2006

Estimation of Large Insurance Losses: A Case Study

Tine Buch-Kromann
Codan Insurance, tbl@codan.dk

Estimation of Large Insurance Losses: A Case Study

Tine Buch-Kromann*

Abstract†

This paper demonstrates an approach to analyzing liability data recently developed by a Danish insurance company. The approach is based on a Champernowne distribution, which is corrected with a non-parametric estimator. The correction estimator is obtained by transforming the data set with the estimated modified Champernowne cdf and then estimating the density of the transformed data set by using the classical kernel density estimator. Our approach is illustrated by applying it to an actual data set.

Key words and phrases: *Semiparametric kernel density estimator, corrected modified Champernowne method, heavy-tailed distributions, Champernowne distribution, extreme value theory, generalized Pareto distribution*

1 Introduction

This paper demonstrates a unified approach to large loss estimation recently developed in a Danish insurance company. A unified approach was needed because actuaries and statisticians were spending too much time trying to develop parametric models of losses. Thus, they often

*Tine Buch-Kromann (formerly Tine Buch-Larsen), MA., is a business researcher at Codan Insurance. She received her MA. in actuarial science from the University of Copenhagen in 2003 and is currently enrolled as Ph.D. student at the same university. She has been with Codan Insurance since 2002 and has for Codan's British parent company Royal & SunAlliance's Commercial Department.

Ms. Buch-Kromann's address is: Codan Insurance, Gammel Kongevej 60, DK-1790 Copenhagen V, DENMARK. E-mail: *tbl@codan.dk*

†The author thanks the Royal & SunAlliance's Commercial Department for providing the data set used in this paper. Also, thanks to the anonymous referees for their comments that have substantially improved this paper.

decided to estimate small and large losses separately because no single parametric model seemed to fit both small and large losses. Apart from the usual challenges such as choosing the appropriate parametric model and identifying the best way of estimating the parameters, a big problem was in determining the threshold between small and large losses, if they are to be estimated separately. Clearly the solution to this problem is fundamentally important to the quality of the estimation.

One approach is to use extreme value theory and generalized Pareto distributions, as described in Embrechts, Klüppelberg, and Mikosch (1997) and Cebrián, Denuit, and Lambert (2003), to analyze the loss data. As this approach, however, is mainly concerned with the estimation of large losses, it maintains the necessity to determine the threshold between small and large losses.

The approach adopted by the Danish insurance company is based on Buch-Larsen et al. (2005) who developed a unified method based on a semi-parametric estimator, i.e., a parametric estimator corrected with a non-parametric correction estimator.[1] The semi-parametric estimator is obtained by transforming the data set with the transform function, $T(x)$, which is the cdf of a modified Champernowne distribution. If $X_1, \ldots, X_N$ represent the data set, then the transformed data set is $Z_1, \ldots, Z_N$ where $Z_i = T(X_i)$ for $i = 1, \ldots, N$. The density of the transformed data set is estimated by means of a classical kernel density estimator [Wand and Jones (1995, page 11)]:

$$\hat{g}(z) = \frac{1}{Nb} \sum_{i=1}^{N} K\left(\frac{z - Z_i}{b}\right) \tag{1}$$

where K is the kernel function and b is the bandwidth. The estimator for the original data set is obtained by an inverse transformation of $\hat{g}(z)$. This results in an estimator that is close to a parametric estimator for small values of N and "more" non-parametric as N increases. The estimator $\hat{g}(z)$ is flexible in that it provides good estimates for many different shapes of loss distributions.

[1] Semiparametric estimators were introduced in the statistics literature by Wand, Marron, and Ruppert (1991) who demonstrated that the classical kernel density estimator could be improved by transforming the data set with a shifted power transformation. Since then semiparametric estimators have been used by other authors including Hjort and Glad (1995), Jones, Linton, and Nielsen (1995), Yang and Marron (1999), and Bolancé, Guillén, and Nielsen (2003). Clements, Hurn, and Lindsay (2003) have developed semiparametric estimators based on a Möbius-like transformation, which is a special case of the Champernowne distribution. This method was further developed by Buch-Larsen et al. (2005) using a modified Champernowne distribution for greater flexibility.

In this paper we will provide a detailed outline of the Buch-Larsen et al. (2005) method, which we have called the corrected modified Champernowne method. In addition, we will introduce an alternative parameter estimation method, called the QM method, which provides better estimates of conditional right-tail expected losses compared to those based on maximum likelihood parameter estimation. Moreover, we compare the corrected modified Champernowne method to the generalized Pareto distribution method of Cebrián, Denuit, and Lambert (2003).

2 Estimation of Parameters

The modified Champernowne distribution is a generalization of the Champernowne distribution (Brown, 1937 and Champernowne, 1952) with an extra parameter c to ensure that the pdf of the modified Champernowne distribution is positive at 0 for all α when $c > 0$ and is zero when $c = 0$. The *modified Champernowne distribution* is defined as:

$$T_{\alpha,M,c}(x) = \frac{(x+c)^\alpha - c^\alpha}{(x+c)^\alpha + (M+c)^\alpha - 2c^\alpha} \tag{2}$$

for $x \geq 0$, with parameters $\alpha > 0$, $M > 0$ and $c \geq 0$ and density

$$\frac{\mathrm{d}T_{\alpha,M,c}(x)}{\mathrm{d}x} = \frac{\alpha(x+c)^{\alpha-1}((M+c)^\alpha - c^\alpha)}{((x+c)^\alpha + (M+c)^\alpha - 2c^\alpha)^2}. \tag{3}$$

The inverse cdf is

$$T^{-1}_{\alpha,M,c}(z) = \left[\frac{z(M+c)^\alpha - (2z-1)c^\alpha}{1-z}\right]^{1/\alpha} - c. \tag{4}$$

Buch-Larsen et al. (2005) have shown that the modified Champernowne distribution is a heavy-tailed distribution that converges to a Pareto distribution in the tail.

Two estimation methods are used for the parameters α, M, and c of the modified Champernowne distribution: the well-known maximum likelihood method and the quantile-mean method, which selects parameters in a way that emphasizes the goodness of fit in the right tail. As $T_{\alpha,M,c}(M) \equiv 0.5$ for all c and α, M is assumed to be equal to the empirical (sample) median in both of these methods. Although this gives a sub-optimal estimate of M, Clements, Hurn, and Lindsay (2003) have argued that it is reasonable to assume that the empirical median is close to the maximum likelihood estimate of M. The empirical median has a further advantage: it is a robust estimator, especially for heavy-tailed distributions (Lehmann, 1991). After the parameter M has been estimated, the estimate of (α, c) is found by each of the methods.

The *maximum likelihood estimate* (MLE) is found by maximizing the log likelihood function:

$$l(\alpha, c) = N \log \alpha + N \log((M + c)^{\alpha} - c^{a}) + (\alpha - 1) \sum_{i=1}^{N} \log(X_i + c) \\ - 2 \sum_{i=1}^{N} \log\left((X_i + c)^{\alpha} + (M + c)^{\alpha} - 2c^{a}\right).$$

The properties of the MLE are well-known: it is efficient and ensures the best fit over the entire range of the distribution.

Because the risk of large losses lies in the tail of the loss distribution, we have also tested the *quantile-mean method*, which is a heuristic parameter estimation method. In this method we first select the parameter α so that the 95 quantile point of the empirical or sample cdf and of the estimated modified Champernowne distribution are equal. The parameter c is then chosen so that the mean of the estimated modified Champernowne distribution is as close as possible to the empirical mean.

Though there may be better ways of choosing α and c, it is important to choose parameters that result in accurate estimates of the number of large losses and the mean because these statistics are important in determining premiums.

3 An Illustration of Density Estimation

The data are losses (claims) from employer's liability line of business at Royal & SunAlliance, a British company. The data consist of 34,493 losses ranging from £0 to £4,213,057 without truncations or censoring, i.e., before deductibles and policy limits are applied. The use of untruncated and uncensored loss data is critical to the application of the proposed method.[2] The average loss size is £26,597. The employers are subdivided into 13 trade groups as shown in Table 1. For each

[2]For an analysis of losses with truncation and censoring see, for example, Cebrián, Denuit, and Lambert (2003) and Denuit, Purcaru, and Van Keilegom (2006).

trade group, the problem is to calculate the expected loss size for a deductible of d (left truncation) and a policy limit (or retention limit) of u (right censoring) where $d < u$.

The employer's liability data set is heavy-tailed, which can be seen by the upward tendency of the empirical mean excess function in Figure 1 (left) and the concave departure of the exponential QQ-plot in Figure 1 (right).

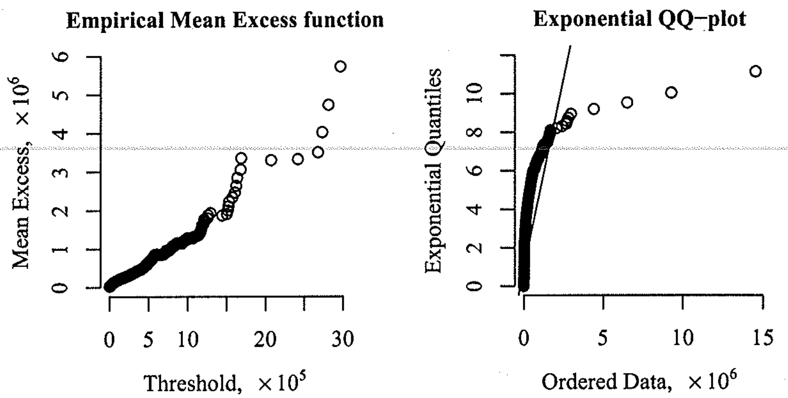

Figure 1: Empirical Mean Excess (Left) and Exponential QQ-Plot (right)

Table 1 shows the MLE and QM estimates of the parameters for the liability data set for each trade group. The M parameters for MLE and QM are equal because they are estimated in the same way. For the α parameters, no clear tendency is seen, whereas the c-parameters seem to be larger with the QM method than with the MLE method.

The estimation method proposed by Buch-Larsen et al. (2005), called the *corrected modified Champernowne* (CMC) method, is demonstrated by applying it to the data set. The CMC method is essentially a semiparametric transformation kernel density estimator, which is computed by transforming the data set with a modified Champernowne distribution and applying a non-parametric classical kernel density estimator to the transformed data set. The kernel smoothing function is a correction to the parametric modified Champernowne transformation function. Because of the properties of kernel smoothing, the correction will be

Table 1
Estimated Modified Champernowne Parameters

Trade	Sample	MLE Estimates			QM Estimates		
Group i	Size N_i	$\widehat{\alpha}_{\mathrm{MLE}}$	$\widehat{M}_{\mathrm{MLE}}$	$\widehat{c}_{\mathrm{MLE}}$	$\widehat{\alpha}_{\mathrm{QM}}$	$\widehat{M}_{\mathrm{QM}}$	$\widehat{c}_{\mathrm{QM}}$
1	1,668	1.610	13,616	6,808	1.400	13,616	27,232
2	597	1.401	12,437	0	1.653	12,437	24,874
3	2,112	1.563	8,532	0	1.563	8,532	4,266
4	537	1.563	8,867	0	1.808	8,867	17,733
5	1,083	1.726	9,596	0	1.774	9,596	4,798
6	2,054	1.888	8,777	4,388	1.913	8,777	17,554
7	707	1.458	9,744	0	1.455	9,744	19,487
8	3,620	2.108	8,858	4,429	1.967	8,858	13,287
9	931	1.481	9,423	0	1.629	9,423	14,135
10	6,297	1.935	9,268	4,634	1.950	9,268	13,902
11	1,022	1.656	11,041	0	1.562	11,041	0
12	5,668	1.865	10,629	5,315	1.934	10,629	21,259
13	8,197	1.574	10,790	5,395	1.493	10,790	21,581

weak if there are few data points and becomes more pronounced as the sample size increases. This means that the transformed kernel density estimator resembles a parametric estimator for small sample sizes and a non-parametric estimator for larger sample sizes.

Let $X_1^i, ..., X_{N_i}^i$ be the data set with sample size N_i for trade group i with an unknown cdf $F_i(x)$ and density $f_i(x)$. We will use a detailed numerical illustration for trade group 1 only, where $N_1 = 1668$. Figure 2 illustrates the four steps of the CMC estimation with QM parameters of f_1.[3] These steps are described in general as follows:

Step 1: Estimate the parameters (α, M, c) of the modified Champernowne distribution as described in Section 2 using either the MLE or QM method. These estimates are displayed in Table 1. Figure 2(1) shows a histogram for the raw data for trade group 1 and the estimated modified Champernowne distribution with QM parameters (dotted line).

[3]The corresponding figure for the CMC estimator with MLE parameters is available from <http://www.math.ku.dk/~tbl/joap06.html>.

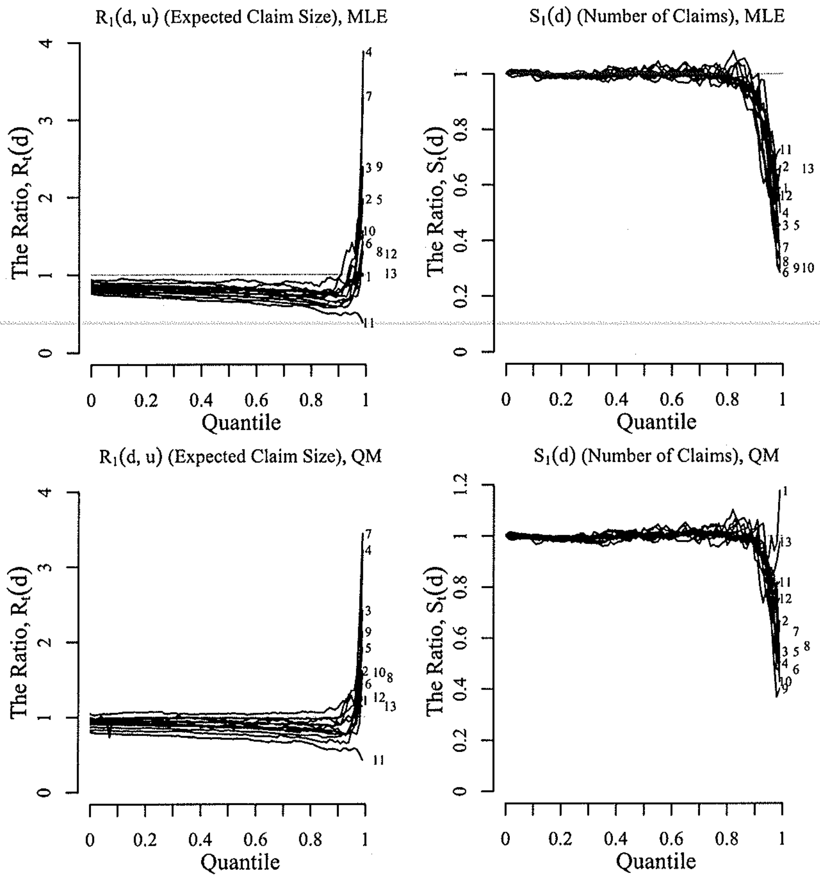

Figure 2: Steps in Density Estimation Using the CMC Transformation with QM Parameter Estimates for Trade Group 1

Step 2: Transform the data set $X_1^i, ..., X_{N_i}^i$ into $Z_1^i, ..., Z_{N_i}^i$ using $Z_j^i = \widehat{T}_i(X_j^i)$ for $j = 1, ..., N_i$ where $T_{\widehat{\alpha}_i,\widehat{M}_i,\widehat{c}_i}(x) \equiv \widehat{T}_i(x)$ is given in equation (2). Figure 2(2) shows the histogram for the transformed trade group 1 data.

Step 3: If the unknown distribution $F_i(x)$ is a modified Champernowne distribution, the transformed data set will be uniformly distributed.[4] Even if $F_i(x)$ is not a modified Champernowne distribution, however, the transformed data set is usually close to a uniform distribution because the modified Champernowne distribution is fitted to the data set. Under the assumption that the transformed distribution is close to a uniform distribution on $(0, 1)$, we can use a constant bandwidth when computing the correction estimator by means of a classical kernel density estimator for $Z_1^i, ..., Z_{N_i}^i$:

$$\widehat{g}_i(z) = \frac{1}{N_i \cdot k_{b_i}(z)} \sum_{i=1}^{N_i} K_{b_i}(z - Z_i^i) \tag{5}$$

where $K_{b_i}(\cdot)$ is the Epanechnikov kernel function defined in equation (8) and $k_{b_i}(z)$ is the boundary correction, which is needed because the Z_j^i's are constrained on the interval $(0, 1)$. The boundary correction $k_{b_i}(z)$ is defined as

$$k_{b_i}(z) = \int_{\max\left(-1,-\frac{z}{b_i}\right)}^{\min\left(1,\frac{1-z}{b_i}\right)} K(u)\,du.$$

The kernel estimator is illustrated in Figure 2(3). Notice that near 0, the kernel estimator is below 1, which means that the resulting estimator for f_1 is lower than the density of the estimated modified Champernowne distribution from Step 1. In the interval from 0.25 to 0.6, the kernel estimator is above 1, which means that the kernel estimator has raised the modified Champernowne distribution.

Step 4: The kernel estimator, $\widehat{g}_i$, can be interpreted as the final estimator on the transformed axis. The estimated density for the original data set $X_1^i, ..., X_{N_i}^i$ is obtained by an inverse transform such that

[4] Uniformity can be tested with a chi-square test or Kolmogorov-Smirnov test.

$$\hat{f}_i(x) = \frac{\hat{g}_i\left(\hat{T}_i(x)\right)}{\left|\frac{d\hat{T}_i^{-1}}{dz}\Big|_{z=\hat{T}_i(x)}\right|}. \tag{6}$$

The resulting estimator for the data from trade group 1 is shown in Figure 2(4). The corrected modified Champernowne estimator (solid line) seems to provide a better estimate for the data set than the uncorrected modified Champernowne distribution (dotted line) from Step 2.

These steps can be summarized into the following expression for the final estimator for f_i:

$$\hat{f}_i(x) = \frac{1}{N_i k_{b_i}\left(\hat{T}_i(x)\right)} \sum_{j=1}^{N_i} K_{b_i}\left(\hat{T}_i(x) - \hat{T}(X_i)\right) \hat{T}_i'(x). \tag{7}$$

As mentioned in Step 3, the Epanechnikov kernel function is used in the kernel estimator. This kernel function is the optimal kernel with respect to efficiency (Wand and Jones, 1995, page 31), i.e., for a fixed number of observations, the Epanechnikov kernel function leads to a better kernel estimator than any other kernel function. The Epanechnikov kernel function has the form

$$K(x) = \begin{cases} \frac{3}{4}\left(1 - x^2\right) & \text{if } -1 < x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

and for bandwidth b,

$$K_b(x) = \frac{1}{b} K\left(\frac{x}{b}\right).$$

The choice of bandwidth determines the smoothness of the estimator. The simple normal scale bandwidth selection is used (Wand and Jones, 1995, page 60):

$$b = \left(\frac{40\pi^{\frac{1}{2}}}{N}\right)^{\frac{1}{5}} \hat{\sigma}$$

where N is the number of observations and $\hat{\sigma}$ is the standard deviation; this is optimal when f is a normal distribution. For fixed $\hat{\sigma}$, the bandwidth is decreasing when N increases, and vice versa. Thus, a small data set results in a large bandwidth and a great amount of smoothing in the kernel estimator, and hence a small correction. This ensures that the final estimator $\hat{f}(x)$ is close to the modified Champernowne distribution from Step 1. A large data set results in a small bandwidth and, hence, a potentially stronger correction by the kernel estimator to the modified Champernowne distribution from Step 1. The asymptotic behavior of the transformation kernel density estimator is described in Buch-Larsen et al. (2005).

Table 2 shows the values of the Kolmogorov-Smirnov tests for the modified Champernowne distributions MC_{MLE} and MC_{QM} from Step 1 and the corresponding CMC distributions CMC_{MLE} and CMC_{QM} are stated for each trade group. In almost all trade groups, the test does not reject the modified Champernowne distribution from Step 1 with MLE parameters, whereas the QM parameters result in a rejection in more than half of the trade groups, using 0.05 as the rejection threshold. This confirms the well-known result that MLE produces the best overall fit. However, the test neither rejects the kernel-smoothed CMC_{MLE} estimates with MLE parameters, nor the CMC_{QM} estimates with QM parameters.

Table 2
Kolmogorov-Smirnov Tests for Corrected (CMC) and Uncorrected Modified Champernowne (MC)

Trade Group i	MC_{MLE}	MC_{QM}	CMC_{MLE}	CMC_{QM}
1	0.005	0.009	0.481	0.550
2	0.248	0.010	0.620	0.336
3	0.417	0.065	0.535	0.531
4	0.484	0.159	0.559	0.487
5	0.519	0.176	0.408	0.582
6	0.085	0.018	0.597	0.516
7	0.279	0.090	0.354	0.413
8	0.087	0.038	0.519	0.495
9	0.619	0.184	0.600	0.475
10	0.073	0.000	0.437	0.430
11	0.403	0.253	0.526	0.592
12	0.103	0.013	0.383	0.632
13	0.066	0.002	0.548	0.599

Next we demonstrate the calculation of conditional means. To avoid numerical problems,[5] all calculations are performed on the transformed

[5]Problems often arise in numerical integration over the interval 0 to ∞ (we assume the integral is convergent). Some (but not all) of these problems can eliminated by an

axis. We first estimate the conditional densities of losses from group i given that they are larger than the deductible. Let $F_j(x|X_j^i > d) = \mathbb{P}\left[X_j^i \leq x|X_j^i > d\right]$. It follows that

$$\hat{F}_j(x|X_j^i > d) = \frac{\int_d^x \hat{f}_i(y)\mathrm{d}y}{\int_d^\infty \hat{f}_i(y)\mathrm{d}y} = \frac{\int_{\hat{T}_i(d)}^{\hat{T}_i(x)} \hat{g}_i(z)\mathrm{d}z}{\int_{\hat{T}_i(d)}^{1} \hat{g}_i(z)\mathrm{d}z} \tag{9}$$

where $\hat{g}_i(z)$ is the classical kernel density estimator given in equation (5) and $\hat{f}_i(x)$ is defined in equation (6).

Let $X_j^i(d,u)$ denote the insurer's actual loss paid by the insurer that results from the loss X_j^i given a deductible d and a policy limit u, then

$$\mathbb{E}[X_j^i(d,u)] = \frac{\int_d^u (x-d)\hat{f}_i(x)\mathrm{d}x + (u-d)\int_u^\infty \hat{f}_i(x)\mathrm{d}x}{\int_d^\infty \hat{f}_i(x)\mathrm{d}x} \tag{10}$$

$$= \frac{\int_{\hat{T}(d)}^{\hat{T}(u)} \hat{T}^{-1}(z)\hat{g}_i(z)dz + u\int_{\hat{T}(u)}^{1} \hat{g}_i(z)dz}{\int_{\hat{T}(d)}^{1} \hat{g}_i(z)dz} - d \tag{11}$$

In order to test the goodness of fit, we will now compute $R_i(d,u)$ and $S_i(d)$, which are ratios of estimated and observed expected conditionals for each trade group, i.e.,

$$R_i(d,u) = \frac{\mathbb{E}[X_j^i(d,u)]}{\overline{X}_j^i(d,u)} \quad \text{and} \quad S_i(d) = \frac{\mathbb{E}[N_j^i(d)]}{\overline{N}_j^i(d)} \tag{12}$$

where, for trade group i with deductible d and policy limit u, $\overline{X}_j^i(d,u)$ is the observed conditional expected loss, $N_j^i(d)$ is the number of losses in excess of d, and $\overline{N}_j^i(d)$ is the observed number of losses in excess of d. Figure 3 shows plots of $R_1(d,u)$ and $S_1(d)$ for various values of d and $u = 5{,}000{,}000$. The parameters are estimated by means of the MLE method in the two upper plots and by means of the QM method in the two lower plots.

---

appropriate transformation so that the integration is done over the interval 0 to 1. For more on numerical integration see, for example, Ralston and Rabinowitz (1978, Chapter 4).

Table 3
Conditional Expected Losses for
Corrected Modified Champernowne (CMC)
Under QM Method with Policy Limit
$u = 5,000,000$ and Various Deductibles

Trade Group i	Deductibles 0	25,000	50,000	100,000
1	46,395	103,932	158,935	247,935
2	32,272	69,969	109,668	175,914
3	20,165	59,234	97,517	170,610
4	19,717	55,965	87,462	143,640
5	18,350	44,742	73,808	132,056
6	18,469	53,439	79,196	128,825
7	27,659	82,559	132,448	217,471
8	17,954	44,303	69,050	117,155
9	21,805	62,074	101,939	169,801
10	18,882	47,763	72,662	120,355
11	22,930	49,061	88,242	163,350
12	23,759	54,219	81,856	130,384
13	32,430	88,206	138,624	216,229
	250,000	500,000	1,000,000	2,500,000
1	476,618	783,787	1,207,821	1,519,513
2	357,761	619,579	1,013,470	1,399,530
3	370,772	651,681	1,062,555	1,435,935
4	302,572	539,661	913,017	1,331,167
5	298,340	542,768	924,039	1,342,388
6	274,763	496,533	855,227	1,288,418
7	439,452	737,475	1,157,059	1,490,929
8	257,922	472,998	825,105	1,266,293
9	357,251	623,971	1,023,078	1,408,028
10	262,505	479,694	833,971	1,273,022
11	365,036	647,448	1,060,089	1,435,471
12	273,758	492,010	846,952	1,281,096
13	425,908	714,913	1,128,817	1,473,290

Table 4
Observed Average Losses ($\overline{X}_j^i(d,u)$) with Policy Limit $u = 5,000,000$ and for Various Deductibles

Trade Group i	Deductibles			
	0	25,000	50,000	100,000
1	44,435	99,421	150,588	208,369
2	35,084	80,771	124,326	207,293
3	21,469	66,863	102,769	147,010
4	20,515	62,918	79,133	116,311
5	20,145	55,599	91,734	114,229
6	21,268	73,225	103,454	150,448
7	28,320	86,489	148,584	172,529
8	19,554	54,378	88,113	107,760
9	26,281	92,743	153,164	213,622
10	20,813	59,815	94,689	156,765
11	32,765	97,685	202,911	389,410
12	24,865	60,025	92,774	133,077
13	34,128	97,010	152,635	220,197
	250,000	500,000	1,000,000	2,500,000
1	364,572	660,494	744,944	242,939
2	279,611	359,043	180,221	0
3	168,918	267,415	0	0
4	89,598	0	0	0
5	124,775	358,410	0	0
6	196,683	198,835	0	0
7	193,729	191,193	0	0
8	152,140	154,640	33,850	0
9	224,949	351,758	533,632	0
10	190,388	209,242	200,246	0
11	1,699,379	2,124,883	3,022,845	6,792,342
12	209,587	850,056	803,610	464,448
13	441,802	835,375	1,592,551	4,550,394

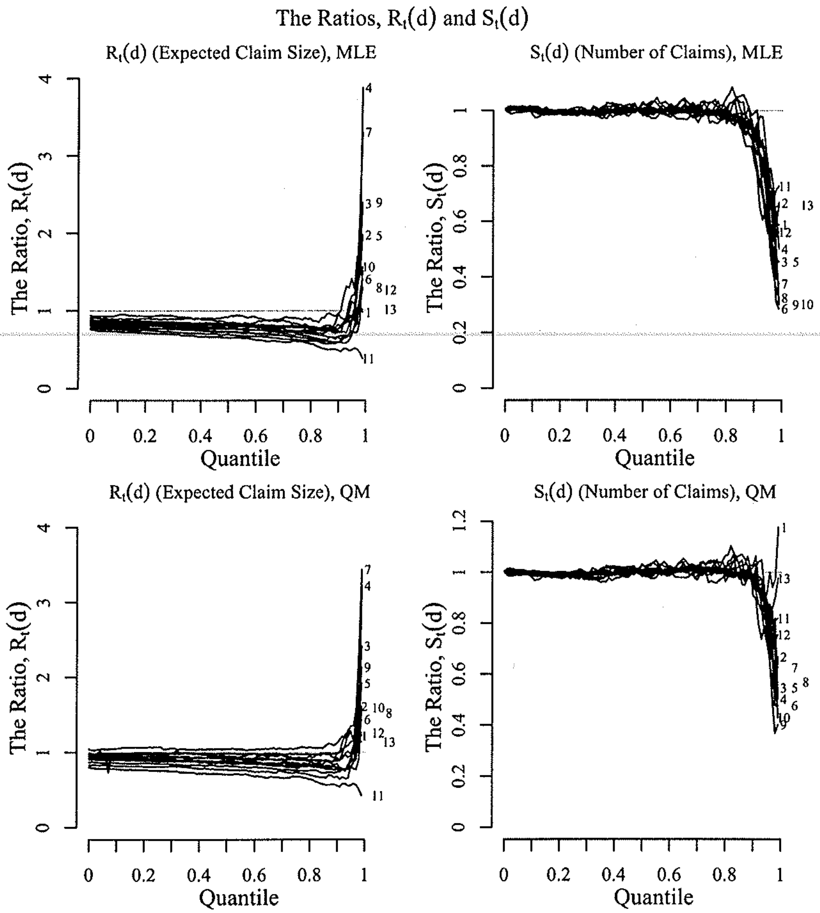

Figure 3: Comparing Ratios $R_1(d,u)$ (Left Plots) and $S_1(d)$ (Right Plots) Using MLE and QM Methods Versus Quantiles

The plots of $S_t(d)$ show that both the MLE and QM parameters result in reasonable estimates of the number of observations. However, the plots of $R_t(d)$ show that the MLE parameters lead to underestimation of the expected loss in all trade groups, whereas the QM parameters are slightly better in this respect. This may be because MLE estimation assigns equal weight to small and large losses, whereas QM estimation places more emphasis on the tail, which has the biggest effect on the estimated loss. Thus, insurers would be wise to choose estimation methods that put greater emphasis on the tail losses. Notice that the bottom half of Figure 3 shows that the underestimation of the conditional mean is less distinct for the CMC_{QM}.

The CMC_{QM} estimators are therefore used to estimate the conditional expected loss for each trade group and for various deductibles; they are shown in Table 3, while the actual observed average losses are in Table 4. For a general insurance company, these statistics can be used to estimate the rates within each trade group.

To continue this illustration, let us compare the corrected modified Champernowne estimation procedure with the *generalized Pareto distribution* approach (GPD) as exemplified by Cebrián, Denuit, and Lambert (2003). A loss from trade group i is said to follow a generalized Pareto distribution if its cdf is given by

$$F_i(x) = \begin{cases} 1 - (1 + \xi_i x)^{-\frac{1}{\xi_i}} & \text{if } \xi_i \neq 0 \\ 1 - e^{-x} & \text{if } \xi_i = 0 \end{cases} \tag{13}$$

for $\xi_i, x > 0$.

According to Cebrián, Denuit, and Lambert (2003), we must find the threshold u separating small and large losses by means of one or more graphical tools: (i) an empirical mean excess function plot, (ii) a GPD index plot, or (iii) a Gertensgarbe plot. In the empirical mean excess function plot, the empirical mean excess function is approximately linear for $x \geq u$. In the GPD index plot, we compute the maximum likelihood estimator for increasing thresholds and identify u as the point from which the MLE estimator becomes approximately constant. The Gertensgarbe plot is based on the assumption that the extreme threshold can be found as a change point in the ordered series of claims and that the change point can be identified by means of a sequential version of the Mann-Kendall test as the intersection point between a normalized progressive and retrograde rank statistics. The progressive and retrograde curves in the Gertensgarbe plot, however, do not in all cases produce an intersection point: in particular, our data set did not lead to an intersection point, and our choice of thresholds is therefore based on the first two methods. Figure 4 shows the GPD index plot and the empirical mean excess plot for trade group 1. In the GPD index plot the chosen threshold corresponds to the 85% quantile where there are 251 observations exceeding the threshold. In the empirical mean excess plot the chosen threshold is 53,571.[6] Table 5 shows the chosen thresholds in quantile terms (u_{quan}), in absolute terms (u_{value}), and in number of observations exceeding the threshold (u_{exc}), as well as the estimated GPD parameters, and the Kolmogorov-Smirnov test probabilities. Table 5 shows that the estimated GPD's are not rejected by the Kolmogorov-Smirnov tests in any trade group.

Table 5
Thresholds, Estimated Parameters, and Kolmogorov-Smirnov Tests for Generalized Pareto Distribution

Trade	Thresholds			Parameters		K-S
Group i	u_{quan}	u_{value}	u_{exc}	$\hat{\xi}$	$\hat{\beta}$	Test
1	85.0%	53,571	251	0.576	72,494	0.696
2	56.0%	14,621	263	0.876	15,348	0.629
3	90.5%	39,040	201	0.537	48,625	0.769
4	88.0%	28,840	65	0.309	50,974	0.810
5	95.3%	68,107	51	0.149	91,930	0.760
6	90.5%	38,897	196	0.525	49,691	0.570
7	91.0%	48,315	64	0.318	102,541	0.642
8	94.0%	54,866	218	0.257	68,954	0.567
9	95.5%	96,062	42	0.210	164,404	0.770
10	88.0%	31,888	755	0.612	32,577	0.434
11	84.0%	28,339	164	0.787	22,821	0.645
12	95.0%	87,678	284	0.372	75,536	0.490
13	90.0%	57,966	820	0.538	73,313	0.612

Table 6 displays the conditional means for various deductibles using the estimated GPD parameters. If we compare the conditional expected losses estimated by means of GPD and CMC_{QM} in Tables 6 and 3, respectively, with the observed conditional expected losses in Table 4, we notice that the GPD estimates are closer to the observed means in approximately half of the trade groups, the CMC_{QM} estimates are closer

[6]Analogous plots for the remaining trade groups are available from <http://www.math.ku.dk/~tbl/joap06.html.>

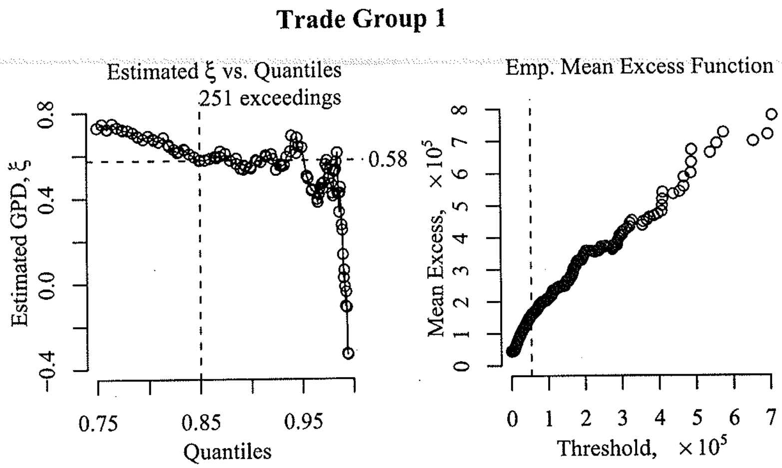

Figure 4: GPD Index Plot (left) and Empirical Mean Excess Plot (right) for Trade Group 1

in three others, and the GPD and CMC_{QM} estimates are similar in the others. GPD estimation, however, has some obvious disadvantages:

- It cannot be used to estimate conditional means when the deductible is smaller than the threshold. In such cases the distribution for small losses must be estimated separately;
- No automatic procedure exists for finding the optimal threshold; and
- The GPD only works for heavy-tailed data. For moderately light tails (such as the lognormal distribution), GPD estimation will often result in an estimator with finite support (Buch-Larsen et al., 2005).

The final phase of the illustration is the validation phase. Whereas a goodness of fit test measures how well the estimation fits claims in the data set, a validation study measures how well the method predicts future claims. Therefore, to get a better comparison of the CMC and GPD methods, the data set is randomly partitioned into two parts: one for estimating model parameters and the other for validation. In other words, the first data set is used to estimate the CMC_{QM} and GPD parameters.

Table 6
Conditional Expected Losses for GPD with Policy Limit $u = 5,000,000$ and Various Deductibles

Trade Group i	Deductibles 0	25,000	50,000	100,000
1	$< u_1$	$< u_1$	$< u_1$	275,744
2	$< u_2$	147,621	217,969	342,505
3	$< u_3$	$< u_3$	155,549	207,875
4	$< u_4$	$< u_4$	96,106	118,417
5	$< u_5$	$< u_5$	$< u_5$	125,509
6	$< u_6$	$< u_6$	153,549	203,988
7	$< u_7$	$< u_7$	172,940	195,924
8	$< u_8$	$< u_8$	$< u_8$	127,276
9	$< u_9$	$< u_9$	$< u_9$	234,495
10	$< u_{10}$	$< u_{10}$	149,165	215,580
11	$< u_{11}$	$< u_{11}$	197,056	299,797
12	$< u_{12}$	$< u_{12}$	$< u_{12}$	178,230
13	$< u_{13}$	$< u_{13}$	$< u_{13}$	255,703

	250,000	500,000	1,000,000	2,500,000
1	435,556	668,774	1,027,518	1,386,478
2	642,293	1,005,860	1,453,782	1,654,455
3	356,409	577,695	929,360	1,326,089
4	185,071	294,617	502,810	920,995
5	151,751	195,476	282,747	526,584
6	347,741	563,296	909,159	1,310,492
7	264,350	375,953	584,624	977,379
8	178,946	264,538	431,202	808,021
9	274,149	339,849	468,285	768,983
10	398,335	658,551	1,047,326	1,415,323
11	559,587	892,057	1,327,369	1,586,321
12	264,357	402,829	653,364	1,072,024
13	401,255	617,947	961,323	1,338,851

Notes: $< u_i$ denotes the deductible is smaller than the threshold.

These estimated parameters are then used to calculate conditional expected losses under the CMC_{QM} and GPD methods, which are then compared to the observed conditional expected losses contained in the second data set. The validation study shows that in terms of prediction, which is essential for a general insurance company, the CMC_{QM} performs as well as the GPD method. The results from these validation comparisons are available from
`<http://www.math.ku.dk/~tbl/joap06.html>`.

4 Summary and Closing Comments

When dealing with heavy-tailed loss distribution data, maximum likelihood estimation of parameters tends to lead to an underestimation of conditional expected losses. For this reason, an alternative, called the quantile-mean method (QM) of parameter estimation, was introduced. The Buch-Larsen et al. (2005) corrected modified Champernowne method (CMC) is combined with the QM method to produce decent results. Comparing the CMC method with the generalized Pareto distribution (GPD) method shows that the GPD performs better than the CMC in terms of goodness of fit, whereas our validation study shows that the two methods are comparable in terms of predicting future claims.

The CMC method also has some advantages that makes it an attractive alternative compared to GPD: The CMC method estimates the density of the whole range of losses, whereas in GPD estimation, we need to estimate small and large losses separately, which involves finding a threshold from where the data set is GPD. This is normally done by graphical methods, which are difficult to automatize. Finally, the GPD can only be used for heavy-tailed distributions, whereas the CMC also works for lighter-tailed distributions because it always has infinite support.

One area for further research is in improving the parameter estimation method and including more sophisticated boundary correction methods. For example, one can combine our work with the methods proposed by Chen (1999 and 2000) and Scaillet (2004). We also hope to integrate insights from recent developments in density estimation, such as Hagmann and Scaillet (2004), and to extend our estimation method to handle covariates.

References

Bolancé, C., Guillén, M., and Nielsen, J.P. "Kernel Density Estimation of Actuarial Loss Functions." *Insurance: Mathematics and Economics* 32 (2003): 19-36.

Brown, E.H.P. "Report of the Oxford Meeting September 25-29, 1936." *Econometrica* 5, no. 4 (1937): 361-383.

Buch-Larsen, T., Nielsen, J.P., Guillén, M., and Bolancé, C. "Kernel Density Estimation for Heavy-Tailed Distributions Using the Champernowne Transformation." *Statistics* 39 (2005): 503-518.

Cebrián, A.C., Denuit, M., and Lambert, P. "Generalized Pareto Fit to the Society of Actuaries' Large Claims Database." *North American Actuarial Journal* 7, no. 3, (July 2003): 18-36.

Champernowne, D.G. "The Graduation of Income Distributions." *Econometrica* 20, no. 4 (1952): 591-615.

Chen, S.X. "Beta Kernel Estimator for Density Functions." *Computational Statistics and Data Analysis* 31 (1999): 131-145.

Chen, S.X. "Probability Density Function Estimation Using Gamma Kernels." *Annals of the Institute of Statistical Mathematics* 52 (2000): 471-480.

Clements, A.E., Hurn, A.S., and Lindsay, K.A. "Möbius-Like Mappings and Their Use in Kernel Density Estimation." *Journal of the American Statistical Association* 98, no. 466 (2003): 993-1000(8).

Denuit, M., Purcaru, O., and Van Keilegom, I. "Bivariate Archimedean Copula Models for Censored Data in Non-Life Insurance." *Journal of Actuarial Practice* 13 (2006): 5-32.

Embrechts, P., Kluppelberg, C., and Mikosch, T. *Modelling Extremal Events for Insurance and Finance.* Berlin, Germany: Springer Verlag, 1997.

Hagmann, M. and Scaillet, O. "Local Multiplicative Bias Correction for Asymmetric Kernel Density Estimators." *Royal Economic Society Annual Conference 2004*, 25. Royal Economic Society, 2004. Available online at <http://repec.org/res2004/HagmannScaillet.pdf>

Hjort, N.L. and Glad, I.K. "Nonparametric Density Estimation with a Parametric Start." *Annuals of Statistics* 23, no. 3 (1995): 882-904.

Jones, M., Linton, O., and Nielsen, J.P. "A Simple Bias Reduction Method for Density Estimation." *Biometrica* 82, no. 2 (1995): 327-338.

Lehmann, E.L. *Theory of Point Estimation.* Belmont, CA: Wadsworth and Brooks Cole, 1991.

Ralston, A. and Rabinowitz, A. *First Course in Numerical Analysis.* Tokyo, Japan: McGraw-Hill, 1978.

Scaillet, O. "Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels." *Journal of Nonparametric Statistics* 16 (2004): 217-226.

Wand, M.P. and Jones, M.C. *Kernel Smoothing.* Boca Raton, FL: Chapman & Hall/CRC, 1995.

Wand, M.P., Marron, J.S., and Ruppert, D. "Transformations in Density Estimation." *Journal of the American Statistical Association* 86, no. 414 (1991): 343-361.

Yang, L. and Marron, J.S. "Iterated Transformation-Kernel Density Estimation." *Journal of the American Statistical Association* 94, no. 446 (1999): 580-589.