

## Using External Data in Operational Risk

Montserrat Guillen<sup>a</sup>, Jim Gustafsson<sup>b</sup>, Jens Perch Nielsen<sup>c</sup> and Paul Pritchard<sup>d</sup>

<sup>a</sup>University of Barcelona, Econometrics, RFA-IREA, Diagonal 690, Barcelona E-08034, Spain.  
E-mail: mguillen@ub.edu

<sup>b</sup>Codan Insurance, Copenhagen, Denmark.

E-mail: jgu@codan.dk

<sup>c</sup>University of Copenhagen & Festina Lente, Copenhagen, Denmark.

E-mail: festinalente@nielsen.mail.dk

<sup>d</sup>Royal&SunAlliance, London, U.K.

E-mail: paul.pritchard@gcc.royalsun.com

We present a method to combine expert opinion on the likelihood of under-reporting with an operational risk data set. Under-reporting means that not all losses are identified and therefore an incorrect distributional assumption may be made, and ultimately an incorrect assessment made of capital required. Our approach can be applied to help insurers and other financial services companies make better assessments of capital requirements for operational risk using either external or internal sources. We conclude that operational risk capital evaluation can be significantly biased if under-reporting is ignored.

*The Geneva Papers* (2007) 32, 178–189. doi:10.1057/palgrave.gpp.2510129

**Keywords:** operational risk; under-reporting function; loss data; capital requirements

---

### Introduction

Many financial services companies are now utilizing loss data for the purposes of calculating operational risk capital requirements, potentially arising from either regulatory requirements or indeed from a desire to integrate capital sensitive management within their organizations. Instinctively, the use of internal loss experience directly or as a means of deriving distribution parameters from which simulations can be made is most appealing. However, several factors mitigate against its effectiveness when considered alone: firstly, the data is backward looking based on historical events – the company profile may have changed, and should any large losses have occurred it is likely that controls will have been improved to prevent a reoccurrence. A greater problem nonetheless is that the regularly encountered losses may provide limited information on the size and frequency of large, rarely occurring losses that are the major factor in determining capital requirements. With this in mind organizations have recognized the value of obtaining loss data from outside their company, either through data sharing consortia or through publicly reported losses. In this paper, we focus on the use of publicly reported (also commonly called external) loss data to supplement internal loss experience.

Utilizing publicly reported operational losses for modeling purposes involves a number of considerations. The losses chosen for modeling should be representative of the organization as far as is practicable. This should facilitate the use of both

internal and external data in the modeling since they could then show similar behavior in terms of size distribution (i.e., they come from the same distribution). This might involve, for example, filtering processes whereby events are removed that are not representative of the company in question (i.e. they could not happen to the company). For example, certain losses may be associated with a particular business activity, geography or even size. Even after this filtering the remaining events do not necessarily convey information about the control standards being applied. It might, therefore, involve some kind of scaling process to reflect the control standard in the organization using the data.

Assuming, however, that the filtering process has been applied rigorously, further challenging issues arise in the use of publicly reported data. These concern the quality and classification of reported losses and the notion of under-reporting. While it is possible for any organization to collect and use loss data taken directly from public sources such as newsfeeds, the challenges associated with ensuring data quality (including updates) and consistent categorization are not inconsiderable. Commercial database vendors can apply suitable quality assurance protocols to provide comfort on these matters. The second issue, that of under-reporting, is more difficult since not all losses will necessarily reach the public domain. Furthermore, the likelihood of the loss being reported could vary by both risk type and by size. For example, an incident involving legal proceedings will almost certainly be reported, similarly large losses intuitively are more likely to reach the public domain. The effect of this under-reporting is that any capital calculation that does not account for this factor could be skewed towards larger losses and therefore could result in extra capital being held.

However, also the frequency of the reported losses is affected by under-reporting. By ignoring under-reporting the high frequently small losses are not included in the capital held, which could add up to a huge amount, and in a worse case scenario lead to an insolvent business. The under-reporting effect is different from the stated minimum reporting threshold in a database, for example  $\sim$  £0.5 million (\$1 million) which could be accounted for relatively easily.

An approach for dealing with under-reporting has been described by Frachot and Roncalli<sup>1</sup> and Baud *et al.*,<sup>2</sup> who regard the problem as being one in which the publicly reported data is subject to an unknown (and varying) lower reporting threshold. In this paper, we propose an alternative, simpler approach, drawing on the input of risk experts opinion in the estimation of under-reporting to derive estimators of the underlying true publicly reported loss data size distribution. Subject matter experts within an internationally active insurance company were asked to provide percentage estimates of reporting likelihood across the Basel operational risk categories at different loss sizes. These factors are then used to derive an under-reporting function that can be combined with the publicly reported data to give an estimator of the underlying true loss distribution. Requirements that apply to the under-reporting function are that it must pass through certain points (reflecting the responses given by the subject matter experts), that it should be a continuous function, and that it should

---

<sup>1</sup> Frachot and Roncalli (2002).

<sup>2</sup> Baud *et al.* (2002).

be mathematically well behaved beyond the final reporting point. The mathematical approach applied in this situation is termed spline theory, which seeks to interpolate values between the reported values (threshold points). Beyond the final reporting value, we use extrapolation. More precisely, we utilize a cubic spline approximation, fitting piecewise cubic polynomials between the fixed report levels (termed knots in spline theory). As a result, the approach provides an indication on the true size of the losses. The exact same function is also involved in the frequency setup, but here in a slightly different form, as seen in the forthcoming sections. For an updated overview about quantifying operational risk in general insurance company, see Tripp *et al.*<sup>3</sup> and the discussion therein.

The strategy is to first formulate the under-reporting functions on each event risk category based on the judgement of the risk experts opinion on the likelihood of reporting losses. Thereafter, a distribution is estimated for the available data combined with the under-reporting information to produce an estimate of the true loss distribution that counteracts the impact of under-reporting. For illustration, we present an application utilizing publicly reported loss data and introduce different true distributions as benchmarks. We compare the different loss distributions with and without the under-reporting effect and evaluate the consequences for operational risk capital requirements in the different scenarios.

### **The under-reporting function**

When modeling a publicly reported database one needs to fit the operational risk exposure to a company's own situation. This is done here by introducing an under-reporting function into the model, which could be thought of as a scaling function explaining the likelihood of a loss being reported at different threshold levels. A very simple interpretation of how the function affects the distribution is that in the beginning and main body is up-weighted to include more probability mass in that domain, while larger losses are down-weighted.

The under-reporting function should, as well as being continuous, pass exactly through the predetermined reporting level values. The requirement is mathematically termed as interpolating the report levels, which is a prediction made through and between the report levels. A further requirement is to have a continuous and well-defined function after the last observed report level. Mathematically, this is defined as extrapolation – a prediction made beyond the last observed report level. Of course, a prediction based upon extrapolation is potentially less reliable than a prediction between report levels since one must assume an historical pattern will continue outside the reported value domain.

Spline theory meets these requirements. We utilize a cubic spline approximation constructed by fitting piecewise cubic polynomials between the report levels (the prediction beyond is captured linearly). A summary is provided in the Appendix and deeper treatment of the subject can be found in Fan and Gijbels<sup>4</sup> and de Boor.<sup>5</sup>

---

<sup>3</sup> Tripp *et al.* (2004).

<sup>4</sup> Fan and Gijbels (1996).

<sup>5</sup> de Boor (2001).

## The estimated under-reporting function

The Basel operational risk categories were used as a standard basis from which to estimate the under-reporting functions. Several subject matter risk experts from a major international insurance company provided estimates of under-reporting.<sup>6</sup> This lead to percentage estimates of the reporting likelihood at different predetermined loss sizes. The risk event categories are shown in Table 1.

Data provided by the expert opinion on the probability of reporting a loss of a given amount for each type of risk has been used to produce the under-reporting functions. Figure 1 presents the estimated probability of reporting for each risk category. It can be seen that the event risk category 7 (Execution, Delivery and Process Management) has the lowest reporting likelihood, which means that losses for this kind of operational risk are likely to be under-reported. On the other hand, smaller losses that belong to category 5 (Clients, Products and Systems Failure) have a much higher probability of being reported than small losses for other types of risks. Also noteworthy is that for all event risk categories, the expert opinion assesses that the probability of reporting a loss is smaller than or equal to 99 per cent, even for very large losses. In fact, this acts as a bound, and it implies that there is at least a 1 per cent chance that a very large loss occurs and is not reported. The estimation of the under-reporting function is essential in our approach, but unfortunately very little is known about the magnitude of under-reporting and even less on the validity of experts' estimation. We think that further research needs to be done on the stability of these estimates and we believe that here we provide a starting point for discussion.

## A parametric distribution for loss data

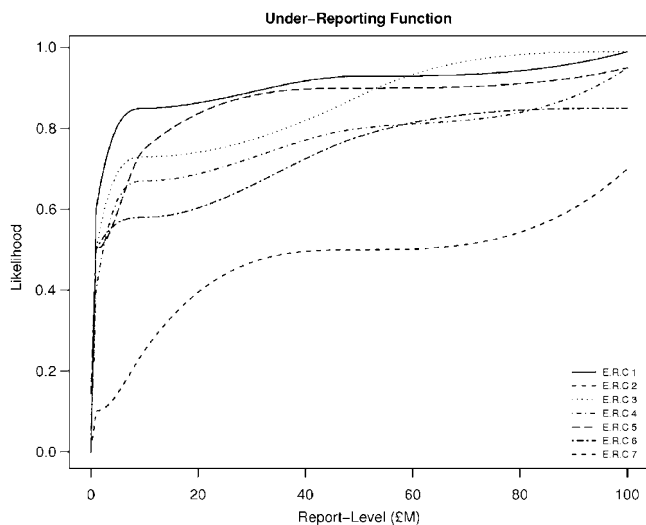
As a starting point we use a probability distribution for the true operational risk losses. Since we do not have full public reporting of all losses, the publicly available

**Table 1** Event risk categories

- 
1. Internal fraud
  2. External fraud
  3. Employment practices and workplace safety
  4. Business disruption
  5. Clients, products and system failures
  6. Damage to physical assets
  7. Execution, delivery and process management
- 

---

<sup>6</sup> Eight top risk managers from different regions (Ireland, Canada, U.K., Scandinavia, etc.) were interviewed. The experts currently practice within an international insurer and are exposed to routine losses within the organization as well as having access to the vendors' data set and any other information in the public domain. They have all been involved with estimating severities and frequencies as part of the Advanced Methodology approach (scenario analysis) operational risk capital calculation for regulatory purposes. This was done initially as part of group activities to arrive at severities and frequencies – for this paper they responded on under-reporting as individuals.



**Figure 1.** The estimated under-reporting functions explaining the likelihood that a loss being reported on each event risk category (ERC).

loss data will not follow the true distribution. We assume two benchmark parametric distributions (lognormal and Weibull) that are potentially suitable and widely used in the industry for the true losses. We also consider the generalized three-parameter Champernowne distribution whose characteristics are appealing when working with operational risk losses. The distribution shows a similar shape to the lognormal distribution in the beginning and middle part of the domain and a convergence to the heavy tail Pareto distribution for very large values (more details are given in the Appendix).

### Publicly reported loss data

The publicly reported loss database used for this exercise comprises information on over 10,000 events including financial loss suffered, dates, locations, description and loss category assignment across a range of global organizations. It is subject to quality assurance and continual updating by the vendor.<sup>7</sup> Of course, an organization could compile its database using newsfeeds if it wished. There is a considerable variation in both the numbers and sizes of losses reported across categories. In Table 2, we give a summary on each event risk categories statistics.

The second column shows the number of observations for each event risk category. There is considerable variation in the number of losses; event risk categories 5 and 6 have 6,526 and 2,395 reported losses compared to category 4 that has been exposed to

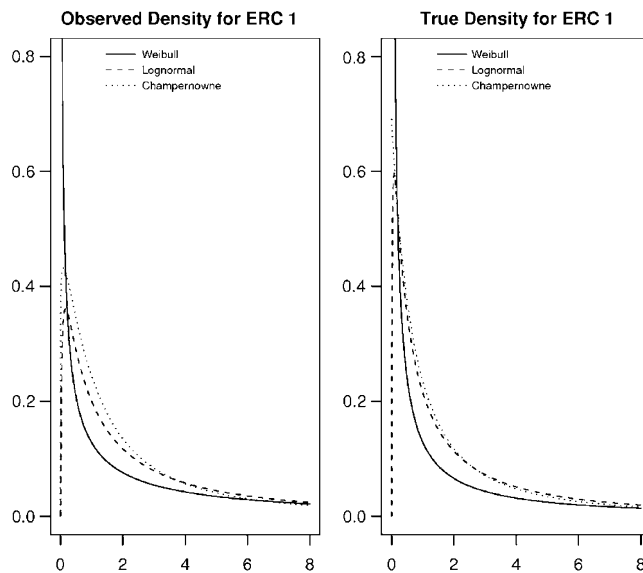
<sup>7</sup> Publicly available data were kindly provided by a specialized firm, which requested not to be cited.

45 losses only. Columns 3–6 show some empirical results on each event risk category. Note that the mean is significantly larger than the median in all cases, consistent with right skewed distributions.

Figure 2 presents the estimated densities on event risk category 1 under different scenarios. Three different probability distributions were fitted (lognormal, Pareto and generalized Champernowne). On the left (observed density for ERC1), there is no correction for under-reporting, while on the right (“true” density for ERC1), the estimation accounts for the existing under-reporting function. For the other six event risk categories, a similar shape is expected.

**Table 2** Descriptive statistics of the seven event risk categories reported losses (£million)

<i>Event risk category</i>	<i>Number of losses</i>	<i>Maximum loss</i>	<i>Sample median</i>	<i>Sample mean</i>	<i>Standard deviation</i>
1. Internal fraud	1,247	6,683.8	1.82	32.24	269.43
2. External fraud	538	910.6	2.14	15.60	69.68
3. Employment practices and workplace safety	721	221.9	1.98	7.84	20.04
4. Business disruption	45	117.6	5.88	22.46	33.25
5. Clients, products and system failures	6,526	11,228.7	3.74	36.59	268.28
6. Damage to physical assets	2,395	39,546.4	2.35	74.91	1,192.55
7. Execution, delivery and process management	75	104.6	1.56	7.39	17.72



**Figure 2.** Estimated densities on event risk category 1. Dotted line represents the generalized Champernowne model, broken line the Lognormal model and solid line the Weibull model. The left-hand plot is the model without under-reporting effects, and the right-hand plot is the model that includes the under-reporting estimates.

## Evaluating an internal operational risk loss distribution based on the publicly reported losses

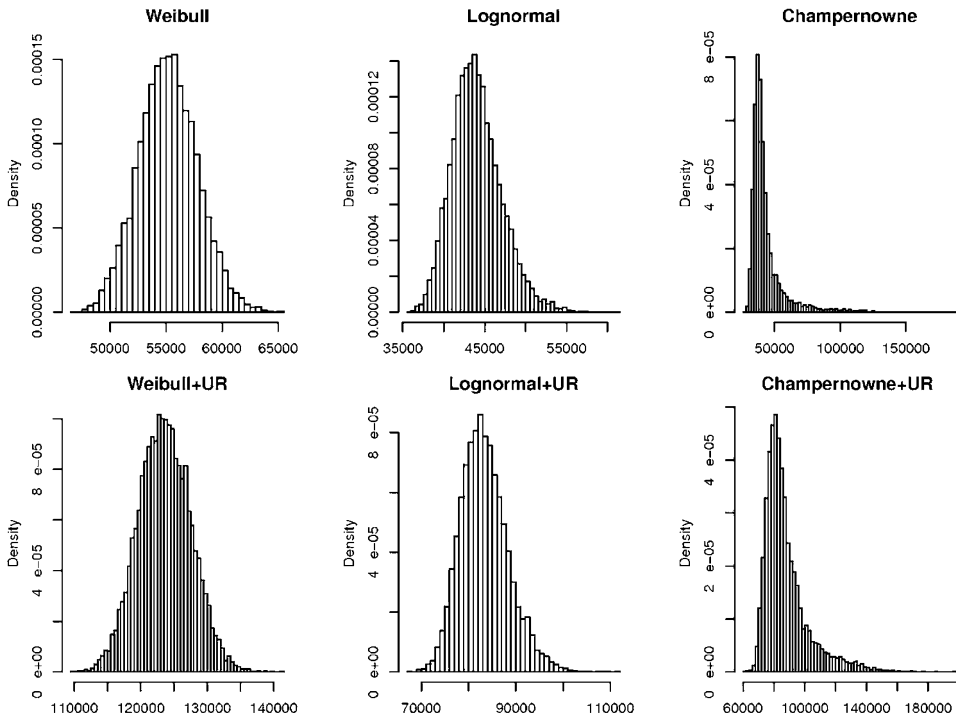
In this section, we simulate an operational risk loss distribution based on both corrected and uncorrected publicly reported losses for every event risk category. This is obtained using a severity distribution (one of the six shown in Figure 2 for event risk category one and correspondingly for all the other events) together with a Poisson-based frequency through 20,000 draws for one-year operational loss. Summation of amounts from the seven event risk categories provides a single value for the simulated distribution. For the severity distributions, we have a total of six different model assumptions, three distributions with and without under-reporting effect. For the frequency, we have four different scenarios, the observed and the corrected frequencies for each probability distribution. Table 3 presents the different frequencies for the different event risk categories.

Figure 3 shows six histograms of operational risk exposure. The top-row is total loss distributions without an under-reporting effect. The second row includes the under-reporting effect. Note that the generalized Champernowne model, both with and without under-reporting, presents a heavy tail characteristic.

Table 4 shows the value at risk (VaR) and tail value at risk (TVaR) with level  $\alpha = \{0.95, 0.99, 0.999\}$  for the three specific models when we account for under-reporting as estimated by the experts. Table 5 shows what happens if it is assumed that there is no under-reporting. For increasing values of  $\alpha$ , the VaR and TVaR increase as expected. It is also shown that when under-reporting is neglected, two of the three models have a similar behavior. Only the generalized Champernowne distribution, which has a heavy-tailed behavior, would imply larger capital requirements, especially for the 0.999 level of significance. Under-reporting brings effects to capital requirements calculation. The greater reporting of lower value losses will lead to estimated distributions which could have lighter tailed characteristics and hence lower capital estimates. On the other hand, where there is under-reporting at higher loss values the effect will be to increase capital required. In our study, it appears that the latter effect outweighs the former; however, the impact in company-specific situations will vary with the specific data set being considered and the degree of under-reporting estimated. Interestingly, the generalized Champernowne model differs significantly

**Table 3** Frequency for each event risk category with and without a correction for under-reporting

<i>Frequency model</i>	<i>Event risk category</i>						
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
Without correction	1,247	538	721	45	6,526	2,395	75
With correction							
Lognormal	1,708	742	1,156	74	10,282	4,318	390
Weibull	1,754	757	1,168	72	11,668	4,501	363
Champernowne	1,798	760	1,206	73	9,720	4,517	419



**Figure 3.** Histograms of total loss distribution using Monte-Carlo simulation with the six scenarios. The first row shows three different models without under-reporting effects. The second shows the models including under-reporting.

from the other two models, because the values for VaR and TVaR are as much as two times the ones obtained for the other two models.

### Conclusions

We have addressed the question of quantifying operational risk when there are two sources of information. One comes from data on the frequency and amount of losses, which may be internal or external. The other source of information is based on the risk experts' opinion on the likelihood of reporting losses with respect to the loss amount. We have used professionals to agree on an estimate of the probability of reporting a loss for several risk categories. We observe that the profiles vary significantly for one type of risk to another. As expected, large amounts are more likely to be reported than small amounts.

We have then looked at the risk evaluation of the sum of risks and have confirmed that the use of a distribution like the generalized Champernowne distribution that is heavy-tailed, substantially increases the estimated capital requirements under the conditions of our study. On the other hand, we have been able to assess that ignoring



**Table 4** Statistics for the loss distribution (£million), with a correction for under-reporting

<i>Model specification</i>	<i>Measure</i>					
	$VaR_{0,95}$	$VaR_{0,99}$	$VaR_{0,999}$	$TVaR_{0,95}$	$TVaR_{0,99}$	$TVaR_{0,999}$
Lognormal	92,411	97,021	102,492	95,271	99,537	106,226
Weibull	130,061	132,951	136,305	131,795	134,425	138,030
Champernowne	118,986	140,265	164,692	132,577	151,203	174,729

**Table 5** Statistics for the loss distribution (£million), without a correction for under-reporting

<i>Model specification</i>	<i>Measure</i>					
	$VaR_{0,95}$	$VaR_{0,99}$	$VaR_{0,999}$	$TVaR_{0,95}$	$TVaR_{0,99}$	$TVaR_{0,999}$
Lognormal	49,392	52,648	55,636	51,306	54,154	57,565
Weibull	54,594	61,560	63,652	60,799	62,584	64,525
Champernowne	72,033	107,711	130,019	92,977	120,264	147,521

the under-reporting phenomenon can lead to a biased estimate of the VaR and TVaR. The approach described here is general in that it could be applied within any organization. In our study, we have found that the overall effect of considering under-reporting is to increase the capital requirement. However, this finding may well not be universal since it appears that two opposing effects are in operation and therefore in different circumstances could lead to a lower capital requirement. The overall effect will depend on both the expert judgement on reporting and the external losses chosen for modeling.

We have offered a way of evaluating the difference between the capital needs to cover operational risk in practice and the capital standards obtained from data that experts believe to be biased. We call for the further consideration of such methods and also for deeper studies on the under-reporting phenomenon.

## References

- Baud, N., Frachot, A and Roncalli, T. (2002) 'Internal Data, External Data and Consortium Data for Operational Risk Measurement: How to Pool Data Properly?', from <http://gro.creditlyonnais.fr/content/wp/oprisk-data-light-version.pdf>.
- Buch-Larsen, T., Nielsen, J.P., Guillen, M. and Bolance, C. (2005) 'Kernel density estimation for heavy tailed distributions using the Champernowne distribution', *Statistics* **39**(6): 503–518.
- Champernowne, D.G. (1937) 'The theory of income distribution', *Econometrica* **5**: 379–381.
- Champernowne, D.G. (1952) 'The graduation of income distributions', *Econometrica* **20**: 591–615.
- de Boor, C. (2001) *A practical guide to splines* (Revised edition), Applied Mathematical Science, Vol. 27, New York, NY: Springer-Verlag.
- Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall.

- Frachot, A. and Roncalli, T. (2002) 'Mixing Internal and External Data for Managing Operational Risk', from <http://gro.creditlyonnais.fr/content/wp/mixing-riskop.pdf>.
- Gustafsson, J., Nielsen, J.P., Pritchard, P. and Roberts, D. (2006) 'Quantifying operational risk guided by kernel smoothing and continuous credibility: A practitioner's view', *The Journal of Operational Risk* **1**(1): 43–57.
- Hjort, N.L. and Glad, I.K. (1995) 'Nonparametric density estimation with a parametric start', *The Annals of Statistics* **23**(3): 882–904.
- Jones, M.C., Linton, O. and Nielsen, J.P. (1995) 'A simple bias reduction method for density estimation', *Biometrika* **82**(2): 327–338.
- Tripp, M.H., Bradley, H.L., Devitt, R., Orros, G.C., Overton, G.L., Pryor, L.M. and Shaw, R.A. (2004) 'Quantifying operational risk in general insurance companies. Developed by a Giro working party', *British Actuarial Journal* **10**(5): 919–1012.

## Appendix

### Piecewise cubic polynomial as under-reporting function

Our approach uses piecewise cubic polynomials for interpolating which are by construction twice-continuous differentiable functions. In order to apply this to our situation, we let  $\{x_{ir}\}_{i=1, r=1}^{k, m_i}$  be the report levels, where  $i=1, \dots, k$  indicates event risk category and  $r=1, \dots, m_i$  specifies the predetermined report levels. Also, we let the partition  $0 = x_{i1} < x_{i2} < \dots < x_{im_i}$  be fulfilled in the domain  $[x_{i1}, x_{im_i}]$ . Furthermore, let the relation  $x_{ir} \mapsto y_{ir}$  explain the corresponding likelihood of a loss being reported, and assuming  $0 = y_{i1} \leq y_{i2} \leq \dots \leq y_{im_i} < 1$  holds. This realizes an under-reporting function that is monotonically increasing. On each interval  $[x_{ir}, x_{ir+1}]$  for  $r=1, \dots, (m_i-1)$  a third-order polynomial is estimated, resulting in a series of  $(m_i-1)$  equations. The compound curve of all piecewise polynomials is then denoted as  $u_i(\cdot)$  for each event risk category  $i$ . The spline method thus results in an underreporting function for each risk event category that is continuous, smooth and passes through all report levels<sup>8</sup>.

### Three probability distributions for loss data

Champernowne<sup>9</sup> first highlighted the distribution in context of economic income. A deeper treatment was presented in Champernowne.<sup>10</sup> Furthermore, Buch-Larsen *et al.*<sup>11</sup> demonstrated in a simulation study that the flexibility of the generalized Champernowne distribution outweighs the advantages of the stability in the original case with scale parameter  $c=0$ , and also discuss the properties of this distribution and its desirable properties in heavy tailed situations. The generalized Champernowne

---

<sup>8</sup> We utilize a linear prediction beyond the last observed report level  $x_{im_i}$  for which we need to find an intercept and a slope. Since the linear function should start from the last observed report level, this point is naturally chosen as the intercept. Utilizing the derivative of the piecewise polynomial corresponding to the intercept point, we find the slope by the expression  $u'_{i(m_i-1)}(x_{im_i})$ .

<sup>9</sup> Champernowne (1937).

<sup>10</sup> Champernowne (1952).

<sup>11</sup> Buch-Larsen *et al.* (2005).

distribution is defined for positive values and with parameter vector  $\theta_i = \{\alpha_i, M_i, c_i\}$  on each event risk category. The generalized Champernowne density takes the form

$$g(x; \theta_i) = \frac{\alpha_i(x + c_i)^{\alpha_i - 1} [(M_i + c_i)^{\alpha_i} - c_i^{\alpha_i}]}{[(x + c_i)^{\alpha_i} + (M_i + c_i)^{\alpha_i} - 2c_i^{\alpha_i}]^2}$$

Let

$$h(x; \eta_i) = \frac{e^{-1/2(\ln x - \mu_i/\sigma_i)^2}}{x\sigma_i\sqrt{2\pi}}$$

define the lognormal density with parameter vector  $\eta_i = \{\mu_i, \sigma_i\}$ , and let

$$\omega(x; \zeta_i) = \frac{\gamma_i}{\beta_i} \left(\frac{x}{\beta_i}\right)^{\gamma_i - 1} e^{-(x/\beta_i)^{\gamma_i}}$$

be the Weibull density with parameter vector  $\zeta_i = \{\gamma_i, \beta_i\}$ . The lognormal and the Weibull are two widely used distributions in operational risk practice and should benchmark the result from the unknown generalized Champernowne distribution.

#### *The correction for under-reporting for severities*

For a correction with under-reporting information on the model assumption, we combine the under-reporting functions discussed in the second section with the parametric distributions for true losses from the previous section. This is obtained by utilizing a multiplicative model defined for each event risk category  $i$  as

$$f(x; \theta_i) = \frac{u_i(x)g(x; \theta_i)}{\int_{\mathfrak{R}_+} u_i(\xi)g(\xi; \theta_i)d\xi}, x \in \mathfrak{R}_+$$

In this structure, the generalized Champernowne density can easily be replaced by the lognormal density  $h(x; \eta_i)$  or the Weibull density  $\omega(x; \zeta_i)$  as required. The resulting density  $f(x; \theta_i)$  is the density of the publicly available information, that is the density estimated on the observed data. By fitting publicly available data, we will be able to find an estimate of the observed parameters ( $\theta_i$  for the modified Champernowne,  $\eta_i$  for the lognormal and  $\zeta_i$  for the Weibull).

The multiplicative construction above is known both from the bias reduction literature<sup>12</sup> and also in the credibility literature.<sup>13</sup> The difference from the latter is that stochastic processes are used in the structure instead of an under-reporting function to estimate the multiplicative error nonparametrically. In the first two references, the multiplicative structure is used as a means to reduce bias while keeping the variance stable on the same order. Note the appearance of the denominator since we want a well-defined estimator, meaning that it should integrate to one. This is often termed as a normalizing constant in the statistical literature. The parameters for the loss distributions are estimated by the maximum likelihood method.

<sup>12</sup> Hjort and Glad (1995) and Jones *et al.* (1995).

<sup>13</sup> Gustafsson *et al.* (2006).

### The correction for under-reporting on frequencies

For the frequency, we assume that the true number of claims  $X_1, \dots, X_{m_i} \forall i$  is *iid* and follows a Poisson distribution with intensity  $\lambda_i$ . The frequency distribution including under-reporting is then defined as follows. Let  $I(j), j=1, \dots, m_i$  be a standard indicator function taking the value zero or one if the  $i$ th claims is not observed or observed, respectively. The observed number of claims are then defined by the summation  $N_i = \sum_{j=1}^{m_i} I(j)$ . If we let  $\tilde{X}_1, \dots, \tilde{X}_{N_i} \forall i$  be the observed number of claims from the publicly reported loss database, the under-reporting is then expressed as  $u_i(x) = P(I(1)=1 | X_1=x)$ . Then we have that the  $\tilde{X}_1, \dots, \tilde{X}_{N_i}$  follows a Poisson distribution with intensity  $\lambda_i P_{i,g}$ , where  $P_{i,g} = \int u_i(w) g(w; \theta) dw$ . If the model assumption is Weibull or lognormal, then we change the density to  $\omega$  or  $h$ . This means that when the frequencies are corrected with under-reporting, the observed intensities are divided by the corresponding  $P_s$ . The authors can provide more details for a deeper mathematical treatment on severity and frequency.

### About the Authors

**Montserrat Guillen** is full professor of econometrics at the University of Barcelona, where she obtained her Ph.D. in Economics and her M.Sc. in Mathematics. She received an M.A. in Risk Management and Insurance from UNED (Spain). She is currently the Director of the research group in Risk in Insurance and Finance at IREA, University of Barcelona.

**Jim Gustafsson** is an actuary and is currently working as the Manager of Research & Development at Codan Insurance in Copenhagen. In combination with his work, he writes a Ph.D. in Applied Mathematics at Copenhagen University, Denmark. He received his B.Sc. in Mathematical Statistics and M.Sc. in Mathematics at Lund University, Sweden.

**Jens Perch Nielsen** is adjunct professor of actuarial science at University of Copenhagen, Denmark, where he obtained his master in actuarial science. He also holds a Ph.D. in Biostatistics from University of California, Berkeley, and he is CEO of the Danish knowledge company Festina Lente.

**Paul Pritchard** has a Ph.D. in biology and is expert in environmental risk and operational risk at Royal&SunAlliance, London.