A Note on Parameter Risk

By Gary Venter, FCAS, MAAA, CERA, and Rajesh Sahasrabuddhe, FCAS, MAAA[1]

June 21, 2012

Contents

1

1 Introduction

For most actuarial modeling applications, model parameters are unknown and must be estimated. If the associated parameter estimation error is not recognized in the modeling, there is a good chance that a substantial portion of the adverse (and favorable) loss potential will appear to be diversified away in the aggregation process.

There is an old fable about buying eggs at 10¢ each and selling them for $1.00 per dozen, making up the difference by doing high volume. The misestimation of the required price is not diversified by volume. Rather, it is a systematic risk that has to be analyzed separately. Similarly parameter risk is a form of systematic risk that does not diversify with volume, although it may diversify across portfolios to some degree.

1.1 Sources of Uncertainty

Parameter risk is the uncertainty as to whether the parameters are appropriate for the phenomenon that we are attempting to model. This uncertainty results from the following factors:

Sampling risk Parameters are estimated from an observed sample. Parameter uncertainty results from differences between that sample and the population.

Data bias Parameters that are used to model outcomes of events that occur during an *exposure period* are estimated from observations from an *experience period*. We often adjust these observations in an attempt to correct for differences between the experience and exposure periods. The most common such adjustment is the trending of claims amounts. This adjustment is intended to remove this bias created by cost level differences. However, if the data are not adjusted correctly then a bias may persist or possibly even be exacerbated. Furthermore, if the amount of the adjustment itself is uncertain, then it should be treated as an additional parameter in the model.

The purpose of this Study Note is to demonstrate that for common approaches for determining mean estimates of actuarial model parameters there exist associated parameter uncertainty models. These uncertainty models are intended to address Sampling Risk. However, this Study Note does not include details regarding the theory and derivation of those uncertainty models. Readers should consult appropriate sources for that information.

There are (at least) four additional sources of uncertainty that should be recognized.

2

Process risk refers to the inherent uncertainty of the insurance claims process. Process risk can diversify away as discussed in Section 1.2.

Model misspecification is the risk that the wrong model is being estimated and applied. For example, this is the risk that we use an exponential model when the phenomenon follows a Pareto distribution. Insufficient parameter identification is also a type of model misspecification.

Actuarial model risk is a broad form of misspecification risk that results from the possibility that the entire actuarial modeling framework may not be appropriate for the phenomenon being modeled. For example, we may model ultimate losses using a loss development model when ultimate claim amounts are not proportional to claim amounts as of the valuation date. Discussion of this risk, which may be significant, is beyond the scope of this Study Note.

Insufficient parameter identification results when we fail to recognize relationships in our models or fail to recognize that certain elements of our model are subject to uncertainty. Examples include:

- Our model may not recognize correlations between development factors in adjacent intervals.

- We may not recognize that relativity between the frequency for a class and the frequency for a base class is an estimated parameter.

1.2 Principles of Diversification

One ad-hoc adjustment sometimes applied in order to capture parameter risk is to add further spread to the frequency and severity distributions. However this approach only adds process risk which will wash out with diversification.

To illustrate the problem, consider applying uncertain trend to the collective risk model. Let N be the random variable for the number of claims, and denote amount of the jth claim as X_j, where the claims amounts are all independent and identically distributed (IID) and independent of N. We then have:

$$L = \sum_{j=1}^{N} X_j \tag{1.1}$$

$$E(L) = E(N)E(X) \tag{1.2}$$

$$Var(L) = E(N)Var(X) + E(X)^2 Var(N) \tag{1.3}$$

To understand the effect of diversification, consider the coefficient of variation (CV, the ratio of standard deviation to mean) of L as a proxy for model uncertainty. It is more convenient to calculate square of the CVs $[CV(L)^2]$ which is

3

the ratio of the variance divided by the mean squared $= Var(L)/E(L)^2$:

$$CV(L)^2 = \frac{Var(L)}{E(L)^2} \tag{1.4}$$

$$= \frac{E(N)Var(X) + E(X)^2 Var(N)}{E(N)^2 E(X)^2}$$

$$= \frac{Var(X)}{E(N)E(X)^2} + \frac{Var(N)}{E(N)^2} \tag{1.5}$$

Actuaries often assume that the CV is constant for severity distributions.

Likewise, for frequency distributions the ratio of variance to mean is often assumed to be constant. We denote that ratio as VM and offer the following examples:

- For a Poisson Distribution , VM is equal to 1.

- For the negative binomial distribution with parameters r and β, with mean r and variance $r(1 + \beta)$, VM is $1 + \beta$, which is often taken as a constant as volume changes.

In any case, VM is constant under the addition of IID exposure units.

By substitution, we have

$$CV(L)^2 = \frac{CV(X)^2}{E(N)} + \frac{VM}{E(N)} \tag{1.6}$$

The numerators of (1.6) are constant under increase in exposure units and inflation, so $CV(L)^2$ decreases proportionally to the inverse of the expected number of claims, and thus can get quite small as volume increases. This is the problem with the collective risk model without parameter uncertainty. The volatility can get unrealistically low leading the actuary to believe that there is no risk in large insurance portfolios. This is a dangerous conclusion as it would lead the insurer to write more business. If we also consider the risk that models for X and N may be incorrectly specified (see the example of the eggs), we understand that potential financial loss actually *increases* with volume.

1.2.1   Uncertain Trend Example

We provide the following example to demonstrate how the aggregate claims random variable is affected by uncertain trend. Including the risk of uncertain trend or other systematic risk will put a minimum on $CV(L)$ that cannot be reduced by diversification (i.e. it is not inversely proportional to $E(N)$).

Let J denote a random trend factor with mean 1.00. We then have the following relationships:

4

$$E(J) = 1 \tag{1.7}$$

$$CV(J)^2 = \frac{Var(J)}{E(J)^2}$$
$$= Var(J) \tag{1.8}$$

Our claims model and its characteristic functions for the trended claim amount K may be expressed as follows:

$$K = JL \tag{1.9}$$

$$E(K) = E(JL)$$
$$= E(J)E(L)$$
$$= E(L) \tag{1.10}$$

$$Var(K) = Var(JL)$$
$$= E(J)^2 Var(L) + E(L)^2 Var(J) + Var(J)Var(L)$$
$$= Var(L) + E(L)^2 CV(J)^2 + CV(J)^2 Var(L) \tag{1.11}$$

$$CV(K)^2 = \frac{Var(L) + E(L)^2 CV(J)^2 + CV(J)^2 Var(L)}{E(L)^2}$$
$$= CV(L)^2 + CV(J)^2 + CV(L)^2 \times CV(J)^2 \tag{1.12}$$

We can now observe that $CV(K)$ has a minimum of $CV(J)$ even if $CV(L)^2$ goes to zero (as $E(N)$ is large). That is, the uncertainty in the trend parameter is not diversified away.

2 Parameter Estimation Methods

We address three common approaches of parameter estimation in this Study Note. For the first two approaches there is a formal methodology for modeling the distribution of parameter fitting errors. This provides quantification of estimation risk.

Regression analysis is used to estimate the parameters of a dependency relationship. Although the category of regression analysis includes non-linear approaches, this Study Note focuses on linear approaches.

<div align="center">5</div>

Maximum likelihood estimation is most commonly used in estimating frequency and severity distributions. The resulting parameters are referred to as maximum likelihood estimators (MLEs).

Although the last approach is less formal, it is no less subject to parameter risk and in fact, it may be subject to greater parameter risk.

Model free methods are commonly used by actuaries in certain applications such as estimation of claim development factors.

3   Parameter Uncertainty Models

3.1   Uncertainty in Regression Parameters

When the data displays dependencies and is (approximately) normally distributed after accounting for those dependencies, actuaries will often use regression to estimate parameters. A common example exists with the modeling of the relationship between claim amounts (X) and time (t) which is often modeled using the following relationship:

$$Y_i = lnX_i = \beta_0 + \beta_1 t_i + \epsilon_i \tag{3.1}$$

where β_0 is often referred to as the intercept and β_1 is often referred to as the slope or regression coefficient.

We observe the following about this relationship:

- Using the log-transform of claim amounts implies that claim values are log-normally distributed. This may be appropriate if the X_is are individual claim observations but possibly not if they are averages.

  It also implies that the growth in claim amounts is exponential rather than linear. This is a generally accepted assumption.

- Exponentiation of the regression coefficient β_1 less unity (i.e. $e^{\beta_1} - 1$) represents an estimate of the annual rate of severity trend.

- $E(Y_i|t_i) = \beta_0 + \beta_1 t_i$, often written as μ_i, is the mean of the distribution of the logs of the claim amounts at time t_i.

We should recognize that regression techniques not only provide estimates of parameters such as β_1 and quantities such as μ_i but also the uncertainty of those estimates. More specifically, for a regression on N data points, the estimated standard deviation of the regression error term, ϵ_i, of the regression may be expressed as:

$$\widehat{\sigma_y} = \sqrt{\frac{SSE}{N-2}} \tag{3.2}$$

6

We denote the sample standard deviation[1] of the observed times (t_is) as σ_t. The estimators then have the following properties which are discussed in textbooks on regression.

- The standard error of b_1 (the estimator of β_1) may be estimated as

$$\widehat{\sigma_{b_1}} = \frac{\widehat{\sigma_y}}{\sigma_t \sqrt{N-1}}. \tag{3.3}$$

  The residuals of b_1 after subtracting β_1 and scaling by the standard error of b_1 follow a Student's t-distribution with $N-2$ degrees of freedom.

- The $(1-\alpha)\%$ confidence interval is equal to

$$b_1 \pm t_{N-2, 1-\frac{\alpha}{2}} \widehat{\sigma_{b_1}} \tag{3.4}$$

- The standard error of m_i, the estimator of μ_i obtained by substituting bs for βs, is calculated as follows:

$$\widehat{\sigma_{\mu_i}} = \widehat{\sigma_y} \sqrt{1 + \frac{1}{N} + \frac{(t_i - \bar{t})^2}{(N-1)\sigma_t^2}} \tag{3.5}$$

  Similar to equation 3.3, the scaled residuals of μ_i also follow Student's t-distribution with N-2 degrees of freedom.

  - We can observe that, as N becomes large, $\widehat{\sigma_{\mu_i}}$ approaches $\widehat{\sigma_y}$.

  - The standard error increases as t_i is further from $\bar{t}$.

  - The $(1-\alpha)\%$ prediction interval is equal to

$$\widehat{Y_i} \pm t_{N-2, 1-\frac{\alpha}{2}} \widehat{\sigma_{\mu_i}} \tag{3.6}$$

Particularly when fitting regression models to average values, N (and, by extension, $(N-2)$) may be "small" which leads to a Student's t-distribution with considerable dispersion. This may result in "unreasonable" parameter values for the regression parameters at higher or lower percentile levels. Excessive dispersion of estimators of parameters is consistent with lack of statistical significance of regression parameters. Issues related to the significance of regression parameters are outside the scope of this Study Note. Readers should consult textbooks on regression analysis for the derivation of the formulae above or for a more complete understanding of the development of the uncertainty model.

---

[1]This is the unbiased standard deviation with denominator $N-1$.

3.2 Uncertainty in Parameters Estimated by Maximum Likelihood

The likelihood function (L) represents the probability that a sample is observed given a model and parameters. It is calculated as the product of probability functions in the discrete case or density functions in the continuous case. As it is computationally more efficient, we generally work with the negative of the log-likelihood (NLL) which is the negative value of the sum of the logarithms of the probability (density) functions . Specifically for a continuous model with density function f, we have:

$$L(x; \theta) = \prod f(x_i) \tag{3.7}$$

$$NLL(x; \theta) = -\sum \ln f(x_i) \tag{3.8}$$

The maximum of L occurs at the minimum of NLL. The minimum of NLL can often be calculated by setting its derivatives with respect to the parameters of the probability (density) function to zero and solving for the parameters. However in more complicated models the minimization must be done numerically.

3.2.1 Large Samples

As described in *Loss Models* [2], **for large** N, the distribution of the parameter estimates is asymptotically normal and the inverse of the **Hessian matrix** (also referred to as the **Hessian** and denoted $\boldsymbol{H}$) provides the variances and covariances of the parameters. The Hessian is comprised of the second partial derivatives of a function of interest, in this case the NLL. The Hessian of the NLL function is also referred to as the information matrix.[2]

3.2.2 Pareto Example

In this section, we demonstrate the calculation for the Pareto distribution with the following properties:

$$F(x) = 1 - x^{-\alpha} \tag{3.9}$$

$$f(x) = \alpha x^{-\alpha-1} \tag{3.10}$$

$$\ln(f(x)) = \ln(\alpha) + (-\alpha - 1)\ln(x) \tag{3.11}$$

---

[2]Most optimization software will numerically calculate the information matrix.

8

We then calculate the NLL as follows:

$$NLL = -\sum_{i=1}^{n} \ln f(x_i)$$

$$= -\sum_{i=1}^{n} \left(\ln(\alpha) + (-\alpha - 1) \ln(x_i) \right)$$

$$= -\sum_{i=1}^{n} \ln(\alpha) + (\alpha + 1) \sum_{i=1}^{n} \ln(x_i)$$

$$= -n\ln(\alpha) + (\alpha + 1) \sum_{i=1}^{n} \ln(x_i) \tag{3.12}$$

To solve for the MLE of α, we taking the derivative of the NLL with respect to α and solve:

$$\frac{dNLL}{d\alpha} = \frac{-n}{\alpha} + \sum_{i=1}^{n} \ln(x_i) = 0$$

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln(x_i)} \tag{3.13}$$

To determine the variance of the MLE, we take second partial derivatives of the NLL as follows:

$$\frac{\partial^2 NLL}{\partial \alpha^2} = \frac{n}{\alpha^2} \tag{3.14}$$

With only one parameter, the $\boldsymbol{H}$ is a 1×1 matrix.

$$H = \left[\begin{array}{c} \frac{n}{\alpha^2} \end{array} \right] \tag{3.15}$$

$$H^{-1} = \left[\begin{array}{c} \frac{\alpha^2}{n} \end{array} \right] \tag{3.16}$$

So for large n, the maximum likelihood estimator of the Pareto parameter is normally distributed with mean $= \hat{\alpha}$ and estimated variance $= \hat{\alpha}^2/n$.

$\square$

We leave it to the reader to verify the uncertainty models for the exponential and lognormal distributions below.

3.2.3   Limited Samples Sizes

For insurance samples the sample size is usually not asymptotic to infinity and the normal distribution often is inappropriate. For instance, a normal distribution might imply too high a probability of negative values for parameters and

9

Table 1: Examples

Model	Lognormal(μ, σ)	exponential(λ)
Mean	$e^{\mu + \sigma^2/2}$	$1/\lambda$
MLE	$\hat{\mu} = \dfrac{\sum \ln x_i}{n},$	$\hat{\lambda} = \dfrac{n}{\sum x_i}$
	$\hat{\sigma} = \dfrac{\sum (\ln x_i - \hat{\mu})^2}{n}$	
H	$\begin{bmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{2n}{\sigma^2} \end{bmatrix}$	$\dfrac{n}{\lambda^2}$
H^{-1}	$\begin{bmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{\sigma^2}{2n} \end{bmatrix}$	$\dfrac{\lambda^2}{n}$

functions of parameters that have to be positive. A reasonable alternative in that case is to use the gamma distribution for each parameter, with the correlation structure of the multivariate normal. This can be implemented using the normal copula with gamma marginal distributions. As the sample sizes get larger, the gamma approaches the normal, so using it is consistent with the asymptotic theory.

3.2.4 The Pareto Example

Returning to our Pareto example, we recall that the log of a Pareto variate is exponentially distributed and the sum of exponentials is gamma. From 3.13, we recognize that the Pareto variates are in the denominator of the MLE of α. As a result, we understand that $\hat{\alpha}$ is inverse gamma distributed with mean and variance of estimators being $\hat{\alpha}$ and $\hat{\alpha}^2/n$, respectively. This agrees what was calculated is Section 3.2.2. The associated inverse gamma shape and scale parameters would be $n + 2$ and $\alpha(n + 1)$, respectively.

It would be tempting to use this inverse gamma as the distribution of the true parameter given the fit. However it is just the opposite - that inverse gamma is the distribution of the estimator given the true parameter. Especially with skewed distributions like the inverse gamma, these two distributions are not the same.

This is a natural setup for Bayesian analysis. We know the distribution of the

10

estimator given the parameters but want the distribution of the parameters given the estimator. If the MLE were also the Bayes estimate from some prior distribution of the parameters, then Bayes Theorem would provide the posterior distribution of the parameters given the estimate. This happens in one setting, and the resulting posterior distribution of the parameters turns out to be gamma in that case.

3.2.5   Bayes Theorem

Bayes Theorem provides a formula for the posterior distribution for Y given X, using the distributions of X, Y and X given Y. That is:

$$f(Y|X) = f(X|Y)\frac{f(Y)}{f(X)} \qquad (3.17)$$

We can think of Y as the true parameter, which is considered a random variable since it is not known, and X as the data. Then, the prior distribution of Y is $f(Y)$ and $f(X|Y)$ is the conditional distribution of the data given the parameter. We want to find the conditional distribution of Y given X, and in that context $f(X)$ in equation 3.17 can be considered as a normalizing constant (not a function of Y) needed to make the distribution integrate to unity. As such, Bayes Theorem can also be expressed as:

$$f(Y|X) \propto f(X|Y)f(Y) \qquad (3.18)$$

Where $\propto$ indicates proportionality - meaning equal up to factors not containing Y. This formulation allows the use of so-called non-informative priors - such as, in this case $f(Y)$. The prior $f(Y)$ is thus expressed by suppressing factors not containing Y. This allows the prior $f(Y)$ itself to be expressed up to a constant factor, and in fact does not even have to integrate to a finite number as long as $f(Y|X)$ does. This gives the possibility of prior distributions that are very spread out on the real line and so have little or no impact on the estimated parameters.

Common examples are $f(Y) \propto 1$ on the whole real line, or $f(Y) \propto 1/Y$ on the positive reals. These can be expressed as limits of the same distributions on $(-M, M)$ or $(1/M, M)$ as M grows without limit. Thus they are very diffuse. Such non-informative priors can give insights into the estimation uncertainty.

For the Pareto, the prior is for the parameter α, and for a positive parameter a useful non-informative prior is $f(\alpha) \propto 1/\alpha$. The anti-derivative of this prior is $\ln(\alpha)$, which slowly diverges at both ends of the positive real line. Thus it has infinite weight at both ends of the range, and as a result does not bias the parameter either up or down. In comparison, for a positive parameter, the prior $f(\alpha) \propto 1$ only diverges at the right end of the range, and tends to pull parameters up.

11

In this example $f(X|\alpha)$ is the distribution of the observations given α. If P is the product of the observations, it is easy to show that

$$f(X|\alpha) \propto \alpha^n / P^{\alpha+1} \tag{3.19}$$

If we substitute $\beta = -ln1/P$, we have:

$$f(X|\alpha) \propto \alpha^n \exp(-\beta\alpha) \tag{3.20}$$

Comparing this to the gamma density shows that the distribution of the parameter given the data is a gamma distribution with shape parameter n and mean $= 1/\text{average}[\ln x_j]$. This mean is the MLE for α, which supports the use of this particular non-informative prior. This gamma distribution is thus the posterior distribution for the true α, with mean equal to the MLE estimate.

A similar exercise for the Poisson with mean λ and n samples which have sum of observations S gives a gamma posterior distribution for λ with mean S/n and shape parameter S. This again agrees with the MLE and has a gamma distribution for the true parameter. Both examples support the idea of using gamma distributions for the parameter uncertainty.

3.3   Uncertainty in Model Free Estimators

Development factors can be calculated within a parametric or model-free framework. The factors themselves are parameters, but the distinction is whether or not a distribution is assumed for the deviation of the losses from what would be estimated by applying the factors, that is, for the distribution of the residuals of the development factor approach.

One method for quantifying the estimation errors of the factors is bootstrapping. This method resamples the residuals and uses them to create new, artificial triangles. The factors are repeatedly estimated from these artificial triangles, and an empirical distribution of the factors is thus built up. Bootstrapping is a straightforward approach but has potential pitfalls that require some care.

- For example, it should be recognized that there are a different number of observations used in the estimation of successive incremental development factors, so each "parameter" has its own number of degrees of freedom. The degrees of freedom is an input to the resampling process.

  In nonlinear models, the degrees of freedom can be estimated by Ye's method of generalized degrees of freedom[3] (gdf). The gdf for an observed point, for an estimation procedure, is the derivative of the fitted point with respect to the observed point. If that derivative is one, the observed point has the power to pull the model to it with an exact match. This would show up for instance in fitting a quintic polynomial to 6 points, which it can fit exactly, using up all the degrees of freedom. The gdf agrees

12

with the usual notion of degrees of freedom in linear models, and is more appropriate in nonlinear models.

Even when using the gdf degrees of freedom for each point's residual, however, bootstrapping is regarded as unreliable in small samples (e.g., less than 40 observations per fitted parameter). There are too few residuals to get a representative resample. This leads to the method of parametric bootstrapping, which draws from fitted distributions instead of the observed residuals. This would only be applicable in the case where there is a parametric model for the residuals. For instance, if residuals are assumed to be over-dispersed Poisson, resampling can be done from this distribution.

■ The approach outlined in *England and Verall* (2002) uses Pearson residuals, r_p, which are calculated using the following approach:

$$r_p = \frac{\text{observation} - \text{estimated parameter}}{\text{estimated parameter}^{1/2}} \qquad (3.21)$$

■ A technical problem is that bootstrapping gives the distribution of the estimated parameters given the true parameters, but what is needed is the distribution of the true parameters given the estimated parameters. This difference will be important especially with asymmetric distributions. This is the same problem that was encountered in the Pareto example, and which there led to replacing the inverse gamma distribution by the gamma. This is a known problem with bootstrapping which is addressed in textbooks on the subject, but is beyond the scope of this Study Note.

■ In development triangles another pitfall of resampling is that the model might not hold for the data.

  - For instance, in slowly developing lines, the first report claim amounts might often be near zero. The second report might then be well modeled as a constant (for the initial valuation of claims that are true IBNR at the first report) plus a factor times first report (for development of the small number of reported claims). If the model uses just a factor, there might be some very high observed factors that would not apply in general but might when the first report is very low. Resampling can generate obviously inappropriate development in this case - such as a large residual combined with a large initial value - basically because the wrong model is being used to estimate claims at second report.

  - Also if there are calendar-year effects in the data but not in the model, bootstrapping can again be distorted because it is resampling residuals of a model that does not apply.

If the development factors are estimated by MLE from a parametric model, the inverse of the Hessian (information matrix) can be used to quantify the

13

parameter uncertainty in the factors, just as in any other MLE case. *Clark*(2006) [1] gives an example of this. Comparison studies have found the results of this method to be comparable to bootstrapping the parameter uncertainty, and using the information matrix in this way avoids many of the pitfalls of bootstrapping.

4 Incorporating parameter risk in simulation models

Actuaries typically use simulation to model risk and uncertainty. Parameter estimation is easily incorporated in a simulation through a two-stage process: in each scenario, we first simulate the parameters from the parameter-risk distributions, and then simulate the process from the simulated parameters. Examples of this approach are as follows:

- In our example of uncertain trend from Section 2, we would first simulate aggregate claims from the collective risk model, and then simulate J which is then multiplied by the aggregate claims. This approach results in a similar floor imposed on the simulated claims $CV(K)$.

- In our Pareto example, we first simulate the parameter value and then simulate claims based on that parameter.

Even if the process risk diversifies away, the parameter risk will not.

5 Conclusion

It should be noted that this approach assumes that:

Parameter risk is one of the principal elements that have to be quantified to obtain reasonable representations of risky processes. As we demonstrated, in a loss simulation environment, simulating from the collective risk model without recognizing parameter risk can wash out most of the actual risk. This is particularly true for high-volume lines.

In this Study Note, we have provided an overview of approaches to estimate parameter uncertainty based on the manner in which the parameters are estimated. Interested readers should consult textbooks and other papers for details related to the theory on the parameter uncertainty models.

References

[1] David R. Clark. Variance and covariance due to inflation. *Casualty Actuarial Society Forum*, pages 61–95, Fall 2006.

14

[2] Harry H. Panjer Klugman, Stuart A. and Gordon E. Willmot. *Loss Models: From Data to Decisions*. Wiley, third edition, 2008.

[3] Jianming Ye. On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, pages 120–131, March 1998.

15