The Unearned Premium Reserve for Warranty Insurance

Richard L. Vaughan, FCAS, FSA, MAAA

Abstract. The Unearned Premium Reserve (UPR) is the largest liability on the balance sheet of most writers of Warranty Insurance. Despite the specialized nature and small size of the line, the NAIC has seen fit in recent years to discuss the UPR for Warranty Insurance in its Regulatory Guidance on Property and Casualty Statutory Statements of Actuarial Opinion.

The UPR is subject to the rules set out for long-duration contracts in Statement of Statutory Accounting Principles 65 (SSAP 65). Because of the high frequency and narrow size-of-loss distribution of Warranty claims, conventional reserve estimators such as Bornhuetter-Ferguson work quite well to estimate the UPR, but to be applied properly they require special modifications. In particular, it is necessary to adjust for unreported losses in recent diagonals of the issue-versus-breakdown lag triangle, to adjust for exposures declining by development month because of cancellations, to estimate appropriate tail factors, to modify expected emergence patterns for coverage of an obligor's failure to perform, and to reserve appropriately for unpaid future refunds; the last two items are not specifically addressed in the regulations.

This paper discusses the purpose and structure of the UPR for Warranty Insurance in general, describes the necessary modifications of conventional actuarial methods in detail, and illustrates them with examples.

1. INTRODUCTION

1.1 Warranty Insurance

Manufacturers of consumer products are required, either by common law or by the Uniform Commercial Code or similar legislation, to warrant that their products are reasonably fit for their intended use. Making a virtue of necessity, nearly all manufacturers formalize this "implied warranty of merchantability" as an express warranty for a stated period, and often advertise such *factory*, or *manufacturers'*, *warranties* as a guarantee of satisfaction and as evidence of their own confidence in their products.

An extensive market has arisen for contracts that supplement factory warranties by running for longer terms and/or by covering more parts and services. Most commonly such *extended warranties* or *extended service contracts* are purchased at the same time as the underlying product, or shortly thereafter. Their appeal to the individual consumer depends on his or her risk aversion and his or her liquidity relative to the cost of repairing or replacing the particular product should it prove defective.

The classic example of extended warranties with wide consumer appeal are those covering automobiles, with their high initial cost, their many components subject to failure, the high price of replacement parts, and the large labor component of repairs. This paper discusses Warranty reserving mainly from the viewpoint of contracts covering automobiles, because they have special features, such as terms defined in months and miles and separate manufacturers' warranties for power train and for other parts, which, when incorporated in a model, also allow that model to cover simpler warranty contracts as special cases.

While factory warranties attach automatically to every product sold, obligate the manufacturer, and are embedded in the price of the product, extended warranties are usually optional, obligate the retailer or a third party, and are paid for by the consumer with a single premium. Retailers offering such contracts on the products they sell, and assuming such service obligations themselves, are known as *obligors*, and in most states are not regulated as insurers. For tax reasons, many retailers actually issue their extended warranties through affiliated warranty companies. Obligors of this kind are also not regulated as insurers in most states, provided they insure their obligations with a licensed insurer.

Most manufacturers retain the risks associated with their warranties and keep the associated loss data proprietary. Therefore manufacturers' warranties, while they are essential to the design, pricing, and reserving of other Warranty contracts, are not themselves part of the Warranty Insurance marketplace. On the other hand, most warranty companies insure all or some of the risks of their extended service contracts. They may do so to satisfy legal requirements, for the usual benefits of risk transfer, or to obtain the expertise of the insurer in administration, data management, ratemaking, and reserving. Perhaps the most critical specialized expertise of a Warranty underwriter is the ability to calculate the Unearned Premium Reserve (UPR), on which both solvency and the ability to measure rate adequacy depend. The remainder of this paper discusses the UPR in the context of other reserves, of statutory requirements, and of the special adaptations necessary to estimate it for Warranty Insurance.

The extended service contracts we are concerned with in this paper are supplemental to the factory warranty. They may extend it by covering a longer term in months or miles, more components of the product, additional services such as towing, a wider selection of repair facilities, or peripheral contingencies such as those envisioned by Guaranteed Asset Protection (GAP) and Involuntary Unemployment Insurance (IUI) policies. However, the principles of reserving for extended warranties are perfectly applicable to factory warranties and might equally well be used by manufacturers in accounting for their warranty liabilities.

In recent years, Warranty Insurance, despite its relatively small and specialized nature, has drawn the attention of regulators because of the exposure of insurers to the failure of obligors for which they have written Warranty Insurance on a failure-to-perform basis. In 2009, the NAIC introduced a discussion of this issue in its Regulatory Guidance on Property and Casualty Statutory Statements of Actuarial Opinion, which was included as Appendix 9a of the American Academy's Practice Note on the Statements of Actuarial Opinion.

1.2 Loss Reserves and the UPR

Like any other insurance product, Warranty Insurance gives rise to obligations for claims incurred but not reported (IBNR) and reported but not paid (RBNP), together making up the reserve for losses incurred but not paid (IBNP). Except for GAP insurance, reporting and settlement lags are short, the size-of-loss distribution is narrow, and frequencies are high, so these loss reserves are both modest in size and straightforward to estimate using conventional triangle-based actuarial techniques.

Unlike many other insurance products, Warranty Insurance policies tend to be of long duration and are usually purchased with a single premium, giving rise to a very significant obligation for losses "paid for" by the single premium but not yet incurred. This obligation is provided for by the Unearned Premium Reserve (UPR). The UPR is by far the largest liability of most Warranty insurers, and is therefore of paramount importance in evaluating their solvency and the solvency of individual Warranty programs. Moreover, the UPR is complementarily related to cumulative earned premium, so it is also of paramount importance when evaluating loss ratios and rate adequacy.

As with other lines of insurance, we may analyze Warranty loss triangles (or similar arrays of more than two dimensions) to establish the UPR and/or loss reserves or to evaluate the adequacy of carried reserves determined in other ways. Typically we analyze the UPR on a policy-month basis (which we shall call *issue month*), the IBNR or IBNP reserves on an accident-month basis (which we shall call *breakdown month*), and the RBNP reserves on a report-month basis.

Many standard actuarial techniques, such as chain-ladder and Bornhuetter-Ferguson loss development, the Cape Cod estimator of expected loss ratios, and the analysis and application of trend and seasonality, may be applied successfully to Warranty Insurance. However, the necessity of issue-month loss development, the importance of cancellations and refunds, and the prevalence of coverage on a failure-to-perform basis in whole or in part, all require certain technical adjustments to the standard techniques, which are a main focus of this paper.

2. BACKGROUND AND METHODS

2.1 Regulatory Requirements

Most Warranty contracts are issued for terms longer than one year. Such *long-duration* contracts are subject to the requirements of Statement of Statutory Accounting Principles 65 (SSAP 65), which may be paraphrased as follows: the UPR must be at least as great as the greatest of (1) the amount payable if all policyholders surrendered their contracts for refund on the accounting date, (2) the sum over all in-force policies of the gross premium times the expected fraction of ultimate losses not yet incurred as of the accounting date, and (3) the expected present value of future losses, from in-force policies, not yet incurred as of the accounting date. These are called Tests 1, 2, and 3. Test 1 values the surrender option, albeit very conservatively; Test 2 recognizes earnings as risk is borne and services performed; Test 3 addresses claim-paying ability.

All three tests apply prospectively to the portfolio of contracts remaining in force at the valuation date after earlier cancellations. Test 1 assumes that all of these policies cancel immediately, and Test 2 by implication assumes no further cancellations. Test 3 may take into account the effect of future cancellations on expected losses, but it does not measure the expected refunds payable for such cancellations. Therefore Test 3 is an incomplete measure of unpaid future obligations and for that purpose we believe it should be supplemented with an estimate of expected future refunds.

For Warranty business, Test 2 is usually dominant. It is normally greater than Test 1 because the UPR on most policies is greater than the required refund; it is normally greater than Test 3 because most policies are priced to produce a loss ratio less than 1.00. Moreover, the formulas or vectors of monthly factors used by Warranty insurers to calculate the UPR for individual contracts are normally calibrated to match the unincurred fraction of ultimate losses, in the manner of Test 2, since only this may be converted into earned premium suitable for loss ratios or other measures of performance.

A few Warranty contracts are sold for terms of 12 months or less. The UPR for such contracts may simply be taken as pro rata, i.e., gross premium times the unelapsed fraction of the total term, as is usual for other lines. However, since Warranty insurers have in place a mechanism for calculating more precise UPR's satisfying SSAP 65 for their long-duration contracts, they may apply the same technique to shorter contracts, with results that are more accurate and, usually, a bit more conservative.

SSAP 65 applies to an insurer's long-duration business in aggregate and need not be satisfied for any given contract or program. However, it is good practice to attempt to satisfy Test 2 for each program considered separately, for then the aggregate will automatically satisfy Test 2 and the inception-to-date loss ratio for each program will be a reasonable predictor of the ultimate loss ratio. Test 3 may fail for a few programs running loss ratios greater than 100%; correcting this is usually a pricing issue. Test 1 may fail for coverages such as GAP that earn more rapidly at first than pro rata. This will only be a problem in aggregate for companies that write mainly GAP or similar coverages. For them, the system UPR should probably continue to be on a Test 2 basis, but the UPR shown on the books may need to be taken from Test 1.

2.2 Reserve Structure

The UPR is chronologically the first component of a reserve structure ultimately designed to recognize all obligations "paid for" but not yet paid; it addresses those obligations not yet incurred. For this reason, in analyzing the UPR, the second, or development, dimension of a loss triangle is usually issue-to-breakdown lag. However, for some sublines of Warranty Insurance the insurer may find it more convenient to analyze the combined UPR and IBNP reserves by issue month versus issue-to-payment lag, or the combined UPR and IBNR reserves by issue month versus report lag, in effect giving rise to several possible reserve structures: (a) UPR, IBNR, and RBNP, or (b) (UPR+IBNR, called UPR) and RBNP, or (c) UPR and (IBNR+RBNP, called IBNR), or (d) (UPR+IBNR+RBNP, called UPR). Because the statutory UPR is governed not only by expected unincurred losses but by the expected emergence pattern of losses *applied to gross premium*, it is usually conservative, so that when UPR principles are applied to some or all of the loss reserves, they in turn become conservative. While this practice is therefore benign, it may be necessary to separate the total reserve into "proper" components for annual statement purposes.

2.3 Earnings

The UPR governs the recognition of earnings from an individual policy or cohort of policies, through the general formula (earned premium) = (written premium) – (change in UPR). At the moment of writing, the UPR equals the written premium, having previously been zero, so no earnings are recognized immediately. Thereafter the UPR for a contract is monotonic non-increasing, change in UPR is nonpositive, and earned premium is nonnegative. When the UPR reaches zero, premium earned since inception equals the original written premium and the contract is said to be fully earned.

2.4 Cancellations and Refunds

A characteristic feature of Warranty policies is that they are subject to cancellation for refund throughout their term. The amount refunded is specified by law or by contract and is usually different from the UPR carried at the moment of cancellation, giving rise to a gain or loss. Cancellations mean that in-force exposure for a cohort of contracts already issued may decline from month to month. By itself this creates a problem for issue-month loss development, which we address below.

Moreover, the usual accounting treatment of cancellations, which nets refunds against premiums written in the same calendar period, even though the refunds may have arisen from earlier contracts, is awkward and creates difficulties for actuarial analysis, especially when the actuary is attempting to measure the performance of a cohort of written contracts.

We have found it useful to distinguish premium earned by providing coverage from premium earned by cancellation, and we digress here to discuss this in some detail because of its close connection to the UPR. First, we define *UPR released by cancellation* to be the UPR on canceling policies just before cancellation. Then the following definitions relate to premium used to provide coverage:

(pure written premium) = premium for new policies, gross of refunds

(pure earned premium) = (pure written premium) – (UPR released) – (change in UPR)

(pure loss ratio) = (incurred losses) / (pure earned premium)

The corresponding figures in the financial statements are defined somewhat differently:

(statement written premium) = premium for new policies, net of refunds paid in month (statement earned premium) = (statement written premium) – (change in UPR)

(statement loss ratio) = (incurred losses) / (statement earned premium)

The "pure" definitions are very useful when tracking a cohort, for which there is no written premium after the first month. The "statement" definitions are confusing for this purpose, since the actual financial statements would show the initial written premium reduced by refunds paid in the same calendar month on behalf of policies written in that month or earlier months, and would show refunds paid in later months as part of the written premium of later cohorts.

Note that (statement earned premium) = (pure earned premium) + (gain from cancellations), where (gain from cancellations) = (UPR released) – (refunds). For Warranty Insurance on automobiles, where refunds are close to pro rata while UPR declines more slowly, (gain from cancellations) is usually positive, the statement earned premium is greater than the pure earned premium, and the statement loss ratio is less than the pure loss ratio.

The "pure" definitions above involve losses but not refunds. We can create similar definitions treating refunds, or refunds plus losses, in a manner parallel to losses, as follows. Here the refunds are only for policies in the cohort being tracked:

(refund ratio) = (refunds) / (UPR released by cancellation)

(payout ratio) = (incurred losses + refunds) / ((pure WP) - (change in UPR))

Note that UPR released by cancellation is analogous to pure earned premium except that it measures premium earned through cancellation rather than through coverage. Therefore the refund ratio is analogous to the pure loss ratio. Finally, the payout ratio is also similar to the pure loss ratio except that refunds are treated as equivalent to losses; its denominator, (pure WP) – (change in UPR), is the total premium earned either by coverage or by cancellation.

If we make the following two definitions,

(cancelled UPR ratio) = (UPR released by cancellation) / ((pure WP) - (change in UPR))

(earned premium ratio) = (pure earned premium) / ((pure WP) - (change in UPR))

then

(payout ratio) = (cancelled UPR ratio)(refund ratio)+ (earned premium ratio)(loss ratio)

showing how the parallel treatments of losses and refunds fit together. Here the cancelled UPR ratio is the fraction of the total premium earned in *some* manner that is earned by cancellation, and earned premium ratio is the complementary fraction earned by coverage.

These ratios are most useful over the time period from inception to ultimate, for then the initial UPR, the final UPR, and the change in UPR are all zero, the "big" denominators just equal the pure written premium, and the ultimate payout ratio gives a good measure of premium adequacy:

(ultimate payout ratio) = (ultimate losses plus refunds) / (pure written premium)

By contrast,

(ultimate statement loss ratio) = (ultimate losses) / ((pure written premium) – refunds)

which will not, in general, equal the ultimate payout ratio even if the refunds are those paid over time on behalf of the cohort rather than those from any source paid in the initial calendar month.

2.5 Carried Reserves

In principle, the management of a Warranty insurer could establish its UPR and loss reserves shortly after the end of each accounting period by developing inception-to-date experience data. But most insurers find it more practical (a) to embed formulas or strings in their administrative systems to generate the UPR automatically from in-force contract data at each month's end, (b) to pull the RBNP reserve directly from reported loss data, and (c) to calculate the IBNR as a factor times RBNP, recent paid losses, or some similar base quantity. These insurers still analyze inception-to-date experience, but do so to validate or modify their strings and formulas, rather than to establish the carried reserves directly.

Strings are vectors of UPR factors that are stored with the data for each contract and multiplied by gross written premium for that contract to obtain the UPR at each elapsed month's end from issue to expiration. These factors are normally adjusted so as to apply to an entire month's cohort of similar contracts, on the assumption that contracts are issued uniformly throughout the month. The usual adjustment at lag n is to take an average of the factor that would be held at lag n and the factor that would be held at lag n-1, if all policies were written at the start of the month; this is the so-called one-half-month adjustment.

In principle the string could be extended beyond the term in months to accommodate "goodwill" claims paid after expiration, but commonly the length of a string equals the term in months. The factor for the moment of issue is 1, and need not be stored with the string; similarly, the factor for any earlier month's end is 0 (no UPR is needed yet since no premium has been received) and the factor for the end of any month beyond the end of the string is also 0 (contract fully earned).

There may be many strings associated with a Warranty program. The choice of string for a given contract depends on characteristics such as term, manufacturer's warranty, and type of product insured. The string is assigned when the contract is first entered into the administrative system, and is usually fixed thereafter, ensuring a stable accounting treatment through the life of the contract.

Formulas are rules for determining UPR factors from contract characteristics and elapsed months. They may be mathematical formulas in the usual sense, such as pro rata or Rule of 78, or they may be lookup tables in every respect analogous to a set of strings, except not stored with the contract. They normally include the one-half-month adjustment. The factors are generated fresh at each valuation; making it possible to change formulas easily and keep UPR patterns responsive to current estimates, for all contracts, old as well as new. The formula UPR for a contract equals its initial written premium multiplied by the formula UPR factor.

Just as formulas may in fact be lookups of strings, strings may have originally been derived from formulas. A set of UPR factors, whether described by string or by formula, may be called a *UPR curve* because of its appearance when displayed graphically.

2.6 Graphical Representation of UPR and Earnings

The UPR for a block of contracts of term T months may be plotted against lag t; it will equal the written premium at lag t=0 and will equal zero at full term, t=T, or at t=T+1 with the one-half month adjustment. Usually we rescale such a graph to represent UPR factors, starting with 1 at t=0 and reaching 0 on or before t=T+1.

Such a graph need not be continuous; it is possible to conceive of products for which claims, if any, emerge at a few discrete times, creating a step-function UPR curve. We shall ignore such special cases here. Actually, in practice all strings and many formulas *are* step functions with discontinuities at the end of each elapsed month. For our purposes it is convenient to "connect the dots" and treat such strings and formulas as continuous for nonnegative real t.

As mentioned above, in the absence of new written premium, earnings are measured by negative change in UPR. Therefore, (a) the complement of the UPR curve is the cumulative earnings factor curve, (b) the negative slope of the UPR curve is proportional to the earnings rate, (c) the cumulative earnings factor curve is an ogive like that of a cumulative distribution function, and (d) the UPR curve is a mirror image of the cumulative earnings curve around the line y=0.5. *Incremental* earnings are often step functions, for example, assuming one level while the full manufacturer's warranty is in effect, another level when only the power-train warranty is in effect, and still another level when both parts of the manufacturer's warranty have expired. This produces changes of slope in the UPR curve but leaves it continuous.

One example of a UPR formula is linear or "pro rata", equivalent to a uniform incidence of losses over an earning period defined by the term and manufacturer's warranty. This is theoretically correct for equipment each of whose components has an exponential distribution of time to first failure, where each failed component is replaced with an identical one, and where there is neither trend nor "breakage" (contract abandonment). Figure 1 shows (to different scales) a uniform earnings pattern over 24 months and the corresponding UPR factors; this graph also illustrates the effect of the one-half-month adjustment for uniform writings throughout the month.

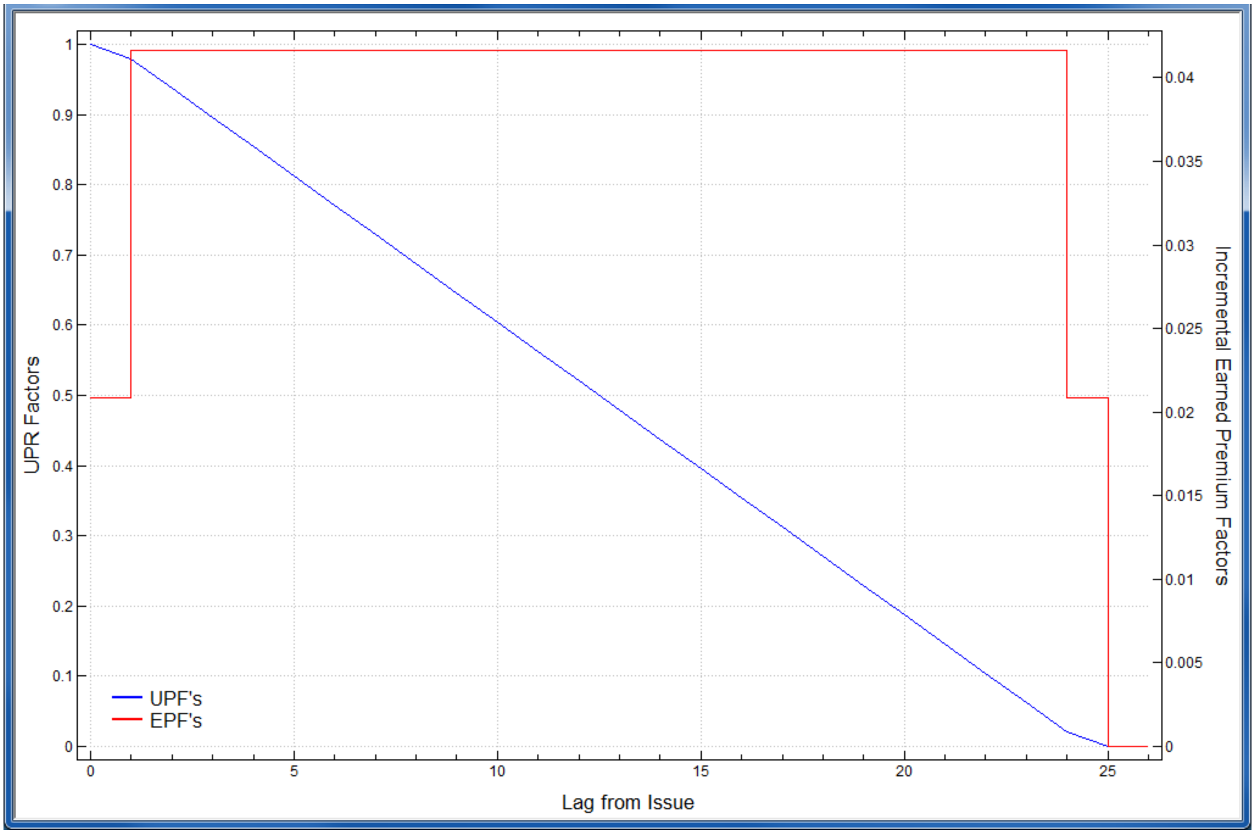


Fig. 1. Uniform earnings and pro rata UPR factors for term 24 months

Note that the one-half-month adjustment as graphed is only an approximation to uniform writings *within* the first and the last months; the true earnings pattern would not be constant, and the true UPR pattern within those months would be a second-degree curve.

The pro rata UPR formula is often modified to start at the end of the manufacturer's warranty (MW) instead of at issue. Warranties on so-called "brown and white" goods (electronics and appliances) are often reserved in this manner, with considerable justification from experience. Figure 2 illustrates 24-month pro rata earnings from the end of a 12-month MW.

Another example is sum-of-digits or Rule of 78, a second-degree curve whose first differences (evaluated at successive months' ends) are proportional to the number of months remaining in the term. This is theoretically appropriate for situations where the size of loss decreases linearly to zero at expiration while the probability of a loss remains constant, again with neither trend nor breakage. An example of such coverage would be a warranty covering failure of parts subject to normal wear and tear, where reimbursement reflects the amount of use already received from the part. This pattern is illustrated in Figure 3. Again, earnings have been shown as a step function and the UPR factors as a corresponding stepwise-linear function, reflecting the common practice of calculating UPR factors only at months' ends.

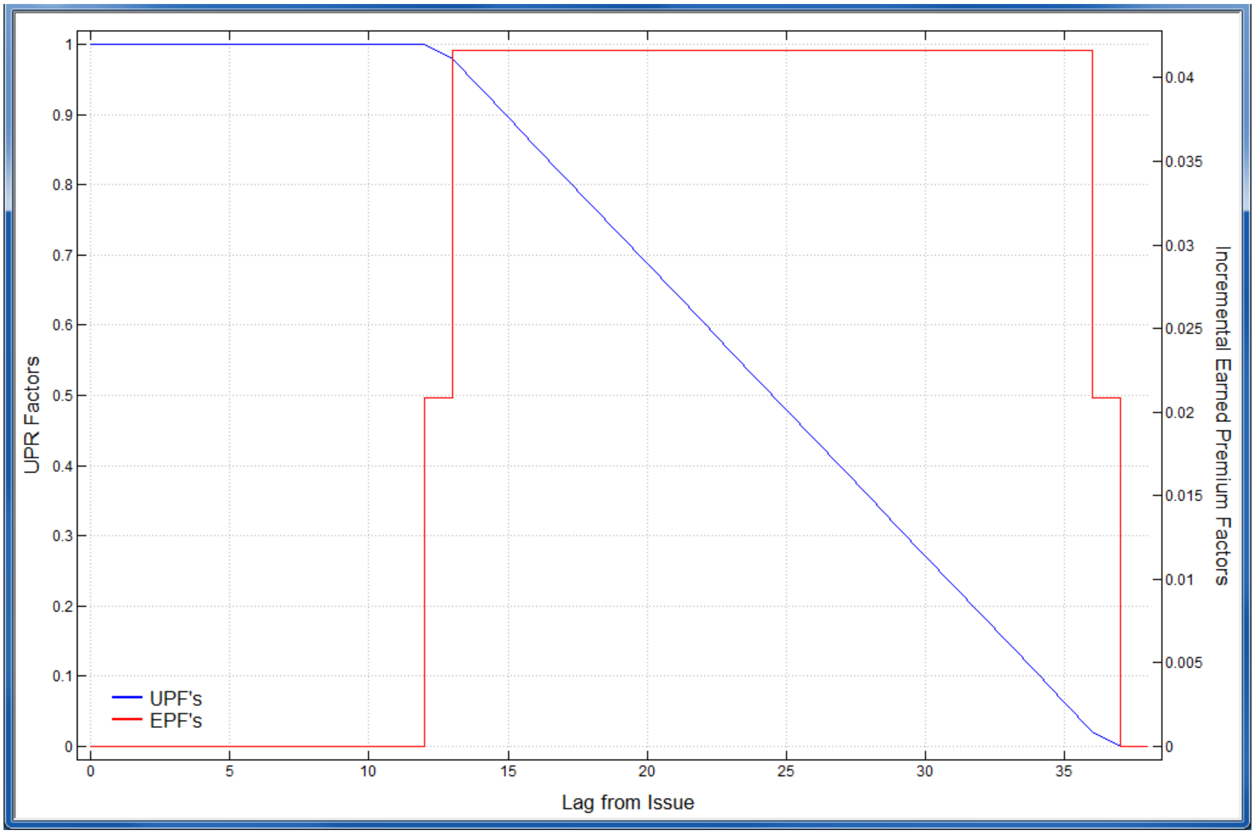


Fig. 2. Pro rata from end of manufacturer's warranty, with one-half-month adjustment.

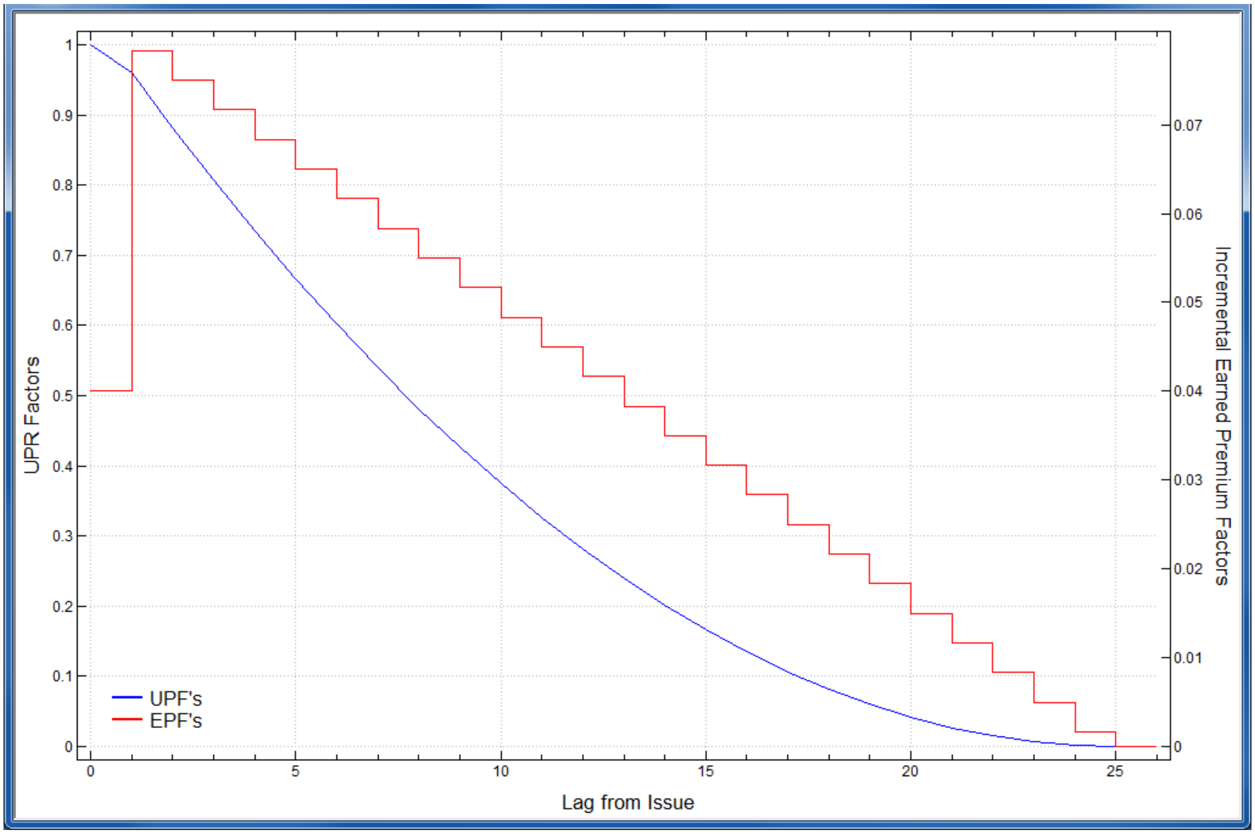


Fig. 3. Rule of 78, with one-half-month adjustment.

Here, with a term of 24 months, the "Rule of 78" should more properly be called "Rule of 300", or just "sum of digits". The Reverse Rule of 78 – first differences proportional to number of months elapsed – is sometimes used as a UPR formula, with little theoretical justification, other than that its slow early earnings produce a "shoulder" in the UPR curve that resembles that produced by the manufacturer's warranty for automobile business. This is illustrated in Figure 4. Interestingly, the Reverse Rule of 78 *is* theoretically appropriate as a UPR formula for Warranty Insurance providing contractual liability coverage for an obligor, without coverage of the underlying contracts unless the obligor fails to perform, where the underlying contracts earn uniformly, and where the probability of failure to perform (i.e., bankruptcy) approaches zero.

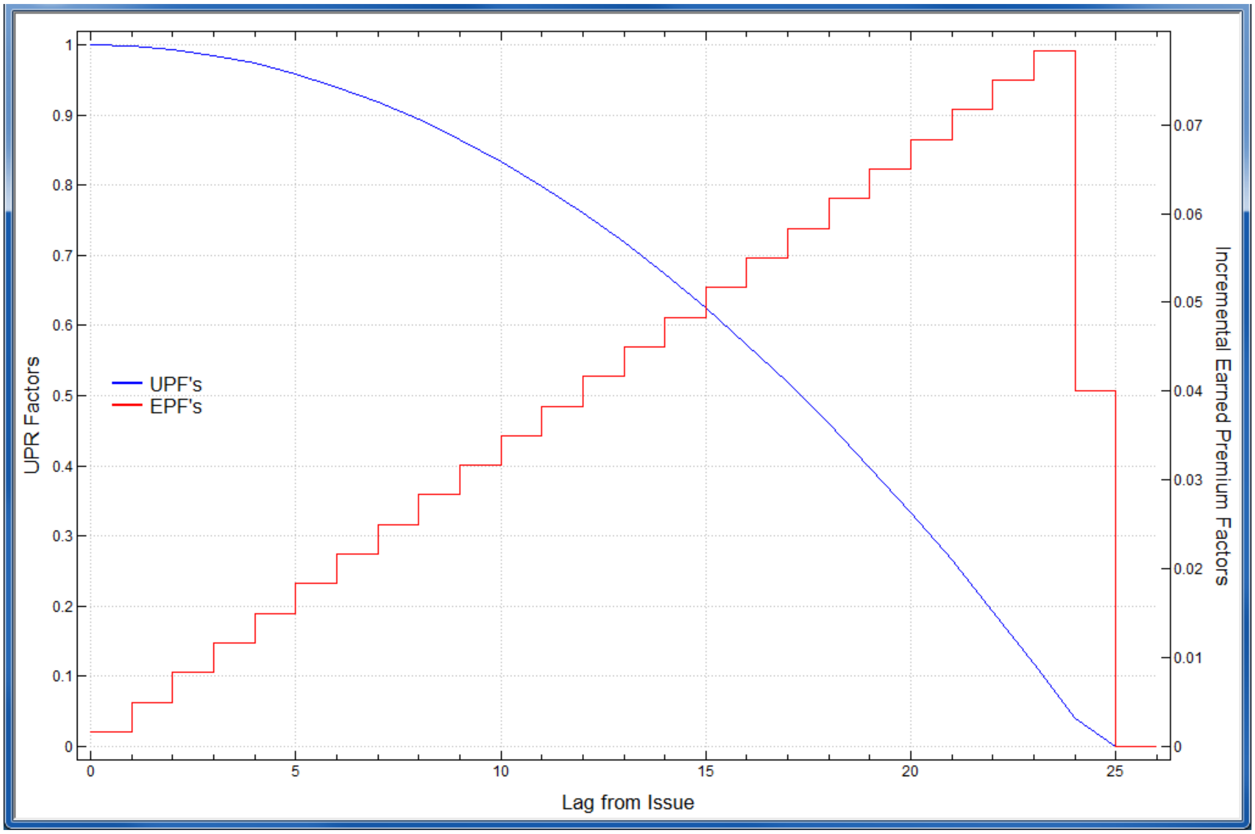


Fig. 4. Reverse Rule of 78, for term 24 months, with one-half-month adjustment.

Some Warranty administrators are known to operate with just three possible choices for UPR formula: pro rata from end of MW, Rule of 78, and Reverse Rule of 78.

2.7 Indicated Reserves

By an *indicated reserve* we mean an estimate of a liability based on analysis of loss experience. The reliability of such analysis is enhanced when summed over a collection of subdivisions of the data, each large enough to be credible but also as homogeneous as practical. Usually this means grouping contracts by term, although finer subdivisions, involving other contract or product details, may sometimes be necessary. For purposes of this discussion, we assume we have already subdivided the data into one or more reasonably homogeneous and reasonably large collections of contracts and associated claims.

Case reserves are usually good estimators of the RBNP reserve since pending Warranty payments are often known quite accurately at the report date. So loss reserving amounts to estimating either the IBNR reserve from breakdown dates versus report lags or estimating the IBNP reserve from breakdown dates versus payment lag and then subtracting the RBNP to obtain the IBNR reserve. Either estimate may be done with the chain-ladder estimator, not requiring any measure of exposure by breakdown month, or with the Bornhuetter-Ferguson estimator, using the results of a UPR analysis to obtain earned exposure. The actual calculations are straightforward and familiar and we do not discuss them further.

The indicated UPR is more complex and involves choices that may depend on the purpose of the analysis. When evaluating the adequacy of the carried UPR for a single program, a SSAP 65 Test 2 estimate may be appropriate; when evaluating rate adequacy, Test 3; when evaluating compliance with SSAP 65 in aggregate, Tests 1, 2, and 3.

We may evaluate SSAP 65 Test 1 directly by applying the refund formula to each contract in-force at the valuation date and summing the results. The typical refund formula is a flat 100% for 30 or 60 days, thereafter pro rata between issue and expiration (meaning there is a discontinuity at the end of the flat refund period), less a small surrender charge capped as a multiple of premium. This is a routine non-actuarial calculation and we do not discuss it further.

From an actuarial perspective the key to evaluating Test 2 is that we must have a sound technique for estimating issue-to-breakdown lag patterns *after* removing the effect of cancellations. Part of the solution involves modifying Chain-Ladder loss development to accommodate exposure declining across each row of the triangle, part involves adjusting for the deficiency of later diagonals of the issue-breakdown lag triangle because of unreported losses, and part involves special procedures for estimating tail factors. These techniques are discussed in sections 2.11, 2.12, and 2.13, below.

We mention in passing another estimator of loss emergence patterns, which Bühlmann [2] called Complementary Loss Ratio and Stanard [5] called Additive, but which we prefer to call Partial Loss Ratios. This method fits an additive model with only column effects to the triangle of partial loss ratios. Stanard's simulations found it efficient and unbiased. For our purposes, it appears to be affected more than the adjusted Chain Ladder when there is trend in the historical losses that is not well matched with trend in the premium or other measure of exposure.

Evaluating Test 3 requires all of the above plus sound techniques for estimating loss ratios and projecting future in-force exposures. In our work we usually project in-force premiums using a survivorship model analogous to chain-ladder loss development, but applied to the premium surviving cancellations at each lag. Next we use loss development to estimate issue-to-breakdown and breakdown-to-payment lag patterns. With these in hand, we use variations of the Cape Cod technique to estimate expected loss ratios, and then project future losses using a Bornhuetter-Ferguson model. These techniques are discussed in 2.14 below. In some ways they are more easily understood in the Warranty context than for other lines of business; for example, when the first dimension of our triangles is issue month, the Bornhuetter-Ferguson expected emerged exposure is simply earned exposure and the Cape Cod ELR is simply the ITD loss ratio.

Often we are not so much interested in the actual UPR for a block of business as in the average string of UPR factors, one for each lag, that would produce the correct UPR, not only at the given valuation date but at any date. Or we may be interested in the string of UPR factors appropriate for each contract in the data, or for proposed contracts yet to be written. We call such strings, derived from experience, *indicated UPR factors*.

The estimation of SSAP 65 Test 2 or Test 3 produces as an intermediate step an average set of indicated UPR factors for each subdivision of the data. If the subdivision is perfectly homogeneous, the indicated average UPR factors will also be appropriate for each contract in the subdivision. But subdivisions large enough to be credible are seldom perfectly homogeneous, and the indicated average UPR factors may not be quite right for any given contract.

To get factors for individual contracts in the data, or for proposed contracts different from any of those in the data, we describe below a technique which we call "All-Terms Factors": this is based on an exposure definition similar to that of Kerper and Bowron [4], together with an algorithm for using the experience of all contracts in one set of data to obtain a UPR curve for any contract whether in the data or proposed.

2.8 Comparison of Carried and Indicated UPR

At any given valuation date the indicated UPR is likely to differ from the carried UPR; if this difference is large enough the actuary will want to determine its cause and possibly revise the strings or formula used to establish the carried UPR. However, the difference between carried and indicated UPR at a single date gives little information as to what changes are needed, and moreover is highly dependent on the maturity distribution of the contracts that happen to be present in the data.

For example, if the carried UPR is less than the indicated, it may be because the contracts are clustered at the early lags and the early carried UPR factors are inadequate, or because the contracts are clustered at later lags and the later carried UPR factors are inadequate, or because the maturities of the contracts are evenly spread and the entire set of carried UPR factors is inadequate.

We have found it more useful to compare entire average carried UPR curve with entire indicated UPR curves, as shown in Figure 5 below. If the two sets of factors are not close to each other, we can conclude that the carried strings for some or all of the contracts need to be revised. If the carried and indicated factors are close, then we have evidence, though not proof, that the carried strings are indeed satisfactory contract by contract.

Our comparison procedure also generates single statistics that quantify the adequacy of the carried UPR factors independently of the maturity distribution of the data at hand.

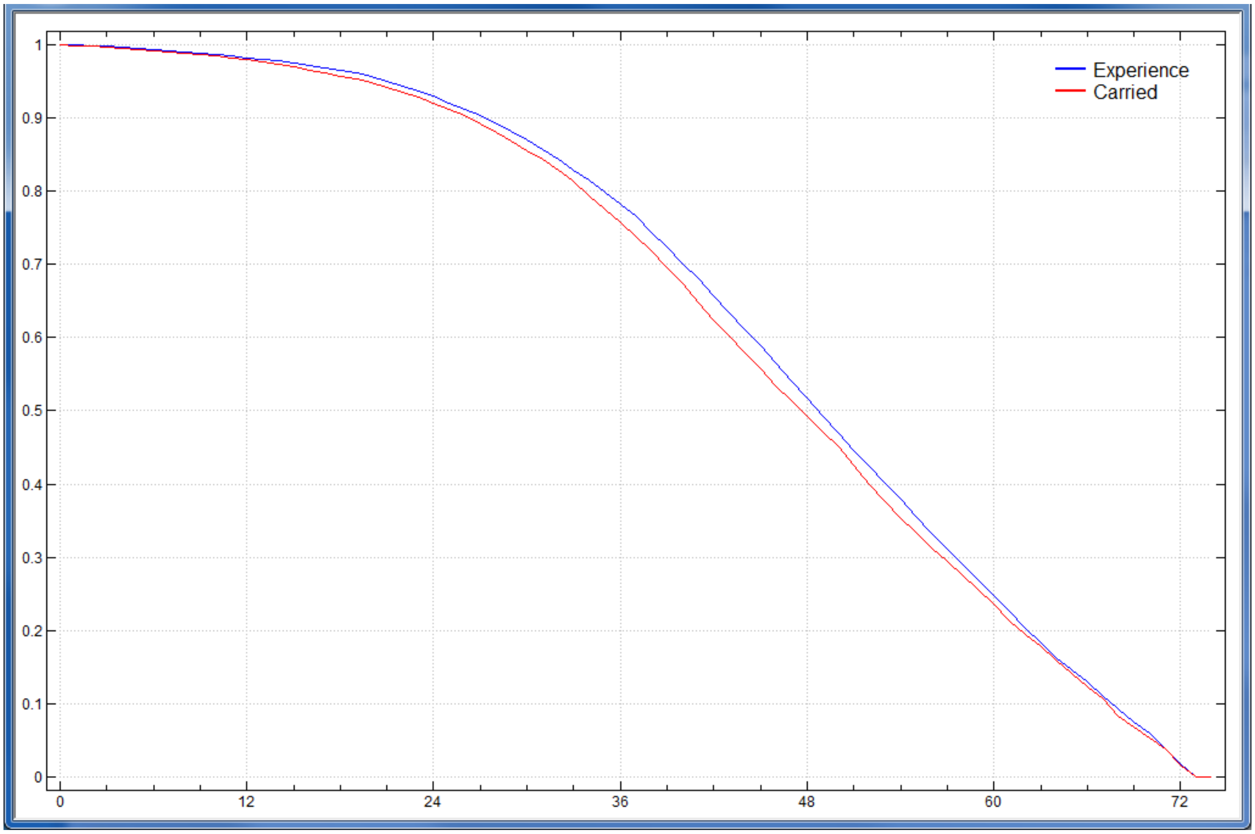


Fig. 5. Comparison of carried UPR factors with Test 2 UPR factors indicated by experience

The comparison in Figure 5 is based on about 272,000 contracts, and 55,000 claims, with nominal term 72 months, from a program insured by a large Warranty underwriter. Like all UPR factor curves, these start with the value 1 at issue and end with the value 0 when all contracts are fully earned, in this case by 73 months. We may read cumulative earnings as the complement of the UPR curve, and the instantaneous earnings rate as proportional to its absolute slope.

The shoulder in both curves reflects the presence of manufacturers' warranties, typically 36 months or longer. The fact that there are some earnings in the first 36 months reflects "extras" covered by the extended warranty above the services provided by the factory warranty, and also the fact that some drivers "mile out" of the manufacturer's warranty before 36 months. There may be an issue of heterogeneity here also, with a few contracts on used cars with little or no manufacturers' warranty remaining.

The Experience curve lies above the Carried curve showing that the latter is slightly inadequate. To assign a measure to this inadequacy we use what we call a "steady state conservatism factor", which is the ratio of the carried to the indicated Test 2 UPR if policies had been written at a constant rate long enough (in this case six years) so that their maturity distribution would thereafter be stationary. In the absence of cancellations the steady-state UPR factor distribution would be precisely that shown by this graph, the steady-state carried and indicated UPR's would be proportional to the areas under the respective curves, and the ratio of carried to indicated would be the ratio of these areas. In the present example the conservatism factor calculated in this manner is 0.980, reflecting about 2% inadequacy. We can also calculate an alternate conservatism factor in which the steady state allows for cancellations; in this case it comes out 0.982, suggesting 1.8% inadequacy.

Bear in mind that this entire comparison is on a Test 2 basis. It is unaffected by loss ratios and says essentially nothing about Test 3, the relationship between the carried UPR and expected unincurred losses. In this example the loss ratio happened to be well below 100% and Test 3 was easily satisfied, but we have no way of knowing, from these Test 2 curves alone, anything about the loss ratio or the adequacy of the carried UPR relative to Test 3.

2.9 Effect of UPR Curves on Estimated Loss Ratios

It is very important to understand the effect on *estimated* loss ratios, especially for immature blocks of business, of carried UPR curves that do not match closely with experience. Using Figure 5 as a model, imagine that there had been a large spike in sales 24 months ago, dominating the contracts in the data set. Then this business would now be at lag 24 months and would have a carried UPR factor of about 0.9196 and an indicated factor of about 0.9298. But then the ratio of cumulative earned premiums would be 0.0804/0.0702, or about 1.145; if the apparent loss ratio were, say, 95% the true loss ratio would be about 108.8%.

If in this case we took the apparent inception-to-date loss ratio as a predictor of the ultimate loss ratio, we might imagine that we were extrapolating the immature business by a factor of 3, from 24 months to 72. In fact we would be extrapolating by a factor of more than 14, from 7% of ultimate losses to 100%. The estimated loss ratio is not only affected by any errors in the UPR curve, also but by random fluctuations in the early incurred losses, greatly leveraged.

In this case we assumed that, although the bulk of the business was clustered at maturity 24 months, there was enough other business to estimate the shape of the earnings curve reliably. The situation is exacerbated when all of the business is immature, for example, when writing started 24 months ago, for then we must rely on tail factors, and the average extrapolation might be from around 12 months (assuming 24 months of uniform writings), or about 1.8% emergence, to 100%, a factor of more than 56.

2.10 Trend and Seasonality

One kind of trend operates purely in the issue-month direction, as changes in product design affect claim frequency and severity in sometimes irregular ways; a second kind operates purely in the calendar-month direction that is along both the issue-month and the development-month axes. This is usually the result of inflation, affects both parts and labor, and drives severity generally upward while having little effect on frequency. There is actually a third kind of trend, purely in the development direction; this we usually regard as simply part of the earnings or UPR pattern, but if not controlled for it may affect our estimates of calendar trend.

Loss emergence patterns are affected by calendar trend, in that increasing severity in the development direction tends to defer the emergence of losses relative to the total. This has only a modest effect on SSAP 65 Test 2 but a much greater potential effect on Test 3. To illustrate, Figure 6 shows a 48-month pro rata UPR curve after a manufacturer's warranty of 12 months, with trends of 0%, 5%, 10%, and 15% per annum. The steady-state SSAP 65 Test 2 UPR's are greater than the no-trend case by about 2.2%, 4.2%, and 6.3%, respectively. Note that the fact that the trend extends through the manufacturer's warranty has no impact on these figures.

By contrast, the Test 3 expected unincurred losses, valued at inception of the contract, would increase by factors of about 15.5%, 32.8%, and 52.3%, respectively, and part of these increases would be the result of trend during the manufacturer's warranty. Moreover, using a UPR curve with incorrect trend built in would have the same amplified effect on loss ratios estimated from immature data as was described the indicated-versus-carried comparison above.

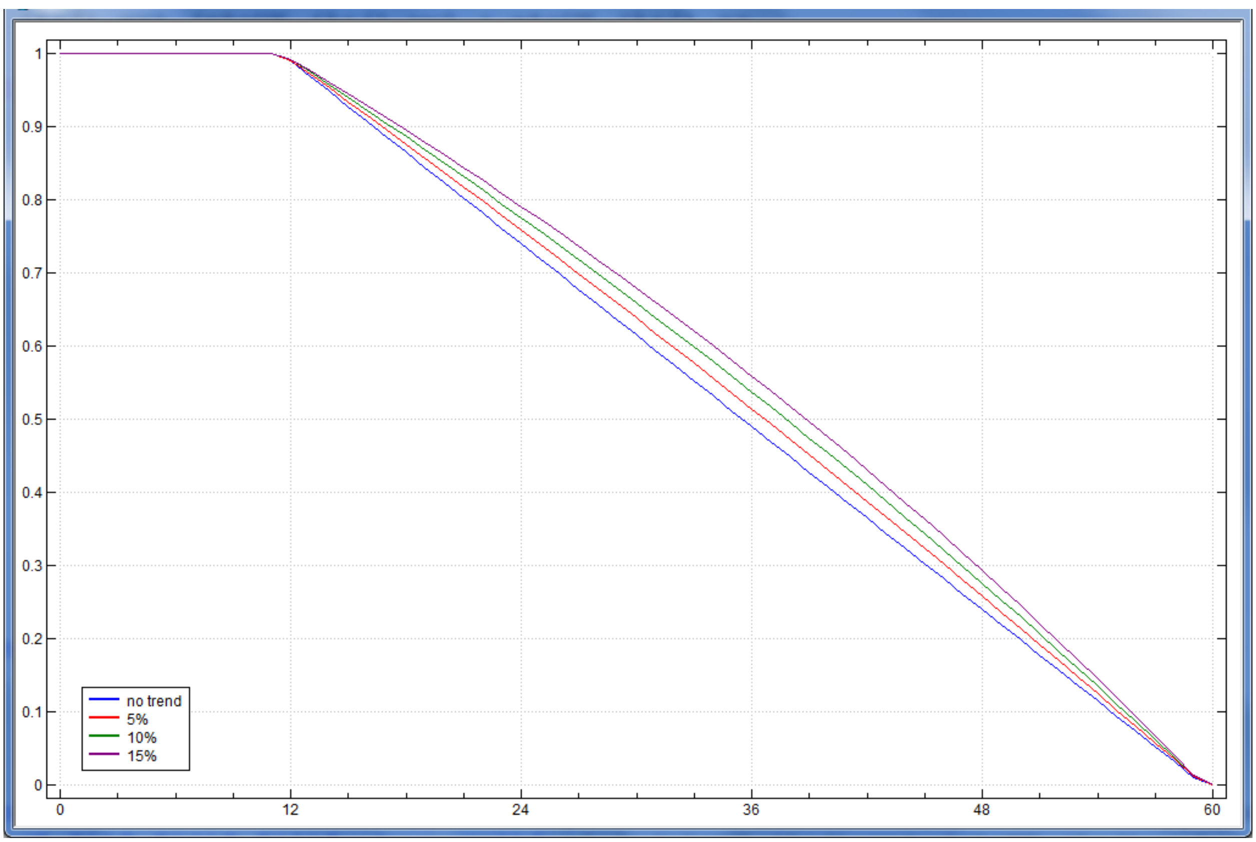


Fig. 6. Effect of 5%, 10% and 15% annual trend on pro rata UPR factors after MW

Carried UPR's by simple formulas like pro rata and Rule of 78 clearly do not anticipate trend, except possibly implicitly, as an offset to "breakage" (failure of eligible policyholders to present claims, which can be modeled as a negative trend). On the other hand, strings obtained empirically by analysis of experience are likely to have embedded in them whatever trend was present in that experience. If it is thought desirable to include a different trend in future strings, the experience can be deflated prior to estimating the earnings pattern, and the results re-inflated afterward. Or the strings from the usual analysis may be rendered more conservative by converting the UPR factors to earnings factors, inflating these by the difference between future and past trends, and converting once again to UPR factors.

Unlike trend, seasonality (which we take to have a period of 12 months) has rather a greater effect on Test 2 than on Test 3. We illustrate seasonality with an extreme, but nevertheless real, example: extended warranties on snowmobiles. In the Northern Hemisphere these may be expected to produce nearly all warranty claims during the season from November through, say, April, with the greatest concentration shortly after the equipment is first put in use for the season, say during December and January.

This pattern has little impact on expected ultimate losses, since most contracts are sold for a multiple of 12 months and thus include an equal number of Decembers and Januaries no matter in what month they are issued. On the other hand, if it is not provided for in the UPR factors, seasonality may lead to serious error in the estimation or projection of loss ratios. This problem may be avoided in two ways. One is to deseasonalize the loss data, estimate a constant loss ratio, then restore seasonality to the projections; the other is to build seasonality into the UPR strings and assign a different string to each contract depending on the month in which it is sold. Figure 7 illustrates the strings that might be assigned to two snowmobiles, one issued in November and one in March, under one plausible set of seasonality assumptions. Note that it is the slopes that are cyclical.

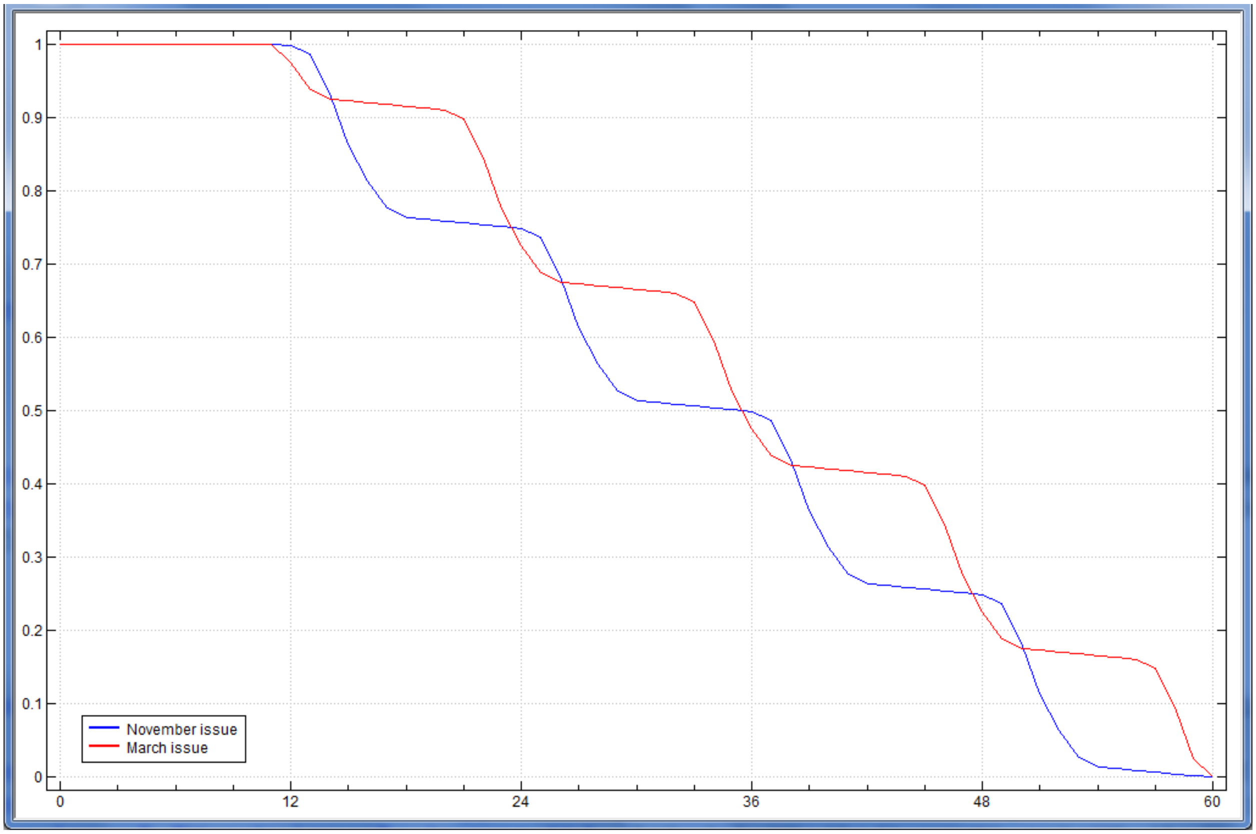


Fig. 7. Seasonalized UPR curves for snowmobiles, issued in November versus March

Implementing seasonalized UPR strings, even via a lookup-table formula, would appear to multiply storage requirements by 12; however, it may be done by factoring the strings into an underlying non-seasonal component and a single set of twelve seasonality factors. Estimating seasonality factors from the data may be done by first estimating the non-seasonal component with a 13-month moving weighted average and removing it from the raw data. If desired the residuals may be credibility-weighted against an a-priori set of seasonality factors and/or smoothed by application of a "circular" variation of Whittaker-Henderson graduation. One should to be cautious about smoothing, however, since some monthly spikes are real, such as those motivated by the timing of annual sales or service promotions.

2.11 Loss Development Factors Excluding the Effect of Cancellations.

An important difference between loss development for Warranty Insurance and for conventional lines is that for SSAP 65 Test 2 we need the lag factors that would be observed if there were no cancellations. In effect we need to decompose the total emergence pattern of losses into a component due to cancellations, which for UPR purposes we discard, and a component due to loss emergence proper, which becomes the basis for Test 2. This section describes one approach to this problem.

Assume we have a triangle of exposures E_{ij} that reflects cancellations, so that the E_{ij}'s may decrease, but not increase, along each the row for each issue month *i*. By exposures we mean such things as in-force premium, loss fund (i.e., provision in the premium for losses), or contracts. E_{i0} is the exposure at the moment of writing; E_{i1} the exposure at the beginning of the next month, and so forth. Also assume that we have a triangle L_{ij} of incremental losses. Here L_{i0} are the losses in the month of writing, L_{i1} in the next month, etc. Thus the exposures E_{ij} give rise to the losses L_{ij}. To keep things simple we assume that a full month of exposure is earned by each contract in the month of cancellation. In practice we may refine the model slightly by adjusting the E_{ij}'s so as to recognize only partial exposure (usually one-half) in the month of cancellation.

We can convert L_{ij} to a cumulative triangle with cells $C_{ij} = L_{i0} + L_{i1} + \ldots + L_{ij}$, and calculate development factors from cell (i,j) to cell $(i,j+1)$ in the usual way as $F_{ij} = C_{i(j+1)} / C_{ij} = 1 + L_{i(j+1)} / C_{ij}$. These factors will reflect both ordinary loss development and cancellations. This is fine for some purposes, including obtaining a rough estimate of Test 3. But for other purposes, such as evaluating UPR curves, we need an efficient way to measure just the loss development for reasons *other* than cancellation, and these simple cumulative factors will not suffice.

A good start would be to adjust the $C_{i(j+1)}$'s and C_{ij}'s to the same exposure level. It is tempting to try to adjust C_{ij} from its reduced exposure level E_{ij} to the original exposure E_{i0} before any cancellations. We can't do this directly, since C_{ij} corresponds not just to E_{ij} but to the whole history of exposures $E_{i0}, E_{i1}, \ldots, E_{ij}$, and these may vary independently. The only way to adjust C_{ij} is to adjust each of its incremental pieces $L_{i0}, L_{i1}, \ldots, L_{ij}$ and recombine them. The adjusted L_{ik} may be written as $L^*_{ik} = L_{ik}E_{i0}/E_{ik}$. Accumulating the L^*_{ik}'s to get $C^*_{i(j+1)}$ and C^*_{ij} and taking their quotient we get a development factor F^*_{ij} with the effect of cancellations removed.

But even this is not quite what we need. We are not usually interested in the triangle of cell by cell development factors, but in some sort of average down each column, to be used as an estimate of the underlying population development factors. For such an estimate to be efficient, we should give greater weight to those cells with greater losses or with greater exposure remaining after all earlier cancellations. Otherwise the larger variation in the smaller cells will contribute disproportionately to the variability of the average development factors.

The most common weighting for conventional loss development is by the losses in the denominator, which makes the average development factor in each column equal to the sum of the numerators divided by the sum of the denominators. (Other weightings involving powers of the losses in the denominator are appropriate for particular models of the loss process. Here we are concerned only with a preliminary weighting to correct for changes in exposure.) The conventional weighting works nicely when the denominators are the C_{ij}'s, but not when they are the C^*_{ij}'s, because the adjustment of L_{ij} *to* L^*_{ij} flattens out differences in volume of losses by issue month that result from different numbers of cancellations, and the weighting doesn't accomplish much.

A way around this problem is to put each L_{ik} on the same exposure basis as $L_{i(j+1)}$ (that is $E_{i(j+1)}$) by letting $L'_{ik} = L_{ik}E_{i(j+1)} / E_{ik}$. The adjusted cumulative losses become $C'_{ik} = L'_{i0} + L'_{i1} + \ldots + L'_{ik}$, and the adjusted development factor becomes $F'_{ij} = C'_{i(j+1)} / C'_{ij}$. This is exactly the same as F^*_{ij}, but this time the weighted average works in the natural way:

$$\Sigma_i C'_{ij} F'_{ij} / \Sigma_i C'_{ij} = \Sigma_i C'_{i(j+1)} / \Sigma_i C'_{ij} = (\Sigma\text{numerators})/(\Sigma\text{denominators})$$

This looks as if it would be cumbersome to calculate. We have to loop through the j's and for each one adjust all the earlier L_{ik}'s and sum them to get the C'_{ij}'s. But notice that

$$C'_{ij} = \Sigma_k(L_{ik}E_{i(j+1)}/E_{ik}) = E_{i(j+1)}\Sigma_k(L_{ik}/E_{ik}) = E_{i(j+1)}\Sigma_k P_{ik} = E_{i(j+1)}Q_{ij},$$

where P_{ik} = partial loss ratio and Q_{ik} = cumulative partial loss ratio. Similarly,

$$C'_{i(j+1)} = E_{i(j+1)}Q_{i(j+1)},$$

so that

$$F'_{ij} = C'_{i(j+1)} / C'_{ij} = Q_{i(j+1)} / Q_{ij}$$

No adjustment here, just developing the triangle of cumulative partial loss ratios, but we get the same factors!

We do not want to use an unweighted average of these factors since all differences in exposure, whether due to cancellations or otherwise, have been flattened out of them. Two approaches come to mind that make the weights reflect volume of experience.

If we weight the ratios $Q_{i(j+1)}/Q_{ij}$ by the $E_{i(j+1)}$'s, we get the following average:

$$(\text{Avg } F)_j = \Sigma_i(Q_{i(j+1)}/Q_{ij})E_{i(j+1)})/\Sigma_i E_{i(j+1)},$$

which equals (Σnumerators)/(Σdenominators) if we write each development factor as

$$F'_{ij} = (Q_{i(j+1)}/Q_{ij})E_{i(j+1)}) / E_{i(j+1)}$$

On the other hand, if we weight the ratios $Q_{i(j+1)}/Q_{ij}$ by the adjusted losses C'_{ij}, we get

$$(\text{Avg } F)_j = \Sigma_i(Q_{i(j+1)}/Q_{ij})(C'_{ij} / \Sigma_i C'_{ij}) = \Sigma_i Q_{i(j+1)}E_{i(j+1)}/\Sigma_i Q_{ij}E_{i(j+1)},$$

once again a "sum of numerators over sum of denominators" situation if we write each development factor as

$$F'_{ij} = Q_{i(j+1)}/Q_{ij} = Q_{i(j+1)}E_{i(j+1)}/Q_{ij}E_{i(j+1)}$$

Both of these formulas for $(\text{Avg } F)_j$ are easy to implement in software since we already have the exposure triangle. The first average is weighted by exposures, the second is weighted by losses. They are not identical, but they are consistent with each other because expected losses are proportional to exposures.

A simulation study involving extreme exposure patterns (some issue months nearly fully cancelled, others with few cancellations) shows that (a) the weighting by exposures appears to be more stable than the weighting by losses, (b) the weighting by losses at the exposure level of the numerator is practically indistinguishable from the weighting by losses at the initial written exposure level, (c) all of these variations of the loss development method are more stable than the Partial Loss Ratio (Stanard's "Additive") approach, (d) there is no evidence of bias in any of the methods, and (e) the differences among methods are small and would probably be immaterial with real-world data.

2.12 Calculations Involving Issue Date versus Breakdown Lag Triangles

In principle, the UPR covers the interval between issue and breakdown, the IBNR reserve covers the interval between breakdown and reporting, and the RBNP reserve covers the interval between reporting and payment.

We can estimate different components of the total reserve from different triangles, as follows:

	Cell contents	**Time 1**	**Time 2**	**Resulting reserve**
A	Paid	Issue	Payment	UPR + IBNR + RBNP
B	Paid	Breakdown	Payment	IBNR + RBNP (i.e., IBNP)
C	Paid	Report	Payment	RBNP
D	Paid + Pending	Issue	Report	UPR + IBNR
E	Paid + Pending	Breakdown	Report	IBNR
F	Paid	Issue	Breakdown	UPR
G	Paid + Pending	Issue	Breakdown	UPR

(A,D,F,G) UPR here means expected losses not yet incurred, or SSAP 65 Test 3, roughly equal to the usual Test 2 or gross UPR times an expected loss ratio.
(D,E) Relies on the pending reserve's being exactly equal to the eventual payment, in which case the paid-plus-pending data is equivalent to a case-basis incurred triangle.
(F) Unreliable without adjustment because some losses in the last few diagonals may not have been paid by the ending date of the data
(G) Unreliable without adjustment because some losses in the last few diagonals may not have been reported by the ending date of the data

We may use the technique described in 2.11 to handle triangles of types A and D, and ordinary loss development to handle B, C, and E. But triangles of types F and G require special treatment.

The problem is that the triangle of losses by issue month versus breakdown lag is not stable, in the sense that the next month's experience will not simply add a new diagonal and leave the existing triangle unchanged. Instead, the next month's experience may contain newly-reported losses that belong in interior cells of the existing triangle, usually somewhere in the last few diagonals. If we are not careful, developing an issue-versus-breakdown triangle may understate the true reserve because of these incompletely-paid or incompletely-reported losses in the recent diagonals. Note that case F (paid losses) and case G (reported losses) are structurally identical, except that F, in effect, treats each claim as if its report date were equal to its payment date.

One solution to this problem is to eliminate the latest diagonals from the triangle and obtain development and earnings factors from the curtailed triangle. This works, but it sacrifices information, especially for situations where report or payment lags stretch for more than three or four months.

A second approach is to use the breakdown-date versus payment-lag or report-lag triangle to obtain payment or report lag factors, cumulate these to obtain the expected fraction of losses paid or reported at the end of one, two, … months, apply these factors to the triangle of exposures (working in on each row from the latest diagonal) to obtain the "expected paid" or "expected reported" fraction of exposure, and calculate earnings factors from the original losses and these adjusted exposures. This is the most robust approach and makes full use of the latest information.

Still another approach is to use a single issue-date versus breakdown-lag triangle to estimate simultaneously the earnings factors (in the breakdown-date direction, a "column effect") and the fraction unpaid or unreported (in the calendar-month direction, a "diagonal effect"). From the latter we may derive lag factors for losses IBNP or IBNR, *even when we do not have payment or report dates of individual losses or any triangle by breakdown date versus payment or report lag.* Estimation proceeds iteratively, alternately using exposure-adjusted loss development as described in section 2.11 to estimate the earnings factors, then using the ratio of actual to fitted losses in the last few diagonals to estimate the cumulative fractions of losses reported at each lag, and then using these fractions to adjust the exposure in those diagonals for the next iteration.

This "one-triangle" approach is occasionally needed, produces good results with enough data, and should be part of the software tool kit of any Warranty insurer. But, whenever possible, it is better to use the breakdown-month versus report-lag triangle to estimate the report lag factors, since this makes use of all the claims in the data, not just those claims that happen to have been incurred within a few months of the valuation date. For example, in the case of 120 months' steady-state data with constant loss emergence, if all claims are settled within three months so there are just two independent payment lag factors to estimate, only the latest two diagonals of the issue-to-breakdown lag triangle will yield information relevant to estimating the "diagonal effect" of unreported losses, and these diagonals will contain only about 3.3% of the total claims that would be included in the full breakdown-to-report (or breakdown-to-payment) lag triangle.

2.13 Tail Factors

It might be thought that tails would be less important when estimating UPR factors for Warranty insurance than when projecting accident-year runoff for other lines, because the duration of Warranty runoff is limited by the term of the contracts. However, when we calculate tail factors we are not so much trying to answer "how much will be paid each month in the future" as "how much remains to be paid relative to what has already been paid". The results of an analysis of immature data are sensitive to this question no matter how long the tail may be in months. The only advantage of Warranty insurance over other lines in this respect is that the terms limit the number of distinct issue months for which tails need to be calculated.

The need for tail factors only arises when estimating indicated UPR factors, since the tail is already incorporated in strings or formulas used for carried UPR factors. The strings and formulas are usually known at the time of the analysis, so it is tempting to consider the problem solved: just append a tail with the same shape as the average tail of the strings or formulas. However, this does not answer the question of how large the entire tail should be relative to what came before.

To illustrate, suppose we have observed losses for 27 months in a new program with term 72 months and carried reserve pro rata. The losses have emerged irregularly but moderately increasing with lag from issue. If we assume that losses in the tail after 27 months will be constant, producing a pro rata UPR curve, we know that the shape of the tail will be a straight line, but we don't know how steep that line will be. Figure 8 shows three possibilities.

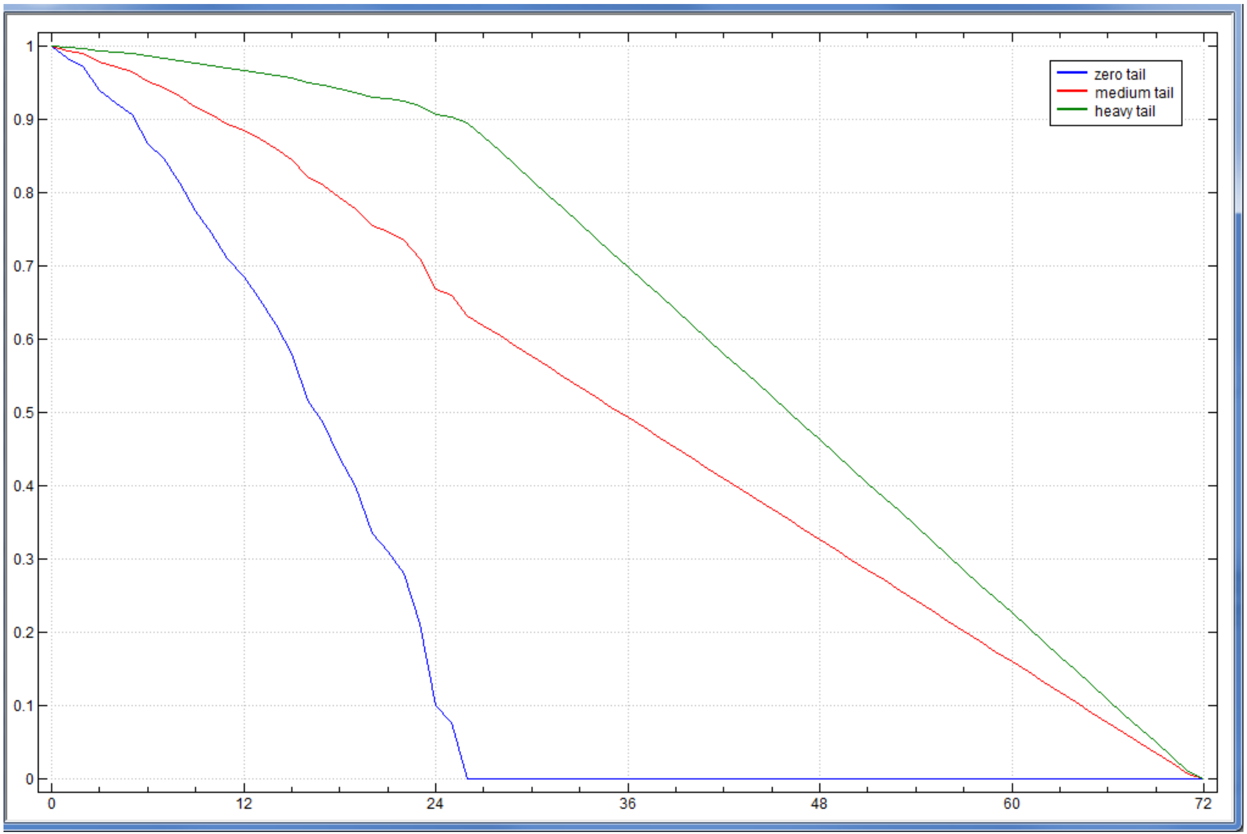


Fig. 8. Possible pro rata tails starting from the same observed losses through 27 months.

The "medium tail" in Figure 8 looks more reasonable than the other two. The key is that its earnings rate was derived from the known data; in fact it equals the mean rate of loss emergence over the entire 27 months of known data.

It might be objected that the slope of the tail thus selected is too gentle, since the known part of the curve exhibits a steepening in the last few months. One way to address this problem is to shorten the "lookback period" for averaging earnings to something shorter than all the known months. Another is to apply linear regression to the known earnings factors and either extrapolate them to 72 months or (as we prefer for stability) letting the earnings rate through the tail be constant at the fitted value for 27 months. A third way is to make the tail steeper using a factor selected by judgment. These possibilities are illustrated in Figure 9, using, respectively, the latest 6-month mean of the observed earnings, regression to month 27 of all 27 observed earnings, and a doubling of the earnings rate from the original 27-month mean. This does not double the slope of the tail, but doubles the ratio of the earnings in the tail to the earnings before the tail, since all of the UPR curves are based on earnings normalized to total 1.00.

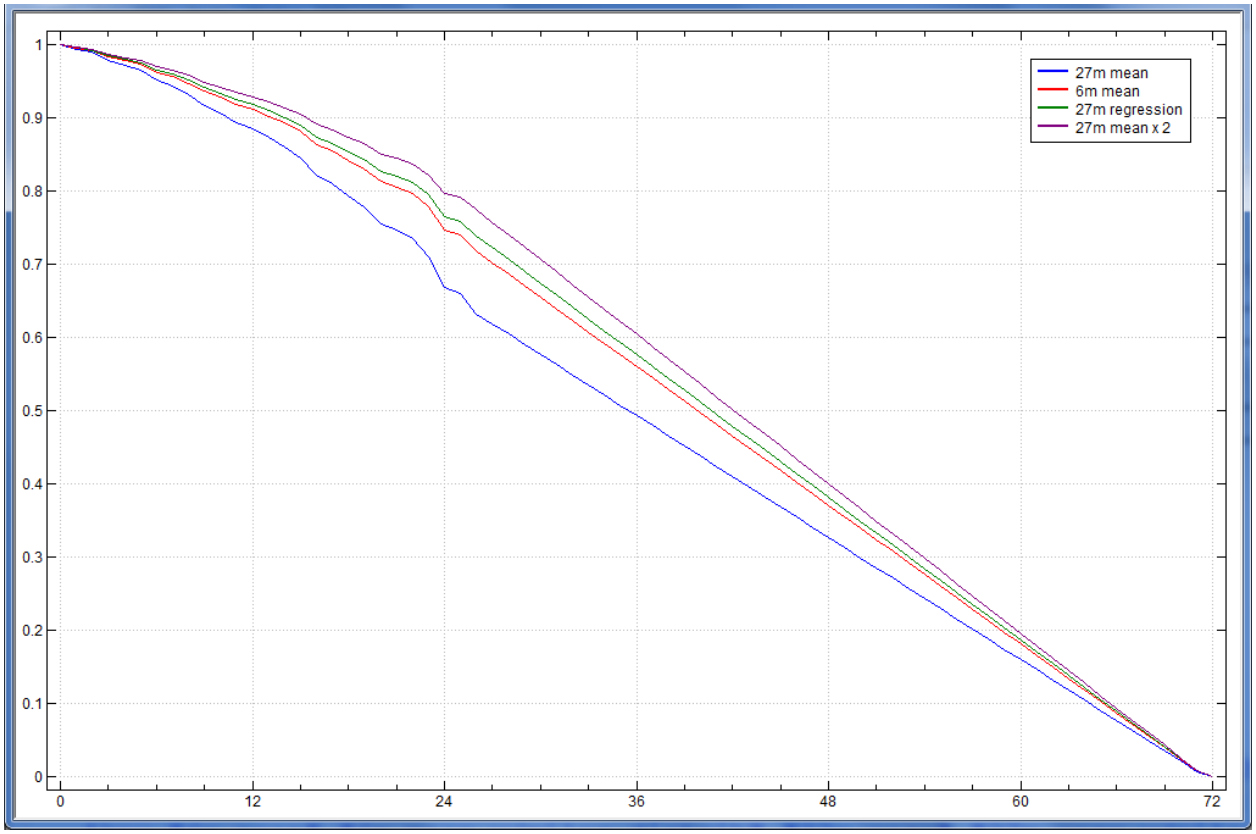


Fig. 9. Different determinations of initial tail slope

It is not necessary that the tail start with the first lag following the known data; in cases where the last few known lags are irregular, the tail may be made to start a few months earlier.

Nor is it necessary that the tail be linear. When dealing with products for which the natural UPR curve earns more rapidly at first than later, we may fit Rule of 78 tails; when breakage is apparent we may fit a curve with exponentially declining first differences; when "miling out" is a factor we may fit a curve that is a weighted average of tails for the various tail lengths corresponding to, say, a 50-point discrete distribution of miles driven per annum. The latter adjustment makes the UPR curves less conservative, since the miling-out of some contracts reduces the earnings in the later lags; it is illustrated in Figure 10. The two lower curves actually fit linear tails just like the two top curves, except using a weighted mixture of term lengths.

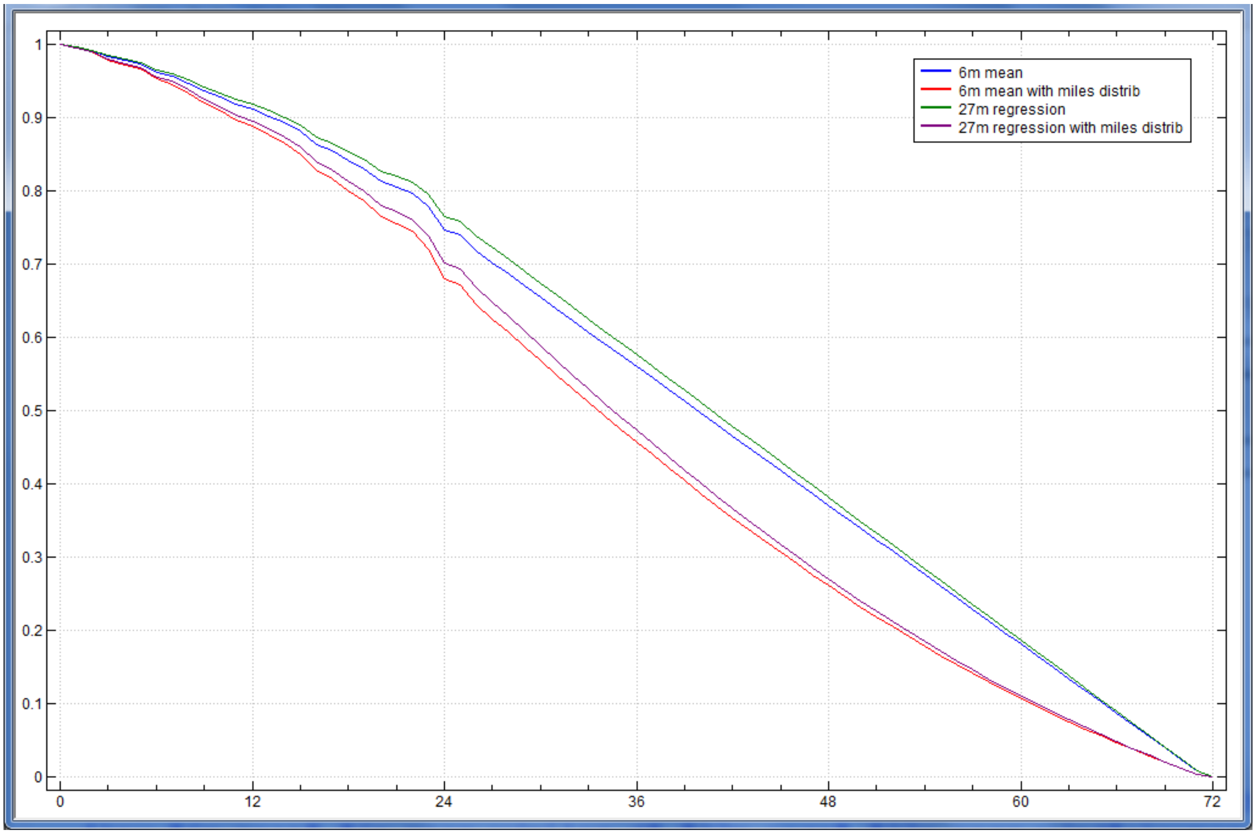


Fig. 10. Effect of using miles-per-annum distribution to create a mixture of tail terms.

In practical work the actuary may need rules for calculating tails, when necessary, for multiple programs and/or terms being analyzed in a batch, without halting execution for insertion of judgment parameters. Fitting tails for Warranty contracts is as much an art as a science (just as is the case for other lines) and cannot eliminate the uncertainty inherent in any extrapolation. But it can be systematized. Our preference is to divide the problem into three parts, determining for each tail (1) its starting point, (2) its shape, and (3) its weight.

When batch-processing multiple programs, we have generally found it satisfactory to start the tails a month or two earlier than the last available lag, to make the tail pro rata for each group of contracts expiring at the same time, using an estimated miles-per-annum distribution, and to average the earnings over a lookback period of about one quarter of the observations to determine the initial slope of the tail. When fine-tuning a single projection, we often prefer to use an initial earnings rate determined by regression, in order to join the tail smoothly to the known data. This requires inspection and possible judgment intervention, however, lest the fitted value at the last lag before the tail become negative or otherwise unreasonable.

Possible tail shapes are not limited to pro rata and pro rata summed across remaining terms of different lengths. For some lines of business, such as Tire & Wheel, we may let the earnings pattern for each remaining term decline linearly to zero, in a sum-of-digits fashion. And it is possible to let the *carried* UPR for each contract in a block of business, summed across contracts to the end of each term, determine the shape of the tail when otherwise estimating the UPR from experience and ignoring the carried UPR factors.

2.14 SSAP 65 Test 3 Estimates

For SSAP 65 Test 3 we need to convert the cancellation survivorship patterns and loss emergence patterns observed in estimating Test 2, together with incurred losses, into estimates of losses for existing contracts unincurred at the accounting date.

While the Chain-Ladder estimator, with the adjustments described in 2.11, is well adapted to estimating the Test 2 UPR pattern (and therefore the Test 2 UPR) of a block of Warranty business, it is not usually satisfactory for the Test 3 UPR, expected unincurred losses. The reason is that many extended Warranty contracts earn very little during their early months, causing projection of expected losses for the latest issue months to be impossible or erratic.

On the other hand, we can adapt the Bornhuetter-Ferguson method (BF) to produce stable projections of the Test 3 UPR, to accommodate known trends or seasonality in the calendar-month direction, and to be responsive to unknown or variable trends in the issue-month direction.

The adaptations involve (a) using Hans Bühlmann's Cape Cod estimator [2], adjusted for declining exposures, to obtain an expected loss ratio (ELR), (b) adjusting this procedure as necessary using Spencer Gluck's decay factors, to increase responsiveness to changes in the ELR by issue month, and (c) detrending and/or deseasonalizing the data entering the calculations and restoring trend and/or seasonality to the results.

2.14.1 Cape Cod

In the traditional calendar-accident year context, the Cape Cod estimate of the ELR equals losses reported (or paid) to date divided by the portion of premium (or other measure of exposure) expected to have emerged as losses over the same time period. The numerator may simply be summed over the entire incremental loss triangle, i.e., down the latest diagonal of the cumulative loss triangle. The denominator is usually summed across accident periods, with each accident period contributing its initial earned premium multiplied by the reciprocal of its cumulative development factor.

In the Warranty context, with issue months instead of accident periods, with in-force premium that declines across development months, and with recent diagonals deficient because of unreported losses, the denominator of the ELR must be adjusted. It is now summed across all cells of the triangle, with each cell contributing its initial in-force premium multiplied by the applicable incremental lag factor and, if we are working with an issue-month versus breakdown lag loss triangle, by the applicable factor for the expected fraction reported through the valuation date. No adjustment is necessary to the numerator; it remains just the sum of paid losses across all cells of the incremental loss triangle.

2.14.2 Gluck factors

In a 1997 PCAS article [3], Spencer Gluck proposed an enhancement of the Cape Cod estimator that puts BF/Cape Cod at one end of a spectrum with the pure Chain-Ladder at the other. Gluck's contribution may be thought of as the third in a series of simple, non-stochastic, but eminently practical approaches to loss reserving, the other two being the Bornhuetter-Ferguson method itself [1] and the Cape Cod estimator of the expected loss ratio (1983). We have found Gluck's procedure to be particularly useful for Warranty insurance, where product redesigns and technical changes result in irregular trends in the issue-month direction.

Gluck's approach produces a separate ELR for each issue month. Notice that the Cape Cod estimator, the sum of the losses divided by the sum of the adjusted exposures, is actually an average of the loss *ratios* in each cell, weighted by adjusted exposures. Gluck multiplies these weights by a "Gluck weight" dependent on the distance between each issue month and the target month via a geometric decay factor g. The scale is immaterial so we may assume that the target month receives Gluck weight 1. Then the adjacent months receive Gluck weight g, the next further months receive Gluck weight g^2, the next further months g^3, and so forth. For each target month, the ELR equals the sum of g times losses divided by the sum of g times adjusted exposure.

If the decay factor g equals 1, this gives the pure BF/Cape Cod; if g equals 0, it gives the pure Chain-Ladder; as g moves from 1 toward 0, it gives an ELR that becomes increasingly responsive to local trends.

2.14.3 Trending and seasonality

The procedures to account for trend and seasonality apply to both Test 2 and Test 3, but these considerations have a much greater impact on Test 3 than on Test 2. Our preferred approach is to deflate and deseasonalize the losses in the source data, based on breakdown date, and, to the extent possible, recalculate premiums in the source data, usually putting both losses and premiums on the level of the valuation date. If we have correctly measured the trend and seasonality to remove, the data becomes stationary and yields UPR factors, ELR's, calendar-month projected losses, and Test 3 UPR's with no embedded trend or seasonality. We then invert the process, restoring the known losses to their historical values and trending and seasonalizing the future losses, producing a trended and seasonalized estimate of Test 3. The future trend and seasonality factors need not be identical to the historical ones.

This whole procedure depends on estimates of historical trend and seasonality. In principle there could be one cumulative trend factor – or even one cumulative seasonality and trend factor – for each historical month. In practice seasonality is usually expressed as a set of 12 factors, one for each calendar month, and trend is usually expressed via a single annual trend factor, presumably applicable to the entire time period from the earliest included month through the valuation date.

Estimating trend and seasonality is complicated by the fact that many quantities we would like to measure can only properly be compared with earned premium or earned contract counts, running the risk that errors in earnings pattern may be confused with trends. These quantities, for which measured trends are suspect, include frequency, loss costs, and loss ratios. On the other hand severity is not dependent on earnings and calendar-month severity trend may be estimated fairly easily and reliably and used as a proxy for total trend. We usually calculate average severity by breakdown month, fit both linear and exponential curves to it, and inspect the plots for evidence of discontinuities which might justify separate trends for different time periods. As for seasonality, we usually deseasonalize the paid severity data using a 13-month moving weighted average and estimate the seasonality by averaging the residuals over a multiple of whole years grouped in 12 columns by calendar month.

Because issue month, calendar month, and development lag are multicollinear we cannot obtain unique simultaneous estimates of pure trends all three of these directions, but with reasonable assumptions serving as constraints, we can use GLM's to estimate issue and development frequency trends and issue and calendar severity and loss-cost trends, and take the product of calendar trend for severity and the development trend for frequency as another estimate of total development trend to be used in deflating UPR curves.

2.14.4 Contractual Liability policies

In many states, service contracts issued by a retailer directly are not regulated as insurance and service contracts issued through an affiliated obligor are not regulated as insurance provided the obligor itself insures its obligations. For this reason Warranty insurers find themselves underwriting some Warranty programs from the ground up, others entirely conditional on the obligor's failure to perform (FTP), and still others so as to cover a share of the risk from the ground up and the remaining share on an FTP basis.

Coverage written from the ground up is assigned the Warranty line of business for annual statement purposes; coverage entirely on an FTP basis is considered Contractual Liability Insurance under the line of business Other Liability – Occurrence. The most reasonable analysis of coverage partially insured with the rest on an FTP basis is that part of the premium (often a small part) should be ascribed to Contractual Liability and the remainder to Warranty, all claims should be ascribed to Warranty unless and until the program is in FTP status, and thereafter the insured share of claims should be allocated to Warranty and the remaining claims to Contractual Liability.

Typically a Warranty insurer will impose financial requirements on any obligor insured on an FTP basis to protect its interest in that obligor's solvency, and the insurer may administer the entire program and its cash flow. With such protection in place the actual FTP premium may be only a small fraction of the amount that would be required on a first-dollar basis. For this reason, SSAP 65 Test 2 will never be more than a small fraction of SSAP 65 Test 2 applied to the obligor's underlying contracts – although the UPR factors should be somewhat greater than the factors for the underlying contracts because expected loss emergence is deferred by being conditional on future bankruptcy.

If the FTP is priced correctly (which on account of the catastrophe risk would justify a low expected loss ratio) then Test 3 should be even smaller than Test 2; but in the event the program enters FTP status, Test 3 rises immediately to the full Test 3 at the obligor level. In principle, the evaluation of Test 3 for a program not in FTP status requires estimation of the probability of first entering insolvency at each future month together with the expected unincurred losses at that time.

2.15 All-Terms Factors

Suppose we need strings for a set of target contracts differing from each other in term, manufacturer's warranty, odometer reading at issue, and so forth. Some or all of the target contracts may not be represented in the available data, or may be represented but in too small a volume to produce reliable indicated UPR factors. Or reasonably credible data may be available and each indicated string may appear reasonable in isolation but the strings for different contracts may not be related to each other in a logical way.

For example, Warranty contracts are usually marketed for a few distinct terms, such as 60 months or 60,000 miles (conventionally written 60060), 72 months or 90,000 miles (72090), and so forth. When these *nominal terms* are effective at issuance of the contract, the data may easily be broken into a few homogeneous term groupings. But when the nominal terms are effective on expiration of the manufacturer's warranty, there may be a great many *actual terms* from issue to expiration, depending on how much of the manufacturer's warranty remains at time of issue. In this case it is unlikely that the available data would contain enough contracts to estimate UPR curves for each separate actual term, not to mention also account for differences in remaining manufacturer's warranty.

Here we describe a technique that uses an entire body of data, with many terms, to derive a *family* of UPR curves, each curve reflecting the experience of *all* the terms in the data, weighted by proximity to the term being evaluated. We call the model that generates such a family of curves *All-Terms Factors,* or ATF. It is in fact a comprehensive model of the loss emergence process for contracts on automobiles, but it is also applicable in simplified form to other contracts.

The ATF model derives a theoretical loss emergence pattern based on our contract, the manufacturer's warranty (MW), trend, breakage (i.e., policyholders' failure to present eligible claims), and certain other factors, all conditional on the car's being driven a given number of miles per annum. It then integrates over an assumed distribution of miles per annum to obtain the expected loss emergence pattern for the usual case where the driver's average usage is not known in advance. This emergence pattern, when converted to a UPR curve, is by itself useful in reserving, and, as explained below, may be thought of as a broad generalization of the concept of earning pro rata from the end of the manufacturer's warranty. This part of our ATF model turns out to be a modest extension of Kerper and Bowron's exposure model [4], mentioned above. We explain it here in detail, using our own terminology and notation; it has certain enhancements but in its essentials is nearly identical to their model.

Our algorithm for applying the model derives the expected loss emergence pattern in units proportional to expected losses, before eventually normalizing it to total 1.00. Therefore in addition to the earnings pattern *per se*, it yields estimates of relativities between any pair of proposed contracts. These relativities, when taken to a common base contract, are very useful in ratemaking.

We may apply the model of loss emergence directly to a proposed contract, on the assumption that it accounts for all relevant factors. *Alternatively, we may fit the model to a body of data and obtain a family of curves, one for each term in months, measuring only the effects of any factors* not *included in the model.* We call these curves *residual ATFs*. We may then derive the UPR curve for any contract, whether part of the data base or not, by applying the contract's theoretical loss emergence pattern to the residual emergence pattern for the same term. The residual ATFs for any given term are derived not just from the experience of contracts with that term, but from *all* contracts in the data, weighted by proximity to the target term.

If the original model really does fully explain the variability of loss emergence patterns, the residual ATFs will all be straight lines and using them will make no difference. Otherwise, starting with the residual emergence pattern will improve the performance of the model on new contracts, provided the residuals were based on a large enough volume of relevant cxperience.

We can, if we wish, ignore some or all of the usual explanatory variables and let their effect (if any) appear in the residual ATFs. If we ignore all variables except term in months, the residual ATF's become a family of average UPR curves, one for each term. We call this the *simplified* ATF model. It is appropriate as an aid in reserving for whole term groupings in the same body of data from which the ATF's are derived, but it will not usually produce reliable UPR curves for individual contracts or for different bodies of data. For these purposes the general ATF model, with its explanatory variables, expected loss emergence pattern, and residual ATF's, is required.

The ATF model starts with the assumption that parts failures will occur whether or not a car is covered by a warranty, and that the number or cost of such failures may be subject to trend and may depend on the car's usage as well as on the passage of time. Then the model considers how parts failures translate into claims against our contract; this depends on the contract term in months and miles, the MW in months and miles, the relationship between services covered by our contract and by the MW, factors such as breakage and pre-expiration claims spikes, and the distribution of miles driven per annum. Finally the model provides procedures for estimating the residual UPR curves, and other parameters, from experience.

2.15.1 Emergence of failures

We assume that any car is subject to parts failures of two kinds:

a. Those that depend on use (e.g., power train)
b. Those that depend on time (e.g., paint and trim)

Here we consider all failures, whether ultimately paid for by the manufacturer, the insurer, or the customer. The cost and the earnings pattern of a contract to the insurer will eventually depend on how this responsibility is allocated, but it all starts from the initial failure of some part of the car. The failure emergence rate is therefore the starting point of the All-Terms Factors model.

Let s be distance driven in miles and t be time elapsed in months. Assume that failures of type (a) emerge at a rate λ_a per mile and failures of type (b) emerge at a rate λ_b per month. We may conceive of these rates as Poisson lambda parameters for failure frequency, or as means per unit distance or time of some distribution of costs.

For a car driven m miles per annum, $ds/dt = m/12$ and, if y = number or cost of failures in time t,

$$E(y) = ((m/12)\,\lambda_a + \lambda_b)t$$

or

$$E(y) = (\lambda_a + (12/m)\lambda_b)s$$

For our purposes we do not need the values of λ_a and λ_b in absolute terms so for simplicity we replace them with a single assumption: the fraction f_{mi} of failures that emerge in proportion to miles driven, with the complement $1\text{-}f_{mi}$ emerging in proportion to time elapsed. If we let m_{avg} = average miles driven per annum, we express this as

$$f_{mi} = (m_{avg}/12)\,\lambda_a \,/\, ((m_{avg}/12)\,\lambda_a + \lambda_b)$$

so that

$$\lambda_b = (m_{avg}/12)\,\lambda_a\,((1\,/\,f_{mi})\text{ -}1) = (m_{avg}/12)\,\lambda_a\,(1\text{ -}f_{mi})\,/\,f_{mi}$$

and

$$E(y) = (\lambda_a\,(1 + (m_{avg}/m)\,(1\text{ -}f_{mi})\,/\,f_{mi}))s$$
$$= (K\,(f_{mi} + (m_{avg}/m)\,(1\text{ -}f_{mi})))s$$

which implies

$$E(y) = (K(m/12)\ (f_{mi} + (m_{avg}/m)\ (1 - f_{mi})))t$$

$$= (K((m/12)\ f_{mi} + (m_{avg}/12)\ (1 - f_{mi})))t$$

where $K = \lambda_a / f_{mi}$ is a constant of proportionality.

The 12's appear in this formula to convert mileage from annual to monthly, because the conventional measure of auto usage is miles per annum, but our time t is measured in months. Notice that K does not depend on miles driven, but that the remaining (parenthesized) factor does, so that the expected failures of cars driven at different rates are proportional to this factor. However, K may be generalized to be a function of calendar month, months elapsed since the car was put in service, or attained odometer reading, so that the loss emergence pattern and/or the relative expected costs of two cars may reflect various trends. For convenience we implement these trends at the same time as we implement factors representing plan design, manufacturer's warranty, breakage, etc., but if you wish you may think of them as a distinct step modifying the emergence of failures.

2.15.2 Emergence of claims

The heart of our exposure model is how we adjust the failure emergence rate to convert it to a claims emergence rate, reflecting the conditions of our contract, the manufacturer's warranty, the miles-per-annum distribution, time trend, odometer trend, breakage, claims spikes, and "extras".

Because of the requirements of SSAP 65 Test 2, the emergence pattern of claims is a guide to the desired earnings pattern of premium, and we shall speak of earnings patterns and emergence patterns interchangeably. The ATF model generates these emergence patterns up to a constant of proportionality times total losses, comparable across plans, so we may obtain valid relativities. For the UPR pattern we normalize the emergence factors to total 1.00 and then sum stepwise backwards from the last elapsed month to the first.

In a sense our entire adjustment process is nothing more than an elaboration of the simple pro rata, or constant-earnings, model. It results in a step function for earnings, and a piecewise-linear function for UPR factors, but there may be so many steps that the UPR function is essentially indistinguishable from a smooth curve.

We illustrate the successive elaborations in turn.

Straight pro rata. Suppose the term of a contract is 60 months or 60,000 miles, with no remaining manufacturer's warranty, and that it is applied to a car driven 12,000 miles per annum. Then the contract will expire on account of both time and mileage at the end of its 60th month. Assume for now that the average number of miles driven per annum (across all cars) is 15,000 and that claims emerge 80% in proportion to miles and 20% in proportion to time. Then all emerging failures are covered by the contract, the earnings pattern is constant for 60 months, equal to $K\,((m/12)\,f_{mi} + (m_{avg}/12)\,(1 - f_{mi})) = 1050K$, and is zero thereafter, and the UPR pattern is pro rata starting from 1 at 0 months and reaching 0 at 60 months. Its total earnings, proportional to its total expected loss cost, will be 60(1050*K*) or 63000*K*.

Now suppose we have another car driven 18,000 miles per annum but otherwise identical to the first. This car will mile out at 40 months, will earn $K\,((m/12)\,f_{mi} + (m_{avg}/12)\,(1 - f_{mi})) = 1450K$ per month for 40 months and zero thereafter, and its total earnings will be 40(1450*K*) = 58000*K*. The expected loss cost of the second car relative to the first is therefore 58000/63000 = 0.921. The UPR pattern of the second car is pro rata starting with 1 and reaching 0 at 40 months.

Similarly a third car driven 30,000 miles per annum but otherwise identical to the first will mile out at 24 months, will have total earnings of 54000*K*, and a relativity to the first car of 0.857.

The emergence pattern of the first car (reduced proportionally to start at 1) is shown in Figure 11a; the UPR patterns of all three cars are shown in Figure 11b.

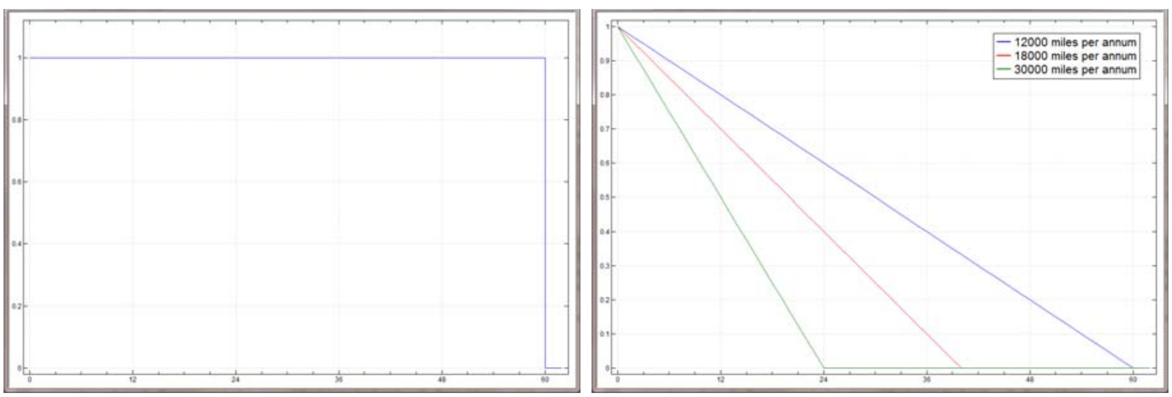


Fig. 11a and 11b. Constant earnings over term of contract, and straight pro rata UPR factors.

Pro rata from end of MW. Suppose each of these cars is subject to a manufacturer's warranty of 36 months or 36,000 miles. For the first car, this warranty will expire at the end of the 36th month. The second car will mile out of its MW after 24 months, and the third car after 14.4 months. We assume no earnings for our contract during the MW. The earnings pattern (Figure 12a) now has three steps, zero from issue to end of MW, non-zero constant from end of MW to expiration, and zero thereafter. The UPR pattern (Figure 12b) is flat until the MW expires, then pro rata to 60 months for the first contract, 40 for the second, and 24 for the third.

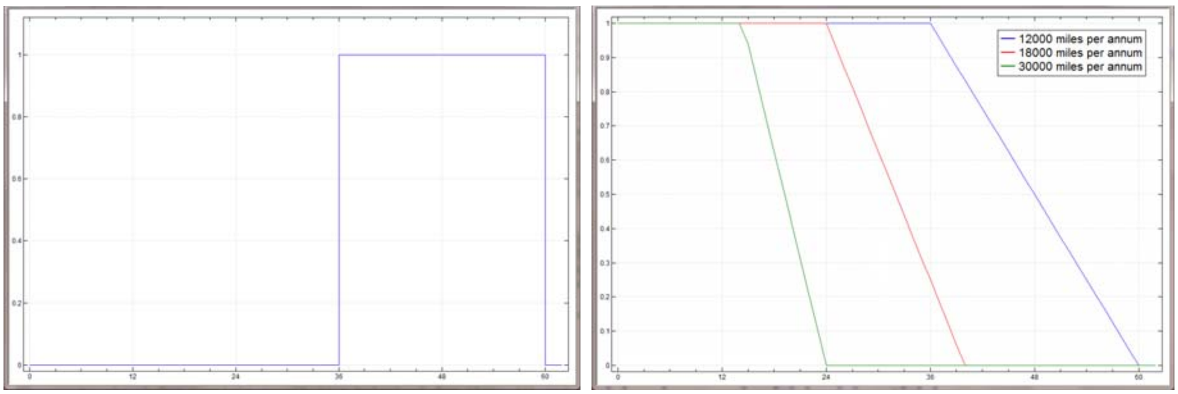


Fig. 12a and 12b. Constant earnings, and pro rata UPR factors, from end of MW.

Power train and non-power train MW's. Now suppose that each car has an extended power train (PT) warranty of 50 months or 50,000 miles, and that 35% of all failures relate to the power train. Now the earnings pattern (Figure 13a) has an additional step and the UPR pattern (Figure 13b) has an additional slope.

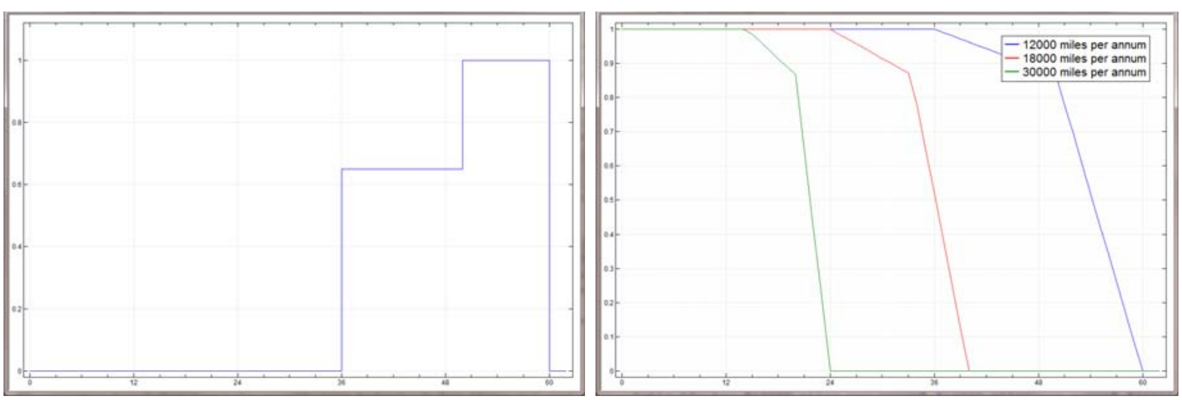


Fig. 13a and 13b. Earnings with PT and non-PT MW's and corresponding UPR factors.

Extras and limitations. Up to now we have assumed that the full cost of a failure, and nothing more, is borne either by the MW or by our contract. But our contract may in fact cover only part of what the MW covers. We adjust for these limitations via factors F_{pt} and F_{npt} applied to the PT and non-PT failures respectively. On the other hand, our contract may cover services, such as towing, that are not part of the MW. We relate these "extras" to y via a factor E. For example, F_{pt} might be

0.99, F_{npt} 0.91, and E 0.03. Figures 14a and 14b show the effect of these values on earnings pattern and UPR factors.

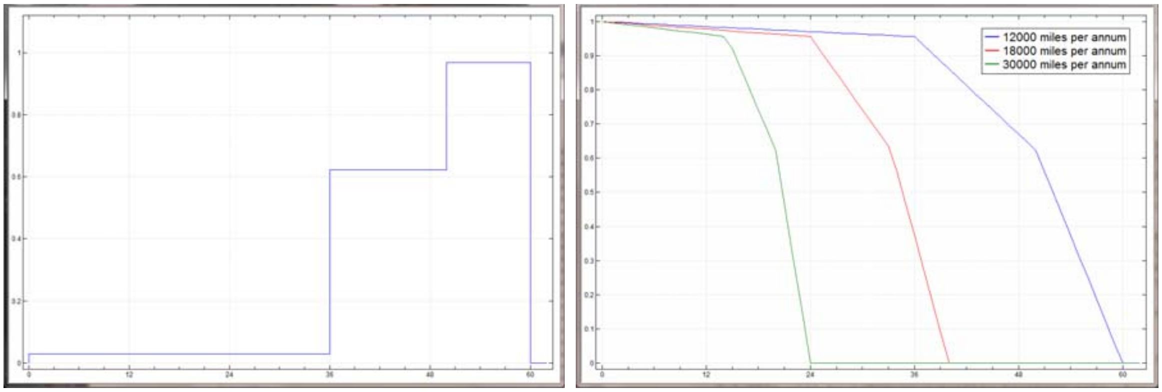


Fig. 14a and 14b. Earnings pattern with extras and limitations, and corresponding UPR curves.

Note that the first earnings step is greater than zero, and the initial slopes of the UPR curves are less than zero.

Pre-expiration spike. As a warranty contract approaches its expiration, policyholders tend to present claims for failures that may have accumulated over some months. If we allow that this increases by 10% the claims presented in the final month of a contract, and (for simplicity) make no adjustment to the claims presented in earlier months, then the resulting earnings and UPR patterns are as shown in Figures 15a and 15b.

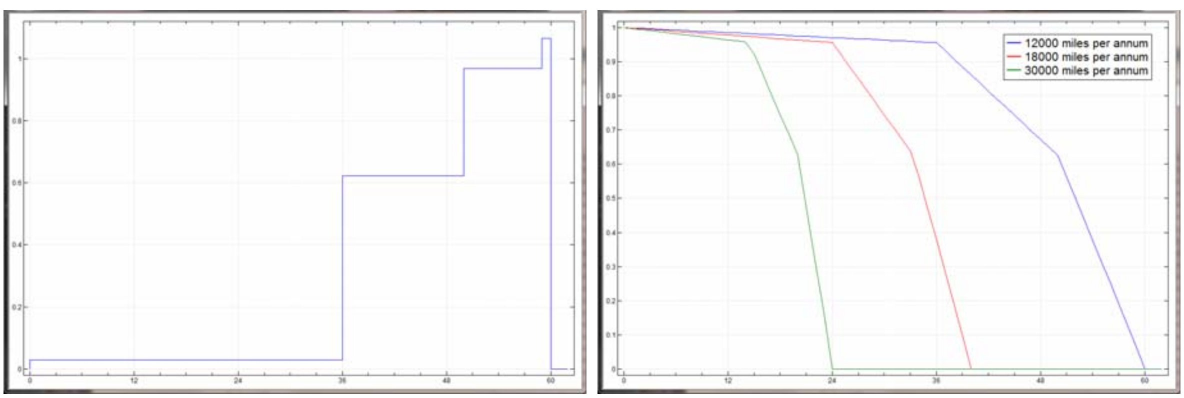


Fig. 15a and 15b. Earnings pattern and UPR factors with 10% pre-expiration spike.

Trend and breakage. Monthly trend and breakage may be represented by monthly growth or decay factors raised to the power of the number of elapsed months and netted against each other. This accounts for the effect of trend on the shape of the earnings and UPR curves. Odometer trend may be represented by a growth factor per (say) 10,000 miles, raised to an appropriate power based on the issue odometer and the number of miles driven per annum. Assuming a trend factor (more precisely a trend-net-of-breakage factor) of 1.05 per annum gives the results shown in Figures 16a and 16b (the vertical scale of Figure 16a is smaller than the previous earnings pattern tables, to accommodate the effect of five years' trend). The effect on the UPR curves is to make them more convex, but only slightly so; trend (when applied in the issue-month axis as well as the development axis) has a much greater effect on loss ratios.

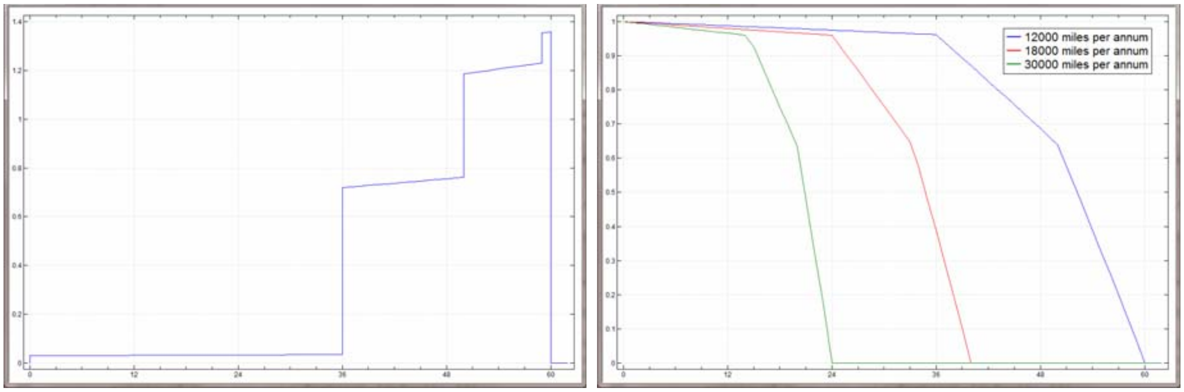


Fig. 16a and 16b. Earnings pattern with 5% annual trend.

Combining miles per annum values. Now suppose that there is a 60% chance that a car will be driven 12,000 miles per annum, a 35% chance of 18,000 and a 5% chance that it will be driven 30,000 miles. Weighting the curves from Figure 16b and averaging them gives us Figure 17.

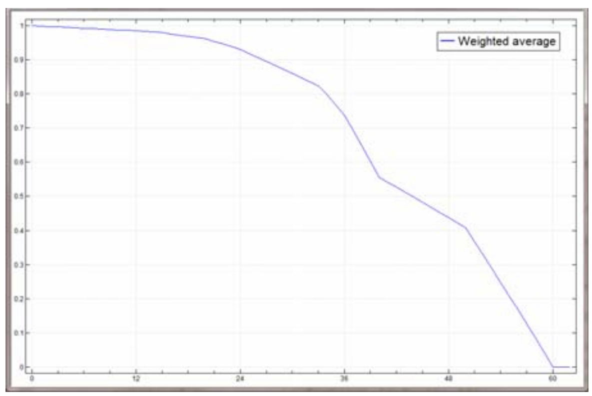


Fig. 17. Weighted-average combination of UPR curves for three miles-per-annum values

Miles-per-annum distribution . In practice we approximate the distribution of miles driven per annum as multinomial with some number N (say 50) of miles values each with an associated probability. Most commonly we select the miles values as midpoints of equiprobable intervals determined by analyzing a large collection of contracts presenting claims, using issue odometer, issue date, odometer at time of first claim, and breakdown date of first claim. This distribution may be adjusted by judgment if it is felt to be biased because of having excluded contracts with no claims. In the case N=50 the probabilities are each 0.02. The result using a representative 50-point miles-per-annum distribution is shown in Figure 18.

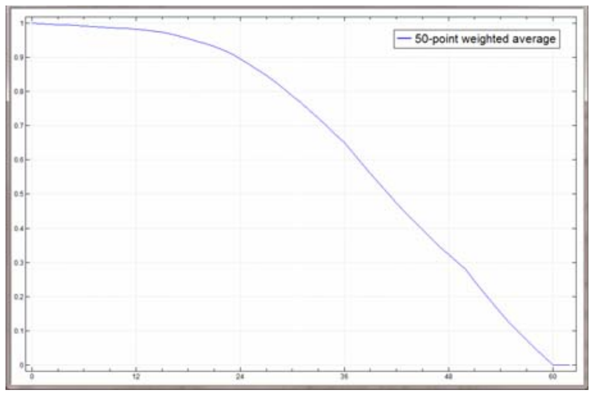


Fig. 18. UPR curve using 50-point miles-per-annum distribution.

The slight irregularity of the UPR curve in Figure 18 beyond lag 48 months reflects the fact that all cars driven less than 12,000 miles per annum will expire at 60 months; therefore there is a large probability mass weighting the 60-month curve, with its inflection point at 50 months when the manufacturer's power-train warranty expires. This is not entirely smoothed out by the contracts with other miles-per-annum values, since many of these mile out earlier than 50 months.

One-half-month adjustment. We normally assume that UPR curves are to be applied to all contracts written in a single month, and that writings are approximately uniform through the month. Figure 19 shows the curve from Figure 18 averaged with the same curve offset by one elapsed month, the so-called one-half-month adjustment.

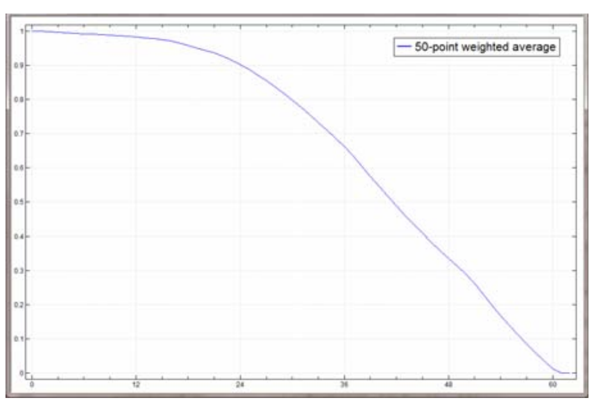


Fig. 19. UPR curve with one-half-month adjustment.

Note that this curve is slightly smoother than the preceding curve and also reaches zero at the end of the 61st month, rather than the 60th, but with slope from 60 to 61 about half that of the slope from 59 to 60. This is the characteristic pattern for one-half-month adjustments.

2.15.3 Residual emergence pattern

Up to now we have talked about how failures, and resulting claims, "ought" to emerge, based on known factors that seem important, such as term, MW, trends, and breakage. If claims really emerge in this way and if we divide actual losses by adjusted exposures lag by lag, for a large collection of contracts with the same term, the result should be constant incremental monthly loss ratios, or straight-line ("pro rata") residual ATF's.

But what if claims do not emerge exactly as expected? For example, what if we assumed no breakage but breakage really is important? Then our theoretical earnings pattern would not decline as much as it should with increasing lag, and our theoretical UPR curve would not be as concave as it should be. If in our body of data we "earn" the premiums according to our theoretical pattern, and then compare losses with this theoretical earned premium, the result will be a residual earnings pattern (proportional to the partial loss ratios to theoretical earned premium) that declines with increasing lag, or residual ATF's that are concave. Now suppose that, when we apply the exposure model to get a theoretical earnings pattern for some proposed contract, we multiply it by the residual earnings pattern for the term of the contract. Then we will be approximately where we would have been had we built the missing factors into our model to start with. Only approximately so, because the residual UPR curve is based on averages, while we are applying it to a particular contract, but in some cases it may be exact, if the missing factors picked up in the residual curve affect all contracts in the same way.

In this way we can use residual ATF's to "true up" our theoretical UPR curves by the ATF model – not just for factors that we might have inadvertently omitted or mis-estimated, but for factors not known to the model, or handled by the model in a simplified way. For example, if breakage were not described by simple exponential decay, but instead tended to increase sharply after, say, year 3, the residual ATF's will pick up this type of variation automatically and will transmit it back to UPR curves calculated for proposed contracts.

Using residual ATFs should improve the fit of the UPR curves for proposed contracts provided (a) the proposed contracts are similar to the contracts in the experience data and (b) the data is of sufficient volume that the residual ATFs represent signal and not noise.

2.15.4 Estimation

Our model provides for estimating residual ATFs for *any* term using *all* terms in the experience data. It does this by loss development using both loss and exposure triangles, assembled separately for each target term *t* from monthly lag triangles of all available terms, weighted by proximity to the target term, but (a) excluding losses that may have emerged after the "official" expiration of a contract (for which the model provides no natural exposure) and (b) with the early rows of the triangle for each contributing term *t** consolidated, if necessary, so that the number of rows equals the number of columns. The latter adjustment allows the triangle to be combined, in the weighted average, with triangles for terms greater than *t** without distorting development patterns. As usual, the exposure triangle reflects the possible decline along rows due to cancellations.

Because the emergence-of-failures patterns are independent of the MW and of our contracts, and the adjustments of the ATF model correct them to the expected emergence-of-claims patterns appropriate for our contracts using a constant of proportionality independent of term, both losses and adjusted exposures may be combined across terms in this way. In effect we first combine the data for all available terms, and then derive residual ATFs for *all* terms, including terms not found in the data. Combining the data and then deriving the curves is more consistent and systematic than deriving curves for each available term and then attempting to reconcile and interpolate them. Our use of a moving weighted average across terms almost always produces a smooth progression of residual ATFs from term to term; it is also tolerant of situations where there is a reasonable volume of data in aggregate but the data is sparse for individual terms (see Figure 20).

The only caveat is that development must essentially cease at the true end of each term. For business not satisfying this condition (e.g., prepaid maintenance, where some dealers informally extend the term to maintain goodwill), judgment adjustments may be necessary either to the data or to the results of the ATF model.

Figure 20 illustrates residual ATFs for a single large program. The residuals are close to straight lines for the shorter terms, but bow outward thereafter, especially for terms 84 months and greater. This suggests that the longer-term contracts may be qualitatively different from the rest in ways that are not explained by the model, and that, in the weighted average of data from different terms, these contracts are pulling those of, say, 60 months, outward. Also the data at the longer terms and later lags is sparse, as shown by the irregularity of the UPR curves.

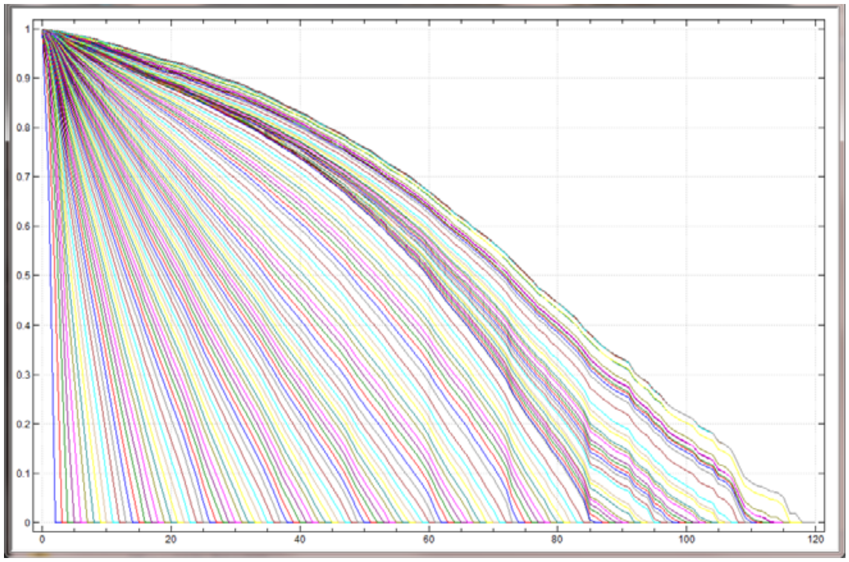


Fig. 20. A set of residual UPR curves based on all data.

Figure 21 shows the residual UPR curves for the same case but with the analysis subdivided so that the curves for terms 1-72 months depend only on data within that range, and likewise for terms 73-120 months. There are now two families of curves with space between.

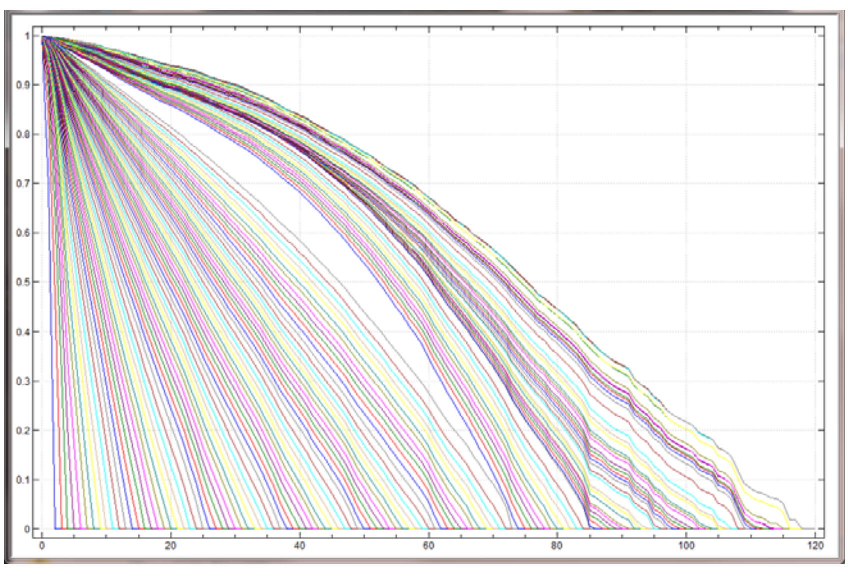


Fig. 21. Residual UPR curves based on data divided at term 72 months.

Figure 22 shows the same curves as in Figure 21, except limited to terms 1 through 72 months and rescaled for clarity. For the terms from about 48 months through 72 months there is still a modest amount of convexity. The ATF model as applied leaves a small amount of the variability of these UPR curves (from pro rata) unexplained.

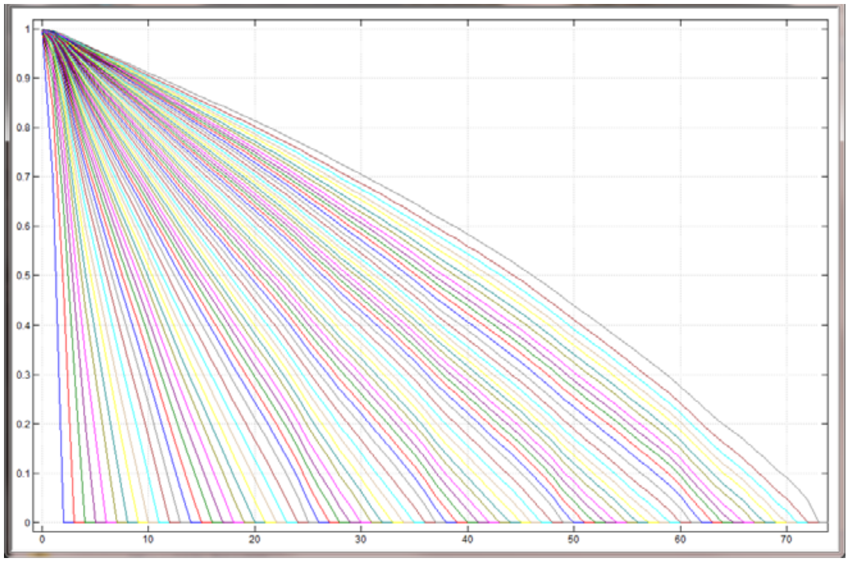


Fig. 22. Detailed look at terms 1-72 from Fig. 21.

Figure 23 shows the residual curves for terms 1 through 72 months after modifying the model to include an annual trend factor of 1.05, thereby removing that amount of trend from the residuals. This eliminates most of the apparent convexity.

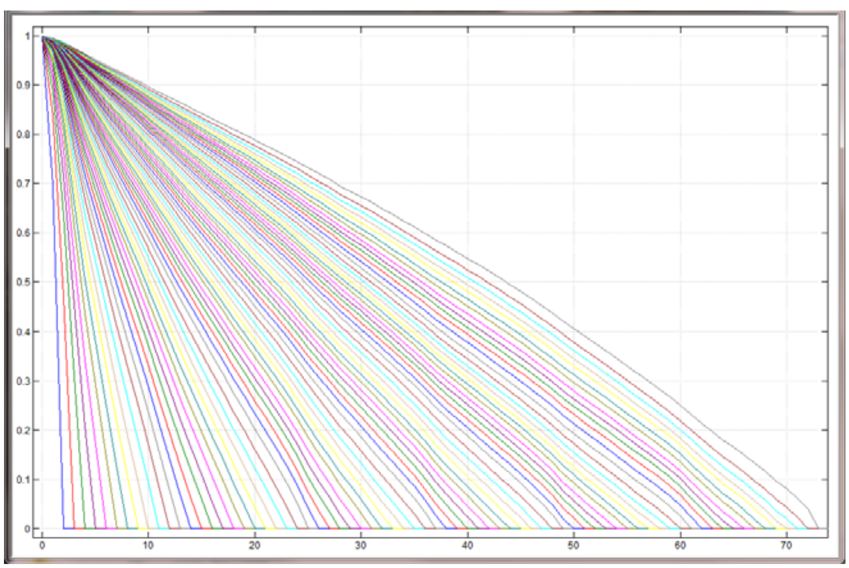


Fig. 23. Residuals after including annual trend of 1.05 in model.

Figure 24 compares the curves for terms 36, 48, 60, and 72 months, with and without the trend, against pro rata lines for the same terms. Clearly the model with the 5% trend parameter brings the residual curves closer to straight lines. However, in estimating UPR curves for proposed contracts, if trend is omitted from the model, it will emerge as part of the residuals, and thus find its way indirectly into the estimated curves, often making them close to what they would have been with trend included.

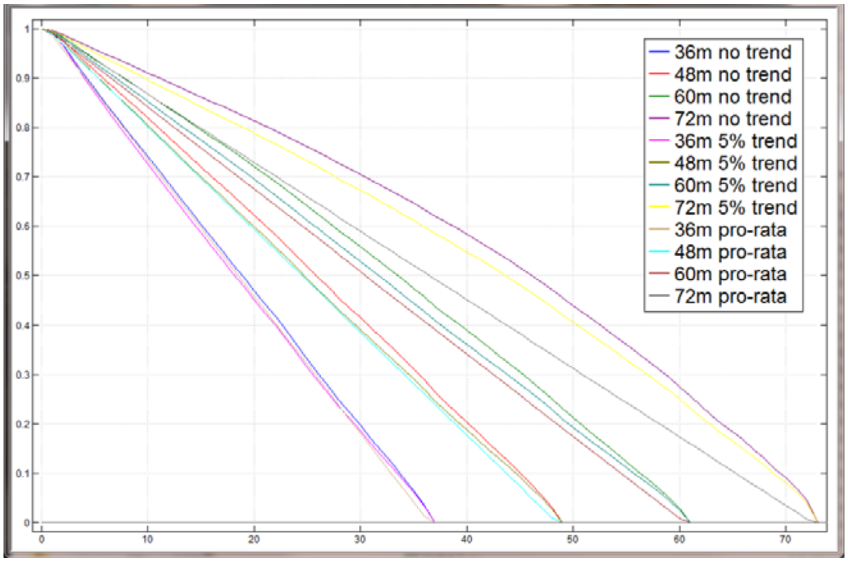


Fig. 24. Comparison of residuals after fitting models with no trend and 5% trend with pro rata.

In summary, when deriving UPR curves for proposed contracts, the ATF parameters involving our contract, the manufacturer's warranty, the power-train, non-power-train, and extras fractions, and the pre-expiration spike, should be chosen carefully, for they affect different contracts in different ways. If there exists a large enough body of relevant experience, it may be used to derive residual ATFs that capture the effect of all other factors, including trend and breakage, at least approximately. The UPR curves for the proposed contracts may be derived by applying the ATF model starting with the residuals. Otherwise, trend and breakage parameters should also be chosen carefully, and the ATF model applied directly to the proposed contracts without using residuals.

2.15.5 Relativities

Our ATF model generates its earnings factors, before normalizing them to total 1, as amounts which are proportional to expected losses across plans as well as across lags within a single plan. Therefore, if we sum these factors across all lags, we obtain sums proportional to the total expected losses for each plan. This facility works identically whether starting from a table of residual ATFs or not.

For example, a plan with term 48048 with underlying MW of 24024 power train and non power train is worth 45.2% more than a plan with term 60060 with MW of 50050 power train and 36036 non power train, using no residuals, or 42.6% more using the residuals from the model fitted to a particular large account with 5% trend.

We make use of relativities of this type when setting up manual rate tables for new programs. Typically such tables involve several different terms, and several different coverage levels, applied to vehicles with several different manufacturers' warranties. For used vehicles the rate table also includes different odometer "bands", for example, 0-999 miles, 1000-9999 miles, 10000-19999 miles, and so forth. Our rating procedure then applies the ATF procedure to proposed contracts with the given terms, MW's, coverage levels, and odometer-band midpoints or centers of gravity. For each such contract the model gives its relativity to a particular base contract. If we have separately estimated the expected loss cost for the base contract, these relativities lead to the expected loss cost for each of the proposed contracts. In this way the ATF model, primarily designed to derive families of UPR curves, may also derive families of expected loss costs for manual ratemaking.

Issues of homogeneity may arise when estimating ATFs. If the contracts for some terms differ from those for other terms in ways not being accounted for in the model, then the residual ATFs for terms close to the boundary may be distorted. We may address this issue by controlling the weighting of data for terms successively more remote from the term being estimated. Or, if we know a priori that the shape of the UPR curves should differ for certain groups of terms (for example because they do or do not have manufacturers' warranties), then we may request that our model calculate the family of curves within each of several term groupings using local data only.

It turns out that the final strings produced by **AllTermsFactors** are not sensitive to assumptions that affect all contracts in the same way, such as breakage and time-based trend. For example, omitting trend entirely, when it exists in the data, will simply transfer it from the model to the residual curves, from which it will be picked up again by the fitted curve for each particular contract.

On the next page is a schematic diagram of the All-Terms Factors process. The names inside blue ellipses are some functions in our system: **AllTermsFactors**, which uses experience to generate ATF curves for all contracts in the data, and a table of residual ATFs, **ATFFormula**, which applies ATFs to obtain UPR for each contract in the data, and **GetATF**, which uses residual ATFs to obtain UPR factors for arbitrary contracts. We do not need to be concerned here with how these functions operate, but need simply to recognize the roles they play in the model.

Note the distinction between contract attributes, global parameters, and estimator controls. The contract attributes are stored in our data. Some of the global parameters may be estimated from the data, or may be selected by judgment and checked for reasonableness by inspecting the residual string table. The estimator controls are largely a matter of judgment.

2.15.6 All-Terms Factors for non-auto contracts

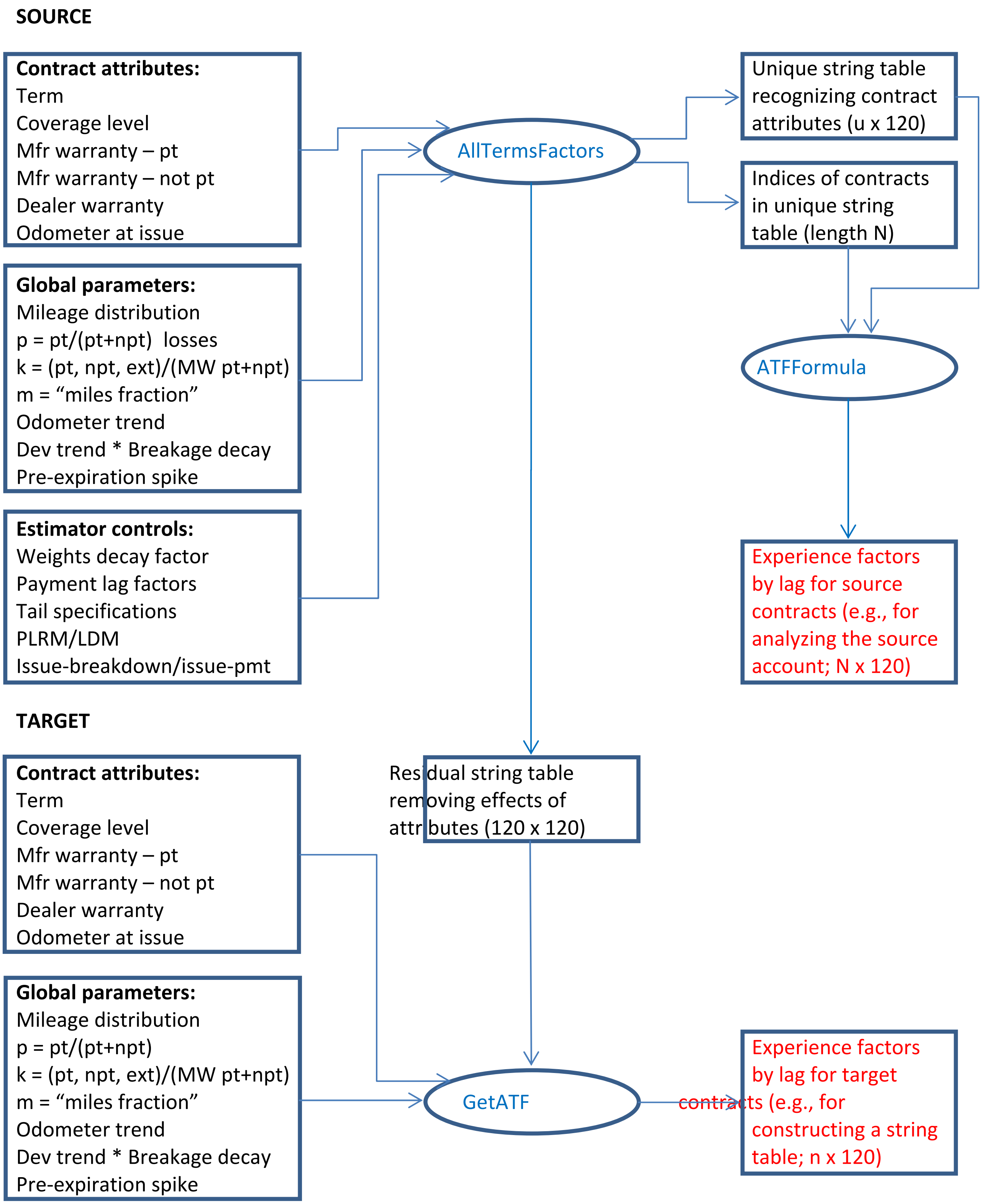
Many of the features described above, such as terms in months and miles, power-train versus non-power-train MW's, and extras, are designed for contracts on automobiles. However, the All-Terms Factors model also may be applied to non-auto business such as electronics or power sports.

One approach is to use the simplified ATF model, creating a table of ATFs by term only. This should be satisfactory when analyzing the UPR term by term for the same data set. If we wish to analyze individual contracts or groupings of contracts other than by term, but recognize differences among contracts in MW's, it may be preferable to use the regular model, interpreted slightly differently from auto, as follows:

- Distinguish parts versus labor instead of power train versus non-power train
- Code all terms and MW's in the data to have no miles limit
- Assume a one-point dummy miles distribution with a low miles value
- Use the "extras" input for Accidental Damage Handling coverage if applicable.

The last point requires explanation. Most non-auto Warranty contracts do not involve extra services, so the coverage factors for extras are normally zero. However, some contracts provide Accidental Damage Handling (ADH) coverage in addition to warranty coverage in case of defects. ADH is different from typical extras in that (a) its expected value may be much higher, often of the same magnitude as the regular repair or replacement coverage, and (b) it may be subject to a months limit shorter than the regular term of the contract. Our ATF model as implemented in software provides for (a) via the "extras" parameter and provides separately for (b), contract by contract.

All-Terms Factors Process



2.15.7 Miles-per-annum distributions

Mileage distributions. The All-Terms Factors model requires that we take into account the number of miles driven per month by the holder of a contract; this is not known at the issue date but must be represented with a probability distribution. Such a distribution is naturally continuous but we have found it useful to approximate it with an n-point discrete distribution.

It is tempting to measure the annual mileage distribution using issue odometer and approximate age of vehicle, based on model year, both usually available in the data for contracts on used cars. Unfortunately this gives the average miles driven by the previous owners rather than by the owners purchasing the contracts in the data. Instead we consider contracts presenting claims, and use the odometer at issue, the odometer at breakdown, the issue date, and the breakdown date of the first claim for each such contract. Using only the first claim avoids counting the down time for service and thereby reducing the apparent mileage. However, using contracts with claims probably biases the result upward, as discussed below.

Figure 25 shows the distribution of miles per annum, to the nearest thousand, for contracts in a large representative program.

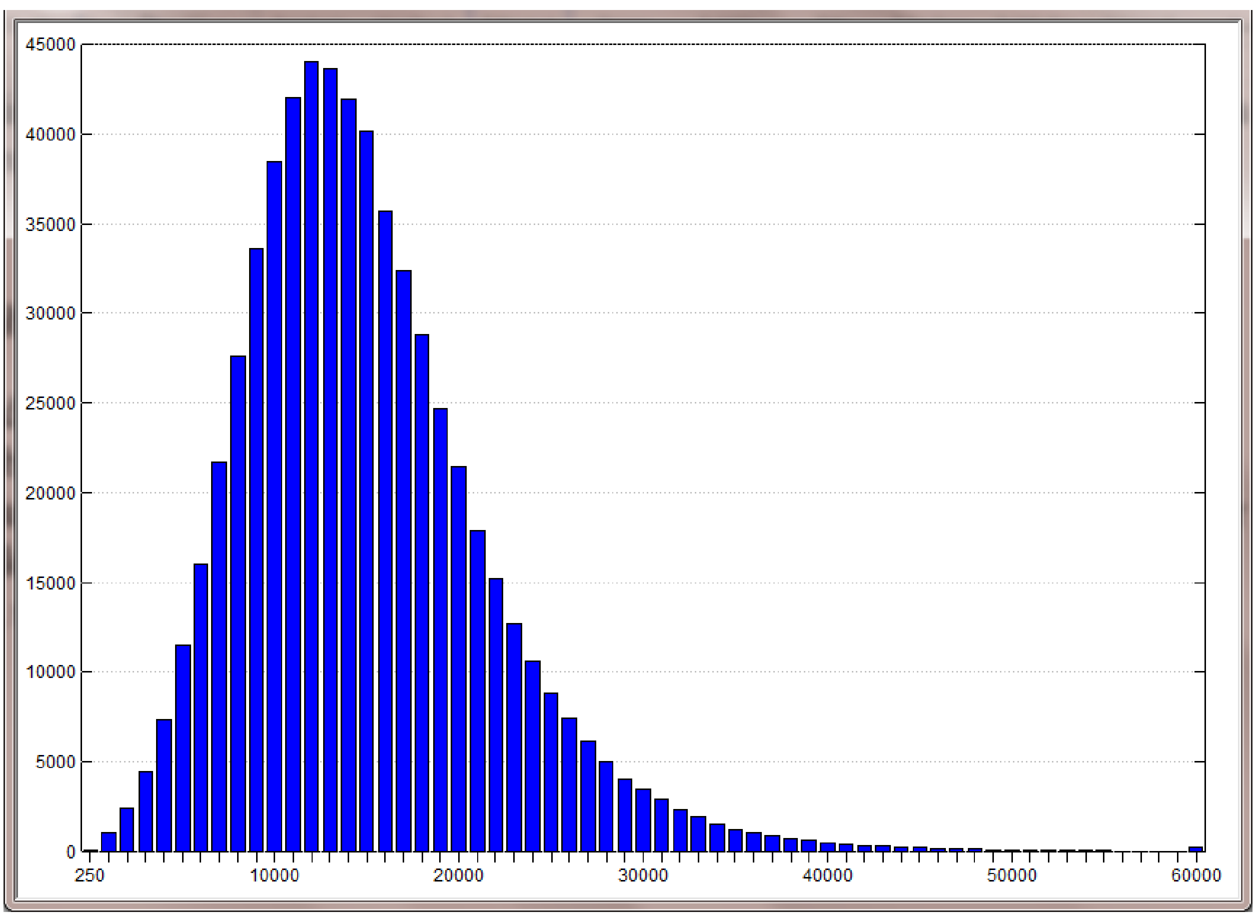


Fig. 25. Distribution of miles per annum, in thousands, based on first claims

The values represented by these bars may be normalized to total 1.00 by dividing by the total number of contracts observed, in this case 535,947, and these probabilities used with the midpoints of the x values in each group to define a discrete distribution. Here the first bar runs from 0 to 500 miles and the last bar covers 59,500 or more miles (so its midpoint should be greater than 60,000); the remaining bars are each of width 1,000, with boundaries 500, 1500, etc.

The function generating the above plot provided the following statistics (from a fragment of a J-language session screen):

```
   m=:20040701 20140630 20140630 1000 0 61 1 MilesDistribBands
''
  Mean (weighted by days) = 14930
  Mean (weighted by cars) = 16442
  Standard deviation (from variance weighted by days) = 6987
  Standard deviation (from variance weighted by cars) = 8173
  Number of observations = 535947
  Maximum = 179488
  Minimum = 365
  Probability weight for each band determined by total years
  Representative value in each band equals center of gravity
  Approximated mean: 14930.44683
  Approximated std dev: 6802.084723
```

Often, instead of using groups of equal width with different probabilities, we use N groups of variable width with equal probabilities. The following lines illustrate this with N=50:

```
   m50=.20040701 20140630 20140630 50 1 MilesDistribNpt ''
  Mean (weighted by days) = 14930
  Mean (weighted by cars) = 16442
  Standard deviation (from variance weighted by days) = 6987
  Standard deviation (from variance weighted by cars) = 8173
  Number of observations = 535947
  Maximum = 179488
  Minimum = 365
  Boundaries of bands determined by accumulated years
  Representative value in each band equals center of gravity
  Approximated mean: 14930.4052
  Approximated std dev: 6712.981379
```

The mean approximated by the discrete distributions is very close to the mean of the entire distribution while the standard deviation is a bit smaller. This is mainly the result of compressing all values in the right tail of the actual distribution – which range as high as 179,488 – into a single value of 38,463. The whole 50-point equally-probable distribution is:

```
   5 10$round >0{m50
   2984 4828 5804 6550 7151 7675 8138 8562 8954 9318
   9663 9992 10313 10628 10930 11224 11519 11808 12089 12373
  12658 12946 13234 13522 13815 14115 14418 14716 15025 15347
  15684 16032 16387 16755 17139 17543 17960 18412 18897 19416
  19973 20588 21279 22064 22971 24051 25402 27200 30004 38463
```

As expected, these values are much closer to each other toward the overall mean than in the extremes, especially in the long right tail.

The use of miles-to-first-claim as a starting point for estimating the miles-per-annum distribution may be the most practical available method but is not ideal. In particular, it ignores the miles driven by any cars that have not presented claims. Since claims emerge largely as the result of usage, cars presenting claims may be expected to be driven more, on average, than all cars, and produce a distribution that is biased upward. One way to offset this tendency is to determine the *N* mileage intervals and their means using time as a weight, as shown above.

Miles per annum might well be represented by a continuous distribution and we use a discrete approximation mainly for simplicity. If we plot the distribution from Figure 25 as a curve and superimpose a lognormal distribution fitted by the method of moments, we obtain the comparison shown in Figure 32.

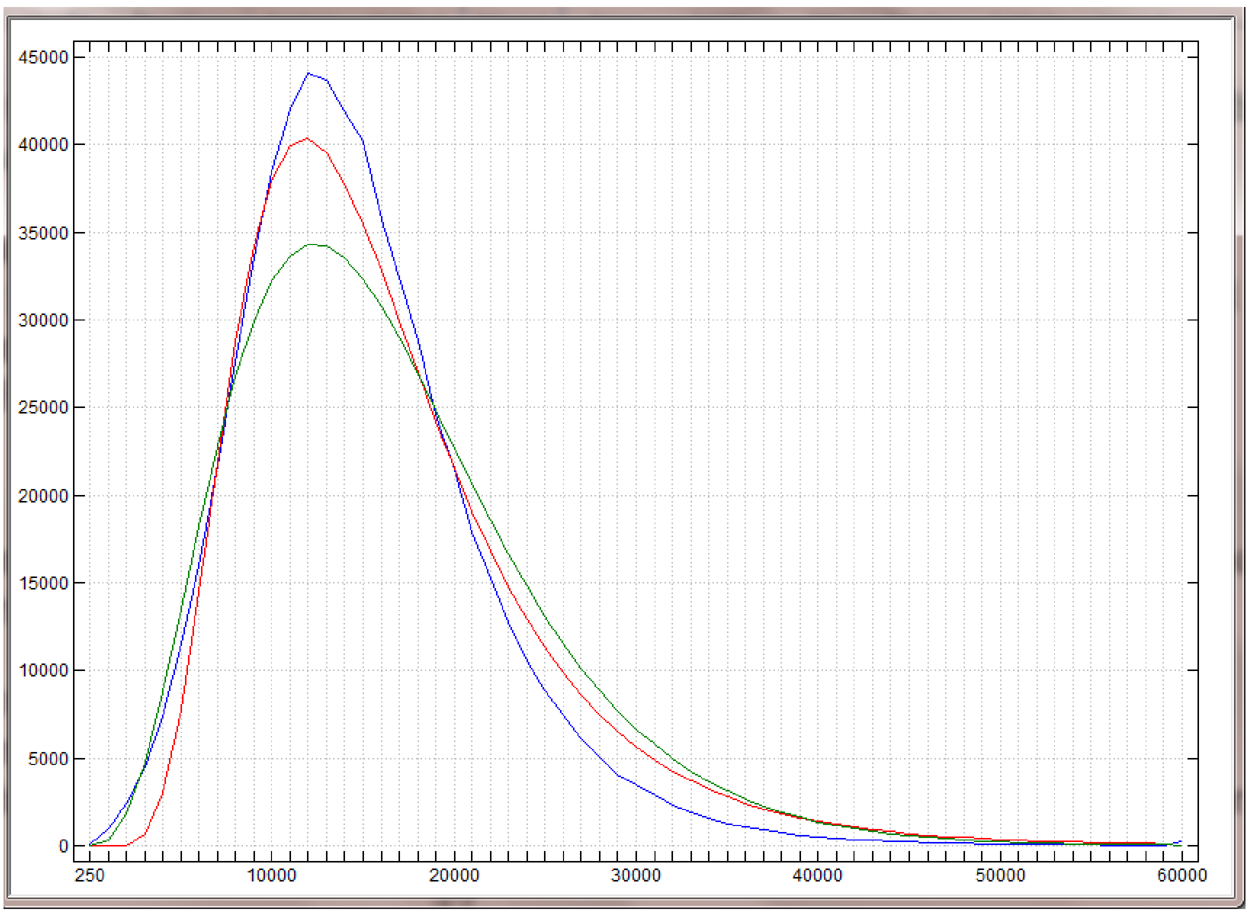


Figure 32. Observed miles-per-annum distribution (blue) compared with lognormal distribution (red) and Gamma distribution (green).

The lognormal does not fit this distribution well, especially in the tails; the Gamma is rather better in the left tail (in which we are not particularly interested) but worse in the central area and the right tail; moreover, there is some area under both of these fitted distributions arbitrarily far to the right, whereas miles driven have a practical upper bound based on speed limits and the finite number of hours in the day. For our purposes we find the discrete distributions more satisfactory.

3. CONCLUSIONS

In conclusion, the UPR is critically important for Warranty Insurance but there are reliable techniques for calculating it. These techniques are similar to those used for loss reserves in other lines of insurance, but require adaptations to address the special characteristics of Warranty contracts. Among these characteristics are exposure that declines over the life of a cohort of policies, loss development triangles that must be adjusted for incompletely reported losses in recent diagonals, tails that are of limited duration but contain significant probability mass, and the need to calculate internally consistent sets of UPR curves for entire families of contracts. When properly conceived and programmed, procedures for assigning and testing UPR factors may be efficiently applied to large numbers of Warranty programs each split into homogeneous subdivisions.

4. REFERENCES

[1] Bornhuetter, Ronald and Ronald Ferguson, The Actuary and IBNR, PCAS LIX, 1972: 181-195

[2] Bühlmann, Hans, Vereinigung Schweizerischer Versicherungsmathematiker / Association des Actuaires Suisses, Ecole d'éte 1983, Estimation of IBNR Reserves by the Methods Chain Ladder, Cape Cod, and Complementary Loss Ratio, unpublished

[3] Gluck, Spencer M., Balancing Development and Trend in Loss Reserve Analysis, PCAS LXXXIV, 1997: 482-532

[4] Kerper, John and Lee Bowron, An Exposure-Based Approach to Automobile Warranty Ratemaking and Reserving, CAS Forum 2007: 29-43

[5] Stanard, James N., A Simulation Test of Prediction Errors for Loss Reserve Estimation Techniques, PCAS LXXII: 124-153