Extreme value estimation of the conditional risk premium in reinsurance

Goegebeur, Yuri; Guillou, Armelle; Qin, Jing

Published in:
Insurance: Mathematics and Economics

DOI:
10.1016/j.insmatheco.2020.10.010

Publication date:
2021

Document version:
Submitted manuscript

Citation for pulished version (APA):
Goegebeur, Y., Guillou, A., & Qin, J. (2021). Extreme value estimation of the conditional risk premium in reinsurance. *Insurance: Mathematics and Economics*, *96*, 68-80.
https://doi.org/10.1016/j.insmatheco.2020.10.010

Extreme value estimation of the conditional risk premium in reinsurance

Yuri Goegebeur[(1)], Armelle Guillou[(2)], Jing Qin[(1)]

[(1)] Department of Mathematics and Computer Science, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
[(2)] Institut Recherche Mathématique Avancée, UMR 7501, Université de Strasbourg et CNRS, 7 rue René Descartes, 67084 Strasbourg cedex, France

Abstract

In the paper we study the estimation of reinsurance premiums when the claim size is observed together with additional information in the form of random covariates. Using extreme value arguments, we propose an estimator for the risk premium conditional on a value for the covariate, and derive its asymptotic properties, after suitable normalization. The finite sample behavior is evaluated with a simulation experiment, and we apply the methodology to a dataset of automobile insurance claims from Australia.
Keywords: Pareto-type distribution, tail index, reinsurance premium, risk.

1 Introduction

In reinsurance, a popular premium calculation method is the net premium principle. If the claim amount or the loss of a reinsurance policy is modelled as a non-negative random variable Y then the reinsurance premium Π at retention level R is defined as

$$\Pi(R) = \mathbb{E}[(Y - R)_+],$$

where $x_+ := \max\{0, x\}$. By standard calculations the above equation can be expressed directly in terms of the distribution function F of Y as

$$\Pi(R) = \int_R^\infty (1 - F(y))dy. \tag{1}$$

Under this premium principle, the reinsurer charges as premium the expected payment, and hence in theory the reinsurer would on average not loose money. In practice, reinsurers must adjust for, among others, the potential risk in fluctuations of actually experienced losses from their expected values and hence they will charge a premium that is at least $\mathbb{E}[(Y - R)_+]$. We refer to Albrecher et al. (2017) for a comprehensive discussion of premium principles in reinsurance.

The net premium principle in (1) can be generalized by introducing a distortion function g to

$$\Pi(R) = \int_R^\infty g\left(1 - F(y)\right) dy, \tag{2}$$

where g is an increasing, concave function that maps $[0, 1]$ onto $[0, 1]$, see, e.g., Wang (1996). In a reinsurance setting, focus is often on extreme events, that is, events with a low frequency of occurrence but a high, often catastrophic, impact. As such, the retention level R is typically large, near the largest observed claim amounts, or even outside the observed data range, which calls for an accurate estimation of the upper tail of the loss or claim size distribution. Extreme value statistics offers the natural toolbox required for this type of tail analysis. We refer to Beirlant et al. (2004) and de Haan and Ferreira (2006).

In this paper, we will consider the estimation of the conditional risk premium when the random variable of main interest Y is recorded together with a random covariate $X \in \mathbb{R}^d$. We will denote by $F(\cdot|x)$ the continuous conditional distribution function of Y, given $X = x$, and use the notation $\overline{F}(\cdot|x)$ for the conditional survival function and $U(\cdot|x)$ for the associated tail quantile function defined as $U(\cdot|x) = \inf\{y : F(y|x) \geqslant 1 - 1/\cdot\}$. Also, we will denote by f_X the density function of the covariate X and by x_0 a reference position such that $x_0 \in \text{Int}(S_X)$, the interior of the support $S_X \subset \mathbb{R}^d$ of f_X, which is assumed to be non-empty. Our aim will be to estimate the conditional risk premium, given $X = x_0$, and defined as

$$\Pi(R_n|x_0) = \int_{R_n}^\infty g\left(\overline{F}(y|x_0)\right) dy, \tag{3}$$

where R_n is a non-random value which tends to ∞ as $n \to \infty$. Taking covariate information into account allows reinsurers to differentiate the risks they are exposed to according to the value of the covariate, which leads to more accurate premium determination.

The estimation of reinsurance premiums under the net premium principle (1) using univariate extreme value methods was studied in Beirlant et al. (2001). Vandewalle and Beirlant (2006) generalized Beirlant et al. (2001), by considering the estimation of (2), along with deriving the asymptotic properties of the proposed estimator under a second order extreme value framework. In El Methni et al. (2014) the nonparametric estimation of extreme risk measures in a heavy-tailed context with random covariates was studied, among others they proposed an estimator for (1) in presence of random covariates. See also El Methni et al. (2018) for a study of conditional risk measures in the general max-domain of attraction.

The paper is organized as follows. In the next section we introduce our estimator for (3) and study its asymptotic properties in the framework of conditional heavy-tailed losses. The finite sample behavior of the estimator is studied in Section 3 by a simulation experiment, and in Section 4 we illustrate the methodology on a dataset of automobile claims data from Australia. All the proofs are postponed to the Appendix.

2 Estimator and asymptotic properties

In this section we will introduce the estimator for $\Pi(R_n|x_0)$ and study its asymptotic properties.

In the sequel, the function g is assumed to be regularly varying at zero with index $-\rho$, $\rho < 0$, i.e., $g(y) = y^{-\rho}\ell_g(y)$, where $\ell_g(y) = C\{1 + \Delta(y)\}$ with $C > 0$ and $\Delta(\cdot)$ is ultimately of constant sign and $|\Delta(\cdot)|$ is regularly varying with index $-\rho^*$, $\rho^* < 0$, i.e., $|\Delta(y)| = y^{-\rho^*}\ell_\Delta(y)$, ℓ_Δ being a slowly varying function. A function ℓ is said to be slowly varying at infinity if $\lim_{t\to\infty} \ell(ty)/\ell(t) = 1$, $\forall y > 0$. The class of regularly varying functions with index α will be denoted as RV_α.

Additionally, we assume that Y follows a conditional Pareto-type model.

Assumption ($\mathcal{D}$) *For all $x \in S_X$, the conditional survival function of Y given $X = x$, satisfies*

$$\overline{F}(y|x) \quad = \quad A(x)y^{-1/\gamma(x)}\left(1 + \frac{1}{\gamma(x)}\delta(y|x)\right),$$

where $A(x) > 0$, $\gamma(x) > 0$, and $|\delta(.|x)|$ is normalized regularly varying at infinity with index $-\beta(x)$, $\beta(x) > 0$, i.e.,

$$\delta(y|x) \quad = \quad B(x)\exp\left(\int_1^y \frac{\varepsilon(u|x)}{u}du\right),$$

with $B(x) \in \mathbb{R}$ and $\varepsilon(y|x) \to -\beta(x)$ as $y \to \infty$. Moreover, we assume $y \to \varepsilon(y|x)$ to be a continuous function.

Clearly, Assumption ($\mathcal{D}$) implies that $U(\cdot|x)$ satisfies

$$U(y|x) = [A(x)]^{\gamma(x)}\, y^{\gamma(x)}\,(1 + a(y|x))$$

where $a(y|x) = \delta(U(y|x)|x)(1 + o(1))$, and thus $|a(.|x)| \in RV_{-\beta(x)\gamma(x)}$.

Note that under our assumptions, g and $\overline{F}(.|x)$ both satisfy the commonly used second order condition from extreme value theory, see, e.g., Theorem 2.3.9 in de Haan and Ferreira (2006). A second order condition on the tail behavior is typically needed for obtaining the limiting distribution of estimators for tail parameters.

Remark that we have the following decomposition

$$\begin{aligned}\Pi(R_n|x_0) &= \int_{R_n}^{\infty} y^{\frac{\rho}{\gamma(x_0)}}\ell(y|x_0)dy \\ &= \ell(R_n|x_0)\int_{R_n}^{\infty} y^{\frac{\rho}{\gamma(x_0)}}dy + \ell(R_n|x_0)\int_{R_n}^{\infty} y^{\frac{\rho}{\gamma(x_0)}}\left[\frac{\ell(y|x_0)}{\ell(R_n|x_0)} - 1\right]dy\end{aligned}$$

where $\ell(y|x_0) := [A(x_0)]^{-\rho}\left(1 + \frac{1}{\gamma(x_0)}\delta(y|x_0)\right)^{-\rho}\ell_g(\overline{F}(y|x_0))$ is a slowly varying function at infinity. Now, assuming $\gamma(x_0) < -\rho$, according to Proposition B.1.10 in de Haan and Ferreira

(2006), for any $\varepsilon > 0$ and $\zeta \in (0, -1-\rho/\gamma(x_0))$, we have for n large and for some constant $L > 0$

$$\begin{aligned}\ell(R_n|x_0)\left|\int_{R_n}^{\infty} y^{\frac{\rho}{\gamma(x_0)}}\left[\frac{\ell(y|x_0)}{\ell(R_n|x_0)}-1\right]dy\right| &\leqslant \varepsilon\ell(R_n|x_0)\int_{R_n}^{\infty} y^{\frac{\rho}{\gamma(x_0)}}\left(\frac{y}{R_n}\right)^{\zeta}dy \\ &\leqslant L\varepsilon\,\ell(R_n|x_0)R_n^{1+\frac{\rho}{\gamma(x_0)}},\end{aligned}$$

and hence,

$$\ell(R_n|x_0)\int_{R_n}^{\infty} y^{\frac{\rho}{\gamma(x_0)}}\left[\frac{\ell(y|x_0)}{\ell(R_n|x_0)}-1\right]dy = o\left(R_n g\left(\overline{F}(R_n|x_0)\right)\right), \tag{4}$$

as $n \to \infty$. From this, we deduce that, for $n \to \infty$,

$$\Pi(R_n|x_0) \sim -\frac{\gamma(x_0)}{\gamma(x_0)+\rho}R_n\, g\left(\overline{F}(R_n|x_0)\right) =: \widetilde{\Pi}(R_n|x_0). \tag{5}$$

Hence, to estimate $\widetilde{\Pi}(R_n|x_0)$, we clearly need to estimate $\gamma(x_0)$ and $\overline{F}(R_n|x_0)$.

Let $(Y_i, X_i), i = 1, \ldots, n$, be independent copies of (Y, X). Concerning $\gamma(x_0)$, we use the following estimator, studied in Goegebeur et al. (2020b),

$$\widehat{\gamma}(x_0) := \frac{\frac{1}{k}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)\left(\ln Y_i - \ln \widehat{U}(n/k|x_0)\right)1\!\!1_{\{Y_i \geqslant \widehat{U}(n/k|x_0)\}}}{\widehat{f}_n(x_0)} \tag{6}$$

where $K_{h_n}(.) := K(./h_n)/h_n^d$, with K a joint density function on $\mathbb{R}^d$, k an intermediate sequence such that $k \to \infty$ with $k/n \to 0$, h_n a positive non-random sequence of bandwidths with $h_n \to 0$ if $n \to \infty$, $1\!\!1_A$ the indicator function on the event A, and $\widehat{f}_n(x_0) := 1/n\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)$ is a classical kernel density estimator. Here $\widehat{U}(.|x_0)$ is an estimator for $U(.|x_0)$, defined as $\widehat{U}(.|x_0) := \inf\{y : \widehat{F}_n(y|x_0) \geqslant 1 - 1/.\}$ where

$$\widehat{F}_n(y|x_0) := \frac{\frac{1}{n}\sum_{i=1}^{n} K_{h_n}(x_0 - X_i)1\!\!1_{\{Y_i \leqslant y\}}}{\widehat{f}_n(x_0)}.$$

This estimator for $\gamma(x_0)$ can be seen as a local version of the well-known Hill estimator (Hill, 1975), initially developed for the univariate context.

Concerning $\overline{F}(R_n|x_0)$, we first remark that for a high threshold t such that $R_n > t$, we have, under assumption $(\mathcal{D})$,

$$\overline{F}(R_n|x_0) = \overline{F}(t|x_0)\frac{\overline{F}(R_n|x_0)}{\overline{F}(t|x_0)} \simeq \overline{F}(t|x_0)\left(\frac{R_n}{t}\right)^{-\frac{1}{\gamma(x_0)}}. \tag{7}$$

Let $t := U(n/k|x_0)$, and estimate $\overline{F}(R_n|x_0)$ by

$$\widehat{\overline{F}}(R_n|x_0) := \frac{k}{n}\left(\frac{R_n}{\widehat{U}(n/k|x_0)}\right)^{-\frac{1}{\widehat{\gamma}(x_0)}}.$$

Combining (5) with (6) and (7), we propose to estimate $\Pi(R_n|x_0)$ by

$$\widehat{\Pi}(R_n|x_0) := -\frac{\widehat{\gamma}(x_0)}{\widehat{\gamma}(x_0)+\rho} R_n\, g\left(\frac{k}{n}\left(\frac{R_n}{\widehat{U}(n/k|x_0)}\right)^{-\frac{1}{\widehat{\gamma}(x_0)}}\right).$$

We can decompose

$$\frac{\widehat{\Pi}(R_n|x_0)}{\Pi(R_n|x_0)} - 1 = \left\{\frac{\widehat{\Pi}(R_n|x_0)}{\widetilde{\Pi}(R_n|x_0)} - 1\right\} + \left\{\left[\frac{\widetilde{\Pi}(R_n|x_0)}{\Pi(R_n|x_0)} - 1\right]\frac{\widehat{\Pi}(R_n|x_0)}{\widetilde{\Pi}(R_n|x_0)}\right\}. \tag{8}$$

To study the two terms in the right-hand side of (8), we need some assumptions due to the regression context. In particular, f_X and the functions appearing in $\overline{F}(y|x)$ are assumed to satisfy the following Hölder conditions. Let $\|.\|$ denote some norm on $\mathbb{R}^d$.

Assumption $(\mathcal{H})$ *There exist positive constants M_{f_X}, M_A, M_γ, M_B, M_ε, η_{f_X}, η_A, η_γ, η_B and η_ε, such that for all $x, z \in S_X$:*

$$\begin{aligned}
|f_X(x) - f_X(z)| &\leqslant M_{f_X}\|x - z\|^{\eta_{f_X}},\\
|A(x) - A(z)| &\leqslant M_A\|x - z\|^{\eta_A},\\
|\gamma(x) - \gamma(z)| &\leqslant M_\gamma\|x - z\|^{\eta_\gamma},\\
|B(x) - B(z)| &\leqslant M_B\|x - z\|^{\eta_B},\\
\sup_{y\geqslant 1}|\varepsilon(y|x) - \varepsilon(y|z)| &\leqslant M_\varepsilon\|x - z\|^{\eta_\varepsilon}.
\end{aligned}$$

We also impose a condition on the kernel function K, which is a standard condition in local estimation.

Assumption $(\mathcal{K})$ *K is a bounded density function on $\mathbb{R}^d$, with support S_K included in the unit ball in $\mathbb{R}^d$.*

As a first result, we make the error explicit when $\widetilde{\Pi}(R_n|x_0)$ is used as approximation to $\Pi(R_n|x_0)$.

Theorem 2.1 *Assume $(\mathcal{D})$. If $\gamma(x_0) < -\rho$, then, we have for $n \to \infty$*

$$\begin{aligned}
\frac{\Pi(R_n|x_0)}{\widetilde{\Pi}(R_n|x_0)} - 1 &= -\frac{\rho\beta(x_0)}{\gamma(x_0)+\rho-\gamma(x_0)\beta(x_0)}\delta(R_n|x_0) - \frac{\rho^*}{\gamma(x_0)+\rho+\rho^*}\Delta(\overline{F}(R_n|x_0))\\
&\quad + o(\delta(R_n|x_0)) + o(\Delta(\overline{F}(R_n|x_0))).
\end{aligned}$$

As mentioned above, our estimator for $\Pi(R_n|x_0)$ depends on estimators for $\gamma(x_0)$ and $\overline{F}(R_n|x_0)$. The asymptotic properties of these will be given the next two theorems.

The weak convergence of $\widehat{\gamma}(x_0)$, after proper normalization, was obtained by Goegebeur et al. (2020b). Since this result is used several times in the proofs of our theorems, we repeat it here for completeness. In the sequel, weak convergence will be denoted by the arrow $\rightsquigarrow$.

Theorem 2.2 *Assume* $(\mathcal{D})$, $(\mathcal{H})$, $(\mathcal{K})$, $x_0 \in Int(S_X)$ *with* $f_X(x_0) > 0$, *and* $y \to F(y|x_0)$, *is strictly increasing. Consider sequences* $k \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$, *in such a way that* $k/n \to 0$, $kh_n^d \to \infty$, $h_n^{\eta_\varepsilon} \ln n/k \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_{f_X} \wedge \eta_A} \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_\gamma} \ln n/k \to 0$, $\sqrt{kh_n^d}|\delta(U(n/k|x_0)|x_0)| \to 0$. *Then we have,*

$$\sqrt{kh_n^d}\,(\widehat{\gamma}(x_0) - \gamma(x_0)) \rightsquigarrow \frac{\gamma(x_0)}{f_X(x_0)} \left[\int_0^1 W(z)\frac{1}{z}dz - W(1)\right],$$

where $W(z)$ *is a zero centered Gaussian process with covariance function*

$$\mathbb{E}(W(z)W(\overline{z})) = \|K\|_2^2 f_X(x_0)\,(z \wedge \overline{z})\,.$$

Note that the variance of the limiting distribution of $\widehat{\gamma}(x_0)$, after normalization, is given by $\gamma^2(x_0)\|K\|_2^2/f_X(x_0)$, compared to an asymptotic variance of γ^2 for the Hill estimator in the univariate context. The asymptotic variance of $\widehat{\gamma}(x_0)$ is inversely proportional to $f_X(x_0)$, which makes intuitively sense, as for x_0 where the density value is low, we expect fewer observations, leading to an increased variability of the estimator.

Next we establish the limiting distribution of the estimator $\widehat{\overline{F}}(R_n|x_0)$, after normalization.

Theorem 2.3 *Assume* $(\mathcal{D})$, $(\mathcal{H})$, $(\mathcal{K})$, $x_0 \in Int(S_X)$ *with* $f_X(x_0) > 0$ *and* $y \to F(y|x_0)$ *is strictly increasing. Consider sequences* $k \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$, *in such a way that* $k/n \to 0$, $h_n^{\eta_\varepsilon} \ln n/k \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_{f_X} \wedge \eta_A} \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_\gamma} \ln n/k \to 0$, $\sqrt{kh_n^d}|\delta(U(n/k|x_0)|x_0)| \to 0$. *Then, if* $R_n \to \infty$ *such that* $\frac{k}{n\overline{F}(R_n|x_0)} \to \infty$ *and* $\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]} \to \infty$, *as* $n \to \infty$, *we have*

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]} \left\{\frac{\widehat{\overline{F}}(R_n|x_0)}{\overline{F}(R_n|x_0)} - 1\right\} \rightsquigarrow \frac{1}{f_X(x_0)} \left[\int_0^1 W(z)\frac{1}{z}dz - W(1)\right].$$

The result of Theorem 2.3 indicates that the estimator $\widehat{\overline{F}}(R_n|x_0)$ inherits its asymptotic behavior from the estimator $\widehat{\gamma}(x_0)$, up to the factor $\gamma(x_0)$.

The aim of the next result is to handle the first term in the right-hand side of (8).

Theorem 2.4 *Assume* $(\mathcal{D})$, $(\mathcal{H})$, $(\mathcal{K})$, $x_0 \in Int(S_X)$ *with* $f_X(x_0) > 0$ *and* $y \to F(y|x_0)$ *is strictly increasing. Consider sequences* $k \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$, *in such a way that* $k/n \to 0$, $h_n^{\eta_\varepsilon} \ln n/k \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_{f_X} \wedge \eta_A} \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_\gamma} \ln n/k \to 0$, $\sqrt{kh_n^d}|\delta(U(n/k|x_0)|x_0)| \to 0$. *Then, if* $R_n \to \infty$ *such that* $\frac{k}{n\overline{F}(R_n|x_0)} \to \infty$, $\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]} \to \infty$ *and* $\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\Delta\left(\overline{F}(R_n|x_0)\right) \to 0$, *as* $n \to \infty$, *we have for* $\gamma(x_0) < -\rho$

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]} \left\{\frac{\widehat{\Pi}(R_n|x_0)}{\widetilde{\Pi}(R_n|x_0)} - 1\right\} \rightsquigarrow -\frac{\rho}{f_X(x_0)} \left[\int_0^1 W(z)\frac{1}{z}dz - W(1)\right].$$

Decomposition (8), combined with Theorem 2.1 and Theorem 2.4, yields the main result of the paper.

Theorem 2.5 *Assume* $(\mathcal{D})$, $(\mathcal{H})$, $(\mathcal{K})$, $x_0 \in Int(S_X)$ *with* $f_X(x_0) > 0$ *and* $y \to F(y|x_0)$ *is strictly increasing. Consider sequences* $k \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$, *in such a way that* $k/n \to 0$, $h_n^{\eta_\varepsilon} \ln n/k \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_{f_X} \wedge \eta_A} \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_\gamma} \ln n/k \to 0$, $\sqrt{kh_n^d}|\delta(U(n/k|x_0)|x_0)| \to 0$. *Then, if* $R_n \to \infty$ *such that* $\frac{k}{n\overline{F}(R_n|x_0)} \to \infty$, $\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]} \to \infty$ *and* $\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\Delta\left(\overline{F}(R_n|x_0)\right) \to 0$, *as* $n \to \infty$, *we have for* $\gamma(x_0) < -\rho$

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\widehat{\Pi}(R_n|x_0)}{\Pi(R_n|x_0)} - 1\right\} \rightsquigarrow -\frac{\rho}{f_X(x_0)}\left[\int_0^1 W(z)\frac{1}{z}dz - W(1)\right].$$

From this theorem we can see that the limiting distribution of $\widehat{\Pi}(R_n|x_0)$ is normal with mean zero and variance $\rho^2\|K\|_2^2/f_X(x_0)$.

3 Simulation experiment

In this section we evaluate the finite sample properties of the proposed estimator by a simulation experiment. In this, two premium principles will be considered:

- the net premium principle, corresponding to $g(x) = x$, for which $\rho = -1$,

- the dual-power premium principle, corresponding to $g(x) = 1 - (1-x)^\alpha$, for which $\rho = -1$. We set the loading parameter α at 1.366, as in Wang (1996), and as also used by Vandewalle and Beirlant (2006).

We assume that the conditional distribution of Y given $X = x$ is a Burr distribution with

$$\overline{F}(y|x) = \left(\frac{\beta}{\beta + y^{\tau(x)}}\right)^\lambda, y > 0, \quad \lambda, \beta, \tau(x) > 0.$$

This model satisfies assumption $(\mathcal{D})$ with $\gamma(x) = 1/(\lambda\tau(x))$ and $\beta(x) = \tau(x)$. The covariate X is assumed to be uniformly distributed on $[0,1]$. In our simulation we consider the following settings

- Setting 1: $\beta = \lambda = 1$ and

$$\tau(x) = 2\left[(0.1 + \sin(\pi x))\left(1.1 - 0.5\exp\left(-64\left(x - \frac{1}{2}\right)^2\right)\right)\right]^{-1}.$$

- Setting 2: $\beta = \lambda = 1$ and

$$\tau(x) = 1/\left(-0.43 + 0.48\sqrt{1+x}\right).$$

In Figure 1 we show the graphs of $\gamma(x_0)$ and Figure 2 displays the risk premiums $\Pi(R_n|x_0)$ for the values of R_n considered in the simulation experiment, for both of the settings.

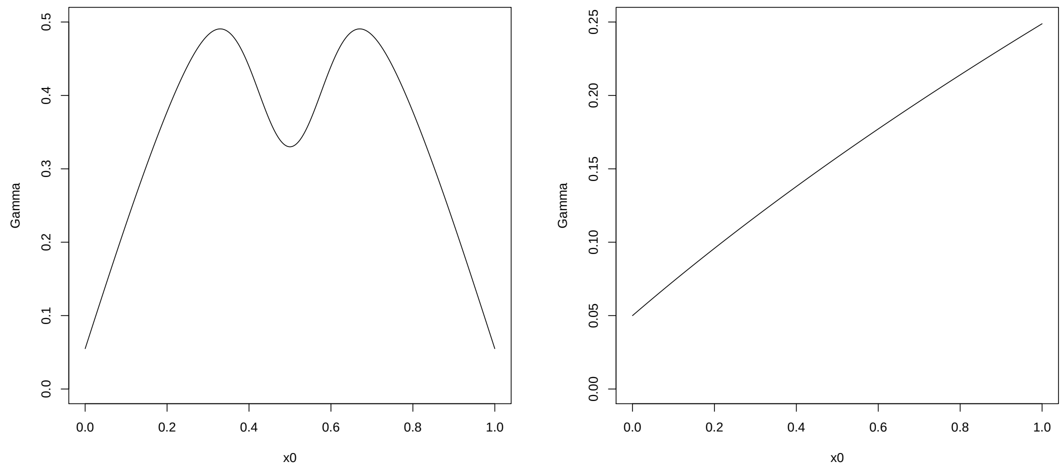

Figure 1: Burr simulation. $\gamma(x_0)$ for Setting 1 (left) and Setting 2 (right).

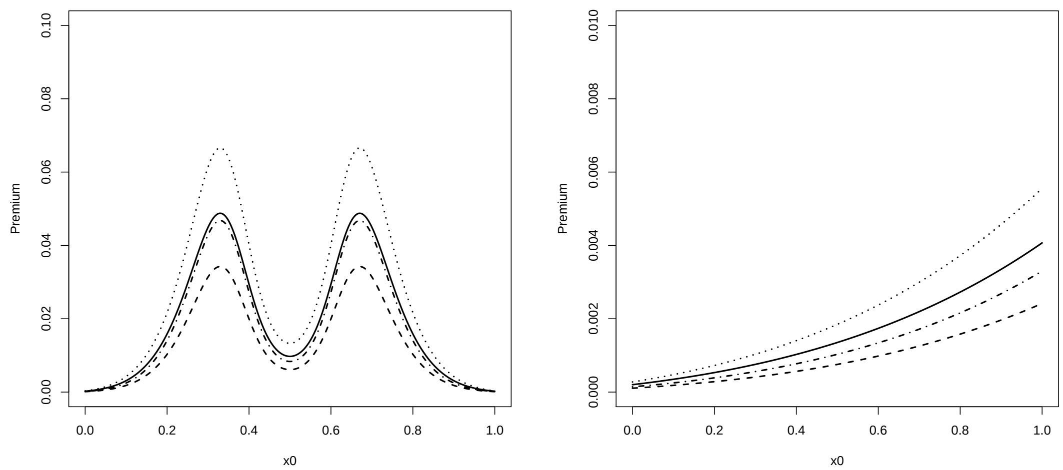

Figure 2: Burr simulation. $\Pi(R_n|x_0)$ for Setting 1 (left) and Setting 2 (right): net premium principle with $R_n = U(0.35n|x_0)$ (solid line) and $R_n = U(0.7n|x_0)$ (dashed line), and the dual-power premium principle with $R_n = U(0.35n|x_0)$ (dotted line) and $R_n = U(0.7n|x_0)$ (dashed-dotted line), where $n = 1000$.

We implement the proposed estimator with the bi-quadratic kernel function given by

$$K(x) = \frac{15}{16}(1 - x^2)^2 \mathbb{1}_{\{x \in [-1,1]\}}.$$

The parameter h_n is selected in a data-driven way by using the following cross-validation criterion

$$h_{cv} := \underset{h_n \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(1\!\!1_{\{Y_i \leqslant Y_j\}} - \widehat{F}_{n,-i}(Y_j|X_i) \right)^2,$$

where $\mathcal{H}$ is a grid of values for h_n and

$$\widehat{F}_{n,-i}(y|x) := \frac{\sum_{j=1, j\neq i}^{n} K_{h_n}(x - X_j) 1\!\!1_{\{Y_j \leqslant y\}}}{\sum_{j=1, j\neq i}^{n} K_{h_n}(x - X_j)},$$

is a cross-validation kernel estimator for $F(y|x)$. This criterion was introduced in Yao (1999), and considered in an extreme value context by Daouia et al. (2011, 2013) and Escobar-Bach et al. (2018). We set $\mathcal{H} = \{0.05, 0.06, \ldots, 0.15\} \times R_X$, where R_X is the range of X.

Next to the selection of h_n, our estimator also requires to determine k. In extreme value statistics, a reasonable value of k is typically obtained by plotting an estimate for an extreme value parameter of interest, here $\Pi(R_n|x_0)$, as a function of k, whereafter k is determined by a visual inspection of the obtained plot for a stable horizontal part. For an automated selection of the good k-range for estimation of $\Pi(R_n|x_0)$, we propose the following data-driven method:

1. Compute $\widehat{\Pi}(R_n|x_0)$ for $k = 30, \ldots, n^* - 1$, where n^* is the number of observations in the neighborhood of x_0 with radius h_n,
2. compute the standard deviation of the $\widehat{\Pi}(R_n|x_0)$ values in a moving block of 50 successive k-values,
3. select the block with the smallest standard deviation,
4. within the block selected in step 3, repeat the steps 2 and 3, now with block size 25,
5. the estimate for $\Pi(R_n|x_0)$ is the median of the $\widehat{\Pi}(R_n|x_0)$ in the finally selected block.

This data-driven method was introduced in the related context of estimation of $\gamma(x_0)$ by Goegebeur et al. (2019). Clearly, the method tries to identify a k-range in the stable region of the plot $(k, \widehat{\Pi}(R_n|x_0))$ in an automatic way.

We simulate 500 datasets of size $n = 1000$, and consider the premium estimation at $x_0 = 0.1, 0.2, \ldots, 0.9$. In Figure 3 we show for Setting 1 the boxplots of the estimates $\widehat{\Pi}(R_n|x_0)$ obtained in the 500 simulations at the different x_0, along with the true premium $\Pi(R_n|x_0)$ at these positions (connected by a blue line), with $R_n = U(0.35n|x_0)$ (left) and $R_n = U(0.7n|x_0)$ (right), and for the net premium principle (top row) and the dual-power premium principle (bottom row). Figure 4 shows the same information but now for Setting 2.

From the simulations we can draw the following conclusions:

- Overall, the boxplots of the premium estimates follow the shape of the true premium function $\Pi(R_n|x)$, with the true premium typically located inside the central box.

- Although R_n was set to $U(0.35n|x_0)$ and $U(0.7n|x_0)$, respectively, this leads already to premium calculations with extrapolations outside the data range. Indeed, for the two settings under consideration, the bandwidths obtained by the cross-validation are typically of the order 0.10-0.14, so that under a $U[0,1]$ distribution for the covariate X one expects around 200-300 observations locally. Despite this the proposed method works quite well.

- The estimation is more difficult in case of Setting 1 compared to Setting 2. This can be expected since the function $\gamma(x_0)$ is more complicated in Setting 1 where it shows local maxima and minima, while it is monotone in Setting 2. From the results for Setting 1 one can, e.g., see that at the local maxima, the median of the estimates is below $\Pi(R_n|x_0)$. This can be explained by the local nature of the estimation: indeed, for x_0 at or nearby a local maximum of $\gamma(x_0)$ one uses in the estimation mainly observations from distributions with a smaller $\gamma(x_0)$ value, leading to an underestimate of the premium. A similar comment can be made about the estimation at the local minima, where the estimate tends to be slightly upwards biased.

- The premium $\Pi(R_n|x_0)$ is decreasing in R_n, and is larger for the dual-power principle than for the net premium principle. The proposed method seems to perform equally well for the different combinations of R_n and premium principle.

4 Application to automobile claims data

In this section we illustrate the developed methodology on a dataset of automobile claims from Australia. As in many countries, third party insurance is a mandatory insurance for vehicle owners in Australia. This type of insurance protects vehicle owners against claims due to injury caused to other drivers, passengers or pedestrians, as a result of an accident. The dataset is available under the name `ausprivauto0405` in the `R` package CASdatasets, which is a collection of datasets from Charpentier (2014). It is based on one-year vehicle insurance policies taken out in 2004 or 2005, and contains information on variables like vehicle value (in thousands of Australian Dollars), vehicle age, gender and age of the policyholder, and the claim amount, from 67856 policies, of which 4624 had at least one claim. We focus here the analysis on the claim amounts which are positive and the corresponding vehicle values. In Figure 5 we show the scatterplot of claim size versus vehicle value (left) and versus ln(vehicle value) (right). Due to the skewness of the distribution of vehicle value, we will focus the analysis on using the covariate ln(vehicle value). Also, we restrict the estimation to the data with ln(vehicle value) $\in [-1, 1.5]$, since outside this interval the data are scarce. In order to validate the assumption of underlying conditional Pareto-type distributions for the claim sizes, we constructed local Pareto quantile plots of claim sizes for which the covariate is in a neighborhood of ln(vehicle value) $= -0.25$, 0.25 and 1, see Figure 6. If a dataset originates from a Pareto-type distribution, then the Pareto quantile plot will become linear in the largest observations (Beirlant et al., 2004, Section 2.3.5). The local Pareto quantile plots displayed in Figure 6 show an approximate linear pattern in the largest observations, indicating underlying Pareto-type distributions for the claim sizes. This is further confirmed by the plot of $\widehat{\gamma}$(ln(vehicle value)) versus ln(vehicle value) shown in Figure 7, where the estimates $\widehat{\gamma}$(ln(vehicle value)) are in the range $[0.2, 0.5]$, a range that is moreover

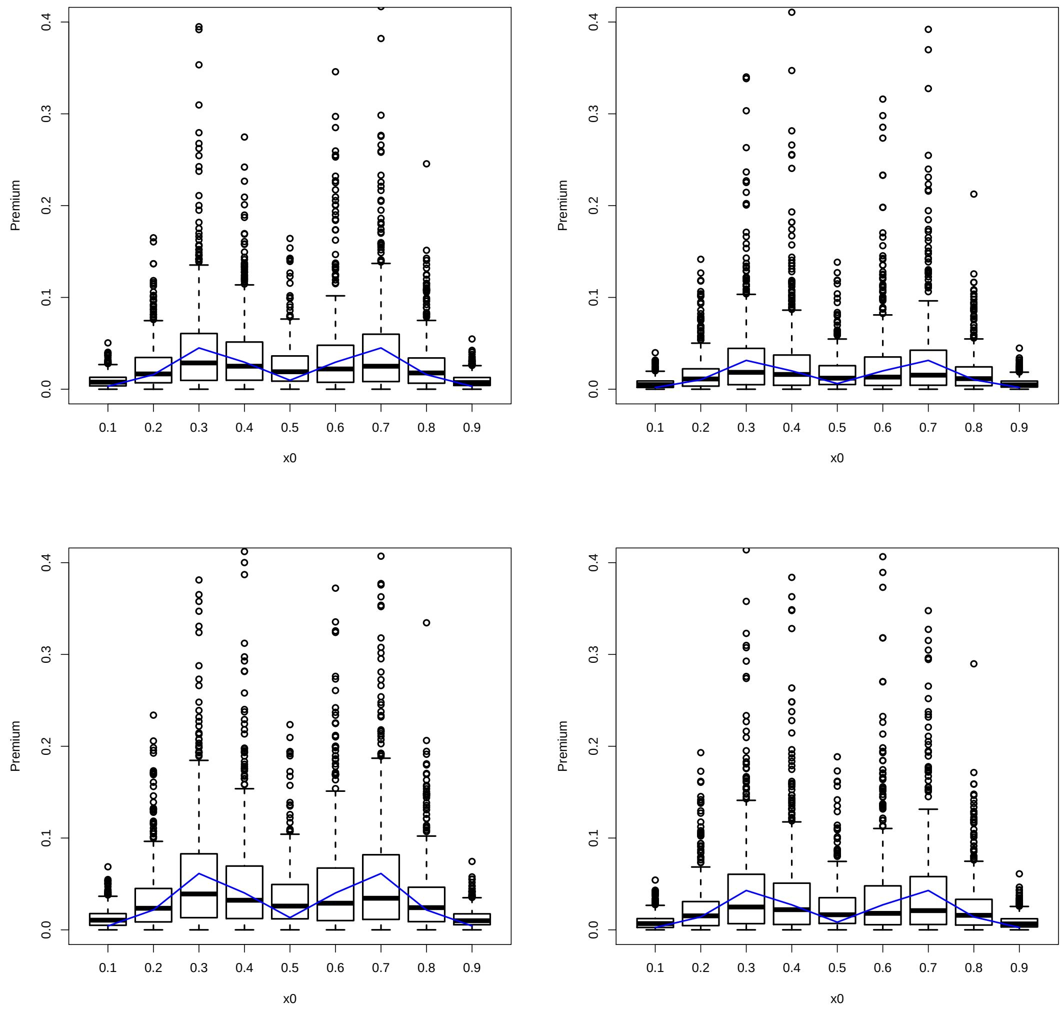

Figure 3: Setting 1. Boxplots of $\widehat{\Pi}(R_n|x_0)$, with $R_n = U(0.35|x_0)$ (left) and $R_n = U(0.70|x_0)$ (right), for the net premium principle (top row) and the dual-power premium principle (bottom row). The true premium values $\Pi(R_n|x_0)$ are connected with a blue line.

reasonable for actuarial applications. These estimates for the tail index are computed with the h_n and k values that were obtained from the premium determination algorithm. Next we estimate the reinsurance premiums with the methodology presented in this paper. In Figure 8 we show the estimate $\widehat{\Pi}(R_n|\ln(\text{vehicle value}))$ as a function of ln(vehicle value) for the net premium principle (left) and the dual-power premium principle with $\alpha = 1.366$ (right). We consider the retention levels $R_n = \widehat{U}(0.7n^*|\ln(\text{vehicle value}))$ (solid line) and $R_n = \widehat{U}(0.9n^*|\ln(\text{vehicle value}))$ (dashed line). The reported premium estimates are obtained with the algorithms for h_n and k selection described in the simulation section. Overall, the premium estimate follows the pattern in the

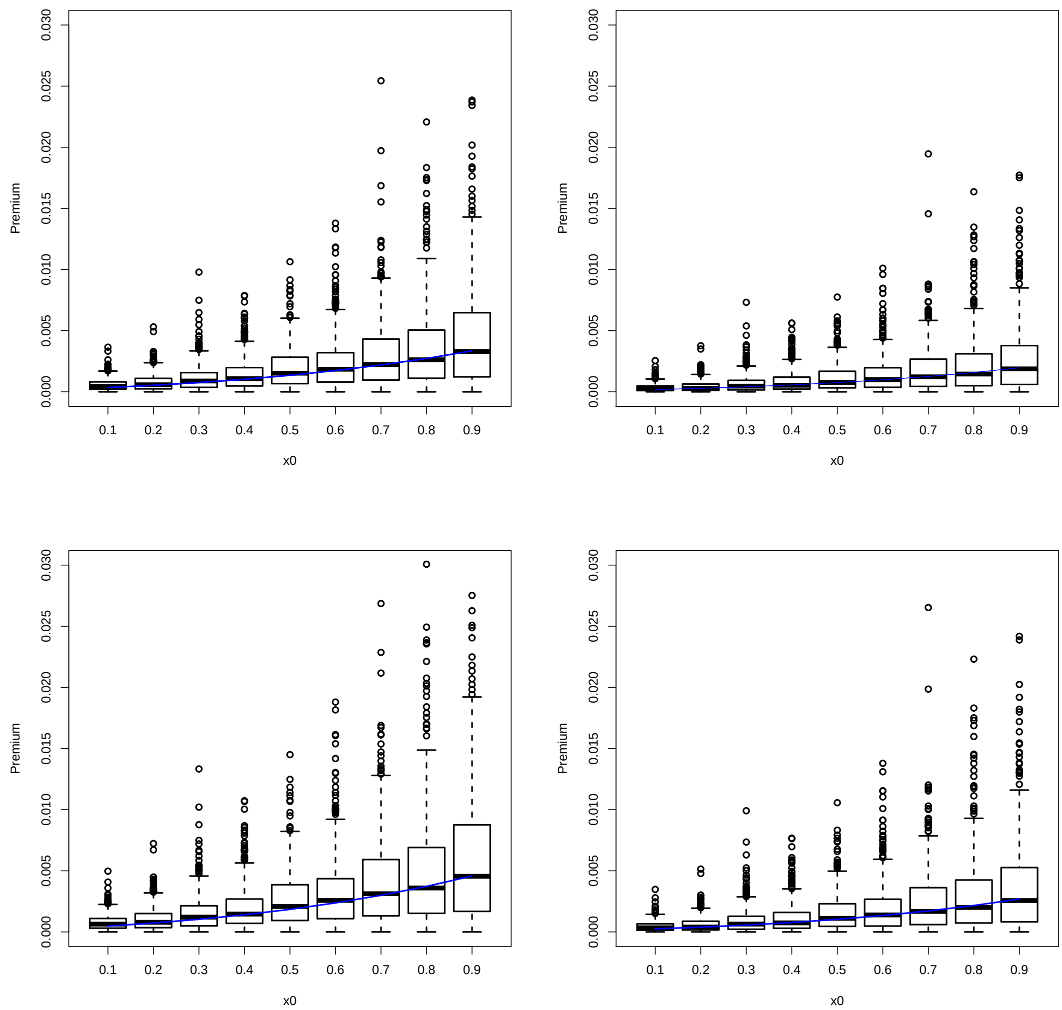

Figure 4: Setting 2. Boxplots of $\widehat{\Pi}(R_n|x_0)$, with $R_n = U(0.35|x_0)$ (left) and $R_n = U(0.70|x_0)$ (right), for the net premium principle (top row) and the dual-power premium principle (bottom row). The true premium values $\Pi(R_n|x_0)$ are connected with a blue line.

data. Indeed, at locations where the spacings of the top observations are large, we have a large tail index, leading in turn to a large premium estimate. Also, as expected, the dual-power premium principle leads to higher premium estimates compared to the net premium principle.

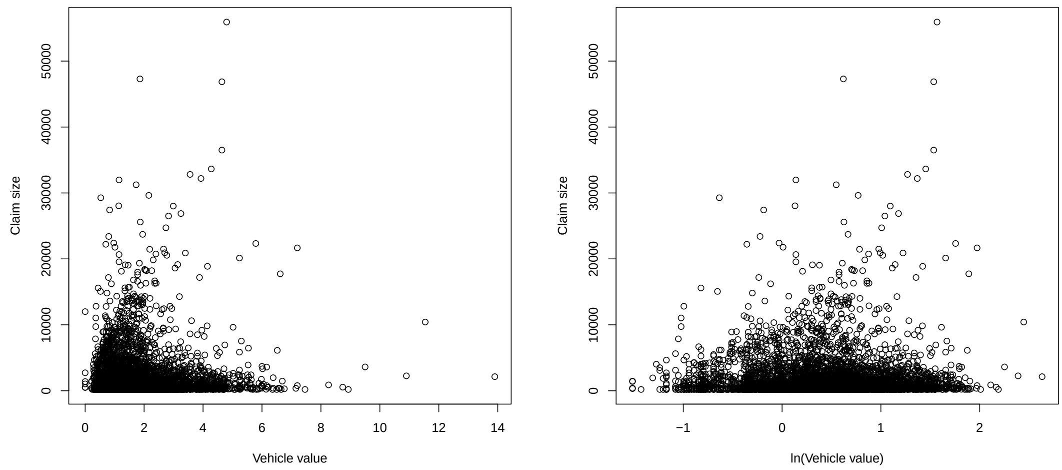

Figure 5: Australian automobile claims data. Scatterplot of claim size versus vehicle value (left) and versus ln(vehicle value) (right).

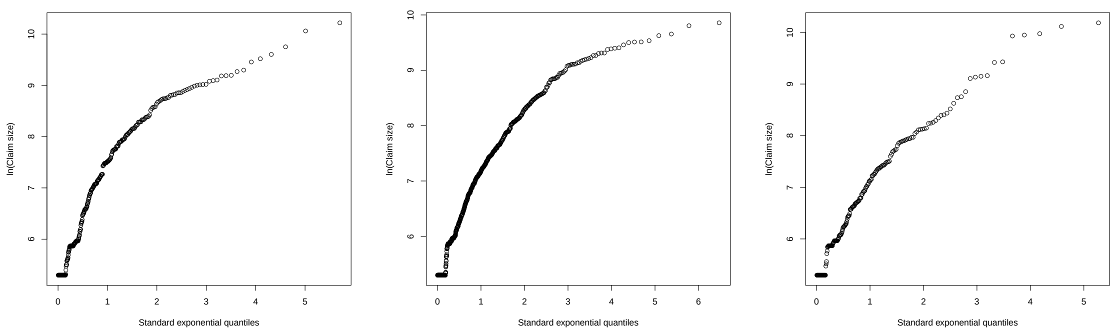

Figure 6: Australian automobile claims data. Local Pareto quantile plots of claim size at ln(vehicle value) $= -0.25$ (left), ln(vehicle value) $= 0.25$ (middle) and ln(vehicle value) $= 1$ (right).

Appendix

To be self contained, we recall below Lemma 5.6 from Goegebeur et al. (2020a), which is used several times in our proofs, and which states the weak convergence of $\widehat{u}_n := \widehat{U}(n/k|x_0)/U(n/k|x_0)$.

Lemma 4.1 *Assume* $(\mathcal{D})$, $(\mathcal{H})$, $(\mathcal{K})$, $x_0 \in Int(S_X)$ *with* $f_X(x_0) > 0$ *and* $y \to F(y|x_0)$ *is strictly increasing. Consider sequences* $k \to \infty$ *and* $h_n \to 0$ *as* $n \to \infty$*, in such a way that* $k/n \to 0$, $kh_n^d \to \infty$, $h_n^{\eta_\varepsilon} \ln n/k \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_{f_X} \wedge \eta_A} \to 0$, $\sqrt{kh_n^d}\, h_n^{\eta_\gamma} \ln n/k \to 0$, $\sqrt{kh_n^d}|\delta(U(n/k|x_0)|x_0)| \to$

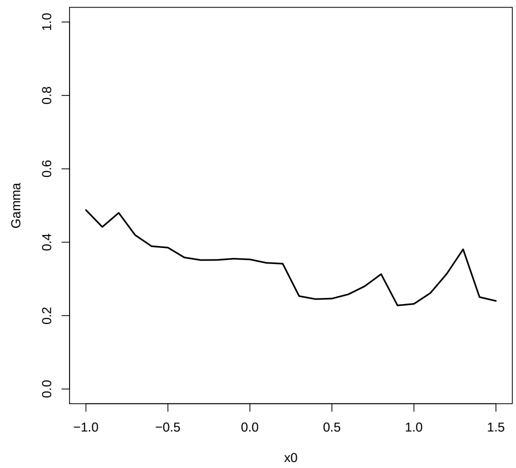

Figure 7: Australian automobile claims data. $\widehat{\gamma}(\ln(\text{vehicle value}))$ versus $\ln(\text{vehicle value})$.

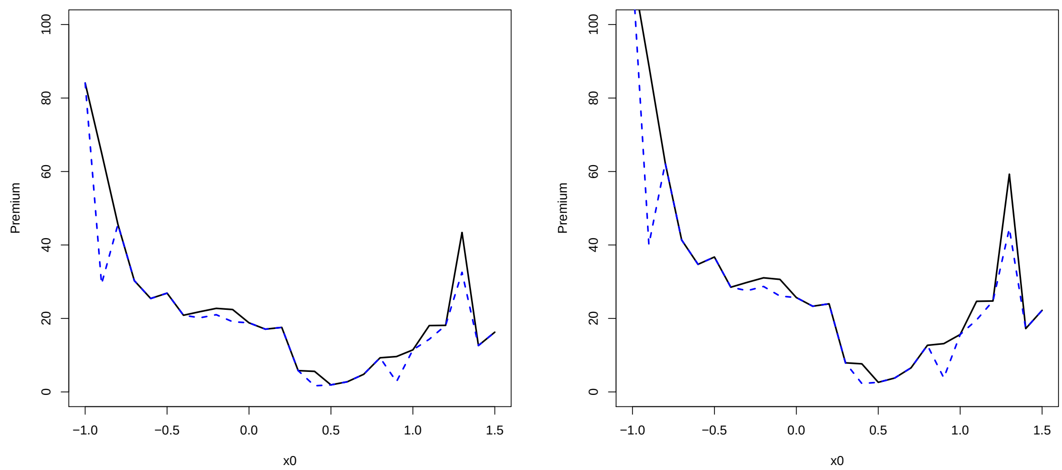

Figure 8: Australian automobile claims data. $\widehat{\Pi}(R_n | \ln(\text{vehicle value}))$ as a function of $\ln(\text{vehicle value})$ for the net premium principle (left) and the dual-power premium principle with $\alpha = 1.366$ (right), with $R_n = \widehat{U}(0.7n^* | \ln(\text{vehicle value}))$ (solid line) and $R_n = \widehat{U}(0.9n^* | \ln(\text{vehicle value}))$ (dashed line).

0. *Then, as $n \to \infty$, we have*

$$\sqrt{kh_n^d}\,(\widehat{u}_n - 1) \rightsquigarrow \frac{\gamma(x_0)W(1)}{f_X(x_0)}.$$

Proof of Theorem 2.1.

Note that

$$\begin{aligned}
\frac{\Pi(R_n|x_0)}{\widetilde{\Pi}(R_n|x_0)} &= -\frac{\gamma(x_0)+\rho}{\gamma(x_0)}\frac{1}{R_n}\int_{R_n}^{\infty}\frac{g(\overline{F}(y|x_0))}{g(\overline{F}(R_n|x_0))}dy \\
&= -\frac{\gamma(x_0)+\rho}{\gamma(x_0)}\int_1^{\infty}\left(\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho}\frac{\ell_g(\overline{F}(zR_n|x_0))}{\ell_g(\overline{F}(R_n|x_0))}dz \\
&= -\frac{\gamma(x_0)+\rho}{\gamma(x_0)}\int_1^{\infty}\left(\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho}dz \\
&\quad -\frac{\gamma(x_0)+\rho}{\gamma(x_0)}\int_1^{\infty}\left(\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho}\left[\frac{\ell_g(\overline{F}(zR_n|x_0))}{\ell_g(\overline{F}(R_n|x_0))}-1\right]dz \\
&= -\frac{\gamma(x_0)+\rho}{\gamma(x_0)}\int_1^{\infty}\left(\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho}dz \\
&\quad -\frac{\Delta(\overline{F}(R_n|x_0))}{1+\Delta(\overline{F}(R_n|x_0))}\frac{\gamma(x_0)+\rho}{\gamma(x_0)}\int_1^{\infty}\left(\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho}\left[\left(\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho*}-1\right]dz \\
&\quad -\frac{\Delta(\overline{F}(R_n|x_0))}{1+\Delta(\overline{F}(R_n|x_0))}\frac{\gamma(x_0)+\rho}{\gamma(x_0)}\int_1^{\infty}\left(\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho-\rho^*}\left[\frac{\ell_\Delta(\overline{F}(zR_n|x_0))}{\ell_\Delta(\overline{F}(R_n|x_0))}-1\right]dz \\
&=: T_{1,n}+T_{2,n}+T_{3,n}.
\end{aligned}$$

Note that a slight modification of Proposition 2.3 in Beirlant et al. (2009) gives

$$\sup_{z\geqslant 1} z^{1/\gamma(x_0)}\left|\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}-\overline{G}(z;\gamma(x_0),\delta(R_n|x_0),\beta(x_0))\right| = o(|\delta(R_n|x_0)|), \qquad R_n\to\infty.$$

where the extended Pareto distribution function G is defined by

$$G(z;\gamma,\delta,\beta)=\begin{cases} 1-[z(1+\delta-\delta z^{-\beta})]^{-1/\gamma}, & z>1,\\ 0, & z\leqslant 1.\end{cases}$$

This implies

$$\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)} = z^{-1/\gamma(x_0)}\left\{1-\frac{\delta(R_n|x_0)}{\gamma(x_0)}\left[1-z^{-\beta(x_0)}\right]+o(\delta(R_n|x_0))\right\}, \tag{9}$$

where the error term is uniform in $z\geqslant 1$.

Clearly, using (9), we have for $\gamma(x_0)<-\rho$

$$T_{1,n} = 1-\frac{\rho\beta(x_0)}{\gamma(x_0)+\rho-\gamma(x_0)\beta(x_0)}\delta(R_n|x_0)+o(\delta(R_n|x_0)), \tag{10}$$

$$T_{2,n} = -\frac{\rho^*}{\gamma(x_0)+\rho+\rho^*}\Delta(\overline{F}(R_n|x_0))+o(\Delta(\overline{F}(R_n|x_0))). \tag{11}$$

Now, concerning $T_{3,n}$, according to Proposition B.1.10 in de Haan and Ferreira (2006), for any $\varepsilon > 0$ and $\zeta \in (0, -\rho^*]$, we have for n large and for some constant $L > 0$

$$|T_{3,n}| \leqslant L\,\varepsilon\,|\Delta(\overline{F}(R_n|x_0))| \int_1^\infty \left(\frac{\overline{F}(zR_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho-\rho^*-\zeta} dz.$$

Hence, again using (9), we have that

$$T_{3,n} = o\left(|\Delta(\overline{F}(R_n|x_0))|\right). \tag{12}$$

Combining (10), (11) and (12) yields Theorem 2.1.

Proof of Theorem 2.3.

We need to study

$$\begin{aligned}
&\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\widehat{\overline{F}}(R_n|x_0)}{\overline{F}(R_n|x_0)} - 1\right\}\\
&= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{k}{n\overline{F}(R_n|x_0)}\left[\frac{U\left(\frac{n}{k}\frac{k}{n\overline{F}(R_n|x_0)}\Big|x_0\right)}{U\left(\frac{n}{k}|x_0\right)}\right]^{-1/\widehat{\gamma}(x_0)} \widehat{u}_n^{1/\widehat{\gamma}(x_0)} - 1\right\}\\
&= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\left(\frac{k}{n\overline{F}(R_n|x_0)}\right)^{\frac{\widehat{\gamma}(x_0)-\gamma(x_0)}{\widehat{\gamma}(x_0)}}\left(\frac{1+a\left(\frac{n}{k}|x_0\right)}{1+a\left(\frac{1}{\overline{F}(R_n|x_0)}\Big|x_0\right)}\right)^{1/\widehat{\gamma}(x_0)} \widehat{u}_n^{1/\widehat{\gamma}(x_0)} - 1\right\}\\
&=: \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{T_{4,n}T_{5,n}T_{6,n} - 1\right\}\\
&= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{(T_{4,n}-1)T_{5,n}T_{6,n} + (T_{5,n}-1)T_{6,n} + (T_{6,n}-1)\right\}.
\end{aligned} \tag{13}$$

We have

$$\begin{aligned}
&\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}(T_{4,n}-1)\\
&\quad= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left[\exp\left(\sqrt{kh_n^d}\,\frac{\widehat{\gamma}(x_0)-\gamma(x_0)}{\widehat{\gamma}(x_0)}\,\frac{\ln[k/(n\overline{F}(R_n|x_0))]}{\sqrt{kh_n^d}}\right) - 1\right],
\end{aligned}$$

which gives, after applying a Taylor series expansion,

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}(T_{4,n}-1) = \exp(\Gamma_n)\frac{\sqrt{kh_n^d}}{\widehat{\gamma}(x_0)}\left(\widehat{\gamma}(x_0)-\gamma(x_0)\right),$$

where Γ_n is a random value between zero and

$$\sqrt{kh_n^d}\frac{\widehat{\gamma}(x_0)-\gamma(x_0)}{\widehat{\gamma}(x_0)}\frac{\ln[k/(n\overline{F}(R_n|x_0))]}{\sqrt{kh_n^d}}.$$

Thus, under our assumptions and using Theorem 2.2, together with Slutsky's theorem, we have

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}(T_{4,n}-1) \rightsquigarrow \frac{1}{f_X(x_0)}\left[\int_0^1 W(z)\frac{1}{z}dz - W(1)\right]. \tag{14}$$

Now we turn to $T_{5,n}$. Again by a Taylor series expansion we obtain

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}(T_{5,n}-1) = \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\frac{1}{\widehat{\gamma}(x_0)}\left[a(n/k|x_0) - a(1/\overline{F}(R_n|x_0)|x_0)\right](1+o_{\mathbb{P}}(1)).$$

Using the fact that $a(y|x_0) = \delta(U(y|x_0)|x_0)(1+o(1))$ gives

$$\begin{aligned}\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}a(n/k|x_0) &= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\delta(U(n/k|x_0)|x_0)(1+o(1)) \\ &\to 0,\end{aligned}$$

under our assumptions. As for the term involving $a(1/\overline{F}(R_n|x_0)|x_0)$, we use Potter's bounds (see Proposition B.1.9 (5) in de Haan and Ferreira, 2006) to obtain, for some constant $L>0$ and $0<\xi<\beta(x_0)$, and for n large,

$$\begin{aligned}&\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left|\delta\left(U\left(\frac{1}{\overline{F}(R_n|x_0)}\middle|x_0\right)\middle|x_0\right)\right| \\ &= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left|\delta\left(U\left(\frac{n}{k}\middle|x_0\right)\middle|x_0\right)\right|\left|\frac{\delta\left(U\left(\frac{1}{\overline{F}(R_n|x_0)}\middle|x_0\right)\middle|x_0\right)}{\delta\left(U\left(\frac{n}{k}\middle|x_0\right)\middle|x_0\right)}\right| \\ &\leqslant L\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left|\delta\left(U\left(\frac{n}{k}\middle|x_0\right)\middle|x_0\right)\right|\left(\frac{U\left(\frac{1}{\overline{F}(R_n|x_0)}\middle|x_0\right)}{U\left(\frac{n}{k}\middle|x_0\right)}\right)^{-\beta(x_0)+\xi} \\ &\leqslant L\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left|\delta\left(U\left(\frac{n}{k}\middle|x_0\right)\middle|x_0\right)\right|,\end{aligned}$$

and hence

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}a\left(\frac{1}{\overline{F}(R_n|x_0)}\middle|x_0\right) = o(1).$$

This leads to

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}(T_{5,n}-1) = o_{\mathbb{P}}(1). \tag{15}$$

Finally,

$$\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}(T_{6,n}-1) = o_{\mathbb{P}}(1), \tag{16}$$

by Lemma 4.1.

Combining (13) with (14), (15) and (16) establishes Theorem 2.3.

Proof of Theorem 2.4.

We have the decomposition

$$\begin{aligned}
\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\widehat{\Pi}(R_n|x_0)}{\widetilde{\Pi}(R_n|x_0)}-1\right\} &= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\widehat{\gamma}(x_0)}{\gamma(x_0)}\,\frac{\gamma(x_0)+\rho}{\widehat{\gamma}(x_0)+\rho}\,\frac{g\left(\widehat{\overline{F}}(R_n|x_0)\right)}{g\left(\overline{F}(R_n|x_0)\right)}-1\right\}\\
&= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\widehat{\gamma}(x_0)}{\gamma(x_0)}-1\right\}\frac{\gamma(x_0)+\rho}{\widehat{\gamma}(x_0)+\rho}\,\frac{g\left(\widehat{\overline{F}}(R_n|x_0)\right)}{g\left(\overline{F}(R_n|x_0)\right)}\\
&\quad+\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\gamma(x_0)+\rho}{\widehat{\gamma}(x_0)+\rho}-1\right\}\frac{g\left(\widehat{\overline{F}}(R_n|x_0)\right)}{g\left(\overline{F}(R_n|x_0)\right)}\\
&\quad+\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{g\left(\widehat{\overline{F}}(R_n|x_0)\right)}{g\left(\overline{F}(R_n|x_0)\right)}-1\right\}\\
&=: T_{7,n}+T_{8,n}+T_{9,n}.
\end{aligned}$$

Clearly, by Theorem 2.2 combining with our assumptions, we have

$$\begin{aligned}
T_{7,n} &= o_{\mathbb{P}}(1),\\
T_{8,n} &= o_{\mathbb{P}}(1).
\end{aligned}$$

Note also that

$$\begin{aligned}
T_{9,n} &= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\left(\frac{\widehat{\overline{F}}(R_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho}\frac{\ell_g\left(\widehat{\overline{F}}(R_n|x_0)\right)}{\ell_g\left(\overline{F}(R_n|x_0)\right)}-1\right\}\\
&= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\left(\frac{\widehat{\overline{F}}(R_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho}-1\right\}\\
&\quad+\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\frac{\Delta(\overline{F}(R_n|x_0))}{1+\Delta(\overline{F}(R_n|x_0))}\left[\frac{\Delta(\widehat{\overline{F}}(R_n|x_0))}{\Delta(\overline{F}(R_n|x_0))}-1\right]\left(\frac{\widehat{\overline{F}}(R_n|x_0)}{\overline{F}(R_n|x_0)}\right)^{-\rho}\\
&=: T_{9,1,n}+T_{9,2,n}.
\end{aligned}$$

By Theorem 2.3 and the delta method (see, e.g., van der Vaart, 1998) we have

$$T_{9,1,n} \rightsquigarrow -\frac{\rho}{f_X(x_0)}\left[\int_0^1 W(z)\frac{1}{z}dz - W(1)\right].$$

As for $T_{9,2,n}$ we use Proposition B.1.10 in de Haan and Ferreira (2006) to obtain, with $\varepsilon, \xi > 0$, arbitrary, and for n large, with arbitrary large probability

$$\begin{aligned}\left|\frac{\Delta(\widehat{\overline{F}}(R_n|x_0))}{\Delta(\overline{F}(R_n|x_0))} - 1\right| &\leqslant \left|\frac{\Delta(\widehat{\overline{F}}(R_n|x_0))}{\Delta(\overline{F}(R_n|x_0))} - \left(\frac{\widehat{\overline{F}}(R_n|x_0))}{\overline{F}(R_n|x_0)}\right)^{-\rho^*}\right| + \left|\left(\frac{\widehat{\overline{F}}(R_n|x_0))}{\overline{F}(R_n|x_0)}\right)^{-\rho^*} - 1\right| \\ &\leqslant \varepsilon\left(\frac{\widehat{\overline{F}}(R_n|x_0))}{\overline{F}(R_n|x_0)}\right)^{-\rho^*\pm\xi} + \left|\left(\frac{\widehat{\overline{F}}(R_n|x_0))}{\overline{F}(R_n|x_0)}\right)^{-\rho^*} - 1\right|,\end{aligned}$$

where the notation $a^{\pm\bullet}$ means $a^{\bullet}$ if $a \geqslant 1$ and $a^{-\bullet}$ if $a < 1$.

Hence

$$\left|\frac{\Delta(\widehat{\overline{F}}(R_n|x_0))}{\Delta(\overline{F}(R_n|x_0))} - 1\right| = o_{\mathbb{P}}(1),$$

and thus $T_{9,2,n} = o_{\mathbb{P}}(1)$. Then, by Slutsky's theorem we have that

$$T_{9,n} \rightsquigarrow -\frac{\rho}{f_X(x_0)}\left[\int_0^1 W(z)\frac{1}{z}dz - W(1)\right].$$

This achieves the proof of Theorem 2.4.

Proof of Theorem 2.5.

We have by (8) and Theorem 2.1

$$\begin{aligned}&\frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\widehat{\Pi}(R_n|x_0)}{\Pi(R_n|x_0)} - 1\right\} \\ &= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\widehat{\Pi}(R_n|x_0)}{\widetilde{\Pi}(R_n|x_0)} - 1\right\} \\ &\quad + \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\rho\beta(x_0)}{\gamma(x_0)+\rho-\gamma(x_0)\beta(x_0)}\delta(R_n|x_0) + \frac{\rho^*}{\gamma(x_0)+\rho+\rho^*}\Delta(\overline{F}(R_n|x_0))\right\}(1+o_{\mathbb{P}}(1)) \\ &= \frac{\sqrt{kh_n^d}}{\ln[k/(n\overline{F}(R_n|x_0))]}\left\{\frac{\widehat{\Pi}(R_n|x_0)}{\widetilde{\Pi}(R_n|x_0)} - 1\right\} + o_{\mathbb{P}}(1)\end{aligned}$$

under our assumptions. Theorem 2.4 achieves the proof of Theorem 2.5.

Acknowledgement

The research of Armelle Guillou was supported by the French National Research Agency under the grant ANR-19-CE40-0013-01/ExtremReg project and an International Emerging Action (IEA-00179). Computation/simulation for the work described in this paper was supported by the DeIC National HPC Centre, SDU.

References

Albrecher, H., Beirlant, J. and Teugels, J.L. (2017). Reinsurance: Actuarial and Statistical Aspects. Wiley.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J.L. (2004). Statistics of Extremes - Theory and Applications. Wiley.

Beirlant, J., Joossens, E. and Segers, J. (2009). Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. Journal of Statistical Planning and Inference, 139, 2800–2815.

Beirlant, J., Matthys, G. and Dierckx, G. (2001). Heavy-tailed distributions and rating. ASTIN Bulletin, 31, 37–58.

Charpentier, A., (2014). Computational actuarial science with R. Chapman & Hall/CRC.

Daouia, A., Gardes, L. and Girard, S. (2013). On kernel smoothing for extremal quantile regression. Bernoulli, 19, 2557–2589.

Daouia, A., Gardes, L., Girard, S. and Lekina, A. (2011). Kernel estimators of extreme level curves. Test, 20, 311–333.

El Methni, J., Gardes, L. and Girard, S. (2014). Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions. Scandinavian Journal of Statistics, 41, 988-1012.

El Methni, J., Gardes, L. and Girard, S. (2018). Kernel estimation of extreme regression risk measures. Electronic Journal of Statistics, 12, 359–398.

Escobar-Bach, M., Goegebeur, Y. and Guillou, A. (2018). Local estimation of the conditional stable tail dependence function. Scandinavian Journal of Statistics, 45, 590–617.

Goegebeur, Y., Guillou, A., Ho, N.K.L. and Qin, J. (2020a). Conditional marginal expected shortfall. https://hal.archives-ouvertes.fr/hal-02272392.

Goegebeur, Y., Guillou, A., Ho, N.K.L. and Qin.J. (2020b). A Weissman-type estimator of the conditional marginal expected shortfall. https://hal.archives-ouvertes.fr/hal-02613135.

Goegebeur, Y., Guillou, A. and Qin, J. (2019). Bias-corrected estimation for conditional Pareto-type distributions with random right censoring. Extremes, 22, 459–498.

de Haan, L. and Ferreira, A. (2006). Extreme value theory. An introduction. Springer.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. Annals of Statistics, 5, 1163–1174.

van der Vaart, A.W. (1998). Asymptotic statistics, Cambridge Series in Statistical and Probabilistic Mathematics, 3, Cambridge University Press, Cambridge.

Vandewalle, B. and Beirlant, J. (2006). On univariate extreme value statistics and the estimation of reinsurance premiums. Insurance: Mathematics and Economics, 38, 441–459.

Wang, S. (1996). Premium calculation by transforming the layer premium density. ASTIN Bulletin, 26, 71–92.

Yao, Q. (1999). Conditional predictive regions for stochastic processes. Technical report, Institute of Mathematics and Statistics, University of Kent at Canterbury.