

The Theory of Split Credibility

by Ira Robbin

ABSTRACT

This paper tackles the question: why should split credibility be better than credibility without a split? It corrects previous misunderstandings and presents new formulas showing how parameter uncertainty is reduced by use of unsplit credibility and then how it might be further reduced by introduction of a split. It derives the formulas for unsplit and split credibility when losses follow the widely used collective risk model (CRM). It then demonstrates that split credibility can sometimes be ineffective in a CRM context and can sometimes produce negative credibility values or inversions of the primary and excess credibilities. The paper concludes with a call for further research to find a stronger conceptual justification for the split credibility plan used in practice.

KEYWORDS

Credibility, experience rating, split plan, Workers Compensation

1. Introduction

In an experience-rating plan with a primary-excess split, such as the one promulgated by the National Council on Compensation Insurance (NCCI) (2002) for rating workers compensation risks, individual risk losses are divided into primary and excess components. A credibility-weighted estimate of each component is obtained and the two estimates are added together to produce the final experience-adjusted estimate of total loss.¹

1.1. Conceptual foundation

But does this splitting procedure lead to a better estimate than credibility weighting without a split? Gillam (1989; 1992) and others have presented strong empirical evidence that a split actually *does* work better. However, we have found no paper that correctly and completely explains *why* it should work better. The first purpose of this paper is to provide a rigorous conceptual foundation for split credibility. Then we will use that as a base to arrive at a clear understanding of the conditions needed for splitting to produce materially superior estimates.

1.2. Incorrect intuitive justification

We start by explaining what is wrong with an intuitive justification for split credibility that is given in the literature. The incorrect justification is that use of a split breaks total loss into two components, each *separately* less volatile and, thus the argument goes, more credible than the total.² However, as is generally accepted, excess layer loss is inherently more

volatile than total loss.³ To be more precise, as shown in Appendix A, excess layer loss has a process risk coefficient of variation (CV) at least as large as the corresponding CV for total loss.

1.3. Volatility alone does not determine credibility

Because the excess layer is more volatile, can we therefore conclude it is less credible than the primary layer? Many readers may be thinking the answer is an obvious “yes.” Recalling further that such a relation between primary and excess layer credibility is designed into the NCCI plan,⁴ they might feel even more certain that excess layer credibility must be less than primary layer credibility. However, as we will later see, there is nothing in the general mathematics that forces such a relation. The reason is simply that “high volatility” is not synonymous with “low credibility.” Rather, credibility is conceptually the weight given to observed data, as opposed to the weight given to prior belief. It depends not only on volatility (process risk), but also on the uncertainty in our initial belief (parameter risk).

An often-underappreciated aspect of credibility is that credibility is positively correlated with initial ignorance (parameter risk): the less we think we know in advance, the more willing we are to be swayed by the observed data, even if it is noisy. So, when we try to assess credibility in a split plan, we need to examine not only the process variances, but also the parameter variances of the split components. In general, with an arbitrary loss model, there is nothing to prevent a split from allocating a relatively larger portion of parameter risk than process risk to the excess layer. If that happens, the excess layer may end up having *more* credibility than the primary layer.

¹The actual plan uses a formula that contains ballast and weight values. Venter (1987) shows this is equivalent to adding together separate credibility-weighted estimates of the primary and excess losses.

²In an otherwise excellent article that has been on the Casualty Actuarial Society examination syllabus for many years, Venter (1987) wrote that “both the primary and excess losses are less heavy-tailed than total losses: this seems obvious for primary losses. For excess losses, by eliminating the smaller portion, enough losses are eliminated to bring up the average value and to reduce the probability of a loss being a large multiple of the average. This makes the excess losses less heavy-tailed and thus more predictable than total losses.”

³For example, Teng (1994) argues that workers compensation large dollar deductible and excess programs are riskier than full coverage programs due to the greater variability of excess losses.

⁴Using the formula from Venter (1987), $z_e = w \cdot z_p$ where w is the weighting value, we see that the design of the plan forces the excess credibility to always be less than the primary credibility.

In addition, when analyzing these components and their credibility, it is insufficient to consider them in isolation: their process covariance and parameter covariance⁵ both need to be considered.

1.4. Risk allocation and effective splitting

When we split losses, we induce allocations of the process risk and the parameter risk to the separate components and to their process covariance and parameter covariance. Based on the realization that splitting leads to such allocations, we can then see that an effective split plan necessarily entails a trade-off in which one component gets a higher credibility and the other, a lower credibility. The key to achieving an effective plan is thus to define components so as to separate a less predictable portion of losses from a more predictable one. To put it another way, a split will work if it helps us concentrate on the signal and ignore the noise.

1.5. Reduction of mean square parameter error with optimal credibility

To show that these understandings are, in fact, correct, we will start first by examining experience rating when there is no split. Using minimal mean-squared error as the criterion for optimality, we will show that use of the optimal credibility value reduces the expected square error of the estimate of the mean (the parameter risk) by a ratio equal to that optimal credibility value.

For example, if the optimal credibility is 40% and the original expected parameter variance is 100, the optimally credibility-weighted estimate of the mean will have an expected square error of 60.

⁵Calculations have been done without including covariance. In particular, when Gillam (1989) computed parameters for the NCCI workers compensation split experience-rating plan, he chose to omit the covariance terms from Mahler's equations. His decision was based on a simplifying assumption that he stated was "defensible more on the basis of its usefulness than its veracity." Given the practical focus of Gillam's work, this was a reasonable choice, but that should not be read as a theoretical justification for ignoring covariance.

Thus optimal credibility has a dual role:

- It is the best weight to assign to observed data as opposed to prior belief in arriving at an estimate of the mean, and
- It is the percentage by which parameter variance is reduced by using the optimal weight on experience in computing the experience-adjusted estimate.

1.6. Reduction of mean square error with optimal split credibilities

We will then turn to an arbitrary split plan, where the split is any manner of dividing losses. We will derive optimal credibility formulas, where optimality here again denotes maximal reduction in the expected square error of the estimate of the mean. Our formulas are equivalent to formulas previously presented by Mahler (1987) with notation modified to facilitate interpretation. We will then study the reduction in mean square error when optimal credibility values are used, leading to a split credibility version of the error reduction formula. In the split model formula, we first allocate the original parameter variance to the components. Under this particular allocation,⁶ each component gets its own variance plus the covariance. When optimal credibilities are used, the square error for each allocated component is reduced by its credibility.

Returning to the example in Section 1.5, suppose we split the losses in two, and suppose further the components have parameter variances of 60 and 20, respectively, and a parameter covariance of 10. Note this reconciles with a parameter variance of 100 for the unsplit total, since $60 + 20 + 2 * 10 = 100$. The allocations of the total parameter variance are therefore 70 ($70 = 60 + 10$) and 30 ($30 = 20 + 10$). If the optimal credibilities are 50% and 20%, then the mean square error is $70 * (100\% - 50\%) + 30 * (100\% - 20\%) = 35 + 24 = 59$. Recall that the optimal unsplit credibility was 40% and that use of unsplit

⁶Note there are many possible ways of allocating covariance back to the individual components.

credibility thus reduced square error from 100 to 60. Introduction of the split has reduced parameter error in this example, but only a modest amount from 60 to 59.

1.7. Differential risk allocation determines split effectiveness

To study, in general, what might be gained by adopting a split plan, we will take the difference in the mean square estimation errors between the optimal non-split and split plans. Based on the resulting formula, we will show split credibility is most effective at reducing mean square estimation error when the two components have relatively different amounts of process and parameter risk. If there is such a differential allocation, the component with the lion's share of parameter risk ends up with optimal credibility larger than the optimal credibility for the unsplit losses while the component with the lion's share of process risk has optimal credibility smaller than the optimal credibility for the unsplit losses. If the split does not produce such a differential allocation of process and parameter risk, it need not be appreciably more effective than a no-split plan.

1.8. Primary-excess splits

While an arbitrary split might not produce much of an improvement, one might hope a reasonable primary-excess split would do better. Such a split will allocate the volatile tail of severity to the excess layer so the excess layer will receive a disproportionate share of the overall process risk. However, as argued previously, we can say nothing about whether the split is effective unless we also know how the parameter risk gets allocated. That, in turn, depends on the structure of the loss model and its priors. With an arbitrary loss model and arbitrary priors, there is no reason the split could not allocate a proportion of the parameter risk that is smaller to, equal to, or greater than the proportion of the process risk allocated to the excess layer. As a result, we arrive at the possibly disappointing conclusion that a primary-excess split does not, in general, significantly improve accuracy.

The key to whether the split is effective depends critically on how much parameter uncertainty there is with respect to the severity of losses. In the extreme case where mean severity is fixed and only the mean claim counts are uncertain, then actual excess losses are just a noisy distraction from the true signal emanating from the primary loss. When that is the case, a split is very effective, and the smaller the split point the better. However, when severity is subject to significant parameter risk, splitting may not accomplish much at all.

1.9. Misbehavior of optimal split credibility values

Under the NCCI plan, credibility values are well-behaved in two respects:

- Both primary and excess credibilities are between zero and unity and there are no negatives or values over 100%, and
- There are no inversions: primary credibilities are always less than or equal to excess credibilities.

However, optimal split plan credibilities under the minimal MSE criteria do not necessarily obey these guidelines. If mean frequency is known to a fair degree of accuracy, while mean severity is quite uncertain, optimal credibility values may become inverted. Intuitively, in such a scenario, the primary layer results carry little information about severity, but it is information about severity that is needed arrive at a better estimate of mean loss.

Mathematically, a negative credibility value for either the primary or the excess layer can emerge as a solution to the optimal mean square error equations. Intuitively, this could occur when a split allocates most process risk to one component and it also induces a sizeable parameter covariance. In such a situation, results from the non-volatile component may provide better information about the other component than its own results.

1.10. Split credibility when losses follow the collective risk model

We will examine split credibility under the Heckman and Meyers (1983) collective risk model (CRM). In

that model, claim counts are assumed to be conditionally Poisson with a Gamma prior. Parameter risk for the claim counts is driven by the “contagion” parameter. Claim severities are conditionally exponential and also have a Gamma prior. Parameter risk for severity is captured in the “mixing” parameter, which quantifies uncertainty about the scale. We will derive equations for split credibility under CRM. Our equations are equivalent to Mahler’s (1987), though we use a different notation to facilitate interpretation.

As might be expected based on prior discussion, a split does not automatically confer any great advantage when the underlying losses follow the CRM. With some sets of parameters it works fairly well; with others it confers modest or even no improvement at all over unsplit credibility. In some cases, a CRM can produce primary-excess credibility inversions in which the optimal excess layer credibility is larger than the optimal primary layer credibility. For example, a primary-excess credibility inversion would be present if the optimal primary layer credibility was 25% while the optimal excess layer credibility was 40%. In still other cases, one can have primary credibilities over 100% and excess credibilities that are negative. Stranger still, there are scenarios in which the primary credibility is negative. The interplay of contagion, mixing and split point governs which scenario will prevail.

1.11. Loss capping and mod extension

Under the NCCI experience rating plan individual accidents are subject to an accident limit, the State Accident Limit (SAL⁷), before being split into primary and excess components by the split point. The sum of credibility weighted primary and excess losses is compared to a calculated value of expected loss⁸ that reflects the accident limit. This produces the experi-

⁷Under the NCCI plan (2002), the State Accident Limit (SAL) is also used to define caps on multi-person accidents and occupational disease losses.

⁸The expected loss is derived from expected loss rates that are also adjusted to the development and law level of the experience period losses. See Gillam (1992) for an excellent and detailed description of the procedures used to put the loss rates at the level of the ratable losses.

ence modification factor (Mod) for a risk. While the Mod is obtained from losses that are capped at the accident limit, it is then applied to initial expected losses that are uncapped to arrive at the final estimate of experience adjusted expected losses.

There are theoretical and practical justifications⁹ for this capping and mod extension procedure. Conceptually, it tames the severity tail that is often the key driver of overall process risk. This stabilizing effect tends to increase the credibility of the excess layer. It comes at a price, however; there is uncertainty in extrapolating from capped losses to uncapped losses. One course for future research is to use the methods developed in this paper to analyze optimal mean square error for Mod extension estimates.¹⁰

1.12. MSE derivation and CRM support for primary-excess split

Our conclusion is that the minimal mean square error credibility derivation does not provide strong conceptual support for a primary-excess split. Further, with CRM losses, split credibility may or may not do appreciably better than unsplit credibility. In addition, optimal credibilities may not be well-behaved: the model may produce negative credibilities or primary-excess credibility inversions. Later we will end with very brief speculation on what could be done to provide more support for primary-excess split credibility.

2. No-split credibility

We start with a general no-split plan. Let A be the random variable representing actual historical loss. We suppose A is dependent on a possibly multi-dimensional parameter, θ , and define $\mu(\theta) = E[A|\theta]$ and $\sigma^2(\theta) = \text{Var}(A|\theta)$. Let h be the prior distribution of θ and use h to define $E = E[\mu(\theta)]$, $\sigma^2 = E[\sigma^2(\theta)]$, and $\tau^2 = \text{Var}(\mu(\theta))$. Under this notation, σ^2 is a measure

⁹One practical advantage is that capping produces results less sensitive to the anomalies of loss development and claims reserving practices.

¹⁰Developing an estimate on limited losses and extending it using a mod factor to unlimited losses is another form of experience rating that is often done in its own right, without any subsequent splitting.

of the process risk and τ^2 is a measure of parameter risk. We also set $\lambda^2 = \sigma^2 + \tau^2$ so that λ^2 is the total variance of A .

In this construction each risk has a particular θ value that we have no way of knowing in advance. Our initial knowledge is only about the distribution of the parameter, θ .

Given an observation of A for a particular risk, we could use Bayes Theorem to obtain the posterior distribution, $h(\theta|A)$.¹¹ From this, we could in principle compute the conditional expected value, $E[\mu(\theta)|A]$. However, the conditional expected value may be difficult to compute and so a linear mod formula is often used. Regarding z as a variable, the resulting linear estimate of the expected value of A is given as

$$\hat{A} = z \cdot A + (1 - z) \cdot E. \quad (2.1)$$

Here credibility, z , is the weight given to the actual experience. We use the notation, z^* , to denote the optimal credibility value under the least mean square error criterion. To find this optimal credibility, we first write the mean square error as a function of the credibility:

$$\begin{aligned} \epsilon^2 &= E[(zA + (1 - z)E - \mu(\theta))^2] \\ &= z^2 \cdot E[(A - \mu(\theta))^2] + (1 - z)^2 \\ &\quad \cdot E[(E - \mu(\theta))^2] \\ &= z^2 \sigma^2 + (1 - z)^2 \tau^2. \end{aligned} \quad (2.2)$$

The expectation is with respect to θ and then with respect to A given θ . In simplifying Equation (2.2), various cross terms vanish under the assumption the sampling deviation of actual results from the mean for a risk is independent of the deviation of the risk mean from the population mean. Specifically, we have assumed:

$$E[(A - \mu(\theta))(E - \mu(\theta))] = 0. \quad (2.3)$$

This assumption is plausible because in the CRM and most other loss model constructions the param-

¹¹In a more complete treatment, the actual experience, A , would explicitly depend on the number of years of observations. That would not change the qualitative conclusions to be reached in this paper.

eters are first randomly selected from the priors and then the values of the losses and sampled from the loss distributions with those selected parameters. Such a sequential procedure guarantees theoretical independence between parameter error and conditional value error as expressed in Equation (2.3).

We next use standard techniques of basic calculus to find the credibility value that minimizes the square error. Taking the derivative of the square error with respect to z , we find:

$$\frac{d\epsilon^2}{dz} = 2z\sigma^2 - 2(1 - z)\tau^2. \quad (2.4)$$

Next we set the derivative to zero and solve:

$$\begin{aligned} \frac{d\epsilon^2}{dz} = 0 &\Rightarrow z(\sigma^2 + \tau^2) = \tau^2 \\ \Rightarrow z &= \frac{\tau^2}{(\tau^2 + \sigma^2)}. \end{aligned} \quad (2.5)$$

So the credibility, z^* , that minimizes mean square error is given as:

$$z^* = \frac{\tau^2}{\tau^2 + \sigma^2} = \frac{\tau^2}{\lambda^2}. \quad (2.6)$$

Using Equations (2.2) and (2.6), we see that the minimum mean squared error for the non-split linear estimator is given as:

$$\begin{aligned} \epsilon_0^2(\text{NS}) &= \left(\frac{\tau^2}{\tau^2 + \sigma^2}\right)^2 \cdot \sigma^2 + \left(\frac{\sigma^2}{\tau^2 + \sigma^2}\right)^2 \cdot \tau^2 \\ &= \left(\frac{\tau^2 \sigma^2}{(\tau^2 + \sigma^2)^2}\right) \cdot (\tau^2 + \sigma^2) \\ &= \frac{\tau^2 \sigma^2}{(\tau^2 + \sigma^2)}. \end{aligned} \quad (2.7)$$

The “NS” label stands for “No-Split.” Using Equation (2.6), this minimal square parameter error can be written as:

$$\epsilon_0^2(\text{NS}) = \frac{\tau^2 \sigma^2}{\tau^2 + \sigma^2} = \tau^2 \left(1 - \frac{\tau^2}{\lambda^2}\right) = \tau^2(1 - z^*). \quad (2.8)$$

Table 1. General split loss model notation

Loss Component	Notation			
	A	A ₁	Co	A ₂
Mean	μ	μ ₁		μ ₂
Process Variance	σ ²	σ ₁ ²	ρ	σ ₂ ²
Parameter Variance	τ ²	τ ₁ ²	π	τ ₂ ²
Total Variance	λ ²	λ ₁ ²	κ	λ ₂ ²

Since the initial square parameter error before any observations are made is τ², this equation says use of the optimal credibility value reduces mean square parameter error by a proportion that is equal to that optimal credibility value.

3. General split plan credibilities

Assume A can be written as the sum of two loss random variables: A = A₁ + A₂. In this generality, the split is not necessarily between primary and excess losses: it could be any way of splitting losses. We suppose each A_i is dependent on a possibly multi-dimensional parameter, θ, and define μ_i(θ) = E[A_i(θ)] and σ_i²(θ) = Var(A_i(θ)). Also let C(θ) = Cov(A₁(θ), A₂(θ)). Assume h is the prior distribution of θ and use h to define E_i = μ_i = E[μ_i(θ)], σ_i² = E[σ_i²(θ)], ρ = E[C(θ)], τ_i² = Var(μ_i(θ)), λ_i² = σ_i² + τ_i², and π = Cov(μ₁(θ), μ₂(θ)). Note that, in addition to the process and parameter risk terms for each loss component, we have also defined expected process covariance and parameter covariance terms. Set κ = ρ + π so that κ is the total covariance. Define σ² as the total process variance, τ² as the total parameter variance, and λ² as the total variance. We observe that σ² = σ₁² + σ₂² + 2ρ, τ² = τ₁² + τ₂² + 2π, and λ² = λ₁² + λ₂² + 2κ. The notation is summarized in Table 1.

The split credibility Mod formula¹² is:

$$MOD = \frac{z_1 A_1 + (1 - z_1) E_1 + z_2 A_2 + (1 - z_2) E_2}{E} \quad (3.1)$$

¹²As before, a more complete treatment would explicitly show dependence on the number of years of observations. That would not change the qualitative conclusions to be reached in this paper.

As before, we derive a formula for the mean square parameter error:

$$\begin{aligned} \varepsilon^2 &= E[(z_1 A_1 + (1 - z_1) E_1 - \mu_1(\theta) + z_2 A_2 \\ &\quad + (1 - z_2) E_2 - \mu_2(\theta))^2] \\ &= z_1^2 \cdot E[(A_1 - \mu_1(\theta))^2] + (1 - z_1)^2 \\ &\quad \cdot E[(E_1 - \mu_1(\theta))^2] + z_2^2 \cdot E[(A_2 - \mu_2(\theta))^2] \\ &\quad + (1 - z_2)^2 \cdot E[(E_2 - \mu_2(\theta))^2] \\ &\quad + 2z_1 z_2 E[C(\theta)] + 2(1 - z_1)(1 - z_2) \\ &\quad \text{Cov}(\mu_1(\theta), \mu_2(\theta)). \end{aligned} \quad (3.2)$$

In obtaining this expression, we have assumed that the sampling deviation of actual results from the mean for each variable is independent of the deviation of the risk mean from the population mean for both variables. In mathematical notation, these assumptions can be written as:

$$\begin{aligned} 0 &= E[(A_1 - \mu_1(\theta)) \cdot (\mu_1(\theta) - \mu_1)] \\ &= E[(A_1 - \mu_1(\theta)) \cdot (\mu_2(\theta) - \mu_2)] \\ &= E[(A_2 - \mu_2(\theta)) \cdot (\mu_1(\theta) - \mu_1)] \\ &= E[(A_2 - \mu_2(\theta)) \cdot (\mu_2(\theta) - \mu_2)]. \end{aligned} \quad (3.3)$$

In the derivation of Equation (3.2), the assumptions in (3.3) are used to eliminate various cross terms. Note that the square error formula has other terms which do not vanish but which depend on the process and parameter covariance. These terms are present in Mahler's formula, though in different notation. We express (3.2) using our notation, next expand expressions to arrive at terms that are polynomials of the credibilities, and then group them as follows:

$$\begin{aligned} \varepsilon^2 &= z_1^2 \sigma_1^2 + (1 - z_1)^2 \cdot \tau_1^2 + z_2^2 \sigma_2^2 + (1 - z_2)^2 \tau_2^2 \\ &\quad + 2z_1 z_2 \rho + 2(1 - z_1)(1 - z_2) \pi \\ &= z_1^2 \sigma_1^2 + \tau_1^2 - 2z_1 \tau_1^2 + z_1^2 \tau_1^2 + z_2^2 \sigma_2^2 + \tau_2^2 \\ &\quad - 2z_2 \tau_2^2 + z_2^2 \tau_2^2 + 2z_1 z_2 \rho + 2\pi - 2z_1 \pi \\ &\quad - 2z_2 \pi + 2z_1 z_2 \pi \\ &= \tau_1^2 + \tau_2^2 + 2\pi + z_1^2 \sigma_1^2 + z_1^2 \tau_1^2 - 2z_1 \tau_1^2 \\ &\quad - 2z_1 \pi + z_2^2 \sigma_2^2 + z_2^2 \tau_2^2 - 2z_2 \tau_2^2 - 2z_2 \pi \\ &\quad + 2z_1 z_2 \rho + 2z_1 z_2 \pi. \end{aligned} \quad (3.4)$$

Using our notation to simplify further, we have:

$$\begin{aligned} \epsilon^2 = & \tau^2 + z_1^2 \lambda_1^2 - 2z_1(\tau_1^2 + \pi) + z_2^2 \lambda_2^2 \\ & - 2z_2(\tau_2^2 + \pi) + 2z_1 z_2 \kappa. \end{aligned} \quad (3.5)$$

We take partials with respect to the credibility parameters:

$$\begin{aligned} \frac{\partial \epsilon^2}{\partial z_1} &= 2z_1 \lambda_1^2 - 2(\tau_1^2 + \pi) + 2z_2 \kappa \\ \frac{\partial \epsilon^2}{\partial z_2} &= 2z_2 \lambda_2^2 - 2(\tau_2^2 + \pi) + 2z_1 \kappa. \end{aligned} \quad (3.6)$$

Setting the partials equal to zero, we obtain the system of equations:

$$\begin{aligned} z_1 \lambda_1^2 + z_2 \kappa &= (\tau_1^2 + \pi) \\ z_2 \lambda_2^2 + z_1 \kappa &= (\tau_2^2 + \pi). \end{aligned} \quad (3.7)$$

Solving we find:

$$\begin{aligned} z_1 &= \frac{\lambda_2^2 (\tau_1^2 + \pi) - \kappa (\tau_2^2 + \pi)}{D} \\ z_2 &= \frac{\lambda_1^2 (\tau_2^2 + \pi) - \kappa (\tau_1^2 + \pi)}{D} \end{aligned} \quad (3.8)$$

where $D = \lambda_1^2 \lambda_2^2 - \kappa^2$.

Example 1 demonstrates the credibility formulas in (3.8). Example 1 is shown with additional information in Exhibit 1 Sheet 1.

Example 1. Split credibility example

Loss Component	Unsplit Credibility	Split Credibility		
	A	A ₁	Covariance	A ₂
Mean	100	50		50
Process Variance	400	100	70	160
Parameter Variance	200	100	25	50
Total Variance	600	200	95	210
MSE Optimal No-split Credibility	33.3%			
D			32,975	
MSE Optimal Split Credibility		58.0%		9.5%

As Mahler (1987) noted, the solutions of (3.8) are not really credibilities in the traditional sense, since, in this generality, one of them could be negative or have a value above unity. As proved in Proposition 2 in Appendix B, the minimal mean square error for the split plan is given as:

$$\epsilon_0^2(SP) = \tau^2 - \frac{1}{D} \begin{pmatrix} \lambda_2^2 (\tau_1^2 + \pi)^2 \\ + \lambda_1^2 (\tau_2^2 + \pi)^2 \\ - 2\kappa (\tau_1^2 + \pi)(\tau_2^2 + \pi) \end{pmatrix}. \quad (3.9)$$

In Appendix B, it is also shown that Equation (3.9) can be reduced to:

$$\begin{aligned} \epsilon_0^2(SP) &= (\tau_1^2 + \pi)(1 - z_1^*) \\ &+ (\tau_2^2 + \pi)(1 - z_2^*). \end{aligned} \quad (3.10)$$

Here we have reintroduced the “*” denoting optimal credibility to emphasize that the formula is only valid when optimal credibility values are used. The “SP” indicates the formula is for a split plan. This formula extends the parameter error reduction formula from the no-split case. The initial mean square parameter error is the parameter risk. Under Formula (3.10), it is split, with each component taking its own parameter variance and the parameter covariance. Since the total covariance portion of the parameter variance is two times the parameter covariance, each component is allocated half of the parameter covariance contribution to the total parameter variance. These allocations are then reduced in proportion to the respective optimal credibility values. There are other ways to allocate the covariances, but under this particular allocation, one arrives at a generalization of Formula (2.8) in which optimal credibility is not only the best weight to use in a linear estimate of the mean, but also it is equal to the percentage reduction in the variance of the estimated mean achieved by using that optimal weight.

4. When does splitting reduce mean square parameter error?

To study whether a split plan reduces minimum mean square parameter error, we first define the reduction in minimal mean square parameter error: $\Delta(\epsilon_0^2) = \epsilon_0^2$

(NS) – ϵ_0^2 (SP). In Corollary 1 of Appendix B it is proved that this can be expressed in terms involving the optimal no-split and split credibilities:

$$\begin{aligned} \Delta \epsilon_0^2 &= \tau^2(1 - z^*) - (\tau_1^2 + \pi)(1 - z_1^*) \\ &\quad - (\tau_2^2 + \pi)(1 - z_2^*) \\ &= (\tau_1^2 + \pi)(z_1^* - z^*) + (\tau_2^2 + \pi)(z_2^* - z^*). \end{aligned} \quad (4.1)$$

In words, the difference in mean square parameter error is the parameter risk allocated to the first component including its share of the parameter covariance times the difference between the optimal credibility of the first component and the optimal credibility of unsplit losses plus the corresponding term for the second component.

4.1. Comparison of credibilities

We immediately see from Equation (4.1) that error improvement at least requires one of the split plan credibility values to be larger than the credibility from the original no-split plan. In this generality, where the split is arbitrary and not necessarily between primary and excess losses, there is no reason why the split plan should reduce mean square parameter error.

We will use Equation (4.1) to derive intuitively accessible formulas that summarize what is required for a split to reduce the minimal mean square parameter error. But, before presenting our main results, it is useful to consider a simple example to hone our intuition.

4.1.1. The even split example

Consider an “even split” where the two components have the same process variance and the same parameter variance as seen, for instance, in Exhibit 1 Sheet 2. In the general case of an even split, we have $\sigma_1^2 = \sigma_2^2$, $\tau_1^2 = \tau_2^2$ and $\pi = \tau_1\tau_2$. It follows that $\lambda_1^2 = \lambda_2^2$ and that $z_1^* = z_2^*$. We derive:

$$\begin{aligned} z_1^* &= \frac{\lambda_2^2(\tau_1^2 + \pi) - \kappa(\tau_2^2 + \pi)}{\lambda_1^2\lambda_2^2 - \kappa^2} \\ &= \frac{\lambda_1^2(\tau_1^2 + \pi) - \kappa(\tau_1^2 + \pi)}{(\lambda_1^2 - \kappa)(\lambda_1^2 + \kappa)} = \frac{(\tau_1^2 + \pi)}{(\lambda_1^2 + \kappa)}. \end{aligned} \quad (4.2)$$

$$\begin{aligned} z^* &= \frac{\tau^2}{\lambda^2} = \frac{\tau_1^2 + \tau_2^2 + 2\pi}{\lambda_1^2 + \lambda_2^2 + 2\kappa} = \frac{(2\tau_1^2 + 2\pi)}{(2\lambda_1^2 + 2\kappa)} \\ &= \frac{(\tau_1^2 + \pi)}{(\lambda_1^2 + \kappa)}. \end{aligned} \quad (4.3)$$

Since all the optimal credibilities are equal, it follows from Equation (4.1) that the split plan does not reduce mean square error. Note this result holds no matter what the process covariance is between the components. So, for example, a split where each component is equal to half the loss gains us nothing. Neither does a plan where we toss a fair coin to decide if a claim belongs to one component or the other. Of course, there is no intuitive reason to expect either of these split plans could improve the accuracy of our credibility-weighted estimate of the mean. More generally, our intuition is that a split cannot improve the accuracy of the final estimate if it does not meaningfully use additional information beyond that which was used for the no-split plan.

4.2. Formula for the difference in minimal mean square error

We are now ready to state several key formulas for the difference in minimal mean square error. The first expresses that difference in terms of the process and parameter variances and covariances:

$$\Delta(\epsilon_0^2) = \frac{1}{D\lambda^2} \left((\tau_1^2 + \pi)(\sigma_2^2 + \rho) - (\sigma_1^2 + \rho)(\tau_2^2 + \pi) \right)^2. \quad (4.4)$$

The proof of this formula is shown in Corollary 2 of Appendix B. The proof also yields the following important formula that expresses the mean square error reduction in terms of the square of difference of the resulting split credibility values:

$$\Delta(\epsilon_0^2) = \frac{D}{\lambda^2} (z_1^* - z_2^*)^2. \quad (4.5)$$

This result is Theorem 2 in Appendix B.

4.3. What makes a split effective?

By definition, we regard a split as effective when it improves our estimate of the mean. More precisely,

when the mean square error criterion is used to judge the quality of an estimate, an effective split is one that leads to a significant reduction in the least mean square error of our estimate in comparison with that obtained using the unsplit plan. So a split is effective if it produces a relatively large value for the difference in optimal mean square errors, $\Delta(\epsilon_0^2)$.

4.3.1. The most effective split possible

Examining Equation (4.4), we see that the most effective split possible would be one that puts all the process risk in one component and all the parameter risk in the other. With a split that extreme, both the process and parameter covariances will be zero. Thus we would have $\pi = \rho = \kappa = 0$ and it would follow from Equation (3.8) that one component would have credibility of 100% and the other would have credibility of 0%. Further, from Equation (3.10) it would follow that the mean square parameter error of the resulting split credibility estimate would be zero! In other words, our split credibility estimate would be exactly right because it is based on a perfect separation of noise from signal. Exhibit 1 Sheet 3 provides a numerical example.

In any realistic scenario it will be impossible to make such a clean split of the process and parameter risk. However, the intuition still holds. The key for a split to be effective is that it must lead to a proportionately different allocation of the total process and total parameter variances.

4.3.2. Ineffective splits and proportional allocation

We have already seen that an even split is ineffective. This can be generalized using the process and parameter variance allocations from Equation (4.4) as displayed in Table 2.

Table 2. Equation (4.4) allocation of process and parameter variance

	A_1	A_2
Allocated Process Variance	$\sigma_1^2 + \rho$	$\sigma_2^2 + \rho$
Allocated Parameter Variance	$\tau_1^2 + \pi$	$\tau_2^2 + \pi$

Each component gets assigned a total process variance equal to the sum of its own process variance plus the process covariance. The parameter variance is allocated the same way. Equation (4.4) implies that if the splitting leads to an allocation where each component has the same ratio of allocated parameter variance to allocated process variance under this particular allocation, then there will be no improvement at all in optimal MSE due to the split. To state this mathematically,

$$\text{If } \frac{\tau_1^2 + \pi}{\sigma_1^2 + \rho} = \frac{\tau_2^2 + \pi}{\sigma_2^2 + \rho}, \text{ then } \Delta(\epsilon_0^2) = 0. \quad (4.6)$$

This is demonstrated with a numerical example in Exhibit 1, Sheet 4.

4.3.3. Different optimal credibility values by component

From Equation (4.5) it follows that an effective split is one that produces optimal credibility values that differ substantially for the two components. One can also argue this makes sense intuitively by reasoning backwards. If a split were to lead to optimal credibility values that were the same for both parts of the split, then one might as well have left them unified and applied a single credibility value to the undivided whole.

4.3.4. The impact of covariance

When comparing two different ways of splitting, the split with the larger process covariance is not necessarily any more or less effective than the other one. Compare the base case, Exhibit 1 Sheet 1, with Exhibit 1 Sheet 5. The split shown in Sheet 5 has a larger process covariance, but it is more effective: it reduces MSE more than the split in Sheet 1. In contrast, the split example in Sheet 6 has the same covariances as the one in Sheet 5, but it is not as effective as the base case split. The conclusion is that the impact of covariance is complicated. The contrary examples in Sheet 5 and Sheet 6 were obtained by adjusting σ_1 , σ_2 , and ρ to exacerbate or diminish the differential allocation of process and parameter risk while obeying the overall constraint that “ σ ” stay fixed in the equation, $\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\rho$. The same non-definitive

result is true for splits with different parameter covariances. Corresponding sets of counterexamples can be readily constructed along the same lines.

4.3.5. Effective split with negative credibility for one component

It is possible to have an effective plan that has a negative credibility value for one component. A particular instance is shown in Exhibit 1, Sheet 7. Such a situation may arise when one split component, the volatile one, is given the lion's share of the process risk, a very modest share of the parameter risk, and the split produces a parameter covariance roughly as large as the parameter variance of the volatile component. Because the volatile component has so much process risk and so little parameter risk, one would not want to give it any weight. The high parameter covariance allows us to gain more accurate information about the mean of the volatile component from the results of its better-behaved sister component than we can gain from the results of the volatile component itself. This provides some intuitive justification for how a negative credibility can occur and how a split with such a negative credibility component may nonetheless be effective.

5. Credibility with losses from a single severity type model

Now we will derive credibility formulas for losses that arise from a model in which claim counts are generated by a single random variable and each claim severity is conditionally an independent sample from the single severity distribution. We will further assume, to simplify the discussion, that our uncertainty about severity is confined to lack of precise knowledge of its scale. We refer to such a model as a Single Severity Type Model with Severity Scale Uncertainty. The CRM is an example of such a model.

To begin the mathematical derivation, let N be the number of claims and write $X(i)$ for the loss from the i th claim. Assume each $X(i)$ is an independent random sample of the severity random variable,

X . Further suppose each $X(i)$ is independent of the claim count. Define the actual loss, A , via: $A = X(1) + X(2) + \dots + X(N)$. Now suppose N is parametrically dependent on a parameter, θ_N , and that X is parametrically dependent on a parameter, θ_X . Assume θ_N and θ_X have prior distributions that are independent. We will abuse notation and usually drop the subscripts, N and X on θ . Define $\mu_N(\theta) = E[N|\theta]$, $\mu_X(\theta) = E[X|\theta]$, $\sigma_N^2(\theta) = \text{Var}(N|\theta)$, and $\sigma_X^2(\theta) = \text{Var}(X|\theta)$. Then take expectations and variances with respect to the priors to define $\mu_N = E[\mu_N(\theta)]$, $\mu_X = E[\mu_X(\theta)]$, $\sigma_N^2 = E[\sigma_N^2(\theta)]$, $\sigma_X^2 = E[\sigma_X^2(\theta)]$, $\tau_N^2 = \text{Var}(\mu_N(\theta))$, and $\tau_X^2 = \text{Var}(\mu_X(\theta))$.

We will now derive the process and parameter variance of loss using terms based on the claim count and claim severity. The conditional mean and variance are given by:

$$\mu_A(\theta) = \mu_N(\theta) \cdot \mu_X(\theta). \quad (5.1)$$

$$\sigma_A^2(\theta) = \mu_N(\theta) \cdot \sigma_X^2(\theta) + \sigma_N^2(\theta) \cdot (\mu_X(\theta))^2. \quad (5.2)$$

Taking expectations with respect to the priors, we find the process and parameter variances:

$$\sigma_A^2 = \mu_N \cdot \sigma_X^2 + \sigma_N^2 \cdot (\tau_X^2 + \mu_X^2). \quad (5.3)$$

$$\tau_A^2 = \tau_N^2 \cdot \tau_X^2 + \tau_N^2 \cdot \mu_X^2 + \mu_N^2 \cdot \tau_X^2. \quad (5.4)$$

Note in Equation (5.3), the expected process variance contains a term that includes the severity parameter variance.

Plugging these into the basic no-split credibility formula, Equation (2.6), we find the optimal credibility is given as:

$$z^* = \frac{\tau_N^2 \cdot \tau_X^2 + \tau_N^2 \cdot \mu_X^2 + \mu_N^2 \cdot \tau_X^2}{\tau_N^2 \cdot \tau_X^2 + \tau_N^2 \cdot \mu_X^2 + \mu_N^2 \cdot \tau_X^2 + \mu_N \cdot \sigma_X^2 + \sigma_N^2 \cdot (\tau_X^2 + \mu_X^2)}. \quad (5.5)$$

If we assume N is conditionally Poisson so that $\sigma_N^2 = \mu_N$, the process variance is:

$$\sigma_A^2 = \mu_N \cdot (\sigma_X^2 + \tau_X^2 + \mu_X^2). \quad (5.6)$$

So with conditionally Poisson claim counts, the formula for optimal credibility is given as:

$$z^* = \frac{\tau_N^2 \cdot \tau_X^2 + \tau_N^2 \cdot \mu_X^2 + \mu_N^2 \cdot \tau_X^2}{\tau_N^2 \cdot \tau_X^2 + \tau_N^2 \cdot \mu_X^2 + \mu_N^2 \cdot \tau_X^2 + \mu_N \cdot (\sigma_X^2 + \tau_X^2 + \mu_X^2)}. \quad (5.7)$$

5.1. Credibility when losses follow the collective risk model

We will now examine the optimal credibility formula, Equation (5.7), when account loss distributions follow the usual collective risk model. Let N be Poisson with parameter $n\chi$, where $E[\chi] = 1$ and $\text{Var}(\chi) = c$. Under these assumptions, we have $\mu_N = n$, $\sigma_N^2 = n$, and $\tau_N^2 = cn^2$. The parameter, c , is called the contagion. Let X be conditionally exponential with mean $s\beta$ where $s > 0$, $E[\beta] = 1$ and $\text{Var}(\beta) = b$. The parameter, b , is called the mixing parameter. With this notation we have $\mu_X = s$, $\sigma_X^2 = E[s^2\beta^2] = s^2(1 + b)$, and $\tau_X^2 = \text{Var}(s\beta) = s^2b$. It follows that $\mu_A = ns$ and:

$$\begin{aligned} \sigma_A^2 &= E[n\chi \cdot E[X^2 | \beta]] = ns^2 \cdot E[2\beta^2] \\ &= 2ns^2 \cdot (1 + b). \end{aligned} \quad (5.8)$$

$$\begin{aligned} \tau_A^2 &= \text{Var}(n\chi s\beta) \\ &= n^2 s^2 \cdot ((1 + c)(1 + b) - 1). \end{aligned} \quad (5.9)$$

Thus the optimal credibility is given as:

$$\begin{aligned} z^* &= \frac{n^2 s^2 \cdot ((1 + c)(1 + b) - 1)}{n^2 s^2 \cdot ((1 + c)(1 + b) - 1) + 2ns^2(1 + b)} \\ &= \frac{n^2 \cdot ((1 + c)(1 + b) - 1)}{n^2 \cdot ((1 + c)(1 + b) - 1) + 2n(1 + b)}. \end{aligned} \quad (5.10)$$

For a specific numerical example, suppose $s = 10$, $n = 10$, $b = .25$, and $c = .20$. Then using 5.8 the process variance is $2 \cdot 10 \cdot 100 \cdot 1.25 = 2,500$ and applying 5.9 the parameter variance is $100 \cdot 100 \cdot (1.25 \cdot 1.20 - 1) = 5,000$. Thus we find the credibility is $^{2/3}$ or about 67%.

6. Split credibility with losses from a single type model

Next we derive comparable split credibility formulas. Given a per occurrence split point, k , and an occurrence of size X , we define $X_p = \min(X, k)$ as the primary severity and $X_e = X - \min(X, k)$ as the excess severity. Observe, under this definition X_e will have a mass point at zero equal to the probability that X is less than or equal to the split point. In other words, X_e is *not* the conditional excess severity. We have adopted this approach so that the primary, excess, and total losses all have the same claim count distribution. This simplifies some derivations. We now define the actual primary loss, $A_p = X_p(1) + X_p(2) + \dots + X_p(N)$ and the actual excess loss, $A_e = X_e(1) + X_e(2) + \dots + X_e(N)$. The primary and excess process and parameter variances are given as:

$$\sigma_{A_p}^2 = \mu_N \cdot \sigma_{X_p}^2 + \sigma_N^2 \cdot (\tau_{X_p}^2 + \mu_{X_p}^2). \quad (6.1)$$

$$\sigma_{A_e}^2 = \mu_N \cdot \sigma_{X_e}^2 + \sigma_N^2 \cdot (\tau_{X_e}^2 + \mu_{X_e}^2). \quad (6.2)$$

$$\tau_{A_p}^2 = \tau_N^2 \cdot \tau_{X_p}^2 + \tau_N^2 \cdot \mu_{X_p}^2 + \mu_N^2 \cdot \tau_{X_p}^2. \quad (6.3)$$

$$\tau_{A_e}^2 = \tau_N^2 \cdot \tau_{X_e}^2 + \tau_N^2 \cdot \mu_{X_e}^2 + \mu_N^2 \cdot \tau_{X_e}^2. \quad (6.4)$$

We can derive the following formulas for the covariances:

$$\begin{aligned} \rho &= E[\text{Cov}(A_p, A_e)] \\ &= (\sigma_N^2 - \mu_N) \cdot \mu_{X_p} \cdot \mu_{X_e} + k \cdot \mu_{X_e} \cdot \mu_N. \end{aligned} \quad (6.5)$$

$$\begin{aligned} \pi &= \text{Cov}(E[A_p | \theta], E[A_e | \theta]) \\ &= (\tau_N^2 + \mu_N^2) \cdot \pi_X + \tau_N^2 \cdot \mu_{X_p} \cdot \mu_{X_e}. \end{aligned} \quad (6.6)$$

Here $\pi_X = \text{Cov}(E[X_p | \theta], E[X_e | \theta])$ denotes the parameter covariance of the primary and excess severities. The derivations are shown in Appendix C.

Assuming claim counts are conditionally Poisson, the process variance terms simplify to:

$$\sigma_{A_p}^2 = \mu_N \cdot (\sigma_{x_p}^2 + \tau_{x_p}^2 + \mu_{x_p}^2). \quad (6.7)$$

$$\sigma_{A_e}^2 = \mu_N \cdot (\sigma_{x_e}^2 + \tau_{x_e}^2 + \mu_{x_e}^2). \quad (6.8)$$

$$\rho = k \cdot \mu_N \cdot \mu_{x_e}. \quad (6.9)$$

6.1. Split credibility formulas under the collective risk model

We will now apply the CRM structure of priors to evaluate the required terms in the split credibility formulas shown in (3.8). We already have the formulas for the variances of the claims counts. Using the CRM assumption that severity is conditionally exponential, we may write the following formulas for the conditional means of the primary and excess severities.

$$\begin{aligned} \mu_{x_p}(\theta) &= s\beta(1 - \exp(-k/(s\beta))). \\ \mu_{x_e}(\theta) &= s\beta \exp(-k/(s\beta)). \end{aligned} \quad (6.10)$$

We can also derive the formulas for the conditional severity process variances:

$$\begin{aligned} \sigma_{x_p}^2(\theta) &= \int_0^k dx \, x^2 \cdot (s\beta)^{-1} \cdot \exp(-x/(s\beta)) \\ &\quad + k^2 \exp(-k/(s\beta)) \\ &\quad - s^2\beta^2 (1 - \exp(-k/(s\beta)))^2 \\ &= s^2\beta^2 - 2s\beta k \cdot \exp(-k/(s\beta)) \\ &\quad - s^2\beta^2 \cdot \exp(-2k/(s\beta)). \end{aligned} \quad (6.11)$$

$$\begin{aligned} \sigma_{x_e}^2(\theta) &= \int_k^\infty dx \, (x - k)^2 \cdot (s\beta)^{-1} \cdot \exp(-x/(s\beta)) \\ &\quad - s^2\beta^2 \cdot \exp(-2k/(s\beta)) \\ &= 2s^2\beta^2 \cdot \exp(-k/(s\beta)) \\ &\quad - s^2\beta^2 \cdot \exp(-2k/(s\beta)). \end{aligned} \quad (6.12)$$

Now, in conformance with the CRM structure, assume that β is such that $\gamma = 1/\beta$ is Gamma distrib-

uted. Let γ have shape parameter α and scale parameter λ such that $E[\gamma] = \alpha/\lambda$ and $\text{Var}(\gamma) = \alpha/\lambda^2$. It follows that $E[\beta] = \lambda/(\alpha - 1)$ and $E[\beta^2] = \lambda^2/\{(\alpha - 1)(\alpha - 2)\}$ as shown in 6.13 and 6.14:

$$\begin{aligned} E[\beta] &= E[1/\gamma] \\ &= \int_0^\infty d\gamma \frac{1}{\gamma} \frac{\lambda^\alpha}{\Gamma(\alpha)} \gamma^{(\alpha-1)} \cdot \exp(-\lambda\gamma) \\ &= \frac{\lambda}{\alpha-1} \int_0^\infty d\gamma \frac{\lambda^{\alpha-1}}{\Gamma(\alpha-1)} \gamma^{(\alpha-1)-1} \cdot \exp(-\lambda\gamma) \\ &= \frac{\lambda}{\alpha-1}. \end{aligned} \quad (6.13)$$

$$\begin{aligned} E[\beta^2] &= E[1/\gamma^2] \\ &= \int_0^\infty d\gamma \frac{1}{\gamma^2} \frac{\lambda^\alpha}{\Gamma(\alpha)} \gamma^{(\alpha-1)} \cdot \exp(-\lambda\gamma) \\ &= \frac{\lambda^2}{(\alpha-1)(\alpha-2)} \cdot \\ &\quad \int_0^\infty d\gamma \frac{\lambda^{\alpha-2}}{\Gamma(\alpha-2)} \gamma^{(\alpha-2)-1} \cdot \exp(-\lambda\gamma) \\ &= \frac{\lambda^2}{(\alpha-1)(\alpha-2)}. \end{aligned} \quad (6.14)$$

It also follows that the density of β is given as:

$$h(\beta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \beta^{-(\alpha+1)} \cdot \exp(-\lambda/\beta). \quad (6.15)$$

With this density, we can derive the unconditional severities and the process and parameter variances and covariances. To ensure the derivations are clearly understood, we will show the first one in some detail.

$$\begin{aligned} \mu_{x_p} &= E[\mu_{x_p}(\theta)] \\ &= E[(s\beta) \cdot (1 - \exp(-k/(s\beta)))] \\ &= s \int_0^\infty d\beta \, \beta \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} \beta^{-(\alpha+1)} \exp(-\lambda/\beta) \\ &\quad - s \int_0^\infty d\beta \, \beta \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} \beta^{-(\alpha+1)} \exp(-(\lambda + (k/s))/\beta) \\ &= s \frac{\lambda}{\alpha-1} \left(1 - \left(\frac{\lambda}{\lambda + k/s} \right)^{\alpha-1} \right). \end{aligned} \quad (6.16)$$

This expression is the formula for the limited expected value of a Pareto severity distribution with scale, $s\lambda$, and shape parameter, α .

Recall we have also assumed in Section 5.1 that $E[\beta] = 1$ and that $\text{Var}(\beta) = b$. It follows immediately from (6.13) that $\lambda = \alpha - 1$ and we can then use (6.14) to show $\alpha = 2 + 1/b$:

$$\begin{aligned} \text{Var}(\beta) &= E[\beta^2] - E[\beta]^2 \\ &= \frac{\lambda^2}{(\alpha-1)(\alpha-2)} - \frac{\lambda^2}{(\alpha-1)^2} = \frac{\lambda}{\alpha-2} - 1 = b \\ &\Rightarrow \frac{\lambda}{\alpha-2} = 1 + b \Rightarrow \frac{\alpha-1}{\alpha-2} = 1 + b \\ &\Rightarrow \frac{\alpha-2+1}{\alpha-2} = 1 + b \Rightarrow \frac{1}{\alpha-2} = b \\ &\Rightarrow \alpha = 2 + \frac{1}{b}. \end{aligned} \tag{6.17}$$

Using this to substitute into (6.16), we have:

$$\begin{aligned} \mu_{x_p} &= s \left(1 - \left(\frac{\lambda}{(\lambda + k/s)} \right)^\lambda \right) \\ &= s \left(1 - \left(1 + \frac{kb}{s(b+1)} \right)^{-(1+1/b)} \right). \end{aligned} \tag{6.18}$$

Given the way we have defined X_e , it follows that:

$$\mu_{x_e} = \mu_x - \mu_{x_p} = s \left(\frac{\lambda}{(\lambda + k/s)} \right)^\lambda = s \left(1 + \frac{kb}{s(b+1)} \right)^{-(1+1/b)}. \tag{6.19}$$

Using (6.11), (6.12), and (6.17) and applying similar logic, we can derive the following formulas for the severity process and parameter variances:

$$\begin{aligned} \sigma_{x_p}^2 &= E[\sigma_{x_p}^2(\theta)] = \\ &= E \left[\begin{aligned} &s^2\beta^2 - 2s\beta k \cdot \exp(-k/(s\beta)) \\ &-s^2\beta^2 \cdot \exp(-2k/(s\beta)) \end{aligned} \right] \\ &= s^2 \cdot (1+b) - 2sk \left(1 + \frac{kb}{s(b+1)} \right)^{-(1+1/b)} \\ &\quad - s^2 \cdot (1+b) \left(1 + \frac{2kb}{s(b+1)} \right)^{-(1/b)}. \end{aligned} \tag{6.20}$$

$$\begin{aligned} \sigma_{x_e}^2 &= E[\sigma_{x_e}^2(\theta)] = \\ &= E \left[\begin{aligned} &2s^2\beta^2 \cdot \exp(-k/(s\beta)) \\ &-s^2\beta^2 \cdot \exp(-2k/(s\beta)) \end{aligned} \right] \\ &= 2s^2 \cdot (1+b) \left(1 + \frac{kb}{s(b+1)} \right)^{-(1/b)} \\ &\quad - s^2 \cdot (1+b) \left(1 + \frac{2kb}{s(b+1)} \right)^{-(1/b)}. \end{aligned} \tag{6.21}$$

$$\begin{aligned} \tau_{x_p}^2 &= \text{Var}(\mu_{x_p}(\theta)) = \text{Var}(s\beta(1 - \exp(-k/(s\beta)))) \\ &= s^2 \cdot (1+b) \left(\begin{aligned} &1 - 2 \left(1 + \frac{kb}{s(b+1)} \right)^{-1/b} \\ &+ \left(1 + \frac{2kb}{s(b+1)} \right)^{-1/b} \end{aligned} \right) \\ &\quad - s^2 \left(1 - \left(1 + \frac{kb}{s(b+1)} \right)^{-(1+1/b)} \right)^2. \end{aligned} \tag{6.22}$$

$$\begin{aligned} \tau_{x_e}^2 &= \text{Var}(\mu_{x_e}(\theta)) = \text{Var}(s\beta \exp(-k/(s\beta))) \\ &= s^2 \cdot (1+b) \left(1 + \frac{2kb}{s(b+1)} \right)^{-1/b} - s^2 \left(1 + \frac{kb}{s(b+1)} \right)^{-2(1+1/b)}. \end{aligned} \tag{6.23}$$

Finally we turn to the process and parameter covariances of the severity. We can derive:

$$\begin{aligned} \rho_x &= E \left[\begin{aligned} &\beta s k \exp(-k/(s\beta)) - \beta s \exp(-k/(s\beta)) \\ &\cdot \beta s (1 - \exp(-k/(s\beta))) \end{aligned} \right] \\ &= sk \left(1 + \frac{kb}{s(b+1)} \right)^{-(1+1/b)} - s^2 (1+b) \left(1 + \frac{kb}{s(b+1)} \right)^{-1/b} \\ &\quad + s^2 (1+b) \left(1 + \frac{2kb}{s(b+1)} \right)^{-1/b}. \end{aligned} \tag{6.24}$$

$$\begin{aligned} \pi_x &= \text{Cov} \left(\begin{aligned} &\beta s (1 - \exp(-k/(s\beta))), \\ &\beta s \exp(-k/(s\beta)) \end{aligned} \right) \\ &= s^2 (1+b) \left(\begin{aligned} &\left(1 + \frac{kb}{s(b+1)} \right)^{-1/b} \\ &- \left(1 + \frac{2kb}{s(b+1)} \right)^{-1/b} \end{aligned} \right) \\ &\quad - s^2 \left(1 - \left(1 + \frac{kb}{s(b+1)} \right)^{-(1+1/b)} \right) \left(1 + \frac{kb}{s(b+1)} \right)^{-(1+1/b)}. \end{aligned} \tag{6.25}$$

While these formulas look forbidding, they are actually not too difficult to program. In the next Section, we will use the formulas in Section 6 to generate unsplit and split credibilities under the CRM structure.

7. Split credibility results for CRM models

With the formulas derived in Section 6, we have enough to compute split credibilities and the error reduction due to a split under CRM. We will look at examples to illustrate that some of the undesirable behaviors that can exist under an arbitrary split plan can also arise under a primary-excess split plan operating on an underlying CRM structure. First, recall the example at the end of Section 5.1 in which the mean claim count is 10, the mean severity is 10, the contagion, c , is .250 and the severity mixing parameter, b , is 0.200. In this example, total mean loss is 100 and we had previously seen the optimal unsplit credibility is 67%. If we now introduce a split point of 10, we find, as shown in Exhibit 2, Sheet 1 that the optimal primary and excess credibilities are both 67%. As we know from Equation (4.5), this implies the primary-excess split is no better than the un-split plan. We can also see from Exhibit 2, Sheet 1 that this is an example of proportional allocation of process and parameter risk in which the primary layer gets 36% of the process variance and 36% of the parameter variance under the covariance allocation in Table 2. Thus, it also follows from Equation (4.6) that this split is ineffective when the loss model is the CRM with the parameters given.

With other parameters, this same split can be effective. In Exhibit 2, Sheet 2, the assumptions are the same as in Sheet 1 except the mixing parameter is reduced from 0.250 to 0.025. With uncertainty about severity dramatically reduced, the excess layer gets a modest allocation of parameter risk and large allocation of process risk. The opposite is true for the primary layer: it has less process risk and more of the parameter risk allocated to it. This

is the scenario under which a primary-excess split will be effective and behave in the usual way actuaries expect. The primary layer in the example has 92% credibility and the excess layer has only 11%. The introduction of the split reduced optimal MSE by roughly 12% of optimal MSE of the unsplit credibility estimate.

However, under the CRM a choice of parameters that puts more parameter risk in the excess layer can produce an inversion of primary and excess credibilities in a plan which is still effective. In Exhibit 2, Sheet 3, the severity mixing parameter is reset back to 0.250 while the contagion is reduced to 0.020. This results in a primary credibility of only 3% and an excess credibility 72%. Such an inversion of primary and excess credibilities can never happen in the NCCI split rating plan. The split in this example reduces MSE by 9% of the MSE of the unsplit estimate.

Finally in Exhibit 2, Sheet 4 is an example of unusual behavior in which the primary credibility is negative and the excess credibility is positive. This is achieved by reducing mean claim counts and count parameter risk while boosting mean severity and severity risk. The primary layer ends up with a very modest parameter variance, one that is smaller than the parameter covariance. In the example, primary credibility is -33% and the excess credibility is 43%.

7.1. Effectiveness and well-behaved primary-excess splits

We say a primary-excess split plan is effective if it appreciably reduces mean square error and we say it is well-behaved if it has no primary-excess credibility inversions and all credibility values are between zero and unity.

The CRM examples show that, with the right set of parameters, there are MSE optimal primary-excess split credibility plans that are both effective and well-behaved. Based on the examples, we see this happens when:

- There is substantial parameter risk due to claim count uncertainty,

- Most process risk is due to volatility of severity,
- The split allocates a disproportionate amount of parameter risk to the primary layer
- The split allocates a disproportionate amount of process risk to the excess layer.

The CRM structure with its single severity subject to scale parameter uncertainty readily allows parameter selections that satisfy these conditions.

However the CRM examples also show that a primary-excess split does not have to produce an effective or well-behaved plan. An effective plan with inversions can result when severity parameter risk drives the overall parameter risk and the split puts a relatively large amount of parameter risk in the excess layer. In such a scenario, the effect of a split intuitively is to deprive the primary layer of much of the information about severity and thus to diminish its allocation of parameter risk and therefore diminish the primary layer credibility. There is nothing in the structure of the model to prevent the primary layer credibility from falling below the excess layer credibility. From an a priori perspective, there is nothing anomalous or bizarre about such scenarios. Why can't we have a fairly small uncertainty about mean claim counts and larger relative uncertainty about mean severity, as in Exhibit 2, Sheet 3? From this perspective, middle scenarios in which split credibility is well-behaved, but only modestly effective, also seem quite reasonable.

8. Conclusion

To summarize, we have shown that analysis of split experience rating requires analysis of the allocation of the process and parameter risk to the primary and excess layers and to the covariances between the layers. We stressed it is insufficient to focus on volatility alone: low volatility is not synonymous with high credibility. We have derived formulas for the mean square parameter errors in the unsplit and split plans. We showed that credibility is the ratio by which parameter risk is reduced in an optimal estimate as well the weight given to experience versus prior belief in the linear estimation formula. We then

extended that error reduction interpretation to apply in a split credibility context.

By taking differences and simplifying, we found a formula for the reduction in mean square parameter error attributable to splitting. Interpreting this formula, we found that mean square error of the estimate would be effectively reduced if the split produced a differential allocation of process and parameter risk. We also saw that credibility for one component had to be bigger and credibility for the other component smaller than the credibility for the unsplit total in order for there to be any error reduction from splitting.

We have shown with examples using the standard CRM that a primary-excess split does not always improve accuracy to any great degree, nor does it always produce non-negative credibility values or a primary layer credibility that is bigger than the excess layer credibility. We saw that these disquieting results were not an artifact of making odd parameter choices but were inherent in the nature of the primary-excess splitting process operating on plausible models with reasonable parameter selections.

To summarize, this work has established a solid mathematical foundation for split credibility. However, it has also highlighted potential weakness: under a standard model such as the CRM, primary-excess splitting does not automatically confer a great advantage; nor does it produce well-behaved credibility values. It is possible a model having different types of claims with different severities or other more complex severity parameter risk structure might allow a more effective split of noise from signal while at the same time preventing inversions and negative credibility values. More support for splitting might also be found by investigating optimality criteria different from minimal mean square error, or perhaps by studying use of optimal credibilities subject to constraints that promote good behavior. Perhaps most promising is to carry the error analysis further to a plan that has a split on capped losses and also uses Mod extension to estimate total uncapped losses. This is the actual type of plan used by the NCCI. Work along these or similar lines might provide a stronger conceptual foundation for the use of split credibility.

Acknowledgments

The author gratefully acknowledges David Clark, Robin Gillam, and Gary Venter for their insights and observations. The members of the review committee also merit commendation for their detailed and thoughtful comments. These contributions helped to motivate and substantially improve the paper.

Disclaimers

The opinions expressed are solely those of the author and are not presented as a statement of the views or practices of any past or present employer or client of the author. The author assumes no liability whatsoever for any damages that may result directly or indirectly from use or reliance on any observation, opinion, idea, or method presented in this paper.

References

- Gillam, W. R., "Parameterizing the Workers Compensation Experience Rating Plan," *Proceedings of the Casualty Actuarial Society* 76, 1989, pp. 21–56.
- Gillam, W. R., "Workers Compensation Experience Rating: What Every Actuary Should Know," *Proceedings of the Casualty Actuarial Society* 79, 1992, pp. 215–239.
- Heckman, P. E., and G. G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *Proceedings of the Casualty Actuarial Society* 70, 1983, pp 22–61.
- Mahler, H. C., Discussion of "An Analysis of Experience Rating," *Proceedings of the Casualty Actuarial Society* 74, 1987, pp 119–189.

- Meyers, G. G., "An Analysis of Experience Rating," *Proceedings of the Casualty Actuarial Society* 72, 1985, pp. 278–316.
- NCCI (National Council on Compensation Insurance), *Experience Rating Plan Manual for Workers Compensation and Employers Liability Insurance*, March 2002.
- Teng, M. T. S., "Pricing Workers Compensation Large Deductible and Excess Insurance," *Casualty Actuarial Society Forum*, Winter 1994, pp. 413–437.
- Venter, G. G., "Experience Rating—Equity and Predictive Accuracy," *NCCI Digest* 2:1, April 1987, pp. 1–9.

Abbreviations and Notations

- b , severity mixing parameter
- c , contagion
- CRM, Collective Risk Model
- CV, Coefficient of Variation
- Experience mod, Experience modification factor
- $\Delta(\epsilon_0^2)$, reduction in minimal mean square estimation error
- κ , total covariance
- μ , the mean
- NCCI, National Council on Compensation Insurance
- NS, Non-Split
- π , parameter covariance
- ρ , process covariance
- σ , process standard deviation
- SP, Split
- τ , parameter standard deviation
- z^* , optimal credibility

Exhibit 1 Sheet 1. Split Credibility Example: Base Case

Loss Component	Unsplit Credibility		Split Credibility	
	A	A ₁	Covariance	A ₂
Mean	100	50		50
Process Variance	400	100	70	160
Parameter Variance	200	100	25	50
Total Variance	600	200	95	210
MSE Optimal No-split Credibility	33.3%			
D			32,975	
MSE Optimal Split Credibility		58.0%		9.5%
Formula 4.4 Variance Allocations				
Allocated Process Variance		170		230
Allocated Parameter Variance		125		75
Total Variance Allocation		295		305
Initial MSE	200		200	
MSE of Credibility Estimate	133		120	
Reduction in MSE	67		80	
Reduction in MSE as % of Initial MSE	33.3%		39.8%	
Addtl reduction in MSE due to split	n/a		13	
Addtl reduction as % of Initial MSE	n/a		6.5%	
Addtl reduction as % of MSE under no-split plan	n/a		9.7%	

Exhibit 1 Sheet 2. Even Split is Ineffective

Loss Component	Unsplit Credibility		Split Credibility	
	A	A ₁	Covariance	A ₂
Mean	100	50		50
Process Variance	400	120	80	120
Parameter Variance	200	90	10	90
Total Variance	600	210	90	210
MSE Optimal No-split Credibility	33.3%			
D			36,000	
MSE Optimal Split Credibility		33.3%		33.3%
Formula 4.4 Variance Allocations				
Allocated Process Variance		200		200
Allocated Parameter Variance		100		100
Total Variance Allocation		300		300
Initial MSE	200		200	
MSE of Credibility Estimate	133		133	
Reduction in MSE	67		67	
Reduction in MSE as % of Initial MSE	33.3%		33.3%	
Addtl reduction in MSE due to split	n/a		0	
Addtl reduction as % of Initial MSE	n/a		0.0%	
Addtl reduction as % of MSE under no-split plan	n/a		0.0%	

Exhibit 1 Sheet 3. Perfect Split of Process and Parameter Risk Produces a Perfect Estimate

Loss Component	Unsplit Credibility	Split Credibility		
	A	A ₁	Covariance	A ₂
Mean	100	50		50
Process Variance	400	—	0	400
Parameter Variance	200	200	0	—
Total Variance	600	200	0	400
MSE Optimal No-split Credibility	33.3%			
D			80,000	
Split Credibility		100.0%		0.0%
Formula 4.4 Variance Allocations				
Allocated Process Variance		—		400
Allocated Parameter Variance		200		—
Total Variance Allocation		200		400
Initial MSE	200		200	
MSE of Credibility Estimate	133		0	
Reduction in MSE	67		200	
Reduction in MSE as % of Initial MSE	33.3%		100.0%	
Addtl reduction in MSE due to split	n/a		133	
Addtl reduction as % of Initial MSE	n/a		66.7%	
Addtl reduction as % of MSE under no-split plan	n/a		100.0%	

Exhibit 1 Sheet 4. Same Proportional Allocation of Process and Parameter Risk is Ineffective

Loss Component	Unsplit Credibility	Split Credibility		
	A	A ₁	Covariance	A ₂
Mean	100	50		50
Process Variance	400	100	80	140
Parameter Variance	200	80	10	100
Total Variance	600	180	90	240
MSE Optimal No-split Credibility	33.3%			
D			35,100	
MSE Optimal Split Credibility		33.3%		33.3%
Formula 4.4 Variance Allocations				
Allocated Process Variance		180		220
Allocated Parameter Variance		90		110
Total Variance Allocation		270		330
Initial MSE	200		200	
MSE of Credibility Estimate	133		133	
Reduction in MSE	67		67	
Reduction in MSE as % of Initial MSE	33.3%		33.3%	
Addtl reduction in MSE due to split	n/a		0	
Addtl reduction as % of Initial MSE	n/a		0.0%	
Addtl reduction as % of MSE under no-split plan	n/a		0.0%	

Exhibit 1 Sheet 5. Increasing Process Covariance Can Improve Effectiveness

Loss Component	Unsplit Credibility	Split Credibility		
	A	A ₁	Covariance	A ₂
Mean	100	50		50
Process Variance	400	90	75	160
Parameter Variance	200	100	25	50
Total Variance	600	190	100	210
MSE Optimal No-split Credibility	33.3%			
D			29,900	
MSE Optimal Split Credibility		62.7%		5.9%
Formula 4.4 Variance Allocations				
Allocated Process Variance		165		235
Allocated Parameter Variance		125		75
Total Variance Allocation		290		310
Initial MSE	200		200	
MSE of Credibility Estimate	133		117	
Reduction in MSE	67		83	
Reduction in MSE as % of Initial MSE	33.3%		41.4%	
Addtl reduction in MSE due to split	n/a		16	
Addtl reduction as % of Initial MSE	n/a		8.1%	
Addtl reduction as % of MSE under no-split plan	n/a		12.1%	

Exhibit 1 Sheet 6. Increasing Process Covariance Can Reduce Effectiveness

Loss Component	Unsplit Credibility	Split Credibility		
	A	A ₁	Covariance	A ₂
Mean	100	50		50
Process Variance	400	110	75	140
Parameter Variance	200	100	25	50
Total Variance	600	210	100	190
MSE Optimal No-split Credibility	33.3%			
D			29,900	
MSE Optimal Split Credibility		54.3%		10.9%
Formula 4.4 Variance Allocations				
Allocated Process Variance		185		215
Allocated Parameter Variance		125		75
Total Variance Allocation		310		290
Initial MSE	200		200	
MSE of Credibility Estimate	133		124	
Reduction in MSE	67		76	
Reduction in MSE as % of Initial MSE	33.3%		38.0%	
Addtl reduction in MSE due to split	n/a		9	
Addtl reduction as % of Initial MSE	n/a		4.7%	
Addtl reduction as % of MSE under no-split plan	n/a		7.1%	

Exhibit 1 Sheet 7. A Split with a Negative z for One Component Can Be Effective

Loss Component	Unsplit Credibility	Split Credibility		
	A	A ₁	Covariance	A ₂
Mean	100	50		50
Process Variance	400	110	75	140
Parameter Variance	200	125	25	25
Total Variance	600	235	100	165
MSE Optimal No-split Credibility	33.3%			
D			28,775	
MSE Optimal Split Credibility		68.6%		-11.3%
Formula 4.4 Variance Allocations				
Allocated Process Variance		185		215
Allocated Parameter Variance		150		50
Total Variance Allocation		335		265
<i>Initial MSE</i>	200		200	
<i>MSE of Credibility Estimate</i>	133		103	
<i>Reduction in MSE</i>	67		97	
<i>Reduction in MSE as % of Initial MSE</i>	33.3%		48.7%	
<i>Addtl reduction in MSE due to split</i>	n/a		31	
<i>Addtl reduction as % of Initial MSE</i>	n/a		15.3%	
<i>Addtl reduction as % of MSE under no-split plan</i>	n/a		23.0%	

Exhibit 2 Sheet 1. Non-Split and Primary-Excess Split Credibility Plans under CRM Ineffective Split

Inputs		Notation	Value			
Mean Claim Count		n	10.000			
Mean Severity		s	10.000			
Severity Mixing Parameter		b	0.250			
Claim Count Contagion		c	0.200			
Split Point		k	10.000			

Results		Total	Non-Split	Primary	Excess	Split
Claim Counts	Mean	10.000		10.000	10.000	
	Process Variance	10.000		10.000	10.000	
	Parameter Variance	20.000		20.000	20.000	
Severity	Mean	10.000		5.981	4.019	
	Process Variance	125.000		12.086	88.025	
	Parameter Variance	25.000		1.200	16.388	
	Process Covariance					12.445
	Parameter Covariance					3.706
Loss	Mean	100.000		59.812	40.188	
	Process Variance	2,500		491	1,206	
	Parameter Variance	5,000		860	2,290	
	Process Covariance					402
	Parameter Covariance					925
	Total Covariance					1,327
	Total Variance	7,500		1,350	3,495	
Credibility	Numerator		5,000	1,971,464	1,971,464	
	Denominator		7,500	2,957,196	2,957,196	
	Optimal z		66.7%	66.7%	66.7%	
Error	MSE	5,000	1,667			1,667
	MSE Reduction		3,333	Addl MSE Reduction		0
Variance Allocation	Process Variance	2,500		892	1,608	
	Process Var Alloc %	33.3%		35.7%	64.3%	
	Parameter Variance	5,000		1,785	3,215	
	Param Var Alloc %	66.7%		35.7%	64.3%	

Exhibit 2 Sheet 2. Non-Split and Primary-Excess Split Credibility Plans under CRM Effective Split

Inputs		Notation	Value			
Mean Claim Count		n	10.000			
Mean Severity		s	10.000			
Severity Mixing Parameter		b	0.025			
Claim Count Contagion		c	0.200			
Split Point		k	10.000			

Results		Total	Non-Split	Primary	Excess	Split
Claim Counts	Mean	10.000		10.000	10.000	
	Process Variance	10.000		10.000	10.000	
	Parameter Variance	20.000		20.000	20.000	
Severity	Mean	10.000		6.277	3.723	
	Process Variance	102.500		12.783	62.935	
	Parameter Variance	2.500		0.167	1.390	
	Process Covariance					13.391
	Parameter Covariance					0.471
Loss	Mean	100.000		62.768	37.232	
	Process Variance	2,050		523	782	
	Parameter Variance	2,300		808	444	
	Process Covariance					372
	Parameter Covariance					524
	Total Covariance					896
Credibility	Total Variance	4,350		1,331	1,226	
	Numerator		2,300	765,277	95,099	
	Denominator		4,350	828,993	828,993	
	Optimal z		52.9%	92.3%	11.5%	
Error	MSE	2,300	1,084			959
	MSE Reduction		1,216	Addl MSE Reduction		125
Variance Allocation	Process Variance	2,050		896	1,154	
	Process Var Alloc %	47.1%		43.7%	56.3%	
	Parameter Variance	2,300		1,332	968	
	Param Var Alloc %	52.9%		57.9%	42.1%	

Exhibit 2 Sheet 3. Non-Split and Primary-Excess Split Credibility Plans under CRM Credibility Inversion: $z_e > z_p$

Inputs		Notation	Value			
Mean Claim Count		n	10.000			
Mean Severity		s	10.000			
Severity Mixing Parameter		b	0.250			
Claim Count Contagion		c	0.020			
Split Point		k	10.000			

Results		Total	Non-Split	Primary	Excess	Split
Claim Counts	Mean	10.000		10.000	10.000	
	Process Variance	10.000		10.000	10.000	
	Parameter Variance	2.000		2.000	2.000	
Severity	Mean	10.000		5.981	4.019	
	Process Variance	125.000		12.086	88.025	
	Parameter Variance	25.000		1.200	16.388	
	Process Covariance					12.445
	Parameter Covariance					3.706
Loss	Mean	100.000		59.812	40.188	
	Process Variance	2,500		491	1,206	
	Parameter Variance	2,750		194	1,704	
	Process Covariance					402
	Parameter Covariance					426
	Total Covariance					828
	Total Variance	5,250		685	2,910	
Credibility	Numerator		2,750	40,533	944,757	
	Denominator		5,250	1,306,291	1,306,291	
	Optimal z		52.4%	3.1%	72.3%	
Error	MSE	2,750	1,310			1,190
	MSE Reduction		1,440	Addl MSE Reduction		119
Variance Allocation	Process Variance	2,500		892	1,608	
	Process Var Alloc %	47.6%		35.7%	64.3%	
	Parameter Variance	2,750		620	2,130	
	Param Var Alloc %	52.4%		22.5%	77.5%	

Exhibit 2 Sheet 4. Non-Split and Primary-Excess Split Credibility Plans under CRM Negative Primary Credibility

Inputs		Notation	Value			
Mean Claim Count		n	4.000			
Mean Severity		s	25.000			
Severity Mixing Parameter		b	0.250			
Claim Count Contagion		c	0.100			
Split Point		k	5.000			

Results		Total	Non-Split	Primary	Excess	Split
Claim Counts	Mean	4.000		4.000	4.000	
	Process Variance	4.000		4.000	4.000	
	Parameter Variance	1.600		1.600	1.600	
Severity	Mean	25.000		4.452	20.548	
	Process Variance	781.250		1.526	761.389	
	Parameter Variance	156.250		0.042	152.014	
	Process Covariance					9.167
	Parameter Covariance					2.097
Loss	Mean	100.000		17.807	82.193	
	Process Variance	6,250		86	5,343	
	Parameter Variance	3,750		32	3,351	
	Process Covariance					411
	Parameter Covariance					183
Credibility	Total Covariance					594
	Total Variance	10,000		118	8,694	
	Numerator		3,750	-224,869	288,835	
	Denominator		10,000	672,661	672,661	
	Optimal z		37.5%	-33.4%	42.9%	
Error	MSE	3,750	2,344			2,305
	MSE Reduction		1,406	Addl MSE Reduction		39
Variance Allocation	Process Variance	6,250		497	5,753	
	Process Var Alloc %	62.5%		7.9%	92.1%	
	Parameter Variance	3,750		216	3,534	
	Param Var Alloc %	37.5%		5.8%	94.2%	

Appendix A: Coefficients of Variation for Excess and Total Loss

We will show the process coefficient of variation (CV) of excess loss is at least as large as the process CV for total loss.

Let N denote the claim count and let X denote the claim severity. Let A stand for total loss so that $A = X(1) + X(2) + \dots + X(N)$, where the $X(i)$ are independent trials of X . Assuming N is Poisson, the mean and variance of A are given as:

$$\mu_A = \mu_N \cdot \mu_X. \tag{A.1}$$

$$\sigma_A^2 = \mu_N \cdot E[X^2]. \tag{A.2}$$

Using CV_A to denote the process Coefficient of Variation of A , it follows that:

$$CV_A^2 = \frac{E[X^2]}{\mu_N \cdot \mu_X^2}. \tag{A.3}$$

Given an attachment, k , define the excess severity, X_e , via $X_e = X - \min(X, k)$, and excess loss, $A_e = X_e(1) + X_e(2) + \dots + X_e(N)$. The square of the process CV of excess loss is:

$$CV_{A_e}^2 = \frac{E[X_e^2]}{\mu_N \cdot \mu_{X_e}^2}. \tag{A.4}$$

We will now mathematically state and prove the result that excess layer CV is greater than CV for total loss.

Proposition 1: The CVs of excess and total layer loss satisfy the inequality:

$$CV_A \leq CV_{A_e}. \tag{A.5}$$

Proof: We will take the derivative of the square of the process CV with respect to the attachment. To do this, we need formulas for the derivatives of the square of the expected excess severity and the expected square of excess severity:

$$\begin{aligned} \frac{\partial(E[X_e])^2}{\partial k} &= 2\mu_{X_e} \frac{\partial}{\partial k} \int_k^\infty (x-k) dF_X(x) \\ &= -2\mu_{X_e} \cdot G_X(k). \end{aligned} \tag{A.6}$$

$$\begin{aligned} \frac{\partial(E[X_e^2])}{\partial k} &= \frac{\partial}{\partial k} \int_k^\infty (x-k)^2 dF_X(x) \\ &= -2\mu_{X_e}. \end{aligned} \tag{A.7}$$

The derivative of the square of the CV of excess loss is therefore:

$$\frac{\partial(CV_{A_e})^2}{\partial k} = \frac{-2(\mu_{X_e})^3 + 2E[X_e^2]\mu_{X_e}G_X(k)}{\mu_N(\mu_{X_e})^4}. \tag{A.8}$$

This derivative will be non-negative if:

$$(\mu_{X_e})^2 \leq E[X_e^2] \cdot G_X(k). \tag{A.9}$$

Now consider the extreme case in which the severity distribution above the attachment, k , consists of a single mass point of probability $G_X(k)$ at a point, $y + k$. It follows in that case that:

$$\mu_{X_e} = y \cdot G_X(k). \tag{A.10}$$

$$E[X_e^2] = y^2 \cdot G_x(k). \quad (\text{A.11})$$

So in the extreme case we have:

$$(\mu_{x_e})^2 = E[X_e^2] \cdot G_x(k). \quad (\text{A.12})$$

This implies the derivative of the CV of excess loss is zero for the extreme case in which the tail consists of a point mass. In any other case, the

inequality will hold strictly. We can thus conclude the derivative of the square of the process CV of excess loss is non-negative, where the derivative is taken with respect to the attachment point for excess loss. Since this is true for any attachment point, it follows that the process CV at any attachment is at least as large as the process CV of total loss. This is because the CV of total loss corresponds to an attachment of zero.

Appendix B: Optimal Mean Square Parameter Error Formula Proofs

Proposition 2: The minimal mean square parameter error for a split plan is given by:

$$\epsilon_0^2(SP) = \tau^2 - \frac{1}{D} \left(\lambda_2^2 (\tau_1^2 + \pi)^2 + \lambda_1^2 (\tau_2^2 + \pi)^2 \right). \quad (\text{B.1})$$

Proof: The general mean square error formula for arbitrary credibility values is given in Equation 3.4 as:

$$\begin{aligned} \epsilon^2 = & \tau^2 + z_1^2 \lambda_1^2 - 2z_1 (\tau_1^2 + \pi) + z_2^2 \lambda_2^2 \\ & - 2z_2 (\tau_2^2 + \pi) + 2z_1 z_2 \kappa. \end{aligned} \quad (\text{B.2})$$

The optimal credibility values as shown in Equation 3.8 are:

$$\begin{aligned} z_1 = & \frac{\lambda_2^2 (\tau_1^2 + \pi) - \kappa (\tau_2^2 + \pi)}{D} \\ z_2 = & \frac{\lambda_1^2 (\tau_2^2 + \pi) - \kappa (\tau_1^2 + \pi)}{D} \end{aligned} \quad (\text{B.3})$$

where $D = \lambda_1^2 \lambda_2^2 - \kappa^2$.

Expand the second term on the right hand side of B.2 as follows:

$$\begin{aligned} z_1^2 \cdot \lambda_1^2 = & \left(\frac{\lambda_2^2 (\tau_1^2 + \pi) - \kappa (\tau_2^2 + \pi)}{D} \right)^2 \cdot \lambda_1^2 \\ = & \frac{1}{D^2} \left(\begin{aligned} & \lambda_1^2 \lambda_2^4 (\tau_1^2 + \pi)^2 \\ & - 2\kappa \lambda_1^2 \lambda_2^2 (\tau_1^2 + \pi) (\tau_2^2 + \pi) \\ & + \kappa^2 \lambda_1^2 (\tau_2^2 + \pi)^2 \end{aligned} \right). \end{aligned} \quad (\text{B.4})$$

Also expand the “ z_1 ” term:

$$\begin{aligned} 2z_1 \cdot (\tau_1^2 + \pi) = & 2 \left(\frac{\lambda_2^2 (\tau_1^2 + \pi)^2 - \kappa (\tau_1^2 + \pi) (\tau_2^2 + \pi)}{D} \right) \\ = & \frac{2}{D^2} (\lambda_2^2 (\tau_1^2 + \pi)^2 - \kappa (\tau_1^2 + \pi) (\tau_2^2 + \pi)) \\ & \cdot (\lambda_1^2 \lambda_2^2 - \kappa^2) \\ = & \frac{1}{D^2} \left(\begin{aligned} & 2\lambda_1^2 \lambda_2^4 (\tau_1^2 + \pi)^2 - 2\kappa^2 \lambda_2^2 (\tau_1^2 + \pi)^2 \\ & - 2\kappa \lambda_1^2 \lambda_2^2 (\tau_1^2 + \pi) (\tau_2^2 + \pi) \\ & + 2\kappa^3 (\tau_1^2 + \pi) (\tau_2^2 + \pi) \end{aligned} \right). \end{aligned} \quad (\text{B.5})$$

Thus it follows:

$$\begin{aligned} z_1^2 \cdot \lambda_1^2 - 2z_1 \cdot (\tau_1^2 + \pi) = & \\ = & \frac{1}{D^2} \left(\begin{aligned} & \lambda_1^2 \lambda_2^4 (\tau_1^2 + \pi)^2 - 2\kappa \lambda_1^2 \lambda_2^2 (\tau_1^2 + \pi) (\tau_2^2 + \pi) \\ & + \kappa^2 \lambda_1^2 (\tau_2^2 + \pi)^2 \end{aligned} \right) \\ & - \frac{1}{D^2} \left(\begin{aligned} & 2\lambda_1^2 \lambda_2^4 (\tau_1^2 + \pi)^2 - 2\kappa^2 \lambda_2^2 (\tau_1^2 + \pi)^2 \\ & - 2\kappa \lambda_1^2 \lambda_2^2 (\tau_1^2 + \pi) (\tau_2^2 + \pi) \\ & + 2\kappa^3 (\tau_1^2 + \pi) (\tau_2^2 + \pi) \end{aligned} \right) \\ = & \frac{1}{D^2} \left(\begin{aligned} & -\lambda_1^2 \lambda_2^4 (\tau_1^2 + \pi)^2 + 2\kappa^2 \lambda_2^2 (\tau_1^2 + \pi)^2 \\ & + \kappa^2 \lambda_1^2 (\tau_2^2 + \pi)^2 - 2\kappa^3 (\tau_1^2 + \pi) (\tau_2^2 + \pi) \end{aligned} \right). \end{aligned} \quad (\text{B.6})$$

Similarly, one can derive:

$$\begin{aligned} z_2^2 \cdot \lambda_2^2 - 2z_2 \cdot (\tau_2^2 + \pi) = & \\ = & \frac{1}{D^2} \left(\begin{aligned} & -\lambda_1^4 \lambda_2^2 (\tau_2^2 + \pi)^2 + 2\kappa^2 \lambda_1^2 (\tau_2^2 + \pi)^2 \\ & + \kappa^2 \lambda_2^2 (\tau_1^2 + \pi)^2 - 2\kappa^3 (\tau_1^2 + \pi) (\tau_2^2 + \pi) \end{aligned} \right). \end{aligned} \quad (\text{B.7})$$

Adding B.6 and B.7 together and grouping like terms leads to:

$$z_1^2 \cdot \lambda_1^2 - 2z_1 \cdot (\tau_1^2 + \pi) + z_2^2 \cdot \lambda_2^2 - 2z_2 \cdot (\tau_2^2 + \pi) = \frac{1}{D^2} \begin{pmatrix} -\lambda_1^2 \lambda_2^4 (\tau_1^2 + \pi)^2 - \lambda_1^4 \lambda_2^2 (\tau_2^2 + \pi)^2 \\ + 3\kappa^2 \lambda_2^2 (\tau_1^2 + \pi)^2 + 3\kappa^2 \lambda_1^2 (\tau_2^2 + \pi)^2 \\ - 4\kappa^3 (\tau_1^2 + \pi)(\tau_2^2 + \pi) \end{pmatrix}. \quad (B.8)$$

Next expand the cross-term in B.2:

$$2\kappa z_1 z_2 = \frac{2\kappa}{D^2} (\lambda_2^2 (\tau_1^2 + \pi) - \kappa (\tau_2^2 + \pi)) \cdot (\lambda_1^2 (\tau_2^2 + \pi) - \kappa (\tau_1^2 + \pi)) = \frac{2}{D^2} \begin{pmatrix} \kappa \lambda_1^2 \lambda_2^2 (\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ + \kappa^3 (\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ - \kappa^2 \lambda_1^2 (\tau_2^2 + \pi)^2 - \kappa^2 \lambda_2^2 (\tau_1^2 + \pi)^2 \end{pmatrix}. \quad (B.9)$$

Plugging B.8 and B.9 into B.2 yields the following formula for the mean square error:

$$\varepsilon^2 = \tau^2 + \frac{1}{D^2} \begin{pmatrix} -\lambda_1^2 \lambda_2^4 (\tau_1^2 + \pi)^2 - \lambda_1^4 \lambda_2^2 (\tau_2^2 + \pi)^2 \\ + 3\kappa^2 \lambda_2^2 (\tau_1^2 + \pi)^2 + 3\kappa^2 \lambda_1^2 (\tau_2^2 + \pi)^2 \\ - 4\kappa^3 (\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ + 2\kappa \lambda_1^2 \lambda_2^2 (\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ + 2\kappa^3 (\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ - 2\kappa^2 \lambda_1^2 (\tau_2^2 + \pi)^2 - 2\kappa^2 \lambda_2^2 (\tau_1^2 + \pi)^2 \end{pmatrix}. \quad (B.10)$$

This simplifies to:

$$\varepsilon^2 = \tau^2 + \frac{1}{D^2} \begin{pmatrix} (\tau_1^2 + \pi)^2 (-\lambda_1^2 \lambda_2^4 + \kappa^2 \lambda_2^2) \\ + (\tau_2^2 + \pi)^2 (-\lambda_1^4 \lambda_2^2 + \kappa^2 \lambda_1^2) \\ + (\tau_1^2 + \pi)(\tau_2^2 + \pi)(-2\kappa^3 + 2\kappa \lambda_1^2 \lambda_2^2) \end{pmatrix}. \quad (B.11)$$

and this further reduces to:

$$\varepsilon^2 = \tau^2 + \frac{1}{D^2} \begin{pmatrix} \lambda_2^2 (\tau_1^2 + \pi)^2 (-\lambda_1^2 \lambda_2^2 + \kappa^2) \\ + \lambda_1^2 (\tau_2^2 + \pi)^2 (-\lambda_1^2 \lambda_2^2 + \kappa^2) \\ + 2\kappa (\tau_1^2 + \pi)(\tau_2^2 + \pi)(-\kappa^2 + \lambda_1^2 \lambda_2^2) \end{pmatrix}. \quad (B.12)$$

Using $D = D = \lambda_1^2 \lambda_2^2 - \kappa^2$ and performing a few basic algebra operations leads to:

$$\varepsilon^2 = \tau^2 - \frac{1}{D} \begin{pmatrix} \lambda_2^2 (\tau_1^2 + \pi)^2 + \lambda_1^2 (\tau_2^2 + \pi)^2 \\ - 2\kappa (\tau_1^2 + \pi)(\tau_2^2 + \pi) \end{pmatrix}. \quad (B.13)$$

This is what was to be proved.

Theorem 1: The minimal mean square parameter error for a split plan is given by:

$$\varepsilon_0^2(SP) = (\tau_1^2 + \pi)(1 - z_1^*) + (\tau_2^2 + \pi)(1 - z_2^*). \quad (B.14)$$

Proof: First expand τ^2 in Equation B.1 to obtain:

$$\varepsilon_0^2(SP) = \tau_1^2 + \tau_2^2 + 2\pi - \frac{1}{D} \begin{pmatrix} \lambda_2^2 (\tau_1^2 + \pi)^2 + \lambda_1^2 (\tau_2^2 + \pi)^2 \\ - 2\kappa (\tau_1^2 + \pi)(\tau_2^2 + \pi) \end{pmatrix}. \quad (B.15)$$

Next regroup terms:

$$\varepsilon_0^2(SP) = \tau_1^2 + \tau_2^2 + 2\pi - \frac{1}{D} \begin{pmatrix} \lambda_2^2 (\tau_1^2 + \pi)^2 \\ - \kappa (\tau_1^2 + \pi)(\tau_2^2 + \pi) \end{pmatrix} - \frac{1}{D} \begin{pmatrix} \lambda_1^2 (\tau_2^2 + \pi)^2 \\ - \kappa (\tau_1^2 + \pi)(\tau_2^2 + \pi) \end{pmatrix}. \quad (B.16)$$

Factor and then substitute the optimal credibility values to derive:

$$\varepsilon_0^2(SP) = \tau_1^2 + \tau_2^2 + 2\pi - \frac{(\tau_1^2 + \pi)}{D} (\lambda_2^2 (\tau_1^2 + \pi) - \kappa (\tau_2^2 + \pi)) - \frac{(\tau_2^2 + \pi)}{D} (\lambda_1^2 (\tau_2^2 + \pi) - \kappa (\tau_1^2 + \pi)) = \tau_1^2 + \tau_2^2 + 2\pi - (\tau_1^2 + \pi) z_1^* - (\tau_2^2 + \pi) z_2^*. \quad (B.17)$$

This leads directly to the result.

Corollary 1: The reduction in minimal mean square parameter error in going from a no-split to a split plan, $\Delta(\epsilon_0^2) = \epsilon_0^2(\text{NS}) - \epsilon_0^2(\text{SP})$, is given as:

$$\Delta\epsilon_0^2 = \tau^2(1 - z^*) - (\tau_1^2 + \pi)(1 - z_1^*) - (\tau_2^2 + \pi)(1 - z_2^*) \\ = (\tau_1^2 + \pi)(z_1^* - z^*) + (\tau_2^2 + \pi)(z_2^* - z^*). \quad (\text{B.18})$$

Proof: The first equality follows directly from 2.8 and B.14.

The second equality is derived from the first as follows.

$$\tau^2(1 - z^*) - (\tau_1^2 + \pi)(1 - z_1^*) - (\tau_2^2 + \pi)(1 - z_2^*) \\ = (\tau_1^2 + \tau_2^2 + 2\pi)(1 - z^*) - (\tau_1^2 + \pi)(1 - z_1^*) \\ - (\tau_2^2 + \pi)(1 - z_2^*) \\ = (\tau_1^2 + \pi)(1 - z^* - (1 - z_1^*)) \\ + (\tau_2^2 + \pi)(1 - z^* - (1 - z_2^*)) \\ = (\tau_1^2 + \pi)(z_1^* - z^*) + (\tau_2^2 + \pi)(z_2^* - z^*). \quad (\text{B.19})$$

Next it is shown the reduction in mean square error is related to the square of the difference in the optimal credibilities.

Theorem 2: The reduction in minimal mean square parameter error in going from a no-split to a split plan, $\Delta(\epsilon_0^2) = \epsilon_0^2(\text{NS}) - \epsilon_0^2(\text{SP})$, is given as:

$$\Delta(\epsilon_0^2) = \frac{D}{\lambda^2}(z_1^* - z_2^*)^2. \quad (\text{B.20})$$

Proof: Consider

$$(z_1^*)^2 = \left(\frac{\lambda_2^2(\tau_1^2 + \pi) - \kappa(\tau_2^2 + \pi)}{D} \right)^2 \\ = \frac{1}{D^2} \left(\begin{array}{l} \lambda_2^4(\tau_1^2 + \pi)^2 \\ - 2\kappa\lambda_2^2(\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ + \kappa^2(\tau_2^2 + \pi)^2 \end{array} \right). \quad (\text{B.21})$$

Similarly, it can be shown that

$$(z_2^*)^2 = \frac{1}{D^2} \left(\begin{array}{l} \lambda_1^4(\tau_2^2 + \pi)^2 \\ - 2\kappa\lambda_1^2(\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ + \kappa^2(\tau_1^2 + \pi)^2 \end{array} \right). \quad (\text{B.22})$$

and that

$$z_1^*z_2^* = \frac{1}{D^2} \left(\begin{array}{l} \lambda_1^2\lambda_2^2(\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ - \kappa\lambda_2^2(\tau_1^2 + \pi)^2 \\ - \kappa\lambda_1^2(\tau_1^2 + \pi)^2 \\ + \kappa^2(\tau_1^2 + \pi)(\tau_2^2 + \pi) \end{array} \right). \quad (\text{B.23})$$

Thus the square difference of the optimal credibilities is

$$(z_1^* - z_2^*)^2 = \frac{1}{D^2} \left(\begin{array}{l} \lambda_2^4(\tau_1^2 + \pi)^2 - 2\kappa\lambda_2^2(\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ + \kappa^2(\tau_2^2 + \pi)^2 + \lambda_1^4(\tau_2^2 + \pi)^2 \\ - 2\kappa\lambda_1^2(\tau_1^2 + \pi)(\tau_2^2 + \pi) + \kappa^2(\tau_1^2 + \pi)^2 \\ + - 2\lambda_1^2\lambda_2^2(\tau_1^2 + \pi)(\tau_2^2 + \pi) \\ + 2\kappa\lambda_2^2(\tau_1^2 + \pi)^2 + 2\kappa\lambda_1^2(\tau_1^2 + \pi)^2 \\ - 2\kappa^2(\tau_1^2 + \pi)(\tau_2^2 + \pi) \end{array} \right). \quad (\text{B.24})$$

This simplifies to

$$(z_1^* - z_2^*)^2 = \frac{1}{D^2} \left(\begin{array}{l} (\tau_1^2 + \pi)^2(\lambda_2^4 + 2\kappa\lambda_2^2 + \kappa^2) \\ + (\tau_2^2 + \pi)^2(\lambda_1^4 + 2\kappa\lambda_1^2 + \kappa^2) \\ - 2(\tau_1^2 + \pi)(\tau_2^2 + \pi)(\lambda_1^2\lambda_2^2 + \kappa\lambda_2^2 + \kappa\lambda_1^2 + 2\kappa^2) \end{array} \right) \\ = \frac{1}{D^2} \left(\begin{array}{l} (\tau_1^2 + \pi)^2(\lambda_2^2 + \kappa)^2 + (\tau_2^2 + \pi)^2(\lambda_1^2 + \kappa)^2 \\ - 2(\tau_1^2 + \pi)(\tau_2^2 + \pi)(\lambda_1^2\lambda_2^2 + \kappa\lambda_2^2 + \kappa\lambda_1^2 + 2\kappa^2) \end{array} \right). \quad (\text{B.25})$$

Now examine the expression for minimal mean square error in B.18:

$$\Delta\epsilon_0^2 = (\tau_1^2 + \pi)(z_1^* - z^*) + (\tau_2^2 + \pi)(z_2^* - z^*). \quad (\text{B.26})$$

Consider the first term can be expanded as follows:

$$(\tau_1^2 + \pi)(z_1^* - z^*) \\ = (\tau_1^2 + \pi) \left(\frac{\lambda_2^2(\tau_1^2 + \pi) - \kappa(\tau_2^2 + \pi)}{D} - \frac{\tau^2}{\lambda^2} \right)$$

$$\begin{aligned}
 &= \frac{1}{\lambda^2 D} (\tau_1^2 + \pi) \begin{pmatrix} \lambda^2 \lambda_2^2 (\tau_1^2 + \pi) \\ -\lambda^2 k (\tau_2^2 + \pi) \\ -D (\tau_1^2 + \tau_2^2 + 2\pi) \end{pmatrix} \\
 &= \frac{1}{\lambda^2 D} (\tau_1^2 + \pi) \begin{pmatrix} (\tau_1^2 + \pi) (\lambda^2 \lambda_2^2 - D) \\ -(\tau_2^2 + \pi) (\kappa \lambda^2 + D) \end{pmatrix}. \quad (\text{B.27})
 \end{aligned}$$

Next derive the formulas:

$$\begin{aligned}
 \lambda^2 \lambda_2^2 - D &= (\lambda_1^2 + \lambda_2^2 + 2\kappa) \lambda_2^2 - \lambda_1^2 \lambda_2^2 + \kappa^2 \\
 &= \lambda_2^4 + 2\kappa \lambda_2^2 + \kappa^2 = (\lambda_2^2 + \kappa)^2. \quad (\text{B.28})
 \end{aligned}$$

And

$$\begin{aligned}
 \kappa \lambda^2 + D &= \kappa (\lambda_1^2 + \lambda_2^2 + 2\kappa) + \lambda_1^2 \lambda_2^2 - \kappa^2 \\
 &= \kappa \lambda_1^2 + \kappa \lambda_2^2 + \kappa^2 + \lambda_1^2 \lambda_2^2. \quad (\text{B.29})
 \end{aligned}$$

Plugging B.28 and B.29 into B.27, one obtains:

$$\begin{aligned}
 &(\tau_1^2 + \pi)(z_1^* - z^*) \\
 &= \frac{(\tau_1^2 + \pi)^2 (\lambda_2^2 + \kappa)^2 - (\tau_1^2 + \pi) (\tau_2^2 + \pi)}{\lambda^2 D} \cdot \quad (\text{B.30}) \\
 &= \frac{(\kappa \lambda_1^2 + \kappa \lambda_2^2 + \kappa^2 + \lambda_1^2 \lambda_2^2)}{\lambda^2 D}.
 \end{aligned}$$

Using this and a similar formula for the second term in B.26, the difference in minimal square error may be written as:

$$\begin{aligned}
 \Delta \epsilon_0^2 &= \\
 &= \frac{1}{\lambda^2 D} \begin{pmatrix} (\tau_1^2 + \pi)^2 (\lambda_2^2 + \kappa)^2 \\ + (\tau_2^2 + \pi)^2 (\lambda_1^2 + \kappa)^2 \\ - 2(\tau_1^2 + \pi) (\tau_2^2 + \pi) \\ (\kappa \lambda_1^2 + \kappa \lambda_2^2 + \kappa^2 + \lambda_1^2 \lambda_2^2) \end{pmatrix}. \quad (\text{B.31})
 \end{aligned}$$

Comparing B.30 to B.25 leads to the desired conclusion.

Corollary 2: The reduction in minimal mean square parameter error in going from a no-split to a split plan is given as:

$$\Delta(\epsilon_0^2) = \frac{1}{D \lambda^2} \left(\frac{(\tau_1^2 + \pi)(\sigma_2^2 + \rho)}{-(\sigma_1^2 + \rho)(\tau_2^2 + \pi)} \right)^2. \quad (\text{B.32})$$

Proof: The overall plan is to start with Theorem 2 and perform substitutions and algebraic operations to arrive at the result. Write

$$\begin{aligned}
 z_1^* - z_2^* &= \frac{\lambda_2^2 (\tau_1^2 + \pi) - \kappa (\tau_2^2 + \pi)}{D} \\
 &\quad - \frac{\lambda_1^2 (\tau_2^2 + \pi) - \kappa (\tau_1^2 + \pi)}{D} \\
 &= \frac{1}{D} \left(\frac{\lambda_2^2 (\tau_1^2 + \pi) - \kappa (\tau_2^2 + \pi)}{-\lambda_1^2 (\tau_2^2 + \pi) + \kappa (\tau_1^2 + \pi)} \right). \quad (\text{B.33})
 \end{aligned}$$

Then expand total variance and covariance terms into process and parameter components and then simplify as follows to get the result:

$$\begin{aligned}
 z_1^* - z_2^* &= \\
 &= \frac{1}{D} \left(\frac{(\sigma_2^2 + \tau_2^2)(\tau_1^2 + \pi) - (\rho + \pi)(\tau_2^2 + \pi)}{-(\sigma_1^2 + \tau_1^2)(\tau_2^2 + \pi) + (\rho + \pi)(\tau_1^2 + \pi)} \right) \\
 &= \frac{1}{D} \left(\frac{(\sigma_2^2 + \tau_2^2 + \rho + \pi)(\tau_1^2 + \pi)}{-(\sigma_1^2 + \tau_1^2 + \rho + \pi)(\tau_2^2 + \pi)} \right) \\
 &= \frac{1}{D} ((\sigma_2^2 + \rho)(\tau_1^2 + \pi) - (\sigma_1^2 + \rho)(\tau_2^2 + \pi)). \quad (\text{B.34})
 \end{aligned}$$

Plug this into Theorem 2 and the result follows.

Appendix C: Parameter and Process Covariance Between Primary and Excess Layer Losses

We will show the process covariance is given as:

Proposition 3:

$$\rho = E[Cov(A_p, A_e)] = (\sigma_N^2 - \mu_N) \cdot \mu_{x_p} \cdot \mu_{x_e} + k \cdot \mu_{x_e} \cdot \mu_N. \quad (C.1)$$

Proof: Consider

$$\begin{aligned} A_p \cdot A_e &= \left(\sum_{i=1}^N X(i)_p \right) \left(\sum_{i=1}^N X(i)_e \right) \\ &= \left(\sum_{i=1}^N X(i)_p X(i)_e \right) + \left(\sum_{i \neq j} X(i)_p X(j)_e \right). \end{aligned} \quad (C.2)$$

Thus

$$\begin{aligned} E[A_p \cdot A_e | \theta] &= E[N | \theta] \cdot E[X_p X_e | \theta] + E[(N^2 - N) | \theta] \\ &\quad \cdot E[X_p | \theta] E[X_e | \theta]. \end{aligned} \quad (C.3)$$

When a claim leads to an excess layer loss that is strictly positive, it follows that the primary loss must consume the whole primary limit. Therefore

$$E[X_p X_e | \theta] = k E[X_e | \theta]. \quad (C.4)$$

Taking expectations to arrive at unconditional values, C.2 implies:

$$E[A_p \cdot A_e] = (E[N^2] - \mu_N) \cdot \mu_{x_p} \cdot \mu_{x_e} + k \cdot \mu_{x_e} \cdot \mu_N. \quad (C.5)$$

So we find

$$\begin{aligned} E[Cov(A_p, A_e)] &= E[A_p \cdot A_e] - E[A_p] \cdot E[A_e] \\ &= (E[N^2] - \mu_N) \cdot \mu_{x_p} \cdot \mu_{x_e} \\ &\quad + \mu_N \cdot k \cdot \mu_{x_e} - \mu_N \mu_{x_p} \cdot \mu_N \mu_{x_e} \end{aligned} \quad (C.6)$$

The result follows immediately.

Now we will show parameter covariance is given as:

Proposition 4:

$$\begin{aligned} \pi &= Cov(E[A_p | \theta], E[A_e | \theta]) \\ &= (\tau_N^2 + \mu_N^2) \cdot \pi_x + \tau_N^2 \cdot \mu_{x_p} \cdot \mu_{x_e}. \end{aligned} \quad (C.7)$$

Proof: Using our notation, we can write

$$\begin{aligned} E[A_p | \theta] \cdot E[A_e | \theta] &= \mu_{A_p}(\theta) \mu_{A_e}(\theta) \\ &= (\mu_N(\theta))^2 \mu_{x_p}(\theta) \mu_{x_e}(\theta). \end{aligned} \quad (C.8)$$

Therefore

$$\begin{aligned} Cov(\mu_{A_p}(\theta), \mu_{A_e}(\theta)) &= E[(\mu_N(\theta))^2] E[\mu_{x_p}(\theta) \mu_{x_e}(\theta)] \\ &\quad - E[\mu_N(\theta)]^2 E[\mu_{x_p}(\theta)] E[\mu_{x_e}(\theta)] \\ &= (\tau_N^2 + \mu_N^2) E[\mu_{x_p}(\theta) \mu_{x_e}(\theta)] - \mu_N^2 \mu_{x_p} \mu_{x_e}. \end{aligned} \quad (C.9)$$

We add and subtract a term, regroup and simplify to obtain the desired result:

$$\begin{aligned} Cov(\mu_{A_p}(\theta), \mu_{A_e}(\theta)) &= (\tau_N^2 + \mu_N^2) E[\mu_{x_p}(\theta) \mu_{x_e}(\theta)] - (\tau_N^2 + \mu_N^2) \mu_{x_p} \mu_{x_e} \\ &\quad + (\tau_N^2 + \mu_N^2) \mu_{x_p} \mu_{x_e} - \mu_N^2 \mu_{x_p} \mu_{x_e} \\ &= (\tau_N^2 + \mu_N^2) \pi_x + \tau_N^2 \mu_{x_p} \mu_{x_e}. \end{aligned} \quad (C.10)$$

This is the result claimed in C.7.