ON MAXIMUM LIKELIHOOD AND PSEUDO-MAXIMUM LIKELIHOOD ESTIMATION IN COMPOUND INSURANCE MODELS WITH DEDUCTIBLES

BY

JOSTEIN PAULSEN AND KNUT STUBØ

ABSTRACT

Non-life insurance payouts consist of two factors: claimsizes and claim frequency. When calculating e.g. next years premium, it is vital to correctly model these factors and to estimate the unknown parameters. A standard way is to separately estimate in the claimsize and the claim frequency models.

Often there is a deductible with each single claim, and this deductible can be quite large, particularly in inhomogeneous cases such as industrial fire insurance or marine insurance. Not taking the deductibles into account can lead to serious bias in the estimates and consequent implications when applying the model.

When the deductibles are nonidentical, in a full maximum likelihood estimation all unknown parameters have to be estimated simultaneously. An alternative is to use pseudo-maximum likelihood, i.e. first estimate the claimsize model, taking the deductibles into account, and then use the estimated probability that a claim exceeds the deductible as an offset in the claim frequency estimation. This latter method is less efficient, but due to complexity or time considerations, it may be the preferred option.

In this paper we will provide rather general formulas for the relative efficiency of the pseudo maximum likelihood estimators in the i.i.d. case. Two special cases will be studied in detail, and we conclude the paper by comparing the methods on some marine insurance data.

KEYWORDS

Compound model; Compound Poisson model; deductibles; maximum likelihood; pseudo-maximum likelihood; asymptotic normality; asymptotic relative efficiency; marine insurance.

1. INTRODUCTION AND MODEL DESCRIPTION

As most practioners in non-life insurance know, making good and credible models for the data at hand is often a difficult task, and if the models become

Astin Bulletin 41(1), 1-28. doi: 10.2143/AST.41.1.2084384

too complicated, statistical estimation can be very hard. Let us list three important issues to that effect. Note that we use the word claim even if it is not reported to the company.

1. Deductibles: Only claims above a specified deductible are reported to the company. This issue becomes even more complicated in the presence of a bonus system, since in order to avoid a loss of bonus, the insured may pay in full herself some claims that exceed the deductible. Another complicating factor is the fact that deductibles and claims (size or frequency) may not be independent.
2. Covariates and random effects: Assuming the data are i.i.d. is often too simple. In car insurance, covariates can be the size of the car, the size of the engine or an indicator of whether the driver is a male or a female. Random effects may be the driver himself, the district the car is registered in or the brand of the car. Covariates and random effects may be present in both the claimsize and the claim frequency parts of the model.
3. Registration errors: Unfortunately data are not always correct due to registration errors, or there may be missing values.

Trying to incorporate some or all of these issues in a model may lead to a very complicated likelihood, and maximizing this to find the MLE (maximum likelihood estimators) becomes hard or even impossible. That makes it necessary to either look for simpler models, or for simpler methods of estimation. One possibility in the second case is to use pseudo-MLE. This is a rather general term, but in this context it will mean that first the parameters in the claimsize model are estimated, and then the estimated probability that a claim exceeds the deductible is used as an offset in the claim frequency estimation.

Another possibility that is frequently used in practice is to only model that part of the claim that exceeds the highest deductible in the portfolio. Under the standard assumption that claimsizes are mutually independent and also independent of the claim numbers, using only the claims that exceed the highest deductible in the estimation, the pseudo-MLE and the full MLE are equal for that part of the claims. However, it does not give any information of the claims below that deductible, but here simpler methods can be used. This may be a good solution when the portfolio is homogeneous, like in car insurance. But when the portfolio consists of heterogeneous risks with very different deductibles, this method is no longer feasible.

A third possibility, which unfortunately is in common use, is to ignore the deductibles altogether and treat the claims as if the deductibles are zero. This can sometimes lead to serious bias and applications of the model may cause some unpleasant financial distress to the company.

To give a more formal description of the problem we are going to study here, consider n risks where risk i has claims X_{ij}^*, $j = 1, \ldots, N_i^*$. Assume that each risk has a deductible D_i, $i = 1, \ldots, n$, and only those X_{ij}^* larger than the corresponding D_i are reported to the insurance company. We will assume throughout that the X_{ij}^*, N_i^* and D_i are all independent random variables. Although we are primarily interested in the distribution of the X_{ij}^*, letting the D_i be random variables is natural in an estimation context. But see Remark 2.3 below for an alternative approach.

The data available to the company are N_i, $i = 1, \ldots, n$, X_{ij}, $j = 1, \ldots, N_i$ and D_i, $i = 1, \ldots, n$. The $\{X_{ij}\}$ consist of those $\{X_{ij}^*\}$ that are larger than their respective D_i's and so $N_i \le N_i^*$. We will assume that for fixed i, the X_{ij}^* are i.i.d. with distribution function $F_i(x; \boldsymbol{\theta}) = P(X_{ij}^* \le x)$, where $\boldsymbol{\theta}$ is an unknown s-dimensional parameter vector. Then

$$P(X_{ij} \le x \mid D_i) = \begin{cases} 0, & x \le D_i, \\ \dfrac{F_i(x; \boldsymbol{\theta}) - F_i(D_i; \boldsymbol{\theta})}{\bar{F}_i(D_i; \boldsymbol{\theta})}, & x > D_i, \end{cases}$$

where $\bar{F}_i(x; \boldsymbol{\theta}) = 1 - F_i(x; \boldsymbol{\theta})$. Assuming that the claim number distribution depends on a t-dimensional parameter $\boldsymbol{\lambda}$, write

$$p_i^*(m; \boldsymbol{\lambda}) = P(N_i^* = m), \quad m = 0, 1, \ldots.$$

Then conditional on D_i, we can write

$$P(N_i = m \mid D_i) = p_i(m; (\boldsymbol{\theta}, \boldsymbol{\lambda}), D_i), \quad m = 0, 1, \ldots,$$

for some function p_i. We have,

$$E[N_i \mid D_i] = E\left[\sum_{j=1}^{N_i^*} 1_{\{X_{ij}^* > D_i\}} \middle| D_i \right] = E[N_i^*]\, \bar{F}_i(D_i; \boldsymbol{\theta}), \tag{1.1}$$

so in particular $E[N_i] = E[N_i^*]\, E[\bar{F}_i(D_i; \boldsymbol{\theta})]$.

Let D_i have distribution function G_i and assume that F_i and G_i have densities f_i and g_i with respect to some measure ν. Then the likelihood is

$$L_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^{n} \left(p_i\left(N_i; (\boldsymbol{\theta}, \boldsymbol{\lambda}), D_i\right) \prod_{j=1}^{N_i} \frac{f_i(X_{ij}; \boldsymbol{\theta})}{\bar{F}_i(D_i; \boldsymbol{\theta})} \right) \prod_{i=1}^{n} g_i(D_i). \tag{1.2}$$

Note that we do not have to know the actual form of the g_i to maximize(1.2) w.r.t. $(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

As an example, if $N_i^* \sim \mathrm{Po}(\lambda_i)$ (Poisson) then conditional on D_i, $N_i \sim \mathrm{Po}(\lambda_i \bar{F}_i(D_i; \boldsymbol{\theta}))$ so that

$$L_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \prod_{i=1}^{n} \left(\frac{\lambda_i^{N_i}}{N_i!} e^{-\lambda_i \bar{F}_i(D_i; \boldsymbol{\theta})} \prod_{j=1}^{N_i} f_i(X_{ij}; \boldsymbol{\theta}) \right) \prod_{i=1}^{n} g_i(D_i). \tag{1.3}$$

As another example, assume that $N_i^* \sim \text{NBin}(\alpha_i, \beta_i)$, i.e. with $\boldsymbol{\lambda} = \{(\alpha_i, \beta_i)\}$,

$$p_i^*(m;\boldsymbol{\lambda}) = \binom{m+\alpha_i-1}{\alpha_i-1}\frac{\beta_i^{\alpha_i}}{\left(1+\beta_i\right)^{m+\alpha_i}}.$$

Then conditional on D_i, $N_i \sim \text{NBin}\left(\alpha_i, \frac{\beta_i}{\bar{F}_i(D_i;\theta)}\right)$ so that

$$L_{(n)}(\boldsymbol{\theta},\boldsymbol{\lambda}) = \prod_{i=1}^{n}\left(\binom{N_i-\alpha_i-1}{\alpha_i-1}\frac{\beta_i^{\alpha_i}}{\left(\bar{F}_i(D_i;\boldsymbol{\theta})+\beta_i\right)^{N_i+\alpha_i}}\prod_{j=1}^{N_i} f_i(X_{ij};\boldsymbol{\theta})\right)\prod_{i=1}^{n} g_i(D_i). \quad (1.4)$$

Now consider the claims only. Conditional on the number of claims and the deductibles the likelihood becomes

$$\tilde{L}_{(n)}(\boldsymbol{\theta}) = \prod_{i=1}^{n}\prod_{j=1}^{N_i}\frac{f_i(X_{ij};\boldsymbol{\theta})}{\bar{F}_i(D_i;\boldsymbol{\theta})}. \quad (1.5)$$

Letting $\tilde{\boldsymbol{\theta}}_n$ maximize (1.5) and plugging it into (1.2) yields the corresponding maximizer $\tilde{\boldsymbol{\lambda}}_n$ of $\boldsymbol{\lambda}$. From (1.2) it is seen that $\tilde{\boldsymbol{\lambda}}_n$ is found by maximizing

$$L_{0(n)}(\tilde{\boldsymbol{\theta}}_n,\boldsymbol{\lambda}) = \prod_{i=1}^{n} p_i\left(N_i;(\tilde{\boldsymbol{\theta}}_n,\boldsymbol{\lambda}),D_i\right) \quad (1.6)$$

w.r.t. $\boldsymbol{\lambda}$. The estimated $(\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\lambda}})$ will be called the pseudo-MLE.

Note that the setup (1.2) and (1.5)-(1.6) does not allow for the presence of random effects, but it can easily be extended to do so. The problem is not the general structure of the likelihood, the problem is to maximize it!

In order for the plug-in method to be useful it is necessary that the pseudo-MLE are not too poor compared to the full MLE. To make comparisons, the notion of asymptotic relative efficiency is useful. To explain, let $h(\boldsymbol{\theta},\boldsymbol{\lambda})$ be a function of $(\boldsymbol{\theta},\boldsymbol{\lambda})$, and assume that asymptotically

$$c_n\left(h(\hat{\boldsymbol{\theta}}_n,\hat{\boldsymbol{\lambda}}_n)-h(\boldsymbol{\theta},\boldsymbol{\lambda})\right) \xrightarrow{d} \mathcal{N}\left(0,\hat{\sigma}_h^2\right) \text{ as } n\to\infty,$$

$$c_n\left(h(\tilde{\boldsymbol{\theta}}_n,\tilde{\boldsymbol{\lambda}}_n)-h(\boldsymbol{\theta},\boldsymbol{\lambda})\right) \xrightarrow{d} \mathcal{N}\left(0,\tilde{\sigma}_h^2\right) \text{ as } n\to\infty.$$

The asymptotic relative efficiency of $h(\tilde{\boldsymbol{\theta}}_n,\tilde{\boldsymbol{\lambda}}_n)$ is then

$$\text{ARE}\left(h(\boldsymbol{\theta},\boldsymbol{\lambda})\right) = \frac{\hat{\sigma}_h^2}{\tilde{\sigma}_h^2}.$$

If the model is sufficiently regular, the asymptotic efficiency of the MLE implies that $\mathrm{ARE}(h(\boldsymbol{\theta},\boldsymbol{\lambda})) \leq 1$. In order for $h(\tilde{\boldsymbol{\theta}}_n,\tilde{\boldsymbol{\lambda}}_n)$ to be useful, $\mathrm{ARE}(h(\boldsymbol{\theta},\boldsymbol{\lambda}))$ should be near 1. If for example $c_n = \sqrt{n}$, which is the typical case, then if $\hat{n}$ i.i.d. risks are necessary to get a certain precision of $h(\boldsymbol{\theta},\boldsymbol{\lambda})$ using the MLE $h(\hat{\boldsymbol{\theta}}_n,\hat{\boldsymbol{\lambda}}_n)$, to obtain the same precision using $h(\tilde{\boldsymbol{\theta}}_n,\tilde{\boldsymbol{\lambda}}_n)$ would require $\hat{n}/\mathrm{ARE}(h(\boldsymbol{\theta},\boldsymbol{\lambda}))$ i.i.d. risks.

The actuarial literature has not paid much attention to the fact that most policies carry a deductible. For example, the book *Loss models. From data to decisions* (2004) by Klugman et al., is to a large extent about statistical methods for estimating insurance losses, but it pays rather lip service to the problem with deductibles. Paulsen et. al. (2008) consider the problem of fitting linear mixed models to insurance claimsize data in the presence of deductibles, and it is clear from their findings that estimation is complicated. In statistical terms, the problem with deductibles is a problem with left truncated data. In biostatistics such data are sometimes called ascertained, and in econometrics they belong to the class of Tobit models. A fairly extensive bibliography of papers dealing with such data in all these fields can be found in Paulsen et al. (2008).

In the actuarial literature a lot more attention has been given to the use of covariates and random effects, and then in particular in the framework of GLM and GLMM. However, deductibles are left out of the consideration. In fact, taking deductibles into account would destroy the GLM and GLMM structure of the model. It would go too far to list all relevant papers here, but several references can again be found in Paulsen et al. (2008). More recent contributions, including extensive bibliographies, are Garrido and Zhou (2009) and Frees et. al. (2009).

An alternative to using maximum likelihood is to take a Bayesian approach. Using MCMC methods, quite complicated models can be estimated, as exemplified in e.g. Dimakos and di Rattalma (2002) and Gschlössl and Czado (2007). However, incorporating deductibles into the framework of MCMC is not a trivial matter, and as far as we know this still remains to be done.

Pseudo-MLE are quite popular in statistics. An example that is of interest to actuaries is in the estimation of copulas. Again the full likelihood can be very complicated, so instead the marginal distributions are first estimated, and then the result from this estimation is plugged into the copula itself so that what remains in the second stage is to estimate the copula parameters only. However, it is shown by Joe (2005) that this method can cause a severe loss of efficiency compared to a full MLE.

The rest of the paper goes as follows. In Section 2 we give formulas for the asymptotic distribution of both the MLE and the pseudo-MLE in the i.i.d. case. The corresponding asymptotic distributions for the net premiums above a given deductible are also found. In Section 3 the results are specialized to the case when the claim numbers are Poisson distributed. Letting both claims and deductibles be exponentially or Pareto distributed, numerical values are given for the asymptotic relative efficiencies of the estimated net premiums for various values of the deductibles. Then in Section 4 some data from marine insurance are analyzed, and the effect on the net premium are compared when using pseudo-MLE versus full MLE. This effect is compared to that of using different claimsize distributions, and also to that of using different covariates.

Compared to the complexities that are discussed at the beginning of this section, the models analyzed in Section 3 are rather simple. In fact, for the models in Section 3 a full MLE is easy. Unfortunately, an in depth analysis of a complicated case where the pseudo-MLE is the only alternative, is intractable. In spite of this gap, we feel that an in depth analysis of a simple model can give a useful insight into the problem of estimation in actuarial models with deductibles.

2. Some general results

We will write $P^{\theta,\lambda}$ for the probability measure when the true parameter is $(\boldsymbol{\theta},\boldsymbol{\lambda})$ and e.g. just P^{θ} when $\boldsymbol{\lambda}$ is not relevant.

For $h(\mathbf{x},\mathbf{y})$ we write e.g. $\nabla_x h(\mathbf{x},\mathbf{y})$ for the gradient w.r.t. $\mathbf{x}$ and $\nabla_{x,y} h(\mathbf{x},\mathbf{y})$ for the gradient w.r.t. both parameters. Furthermore, $H_{xx} h(\mathbf{x},\mathbf{y})$ is the Hessian w.r.t. $\mathbf{x}$ and $H_{xy} h(\mathbf{x},\mathbf{y})$ is the second partial derivative matrix w.r.t. both $\mathbf{x}$ and $\mathbf{y}$. Finally, with $H_{xxx} h(\mathbf{x},\mathbf{y})$ we mean the 3-dimensional array obtained by taking the three times partial derivatives of h w.r.t. $\mathbf{x}$. Similarly $H_{xyy} h(\mathbf{x},\mathbf{y})$ is the 3-dimensional array obtained by taking once partial derivative w.r.t. $\mathbf{x}$ and twice w.r.t. $\mathbf{y}$. For A any number, vector, matrix or 3-dimensional array, by $|A|$ we mean the Euclidean norm.

In addition to the assumptions made in Section 1 we shall assume that the N_i^* are i.i.d., the X_{ij}^* are i.i.d. and the D_i are i.i.d. Consequently we shall omit the subscript i in the distribution, so in particular the loglikelihood of (1.2) becomes

$$l_{(n)}(\boldsymbol{\theta},\boldsymbol{\lambda}) \stackrel{\text{def}}{=} \ln L_{(n)}(\boldsymbol{\theta},\boldsymbol{\lambda}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta},\boldsymbol{\lambda}),$$

where

$$l_i(\boldsymbol{\theta},\boldsymbol{\lambda}) = \ln p\big(N_i;(\boldsymbol{\theta},\boldsymbol{\lambda}),D_i\big) + \sum_{j=1}^{N_i} \ln f(X_{ij};\boldsymbol{\theta}) - N_i \ln \bar{F}(D_i;\boldsymbol{\theta}) + \ln g(D_i). \quad (2.1)$$

Furthermore, the conditional loglikelihood of (1.5) becomes

$$\tilde{l}_{(n)}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \tilde{l}_i(\boldsymbol{\theta}),$$

where

$$\tilde{l}_i(\boldsymbol{\theta}) = \sum_{j=1}^{N_i} \ln f(X_{ij};\boldsymbol{\theta}) - N_i \ln \bar{F}(D_i;\boldsymbol{\theta}). \quad (2.2)$$

Define (provided the relevant differentials exist),

$$\begin{aligned} A &\equiv A(\boldsymbol{\theta}) = E^{\theta}\left[\bar{F}(D;\boldsymbol{\theta})\right], \\ \mathbf{B} &\equiv \mathbf{B}(\boldsymbol{\theta}) = -\nabla_{\theta} A(\boldsymbol{\theta}), \\ \mathbf{C} &\equiv \mathbf{C}(\boldsymbol{\theta}) = E^{\theta}\left[\left(\nabla_{\theta} \ln \bar{F}(D;\boldsymbol{\theta})\right)\left(\nabla_{\theta} \ln \bar{F}(D;\boldsymbol{\theta})\right)^{T} \bar{F}(D;\boldsymbol{\theta})\right], \\ \mathbf{J} &\equiv \mathbf{J}(\boldsymbol{\theta}) = E^{\theta}\left[G(X^{*})\left(\nabla_{\theta} \ln f(X^{*};\boldsymbol{\theta})\right)\left(\nabla_{\theta} \ln f(X^{*};\boldsymbol{\theta})\right)^{T}\right] \\ &= \mathbf{J}_0 - E^{\theta}\left[\bar{G}(X)\left(\nabla_{\theta} \ln f(X^{*};\boldsymbol{\theta})\right)\left(\nabla_{\theta} \ln f(X^{*};\boldsymbol{\theta})\right)^{T}\right], \end{aligned}$$

where

$$\mathbf{J}_0 \equiv \mathbf{J}_0(\boldsymbol{\theta}) = E^{\theta}\left[\left(\nabla_{\theta} \ln f(X^{*};\boldsymbol{\theta})\right)\left(\nabla_{\theta} \ln f(X^{*};\boldsymbol{\theta})\right)^{T}\right]$$

is the information when there are no deductibles. Here and later X_j^* and X^* are generic for the X_{ij}^*, X_j and X for the X_{ij}, N^* for the N_i^*, N for the N_i and D for the D_i.

We also let $\mathbf{K}$ be the $(s+t)\times(s+t)$ symmetric matrix blockdivided as (again assuming the relevant differentials exist),

$$\begin{aligned} \mathbf{K}_{11} &= -E^{\theta,\lambda}\left[H_{\theta\theta} \ln p\left(N;(\boldsymbol{\theta},\boldsymbol{\lambda}),D\right)\right], \\ \mathbf{K}_{12} &= -E^{\theta,\lambda}\left[H_{\theta\lambda} \ln p\left(N;(\boldsymbol{\theta},\boldsymbol{\lambda}),D\right)\right], \\ \mathbf{K}_{22} &= -E^{\theta,\lambda}\left[H_{\lambda\lambda} \ln p\left(N;(\boldsymbol{\theta},\boldsymbol{\lambda}),D\right)\right]. \end{aligned}$$

We need the following rather lengthy set of assumptions. It is assumed that $(\boldsymbol{\theta},\boldsymbol{\lambda}) \in \Theta \times \Lambda$, open sets in R^s and R^t respectively.

B1. For all m and d, $\ln p(m,(\boldsymbol{\theta},\boldsymbol{\lambda}),d)$ are three times continuously differentiable in $(\boldsymbol{\theta},\boldsymbol{\lambda})$.

B2. Interchange of differentiation (w.r.t. $(\boldsymbol{\theta},\boldsymbol{\lambda})$) and integration of $\ln p(n,(\boldsymbol{\theta},\boldsymbol{\lambda}),d)$ is allowed up to the second derivative.

B3. For all x, $\ln f(x;\boldsymbol{\theta})$ and $\ln \bar{F}(x;\boldsymbol{\theta})$ are three times continuously differentiable in $\boldsymbol{\theta}$.

B4. For all d and all $\boldsymbol{\theta}$

$$\nabla_{\theta} \bar{F}(d;\boldsymbol{\theta}) = \int_d^{\infty} \nabla_{\theta} f(x;\boldsymbol{\theta})\,dv(x) \text{ and } H_{\theta\theta} \bar{F}(d;\boldsymbol{\theta}) = \int_d^{\infty} H_{\theta\theta} f(x;\boldsymbol{\theta})\,dv(x).$$

Furthermore

$$\nabla_{\theta} A = E^{\theta}\left[\nabla_{\theta} \bar{F}(D;\boldsymbol{\theta})\right] \text{ and } H_{\theta\theta} A = E^{\theta}\left[H_{\theta\theta} \bar{F}(D;\boldsymbol{\theta})\right].$$

B5. For all $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ there are open sets U and V with $(\boldsymbol{\theta}, \boldsymbol{\lambda}) \in U \times V$ so that for all $(\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1) \in U \times V$,

$$\left| \nabla_\theta \ln \frac{p\big(m; (\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1), d\big)}{p\big(m; (\boldsymbol{\theta}_1, \boldsymbol{\lambda}), d\big)} \right| < M(m, d; \boldsymbol{\lambda}),$$

where $E^{\theta,\lambda}[M(N, D; \boldsymbol{\lambda}_1)] < \infty$.

B6. For all $(\boldsymbol{\theta}, \boldsymbol{\lambda})$ there are open sets U and V with $(\boldsymbol{\theta}, \boldsymbol{\lambda}) \in U \times V$ so that for all $(\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1) \in U \times V$,

$$\begin{aligned} &\left| H_{\theta\theta\theta} \ln p\big(m; (\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1), d\big) \right| + \left| H_{\theta\theta\lambda} \ln p\big(m; (\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1), d\big) \right| + \left| H_{\theta\lambda\lambda} \ln p\big(m; (\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1), d\big) \right| \\ &\quad + \left| H_{\lambda\lambda\lambda} \ln p\big(m; (\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1), d\big) \right| < M(m, d), \end{aligned}$$

with $E^{\theta,\lambda}[M(N, D)] < \infty$.

B7. For all $\boldsymbol{\theta}$ there is an open set U with $\boldsymbol{\theta} \in U$ so that for all $\boldsymbol{\theta}_1 \in U$,

$$\begin{aligned} \left| H_{\theta\theta\theta} \ln f(x; \boldsymbol{\theta}_1) \right| &< M_1(x), \\ \left| H_{\theta\theta\theta} \ln \bar{F}(d; \boldsymbol{\theta}_1) \right| &< M_2(d) \end{aligned}$$

and $E^{\theta}[M_1(X^*) + M_2(D)| < \infty$.

B8. For any $(\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1) \neq (\boldsymbol{\theta}, \boldsymbol{\lambda})$,

$$P^{\theta,\lambda}\Big(p\big(N; (\boldsymbol{\theta}_1, \boldsymbol{\lambda}_1), D\big) = p\big(N; (\boldsymbol{\theta}, \boldsymbol{\lambda}), D\big) \Big) < 1.$$

B9. For any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}$,

$$P^{\theta}\Big(f\big(X^* \mid X^* > D, \boldsymbol{\theta}_1\big) = f\big(X^* \mid X^* > D, \boldsymbol{\theta}\big) \Big) < 1.$$

B10. The quantity A is positive and the matrices $\mathbf{J} - \mathbf{C}$ and $\mathbf{K}_{22}$ are positive definite.

Note that if either $N^* \sim \text{Po}(\lambda)$ or $N^* \sim \text{NBin}(\alpha, \beta)$, then conditions B1, B2, B5 and B6 are satisfied if conditions B3, B4 and B7 are.

Here is the main theoretical result of the paper. Its proof is given in the appendix.

Theorem 2.1. *Assume B1-B10 and that the equations* $\nabla_\theta \tilde{l}_{(n)}(\boldsymbol{\theta}) = 0$ *and* $\nabla_{\theta,\lambda} l_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = 0$ *both have unique solutions* $\tilde{\boldsymbol{\theta}}_n$ *and* $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\lambda}}_n)$ *for all n. Assume also that the equation* $\nabla_\lambda l_{(n)}(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{\lambda}) = 0$ *has a unique solution* $\tilde{\boldsymbol{\lambda}}_n$ *for all n. Then as* $n \to \infty$,

$$\sqrt{n}\Big((\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\lambda}}_n) - (\boldsymbol{\theta}, \boldsymbol{\lambda}) \Big) \xrightarrow{d} \mathcal{N}(0, \Sigma^M) \ \textit{under} \ P^{\theta,\lambda}$$

and

$$\sqrt{n}\Big((\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\lambda}}_n) - (\boldsymbol{\theta}, \boldsymbol{\lambda}) \Big) \xrightarrow{d} \mathcal{N}(0, \Sigma^P) \ \textit{under} \ P^{\theta,\lambda}.$$

Here

$$\Sigma^M = \frac{1}{\mu(\boldsymbol{\lambda})}\begin{bmatrix} (\mathbf{J}-\mathbf{C}')^{-1} & (\mathbf{J}-\mathbf{C}')^{-1}\mathbf{R} \\ \mathbf{R}^T(\mathbf{J}-\mathbf{C}')^{-1} & \mu(\boldsymbol{\lambda})\mathbf{K}_{22}^{-1} + \mathbf{R}^T(\mathbf{J}-\mathbf{C}')^{-1}\mathbf{R} \end{bmatrix},$$

where

$$\begin{aligned} \mu(\boldsymbol{\lambda}) &= E[N^*], \\ \mathbf{C}' &= \mathbf{C} - \frac{1}{\mu(\boldsymbol{\lambda})}\left(\mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}\right), \\ \mathbf{R} &= -\mathbf{K}_{12}\mathbf{K}_{22}^{-1}. \end{aligned}$$

Furthermore Σ^P *is the same as* Σ^M*, but with* $(\mathbf{J}-\mathbf{C}')^{-1}$ *replaced by* $(\mathbf{J}-\mathbf{C})^{-1}$.

Remark 2.1. Since $\mathbf{K}$ is nonnegative definite,

$$(\mathbf{J}-\mathbf{C}') - (\mathbf{J}-\mathbf{C}) = \mathbf{C}-\mathbf{C}' = \frac{1}{\mu(\boldsymbol{\lambda})}\left(\mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}\right)$$

is nonnegative definite as well. In fact, it is positive definite if and only if $\mathbf{K}$ is invertible, and then $\mathbf{C}-\mathbf{C}' = (\mu(\boldsymbol{\lambda})\mathbf{K}^{11})^{-1}$ where $\mathbf{K}^{11}$ is the upper left block in the blockdivision of $\mathbf{K}^{-1}$. It follows from Exercise 9 (iii) p. 70 in Rao (1973) that $\mathbf{V} = (\mathbf{J}-\mathbf{C})^{-1} - (\mathbf{J}-\mathbf{C}')^{-1}$ is nonnegative definite. From Theorem 2.1,

$$\Sigma^P - \Sigma^M = \frac{1}{\mu(\boldsymbol{\lambda})}\begin{bmatrix} \mathbf{V} & \mathbf{V}\mathbf{R} \\ \mathbf{R}^T\mathbf{V} & \mathbf{R}^T\mathbf{V}\mathbf{R} \end{bmatrix}.$$

Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be s and t vectors respectively. Then

$$(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)(\Sigma_P - \Sigma_M)(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T = \frac{1}{\mu(\boldsymbol{\lambda})}(\boldsymbol{\alpha}+\mathbf{R}\boldsymbol{\beta})^T\,\mathbf{V}(\boldsymbol{\alpha}+\mathbf{R}\boldsymbol{\beta}) \geq 0.$$

Therefore, we have proved directly that $\Sigma^P - \Sigma^M$ is nonnegative definite, a fact that also follows from the efficiency of the MLE. For a general sequence $\{Y_n\}$, denoting the asymptotic variance of $\sqrt{n}\,Y_n$ by $\mathrm{aVar}(Y_n)$, note that for any $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} = -\mathbf{R}\boldsymbol{\beta}$,

$$\mathrm{aVar}\left(\boldsymbol{\alpha}^T\hat{\boldsymbol{\theta}}_n + \boldsymbol{\beta}^T\hat{\boldsymbol{\lambda}}_n\right) = \mathrm{aVar}\left(\boldsymbol{\alpha}^T\tilde{\boldsymbol{\theta}}_n + \boldsymbol{\beta}^T\tilde{\boldsymbol{\lambda}}_n\right),$$

i.e. for such particular linear combinations of the parameters the estimators are equally efficient.

Remark 2.2. When $\boldsymbol{\theta}$ is known, the MLE $\check{\boldsymbol{\lambda}}_n$ of $\boldsymbol{\lambda}$ is obtained by maximizing $L_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ w.r.t. $\boldsymbol{\lambda}$. As above we get as $n \to \infty$,

$$\sqrt{n}\left(\check{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{K}_{22}^{-1}\right) \quad \text{under} \quad P^{\theta,\lambda}.$$

It follows from Theorem 2.1 and Remark 2.1 that

$$\mathrm{aVar}(\check{\boldsymbol{\lambda}}_n) \leq \mathrm{aVar}(\hat{\boldsymbol{\lambda}}_n) \leq \mathrm{aVar}(\tilde{\boldsymbol{\lambda}}_n),$$

where for nonnegative definite matrices $\mathbf{A}_1$ and $\mathbf{A}_2$, by $\mathbf{A}_1 \leq \mathbf{A}_2$ we mean that $\mathbf{A}_2 - \mathbf{A}_1$ is nonnegative definite. Inference on $\tilde{\boldsymbol{\lambda}}_n$ using $\mathbf{K}_{22}^{-1}$, or an estimator of it, as variance may give wrong conclusions, for example confidence regions for $\boldsymbol{\lambda}$ will be too narrow. A proper adjustment taking the uncertainty of using $\tilde{\boldsymbol{\theta}}_n$ instead of $\boldsymbol{\theta}$ into account may be difficult for complex non i.i.d. models, and this is a drawback with the pseudo-MLE. One possible way out is to use a bootstrap method, but this can be very time consuming.

Remark 2.3. Instead of assuming the D_i i.i.d., we could let them be deterministic $d_1, d_2, \ldots$ satisfying some proper limit assumptions such as the existence of

$$a = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \bar{F}(d_i; \boldsymbol{\theta})$$

and so on. This would give a result similar to Theorem 2.1 where the terms in Σ^M and Σ^P evaluated as expectations now would be evaluated as limits of averages.

The following not obvious result is easy to prove using Theorem 2.1 and Remark 2.1.

Corollary 2.1. *In addition to the assumptions of Theorem 2.1, let $\boldsymbol{\theta} = \theta$ and $\boldsymbol{\lambda} = \lambda$ be scalars. Then*

$$ARE(\theta) \leq ARE(\lambda),$$

with strict inequality whenever $C > C'$.

A quantity of great interest in insurance is the net premium, i.e. the expected total claim paid by the insurance company for a risk with deductible d,

$$m_d(\boldsymbol{\theta}, \boldsymbol{\lambda}) = E^{\theta,\lambda}\left[\sum_{i=1}^{N^*} (X_i^* - d)_+\right] = \mu(\boldsymbol{\lambda})\int_d^\infty \bar{F}(x; \boldsymbol{\theta})dx \stackrel{\text{def}}{=} \mu(\boldsymbol{\lambda})\mu_d(\boldsymbol{\theta}).$$

In reinsurance $m_d(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is called the net excess of loss (XL) premium. The next result follows directly by using Theorem 2.1 and the delta method on $m_d(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \mu(\boldsymbol{\lambda})\mu_d(\boldsymbol{\theta})$.

Corollary 2.2. *Under the assumptions of Theorem 2.1, as $n \to \infty$,*

$$\sqrt{n}\left(m_d(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\lambda}}_n) - m_d(\boldsymbol{\theta}, \boldsymbol{\lambda})\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{M, m_d}\right) \text{ under } P^{\theta,\lambda}$$

and

$$\sqrt{n}\left(m_d(\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\lambda}}_n) - m_d(\boldsymbol{\theta}, \boldsymbol{\lambda})\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{P,m_d}\right) \text{ under } P^{\theta,\lambda}$$

w.r.t. $P^{\theta,\lambda}$. *Here*

$$\sigma^2_{M,m_d} = m_d^2\left(\frac{1}{\mu(\boldsymbol{\lambda})}\left(\left(\nabla_\theta \ln \mu_d + \mathbf{R}\nabla_\lambda \ln \mu\right)^T (\mathbf{J} - \mathbf{C}')^{-1}\left(\nabla_\theta \ln \mu_d + \mathbf{R}\nabla_\lambda \ln \mu\right)\right)\right.$$
$$\left. + (\nabla_\lambda \ln \mu)^T \mathbf{K}_{22}^{-1}(\nabla_\lambda \ln \mu)\right),$$

where $\mu = \mu(\boldsymbol{\lambda})$, $\mu_d = \mu_d(\boldsymbol{\theta})$ *and* $m_d = m_d(\boldsymbol{\theta}, \boldsymbol{\lambda})$.

Furthermore, σ^2_{P,m_d} *is the same as* σ^2_{M,m_d}, *but with* $(\mathbf{J} - \mathbf{C}')^{-1}$ *replaced by* $(\mathbf{J} - \mathbf{C})^{-1}$.

3. The i.i.d. compound Poisson case

In this section we shall assume that $N^* \sim \text{Po}(\lambda)$ so that $N \sim \text{Po}(\lambda \bar{F}(D; \boldsymbol{\theta}))$. As mentioned in Section 2, this means that conditions B1, B2, B5, B6 and B8 of Section 2 are satisfied if conditions B3, B4 and B7 are.

Taking the logarithm in (1.3) and then differentiating gives for the gradient,

$$\nabla_{\theta,\lambda}\, l_{(n)}(\boldsymbol{\theta}, \lambda) = \sum_{i=1}^{n}\left(\sum_{j=1}^{N_i}\left(\nabla_\theta \ln f(X_{ij}; \boldsymbol{\theta})\right) - \lambda \nabla_\theta \bar{F}(D_i; \boldsymbol{\theta}), \frac{1}{\lambda} N_i - \bar{F}(D_i; \boldsymbol{\theta})\right). \tag{3.1}$$

Set $\nabla_\lambda l_{(n)}(\boldsymbol{\theta}, \lambda) = 0$, solve for λ and insert this λ into $\nabla_\theta l_{(n)}(\boldsymbol{\theta}, \lambda) = 0$. This gives that $\hat{\boldsymbol{\theta}}_n$ solves

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i} \nabla_\theta \ln f(X_{ij}; \boldsymbol{\theta}) - \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} \bar{F}(D_i; \boldsymbol{\theta})} \sum_{i=1}^{n} \nabla_\theta \bar{F}(D_i; \boldsymbol{\theta}) = 0, \tag{3.2}$$

and then

$$\hat{\lambda}_n = \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} \bar{F}(D_i; \hat{\boldsymbol{\theta}}_n)}. \tag{3.3}$$

Taking the logarithm in (1.5) and differentiating gives that $\tilde{\boldsymbol{\theta}}_n$ solves

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i} \nabla_\theta \ln f(X_{ij}; \boldsymbol{\theta}) - \sum_{i=1}^{n} \frac{N_i}{\bar{F}(D_i; \boldsymbol{\theta})} \nabla_\theta \bar{F}(D_i; \boldsymbol{\theta}) = 0. \tag{3.4}$$

Finally, taking the logarithm in (1.6) and differentiating w.r.t. λ then gives

$$\tilde{\lambda}_n = \frac{\sum_{i=1}^n N_i}{\sum_{i=1}^n \bar{F}(D_i;\tilde{\boldsymbol{\theta}}_n)}. \tag{3.5}$$

A comparison of (3.2)-(3.3) with (3.4)-(3.5) shows that basically (see Lemma 3.1 below) $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\lambda}}_n) = (\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\lambda}}_n)$ if and only if $D_1 = \ldots = D_n$. Because of the simple form (3.2)-(3.3), in the i.i.d. case it does not seem more complicated to find the MLE than the pseudo-MLE. However, as discussed in the introduction, this is no longer the case with more general models.

Using that

$$\ln p\big(N;(\boldsymbol{\theta},\lambda),D\big) = N\ln\lambda + N\ln\bar{F}(D;\boldsymbol{\theta}) - \lambda\bar{F}(D;\boldsymbol{\theta}) - \ln N!,$$

some straightforward calculations using Lemma A.1 gives,

$$\mathbf{K}_{11} = \lambda\mathbf{C}, \quad \mathbf{K}_{12} = \mathbf{B} \quad \text{and} \quad \mathbf{K}_{22} = \frac{1}{\lambda}A.$$

Furthermore, $\mathbf{C}' = \frac{1}{A}\mathbf{B}\mathbf{B}^T$.

Lemma 3.1. *Assume that* $\ln\bar{F}(x;\boldsymbol{\theta})$ *is differentiable in* $\boldsymbol{\theta}$ *and that* $E^\theta[\nabla_\theta\bar{F}(D;\boldsymbol{\theta})] = \nabla_\theta A$. *Then the matrix* $\mathbf{C} - \mathbf{C}'$ *is positive definite unless for some nonzero vector* $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^T\mathbf{U}$ *equals a constant, where* $\mathbf{U} = -\nabla_\theta \ln\bar{F}(V;\boldsymbol{\theta})$ *and* V *has the distribution*

$$P^\theta(V \le v) = \frac{1}{A}\int_0^v \bar{F}(d;\boldsymbol{\theta})\,dG(d).$$

This result is proved in the appendix. We saw above that $D_1 = \cdots = D_n$ implies that $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\lambda}}_n) = (\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\lambda}}_n)$ so in this case the ARE equals 1. It is therefore natural that the ARE will decrease as the D_i gets more spread out. The following result, proved in the appendix, shows that such a hypothesis indeed makes some sense.

Corollary 3.1. *Let* U *be as in Lemma 3.1, and assume that* $\boldsymbol{\theta} = \theta$ *is a scalar. Then under the assumptions in Theorem 2.1,*

$$ARE(\theta) = 1 - \frac{Var^\theta[U]}{\frac{1}{A}J - E^\theta[U]^2},$$

$$ARE(\lambda) = 1 - \frac{E^\theta[U]^2\,Var^\theta[U]}{\left(\frac{1}{A}J - Var^\theta[U]\right)\left(\frac{1}{A}J - E^\theta[U]^2\right)},$$

$$ARE\big(m_d(\theta,\lambda)\big) = 1 - \frac{\left(E^\theta[U] + \frac{d}{d\theta}\ln\mu_d\right)^2 Var^\theta[U]}{\left(\frac{1}{A}J - E^\theta[U]^2\right)\left(\frac{1}{A}J - E^\theta[U^2] + \left(E^\theta[U] + \frac{d}{d\theta}\ln\mu_d\right)^2\right)}.$$

Furthermore, $ARE(\theta) \le ARE(\lambda)$ *with strict inequality whenever* $C > \frac{B^2}{A}$, *or equivalently whenever* $Var^\theta[U] > 0$.

Example 3.1. Assume that $D \sim \exp(\alpha)$, and that $X^* \sim \exp(\theta)$, meaning e.g. that $f(x;\theta) = \theta e^{-\theta x} 1_{\{x>0\}}$. Using that $\bar{F}(x;\theta) = e^{-\theta x}$, $x \ge 0$, it is not difficult to show that

$$A = \frac{\alpha}{\alpha + \theta} \quad \text{and} \quad J = \frac{1}{\theta^2} - \frac{\alpha^2 + \theta^2}{\theta(\alpha + \theta)^3}.$$

This then gives that

$$\frac{1}{A} J = \frac{1}{\theta^2} + \frac{2}{(\alpha + \theta)^2}.$$

With reference to the notation of Corollary 3.1, we see that $V = U \sim \exp(\alpha + \theta)$. Therefore,

$$E^\theta[U] = \frac{1}{\alpha + \theta}, \quad E^\theta[U^2] = \frac{2}{(\alpha + \theta)^2} \quad \text{and} \quad \text{Var}^\theta[U] = \frac{1}{(\alpha + \theta)^2}.$$

Finally, $\mu_d = \frac{1}{\theta} e^{-\theta d}$ so that $\frac{d}{d\theta} \ln \mu_d = -(\frac{1}{\theta} + d)$. Some straightforward, but slightly tedious calculations using that $E^\theta[X^*] = \frac{1}{\theta}$ and $E^\theta[D] = \frac{1}{\alpha}$ gives that

$$\text{ARE}(\theta) = 1 - \frac{1}{1 + \left(1 + \frac{E^\theta[X^*]}{E^\theta[D]}\right)^2},$$

$$\text{ARE}(\lambda) = 1 - \frac{1}{\left(1 + \left(1 + \frac{E^\theta[X^*]}{E^\theta[D]}\right)^2\right)^2} = 1 - (1 - \text{ARE}(\theta))^2,$$

$$\text{ARE}(m_d(\theta, \lambda)) = 1 - \frac{\left(\frac{1}{1 + \frac{E^\theta[X^*]}{E^\theta[D]}} + \frac{d}{E^\theta[X^*]}\right)^2}{\left(1 + \left(1 + \frac{E^\theta[X^*]}{E^\theta[D]}\right)^2\right)\left(1 + \left(\frac{1}{1 + \frac{E^\theta[X^*]}{E^\theta[D]}} + \frac{d}{E^\theta[X^*]}\right)^2\right)}.$$

Tables 3.1 and 3.2 give $\text{ARE}(\theta)$, $\text{ARE}(\lambda)$ and $\text{ARE}(m_d(\theta, \lambda))$ for various values of $\frac{E^\theta[D]}{E^\theta[X^*]} = \frac{\theta}{\alpha}$ and $\frac{d}{E^\theta[X^*]}$. From Table 3.2 it is seen that as long as the

TABLE 3.1

VALUES OF ARE(θ) AND ARE(λ) FROM EXAMPLE 3.1 FOR VARIOUS VALUES OF $\frac{E^\theta[D]}{E^\theta[X^*]}$.

$\frac{E^\theta[D]}{E^\theta[X^*]}$	0.1	0.2	0.5	1	2	5	10
ARE(θ)	0.9918	0.9730	0.9000	0.8000	0.6933	0.5902	0.5475
ARE(λ)	0.9999	0.9993	0.9900	0.9600	0.9053	0.8320	0.7953

TABLE 3.2

VALUES OF ARE($m_d(\theta,\lambda)$) FROM EXAMPLE 3.1 FOR VARIOUS VALUES OF $\frac{E^\theta[D]}{E^\theta[X^*]}$ AND $\frac{d}{E^\theta[X^*]}$.

		$\frac{E^\theta[D]}{E^\theta[X^*]}$						
		0.1	0.2	0.5	1	2	5	10
$\frac{d}{E^\theta[X^*]}$	0.1	0.9997	0.9982	0.9842	0.9471	0.8861	0.8092	0.7717
	0.2	0.9994	0.9968	0.9779	0.9342	0.8680	0.7884	0.7504
	0.5	0.9979	0.9917	0.9590	0.9000	0.8333	0.7377	0.6991
	1	0.9955	0.9844	0.9360	0.8615	0.7738	0.6841	0.6449
	2	0.9933	0.9777	0.9155	0.8346	0.7302	0.6356	0.5953
	5	0.9931	0.9739	0.9034	0.8064	0.7016	0.6019	0.5601
	10	0.9919	0.9732	0.9009	0.8018	0.6950	0.5936	0.5513

deductibles are moderately large compared to the claims, i.e. $\frac{E^\theta[D]}{E^\theta[X^*]}$ is not more than 0.5, the loss in efficiency from using the pseudo-MLE is not unduly high when ARE($m_d(\theta,\lambda)$) is the quantity of interest. Finally we notice that ARE($m_d(\theta,\lambda)$) is decreasing in both $\frac{E^\theta[D]}{E^\theta[X^*]}$ and $\frac{d}{E^\theta[X^*]}$.

Example 3.2. Assume that $D \sim \text{Pa}(\beta, q)$ (Pareto distributed) and that $X \sim \text{Pa}(\alpha, p)$, meaning e.g. that $f(x;\alpha,p) = p\alpha^p(\alpha + x)^{-(1+p)}1_{\{x>0\}}$. Then

$$A = p\alpha^p\beta^q\int_0^\infty (\alpha+x)^{-(1+p)}(\beta+x)^{-q}\,dx.$$

This can be integrated numerically. Similar expressions for B, C and J can be derived, for example

$$\frac{\partial}{\partial\alpha}A = p\alpha^{p-1}\beta^q\int_0^\infty (px-\alpha)(\alpha+x)^{-(2+p)}(\beta+x)^{-q}\,dx.$$

Finally

$$\mu_d = \frac{\alpha^p}{p-1}(d+\alpha)^{1-p}.$$

Using the formulas in Corollary 2.2 we can easily calculate numerically $\mathrm{ARE}(m_d(\boldsymbol{\theta},\lambda))$. We have done so both for the case with p known, i.e. $\boldsymbol{\theta}=a$, and the case with both a and p unknown, i.e. $\boldsymbol{\theta}=(a,p)$. This to see if the loss in efficiency is higher or lower in the multiparameter case compared to the one parameter case.

In Tables 3.3 and 3.4, $\mathrm{ARE}(m_d(\boldsymbol{\theta},\lambda))$ values are given for the same values of $\frac{d}{E^{\theta}[X^*]}$ and $\frac{E^{\theta}[D]}{E^{\theta}[X^*]}=\frac{p-1}{q-1}$ as in Table 3.2. The parameters are $\alpha=2$, $p=3$ and $q=8$. In Table 3.3, $\boldsymbol{\theta}=p$, i.e. p is known, while in Table 3.4, $\boldsymbol{\theta}=(\alpha,p)$.

As opposed to the case in Example 3.1, the $\mathrm{ARE}(m_d(\boldsymbol{\theta},\lambda))$ are not independent of the parameters. Therefore, in Tables 3.5 and 3.6 the same quantities are presented, but now with the more heavy tailed distributions with $\alpha=1$, $p=2$ and $q=4$. In Table 3.5, $\boldsymbol{\theta}=p$, while in Table 3.6, $\boldsymbol{\theta}=(\alpha,p)$.

Comparing Tables 3.3 and 3.4 it is seen that the loss in efficiency is always smaller in the two parameter case, and at times it is considerably smaller. This

TABLE 3.3

VALUES OF $\mathrm{ARE}(m_d(\boldsymbol{\theta},\lambda))$ FOR VARIOUS VALUES OF $\frac{E^{\theta}[D]}{E^{\theta}[X^*]}$ AND $\frac{d}{E^{\theta}[X^*]}$.
THE PARAMETERS ARE $\alpha=2$, $p=3$ AND $q=8$. THE PARAMETER p IS ASSUMED KNOWN.

		$\frac{E^{\theta}[D]}{E^{\theta}[X^*]}$						
		0.1	0.2	0.5	1	2	5	10
$\frac{d}{E^{\theta}[X^*]}$	0.1	0.9843	0.9619	0.9130	0.8777	0.8651	0.8891	0.9199
	0.2	0.9834	0.9592	0.9043	0.8604	0.8358	0.8444	0.8690
	0.5	0.9814	0.9539	0.8874	0.8333	0.7769	0.7473	0.7473
	1	0.9798	0.9494	0.8736	0.7992	0.7303	0.6687	0.6450
	2	0.9785	0.9459	0.8630	0.7788	0.6965	0.6138	0.5743
	5	0.9775	0.9432	0.8550	0.7643	0.6732	0.5779	0.5297
	10	0.9771	0.9431	0.8532	0.7591	0.6654	0.5664	0.5156

TABLE 3.4

VALUES OF $\mathrm{ARE}(m_d(\boldsymbol{\theta},\lambda))$ FOR VARIOUS VALUES OF $\frac{E^{\theta}[D]}{E^{\theta}[X^*]}$ AND $\frac{d}{E^{\theta}[X^*]}$.
THE PARAMETERS ARE THE SAME AS IN TABLE 3.3, BUT NOW BOTH α AND p ARE ASSUMED UNKNOWN.

		$\frac{E^{\theta}[D]}{E^{\theta}[X^*]}$						
		0.1	0.2	0.5	1	2	5	10
$\frac{d}{E^{\theta}[X^*]}$	0.1	0.9966	0.9886	0.9587	0.9189	0.8812	0.8734	0.8998
	0.2	0.9965	0.9881	0.9546	0.9072	0.8573	0.8333	0.8508
	0.5	0.9972	0.9891	0.9516	0.8903	0.8132	0.7406	0.7350
	1	0.9986	0.9929	0.9589	0.8934	0.7946	0.6734	0.6202
	2	0.9995	0.9968	0.9734	0.9154	0.8132	0.6535	0.5586
	5	0.9982	0.9960	0.9819	0.9410	0.8548	0.6915	0.5731
	10	0.9964	0.9929	0.9796	0.9453	0.8705	0.7195	0.6009

TABLE 3.5

VALUES OF $\text{ARE}(m_d(\boldsymbol{\theta},\lambda))$ FOR VARIOUS VALUES OF $\frac{E^\theta[D]}{E^\theta[X^*]}$ AND $\frac{d}{E^\theta[X^*]}$.
THE PARAMETERS ARE $\alpha = 1$, $p = 2$ AND $q = 4$. THE PARAMETER p IS ASSUMED KNOWN.

		$\frac{E^\theta[D]}{E^\theta[X^*]}$						
		0.1	0.2	0.5	1	2	5	10
$\frac{d}{E^\theta[X^*]}$	0.1	0.9732	0.9419	0.8837	0.8405	0.8150	0.8206	0.8481
	0.2	0.9708	0.9387	0.8751	0.8316	0.7884	0.7765	0.7915
	0.5	0.9684	0.9330	0.8602	0.7972	0.7430	0.6999	0.6894
	1	0.9667	0.9358	0.8494	0.7779	0.7118	0.6482	0.6203
	2	0.9653	0.9334	0.8415	0.7642	0.6902	0.6138	0.5752
	5	0.9643	0.9333	0.8356	0.7542	0.6749	0.5903	0.5453
	10	0.9639	0.9331	0.8335	0.7506	0.6694	0.5831	0.5351

TABLE 3.6

VALUES OF $\text{ARE}(m_d(\boldsymbol{\theta},\lambda))$ FOR VARIOUS VALUES OF $\frac{E^\theta[D]}{E^\theta[X^*]}$ AND $\frac{d}{E^\theta[X^*]}$.
THE PARAMETERS ARE THE SAME AS IN TABLE 3.5, BUT NOW BOTH α AND p ARE ASSUMED UNKNOWN.

		$\frac{E^\theta[D]}{E^\theta[X^*]}$						
		0.1	0.2	0.5	1	2	5	10
$\frac{d}{E^\theta[X^*]}$	0.1	0.9975	0.9902	0.9596	0.9117	0.8485	0.7778	0.7608
	0.2	0.9975	0.9901	0.9578	0.9057	0.8343	0.7469	0.7149
	0.5	0.9979	0.9910	0.9580	0.9006	0.8155	0.6956	0.6307
	1	0.9983	0.9934	0.9633	0.9058	0.8150	0.6733	0.5801
	2	0.9982	0.9937	0.9686	0.9172	0.8359	0.6763	0.5650
	5	0.9968	0.9932	0.9719	0.9356	0.8497	0.7017	0.5829
	10	0.9955	0.9907	0.9713	0.9320	0.8592	0.7183	0.6009

indicates that for more complex models when the need to use the plug-in method is much more prevalent, the loss in efficiency may be quite insignificant. The same is seen when comparing Tables 3.5 and 3.6, except for the cases with large $\frac{E^\theta[D]}{E^\theta[X^*]}$ and small $\frac{d}{E^\theta[X^*]}$. However, such cases are not commonly encountered in practice.

It is also worth noting from Tables 3.3 and 3.5 that for known p, $\text{ARE}(m_d(\boldsymbol{\theta},\lambda))$ is decreasing in $\frac{d}{E^\theta[X^*]}$ for fixed $\frac{E^\theta[D]}{E^\theta[X^*]}$. However, when $\frac{d}{E^\theta[X^*]}$ is fixed, $\text{ARE}(m_d(\boldsymbol{\theta},\lambda))$ is not necessarily decreasing in $\frac{E^\theta[D]}{E^\theta[X^*]}$, this in contrast to the situation in Example 3.1. When both α and p are unknown, $\text{ARE}(m_d(\boldsymbol{\theta},\lambda))$ is not decreasing in any direction.

4. Analysis of marine insurance data

The idea for this investigation arose from problems in marine insurance. To indicate the complexities involved in this particular business, let there be m shipowners, where shipowner i has n_i ships. For each ship there are covariates such as the value of the ship, tonnage, age, speed and several others, depending on availability, like the number of times it has been detained at a harbour due to poor conditions. Transformations and combinations of these covariates are often useful. Then there are dummy variables like type of ship, type of engine, geographical areas the ship is sailing and so on. Finally, there may be random effects like shipowner, or ships nested within shipowners. Other reasonable random effects could be registration country of the ship or alternatively of the shipowner. All or some of these factors can be present both at the claimsize distribution as well as at the claim frequency distribution. Finding full MLE for such complicated models can be an impossible task, and a two-step procedure may be the only option. Even without introducing random effects, in the above example it may be time saving to explore the effects of covariates on claimsizes and claim frequency separately, and once good covariates have been identified, full MLE can be found. In fact, we have practiced this method for some time.

The models fitted here will be rather simple, but due to the heterogeneity of the data, it is necessary to include some covariates. The model assumptions are, assuming there are n ships,

- $N_i^* \sim \mathrm{Po}(\lambda_i)$ with $\ln \lambda_i = \boldsymbol{\alpha}' \mathbf{w}_i$, $i = 1, \ldots, n$.
- $X_{ij}^* \sim \mathrm{LN}(\boldsymbol{\beta}' \mathbf{z}_i, \sigma^2)$, $j = 1, \ldots, N_i^*$, $i = 1, \ldots, n$,

i.e. $\ln X_{ij}^*$ is normally distributed with expectation $\boldsymbol{\beta}'\mathbf{z}_i$ and variance σ^2. The $\mathbf{w}_i$ and the $\mathbf{z}_i$ are covariate vectors, while $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2)$ are unknown parameters. We will always include an intercept, i.e. $w_{i1} = z_{i1} = 1$.

We have a fair amount of experience with various claimsize distributions, and the conclusion is that the lognormal distribution in general performs very well. In terms of its heavy-tailedness it lies between the two distributions of Examples 3.1 and 3.2.

In order to assess what is most crucial for the estimation of the net premiums, claimsize distributions or estimation method, we tried the full MLE with two more claimsize distributions, and compared the results with the lognormal case. These distributions were Weibull and Pareto, i.e.

- $X_{ij}^* \sim W(a_i, \tau)$ with $\ln a_i = \boldsymbol{\gamma}' \mathbf{z}_i$, $j = 1, \ldots, N_i^*$, $i = 1, \ldots, n$,
- $X_{ij}^* \sim \mathrm{Pa}(b_i, p)$ with $\ln b_i = \boldsymbol{\delta}' \mathbf{z}_i$, $j = 1, \ldots, N_i^*$, $i = 1, \ldots, n$.

By $Y \sim W(a, \tau)$ we mean that

$$F_Y(y) = 1 - e^{-\left(\frac{y}{a}\right)^\tau}, \quad y > 0,$$

and the definition of $\mathrm{Pa}(b,p)$ is given in Example 3.2.
For the net premiums without deductibles, these three distributions give

$$E[N^*]\,E[X^*] = \begin{cases} e^{\alpha'\mathbf{z}+\beta'\mathbf{w}}\,e^{\frac{1}{2}\sigma^2}, & \text{Lognormal,}\\ e^{\alpha'\mathbf{w}+\gamma'\mathbf{z}}\,\Gamma(1+\tau^{-1}), & \text{Weibull,}\\ e^{\alpha'\mathbf{w}+\delta'\mathbf{z}}\,\dfrac{1}{p-1},\quad p>1, & \text{Pareto.}\end{cases}$$

We see that all these expressions are comparable. Since we only compute net premiums, the choice of claim number distribution does not matter much. The corresponding formulas with deductibles are a bit more complicated, for example in the lognormal case

$$\begin{aligned} m_d(\boldsymbol{\alpha},\boldsymbol{\beta},\sigma^2) &= E[N^*]\,E[(X^*-d)_+] \\ &= e^{\alpha'\mathbf{w}}\left(e^{\boldsymbol{\beta}'\mathbf{z}}\Phi\left(\frac{\boldsymbol{\beta}'\mathbf{z}-\ln d}{\sigma}+\sigma\right)-d\Phi\left(\frac{\boldsymbol{\beta}'\mathbf{z}-\ln d}{\sigma}\right)\right), \end{aligned} \tag{4.1}$$

where Φ is the standard normal distribution function.

The data analyzed are partial hull claims, i.e. total losses are excluded since they are usually priced separately. They are provided to us by NHC (Norwegian Hull Club) and cover the years 1997-2003. The data used in Paulsen et al. (2008) is from the same database, but covering a slightly different time period (1995-2001). Since shiptypes are very different, we have concentrated on two types, bulk and tank. According to Wikipedia, bulk ships are specially designed to transport unpacked bulk cargo such as grains, coal, ore and cement, while tank ships are designed to carry liquid in bulks. There were several covariates available, and we have chosen to pick from the following.

- $\ln S$, where S is sum insured, i.e. total value of the ship.
- $\ln(A+2)$, where A is the age of the ship.
- $\ln G$, where G is gross tonnage (GRT).
- I, where I is an indicator that equals 1 if the engine is a 4-stroke and 0 if it is a 2-stroke. Only these two kinds of engines are considered.

This gives for the net premium without deductibles in the lognormal case

$$E[N^*]\,E[X^*] = e^{\alpha_1+\beta_1}S^{\alpha_2+\beta_2}(A+2)^{\alpha_3+\beta_3}G^{\alpha_4+\beta_4}e^{\alpha_5 I}e^{\frac{1}{2}\sigma^2},$$

where in practice several of the a_i and β_i are zero. Thus S, $A+2$ and G enter in a multiplicative fashion. Since A is zero for new ships, we chose to add 2 years to all ships. This is of course a bit arbitrary, but there is some experimentation behind it.

TABLE 4.1

SUMMARY OF DATA USED IN THE STUDY. HERE d IS DEDUCTIBLE, S IS SUM INSURED, A IS AGE AND G IS GROSS TONNAGE. SUM INSURED IS IN DOLLARS. SINCE NUMBERS ARE IN TERMS OF POLICY YEARS, THE SAME SHIP CAN BE COUNTED SEVERAL TIMES.

	POLICIES					
Shiptype	Policy years	Claims	min $\frac{d}{S}$	max $\frac{d}{S}$	min d	max d
Bulk	3957	245	0.00071	0.370	17500	2500000
Tank	5773	440	0.00004	0.083	935	2500000
	COVARIATES					
Shiptype	min S	max S	min A	max A	min G	max G
Bulk	1048600	$1.4 \cdot 10^8$	0	49	309	304000
Tank	1250000	$3.6 \cdot 10^8$	0	46	383	261453

The covariate $\ln S$ is always included in the claimsize distribution, and therefore we have not made any further corrections for inflation.

A summary of the data and the covariates is given in Table 4.1. In addition, 250 of the bulk policies were four stroke and 440 of the tankers. From the table we can see that even within the same shiptype data are very heterogeneous.

After some experimentation, assuming claims lognormally distributed, we ended up with the covariates given in Table 4.2.

TABLE 4.2

COVARIATES CHOSEN IN THE ANALYSIS. BY 1 IS MEANT A CONSTANT TERM.

	Claim frequency	**Claimsize**
Bulk	$1 + \ln S + \ln(A+2) + I$	$1 + \ln S$
Tank	$1 + \ln S$	$1 + \ln S + \ln(A+2) + I$

In Table 4.3 estimated parameters are reported together with their estimated standard errors for shiptype bulk. The estimated standard errors are obtained from the Hessian of the loglikelihoods at the maximum points. However, when calculating the standard error for the pseudo-MLE $\tilde{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$, the estimated $\bar{F}_i(d_i; \tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2)$ is used as a fixed offset, i.e. as if $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$ and $\sigma^2 = \tilde{\sigma}^2$ are known, and as pointed out in Remark 2.2 this underestimates the true standard error of $\tilde{\boldsymbol{\alpha}}$. Nevertheless, these underestimated standard errors are included as a comparison.

We see from Table 4.3 that there are some differences in the estimated parameters. The standard errors of the $\tilde{\beta}_i$ are a little higher than those of the $\hat{\beta}_i$, as expected. Surprisingly, the standard error of $\tilde{\sigma}$ is lower than that of $\hat{\sigma}$. This is no contradiction, these are estimated standard deviations, not the true ones. The standard errors of the $\tilde{\alpha}_i$ are always lower than those of the $\hat{\alpha}_i$, but

TABLE 4.3

ESTIMATED PARAMETERS WITH STANDARD ERRORS FOR SHIPTYPE BULK.
AGAIN 1 IN THE COVARIATE ROW MEANS A CONSTANT.

Parameter (Covariate)	α_1 (1)	α_2 ($\ln S$)	α_3 ($\ln(A+2)$)	α_5 (I)	β_1 (1)	β_2 ($\ln S$)	σ
Estimate full MLE	–9.896	0.704	0.369	0.888	9.036	0.194	1.069
St. errors	2.355	0.182	0.129	0.171	2.223	0.135	0.089
Estimate pseudo-MLE	–8.923	0.683	0.313	0.898	7.410	0.294	1.042
St. errors	2.079	0.179	0.111	0.169	2.316	0.139	0.082

as discussed above, since these are calculated using the wrong assumptions, this is to be expected.

It is difficult to read off from the estimated parameters the impact of their differences when put into practical use. Therefore, in Table 4.4 we have calculated net premiums for various methods and models. To better compare these, we have also calculated the percentage differences

$$r_d = 100\,\frac{\hat{m}_d - \check{m}_d}{\hat{m}_d}. \tag{4.2}$$

Here $\hat{m}_d$ is always $m_d(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\sigma}^2)$ from (4.1), using full MLE and the covariates from Table 4.2. Then $\check{m}_d$ estimates the same quantity, but it can be either using pseudo-MLE in (4.1), or using full MLE but a different distribution for the claimsizes, or also using full MLE and lognormal distribution, but with slightly different covariates, see the table caption for details. All calculations are done for an average ship, i.e. $\mathbf{w}$ and $\mathbf{z}$ are the averages of the $\mathbf{w}_i$ and $\mathbf{z}_i$ in their shiptype category. Instead of giving the results as a function of the deductible d, we give them as a function of the relative deductible, i.e. of $r = \frac{d}{S}$.

Table 4.4 basically speaks for itself, and it is clear that pseudo-MLE generally causes less deviation than using either different claimsize distributions or different covariates in the claimsize distributions. Although no examples are given in the table, it can be mentioned that using different covariates in the claim frequency distribution does not matter so much for the net premium calculations. For small and moderate deductibles, the net premiums do not differ by very much. For (unrealistically) high deductibles, there are fairly large differences, and this matters if e.g. the model is used for pricing XL premiums in reinsurance. In Table 4.5 the distribution of the relative deductibles in the data is given, and we see that 98.73% of the bulk ships have a deductible that is 5% or less of sum insured. The two other rows in Table 4.3 is to facilitate the comparison with the tables in Section 3, where instead of $\frac{d}{S}$, $\frac{d}{E^\theta[X^*]} = \frac{S}{E^\theta[X^*]}\frac{d}{S}$ is used. The factor $\frac{S}{E^\theta[X^*]}$ is obtained by dividing average sum insured by estimated claimsize, using the model from the full MLE with covariates from the average ship, giving an estimated factor of 37.1. This procedure, taking the

TABLE 4.4

ESTIMATED NET PREMIUMS AND RELATIVE DIFFERENCES. RELATIVE DIFFERENCES ARE AS CALCULATED IN (4.2), ALWAYS RELATIVE TO FULL MLE WITH LOGNORMAL CLAIMSIZE DISTRIBUTION. THE COVARIATES ARE AS IN TABLE 4.2, BUT IN THE ROW "$\ln G$ included" UNDER BULK, THE ADDITIONAL COVARIATE $\ln G$ IS USED IN THE CLAIMSIZE DISTRIBUTION. FURTHERMORE, IN THE ROW "Only $\ln S$" UNDER TANK, ONLY A CONSTANT TERM AND $\ln S$ ARE USED AS COVARIATES IN THE CLAIMSIZE DISTRIBUTION. IN BOTH THESE CASES FULL MLE WITH LOGNORMAL CLAIMSIZE DISTRIBUTION IS USED.

Rel. ded. $r = \frac{d}{S}$	0	0.01	0.02	0.05	0.10	0.20	0.30	0.50
	Bulk							
Logn. full MLE	55014	37738	27285	12992	5471	1697	733	215
Logn. pseudo-MLE	54610	37079	26438	12128	4872	1414	583	159
Weibull full MLE	57170	38036	27815	13087	4852	1006	273	30
Pareto full MLE	56952	39005	28646	14438	6853	2764	1536	703
$\ln G$ included	55163	38130	27539	12939	5317	1583	663	185
r_d logn. pseudo-MLE	0.7	1.7	3.1	6.7	10.9	16.7	20.5	26.0
r_d Weibull full MLE	–3.9	–0.8	–1.9	–0.7	11.3	40.7	62.7	85.8
r_d Pareto full MLE	–3.5	–3.4	–5.0	–11.1	–25.3	–62.9	–109.5	–227.2
r_d $\ln G$ included	–0.3	–1.0	–0.9	0.4	2.9	7.2	10.6	16.3
	Tank							
Logn. full MLE	34168	24726	17384	6611	1847	290	73	9
Logn. pseudo-MLE	34453	25053	17673	6738	1876	292	73	9
Weibull full MLE	34953	24856	18097	7460	1913	157	15	0
Pareto full MLE	35474	25483	18733	8314	2794	562	174	32
Only $\ln S$	35944	26503	19243	8189	2698	544	166	28
r_d logn. pseudo-MLE	–0.8	–1.3	–1.7	–1.9	–1.6	–0.6	0.5	4.0
r_d Weibull full MLE	–2.3	–0.5	–4.1	–12.8	–3.6	46.0	79.4	98.0
r_d Pareto full MLE	–3.8	–3.1	–7.8	–25.8	–51.3	–93.8	–137.0	–241.3
r_d only $\ln S$	–5.2	–7.2	–10.7	–23.9	–46.1	–87.5	–126.2	–199.9

TABLE 4.5

DISTRIBUTION OF THE RELATIVE DEDUCTIBLES, I.E. DEDUCTIBLE DIVIDED BY SUM INSURED. ESTIMATED $\frac{E^{\theta}[D]}{E^{\theta}[X^*]}$ FOR THE AVERAGE SHIP EQUALS 0.305 FOR BULK AND 0.155 FOR TANK.

Rel. ded. $r = \frac{d}{S}$	0	0.01	0.02	0.05	0.10	0.20	0.30	0.50
	Bulk							
Est. $\frac{d}{E^{\theta}[X^*]}$	0	0.4	0.7	1.9	3.7	7.4	11.1	18.6
Percentage	0	61.71	89.60	98.73	99.93	99.93	99.98	100
	Tank							
Est. $\frac{d}{E^{\theta}[X^*]}$	0	0.3	0.6	1.4	2.9	5.7	8.6	14.3
Percentage	0	62.66	85.55	95.85	99.87	100	100	100

average of all deductibles within the shiptype to estimate $E^\theta[D]$ also gave for the average ship an estimated ratio $\frac{E^\theta[D]}{E^\theta[X^*]}$ equal to 0.305. Looking at the tables in Section 3 with $\frac{d}{E^\theta[X^*]}$ about 1.9 (the 98.73 quantile) and $\frac{E^\theta[D]}{E^\theta[X^*]}$ about 0.305, we see that the loss in efficiency in the i.i.d. case is never very large. This is of course a different model, but the fact that for the Pareto distribution in Example 3.2 the loss in efficiency was typically smaller when two parameters was estimated than just one gives a reason to believe that the tables of Section 3 are not overly optimistic.

5. Concluding remarks

Under the standard assumption that claim numbers and claimsizes are independently distributed, when there are no deductibles or when the deductibles are the same for all policies, maximum likelihood estimation can be separated into claim number estimation and claimsize estimation. This is no longer the case when policies have different deductibles, now maximum likelihood estimation requires that the two distributions have to be estimated jointly. For simple models that is usually not a big problem. For complex models an alternative is to use pseudo-maximum likelihood which here means that first the claimsize distribution is estimated, and then the estimated probability of a damage exceeding the deductible is used as an offset in the claim number estimation. The question is then how big is the loss in efficiency by using this procedure? Here we have focussed on the impact on the estimated net premiums for various deductibles. We have been able to theoretically quantify the relative efficiency when using the pseudo-MLE versus the full MLE for simple i.i.d. models. Numerical examples showed that the loss in efficiency using the pseudo-MLE is rather modest, at least when the deductibles are not exceedingly high, which is typically the case in applications. We also used data from marine insurance to calculate estimated net premiums for the two estimation procedures. This was done for two different shiptypes. Comparing the results, it again turned out that unless the deductibles are exceedingly high, the estimated net premiums did not differ very much. In fact, changing the distribution class for the claimsizes resulted in much larger differences, and the same happened when the distribution class was held fixed, but different covariates to explain the claimsizes were used.

On the basis of our findings we can conclude that a succesful use of historical data to estimate net premiums depend more on the models chosen than on the estimation procedure. Of course, if a full MLE is possible there is no reason not to use it, but the prospect of having to use pseudo-MLE should not be a deterrent against making realistic and useful models.

6. Acknowledgments

We would like to thank an unknown referee for several good suggestions that led to a considerable improvement of the manuscript.

A. Appendix: Proofs

Let $Z_1, \ldots, Z_n$ be i.i.d. with density $m(z;\boldsymbol{\theta},\boldsymbol{\lambda})$ w.r.t. some σ-finite measure. Let the loglikelihood be

$$\check{l}_{(n)}(\boldsymbol{\theta},\boldsymbol{\lambda}) = \sum_{i=1}^{n} \ln m(Z_i;\boldsymbol{\theta},\boldsymbol{\lambda}).$$

Here $(\boldsymbol{\theta},\boldsymbol{\lambda}) \in \Theta \times \Lambda$ where Θ is an open subset of R^s and Λ an open subset of R^t.

Define, if it exists, $\mathcal{I}$ as the $(s+t)\times(s+t)$ symmetric matrix blockdivided as

$$\begin{aligned}
\mathcal{I}_{11} &= -E^{\theta,\lambda}[H_{\theta\theta} \ln m(Z;\boldsymbol{\theta},\boldsymbol{\lambda})],\\
\mathcal{I}_{12} &= -E^{\theta,\lambda}[H_{\theta\lambda} \ln m(Z;\boldsymbol{\theta},\boldsymbol{\lambda})],\\
\mathcal{I}_{22} &= -E^{\theta,\lambda}[H_{\lambda\lambda} \ln m(Z;\boldsymbol{\theta},\boldsymbol{\lambda})].
\end{aligned}$$

The following are multivariate versions of the assumptions made in Gong and Samaniego (1981).

A1. For all $(\boldsymbol{\theta},\boldsymbol{\lambda})$, $\check{l}_{(n)}(\boldsymbol{\theta},\boldsymbol{\lambda})$ is three times continuously differentiable.

A2. Interchange of differentiantion and integration of m is allowed up to the second derivatives.

A3. $\mathcal{I}$ exists and $\mathcal{I}_{22}$ is positive definite.

A4. For all $(\boldsymbol{\theta},\boldsymbol{\lambda})$ there are open sets U and V with $(\boldsymbol{\theta},\boldsymbol{\lambda}) \in U \times V$ so that for all $(\boldsymbol{\theta}_1,\boldsymbol{\lambda}_1) \in U \times V$,

$$\left| \nabla_\theta \ln \frac{m(z;\boldsymbol{\theta}_1,\boldsymbol{\lambda}_1)}{m(z;\boldsymbol{\theta}_1,\boldsymbol{\lambda})} \right| < M(z;\boldsymbol{\lambda}_1),$$

where $E^{\theta,\lambda}[M(Z;\boldsymbol{\lambda}_1)] < \infty$.

A5. For all $(\boldsymbol{\theta},\boldsymbol{\lambda})$ there are open sets U and V with $(\boldsymbol{\theta},\boldsymbol{\lambda}) \in U \times V$ so that for all $(\boldsymbol{\theta}_1,\boldsymbol{\lambda}_1) \in U \times V$,

$$|H_{\theta\theta\lambda} \ln m(z;\boldsymbol{\theta},\boldsymbol{\lambda}_1)| + |H_{\theta\lambda\lambda} \ln m(z;\boldsymbol{\theta},\boldsymbol{\lambda}_1)| + |H_{\lambda\lambda\lambda} \ln m(z;\boldsymbol{\theta}_1,\boldsymbol{\lambda}_1)| < M(z),$$

where $E^{\theta,\lambda}[M(Z)] < \infty$

A6. For any $(\boldsymbol{\theta}_1,\boldsymbol{\lambda}_1) \neq (\boldsymbol{\theta},\boldsymbol{\lambda})$, $P^{\theta,\lambda}(m(Z;\boldsymbol{\theta},\boldsymbol{\lambda}) = m(Z;\boldsymbol{\theta}_1,\boldsymbol{\lambda}_1)) < 1$.

$$P^{\theta,\lambda}(m(Z;\boldsymbol{\theta},\boldsymbol{\lambda}) = m(Z;\boldsymbol{\theta}_1,\boldsymbol{\lambda}_1)) < 1.$$

The following result is a multivariate generalization of Theorem 2.2 in Gong and Samaniego (1981). The proof is basically the same as that in Gong and Samaniego and is omitted. That $\tilde{\Sigma}_{12} = 0$ is proved in Parke (1986).

Theorem A.1. *Let $Z_1, \ldots, Z_n$ be i.i.d. as above and let $\tilde{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n(Z_1, \ldots, Z_n)$ be such that $(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = O_p(n^{-\frac{1}{2}})$. Assume that the equation $\nabla_\lambda \check{l}_{(n)}(\tilde{\boldsymbol{\theta}}_n, \boldsymbol{\lambda}) = 0$ has a unique solution $\tilde{\boldsymbol{\lambda}}_n$ for all n. Also suppose that as $n \to \infty$,*

$$\sqrt{n}\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta},\ \frac{1}{n}\nabla_\lambda \check{l}_{(n)}(\boldsymbol{\theta},\boldsymbol{\lambda})) \xrightarrow{d} \mathcal{N}(0,\tilde{\Sigma}) \quad under\ P^{\theta,\lambda}$$

and that

$$\frac{1}{\sqrt{n}}\nabla_{\theta,\lambda}\check{l}_{(n)}(\boldsymbol{\theta},\boldsymbol{\lambda}) \xrightarrow{d} \mathcal{N}(0,\mathcal{I}) \quad under\ P^{\theta,\lambda}.$$

Let the $(s+t)\times(s+t)$ *matrix* $\tilde{\Sigma}$ *be blockdivided similar to* $\mathcal{I}$. *Then* $\tilde{\Sigma}_{12} = 0$ *and furthermore as* $n \to \infty$,

$$\sqrt{n}\,((\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\lambda}}_n) - (\boldsymbol{\theta},\boldsymbol{\lambda})) \xrightarrow{d} \mathcal{N}(0,\Sigma) \quad under\ P^{\theta,\lambda},$$

where Σ *is the* $(s+t)\times(s+t)$ *symmetric matrix blockdivided as*

$$\begin{aligned}\Sigma_{11} &= \tilde{\Sigma}_{11},\\ \Sigma_{12} &= -\tilde{\Sigma}_{11}\mathcal{I}_{12}\mathcal{I}_{22}^{-1},\\ \Sigma_{22} &= \mathcal{I}_{22}^{-1} + \mathcal{I}_{22}^{-1}\mathcal{I}_{21}\tilde{\Sigma}_{11}\mathcal{I}_{12}\mathcal{I}_{22}^{-1}.\end{aligned}$$

It is of course not necessary that the Z_i are i.i.d. in order for a result such as Theorem A.1 to hold. It could easily be extended to e.g. a regression-like model, but at the expense of more conditions. In Pierce (1982) results similar to Theorem A.1, but less specific, are given for non i.i.d. data. We think the i.i.d. assumption strikes a reasonable balance between simplicity and illumination of the differences in the estimation methods.

From here on, all relevant definitions are given in Section 2.

Lemma A.1. *Assume B3 and B4. Then*

$$E^{\theta}[G(X^*)] = A, \tag{A.1}$$

$$E^{\theta,\lambda}\left[\sum_{j=1}^{N}\nabla_\theta \ln f(X_j;\boldsymbol{\theta})\right] = -\mu(\boldsymbol{\lambda})\mathbf{B}, \tag{A.2}$$

$$E^{\theta,\lambda}\left[\sum_{j=1}^{N}(\nabla_\theta \ln f(X_j;\boldsymbol{\theta}))(\nabla_\theta \ln f(X_j;\boldsymbol{\theta}))^T\right] = \mu(\boldsymbol{\lambda})\mathbf{J}, \tag{A.3}$$

$$E^{\theta,\lambda}\left[\sum_{j=1}^{N}H_{\theta\theta}\ln f(X_j;\boldsymbol{\theta})\right] = \mu(\boldsymbol{\lambda})E^{\theta}[G(X^*)H_{\theta\theta}\ln f(X^*;\boldsymbol{\theta})] = \mu(\boldsymbol{\lambda})(H_{\theta\theta}A - \mathbf{J}), \tag{A.4}$$

$$E^{\theta,\lambda}[NH_{\theta\theta}\ln\bar{F}(D;\boldsymbol{\theta})] = \mu(\boldsymbol{\lambda})E^{\theta}[\bar{F}(D;\boldsymbol{\theta})H_{\theta\theta}\ln\bar{F}(D;\boldsymbol{\theta})] = \mu(\boldsymbol{\lambda})(H_{\theta\theta}A - \mathbf{C}). \tag{A.5}$$

In particular from (A.4) and (A.5),

$$\mathbf{J} = H_{\theta\theta} A - E^{\theta}[G(X^*) H_{\theta\theta} \ln f(X^*; \boldsymbol{\theta})]$$

and

$$\mathbf{J} - \mathbf{C} = E^{\theta}[\bar{F}(D; \boldsymbol{\theta}) H_{\theta\theta} \ln \bar{F}(D; \boldsymbol{\theta})] - E^{\theta}[G(X^*) H_{\theta\theta} \ln f(X^*; \boldsymbol{\theta})].$$

Proof. Equation (A.1) is just a simple application of Fubinis theorem. For (A.2) we have

$$\begin{aligned} E^{\theta,\lambda}\left[\sum_{j=1}^{N} \nabla_\theta \ln f(X_j; \boldsymbol{\theta})\right] &= E^{\theta,\lambda}\left[E^{\theta,\lambda}\left[\sum_{j=1}^{N} \nabla_\theta \ln f(X_j; \boldsymbol{\theta})\middle| D\right]\right] \\ &= E^{\theta}\left[\mu(\boldsymbol{\lambda}) \bar{F}(D; \boldsymbol{\theta}) E^{\theta}[\nabla_\theta \ln f(X; \boldsymbol{\theta}) \mid D]\right] \\ &= E^{\theta}\left[\mu(\boldsymbol{\lambda}) \bar{F}(D; \boldsymbol{\theta}) \int_D^\infty \frac{1}{f(x; \boldsymbol{\theta})} (\nabla_\theta f(x; \boldsymbol{\theta})) \frac{f(x; \boldsymbol{\theta})}{\bar{F}(D; \boldsymbol{\theta})} dv(x)\right] \\ &= \mu(\boldsymbol{\lambda}) E^{\theta}[\nabla_\theta \bar{F}(D; \boldsymbol{\theta})] = -\mu(\boldsymbol{\lambda}) \mathbf{B}. \end{aligned}$$

Similarly for (A.3)

$$\begin{aligned} & E^{\theta,\lambda}\left[\sum_{j=1}^{N} (\nabla_\theta \ln f(X_j; \boldsymbol{\theta}))\, (\nabla_\theta \ln f(X_j; \boldsymbol{\theta}))^T\right] \\ &= E^{\theta}\left[\mu(\boldsymbol{\lambda}) \bar{F}(D; \boldsymbol{\theta}) \int_D^\infty (\nabla_\theta \ln f(x; \boldsymbol{\theta}))\, (\nabla_\theta \ln f(x; \boldsymbol{\theta}))^T \frac{f(x; \boldsymbol{\theta})}{\bar{F}(D; \boldsymbol{\theta})} dv(x)\right] \\ &= \mu(\boldsymbol{\lambda}) \int_0^\infty \int_d^\infty (\nabla_\theta \ln f(x; \boldsymbol{\theta}))(\nabla_\theta \ln f(x; \boldsymbol{\theta}))^T f(x; \boldsymbol{\theta})\, dv(x)\, dG(d) \\ &= \mu(\boldsymbol{\lambda}) \int_0^\infty G(x)\, (\nabla_\theta \ln f(x; \boldsymbol{\theta}))\, (\nabla_\theta \ln f(x; \boldsymbol{\theta}))^T f(x; \boldsymbol{\theta})\, dv(x) = \mu(\boldsymbol{\lambda}) \mathbf{J}. \end{aligned}$$

Furthermore

$$\begin{aligned} & E^{\theta,\lambda}\left[\sum_{j=1}^{N} H_{\theta\theta} \ln f(X_j; \boldsymbol{\theta})\right] \\ &= E^{\theta,\lambda}\left[\sum_{j=1}^{N} \frac{1}{f(X_j; \boldsymbol{\theta})} H_{\theta\theta} f(X_j; \boldsymbol{\theta}) - (\nabla_\theta \ln f(X_j; \boldsymbol{\theta}))\, (\nabla_\theta \ln f(X_j; \boldsymbol{\theta}))^T\right] \\ &= E^{\theta}\left[\mu(\boldsymbol{\lambda}) \bar{F}(D; \boldsymbol{\theta}) \int_D^\infty \frac{1}{f(x; \boldsymbol{\theta})} H_{\theta\theta} f(x; \boldsymbol{\theta}) \frac{f(x; \boldsymbol{\theta})}{\bar{F}(D; \boldsymbol{\theta})} dv(x)\right] - \mu(\boldsymbol{\lambda}) \mathbf{J} \\ &= \mu(\boldsymbol{\lambda}) E^{\theta} [H_{\theta\theta} \bar{F}(D; \boldsymbol{\theta})] - \mu(\boldsymbol{\lambda}) \mathbf{J} \\ &= \mu(\boldsymbol{\lambda}) H_{\theta\theta} A - \mu(\boldsymbol{\lambda}) \mathbf{J}, \end{aligned}$$

and this proves the equality between the left and the right terms in (A.4). For equality between the middle and right terms we use (A.1), the above arguments and the fact that $H_{\theta\theta} \ln g(\boldsymbol{\theta}) = \frac{1}{g(\boldsymbol{\theta})} H_{\theta\theta} g(\boldsymbol{\theta}) - (\nabla_\theta \ln g(\boldsymbol{\theta}))(\nabla_\theta \ln g(\boldsymbol{\theta}))^T$. Finallly, (A.5) follows by similar arguments.

Lemma A.2. *Assume B1-B4. Then as $n \to \infty$,*

$$\frac{1}{\sqrt{n}} \nabla_{\theta,\lambda} l_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}) \ \text{under } P^{\theta,\lambda},$$

where $\mathcal{I}$ is the $(s+t) \times (s+t)$ matrix blockdivided as

$$\mathcal{I}_{11} = \mathbf{K}_{11} + \mu(\boldsymbol{\lambda})(\mathbf{J} - \mathbf{C}), \quad \mathcal{I}_{12} = \mathbf{K}_{12} \ \text{ and } \ \mathcal{I}_{22} = \mathbf{K}_{22}.$$

Proof. By (2.1),

$$\frac{1}{\sqrt{n}} \nabla_{\theta,\lambda} l_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i,$$

where

$$\mathbf{Y}_i = \begin{bmatrix} \nabla_\theta \ln p(N_i; (\boldsymbol{\theta}, \boldsymbol{\lambda}), D_i) + \sum_{j=1}^{N_i} \nabla_\theta \ln f(X_{ij}; \boldsymbol{\theta}) - N_i \nabla_\theta \bar{F}(D_i; \boldsymbol{\theta}) \\ \nabla_\lambda \ln p(N_i; (\boldsymbol{\theta}, \boldsymbol{\lambda}), D_i) \end{bmatrix}.$$

Since the $\mathbf{Y}_i$ are i.i.d. loglikelihood derivatives, the result follows by differentating once more w.r.t. both $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ and taking expectations, using (A.3) and (A.4).

Lemma A.3. *Assume B1-B10 and that the equation $\nabla_\theta \tilde{l}_{(n)}(\boldsymbol{\theta}) = 0$ has a unique solution for all n. Then as $n \to \infty$,*

$$\sqrt{n}\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}, \frac{1}{n} \nabla_\lambda l_{(n)}(\boldsymbol{\theta}, \boldsymbol{\lambda})) \xrightarrow{d} \mathcal{N}(0, \tilde{\Sigma}) \ \text{under } P^{\theta,\lambda},$$

where

$$\tilde{\Sigma}_{11} = \frac{1}{\mu(\boldsymbol{\lambda})} (\mathbf{J} - \mathbf{C})^{-1}, \quad \tilde{\Sigma}_{12} = 0 \quad \text{and} \quad \tilde{\Sigma}_{22} = \mathbf{K}_{22}.$$

Proof. It follows by uniqueness and Theorem 5.42 in van der Vaart (1998) that $\tilde{\boldsymbol{\theta}}_n$ is consistent. Using Taylor's formula and the definition of $\tilde{\boldsymbol{\theta}}_n$,

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\theta \tilde{l}_i(\tilde{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\theta \tilde{l}_i(\boldsymbol{\theta}) + \left(\frac{1}{n} \sum_{i=1}^{n} H_{\theta\theta} \tilde{l}_i(\boldsymbol{\theta}) \right) \sqrt{n}\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$$
$$+ \frac{1}{2} (\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^T \left(\frac{1}{n} \sum_{i=1}^{n} H_{\theta\theta} (\nabla_\theta \tilde{l}_i(\check{\boldsymbol{\theta}}_n)) \right) \sqrt{n}\,(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}),$$

where $\check{\boldsymbol{\theta}}_n$ lies between $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}_n$. Here, for a three times differentiable function $k(\boldsymbol{\theta})$, $H_{\theta\theta}(\nabla_\theta k(\boldsymbol{\theta}))$ is an $s \times s$ matrix where the i'th row equals $H_{\theta\theta}(\frac{\partial}{\partial\theta_i} k(\boldsymbol{\theta}))$. From (2.2), (A.4) and (A.5) and the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} H_{\theta\theta}\tilde{l}_i(\boldsymbol{\theta}) \overset{\text{a.s.}}{\to} E^{\theta}[H_{\theta\theta}\tilde{l}_i(\boldsymbol{\theta})] = -\mu(\boldsymbol{\lambda})(\mathbf{J}-\mathbf{C}).$$

Therefore, arguing as in e.g. Ch. 4.2.2 in Serfling(1980), using Lemma A.1 we find that w.r.t. $P^{\theta,\lambda}$,

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) = -(\mu(\boldsymbol{\lambda})(\mathbf{J}-\mathbf{C}) + o(1))^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_\theta\tilde{l}_i(\boldsymbol{\theta}).$$

This gives

$$\sqrt{n}\begin{bmatrix}\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta} \\ \frac{1}{n}\nabla_\lambda l_{(n)}(\boldsymbol{\theta},\boldsymbol{\lambda})\end{bmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{Y}_i + o_p(1),$$

where

$$\mathbf{Y}_i = \begin{bmatrix}\mathbf{Y}_{1i} \\ \mathbf{Y}_{2i}\end{bmatrix} = \begin{bmatrix}-\frac{1}{\mu(\boldsymbol{\lambda})}(\mathbf{J}-\mathbf{C})^{-1}\left(\sum_{i=1}^{N_i}\nabla_\theta \ln f(X_{ij};\boldsymbol{\theta}) - N_i\nabla_\theta \ln\bar{F}(D_i;\boldsymbol{\theta})\right) \\ \nabla_\lambda \ln p(N_i;(\boldsymbol{\theta},\boldsymbol{\lambda}),D_i)\end{bmatrix}.$$

Since the $\mathbf{Y}_i$ are i.i.d. asymptotic normality follows. The expressions for $\tilde{\Sigma}_{11}$ and $\tilde{\Sigma}_{22}$ are also straightforward, and that $\tilde{\Sigma}_{12} = 0$ follows from Theorem A.1.

Proof of Theorem 2.1

By uniqueness of solutions of the likelihood equations, the result for $(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\lambda}}_n)$ follows easily from Lemma A.2 and Theorems 5.41 and 5.42 in van der Vaart (1998). The form of $\Sigma^M = \mathcal{I}^{-1}$ is a consequence of a small variation of the formula given in Exercise 2.7 p. 33 in Rao (1973).

That $\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = O_p(n^{-\frac{1}{2}})$ follows again from Theorems 5.41 and 5.42 in van der Vaart (1998). The rest of the result for $(\tilde{\boldsymbol{\theta}}_n, \tilde{\boldsymbol{\lambda}}_n)$ is then a consequence of Theorem A.1 and Lemmas A.2 and A.3.

Proof of Lemma 3.1

The covariance matrix of $\mathbf{U}$ is

$$\text{Cov}^{\theta}(\mathbf{U},\mathbf{U}) = E^{\theta}[\mathbf{U}\mathbf{U}^T] - E^{\theta}[\mathbf{U}]E^{\theta}[\mathbf{U}]^T = \frac{1}{A}(\mathbf{C}-\mathbf{C}')$$

and this proves the first part. Since $E^{\theta}[(\boldsymbol{\alpha}^T\mathbf{U})^2] = E^{\theta}[\boldsymbol{\alpha}^T\mathbf{U}]^2$ if and only if $\boldsymbol{\alpha}^T\mathbf{U}$ is a constant, the rest of the result follows.

Proof of Corollary 3.1

We only prove the result for ARE(θ) as the two others are similar, but more tedious. Note that we have

$$E^{\theta}[U] = \frac{B}{A}, \quad E^{\theta}[U^2] = \frac{C}{A} \quad \text{and therefore} \quad \mathrm{Var}^{\theta}[U] = \frac{1}{A}\left(C - \frac{B^2}{A}\right).$$

Then

$$1 - \mathrm{ARE}(\theta) = 1 - \frac{\frac{1}{\lambda}\left(J - \frac{1}{A}B^2\right)^{-1}}{\frac{1}{\lambda}(J - C)^{-1}} = \frac{C - \frac{1}{A}B^2}{J - \frac{1}{A}B^2} = \frac{\mathrm{Var}^{\theta}[U]}{\frac{J}{A} - E^{\theta}[U]^2}.$$

References

Dimakos, X.K. and di Rattalma, A.F. (2002) Bayesian premium rating with latent structure. *Scandinavian Actuarial Journal*, 162-184.

Frees, E.W., Shi, P. and Valdez, E.A. (2009) Actuarial applications of a hierarchical insurance claims model. *ASTIN Bulletin*, **39**, 165-198.

Garrido, J. and Zhou, J. (2009) Full credibility with generalized linear and mixed models. *ASTIN Bulletin*, **39**, 61-80.

Gong, G. and Samaniego, F.J. (1981) Pseudo maximum likelihood estimation: Theory and applications. *Annals of Statistics*, **9**, 861-869.

Gschlössl, S. and Czado, C. (2007) Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal*, 202-225.

Joe, H. (2005) Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, **94**, 401-419.

Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2004) *Loss models. From data to decisions.* Second Edition, Wiley.

Paulsen, J., Lunde, A. and Skaug, H. (2008) Fitting mixed-effects models when data are left truncated. *Insurance: Mathematics and Economics*, **43**, 121-133.

Parke, W.R. (1986) Pseudo maximum likelihood estimation: The asymptotic distribution. *Annals of Statistics*, **14**, 355-357.

Pierce, D.A. (1982) The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Annals of Statistics*, **10**, 475-478.

Rao, C.R. (1973) *Linear Statistical Inference and its Applications*, 2nd edn., Wiley, New York.

Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

van der Vaart, A.W. (1998) *Asymptotic Statistics*, Cambridge University Press.

Jostein Paulsen
Department of Mathematical Sciences
Universitetsparken 5
University of Copenhagen
DK-2100 Copenhagen
Denmark

Knut Stubø
Kr Olavs Vei 50
2819 Gjøvik
Norway