THE COMPETITIVE MARKET EQUILIBRIUM RISK LOAD FORMULA FOR INCREASED LIMITS RATEMAKING

GLENN G. MEYERS

DISCUSSION BY IRA ROBBIN

1. INTRODUCTION

Glenn Meyers has made a valuable contribution to actuarial literature with his well-written paper on how to load increased limits factors (ILFs) for risk. Given the complexity of the topic, he deserves special commendation for his coherent presentation. Meyers clearly states his fundamental assumptions and provides sufficient background for the reader to understand his results in context. His skill in composing mathematical derivations is also noteworthy. As to substance, Meyers uses the intuitively appealing market paradigm as the foundation for his risk load theory. This represents a conceptual step forward. Unfortunately, Meyers does not carry the theory far enough, and, in my opinion, ends up with an incorrect answer.

2. WHAT IS RISK LOAD?

Before explaining why his answer is wrong, it is necessary to set the stage by first defining risk load. What is risk load? I would define it as an extra component of indicated premium arising from the potential for possible deviations between expected and actual loss results. The indicated premium for a coverage is the sum of the expected loss cost, expense provisions, and usual profit load, plus the risk load.

Why is risk load needed? The most general answer is that the price for insurance coverage ought to somehow depend on the volatility of the actual losses covered by the insurance. In the pricing of assets, such as bonds, it is well accepted that the interest rate demanded by the market rises with the riskiness of the asset. With respect to increased limits, risk load is important because increasing the limit of coverage changes the relative volatility of actual results versus initial expectations.

3. OVERALL LEVEL VERSUS RELATIVITY BY LIMIT OR LINE

When considering risk loads for increased limits, it is useful to split the question into two parts by asking first, "What is the proper overall level of risk load?" and, second, "How should the risk load vary by limit and by line of coverage?" While I have some question about whether Meyers developed a logically consistent theory for setting an appropriate overall level of risk load, my major criticism is that his model produces risk loads that can rise too steeply by limit.

4. PROCESS AND PARAMETER RISK

Before detailing this criticism, I should note my agreement with much of what Meyers has done. In particular, I concur with Meyers that risk loads should reflect not only the stochastic variability of actual results versus expectation (process risk), but also the uncertainty about the loss expectation itself (parameter risk). This is generally accepted in principle by most actuaries knowledgeable in the subject. I also accept the collective risk model Meyers employed to incorporate parameter risk (see Section 4 of his paper). While I might want to quibble with the specific techniques Meyers used for quantifying parameter risk, I will not do so in this discussion.

5. MARKET EQUILIBRIUM THEORY AS A FOUNDATION FOR RISK LOAD

I agree with Meyers that market-based theory can provide a sensible foundation for risk load calculations. The basic idea in using a market-based theory is to apply the "supply-versus-demand" concept of general economics to the pricing of risk. A key advantage of this approach is that it explains the need for risk loads in ILFs using fundamental economic principles. I also believe it provides a basis for conceptually defining what risk loads should be included in ILFs filed by a rating bureau. Under my view, risk loads filed by a rating bureau should be those that would be theoretically charged by a rational market in equilibrium. Actual risk loads charged in the real market could, of course, be different since the market may be irrational or in disequilibrium. However, by relying on a theoretical market, one should obtain bureau risk loads that are sound benchmarks unaffected by market cycles or imperfections. Individual companies can then deviate up or down as they deem appropriate.

The theoretical market approach is not universally accepted as the logical foundation for a theory of risk load. Some, for instance, have insisted on determining what the risk load should really be and not what some hypothetical market says it should be. While this "just give me the real risk load" attitude sounds direct and practical, it leads nowhere. The problem is that there is no inherent notion of how to properly price for risk, either before or after the fact. In contrast, the right price for losses is an amount sufficient, on average, to cover the actual losses. After the fact, we know the actual loss costs and what we should have charged for losses. However, we have no actual "risk" costs to tell us what the charge for risk should have been. If there is a difference between expected and actual losses, we have evidence of volatility and thus proof that some risk load should have been charged. Nonetheless, this evidence alone does not tell us how to measure volatility, nor how to translate volatility, however measured, into a charge for risk.

6. UTILITY THEORY

Utility theory has been used as a conceptual framework for pricing risk. Under utility theory, the minimum premium an insurer is willing to accept is the lowest one for which the insurer's expected utility will not suffer if it provides the coverage. This means that the utility of initial wealth is the same as the expectation of utility of final wealth, where final wealth equals initial wealth plus premium less expenses and actual losses. A key point is that the resulting risk load is, to first approximation, proportional to the variance of losses.

I agree with Meyers that a "single insurer, single insured" implementation of utility theory is too simplistic. It produces risk loads an insurer might want to charge if there were no other insurers competing for the business and if there were no reinsurance market. With only one insurer in the model, there is no market and therefore no room for the forces of supply and demand to operate.

7. THE MEYERS RISK LOAD THEORY

In the Meyers model, there are many insurers and many insureds. So far, so good. Each insurer sets a constraint on the variance of losses it will tolerate on its individual book of business, and seeks to achieve the maximum profit subject to that constraint. The insureds are assumed to have an inelastic demand for insurance coverage by limit. For example, half the market may demand coverage at a $1 million limit and will pay the going rate to obtain it. The insurers then "bid" on the risks. Under the given assumption, and a further hypothesis about the total needed risk load over all lines and limits, the market in each subline and limit will be cleared at an optimal price. With some elegant mathematics, Meyers shows that the market clearing price can be viewed as the price that would have been charged by an average insurer acting alone. This result is shown in his Equation 5.6. Thus, Meyers effectively ends up with a variance-based risk load, and his theory could be regarded as another argument for variance. The central question, then, is whether his argument undercuts the serious criticisms made against variance-based risk loads.

8. THE "UNITS DON'T MATCH" OBJECTION TO VARIANCE-BASED RISK LOADS

Before presenting what I regard as valid criticisms of variance-based risk loads, I would like to switch sides for a moment and refute one set of common arguments made against variance. This set of arguments deals with the units of risk load and with currency translation. Consider that variance is in units of "dollars squared," while the desired cost is in units of "dollars." As any engineering or physics student knows, if the units don't match, there is a mistake somewhere in the derivation. Related to this is the criticism that variance-based risk load formulae lead to nonsensical results if one tries to switch from one currency to another.

Both these criticisms are an attack on the formula: $RL = \lambda \cdot \mathrm{Var}(L)$, where RL is risk load, L is loss, and λ is a constant. If L is in dollars, then $\mathrm{Var}(L)$ is in units of dollars squared. It is true that if λ were a unitless constant, then there would be a mismatch of units. However, λ should carry units of inverse dollars, so that the resulting risk load is, as it should be, in units of dollars.

This also refutes the currency translation paradox. Write $\$L$ for loss in dollars and £L for loss in pounds and let $£L = \kappa \cdot \$L$, where κ is the exchange rate constant. It follows that:

$$\mathrm{Var}(£L) = \kappa^2 \cdot \mathrm{Var}(\$L).$$

Ostensibly confounding results arise if λ is viewed as a unitless constant and not adjusted for exchange rate. However, if $£RL = \lambda_£ \cdot \mathrm{Var}(£L)$, where $\lambda_£ = \lambda_\$/\kappa$, then $£RL = \kappa \cdot \$RL$. Thus, the problem disappears if λ carries units of inverse currency and is properly adjusted to reflect exchange rates.

9. VARIANCE-BASED RISK LOADS MAY LEAD TO INCONSISTENT INCREASED LIMITS FACTORS

Having shown I do not find merit in every argument against variance-based risk loads, I will now turn to one I do regard as valid. In my view, the major flaw with variance-based risk loads is that they can lead to prices that may rise too rapidly by limit. For example, it might cost more to raise the limit from \$2 million to \$3 million than it does to raise the limit from \$1 million to \$2 million. With a variance-based risk load, it would not be impossible to have the following ILF table:

TABLE 1

Limit (\$000)	ILF
1,000	2.50
2,000	3.00
3,000	3.75

To see what is wrong here, consider matters from the point of view of a prospective insured whose basic limits premium is \$10,000. This insured will be asked to pay \$25,000 for \$1 million of coverage, an extra \$5,000 to increase coverage from \$1 million to \$2 million, and an extra \$7,500 to increase coverage from \$2 million to \$3 million. Breaking this down by layer and using an "M" suffix to denote million(s), the insured would see the following:

TABLE 2

Layer	Cost
0 -\$1M	\$25,000
\$1M excess of \$1M	\$5,000
\$1M excess of \$2M	\$7,500

What is exceedingly strange about this is that the loss in the lower excess layer (\$1M excess of \$1M) must be \$1 million before there is even a penny of loss in the upper excess layer (\$1M excess of \$2M). Further, loss in the upper layer can never exceed the loss in the lower layer. How can it make sense to charge more for the upper layer when it never has more loss?

This problem of inverted layer costs is referred to as "inconsistency by layer." Increased limits factors are consistent if they exhibit a pattern of declining marginal increases as the limit of coverage is raised. If the ILF formula is viewed as a function with respect to the coverage limit, and this function has a second derivative, then the ILFs are consistent if, and only if, the second derivative is negative. Consistency implies that excess layer costs decline as the attachment point is raised, assuming all layers carry the same limit of excess coverage, and assuming that excess layer costs are computed by taking differences in appropriate ILFs.

Meyers is well aware of the inconsistency problem. He tries to get around it by defining a new notion of consistency. Under the Meyers definition (see Meyers's Appendix E), one considers any two excess layers with identical limits, but different attachment points. Consistency, according to Meyers, exists if the layer with the higher attachment point always has a lower indicated price. Meyers calculates the indicated price for a layer as the sum of its expected loss plus risk load. His risk load is proportional to the variance of losses in the *layer*. In his paper, Meyers gave a proof that this calculation results in consistency under *his* definition. Subsequently, Meyers has told this reviewer that the proof may not be valid.

However, even if it were valid, it would prove much less than it may seem. Since Meyers does not calculate layer prices by taking the difference between ILFs, his notion of consistency does not apply to the ILFs, but rather to his premium calculation principle. In fact, taking differences of ILFs gives a larger indicated (variance-based) risk load for a layer than would calculating layer risk load based on the variance of layer losses. This happens because the variance in layer losses is always less than the difference in the variance of losses capped at the upper limit minus the variance of losses capped at the lower limit (see Appendix A). This is the essence of "risk reduction through layering." (See Miccolis [2].)

Therefore, Meyers is, at least implicitly, asserting that ILFs are not appropriate for pricing layers. Also inherent in his theory is the idea that the price for coverage up to a limit depends on how the coverage is layered. For example, under Meyers, coverage for the layer from zero to $2 million should have a price different than the sum of prices for coverage on the underlying $1 million plus coverage on the $1-million-excess-of-$1-million layer. Not only does this fail to produce a unique benchmark price for coverage, because it allows that different layerings of coverage could result in different prices, it also leaves open the question of what layering, if any, should be assumed when filing ILFs. Meyers believes that the bureau should file ILFs under the hypothesis that layering is not allowed, and yet also promulgate advisory factors for various possible layerings. I feel the publication of different costs for different layerings of the same coverage is no solution, and only adds to the confusion. Try as he might, Meyers is unsuccessful at defining away the consistency problem. ILFs produced with the variance-based risk load can be inconsistent.

10. VARIANCE REDUCTION THROUGH PRO RATA SHARING

It is also somewhat troubling that variance-based risk loads lead not only to "risk reduction through layering," but also to "risk reduction through sharing." To illustrate this, suppose two insurers decide to become "50-50" partners in writing a risk. Each will take half the premium and pay half the loss. If σ^2 is the variance of the total loss, then $\sigma^2/4$ is the variance of the loss covered by each insurer. If each uses a variance-proportional risk load with a common risk loading scalar, λ, then the total risk load demanded by their syndicate will be $\lambda \cdot \sigma^2/2$. If either insurer had written the risk on its own, then the risk load would have been $\lambda \cdot \sigma^2$. Due to syndication, the risk load has been cut in half. The syndicate operates, in effect, with a reduced risk load constant. More generally, if risk load is in proportion to variance and if expense considerations are neglected, then syndicates ought to be able to take advantage of "risk reduction through sharing" in order to reduce their risk loading constants.

It appears that quota share syndicates would not be allowed in the Meyers's Competitive Market Equilibrium (CME) model. I deduce this from Equation 5.3 in which Meyers derives the theoretical market risk loading constant as the (harmonic) *average* of the λs for the individual insurance companies. If the companies were allowed to form a syndicate and quota share the business, the theoretical market λ would be much lower than the average λ .

11. RESTRICTIONS ON THE "COMPETITIVE MARKET"

In the market posited by Meyers, risk load is proportional to variance and there is no excess layer or quota-share reinsurance. Yet, the variance principle implies that such reinsurance is decidedly advantageous. Implicitly, Meyers has prohibited insurers from entering into transactions that his theory says are beneficial.

Meyers must also have some hidden restrictions on the insureds to prevent them from taking advantage of analogous ways to reduce their variance-based risk loads. For example, instead of buying $2 million of coverage, an insured could opt to save money by buying a primary policy with a limit of $1 million and an excess policy covering $1 million excess of $1 million. As well, the insured could save money by getting two policies with each covering half of the insured's losses.

If Meyers had a free market model, then the insurers would be allowed to reinsure and the insureds would be permitted to package coverage. In either event, the market would soon cease to operate under the original variance-based risk load. As Venter [3] has noted, variance-based risk loads create arbitrage possibilities. Yet one aspect of competitive market theory is that "arbitrage profit possibilities are quickly extinguished by market competition"[3]. Meyers started off to build a "Competitive Market Equilibrium" theory, but ended up with ILFs that could never exist in a free market.

12. PUTTING REINSURANCE INTO THE MODEL

In my opinion, the theory ought to be extended to arrive at the theoretical premium that would prevail in a competitive market that allowed reinsurance. Additional expenses associated with reinsurance must also be considered. Since reinsurance companies do exist, such an extended theory would be a better model of reality than the highly-restricted model proposed by Meyers.

By theoretically allowing coverage to be reinsured, one is not forcing any individual insurer to abandon use of a utility function to price the risk on its net coverage. Each insurer could still use a utility function and end up demanding a risk load proportional to the variance of loss it covers. However, as previously noted, variance reduction through layering or sharing can lead to a market risk load that is not proportional to the variance of the total (unlayered and undivided) loss.

The reduction of variance through layering will lead to a hypothetical market price for coverage by limit in which risk load rises less rapidly than the variance of limited losses. Indeed, if one neglects to consider expenses associated with layering, the continual subdivision into finer layers will drive the process risk component of theoretical market risk loads to zero. This is proven in Appendix A.

However, layering has its costs. At the very least, it leads to additional processing and billing as each reinsurer needs to receive its share of the premium and pay losses in its layer. While such costs may be small, they rise as the number of subdivisions is increased. Intuitively, it is clear that there will come a point when risk load reduction through further layering will be canceled out by the additional costs. At that point, one arrives at the best possible price the theoretical market could offer.

I do not have a general formula for this lowest theoretical price under arbitrary layerings that also reflects the expenses associated with layering. However, I can present to the reader a comparable formula derived by Fred Klinker for the risk-sharing case. As shown in Appendix B, if one allows a syndicate of insurers to take percentage shares of coverage, with each charging a risk load proportional to the *variance* of its share of loss, and add in expenses which rise with the number of syndicate members, then the total risk load is proportional to the *standard deviation* of the total loss. Since layering is more efficient than sharing in reducing process variance, it is likely that the lowest theoretical market price for coverage will increase less rapidly by limit than the standard deviation of limited loss. In summary, if Meyers were to complete his theoretical foundation and allow layering and reflect associated expenses, his theory would produce risk loads whose process risk load components would not increase in proportion to process variance, but to something that rises even less steeply by limit than does the process standard deviation.

Venter [3] has advocated calculation of the risk-loaded premium using the expected value of a risk-adjusted distribution. This distribution is obtained by transforming the original loss distribution. In his paper, Venter deduces some general principles from the requirement that the market be free of arbitrage possibilities. The transformation methodology satisfies these principles. Further, according to Venter, the only premium calculation principles with the desired properties are those that can be generated from transformed distributions. However, it is not clear if any transformation would be equivalent to an extended Competitive Market Equilibrium model which incorporated reinsurance and associated expenses. Once such a model is developed, its equivalence to some transformation of distributions should be investigated.

13. SUPPLY AND DEMAND

The supply assumption in the Meyers CME theory is that each company sets a constraint on the loss variance it will tolerate. While this reviewer knows of no company that actually does this in practice, it is likely that most insurers implicitly operate under such a constraint. Thus, the assumption seems reasonable, especially in the context of modelling a theoretical market.

However, the assumption about demand could be much improved. Recall that the Meyers CME theory effectively ignores the demand side of the "supply and demand" equation under the dubious assumption that demand is inelastic. I would assert on general grounds that demand by limit might well be influenced by the pattern of increase in a set of ILFs. While it is beyond the scope of this discussion to propose a theory incorporating an explicit demand function, future research in this direction would seem worthwhile. Such a theory would require an explicit demand function by limit. Estimation of the demand function might be rather difficult. One problem is that the only data currently available is on the limits of primary policies purchased by insureds. To obtain a true picture of the demand by limit, one would need data on the total coverage afforded by the combination of primary and excess policies purchased by insureds.

The "demand side" perspective also leads to the consideration of the impact of risk retention groups. Under one approach to risk retention financing where the goal is to minimize the probability of ruin, a risk retention group must assess the total risk load proportional to the standard deviation of losses for all members of the group. Within this framework, principles of game theory were used by Lemaire [1] to calculate a fair cost allocation for each member of the group. The resulting risk load assessment for a group member is not proportional to the variance of the member's losses. This strongly suggests that variance-based risk loads would not be theoretically tolerated in a market that allowed the formation of risk retention groups. Note also that proper consideration of risk retention groups would entail consideration of expenses. One might end up with a result not too dissimilar from what one would achieve by incorporating reinsurance layering and associated expenses in the CME theory.

14. CONCLUSION

Meyers has made a valuable contribution to the field of actuarial literature with a well-written and thought-provoking paper. He has laid down some of the foundation for a solid theory of risk load and has artfully applied mathematical techniques to get a result. Meyers was courageous enough to go beyond the "single insurer, single insured" paradigm to develop a competitive market approach. Unfortunately, he posited an artificially restricted market and ended up with variance-based risk load. This is the same answer produced by the "single insurer, single insured" utility theory model. The Meyers theory does nothing to dispel valid criticisms made against variance-based risk load. Most important, his variance-based risk load formula could lead to inconsistent ILFs, where "inconsistent" retains its original and appropriate meaning as given by Miccolis [2]. In conclusion, I believe Meyers did not succeed in obtaining a risk load formula appropriate to use in bureau increased limits ratemaking.

REFERENCES

[1] Lemaire, Jean, "Cooperative Game Theory and Its Insurance Applications," *ASTIN Bulletin*,, Vol. 21, No. 1, 1991, p. 17.

[2] Miccolis, Robert S., "On the Theory of Increased Limits and Excess of Loss Pricing," *PCAS* LXIV, 1977, p. 27.

[3] Venter, Gary G., "Premium Calculation Implications of Reinsurance Without Arbitrage," *ASTIN Bulletin*, Vol. 21, No. 2, 1991, p. 223.

APPENDIX A

IMPACT OF LAYERING ON PROCESS VARIANCE

Suppose coverage up to limit K is achieved by stacking n layers, and assume the indicated risk-loaded pure premium for the coverage is the sum of the indicated risk-loaded pure premiums for the layers. Suppose the indicated risk-loaded pure premium for each layer is calculated as the sum of expected loss in the layer plus risk load, and assume risk load is proportional to variance. Consider the "process" variance component of the risk load. I will first show that this layering reduces the risk load compared to the risk load of the unlayered coverage. I will then show that the sum of process variance risk loads becomes smaller as the layer subdivisions grow ever finer. In other words, if expenses are not included in the analysis, infinite layering will drive process variance risk load to zero.

Let $0 = k_0 < k_1 < \ldots < k_n = K$ be the layer end points.

Let $\Delta k_i = k_i - k_{i-1}$ for $i = 1, 2, \ldots, n$.

Given loss severity random variable, X, define $X_i = \min(X, k_i)$ and let $\mu_i = \mathrm{E}(X_i)$ and $\sigma^2{}_i = \mathrm{Var}\,(X_i)$.

Let Y_i denote the loss in the i^{th} layer so that $Y_i = X_i - X_{i-1}$.

Thus, $\mathrm{E}(Y_i) = \Delta\mu_i = \mu_i - \mu_{i-1}$.

Now, set $\tau_i^2 = \mathrm{Var}\,(Y_i)$.

Proposition

$$\mathrm{Var}\,(X_n) = \sum_{i=1}^{n} \tau_i^2 + 2 \cdot \sum_{i=1}^{n-1} (k_i - \mu_i) \cdot \Delta\mu_{i+1} \qquad \text{(A.1)}$$

Proof: Since $X_n = \sum Y_i$, it follows that:

$$\mathrm{Var}\,(\min(X, K)) = \sum_{i=1}^{n} \tau_i^2 + 2 \cdot \sum_{i<j} \mathrm{Cov}(Y_i, Y_j)\ .$$

Now consider that, for $i < j$,

$$\mathrm{Cov}(Y_i, Y_j) = (\Delta k_i)(\Delta\mu_j) - (\Delta\mu_i)(\Delta\mu_j) = ((\Delta k_i) - (\Delta\mu_i)) \cdot (\Delta\mu_j) .$$

The desired result then follows since:

$$\Delta k_1 + \Delta k_2 + \ldots + \Delta k_i = k_i \text{ and } \Delta\mu_1 + \Delta\mu_2 + \ldots + \Delta\mu_i = \mu_i .$$

Since the latter sum in equation A.1 has only non-negative terms, one can immediately conclude that layering reduces total risk load if risk load is proportional to variance.

Variance Reduction Through Layering

$$\mathrm{Var}(\min(X, K)) \geq \sum_{i=1}^{n} \tau_i^2 \qquad \text{(A.2)}$$

The inequality is strict if there is some non-zero probability of a non-zero loss strictly less than the upper limit, K.

If the subdivision process continues indefinitely, the process risk load will shrink to nothing.

Infinite Layering Drives Process Variance to Zero

$$\text{If } \max \{\Delta k_i : i = 1, 2, \ldots, n\} \to 0 \text{, then } \sum \tau_i^2 \to 0 . \qquad \text{(A.3)}$$

Proof: Consider that $\tau_i^2 < \Delta k_i^2$ so that $\sum \tau_i^2 < \sum \Delta k_i^2$.

Further observe that $\sum \Delta k_i^2 < \max \{\Delta k_i\} \cdot \sum \Delta k_i = \max \{\Delta k_i\} \cdot K$.

The result follows since the latter expression approaches zero by assumption and by the fact that K is unchanged by the layering.

APPENDIX B

RISK LOADS CHARGED BY A PRO RATA SHARING SYNDICATE

(Simplified from the unpublished work of Fred Klinker)

If there is a competitive insurance market with risk sharing among the insurers via pro rata reinsurance, and if the expenses associated with risk sharing are considered, then variance-based risk loads at the level of the individual firm become standard deviation-based at the level of the market. This seemingly paradoxical result stems from a trade-off between the increased expense and decreased "risk" associated with sharing.

Assume each insurance firm charges a risk load proportional to the variance of loss for the coverage it provides. Now suppose there is a syndicate of n insurers in which each takes a fixed share, ρ_i , of the total loss, where the ρ_i are non-negative and sum to one.

The loss experienced by the syndicate as a whole, L, is a random variable with expectation $\mathrm{E}(L)$ and variance $\mathrm{Var}\,(L)$. The loss experienced by the i^{th} insurer is $\rho_i L$ with expectation $\rho_i\mathrm{E}(L)$ and variance $\rho_i^2\,\mathrm{Var}\,(L)$. Neglecting regular underwriting expenses and the usual profit load, the components of the net premium charged by this insurer are:

Expected loss:	$\rho_i\mathrm{E}\,(L)$
Fixed expense (transaction costs):	ε
Risk load:	$\lambda_i\,\rho_i^2\,\mathrm{Var}(L)$

The resulting net premium charged by this insurer is

$$P_i = \rho_i\,\mathrm{E}(L) + \varepsilon + \lambda_i\,\rho_i^2\,\mathrm{Var}\,(L). \tag{B.1}$$

The resulting premium charged by the syndicate of n insurers is, recalling that the ρ_i are weights summing to unity,

$$P(n, \rho_i) = \mathrm{E}(L) + n\varepsilon + \mathrm{Var}(L) \sum_{i=1}^{n} \lambda_i \rho_i^2 . \tag{B.2}$$

Now minimize the premium given in equation B.2, holding n fixed. Equation B.2 will be minimized by minimizing

$$\sum_{i=1}^{n} \lambda_i \rho_i^2 .$$

The minimization is constrained since the ρ_i must sum to unity. The constrained minimization can be solved by the method of Lagrange multipliers.

$$\frac{\partial}{\partial \rho_i} \sum_{j=1}^{n} \lambda_j \rho_j^2 = \Lambda \frac{\partial}{\partial \rho_i} \sum_{j=1}^{n} \rho_j \ \text{ for all } i .$$

$$2\lambda_i \rho_i = \Lambda \ \text{ for all } i .$$

In other words, the optimal ρ_i are proportional to the reciprocals of the λ_i for all i .

Imposing the constraint that the ρ_i sum to one leads to

$$\rho_i = \frac{1}{\lambda_i \left(\sum_{j=1}^{n} \frac{1}{\lambda_j} \right)} . \tag{B.3}$$

It follows that

$$\sum_{i=1}^{n} \lambda_i \rho_i^2 = \frac{1}{\sum_{j=1}^{n} \frac{1}{\lambda_j}} = \frac{1}{n \left\langle \frac{i}{\lambda} \right\rangle} . \tag{B.4}$$

where the angled brackets denote the average value. The minimum premium, for fixed n, from equation B.2 and the above, is

$$P(n) = \mathrm{E}(L) + n\,\varepsilon + \frac{\mathrm{Var}(L)}{n\left\langle \frac{1}{\lambda} \right\rangle}. \tag{B.5}$$

Note the behavior of equation B.5 with respect to n, the number of insurers in the syndicate. There is a term due to fixed transaction costs per insurer which increases linearly with the number of insurers. Another term declines as the reciprocal of the number of insurers, which captures the declining aggregate of the variance-based risk loads of the individual insurers. The decline is due to the increasing spread of risk among more insurers.

Because of the above behavior, there is an optimal n for which the premium, $P(n)$, is minimized. This occurs where the derivative of $P(n)$ with respect to n vanishes.

$$0 = \frac{d}{dn}P(n) = \varepsilon - \frac{\mathrm{Var}(L)}{n^2\left\langle \frac{1}{\lambda} \right\rangle}.$$

Hence,

$$n = \sqrt{\frac{\mathrm{Var}[L]}{\varepsilon\left\langle \frac{1}{\lambda} \right\rangle}} \tag{B.6}$$

and the minimum value of $P(n)$ is

$$P = \mathrm{E}(L) + 2\left(\sqrt{\frac{\varepsilon\,\mathrm{Var}(L)}{\varepsilon\left\langle \frac{1}{\lambda} \right\rangle}}\right) \tag{B.7}$$

Equation B.6 provides the optimal number of insurers and equation B.7 provides the minimum premium over all possible numbers of insurers and all possible pro rata distributions of risk among insurers. The last term of equation B.7 can be identified as the market risk load. In the immediate context of increased limits factors, with increasing limit—hence increasing Var (L)—the optimal number of insurers on the contract will increase according to equation B.6. Also, the market risk load of equation B.7 will increase only as the square root of Var (L); i.e., as the standard deviation of L, where L is the aggregate loss across the syndicate, despite the fact that individual insurer risk loads are variance-based rather than standard deviation-based.