2006), Frees et al. (2014a) and Chapter 9 of Frees et al. (2014b). In fact, the set of possible distributions that could be used for long-tail analyses is much broader than suggested here and good references for these are Chapter 10 of Frees et al. (2014a), Chapter 9 of Frees et al. (2014b) and Section 4.11 of Kleiber and Kotz (2003). We propose in this article the use of the double Pareto lognormal (DPLN) distribution as an alternative model for heavy-tailed events.

The DPLN distribution was introduced by Reed (2003) to model the distribution of incomes. It occurs as the distribution of the stopped wealth where the wealth process is geometric Brownian motion, the initial wealth is lognormally distributed and the random stopping time is exponentially distributed. This parametric model exhibits Paretian behaviour in both tails, among other suitable theoretical properties, and there is favourable evidence of its fit to data in various applications, as demonstrated in Colombi (1990), Reed (2003), Reed and Jorgensen (2004) and Hajargasht and Griffiths (2013) for income data and Giesen et al. (2010) for settlement size data. Particular applications of the DPLN distribution to insurance and actuarial science have previously been given in Ramírez-Cobo et al. (2010) and Hürlimann (2014).

In this paper, the DPLN generalized linear model (GLM) is introduced by setting the location parameter equal to a linear predictor, i.e.,  $v = \beta^\top x$ , where  $\beta \in R^d$  is a vector of regression coefficients

and  $x$  is a vector of explanatory variables. Then the mean of the DPLN GLM is proportional to some exponential transformation of the linear predictor. Then an EM algorithm is developed which solves for the regression parameters.

Particular applications of the DPLN distribution to insurance and actuarial science have previously been given in [Ramírez-Cobo et al. \(2010\)](#) and [Hürlimann \(2014\)](#). However, another practical application of the DPLN GLM, beyond what is demonstrated in this article, is assessing variability of survival rates of employment, as done in [Yamaguchi \(1992\)](#).

Applying this generalized linear model, we model the claim severity for private passenger automobile insurance and claim amounts due to bodily injuries sustained in car accidents, the data sets of which are supplied in [Frees et al. \(2014a\)](#). We compare this predictive model with the generalized linear model derived from the generalized beta distribution of the second kind (GB2), which has been employed in modelling incomes, for example by [McDonald \(1990\)](#), and in modelling insurance claims, for example by [Wills et al. \(2006\)](#) and [Frees and Valdez \(2008\)](#). The EM algorithm has previously been applied to insurance data, for example in [Kocović et al. \(2015\)](#), where it is used to explain losses of a fire insurance portfolio in Serbia.

The rest of the paper is organized as follows. Section 2 explains how the DPLN model is applied to regression by setting the location parameter equal to a linear predictor. In Section 3, details of parameter estimation by the method of maximum likelihood using the related normal skew-Laplace (NSL) distribution are provided, where we develop an EM algorithm for such a purpose. Section 4 gives numerical applications of the models to two insurance-related data sets and makes comparisons of fits with the LN and GB2 distributions using the likelihood ratio test and another test for non-nested models due to [Vuong \(1989\)](#). Out-of-sample performances of the models in regards to capital requirements of an insurer are also provided. Section 5 concludes.

## 2. DPLN Generalized Linear Model

As mentioned previously, the DPLN distribution can be obtained as a randomly stopped geometric Brownian motion whose initial value is lognormally distributed. Therefore, without any mathematical analysis, stopping the geometric Brownian motion at the initial time with probability one will give the lognormal distribution. If the diffusion coefficient of the geometric Brownian motion is set to zero, we have a deterministic geometric motion which is stopped at an exponentially distributed time, giving the PLN distribution. If also the drift coefficient of the geometric Brownian motion is set to zero then the lognormal distribution results. Another degenerate case emerges when the initial value is constant, that is, when its variance is zero, giving the Pareto distribution. This gives us a sensible intuition on the mathematical derivations in this section.

Formally, given a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ , where  $\Omega$  is the sample space,  $\mathcal{F}$  is the  $\sigma$ -algebra of events,  $(\mathcal{F}_t)_{t \geq 0}$  is the filtration of sub- $\sigma$ -algebras of  $\mathcal{F}$  and  $P$  is a probability measure, we consider the adapted stochastic process  $Y$  specified by the following stochastic differential equation (SDE), for  $t \geq 0$ ,

$$dY_t = \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t, \quad (1)$$

where  $\mu$  and  $\sigma \geq 0$  are constants and  $W$  is a Wiener process adapted to the filtration  $(\mathcal{F})_{t \geq 0}$ .

Then for a fixed time  $t \geq 0$ , the random variable  $Y_t$  can be written as

$$Y_t = Y_0 + \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t. \quad (2)$$

Now if  $Y_0$  is a random variable, dependent upon the vector of predictor variables  $x = (x_1, x_2, \dots, x_d)^\top \in R^d$ , such that  $Y_0 \sim N(\nu, \tau^2)$ , where  $\nu = \beta^\top x$ , and if we stop the process  $Y$

randomly at time  $t = T$ , where  $T \sim \text{Exp}(\lambda)$ , then  $Y_T$  is a normal skew Laplace (NSL) distributed random variable regressed on the vector of predictors  $x$ , that is

$$Y_T \sim \text{NSL}(\nu, \tau^2, \lambda_1, \lambda_2), \tag{3}$$

the exponential of which is a DPLN distributed random variable  $V_T$  (see [Reed and Jorgensen \(2004\)](#) for more details) dependent on the same predictors, namely

$$V_T = \exp(Y_T) \sim \text{DPLN}(\nu, \tau^2, \lambda_1, \lambda_2), \tag{4}$$

where  $\nu = \beta^\top x$ . As indicated previously, the particular case of the PLN distribution arises when  $\sigma = 0$  in (1) and the case of the lognormal distribution (LN) arises when  $\mu = 0$  and  $\sigma = 0$  in (1).

The moment generating function (MGF) of  $Y_T$  is

$$\text{MGF}_{Y_T}(s) = \text{MGF}_{\log X_0}(s) \text{MGF}_{T_1 - T_2}(s), \tag{5}$$

where  $T_1 \sim \text{Exp}(\lambda_1)$  and  $T_2 \sim \text{Exp}(\lambda_2)$  are exponentially distributed random variables with

$$\begin{aligned} \lambda_1 &= \frac{1}{\sigma^2} \left[ \sqrt{(\mu - \frac{1}{2}\sigma^2)^2 + 2\lambda\sigma^2} - (\mu - \frac{1}{2}\sigma^2) \right], \\ \lambda_2 &= \frac{1}{\sigma^2} \left[ \sqrt{(\mu - \frac{1}{2}\sigma^2)^2 + 2\lambda\sigma^2} + (\mu - \frac{1}{2}\sigma^2) \right], \end{aligned} \tag{6}$$

as given in [Reed and Jorgensen \(2004\)](#). This product of MGFs demonstrates that the NSL distributed random variable  $Y_T$  can be expressed as the sum

$$Y_T = \log X_0 + T_1 - T_2, \tag{7}$$

where  $\log X_0 \sim N(\nu, \tau^2)$  and  $T_1$  and  $T_2$  are as above.

The probability density function (PDF) of  $V_T = \exp(Y_T)$ , as in (4), is given by

$$\begin{aligned} f_{V_T}(x) &= \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \frac{1}{x} \left[ \exp \left\{ \frac{1}{2} \tau^2 \lambda_1^2 - \lambda_1 (\log x - \nu) \right\} \Phi \left( \frac{\log x - \tau^2 \lambda_1 - \nu}{\tau} \right) \right. \\ &\quad \left. + \exp \left\{ \frac{1}{2} \tau^2 \lambda_2^2 - \lambda_2 (\nu - \log x) \right\} \Phi^c \left( \frac{\log x + \tau^2 \lambda_2 - \nu}{\tau} \right) \right], \end{aligned} \tag{8}$$

also given in [Reed \(2003\)](#). Because we will work with logarithms of DPLN variates we will make use of the PDF of  $Y_T$  given by

$$\begin{aligned} f_{Y_T}(y) &= \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \left[ \exp \left\{ \frac{1}{2} \tau^2 \lambda_1^2 - \lambda_1 (y - \nu) \right\} \Phi \left( \frac{y - \tau^2 \lambda_1 - \nu}{\tau} \right) \right. \\ &\quad \left. + \exp \left\{ \frac{1}{2} \tau^2 \lambda_2^2 - \lambda_2 (\nu - y) \right\} \Phi^c \left( \frac{y + \tau^2 \lambda_2 - \nu}{\tau} \right) \right], \end{aligned} \tag{9}$$

where  $\Phi(\cdot)$  and  $\Phi^c(\cdot)$  represent the cumulative distribution function and survival function of the standard normal distribution respectively.

Additional properties concerning the moments of the DPLN and asymptotic tail behaviour can be found in [Reed \(2003\)](#). When we incorporate a vector of explanatory variables (or covariates)  $x \in R^d$  in our model, the location parameter  $\nu$  is set equal to  $\beta^\top x$ , where  $\beta \in R^d$  is a vector of regression coefficients, and the mean of the response variable is given by

$$E[V_T | x] = \frac{\lambda_1 \lambda_2}{(\lambda_1 - 1)(\lambda_2 + 1)} \exp \left\{ \beta^\top x + \frac{1}{2} \tau^2 \right\}, \tag{10}$$

where  $\tau, \lambda_2 > 0$  and  $\lambda_1 > 1$ . Each regression coefficient can be interpreted as a proportional change in the mean of the response variable per unit change in the corresponding covariate.

For a random sample coming from  $V_T$

$$v_1, v_2, \dots, v_n \quad (11)$$

and corresponding vectors of covariates  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}$ , we will use the maximum likelihood estimation described in the following section to compute the parameters of the DPLN distribution.

### 3. Maximum Likelihood Estimation of Parameters

#### 3.1. Methods of Estimation

Given the random sample in (11) and corresponding vectors of covariates, there are several ways of estimating parameters, such as moment matching, where such moments exist, and maximum likelihood estimation (MLE). As we are dealing with heavy-tailed distributions, moment matching may not be possible and we therefore resort to maximum likelihood estimation. Maximum likelihood estimators are also preferable since for large samples the estimators are unbiased, efficient and normally distributed. The EM algorithm of [Dempster et al. \(1977\)](#) is one approach to performing MLE of parameters, which we describe in the next section. Another approach is based on the gradient ascent method which we discuss in a subsequent section.

#### 3.2. Application of the EM Algorithm to DPLN Generalized Linear Model

##### 3.2.1. The EM Algorithm for the DPLN GLM

Our task is to obtain maximum likelihood estimates of the parameters of the model  $DPLN(\nu, \tau^2, \lambda_1, \lambda_2)$  using the EM algorithm, which was developed in [Dempster et al. \(1977\)](#). Because an NSL random variable is the logarithm of a DPLN random variable, fitting a DPLN distribution to the observations in (11) is the same as fitting the NSL distribution to the logarithms  $y_1, y_2, \dots, y_n$  of these observations. The EM algorithm starts from an initial estimate of parameter values and sequentially computes refined estimates which increase the value of the log-likelihood function. In the following paragraphs we explain how it is applied to the DPLN distribution.

Suppose that  $\theta = (\beta, \tau^2, \lambda_1, \lambda_2)$  is an initial estimate of the parameters of the distribution of the random variable  $Y$  whose density function is  $f_Y$ , given in (9). Let  $\theta'$  denote a refined estimate of the parameters of the distribution, that is, an estimate for which the log-likelihood function  $\ell$  exceeds that of the initial estimate  $\theta$ . In what follows, we demonstrate how to generate a refined estimate for the DPLN GLM. For the refined estimate  $\theta'$ , we can write the log-likelihood function as

$$\ell(\theta') = \sum_{i=1}^n \log f_Y(y_i; \theta') = \sum_{i=1}^n \log \int_{-\infty}^{\infty} f_{Y,Z}(y_i, z; \theta') dz, \quad (12)$$

where  $f_Y(y)$  is the PDF of the random variable  $Y$ ,  $f_{Y,Z}(y, z)$  is the joint density function of the random variables  $(Y, Z)$  and where the random variable  $Z$  is latent and therefore unobserved. In our case, the random variable  $Z$  is a normally distributed random variable having parameters  $\nu$  and  $\tau^2$ , as indicated by the random variable  $\log X_0$  in (7).

We now give the probability density function  $g_i$ , for each  $i = 1, 2, \dots, n$ , for the conditional random variable  $Z|Y = y_i$  which only depends on the initial estimate  $\theta$  of parameters, and not on  $\theta'$ , namely

$$g_i(z) = f_{Y,Z}(y_i, z; \theta) / f_Y(y_i; \theta) = f_{Z|Y=y_i}(z). \quad (13)$$

We then rewrite (12) as

$$\begin{aligned} \ell(\theta') &= \sum_{i=1}^n \log \int_{-\infty}^{\infty} f_{Y,Z}(y_i, z; \theta') dz \\ &= \sum_{i=1}^n \log \int_{-\infty}^{\infty} \frac{f_{Y,Z}(y_i, z; \theta')}{g_i(z)} g_i(z) dz \end{aligned} \tag{14}$$

and applying Jensen’s inequality  $\log E[X] \geq E[\log X]$  gives

$$\begin{aligned} \ell(\theta') &= \sum_{i=1}^n \log \int_{-\infty}^{\infty} \frac{f_{Y,Z}(y_i, z; \theta')}{g_i(z)} g_i(z) dz \\ &\geq \sum_{i=1}^n \int_{-\infty}^{\infty} \log \left\{ \frac{f_{Y,Z}(y_i, z; \theta')}{g_i(z)} \right\} g_i(z) dz \\ &= \sum_{i=1}^n \int_{-\infty}^{\infty} \log \{ f_{Y,Z}(y_i, z; \theta') \} g_i(z) dz \\ &\quad - \sum_{i=1}^n \int_{-\infty}^{\infty} \log \{ g_i(z) \} g_i(z) dz. \end{aligned} \tag{15}$$

So our maximization of likelihood amounts to maximizing

$$\sum_{i=1}^n \int_{-\infty}^{\infty} \log \{ f_{Y,Z}(y_i, z; \theta') \} g_i(z) dz \tag{16}$$

with respect to  $\theta'$ , which is the M-step or maximization-step of the EM algorithm.

In practice, there is an E-step or expectations-step of the algorithm which is performed prior to the M-step, however we continue with the M-step in the next section because this identifies the variables whose expectations are to be computed in the E-step.

### 3.2.2. M-Step

So we need to maximize (16) with respect to the parameter  $\theta'$ . We show how this is done for the double Pareto lognormal distribution by expanding out the terms in (16), giving

$$\begin{aligned} &\sum_{i=1}^n \int_{-\infty}^{\infty} \log \{ f_{Y,Z}(y_i, z; \theta') \} g_i(z) dz \\ &= \sum_{i=1}^n \int_{-\infty}^{\infty} \log \left[ \frac{1}{\sqrt{2\pi(\tau')^2}} \exp \left( -\frac{(z-v'_i)^2}{2(\tau')^2} \right) \right. \\ &\quad \left. \times \frac{\lambda'_1 \lambda'_2}{\lambda'_1 + \lambda'_2} \begin{cases} \exp(\lambda'_2(y_i - z)), & \text{if } z > y_i \\ \exp(-\lambda'_1(y_i - z)), & \text{if } z \leq y_i \end{cases} \right] g_i(z) dz, \end{aligned} \tag{17}$$

which becomes

$$\begin{aligned} &n \log \left( \frac{1}{\sqrt{2\pi(\tau')^2}} \right) - \frac{1}{2(\tau')^2} \sum_{i=1}^n z_i^{(2)} + \frac{1}{(\tau')^2} \sum_{i=1}^n z_i v'_i - \frac{1}{2(\tau')^2} \sum_{i=1}^n (v'_i)^2 \\ &+ n \log \frac{\lambda'_1 \lambda'_2}{\lambda'_1 + \lambda'_2} + \lambda'_2 \sum_{i=1}^n w_i^- - \lambda'_1 \sum_{i=1}^n w_i^+, \end{aligned} \tag{18}$$

where

$$\begin{aligned} z_i &= \int_{-\infty}^{\infty} z g_i(z) dz, \quad z_i^{(2)} = \int_{-\infty}^{\infty} z^2 g_i(z) dz, \\ w_i^+ &= \int_{-\infty}^{\log x_i} (\log x_i - z) g_i(z) dz, \quad w_i^- = \int_{\log x_i}^{\infty} (\log x_i - z) g_i(z) dz. \end{aligned} \tag{19}$$

We arrive at the following Theorem giving the optimum parameter vector  $\theta'$  and whose proof follows the reasoning for the simpler case without explanatory variables given in [Reed and Jorgensen \(2004\)](#).

**Theorem 1.** The components of the parameter vector  $\theta'$  which maximise (16) are

$$\begin{aligned} \beta' &= (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{X}\mathbf{Z}, \quad (\tau')^2 = \frac{1}{n} \left( \mathbf{1}^\top \mathbf{Z}^{(2)} - \mathbf{Z}^\top \mathbf{X}^\top (\mathbf{X}\mathbf{X}^\top)^{-1} \mathbf{X}\mathbf{Z} \right), \\ \lambda'_1 &= \frac{1}{P' + \sqrt{P'Q'}}, \quad \lambda'_2 = \frac{1}{Q' + \sqrt{P'Q'}}, \end{aligned} \tag{20}$$

where

$$P' = \frac{1}{n} \sum_{i=1}^n w_i^+, \quad Q' = -\frac{1}{n} \sum_{i=1}^n w_i^-, \tag{21}$$

$$\mathbf{Z} = (z_1, z_2, \dots, z_n)^\top, \quad \mathbf{Z}^{(2)} = (z_1^{(2)}, z_2^{(2)}, \dots, z_n^{(2)})^\top \tag{22}$$

and  $\mathbf{X}$  is the matrix of predictor variables

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^{(1)} & \mathbf{x}^{(2)} & \dots & \mathbf{x}^{(n)} \end{pmatrix}. \tag{23}$$

**Proof.** See Appendix A.1.  $\square$

### 3.2.3. E-Step

Here we compute the conditional distributions which are used in the E-step. For the set of  $n$  logarithms  $y_1, \dots, y_n$  of observations, the maximum likelihood estimates of the parameters can be obtained using the EM algorithm with as follows,

$$g_i(z) = \frac{f_Z(z; \theta) f_W(y_i - z; \theta)}{f_Y(y_i; \theta)}, \tag{24}$$

where the density functions  $f_Z$ ,  $f_W$  and  $f_Y$  are defined as

$$\begin{aligned} f_Z(z; \theta) &= \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{1}{2\tau^2} (z - \nu)^2 \right\}, \\ f_W(w; \theta) &= \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \begin{cases} \exp(\lambda_2 w) & w < 0 \\ \exp(-\lambda_1 w) & w \geq 0 \end{cases} \\ f_Y(y; \theta) &= \phi((y - \nu)/\tau) \{ R(\lambda_1 \tau - (y - \nu)/\tau) + R(\lambda_2 \tau + (y - \nu)/\tau) \}. \end{aligned} \tag{25}$$

For this choice of  $g_i$ ,  $\log f_{Y,Z}(y_i, z; \theta')$  in (16) becomes

$$\log f_{Y,Z}(y_i, z; \theta') = \log f_Z(z; \theta') + \log f_W(y_i - z; \theta'). \tag{26}$$

Our E-step of the EM algorithm is given in the following Theorem, most of the equations of which are mentioned in [Reed and Jorgensen \(2004\)](#) but for which we supply explicit proofs.

**Theorem 2.** The expectations in our E-step are as follows

$$\begin{aligned} z_i &= \int_{-\infty}^{\infty} z g_i(z) dz = \nu + \tau^2 \frac{\lambda_1 R(p_i) - \lambda_2 R(q_i)}{R(p_i) + R(q_i)}, \\ z_i^{(2)} &= \int_{-\infty}^{\infty} z^2 g_i(z) dz = \nu^2 + \tau^2 - \tau^2 \frac{p_i + q_i}{R(p_i) + R(q_i)} \\ &\quad + \tau^2 \frac{(2\nu\lambda_1 + \lambda_1^2 \tau^2) R(p_i) + (\lambda_2^2 \tau^2 - 2\nu\lambda_2) R(q_i)}{R(p_i) + R(q_i)}, \\ w_i^+ &= \int_{-\infty}^{y_i} (y_i - z) g_i(z) dz = \tau \frac{-p_i R(p_i) + 1}{R(p_i) + R(q_i)}, \\ w_i^- &= \int_{y_i}^{\infty} (y_i - z) g_i(z) dz = \tau \frac{q_i R(q_i) - 1}{R(p_i) + R(q_i)}, \end{aligned} \tag{27}$$

where  $p_i = -(y_i - v)/\tau + \lambda_1\tau$  and  $q_i = (y_i - v)/\tau + \lambda_2\tau$ .

**Proof.** See Appendix A.2.  $\square$

### 3.2.4. Standard Errors

The standard errors of the estimates  $\hat{\theta} = (\hat{\beta}^\top, \hat{\tau}^2, \hat{\lambda}_1, \hat{\lambda}_2)$  can be estimated in the last iteration of the EM algorithm, as shown in Louis (1982). The observed Fisher information matrix evaluated at  $\hat{\theta}$  based on the observations  $\{v_i\}_{i=1}^n$  can be approximated by

$$\mathcal{I}(\hat{\theta}; v_i) \approx \sum_{i=1}^n \frac{\partial}{\partial \theta} \log f_Y(\log v_i; \hat{\theta}) \left[ \frac{\partial}{\partial \theta} \log f_Y(\log v_i; \hat{\theta}) \right]^\top, \tag{28}$$

where  $f_Y$  is as in (9). Since  $E \left[ \frac{\partial}{\partial \theta} \log f_Y(y_i; \theta) \middle| y_i; \hat{\theta} \right] = \frac{\partial}{\partial \theta} \log f_Y(y_i; \hat{\theta})$ , (28) is equivalent to

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} E \left[ \log f_Y(y_i; \theta) \middle| y_i; \hat{\theta} \right] \left( \frac{\partial}{\partial \theta} E \left[ \log f_Y(y_i; \theta) \middle| y_i; \hat{\theta} \right] \right)^\top.$$

In particular,

$$\begin{aligned} & \frac{\partial}{\partial \theta} E \left[ \log f_Y(y_i; \theta) \right] \\ &= \left( \frac{\partial E \left[ \log f(u_i; v_i, \tau) \middle| y_i; \hat{\theta} \right]}{\partial(\beta, \tau)^\top}, \frac{\partial E \left[ \log f_Y(w_i; \lambda_1, \lambda_2) \middle| y_i; \hat{\theta} \right]}{\partial(\lambda_1, \lambda_2)^\top} \right)^\top, \end{aligned}$$

and therefore these expressions are available in the last iteration of the EM algorithm.

### 3.3. Gradient Ascent Method

The gradient ascent method is applied to the likelihood function of the normal skew Laplace distribution in this subsection. Let be  $\mathbf{y} = (y_1, \dots, y_n)$  be a random sample of size  $n$  from a NSL distribution. Its log-likelihood function is:

$$\begin{aligned} \ell(\mathbf{y}; \beta, \tau^2, \lambda_1, \lambda_2) &= n \log \lambda_1 + n \log \lambda_2 - n \log(\lambda_1 + \lambda_2) \\ &+ \sum_{i=1}^n \log \left( \frac{\Phi \left( \frac{y_i - \tau^2 \lambda_1 - v_i}{\tau} \right)}{\exp \left\{ -\frac{1}{2} \lambda_1 \tau^2 + \lambda_1 (y_i - v_i) \right\}} + \frac{1 - \Phi \left( \frac{y_i + \tau^2 \lambda_2 - v_i}{\tau} \right)}{\exp \left\{ -\frac{1}{2} \lambda_2 \tau^2 + \lambda_2 (v_i - y_i) \right\}} \right). \end{aligned} \tag{29}$$

The solutions of the  $d + 3$  score equations that are shown in the Appendix, provide the maximum likelihood estimates of  $\lambda_1, \lambda_2, \tau$  and  $\{\beta_j\}_{j=1, \dots, d}$ , which can be obtained by numerical methods such as Newton-Raphson algorithm. Alternatively, parameter estimates can be obtained directly via a grid search for the global maximum of the log-likelihood surface given by (29), or equivalently by maximizing the log-likelihood function derived from the expression (8). We have used FindMaximum function of Mathematica software package v.11.0. Since the global maximum of the log-likelihood surface is not guaranteed, different initial values of the parametric space can be considered as seed point using different methods of maximization, such as Newton–Raphson method, Principal Axis method and the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS), among others. The standard errors of the estimates have been approximated by inverting the Hessian matrix and the relevant partial derivatives can be approximated well by finite differences.



#### 4. Numerical Applications

In this section, two well-known data sets in the actuarial literature that can be downloaded from Professor E. Frees' personal website<sup>1</sup> will be considered to test the practical performance of the DPLN generalized linear model. For the two data sets considered, the EM algorithm for the DPLN GLM was stopped when the relative change of the log-likelihood function was smaller than  $1 \times 10^{-10}$ . The initial values were calculated by using the estimates of the lognormal GLM and the estimates of the parameters  $\lambda_1$  and  $\lambda_2$  for the model without covariates.

Because we are comparing the DPLN generalized linear model with the GB2 GLM, we give here some rudimentary facts concerning the GB2 distribution. Let  $Z$  be a random variable having the  $Beta(p, q)$  distribution, for  $p, q \in (0, \infty)$ , as defined in Chapter 6 of Kleiber and Kotz (2003). Then, for  $v \in (-\infty, \infty)$  and  $\tau \in (0, \infty)$ , the random variable

$$V = \exp(v)(Z/(1 - Z))^\tau \quad (30)$$

has the GB2 distribution and its probability density function can be written as

$$f_V(v) = \frac{\exp\left\{\frac{p(\log v - v)}{\tau}\right\}}{v \tau B(p, q) [1 + \exp\left\{\frac{(\log v - v)}{\tau}\right\}]^{p+q}}, \quad (31)$$

where  $v \in (0, \infty)$ ,  $v$  is a location parameter,  $\tau > 0$  is a scale parameter,  $p > 0$  and  $q > 0$  are shape parameters and

$$B(p, q) = \int_0^1 z^{p-1}(1 - z)^{q-1} dz = \Gamma(p)\Gamma(q)/\Gamma(p + q)$$

is the Beta function. As for the aforementioned distributions, to include explanatory variables in the model, we let the location parameter be a linear function of covariates, i.e.,  $v = \beta^\top x$ . The  $k$ -th moment is easily seen to be

$$E[V^k] = \exp(kv) \frac{B(p + k\tau, q - k\tau)}{B(p, q)}, \quad (32)$$

where  $k \in (-p/\tau, q/\tau)$ , and looking at the case  $k = 1$  we can interpret each of the regression coefficients  $\beta_i$ ,  $i = 1, \dots, d$ , as being the proportional sensitivity of the mean to the corresponding covariate. Further details of this model can be found in Frees et al. (2014a). Parameter estimation for the GB2 GLM has been performed via a grid search for the global maximum of the log-likelihood surface associated to this model. We have used FindMaximum function of *Mathematica* software package v.11.0.

##### 4.1. Example 1: Automobile Insurance

The first data set pertains to claims experience from a large midwestern (US) property and casualty insurer for private passenger automobile insurance.

The dependent variable is the amount paid on a closed claim, in US\$. The sample includes 6773 claims. The following explanatory variables have been considered to explain the claims amount:

- *GENDER*, gender of operator, takes the value 1 if female and 0 otherwise;
- *AGE*, age of operator;
- *CLASS* rating class of operator as coded in Table 1.

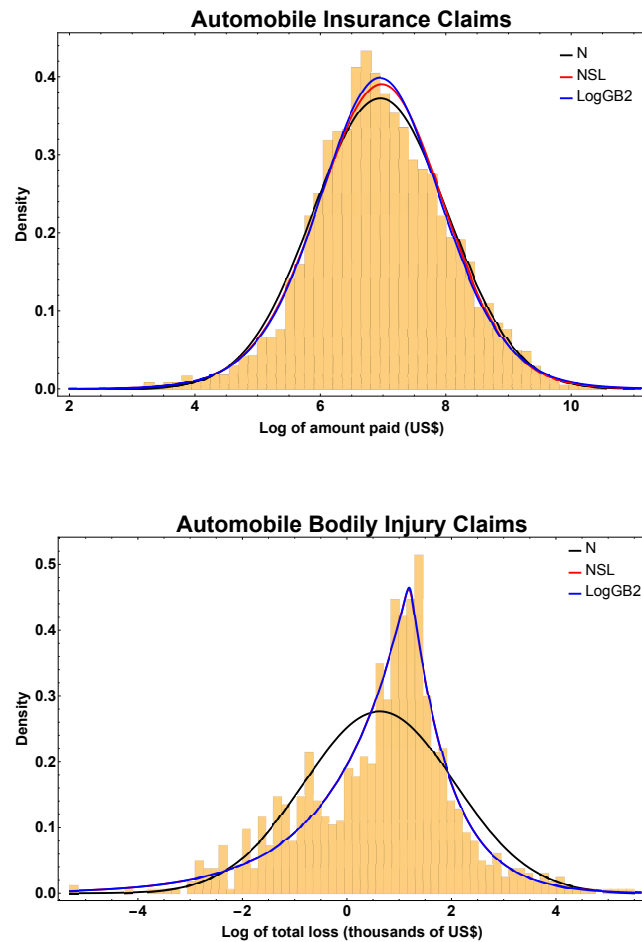
In the top part of Figure 1 the histogram of the Automobile insurance claims is exhibited in logarithmic scale. This dataset is quite symmetrical but it presents a slightly longer lower tail. For that reason the DPLN distribution seems suitable to explain this dataset.

<sup>1</sup> <http://instruction.bus.wisc.edu/jfrees/jfreesbooks/RegressionModeling/BookWebDec2010/data.htm>.



**Table 1.** Parameter estimates, standard errors (S.E.) and  $p$ -values of the  $t$ -test for automobile insurance claims dataset under lognormal distribution (LN), generalized beta distribution of the second kind (GB2) and double Pareto lognormal distribution (DPLN) generalized linear models.

Generalized Linear Model			
Estimate (S.E.)	LN	GB2	DPLN
<i>INTERCEPT</i>	7.184 (0.150)	7.234 (0.163)	7.260 (0.080)
<i>p</i> -value	<0.0001	<0.0001	<0.0001
<i>GENDER</i>	−0.035 (0.027)	−0.012 (0.027)	−0.039 (0.014)
<i>p</i> -value	0.1918	0.6604	0.0073
<i>AGE</i>	−0.004 (0.002)	−0.004 (0.002)	−0.005 (0.001)
<i>p</i> -value	0.0167	0.0110	<0.0001
<i>C1</i>	0.018 (0.118)	−0.002 (0.115)	0.017 (0.063)
<i>p</i> -value	0.8760	0.9877	0.7889
<i>C11</i>	0.063 (0.116)	0.021 (0.114)	0.063 (0.062)
<i>p</i> -value	0.5853	0.8567	0.3146
<i>C1A</i>	−0.076 (0.165)	−0.047 (0.161)	−0.085 (0.088)
<i>p</i> -value	0.6453	0.7687	0.3389
<i>C1B</i>	0.057 (0.122)	0.008 (0.120)	0.055 (0.066)
<i>p</i> -value	0.6411	0.9471	0.4045
<i>C1C</i>	−0.164 (0.206)	−0.154 (0.203)	−0.1392 (0.110)
<i>p</i> -value	0.4267	0.4498	0.2075
<i>C2</i>	−0.134 (0.176)	0.034 (0.170)	−0.132 (0.094)
<i>p</i> -value	0.4450	0.8407	0.1626
<i>C6</i>	0.070 (0.120)	0.033 (0.118)	0.086 (0.065)
<i>p</i> -value	0.5594	0.7767	0.1815
<i>C7</i>	−0.030 (0.116)	−0.028 (0.114)	−0.033 (0.062)
<i>p</i> -value	0.7983	0.8071	0.5960
<i>C71</i>	0.018 (0.115)	−0.029 (0.113)	0.013 (0.062)
<i>p</i> -value	0.8725	0.7941	0.8380
<i>C72</i>	0.239 (0.160)	0.036 (0.157)	0.226 (0.086)
<i>p</i> -value	0.1367	0.8203	0.0087
<i>C7A</i>	0.127 (0.150)	0.225 (0.147)	0.123 (0.080)
<i>p</i> -value	0.3965	0.1249	0.1248
<i>C7B</i>	0.128 (0.118)	0.091 (0.116)	0.129 (0.063)
<i>p</i> -value	0.2806	0.4313	0.042
<i>C7C</i>	0.282 (0.162)	0.173 (0.158)	0.270 (0.087)
<i>p</i> -value	0.0824	0.2735	0.0020
<i>F1</i>	0.103 (0.228)	−0.134 (0.222)	0.132 (0.122)
<i>p</i> -value	0.6499	0.5462	0.2785
<i>F11</i>	−0.087 (0.203)	−0.177 (0.202)	−0.099 (0.109)
<i>p</i> -value	0.6675	0.3798	0.3623
<i>F6</i>	0.058 (0.144)	0.069 (0.142)	0.090 (0.077)
<i>p</i> -value	0.6880	0.6300	0.2434
<i>F7</i>	−0.347 (0.178)	−0.382 (0.172)	−0.351 (0.095)
<i>p</i> -value	0.0508	0.0266	0.0002
$\tau$	1.068 (0.009)	0.968 (0.111)	0.810 (0.006)
<i>p</i> -value	<0.0001	<0.0001	<0.0001
$p$ or $\lambda_1$		2.083 (0.371)	2.127 (0.032)
<i>p</i> -value		<0.0001	<0.0001
$q$ or $\lambda_2$		2.109 (0.427)	1.952 (0.029)
<i>p</i> -value		0.0001	<0.0001
NLL	57,164.4	57,145.2	57,139.3
AIC	11,4370.7	11,4336.6	11,4324.6
BIC	114,513.9	114,493.2	114,481.6
CT	3.0108	95.4570	91.1358



**Figure 1.** Empirical distribution of the logarithm of automobile insurance claims (**above**) and logarithm of automobile bodily injury claims (**below**). The log transformation of the LN (N, black), DPLN (normal skew-Laplace (NSL), red) and GB2 (*LogGB2*, blue) distributions have been superimposed.

#### 4.1.1. Model Without Covariates

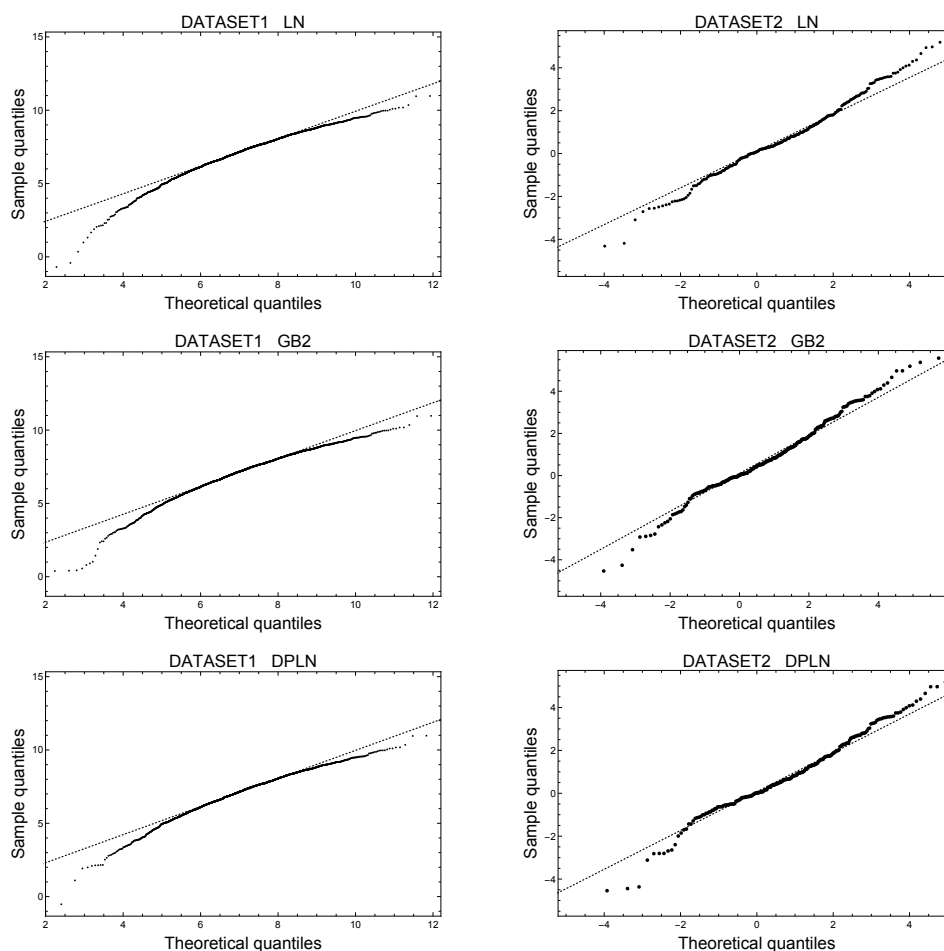
Here, for comparison purposes only, the lognormal, DPLN and GB2 distributions will be used to describe the total losses (e.g., when explanatory variables are not considered). Firstly, the automobile insurance claims dataset is examined. Table 2 summarizes parameters estimates obtained by maximum likelihood with corresponding standard errors (in brackets) for the aforementioned distributions.

In respect of model selection, we also provide the negative of the maximum of the log-likelihood (NLL), Akaike's information criterion (AIC) and Bayesian information criterion (BIC) results in the table. Note that for all three measures of model validation, smaller values indicate a better fit of the model to the empirical data. As expected, the lognormal distribution exhibits the worst performance in terms of all three measures of model validation. In the top part of Figure 2, we have superimposed the log transformation of these three distributions to the empirical distribution of the log of the claims sizes to test the fit in both tails. It is evident that the log transformation of the lognormal distribution (black curve), i.e., Normal distribution (N), provides the worst fit due to asymmetry of the data. The logGB2 (blue curve) and NSL distributions (red curve) give better fit to data as measured by the NLL, AIC and BIC, although the latter model adheres closely to the data. Although it is not shown in Table 2, the PLN distribution, replicates the fit of the LN distribution and the value of the shape parameter that controls the right tail tends to infinity. The computing times (CT) in seconds to estimate the maximum likelihood estimates by directly maximizing the log-likelihood surface for these distributions are shown in the last row of the table. The computing time of the EM algorithm for the DPLN GLM was

1145.86 s using the stopping criterion that the relative change of the log-likelihood function be less than  $1 \times 10^{-4}$ .

**Table 2.** Model fitting results of the LN, GB2 and DPLN distributions regarding automobile insurance claims.

Estimate (S.E.)	Distribution		
	LN	GB2	DPLN
$\nu$	6.956 (0.013)	6.945 (0.074)	7.009 (0.007)
$\tau$	1.071 (0.009)	0.916 (0.089)	0.824 (0.006)
$p$ or $\lambda_1$		1.914 (0.289)	2.191 (0.033)
$q$ or $\lambda_2$		1.897 (0.316)	1.961 (0.029)
NLL	57,185.1	57,162.5	57,161.5
AIC	114,374	114,333	114,331
BIC	114,390	114,360	114,358
CT	0.2340	3.8376	12.5113



**Figure 2.** QQ-plots of the log-residuals for LN (above), GB2 (middle) and DPLN (below) generalized linear models for automobile insurance claims data (left panel) and automobile bodily injury claims (right panel).

### 4.1.2. Comparison of Estimation from Simulations

In this section we compare the methods of estimating parameters by conducting the following simulation experiment. For the lognormal, DPLN and GB2 distributions, we simulate values based on the corresponding parameter estimates given in Table 2, and then, using as appropriate either standard formulae or the EM algorithm, compute parameter estimates from the 1000 simulated data sets of size  $N = 100, 200, 300, 400, 500, 1000$ . The results are shown in Table 3, where it is evident that increasing the sample size increases the accuracy of the parameter estimate. Of course, the true parameter values are given in Table 2, and these are the limits of the estimates as the sample size  $N$  increases. Importantly, the standard errors of the parameter estimates in Table 3 are noticeably smaller for the DPLN distribution, highlighting the consistency of the parameter estimation for the DPLN model. It is noteworthy to observe that the parameter estimates of the GB2 distribution are unstable for small sample sizes, which is not the case for the DPLN model. This highlights an advantage of the DPLN model over the GB2 in this case.

**Table 3.** Results of the simulation experiment involving 1000 simulations of data sets of size  $N$ , with standard errors shown in brackets.

Sample Size $N$	Distribution		
	LN	DPLN	GB2
100	$\hat{\nu} = 6.9551 (0.1017)$ $\hat{\tau} = 1.0599 (0.0779)$	$\hat{\nu} = 7.0095 (0.1401)$ $\hat{\tau} = 0.8091 (0.1063)$ $\hat{\lambda}_1 = 2.3715 (0.5056)$ $\hat{\lambda}_2 = 2.1121 (0.4517)$	$\hat{\nu} = 7.0670 (2.1561)$ $\hat{\tau} = 1.3188 (6.6295)$ $\hat{p} = 146.9750 (2403.6500)$ $\hat{q} = 180.7270 (3075.6600)$
200	$\hat{\nu} = 6.9565 (0.0763)$ $\hat{\tau} = 1.0684 (0.0540)$	$\hat{\nu} = 7.0098 (0.1015)$ $\hat{\tau} = 0.8176 (0.0750)$ $\hat{\lambda}_1 = 2.3024 (0.3785)$ $\hat{\lambda}_2 = 2.0471 (0.3450)$	$\hat{\nu} = 6.9789 (0.5097)$ $\hat{\tau} = 0.5591 (1.7972)$ $\hat{p} = 11.3055 (204.7950)$ $\hat{q} = 12.6484 (231.6380)$
300	$\hat{\nu} = 6.9569 (0.0610)$ $\hat{\tau} = 1.0668 (0.0445)$	$\hat{\nu} = 7.0038 (0.0859)$ $\hat{\tau} = 0.8235 (0.0636)$ $\hat{\lambda}_1 = 2.2636 (0.3131)$ $\hat{\lambda}_2 = 2.0389 (0.2892)$	$\hat{\nu} = 6.9602 (0.5468)$ $\hat{\tau} = 0.3635 (0.3605)$ $\hat{p} = 1.0887 (5.1770)$ $\hat{q} = 1.7380 (26.1729)$
400	$\hat{\nu} = 6.9578 (0.0531)$ $\hat{\tau} = 1.0695 (0.0368)$	$\hat{\nu} = 7.0031 (0.0772)$ $\hat{\tau} = 0.8211 (0.0560)$ $\hat{\lambda}_1 = 2.2368 (0.2761)$ $\hat{\lambda}_2 = 2.0189 (0.2553)$	$\hat{\nu} = 6.9411 (0.2291)$ $\hat{\tau} = 0.3052 (0.2085)$ $\hat{p} = 0.7963 (4.7610)$ $\hat{q} = 0.6528 (0.7841)$
500	$\hat{\nu} = 6.9555 (0.0465)$ $\hat{\tau} = 1.0709 (0.0333)$	$\hat{\nu} = 7.0104 (0.0665)$ $\hat{\tau} = 0.8208 (0.0489)$ $\hat{\lambda}_1 = 2.2289 (0.2450)$ $\hat{\lambda}_2 = 1.9930 (0.2162)$	$\hat{\nu} = 6.9409 (0.0843)$ $\hat{\tau} = 0.2941 (0.1768)$ $\hat{p} = 0.6268 (0.4547)$ $\hat{q} = 0.6036 (0.4101)$
1000	$\hat{\nu} = 6.9553 (0.0335)$ $\hat{\tau} = 1.0699 (0.0242)$	$\hat{\nu} = 7.0077 (0.0488)$ $\hat{\tau} = 0.8208 (0.0358)$ $\hat{\lambda}_1 = 2.2055 (0.1776)$ $\hat{\lambda}_2 = 1.9722 (0.1558)$	$\hat{\nu} = 6.9493 (0.0526)$ $\hat{\tau} = 0.2680 (0.1085)$ $\hat{p} = 0.5409 (0.2539)$ $\hat{q} = 0.5344 (0.2460)$

### 4.1.3. Including Explanatory Variables

Making use of the above additional information, we aim to better explain the total losses in terms of the set of covariates by using the DPLN generalized linear model. For the purpose of comparison, we have also fitted the lognormal and GB2 generalized linear models. Here, we choose the identity link function for the location parameter.

From left to right in Table 1, the parameter estimates, standard errors (S.E.) and the corresponding  $p$ -values calculated based on the  $t$ -Wald statistics for the LN, GB2 and DPLN generalized linear

models are displayed for the automobile insurance claims dataset. The AIC and BIC values for each model are provided in the last two rows of the table. For the  $i$ -th claimant, the number of total amount  $y_i$  follows (10) whose mean depends on the above set of covariates through the identity link function. The exponential of *INTERCEPT* coefficient 7.260 is proportional to the predicted loss amount when the values of the other explanatory variables are equal to 0. This estimate is statistically significant at the usual significance levels, i.e., 5% and 1%. In total, the estimates of 10 out of 23 parameters for the DPLN generalized linear models are statistically significant at the usual levels (i.e., 5% and 1%) including the scale and shape parameters. The results for the LN and GB2 generalized linear model are also exhibited in Table 1 to compare their behaviour with DPLN generalized linear model. As it can be seen the fit provided by the DPLN generalized linear model improves the one provided by GB2. For the DPLN generalized linear model, parameters were estimated by the method of maximum likelihood by maximizing the log-likelihood surface. The same estimates were achieved by the EM algorithm described in Section 3.2. The standard errors of the parameter estimates for the DPLN GLM were computed from the last iteration of the EM algorithm and also approximated by inverting the Hessian matrix. Similar values were obtained. The computing times in seconds to estimate the maximum likelihood estimates by directly maximizing the log-likelihood surface for these generalized linear models are shown in the last row of the table. The DPLN GLM shows a better performance than the GB2 counterpart. The computing time of the EM algorithm for the DPLN GLM was 2239.24 s using the stopping criterion that the relative change of the log-likelihood function be less than  $1 \times 10^{-4}$ .

#### 4.1.4. Model Validation

Now, we analyze model validation from a practical perspective. In this regard, LN generalized linear model can be seen as a limiting case of DPLN generalized linear model when both  $\lambda_1$  and  $\lambda_2$  tend to infinity. We are interested, by means of the likelihood ratio test, in determining whether the LN generalized linear model (null hypothesis) is preferable to DPLN generalized linear model (alternative hypothesis) in describing these datasets. The test statistic is  $T = 2(\ell_{LN} - \ell_{DPLN})$  where  $\ell_{LN}$  and  $\ell_{DPLN}$  represent the maximum of the log-likelihood function for the LN and DPLN generalized linear models respectively. Asymptotically, under certain regularity conditions (see for example Lehmann and Casella (1998))  $T$  follows a chi-square distribution with two degrees of freedom. We have that  $T = 2(-57179.69 + 57155.87) = 50.02$ , therefore the larger model (DPLN) is preferable to the smaller (LN) generalized linear model at the usual significance levels, i.e., 5% and 1% ( $p$ -value less than 0.0001).

Next, the likelihood ratio test proposed by Vuong (1989) for non-nested models will be considered as a tool for model diagnostic. The test statistic is

$$T = \frac{1}{\omega\sqrt{n}} \left( \ell_f(\hat{\theta}_1) - \ell_g(\hat{\theta}_2) - \log n \left( \frac{n_f}{2} - \frac{n_g}{2} \right) \right),$$

where

$$\omega^2 = \frac{1}{n} \sum_{i=1}^n \left[ \log \left( \frac{f(\hat{\theta}_1)}{g(\hat{\theta}_2)} \right) \right]^2 - \left[ \frac{1}{n} \sum_{i=1}^n \log \left( \frac{f(\hat{\theta}_1)}{g(\hat{\theta}_2)} \right) \right]^2$$

is the sample variance of the pointwise log-likelihood ratios and  $f$  and  $g$  represent the probability density function (pdf) of two different non-nested models,  $\hat{\theta}_1$  and  $\hat{\theta}_2$  are the maximum likelihood estimates of  $\theta_1$  and  $\theta_2$  and  $n_f$  and  $n_g$  are the number of estimated coefficients in the model with pdf  $f$  and  $g$  respectively. Note that the Vuong's statistic is sensitive to the number of estimated parameters in each model and therefore the test must be corrected for dimensionality. Under the null hypotheses,  $H_0 : E[\ell_f(\hat{\theta}_1) - \ell_g(\hat{\theta}_2)] = 0$  and  $T$  is asymptotically normally distributed. At the 5% significance level, the rejection region for this test in favor of the alternative hypothesis occurs when  $T > 1.96$ .

Now we compare the GB2 and DPLN generalized linear models in terms of Vuong's test. Under the null hypothesis the two models are equally close to the true but unknown specification. For our data set, the value of the test statistic is  $T = 1.00$  and we fail to reject  $H_0$ , and therefore differences between these two models do not exist.

#### 4.2. Example 2: Automobile Bodily Injury Claims

The second data set deals with automobile bodily injury claims sourced from the Insurance Research Council (IRC), a division of the American Institute for Chartered Property Casualty Underwriters and the Insurance Institute of America. The data, collected in 2002, contains demographic information about the claimants, attorney involvement and the economic losses (in thousands of US\$), among other variables. As some of these explanatory variables contain missing observations, we only consider those data items having no missing values, resulting in a sample of 1,091 losses from a single state. We use as the response variable the claimant's total economic loss. Also, additional information is available to explain the claimants' total economic losses. We employ the following factors as covariates in our model fitting:

- *ATTORNEY*, takes the value 1 if the claimant is represented by an attorney and 0 otherwise;
- *CLMSEX*, takes the value 1 if the claimant is male and 0 otherwise;
- *MARRIED*, takes the value 1 if the claimant is married and 0 otherwise;
- *SINGLE*, takes the value 1 if the claimant is single and 0 otherwise;
- *WIDOWED*, takes the value 1 if the claimant is widowed and 0 otherwise;
- *CLMINSUR*, whether or not the claimant's vehicle was uninsured (= 1 if yes and 0 otherwise);
- *SEATBELT*, whether or not the claimant was wearing the seatbelt/child restraint ing belt's vehicle was uninsured (= 1 if yes and 0 otherwise);
- *CLMAGE*, claimant's age.

The empirical distribution of this variable combines losses of small, moderate and large sizes which makes it suitable for fitting heavy-tailed distributions. It has other features such as unimodality, skewness and a long upper tail, indicating a high likelihood of extremely expensive events. In the bottom part of Figure 1 the histogram of the response variable of this data set is shown again in logarithmic scale. A heavy lower tail is evident when this scale is used.

##### 4.2.1. Model Without Covariates

The results for the bodily injury claims data are shown in Table 4. The GB2 and DPLN distributions give the best fit to data as measured by these three measures of model selection. As expected, the LN distribution has the worst performance due to the asymmetry of the data. Again, although it is not shown in Table 4, the three-parameter PLN model replicates the LN distribution. This is due to the fact that the former model is a limiting case of the latter when shape parameter  $\lambda_1$  tends to infinity. These results are also supported by the bottom part of Figure 1, where it can be seen that the log transformation of the GB2 distribution *LogGB2* (blue curve) and the NSL distribution (red curve) provide almost an identical fit to data. The MLEs for the DPLN distribution were obtained by using the EM algorithm whose starting parameter values  $\lambda_1$ ,  $\lambda_2$ ,  $\nu$  and  $\tau$  are those obtained by moment-matching the first four cumulants. These MLEs were confirmed by those obtained directly from maximizing the log-likelihood surface. The computing times in seconds to estimate the maximum likelihood estimates by directly maximizing the log-likelihood surface for these distributions are shown in the last row of the table. The computing time of the EM algorithm for the DPLN GLM was 1,322.81 seconds using the stopping criterion that the relative change of the log-likelihood function be less than  $1 \times 10^{-4}$ .

**Table 4.** Results of fitting the LN, GB2 and DPLN distributions to automobile bodily injury claims data.

Estimate (S.E.)	Distribution		
	LN	GB2	DPLN
$\nu$	0.620 (0.044)	1.204 (0.052)	1.200 (0.040)
$\tau$	1.445 (0.031)	0.022 (0.186)	0.047 (0.150)
$p$ or $\lambda_1$		0.017 (0.140)	1.324 (0.068)
$q$ or $\lambda_2$		0.030 (0.247)	0.749 (0.025)
NLL	2626.74	2573.47	2573.47
AIC	5257.48	5154.94	5154.94
BIC	5267.47	5174.92	5174.92
CT	0.1716	3.4476	3.0888

#### 4.2.2. Comparison of Estimation from Simulations

In this section we compare the methods of estimating parameters by conducting the following simulation experiment. For the lognormal and DPLN distributions, we simulate values based on the corresponding parameter estimates given in Table 4, and then, using as appropriate either standard formulae or the EM algorithm, compute parameter estimates from the 1000 simulated data sets of size  $N = 100, 200, 300, 400, 500, 1000$ . The results are shown in Table 5, where it is evident that increasing the sample size increases the accuracy of the parameter estimate. Of course, the true parameter values are given in Table 4, and these are the limits of the estimates as the sample size  $N$  increases. Importantly, the standard errors of the parameter estimates in Table 5 are noticeably smaller for the DPLN distribution, highlighting the consistency of the parameter estimation for the DPLN model. However, in attempting to simulate values from the GB2 distribution, calculation of the inverse CDF via the expression in (30) is highly unstable for simulated values of the  $Beta(p, q)$  random variable  $Z$  which are close to unity.

**Table 5.** Results of the simulation experiment involving 1000 simulations of data sets of size  $N$ , with standard errors shown in brackets.

Sample Size $N$	Distribution		
	LN	DPLN	
100	$\hat{\nu} = 0.6220$ (0.1442) $\hat{\tau} = 1.4336$ (0.1054)	$\hat{\nu} = 1.2001$ (0.0028) $\hat{\tau} = 0.0470$ (0.0004)	$\hat{\lambda}_1 = 1.3611$ (0.1934) $\hat{\lambda}_2 = 0.7663$ (0.0861)
200	$\hat{\nu} = 0.6214$ (0.0995) $\hat{\tau} = 1.4397$ (0.0714)	$\hat{\nu} = 1.2001$ (0.0020) $\hat{\tau} = 0.0470$ (0.0002)	$\hat{\lambda}_1 = 1.3472$ (0.1379) $\hat{\lambda}_2 = 0.7581$ (0.0606)
300	$\hat{\nu} = 0.6189$ (0.0820) $\hat{\tau} = 1.4430$ (0.0580)	$\hat{\nu} = 1.2000$ (0.0016) $\hat{\tau} = 0.0470$ (0.0002)	$\hat{\lambda}_1 = 1.3348$ (0.1076) $\hat{\lambda}_2 = 0.7537$ (0.0483)
400	$\hat{\nu} = 0.6201$ (0.0737) $\hat{\tau} = 1.4377$ (0.0507)	$\hat{\nu} = 1.2000$ (0.0015) $\hat{\tau} = 0.0470$ (0.0002)	$\hat{\lambda}_1 = 1.3332$ (0.0921) $\hat{\lambda}_2 = 0.7509$ (0.0419)
500	$\hat{\nu} = 0.6223$ (0.0627) $\hat{\tau} = 1.4430$ (0.0440)	$\hat{\nu} = 1.2000$ (0.0013) $\hat{\tau} = 0.0470$ (0.0002)	$\hat{\lambda}_1 = 1.3335$ (0.0848) $\hat{\lambda}_2 = 0.7520$ (0.0383)
1000	$\hat{\nu} = 0.6210$ (0.0449) $\hat{\tau} = 1.4434$ (0.0329)	$\hat{\nu} = 1.2000$ (0.0009) $\hat{\tau} = 0.0470$ (0.0001)	$\hat{\lambda}_1 = 1.3273$ (0.0581) $\hat{\lambda}_2 = 0.7486$ (0.0273)



#### 4.2.3. Including Explanatory Variables

Table 6, displays the same results for the automobile injury claims dataset. For the  $i$ -th policyholder, the number of total amount  $y_i$  follows (10) whose mean depends on the above set of covariates through the identity link function. The exponential of *INTERCEPT* coefficient 1.023 is proportional to the predicted loss amount when the values of the other explanatory variables are equal to 0. In view of its low  $p$ -value, this estimate is statistically significant at the usual significance levels, 5% and 1%. On the other hand, the indicator *ATTORNEY* is statistically significant at the usual nominal levels, whereas the gender and marital status of the claimant, except that the explanatory variable *SINGLE*, are not significant at the 5% significance level. Similarly, the fact that the vehicle was uninsured is not relevant in the investigation. Both claimant's age and usage of seatbelt/child restraint are highly significant. Three more parameters affect the calculation of the predicted mean: the parameter  $\tau$ , which is also highly significant, and shape parameters  $\lambda_1$  and  $\lambda_2$ . All these three parameters are highly statistically significant at the usual nominal levels, 5% and 1%. For the sake of comparison the results for the LN and GB2 generalized linear model are displayed in Table 6. As it can be observed the fit provided by the GB2 generalized linear model is only marginally better than the DPLN generalized linear model. For the DPLN generalized linear model, parameters were estimated by the method of maximum likelihood by using log-transformed data and the NSL distribution. The maximum of the log-likelihood function was  $-1753.07$  and it was achieved after considering different initial values of likelihood surface by using the `FindMaximum` function of Mathematica software package v.11.0. Similar estimates were obtained by means of the EM algorithm described in Section 3.2. In this case the same value was obtained for the maximum of the log-likelihood function of the NSL GLM. The standard errors of the parameter estimates for the DPLN GLM have been approximated by inverting the Hessian matrix and also from the last iteration of the EM algorithm. Similar values were obtained. The computing times in seconds to estimate the maximum likelihood estimates by directly maximizing the log-likelihood surface for these generalized linear models are shown in the last row of the table. The DPLN GLM shows a better performance than the GB2 counterpart. The computing time of the EM algorithm for the DPLN GLM was 142.57 seconds using the stopping criterion that the relative change of the log-likelihood function be less than  $1 \times 10^{-4}$ .

**Table 6.** Parameter estimates, standard errors (S.E.) and  $p$ -values of the  $t$ -test for automobile bodily injury claims dataset under LN, GB2 and DPLN generalized linear models.

	Generalized Linear Model		
Estimate (S.E.)	LN	GB2	DPLN
<i>INTERCEPT</i>	0.764 (0.382)	1.083 (0.383)	1.023 (0.376)
$p$ -value	0.0458	0.0048	0.0067
<i>ATTORNEY</i>	1.368 (0.075)	1.215 (0.079)	1.213 (0.075)
$p$ -value	<0.0001	<0.0001	<0.0001
<i>CLMSEX</i>	-0.103 (0.076)	-0.135 (0.070)	-0.135 (0.069)
$p$ -value	0.1757	0.0524	0.0516
<i>MARRIED</i>	-0.221 (0.235)	-0.350 (0.233)	-0.352 (0.234)
$p$ -value	0.3464	0.1340	0.1320
<i>SINGLE</i>	-0.378 (0.241)	-0.494 (0.237)	-0.498 (0.237)
$p$ -value	0.1171	0.0374	0.0360
<i>WIDOWED</i>	-0.887 (0.430)	-0.748 (0.417)	-0.744 (0.419)
$p$ -value	0.0393	0.0730	0.0763
<i>CLMINSUR</i>	-0.009 (0.127)	-0.043 (0.116)	-0.041 (0.115)
$p$ -value	0.9448	0.7091	0.7218
<i>SEATBELT</i>	-0.996 (0.278)	-0.785 (0.272)	-0.768 (0.272)
$p$ -value	0.0015	0.0040	0.0048

Table 6. Cont.

Estimate (S.E.)	Generalized Linear Model		
	LN	GB2	DPLN
CLMAGE	0.014 (0.003)	0.013 (0.003)	0.013 (0.003)
<i>p</i> -value	0.0010	<0.0001	<0.0001
$\tau$	1.230 (0.026)	0.448 (0.129)	0.538 (0.110)
<i>p</i> -value	<0.0001	0.0006	<0.0001
<i>p</i> or $\lambda_1$		0.513 (0.185)	1.458 (0.139)
<i>p</i> -value		0.0055	<0.0001
<i>q</i> or $\lambda_2$		0.670 (0.252)	1.112 (0.085)
<i>p</i> -value		0.0079	<0.0001
NLL	2450.54	2429.59	2430.02
AIC	4921.09	4883.18	4884.05
BIC	4971.04	4943.12	4943.98
CT	0.4524	6.2556	3.2488

#### 4.2.4. Model Validation

As done for our first example, we analyze model validation from a practical perspective. The test statistic is  $T = 2(\ell_{LN} - \ell_{DPLN})$  where  $\ell_{LN}$  and  $\ell_{DPLN}$  represent the maximum of the log-likelihood function for the LN and DPLN generalized linear models respectively and  $T$  asymptotically follows a chi-square distribution with two degree of freedom. For the automobile bodily injury claims data set, it is verified that  $T = 2(-2450.54 + 2430.02) = 20.52$ . Then, at the usual significance levels (i.e., *p*-value is less than 0.0001), the null hypothesis is clearly rejected and consequently, the smaller regression (LN) is rejected in favour of the model based on the DPLN distribution.

Also, as done for our first example, Vuong's test statistic is  $T = 1.00$  for our second data set and we fail to reject  $H_0$ , and therefore differences between these two models do not exist.

#### 4.3. Log-Residuals for Assessing Goodness-of-Fit

In the following we consider the log-residuals for assessing the goodness-of-fit of the proposed models for the two datasets considered. As the population moments of order higher than two cannot be derived neither for the DPLN (i.e.,  $\lambda_1 < 2$ ) nor the GB2 (i.e., the condition  $p < 2\tau < q$  is not satisfied in none of the datasets) distribution for the automobile bodily injury claims dataset, we have not examined the Pearson's type residual. In Figure 2, one can see the QQ-plot of the log-residuals for LN, GB2 and DPLN generalized linear models for the automobile insurance claims (left hand side) automobile bodily injure claims data set (right hand side). The alignment along the 45-degree line is better in both the DPLN and GB2 generalized linear models in the central part and both tails of the distribution of the residuals as compared to the LN generalized linear model for the two datasets analyzed.

#### 4.4. Out-of-Sample Validation of Models

We demonstrate the abilities of the models to predict portfolio losses out-of-sample with probability-probability plots shown in Figure 3. The data set  $\{v_1, v_2, \dots, v_n\}$  is partitioned into two halves by sorting the claim sizes ascendingly, that is, we write the data set as

$$v_{i_1} < v_{i_2} < v_{i_3} < \dots < v_{i_n}, \quad (33)$$

where  $i_1, \dots, i_n \in \{1, 2, \dots, n\}$ , and then two data sets are formed,  $A = \{v_{i_1}, v_{i_3}, v_{i_5}, \dots\}$  and  $B = \{v_{i_2}, v_{i_4}, v_{i_6}, \dots\}$ , alternating the data set to which each claim data item in the ordered data set is allocated. In this way the second data set is a good representation of the first data set in respect of

claim distribution, but not necessarily in respect of the corresponding covariates, i.e.,  $\{x^{(i_2)}, x^{(i_4)}, \dots\}$  may not be representative of  $\{x^{(i_1)}, x^{(i_3)}, \dots\}$ . The data set  $A$  is used for fitting the models, whereas the data set  $B$  is used for graphing the probability-probability plots. In Figures 4 and 5 we focus on the lower and upper tails of the distributions respectively, where it is evident that the DPLN and GB2 models provide the best fit and are almost indistinguishable.

In Figure 6 the net losses under the various models are shown, where, as before, we have used half of the data for fitting the models and the other half for computing the net losses and maximum probable losses based on the 99.5-th percentile. It is evident that the DPLN and GB2 models give a higher computed maximum probable loss than for the LN distribution, and thus illustrates the ability of these models to provide adequate solvency levels when extreme claims are experienced by the insurer’s portfolio.

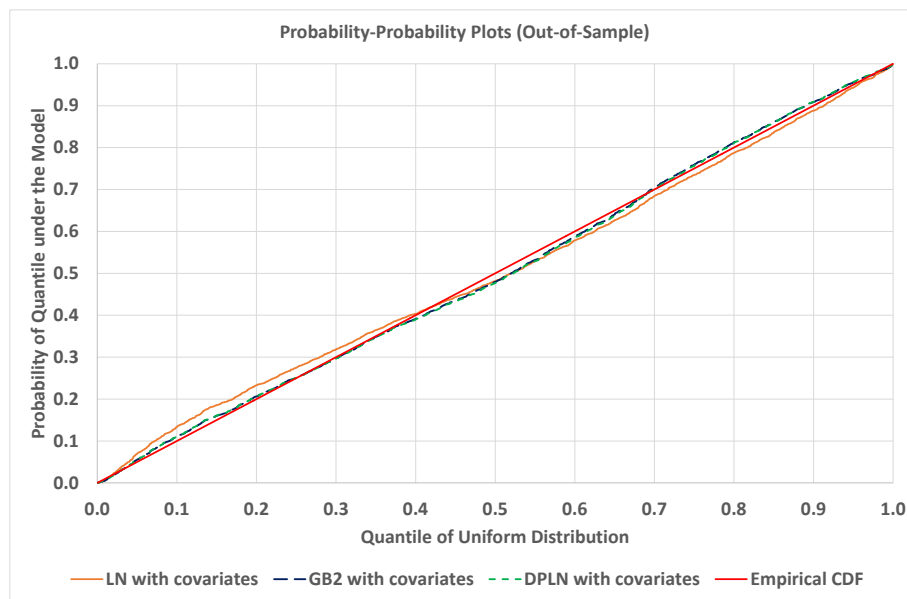


Figure 3. Probability-probability plot (out-of-sample).

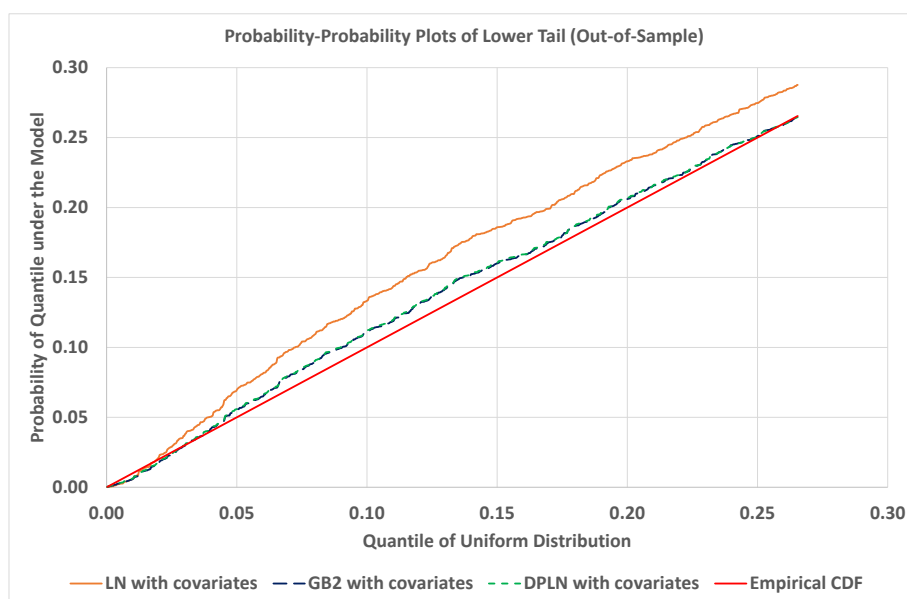


Figure 4. Probability-probability plot of lower tail (out-of-sample).

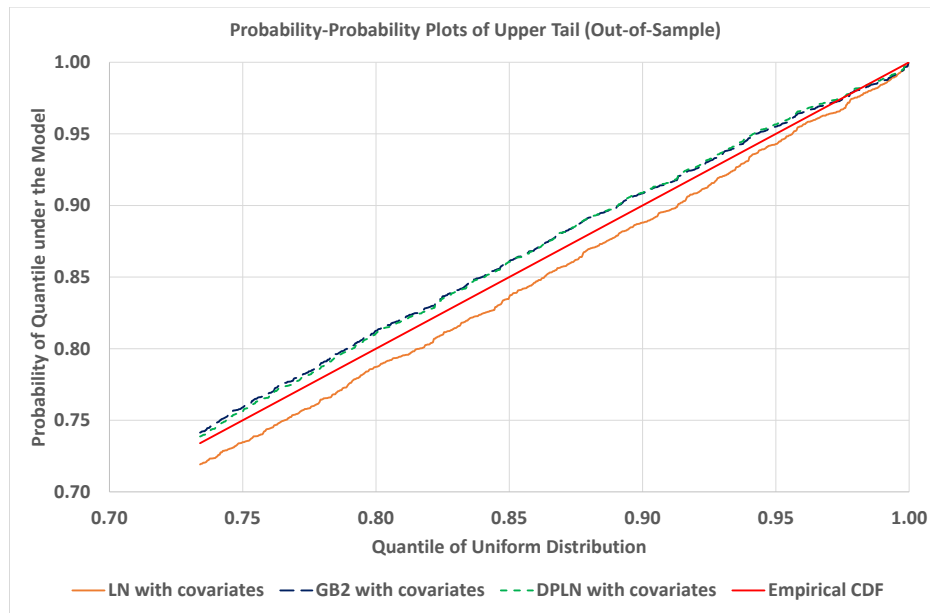


Figure 5. Probability-probability plots of upper tail (out-of-sample).

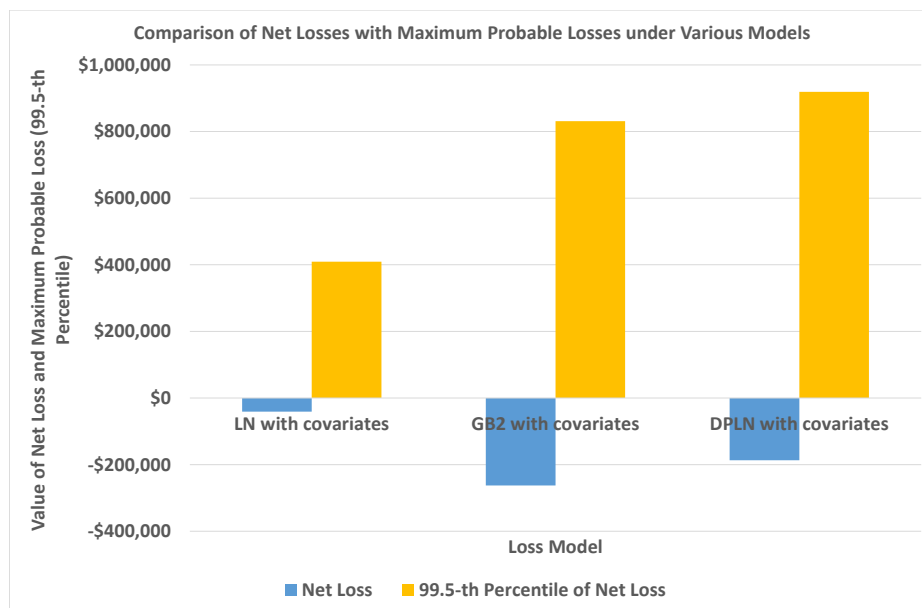


Figure 6. Net losses on a portfolio of automobile bodily injury claims for various models (out-of-sample).

### 5. Conclusions

In this paper, the DPLN generalized linear model was developed and fitted to two data sets, these being private passenger automobile insurance claims data and automobile bodily injury claims data. Several covariates pertaining to various attributes of insurance claimants were combined in the linear predictor of the location parameter  $\nu$ , and were chosen because of their anticipated effect on claim size. This model exhibits Paretian behaviour in both tails and it is shown to provide fits to the two data sets which are comparable to those of the GB2 distribution.

The parameters of the DPLN generalized linear model were estimated via the EM algorithm and independently confirmed by maximizing the log-likelihood surface of the closely related Normal Laplace generalized linear model. The performance of the DPLN model has been compared with the lognormal distribution, a limiting case of the DPLN distribution, and the GB2 generalized linear model according to different model selection criteria. In view of the results obtained, we have found that the proposed DPLN generalized linear model is a valid alternative to other parametric heavy-tailed generalized linear models such as the GB2 GLM.

Potential practical applications of the DPLN GLM, beyond what is demonstrated in this article, include predicting mortality rates for lives where the covariates of the GLM are age, sex, occupation, etc. and predicting hazard rates in reduced-form credit risk models. These will be considered in further work.

**Acknowledgments:** The authors acknowledge the financial support from the Faculty of Business and Economics, University of Melbourne via a grant awarded for 2017. Also, the authors are grateful for the helpful suggestions of the reviewers.

**Author Contributions:** These authors contributed equally to this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

### Appendix A.1. Proof of Theorem 1

Looking at (18) we see that the expression separates into two parts, as mentioned in [Reed and Jorgensen \(2004\)](#), the first being dependent upon  $\beta'$  and  $\tau'$  and the second being dependent upon  $\lambda'_1$  and  $\lambda'_2$ . The first part

$$n \log \left( \frac{1}{\sqrt{2\pi(\tau')^2}} \right) - \frac{1}{2(\tau')^2} \sum_{i=1}^n z_i^{(2)} + \frac{1}{(\tau')^2} \sum_{i=1}^n z_i v_i - \frac{1}{2(\tau')^2} \sum_{i=1}^n (v_i')^2 \tag{A1}$$

can be rewritten using matrix notation as

$$n \log \left( \frac{1}{\sqrt{2\pi(\tau')^2}} \right) - \frac{1}{2(\tau')^2} (\mathbf{1}^\top \mathbf{Z}^{(2)} - 2\mathbf{Z}^\top \mathbf{X}^\top \beta' + \beta'^\top \mathbf{X} \mathbf{X}^\top \beta'). \tag{A2}$$

Viewed as a quadratic form in  $\beta'$ , the optimum value of  $\beta'$  is

$$\beta' = (\mathbf{X} \mathbf{X}^\top)^{-1} \mathbf{X} \mathbf{Z} \tag{A3}$$

and the first part becomes

$$n \log \left( \frac{1}{\sqrt{2\pi(\tau')^2}} \right) - \frac{1}{2(\tau')^2} (\mathbf{1}^\top \mathbf{Z}^{(2)} - \mathbf{Z}^\top \mathbf{X}^\top (\mathbf{X} \mathbf{X}^\top)^{-1} \mathbf{X} \mathbf{Z}), \tag{A4}$$

which is to be maximized with respect to  $\tau'$ . Differentiating this with respect to  $\tau'$  and equating to zero gives the update for  $\tau'$ .

The second part

$$n \log \frac{\lambda'_1 \lambda'_2}{\lambda'_1 + \lambda'_2} + \lambda'_2 \sum_{i=1}^n w_i^- - \lambda'_1 \sum_{i=1}^n w_i^+ \tag{A5}$$

can be rewritten as

$$n \left( \log \frac{\lambda'_1 \lambda'_2}{\lambda'_1 + \lambda'_2} - \lambda'_2 Q' - \lambda'_1 P' \right), \tag{A6}$$

where

$$P' = \frac{1}{n} \sum_{i=1}^n w_i^+, \quad Q' = -\frac{1}{n} \sum_{i=1}^n w_i^-. \tag{A7}$$

The optimum values of  $\lambda'_1$  and  $\lambda'_2$  are solved by equating the first order partial derivatives to zero with closed-form solutions

$$\lambda'_1 = \frac{1}{P' + \sqrt{P'Q'}}, \quad \lambda'_2 = \frac{1}{Q' + \sqrt{P'Q'}}.$$

Appendix A.2. Proof of Theorem 2

The first expectation is computed as follows

$$\begin{aligned} z_i &= \int_{-\infty}^{\infty} z g_i(z) dz \\ &= \frac{1}{f_Y(y_i; \theta)} \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(z - \nu)^2\right\} \\ &\quad \times \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} \begin{cases} \exp(\lambda_2(y_i - z)) & , y_i - z < 0 \\ \exp(-\lambda_1(y_i - z)) & , y_i - z \geq 0 \end{cases} dz \\ &= \frac{1}{f_Y(y_i; \theta)} \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} \\ &\quad \times \left\{ \int_{y_i}^{\infty} z \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(z - \nu)^2\right\} \exp(\lambda_2(y_i - z)) dz \right. \\ &\quad \left. + \int_{-\infty}^{y_i} z \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(z - \nu)^2\right\} \exp(-\lambda_1(y_i - z)) dz \right\} \\ &= \frac{1}{f_Y(y_i; \theta)} \frac{\lambda_1\lambda_2}{\lambda_1 + \lambda_2} \\ &\quad \times \left\{ \int_{y_i}^{\infty} z \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(z^2 + \nu^2 - 2\nu z + 2\tau^2\lambda_2 z - 2\tau^2\lambda_2 y_i)\right\} dz \right. \\ &\quad \left. + \int_{-\infty}^{y_i} z \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(z^2 + \nu^2 - 2\nu z - 2\tau^2\lambda_1 z + 2\tau^2\lambda_1 y_i)\right\} dz \right\} \end{aligned} \tag{A8}$$

The first of these integrals simplifies as

$$\begin{aligned} &\exp\left\{-\frac{1}{2\tau^2}(\nu^2 - (-\nu + \tau^2\lambda_2)^2 - 2\tau^2\lambda_2 y_i)\right\} \\ &\quad \times \int_{y_i}^{\infty} z \frac{1}{\sqrt{2\pi\tau^2}} \exp\left\{-\frac{1}{2\tau^2}(z + (-\nu + \tau^2\lambda_2))^2\right\} dz \\ &= \phi((y_i - \nu)/\tau) / \phi(q_i) \times \int_{y_i}^{\infty} z \frac{1}{\tau} \phi((z - (\nu - \tau^2\lambda_2))/\tau) dz \\ &= \phi((y_i - \nu)/\tau) / \phi(q_i) \times \left\{ (\nu - \tau^2\lambda_2) \Phi^c((y_i - (\nu - \tau^2\lambda_2))/\tau) \right. \\ &\quad \left. + \tau \phi((y_i - (\nu - \tau^2\lambda_2))/\tau) \right\} \\ &= \phi((y_i - \nu)/\tau) \times \left\{ (\nu - \tau^2\lambda_2) R(q_i) + \tau \right\}. \end{aligned} \tag{A9}$$

The second of these integrals simplifies as

$$\begin{aligned}
 & \exp \left\{ -\frac{1}{2\tau^2}(v^2 - (-v - \tau^2\lambda_1)^2 + 2\tau^2\lambda_1 y_i) \right\} \\
 & \times \int_{-\infty}^{y_i} z \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{1}{2\tau^2}(z^2 + (-vz - 2\tau^2\lambda_1)^2) \right\} dz \\
 & = \phi((y_i - v)/\tau) / \phi(p_i) \times \left\{ \int_{-\infty}^{y_i} z \frac{1}{\tau} \phi((z - (vz + 2\tau^2\lambda_1))/\tau) dz \right\} \\
 & = \phi((y_i - v)/\tau) / \phi(p_i) \times \left\{ (vz + 2\tau^2\lambda_1) \Phi((y_i - (vz + 2\tau^2\lambda_1))/\tau) \right. \\
 & \quad \left. - \tau \phi((y_i - (vz + 2\tau^2\lambda_1))/\tau) \right\} \\
 & = \phi((y_i - v)/\tau) / \phi(p_i) \times \left\{ (vz + 2\tau^2\lambda_1) R(p_i) - \tau \right\}.
 \end{aligned} \tag{A10}$$

Combining both integrals in the simplified formula for  $z_i$  gives

$$z_i = v + \tau^2 \frac{\lambda_1 R(p_i) - \lambda_2 R(q_i)}{R(p_i) + R(q_i)}. \tag{A11}$$

The second expectation is computed as follows

$$\begin{aligned}
 z_i^{(2)} & = \int_{-\infty}^{\infty} z^2 g_i(z) dz \\
 & = \frac{1}{f_Y(y_i; \theta)} \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{1}{2\tau^2}(z - v)^2 \right\} \\
 & \quad \times \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \begin{cases} \exp(\lambda_2(y_i - z)) & , y_i - z < 0 \\ \exp(-\lambda_1(y_i - z)) & , y_i - z \geq 0 \end{cases} dz \\
 & = v^2 + \tau^2 - \tau^2 \frac{p_i + q_i}{R(p_i) + R(q_i)} \\
 & \quad + \tau^2 \frac{(2v\lambda_1 + \lambda_1^2\tau^2)R(p_i) + (\lambda_2^2\tau^2 - 2v\lambda_2)R(q_i)}{R(p_i) + R(q_i)}.
 \end{aligned} \tag{A12}$$

The third expectation is computed as follows

$$\begin{aligned}
 w_i^+ & = \int_{-\infty}^{\infty} (y_i - z)^+ g_i(z) dz \\
 & = \frac{1}{f_Y(y_i; \theta)} \int_{-\infty}^{\infty} (y_i - z)^+ \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{1}{2\tau^2}(z - v)^2 \right\} \\
 & \quad \times \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \begin{cases} \exp(\lambda_2(y_i - z)) & , y_i - z < 0 \\ \exp(-\lambda_1(y_i - z)) & , y_i - z \geq 0 \end{cases} dz \\
 & = \tau \frac{-p_i R(p_i) + 1}{R(p_i) + R(q_i)}.
 \end{aligned} \tag{A13}$$

The fourth expectation is computed as follows

$$\begin{aligned}
 w_i^- & = \int_{-\infty}^{\infty} (y_i - z)^- g_i(z) dz \\
 & = \frac{1}{f_Y(y_i; \theta)} \int_{-\infty}^{\infty} (y_i - z)^- \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{1}{2\tau^2}(z - v)^2 \right\} \\
 & \quad \times \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \begin{cases} \exp(\lambda_2(y_i - z)) & , y_i - z < 0 \\ \exp(-\lambda_1(y_i - z)) & , y_i - z \geq 0 \end{cases} dz \\
 & = \tau \frac{q_i R(q_i) - 1}{R(p_i) + R(q_i)}.
 \end{aligned} \tag{A14}$$



Appendix A.3. Score Equations

The score equations to be solved for calculation of the maximum likelihood estimates are given by

$$\frac{\partial \ell(\mathbf{y}; \beta, \tau^2, \lambda_1, \lambda_2)}{\partial \lambda_1} = \frac{n}{\lambda_1} - \frac{n}{\lambda_1 + \lambda_2} + \sum_{i=1}^n \frac{2\mathcal{A}_1(y_i, v_i) \left( \sqrt{\pi}(\tau^2 + 2v_i - 2y_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) - \sqrt{2}\mathcal{B}_1(y_i, v_i) \right)}{\sqrt{\pi} \left( \mathcal{A}_1(y_i, v_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) + \mathcal{C}_1(y_i, v_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right) \right)} = 0,$$

with  $\mathcal{A}_1(y_i, v_i) = \exp\left\{\lambda_2 y_i + \lambda_1 v_i + \frac{\lambda_1 \tau^2}{2}\right\}$ ,  $\mathcal{B}_1(y_i, v_i) = \exp\left\{-\frac{(\lambda_1 \tau^2 + v_i - y_i)^2}{2\tau^2}\right\}$  and  $\mathcal{C}_1(y_i, v_i) = \exp\left\{\lambda_1 y_i + \lambda_2 v_i + \frac{\lambda_2 \tau^2}{2}\right\}$ .

$$\frac{\partial \ell(\mathbf{y}; \beta, \tau^2, \lambda_1, \lambda_2)}{\partial \lambda_2} = \frac{n}{\lambda_2} - \frac{n}{\lambda_1 + \lambda_2} + \sum_{i=1}^n \frac{2\mathcal{C}_2(y_i, v_i) \left( (\tau^2 + 2v_i - 2y_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right) - \sqrt{\frac{2}{\pi}}\mathcal{B}_2(y_i, v_i) \right)}{\mathcal{A}_2(y_i, v_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) + \mathcal{C}_2(y_i, v_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right)} = 0,$$

with  $\mathcal{A}_2(y_i, v_i) = \exp\left\{-\lambda_1 y_i + \lambda_1 v_i + \frac{\lambda_1 \tau^2}{2}\right\}$ ,  $\mathcal{B}_2(y_i, v_i) = \exp\left\{-\frac{(\lambda_2 \tau^2 - v_i + y_i)^2}{2\tau^2}\right\}$  and  $\mathcal{C}_2(y_i, v_i) = \exp\left\{-\lambda_2 y_i + \lambda_2 v_i + \frac{\lambda_2 \tau^2}{2}\right\}$ .

$$\frac{\partial \ell(\mathbf{y}; \beta, \tau^2, \lambda_1, \lambda_2)}{\partial \tau} = \sum_{i=1}^n \frac{4\mathcal{A}_3(y_i, v_i) (\mathcal{B}_3(y_i - v_i - \lambda_2 \tau^2) - y_i + v_i - \lambda_1 \tau^2)}{\sqrt{2\pi}\tau^2 (\mathcal{A}_2(y_i, v_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) + \mathcal{C}_2(y_i, v_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right))} + \sum_{i=1}^n \frac{\tau(\lambda_1 \mathcal{A}_2(y_i, v_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) + \lambda_2 \mathcal{C}_2(y_i, v_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right))}{2 \left( \mathcal{A}_2(y_i, v_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) + \mathcal{C}_2(y_i, v_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right) \right)} = 0,$$

with  $\mathcal{A}_3(y_i, v_i) = \exp\left\{-\frac{(y_i - v_i) + (\lambda_1 - 1)\lambda_1 \tau^4}{2\tau^2}\right\}$  and  $\mathcal{B}_3(y_i, v_i) = \exp\left\{-2\lambda_2(y_i - v_i) + \frac{1}{2}(\lambda_1 - \lambda_2)(\lambda_1 + \lambda_2 - 1)\tau^2\right\}$ .

$$\frac{\partial \ell(\mathbf{y}; \beta, \tau^2, \lambda_1, \lambda_2)}{\partial \beta_j} = \sum_{i=1}^n \frac{2\sqrt{2}x_j (\mathcal{A}_4(y_i, v_i) - \mathcal{B}_4(y_i, v_i))}{\sqrt{\pi}\tau (\mathcal{A}_2(y_i, v_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) + \mathcal{C}_2(y_i, v_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right))} + \sum_{i=1}^n \frac{x_j \left( \lambda_1 \mathcal{A}_2(y_i, v_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) + \lambda_2 \mathcal{C}_2(y_i, v_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right) \right)}{\mathcal{A}_2(y_i, v_i)\Phi\left(\frac{y_i - \lambda_1\tau^2 - v_i}{\tau}\right) + \mathcal{C}_2(y_i, v_i)\Phi^c\left(\frac{y_i + \lambda_2\tau^2 - v_i}{\tau}\right)} = 0,$$

with  $j = 1, \dots, d$ ,  $\mathcal{A}_4(y_i, v_i) = \exp\left\{-\lambda_2(y_i - v_i) + \frac{\lambda_2 \tau^2}{2} - \frac{(y_i - v_i + \lambda_2 \tau^2)^2}{2\tau^2}\right\}$  and  $\mathcal{B}_4(y_i, v_i) = \exp\left\{-\frac{1}{2}\lambda_1 \tau^2(1 - \lambda_1) - \frac{(y_i - v_i)^2}{2\tau^2}\right\}$ .

References

Colombi, Roberto. 1990. A new model of income distribution: The Pareto lognormal distribution. In *Income and Wealth Distribution, Inequality and Poverty*; Edited by C. Dagum and M. Zenga. Berlin: Springer, pp. 18–32.

- Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* 39: 1–38.
- Frees, Edward W., Richard A. Derrig, and Glenn Meyers. 2014a. *Predictive Modeling Applications in Actuarial Science, Volume 1*. New York: Cambridge University Press.
- Frees, Edward W., Richard A. Derrig, and Glenn Meyers. 2014b. *Predictive Modeling Applications in Actuarial Science, Volume 2*. New York: Cambridge University Press.
- Frees, Edward W., and Emiliano A. Valdez. 2008. Hierarchical Insurance Claims Modeling. *Journal of the American Statistical Association* 103: 1457–69.
- Giesen, Kristian, Arndt Zimmermann, and Jens Suedekum. 2010. The size distribution across all cities—Double Pareto lognormal strikes. *Journal of Urban Economics* 68: 129–37.
- Hajargasht, Gholamreza, and William E. Griffiths. 2013. Pareto-lognormal distributions: Inequality, poverty, and estimation from grouped income data. *Economic Modelling* 33: 593–604.
- Hürlimann, Werner. 2014. Pareto type distributions and excess-of-loss reinsurance. *International Journal of Recent Research and Applied Studies* 18: 1.
- Kleiber, Christian, and Samuel Kotz. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken: Wiley.
- Kočović, Jelena, Vesna Čojbašić Rajić, and Milan Jovanović. 2015. Estimating a tail of the mixture of log-normal and inverse gaussian distribution. *Scandinavian Actuarial Journal* 2015: 49–58.
- Lehmann, Erich Leo, and George Casella. 1998. *Theory of Point Estimation*, 2nd ed. New York: Springer.
- Louis, Thomas A. 1982. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B* 44: 226–33.
- McDonald, James B. 1990. Regression model for positive random variables. *Journal of Econometrics* 43: 227–51.
- Ramirez-Cobo, Pepa, R. E. Lillo, S. Wilson, and M. P. Wiper. 2010. Bayesian inference for double pareto lognormal queues. *The Annals of Applied Statistics* 4: 1533–57.
- Reed, William J. 2003. The Pareto law of incomes - an explanation and an extension. *Physica A* 319: 469–86.
- Reed, William J., and Murray Jorgensen. 2004. The Double Pareto-Lognormal Distribution—A new parametric model for size distributions. *Communications in Statistics - Theory and Methods* 33: 1733–53.
- Vuong, Quang H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society* 57: 307–33.
- Wills, M., E. Valdez, and E. Frees. 2006. GB2 regression with insurance claim severities. Paper presented at the UNSW Actuarial Research Symposium, Sydney, New South Wales, Australia, November 9.
- Yamaguchi, Kazuo. 1992. Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of 'permanent employment' in Japan. *Journal of the American Statistical Association* 87: 284–92.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).