

Claim Reserving: Performance Testing and the Control Cycle

by Yi Jing, Joseph Lebens, and Stephen Lowe

ABSTRACT

Fundamentally, estimates of claim liabilities are forecasts subject to estimation errors. The actuary responsible for making the forecast must select and apply one or more actuarial projection methods, interpret the results, and apply judgment. Performance testing of an actuarial projection method can provide empirical evidence as to the inherent level of estimation error associated with its forecasts. Performance testing of alternative methods provides formal assurance that the actuary is using the best methods for the given circumstance, and also provides insight into the appropriate weight to give to the indications produced by each method. Performance testing is an integral part of the actuarial control cycle associated with the loss reserving process. It provides the necessary feedback loop to the actuary, assuring that he or she is not overconfident about his or her forecasts. This paper describes how to construct sound performance tests, consistent with statistical cross-validation, within the reserving control cycle. It illustrates the application of the techniques via a case study, including some interesting empirical results.

KEYWORDS

Chain-ladder, claim reserving, control cycle, cross-validation, estimation error, forecast error, loss reserves, overconfidence, parameter risk, process risk, predictive skill, reserve accuracy, reserving process

1. Introduction

Fundamentally, estimates of claim liabilities are predictions or forecasts. Presented with a body of historical data, and with knowledge of the operation of the business and the exposures underwritten, the actuary is called upon to develop a forecast of the future cash flows associated with the settlement of all unpaid claim obligations. Since claim development involves one or more underlying stochastic processes, the actuary must separate noise from signal in the available data. This necessitates the selection of an underlying model, application of the model to the available data, selection of model parameters, and interpretation of the results. The resulting forecast will be subject to estimation errors resulting from both any misspecification of the model and misestimation of its parameters, as well as random process variation.

Typically the actuary will employ several actuarial projection methods and select an actuarial central estimate of the claim liabilities after reviewing the results of each method, giving appropriate weight to the indication produced by each method and supplementing the mechanically generated indications with judgment. A natural question to ask of the actuary is, “*How do you know that the methods chosen are the most appropriate ones to employ?*” A corollary question is, “*How do you decide the relative weight to be given to the results from each method?*” While some might reply that the selection of methods and the assignment of weights are actuarial judgments made with the benefit of experience, in the current regulatory and accounting environment there is a need to respond with greater clarity to such questions. It therefore seems worthwhile to develop a formal methodology to assess the accuracy of the methods used to make actuarial estimates.

Ongoing *performance testing* of methods can help (a) provide empirical evidence as to the inherent level of estimation error associated with

any particular method; (b) provide insight into the strengths and weaknesses of various methods in particular circumstances; (c) provide assurance that the actuary is using the best methods for the given circumstance; (d) provide useful information regarding the appropriate weights to be given to each method; and (e) ultimately lead to meaningful improvements in actuarial estimates, in the sense of lower estimation errors, as the stronger methods win out over the weaker methods, based on objective evidence as to their performance.

Performance testing can also assist in managing the potential overconfidence of the actuary, a topic discussed by Conger and Lowe (2003). Behavioral scientists have demonstrated that the vast majority of managers are overconfident in their ability to make accurate forecasts—of everything from estimates of next year’s sales revenue, to the cost of implementing a new system, to the delivery date of a new product under development—because they lack *meta-knowledge* about the limitations of their forecasting abilities.¹ Overconfidence is a problem because blown estimates, cost overruns, and missed deadlines hurt business performance. Actuaries are not immune to the overconfidence phenomenon; it manifests itself when actual claims are outside the range more frequently than expected. The formal feedback loop provided by the control cycle helps the actuary develop this meta-knowledge, becoming “well-calibrated” as to the true limitations of his or her actuarial projections.

In addition to improving the quality of the actuarial estimates of claim liabilities, performance testing results are also useful in other contexts such as setting ranges around central estimates

¹Meta-knowledge is an understanding of the limitations of one’s knowledge and capabilities. It includes an appreciation of the limitations on one’s forecasting ability. A forecaster can develop meta-knowledge by receiving continuous feedback as to the accuracy of his or her forecasts. (Such a forecaster is said to be “well-calibrated.”) For example, meteorologists are generally well-calibrated because of the natural feedback they receive about their daily weather forecasts.

and measuring required economic capital for reserve risk.

Since estimates of claim liabilities are forecasts, performance testing of actuarial projection methods should involve proper measurement of prediction errors via a standard statistical technique, referred to as *cross-validation*. This paper provides an overview of cross-validation, defines how it would be employed in the context of property and casualty claim liability estimates as a key element of the control cycle, and provides some illustrative empirical results.

Including this introduction, this paper is organized into four sections. Section 2 of the paper lays out a definitional framework for claim liabilities and the actuarial estimation process. Section 3 discusses performance testing in general, including performance criteria and cross-validation techniques; and introduces the actuarial control cycle, outlining the role that performance testing should play in it. Section 4 illustrates all of the concepts via a case study, using real data from a U.S. insurer, with additional calculation details provided in an Appendix.

The case study generates some interesting conclusions, highlighted here to pique the reader's curiosity.

- At some maturities, a method that projects the ultimate claim liabilities from the *outstanding case reserves* outperforms the reported chain-ladder method.
- Ultimate claim counts can be predicted with much greater accuracy than ultimate dollars. This implies that projection methods that make use of claim counts can be more predictive than those that do not. While the greater stability of claim count projections may be partially offset by greater instability of average claim values, the information value of the claim counts should lead to a net improvement in overall accuracy over projections that ignore claim count information.

- Methods that formally adjust for changing settlement rates or case reserve adequacy have greater predictive accuracy than the analogous methods applied without adjustments. When changes in claim settlement rates and/or changes in case reserve adequacy occur, the accuracy of projection methods that do not adjust for them is materially degraded.
- In assessing which projection methods to employ, and how to combine the results from several methods, the degree of correlation between the estimation errors is an important consideration. To illustrate, if the results of two methods are perfectly correlated but the estimation errors from one are twice the estimation errors of the other, then the appropriate course of action is to drop the method with the higher estimation errors entirely. A different course of action would be implied if the errors were uncorrelated.

Not all of the conclusions from the case study can be generalized beyond its immediate circumstances. However, the case study does demonstrate the kinds of insights that can be obtained through the performance testing process, and the utility of those insights in the control cycle. Performance tests that we have performed on other lines of business and other actuarial projection methods also confirm the value of the exercise, suggesting that this is a fertile field of empirical investigation. The authors would encourage others to pick up where the paper leaves off and add to the body of performance testing research.

2. Defining the claim liability estimation process

Before describing performance testing, it is necessary to establish a definitional framework for the process by which claim liabilities are estimated.

2.1. Claim liabilities

Claim liabilities are the loss and loss adjustment expense obligations arising from coverage provided under insurance or reinsurance contracts. The term may refer either to *actual* claim liabilities (that is, an after-the-fact realization of the claim payments) or to *estimated* claim liabilities produced by an actuarial projection method.

Claim liabilities may be estimated either on a nominal basis or on a present value basis. While estimates are more commonly made on a nominal basis, estimates on a present value basis are more representational of the economic cost of the claims. Present value estimates may be less accurate due to misestimation of the timing of payments; they also may be more accurate because the predicted payments further into the future—which are subject to the greatest uncertainty—are given successively less weight.

Let $C_{i,d}$ represent the actual incremental paid claims on the i th accident year² during the d th development period. Claim development is the realization of the actual claim liabilities, $C_{i,d}$, over time, reflecting the confluence of a number of underlying stochastic processes: claim generation, claim reporting, claim adjusting, and claim settlement. (On a net basis, additional processes would include the identification, pursuit, and recovery of subrogation, salvage, third-party deductibles, and reinsurance.)

Since they are the result of stochastic processes, all $C_{i,d}$ are random variables.

We typically have a triangular array of actual values of $C_{i,d}$, up to the current calendar point in time $t = i + d - 1$, as shown below.³

$$\begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} \\ C_{2,1} & C_{2,2} & C_{2,3} & C_{2,4} & \\ C_{3,1} & C_{3,2} & C_{3,3} & & \\ C_{4,1} & C_{4,2} & & & \\ C_{5,1} & & & & \end{bmatrix}$$

These actual $C_{i,d}$ are a sample partial realization from the underlying stochastic processes.

Let $\hat{C}_{i,d}^{(t)}$ represent⁴ the estimate (more accurately, a forecast) of the expected incremental paid claims on accident year i during development period d , based on information at time t .

The actual aggregate discounted claim liabilities at time t are given by

$$L^{(t)} = \sum (v_{i+d})(C_{i,d}^{(t)}) \quad \text{for all } i + d - 1 > t, \quad (2.1)$$

and the estimated aggregate discounted claim liabilities at time t are given by:

$$\hat{L}^{(t)} = \sum (v_{i+d})(\hat{C}_{i,d}^{(t)}) \quad \text{for all } i + d - 1 > t, \quad (2.2)$$

where v_{i+d} is the discount factor applicable to payments made $i + d - t$ periods after t .

Finally, actual and estimated aggregate claim liabilities are related by

$$L^{(t)} = \hat{L}^{(t)} - e, \quad (2.3)$$

where e is the estimation error, the difference between the estimated and the actual realization of the claim liabilities. Because all of the $C_{i,d}$ are random variables, e is also a random variable.

$\hat{L}^{(t)}$ may be broken down into *component estimates*; for example, the estimated liabilities associated with a particular accident year, or the liabilities that are expected to be paid in a particular future calendar year. Because they are estimates of means, the total of the component estimates is equal to the overall estimate, $\hat{L}^{(t)}$.

²The use of the term accident year is for ease of exposition; the concepts apply equally in an underwriting year, report year, policy year, or other cohort-based configuration of data.

³While the presentation assumes the traditional situation where a triangular array of historical claims is available, the concepts would apply equally in other contexts, for example, where claim liabilities are based on exposure-based methods.

⁴The use of parentheses around the parameter t is to distinguish it from an exponent.

In contrast to claim liabilities, a *claim reserve* is a balance sheet provision for the estimated claim liabilities at a particular statement date. The reserve is a fixed, observable, known quantity, not an estimate. Those who speak of “reserve estimates” are guilty of sloppy language, which hopefully does not translate into sloppy thinking. The reserve is a financial representation of the underlying obligations, reflecting the valuation standards associated with the particular financial statement of which it is an element. These standards are established by the applicable regulatory and accounting bodies within the jurisdiction. Such valuation requirements might not equate the reserve, $R^{(t)}$, with the associated estimated liabilities, $\hat{L}^{(t)}$. For example, regulatory requirements might stipulate that the reserve be set conservatively at a percentile above the mean. Alternatively, the reserve might include a margin for the cost of the economic capital required to support the risk of the associated claim liabilities. It is sufficient for our purposes to simply note that $R^{(t)} = f(\hat{L}^{(t)})$, (or more precisely the distribution around $\hat{L}^{(t)}$) where f is determined by the particular regulatory/accounting regime.

2.2. Actuarial projection methods

An *Actuarial Projection Method* is a systematic process for estimating claim liabilities. It encompasses three fundamental elements:

- An *algorithm*, which is used to project future emergence and development. Any algorithm is predicated on an underlying model, which is a mathematical representation that embodies explicit and/or inferred properties of the claim incidence, reporting, and settlement processes; the algorithm applies the model to a given dataset to produce an estimate of the claim liabilities.
- A *predictor dataset*, which is the input data used by the method. Specification of the data-

set includes the types of data elements, the valuation points of those data elements that vary over time, and the level of detail at which the projection is performed (for example, separate analyses for indemnity and expense, or by type of claim, class of policyholder, or jurisdiction). The dataset may include data external to the entity in addition to internal data; it may also include qualitative information as well as quantitative.

- *Intervention points*, which are points in the process at which judgments are made by the actuary. These include the selection of assumptions (model parameters such as chain-ladder link ratios), overrides to outlier data points, and other adjustments from what would otherwise be produced by mechanically applying the algorithm to the data.

A particular method $m(a, d, p)$ (i.e., consisting of an algorithm, a dataset, and intervention points) is selected by the actuary from the set of all possible methods $M(A, D, P)$ to produce an estimate of the claim liabilities, $\hat{L}_m^{(t)}$, at a valuation point t , for example at the end of a given calendar year.

A projection method might use data with a valuation date that is earlier than the valuation date of the estimated liabilities; in such a case the actuarial projection method would include the “roll forward” procedure that adjusts for the effects of the timing difference.

Most traditional projection methods assume stationarity of the underlying claim processes, such that the historical development in the dataset is assumed to be representative of the conditions that will drive the future development. Often this is not the case, and the data must first be adjusted to reflect the current and anticipated future conditions. Such adjustments, when made, are part of the actuarial projection method.⁵

Wacek (2007) distinguishes between *clinical predictions*, which are conclusions reached by an

⁵Some of the more advanced stochastic projection methods allow for the claim development processes to be dynamic, varying along various dimensions. Some methods also allow for process variation due to exogenous factors, such as monetary inflation.

expert when presented with a set of facts about a problem of a type with which he or she has experience; and *statistical predictions*, which are conclusions indicated by a quantitative or statistical formula or model using a set of quantifiable facts about a problem. He points out that an expert making a clinical prediction may use a statistical model, but if the model results are augmented by consideration of other information and the judgment of the expert, then the prediction would be classified as clinical. Thus the estimation of claim liabilities is generally a clinical rather than a statistical prediction. The inclusion of intervention points in the definition of an actuarial projection method is in recognition of the important role that endogenous information and judgment play. All statistical prediction models/methods have inherent limitations; many of those limitations are routinely addressed by the actuary through the use of intervention points and the insertion of judgment into the estimate.

Performance testing can be done on either statistical predictions (those involving a mechanical application of the algorithm without any judgmental interventions) or on clinical predictions in which we are testing the actuary's performance, not just "the machine's" performance.

2.3. From actuarial projections to the actuarial central estimate

Ultimately, the actuary is responsible for developing an *actuarial central estimate* of unpaid claim liabilities, representing an expected value over the range of reasonably possible outcomes. In developing an actuarial central estimate, the actuary will typically use multiple actuarial projection methods that rely on different data⁶ and

⁶We note, however, that many methods use overlapping data sets. For example, both the incurred chain-ladder method and the paid chain-ladder method make use of cumulative paid loss data. Because of this overlap in data, one would expect that the indications resulting from the paid and incurred chain-ladder methods would exhibit some level of correlation.

require different assumptions. For each method employed, the actuary can produce alternative actuarial estimates by varying the selected parameters and other judgments made at the intervention points in the method. Ultimately, the actuary will settle on parameters and assumptions that produce the central estimate for each method.

One can view the central estimate from each method as a sample mean, developed from a sample of the data. Each sample mean will have a sample variance, depending on the predictive quality of the sample data employed. The actuary will reconcile/combine the sample means by assigning a relative weight to each one to produce the actuarial central estimate.

3. Performance testing and the control cycle

Within the definitional framework articulated in the previous section, we can now turn to the problem at hand.

The first issue facing the actuary is to choose a finite set of actuarial projection methods $\{m_1, m_2, \dots, m_n\}$ from the set of all possible methods M that is "best" for a particular class of business and circumstance.⁷ The second issue is to combine the results of the individual methods together into a single estimate that is "best," giving appropriate weight to each method based on its predictive value in the specific circumstances. As will be seen in subsequent sections, both issues can be addressed more formally via performance testing.

3.1. Performance testing and selection of methods

In establishing a performance testing methodology, one needs to start by establishing appro-

⁷The selected set of methods may vary by maturity. For example, the actuary might use projection methods for complete accident years, but use an expected loss ratio method for the current, incomplete accident year.

appropriate performance criteria. In other words, by what set of criteria do we say that one method is better than another? Preference for one method over another would then be based on performance test results against those criteria in the specific circumstances.

Obviously, the choices of methods at a given point will also be constrained by the available data (both quantity and type) and the available resources to produce and analyze them. However, over time, performance testing can assist in evaluating the benefits against the costs of generating currently unavailable data elements or of adding resources to support more labor-intensive projection methods. Performance testing might even *reduce* actuarial resource costs by redeploying resources away from actuarial methods that do not meet cost-benefit criteria.

Statisticians and forecasters traditionally argue that the best method is the one that produces estimates that are unbiased and minimize the expected squared error. Such methods are described as “BLUE”—Best Least-Square, Unbiased Estimators.

For our purposes, we extend the criteria somewhat beyond BLUE. In evaluating the performance of actuarial projection methods it is appropriate to consider the criteria listed in Table 1.

Our criteria extend the acronym from BLUE to “BLURS-ICE”—Best Least-Square, Unbiased, Responsive, Stable, Independent, and Comprehensive Estimate.

The minimum squared error criterion is obviously the strongest; it is the primary yardstick against which methods are compared. The requirement for unbiased estimates is a weaker criterion, in that we would accept methods with bias if the bias were measurable or immaterial. In other words, we would be willing to trade a lower expected squared error for the introduction of a manageable level of bias. Stability and responsiveness are desirable characteristics, but are also weaker criteria.

Table 1. Performance criteria for selecting actuarial projection methods

Least-Square Error	The method should produce estimates that have minimum expected error, in the least squares sense.
Unbiased	Ideally, the method should produce estimates that are unbiased, with an expected error of zero.
Responsive	The method should respond quickly to changes in underlying trends or operational changes. Given several methods with similar expected errors, we would favor the method that responds to changes in trends or operations most quickly.
Stable	Methods that work well in some circumstances, but “blow up” in other circumstances are less desirable.
Independent	The method should not generate prediction errors that are highly correlated with those of other selected methods. (This criterion is discussed more fully in subsequent sections.)
Comprehensive	The method should use as much data as possible, in an integrated way, so that the maximum available information value is extracted.

Comprehensiveness and independence look at the problem through a slightly different lens than the other criteria. These two criteria relate to the issue of combining multiple estimates together to obtain an actuarial central estimate. The comprehensive criterion stems from our desire to maximize the information value of all of the available data; rather than applying many separate methods to disparate data elements it is preferable to use fewer, more comprehensive methods. Similarly, the independence criterion stems from a desire to not waste resources on methods with highly correlated estimation errors, producing essentially the same answer as another method. If the errors from two projection methods are highly correlated it is preferable to eliminate the one with the higher expected squared error.

3.2. Performance testing using cross-validation

Rather than invent a performance testing methodology from scratch, we need only look to the approaches used in other sciences that are also

called upon to make forecasts. Statistical techniques to measure the accuracy of forecast models are abundant in the literature of other professions. For example, as a result of the ongoing interest in hurricanes, the authors have had occasion to review the literature relating to the accuracy of hurricane forecasts made by climatologists. In climatology, the “skill” of statistical forecasting methods is measured via *cross-validation*. The method of cross-validation is outlined in Michaelson (1987). The validity of the method for estimating forecast skill is discussed in Barnston and van den Dool (1993). The potential pitfalls and misapplication of the method are discussed in Elsner and Schmertmann (1994).

Cross-validation, properly applied, permits a fair comparison of the skill of competing forecast models. For example, forecasts of the expected number of major hurricanes spawned in the Atlantic basin are made annually by Gray (2008) and others, based on statistical models that relate the level of Atlantic hurricane activity to other climate variables such as the state of the El Niño Southern Oscillation (ENSO) in the Pacific. The skill of each of these models can easily be measured and compared via cross-validation. Similar cross-validation tests are performed to measure the skill of storm path forecast models employed by the National Hurricane Center. This helps users of the forecasts to appreciate their limitations, and allows each modeler to place formal ranges around the central estimate of their model’s forecast. It also facilitates the selection of the “best” forecast model and provides direction to the course of further research and enhancements.

Cross-validation is a straightforward, nonparametric approach to estimating the errors associated with any forecast produced by a model. This would include forecasts of future claim payments in estimating claim liabilities. Cross-validation is relatively simple to implement, given sufficient data. It is quite general in its applicability to any

suitable defined actuarial projection method, and has intuitive appeal because it closely mimics the actual claim liability estimation process.

3.3. Cross-validation in general

The basic approach in the context of a general statistical forecasting model is to develop a model using all observations except one, and then to use the model to produce a forecast of the omitted observation. This process is repeated iteratively, generating a forecast from a new model developed by excluding a different observation each time. The cross-validation exercise generates an independent forecast for each observation, based on a model generated by the rest of the data. The skill of the model can then be measured by comparison of the forecasts to the actual outcomes.

For example, let Y represent a random variable that we believe is dependent on some set of independent variables $X = \{a, b, c, \dots\}$. In other words we believe that $Y = f(X)$, although we don’t know the precise form of f . Suppose that we have n historical observations of (X, Y) , and would like to develop a linear regression model that predicts future values of Y given X . Rather than using all n values of (X, Y) to develop the regression equation, the cross-validation approach would develop n regression equations, each one based on $n - 1$ of the observations. Each equation would then be used to predict the Y value of the excluded observation, Y^* , as displayed schematically in Figure 1.

Comparison of the actual values of Y to the predicted values Y^* is a fair way to measure the true ability of a linear regression model to predict Y from X . We are testing the predictive ability of a linear model, not the parameters of the model itself. By comparison, the standard error of the regression is *not* a fair way to measure the equation’s predictive ability. The standard error is a measure of the degree to which the equation explains the variation in Y ; *inevitably the standard*

Figure 1. Cross-validation in a regression analysis context

$X_1 = \{a_1, b_1, c_1, \dots\}$	Y_1	$f_1(X)$	Y^*_1
$X_2 = \{a_2, b_2, c_2, \dots\}$	Y_2	$f_2(X)$	Y^*_2
$X_3 = \{a_3, b_3, c_3, \dots\}$	Y_3	$f_3(X)$	Y^*_3
\vdots	\vdots	\vdots	\vdots
$X_k = \{a_k, b_k, c_k, \dots\}$	Y_k	$f_k(X)$	Y^*_k
\vdots	\vdots	\vdots	\vdots
$X_n = \{a_n, b_n, c_n, \dots\}$	Y_n	$f_n(X)$	Y^*_n

Y^*_k , the prediction of Y_k is made first by developing a regression equation f_k using all data except (X_k, Y_k) , then by applying that equation to X_k .

This process is repeated iteratively to obtain the complete set of Y^* . Comparison of Y to Y^* is a fair measure of the skill of the (X, Y) regression model as a means to predict some unobserved future value, Y_{n+1} from a new observation set X_{n+1} .

error of a regression overstates predictive ability because, in finding the best-fit equation to the data by minimizing the errors, the result inadvertently “explains” some of the noise in addition to the signal in the sample data.

Once the cross-validation tests have been performed, then all of the data can be used to develop the best regression model for use in predicting Y_{n+1} .

Elsner and Schmertmann (1994) point out that the form of the equation, as well as the parameters, must be incorporated into the cross-validation test for it to be fair. Going back to our regression example, let us assume that in developing our regression equation some of the variables in the set X might not be significant, or might create problems of autocorrelation, such that the final regression equation might make use of only a subset of the variables in X . It would not be fair to use all of the data to decide which variables to retain and which to eliminate, and then perform cross-validation on the data with the selected equation. Instead one must start with all of the variables and go through the process of elimination in developing each of the n equations for the cross-validation test to be fair. Cross-validation properly performed, captures model risk as well as parameter and process risk.

Also, in designing the test it is important to consider whether using data on “both sides” of each (X, Y) value will result in a fair test. Sup-

pose that the data is a time series—for example, the numbers of Atlantic hurricanes occurring in each calendar year over a twenty-year period. In the general case, we would use twenty observation sets of nineteen data points to develop twenty forecasts that could be compared to the actual values. This would include using observations from years after a particular year to forecast the results for that year. Of course, in reality the subsequent observations would not have been available at the time we were making the forecast. This is only a serious problem if the data exhibits serial correlation, typically due to a trend, cyclical, or other nonstationary situation. Where serial correlation is present in time-series data, using adjacent data on both sides provides too much information about the omitted value to the forecast model; in such situations sufficient adjacent data points on one side should be excluded so that the test is fair. Where serial correlation is not present, the observations can be treated as a sample set without order significance. This issue is particularly relevant in the context of performance testing of claim liability estimates, as there is very little data in insurance that does not exhibit serial correlation.

The cross-validation approach goes by a variety of other names, depending on the context. Those involved in developing predictive statistical models might refer to the approach as testing the model using out-of-sample data. Meteorologists refer to cross-validation as “hindcast” test-

ing. In the context of claim liability estimation, some actuaries refer to it as retrospective testing.

3.4. Application of cross-validation to estimates of claim liabilities

The cross-validation approach applies quite naturally to the estimation of claim liabilities. It can be applied to an overall actuarial method, used to develop $\hat{L}_m^{(t)}$; or to a component of the method—for example, one that estimates the claim liabilities for an accident year at a particular maturity, or one that forecasts the expected incremental claim payments for an accident year in a particular development year.

One issue with cross-validation in a claim liability estimation context is whether any data after a point in time can be used to validate estimates of claims before that point in time. As was discussed in the previous section, cross-validation tests require careful design to assure that they fairly assess predictive skill.

In the context of testing an actuarial projection method, the selected actuarial projection method is typically applied to a historical dataset that is limited to information that would have been available at the time, and the resulting estimate of the claim liabilities is compared to the actual outcome with the benefit of hindsight reflecting the actual run-off experience. This assures that the test is a realistic and fair test of the performance of the method. The dataset is then brought forward to the next valuation point, and the estimation process is repeated. Schedule P of the U.S. Annual Statement includes this type of cross-validation test of the historical held reserves.

We illustrate the application of the standard performance testing technique in the case study in Section 4, with additional calculation details in the Appendix.

Finally, we note that in most actuarial projection applications one does not have the luxury of knowing the actual value of the claim liabilities,

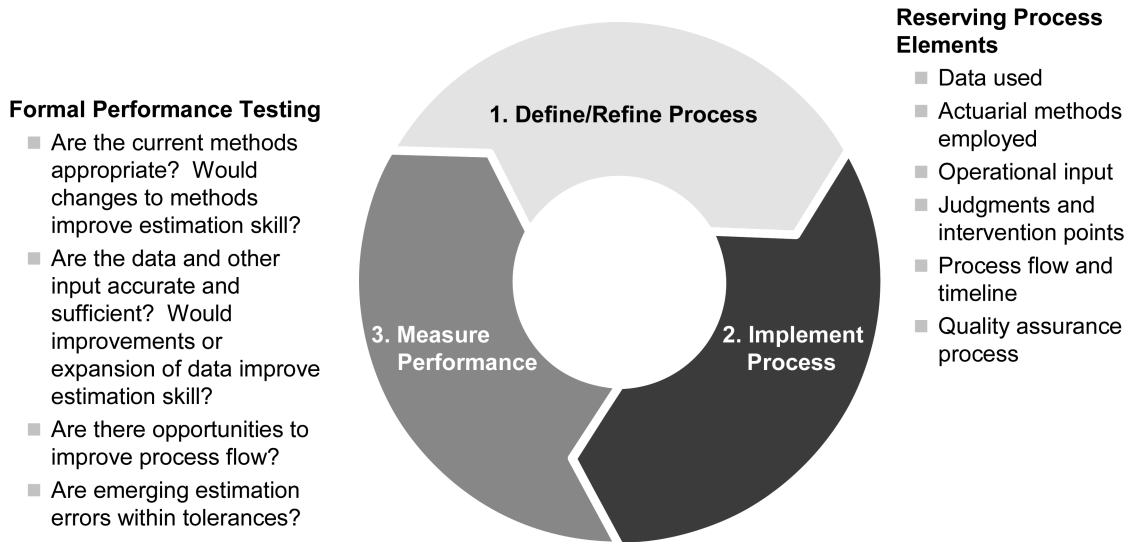
$L^{(t)}$, because the subsequent run-off is not complete. Instead, one typically only knows the actual run-off through the current valuation point, and must combine it with the current actuarial central estimate of any remaining unpaid claim liabilities to obtain a proxy for $L^{(t)}$. This is a practical concession that one must make unless one is prepared to restrict the analysis to very old experience. The validity of the results is a function of the proportion of each “actual” value that is paid versus still estimated unpaid, statistics that should be tracked throughout the analysis. A final step in each performance test exercise will be to be sure that the results and conclusions are not unduly influenced by the current estimates of the remaining liabilities.

3.5. Performance testing as an element of the reserving control cycle

The importance of an actuarial control cycle has been presented by many others. The concept was first presented by Goford (1985) in an address to U.K. actuarial students in the context of life insurance. Since then, the actuarial control cycle has been adopted as an educational paradigm by many actuarial organizations, most notably the Institute of Actuaries in Australia. Most recently, the Casualty Actuarial Society has announced a major revision to its basic education and examination structure, one element of which is the addition of the actuarial control cycle to the new Internet-based exam modules [see for example, Hadidi (2008)]. For a general introduction to the Control Cycle see Bellis (2003).

In the context of loss reserving a control cycle involves identifying, testing, and validating all elements of the process by which claim liabilities are estimated. Figure 2 depicts what the control cycle might look like in this context. Most importantly the control cycle should involve an ongoing assessment of the estimation skill of the actuarial methods currently being employed, and exploration of opportunities to enhance overall

Figure 2. The actuarial control cycle for the claim reserving process



estimation skill by implementing better actuarial projection methods, based on all of the BLURSI-CE criteria. As with other control cycles, the reserving control cycle is critical to the management of reserve risk.

Those with an engineering background will recognize that the actuarial control cycle is a specific application of a more general approach to quality assurance and continuous process improvement. Periodically, in a planned manner, one steps back from the production process to assess its effectiveness and to identify ways to improve it. The control cycle also codifies that, in addition to a production role, the actuary has a developmental role to play. This latter role can often get lost in the crush of production activities.

In the reserving control cycle, performance testing is the primary assessment tool. As part of the control cycle, the actuary would identify those areas where the greatest exposure to estimation errors is present, and seek process improvements to mitigate the exposure to those errors. This can take the form of improvements in the quality, timeliness, and range of data relating to the nature of the underlying exposures and emerging claims, or the introduction of new ac-

tuarial projection methods that make use of new data or better leverage existing data. Occasionally it will take the form of changes to people or systems involved in the claim reserving process.

Implementing performance testing within the claim reserving control cycle can be labor intensive, especially initially when the process is first being established. It is impractical to expect that extensive performance tests would be performed on every possible actuarial method for each class of business, with the detailed analysis updated during every reserving cycle. Fortunately, this is not necessary. Insights gained from the analysis of one class of business often extend to other similar classes. And, since the additional insight gained by updating the data with one new diagonal is incremental, the annual updates for each class do not entail an extensive analysis.

Performance testing can be substantially facilitated by capturing the central estimate from each actuarial projection method, along with the final selected actuarial central estimate, for each class of business in a relational database on an ongoing basis. Reconstructing this information after the fact can be a substantial archaeological exercise, particularly when the methods involve intervention points with judgmental selections. Recon-

struction exercises are often necessary to “seed” the database initially; the database is then updated with new projection results on an ongoing basis. Once this database is established, ongoing performance test results for existing methods can be produced somewhat mechanically. Obviously this will not be the case for new actuarial projection methods, where some “back-casting” of the method is necessary to get an initial view of its performance.

The key to successful implementation of performance testing within the control cycle is a carefully-thought-out test plan that focuses on a few key issues and experiments with a few new methods or new data sources.

4. Illustrative case study

In this section, we move the discussion from concept to practice, by means of a real-world case study based on the actual reserving experience of a major U.S. insurer. While we draw some actual conclusions from the case study, its primary purpose is to illustrate the application of performance testing, cross-validation, and the control cycle. While our work in this area (including performance testing on other datasets) generally confirms the findings in the case study, we can’t say that all of the conclusions drawn from the case study can be generalized to other companies or circumstances.

4.1. The company and its data

The case study is based on the experience of a company that no longer exists in its historical form, having been transformed by a series of subsequent mergers and acquisitions. Since the company was a long-standing client of the authors’ firm, a wealth of historical reserving experience is available. We are appreciative of the successor company’s support in allowing us to publish results based on the historical dataset.

For each reserve segment, accident year loss development data is available for twenty-seven

accident years, from 1972 to 1998, with annual valuations at June 30. Sheet 1 of the Appendix illustrates the complete triangular dataset, displaying reported claim counts. This data facilitates claim liability estimation at each June 30 from 1979 to 1998, a twenty-year span. During this same period, one of the authors provided independent estimates of claim liabilities to the management of the company, so actual historical projections using various methods are also available. Claim liability estimates at each of these twenty valuation points can be compared with “hindsight” estimates based on a special study performed using data through December 31, 2000, that combined actual run-off claim payments with estimates of the remaining unpaid claim liabilities at that time. Thus, for the most recent of the twenty estimates to be tested (i.e., the estimate at June 1998), the hindsight estimate reflects thirty months of actual run-off payments. For the earlier estimates, the hindsight estimate reflects an increasing proportion of actual run-off payments. Assuming that the December 2000 remaining claim liability estimates are a reasonable proxy for the actual remaining unpaid claim liabilities, the claim liabilities at June 1998 are 77% paid; the claim liabilities at June 1997 are 84% paid; the claim liabilities at June 1996 are 92% paid; and the claim liabilities at earlier valuations are between 96% and 100% paid.

For each reserve segment, loss development data consists of the following triangles:

- Reported claim counts (shown on Appendix sheet 1)
- Closed claim counts (combination of closed with payment and closed without payment)
- Paid loss and allocated loss adjustment expenses
- Incurred loss and allocated adjustment expenses (paid plus case reserves)

Fortunately, during this extended period the company’s claim systems did not change in material ways, so the historical data reflect relatively

consistent claim processing, from an IT perspective.

In addition to the claim experience, direct earned premiums by calendar year are also available, allowing us to use methods based on loss ratios.

The company wrote what may be described as “main street business” through independent agents. The data is net of excess of loss reinsurance, at retentions that were relatively low, typically around \$100,000.

The company’s historical experience provides a rich backdrop for illustrating performance testing, for several reasons.

- First, the experience period includes the seventies and early eighties, a period where significant monetary and social inflation caused trends and development patterns to change from their historical levels. Conversely, in the nineties this inflation abated and trends and development patterns stabilized.
- Second, the company made some significant operational changes—particularly to its claim handling procedures—that affected claim closure rates, case reserve adequacy, and the trends in average claim costs across accident years.
- Third, the company’s underwriting posture changed several times during the period. In the early experience years they focused on revenue growth, leading to relatively disastrous underwriting results (in this they were not alone). Eventually it became necessary to re-underwrite the existing business, become more selective about new business, and seek price adequacy.

These factors create some distinct “turning points” in the historical data, creating significant challenges to the claim liability estimation process. While the challenges in the future might take a different form than those observed in the past, the uncertainties created will potentially be of a similar magnitude. Thus we believe perfor-

mance testing over this period is a useful exercise.

The case study focuses on our ability to estimate claim liabilities for Commercial Automobile Bodily Injury Liability (CABI). Estimation for this reserving segment is neither very difficult nor very easy, so it is a good place to start.

For the purposes of the case study we can pretend that it is 2001, and the company’s management has asked the corporate actuary to begin establishing a control cycle (making use of the “final” estimates of claim liabilities as of December 2000). The starting point is a performance testing project focusing on a representative line of business, CABI. The tests will include the two methods currently employed in the reserving process, the paid and incurred chain ladder methods; an alternative method, not currently used by the company, that projects case reserves only; and an additional incurred development method that seeks to adjust for changes in case reserve adequacy. While methods to adjust for case reserve adequacy have been in the literature for years, they have not been implemented at the company due to misgivings about their effectiveness and stability. The project is purposely limited in scope to fit within available resource constraints. As part of the project, a plan for further performance testing is also to be developed based on the initial results from the project.

4.2. Paid chain-ladder development method

The project starts with performance testing of the paid chain-ladder development (PCLD). First, we need to formally define the PCLD method, using our $m(a, d, p)$ construct from the previous section. This is done in summary form in Table 2.

As can be seen in Table 2, this is a highly mechanical method, with little or no intervention, and no exogenous data employed. In essence, this is a test of “the machine,” without the normal dose of actuarial judgment supplied by “the

Table 2. Paid chain-ladder development method

The algorithm	<p>The basic chain ladder approach in which the cumulative paid claim and allocated loss adjustment expenses at each historical maturity is divided by the cumulative paid value at the preceding maturity to obtain a triangle of historical report-to-report development factors. The simple average of the latest four observations⁸ is calculated at each maturity. A tail factor to provide for development beyond 150 months (the last available maturity in the triangle) is based on ratios of the current value of reported claim and allocated expenses to the paid claims and allocated expenses at 150 months for the latest four fully mature years, if available.</p> <p>A “benchmark” set of factors is used to obtain report-to-report factors to the extent that the observed development data does not extend to 150 months.</p> <p>The selected factors are chained together to produce report-to-ultimate development factors. The development factor from 6 months to 18 months is divided by two to account for the accident half-year. Cumulative paid losses as of the current valuation are subtracted from the projected ultimates to obtain estimates of unpaid claim liabilities.</p>
The data	Cumulative paid claim and allocated loss adjustment expenses, in traditional triangular format, with valuations from 6 months to 150 months. Reported claim and allocated expenses as of the current valuation for accident years beyond 150 months. No other data is employed.
Intervention points	Other than selecting the benchmark factors, no interventions are made.

⁸The choice to use the simple average of latest four factors is meant to be illustrative only. In actual practice one could also test the use of weighted averages, including fewer or more observations, excluding high and low values, etc.

man.” This is a scope limitation to the initial performance testing project; performance tests of the actuarial judgments are deferred to subsequent control cycles, after some experience with the performance of the basic methods has been gained.

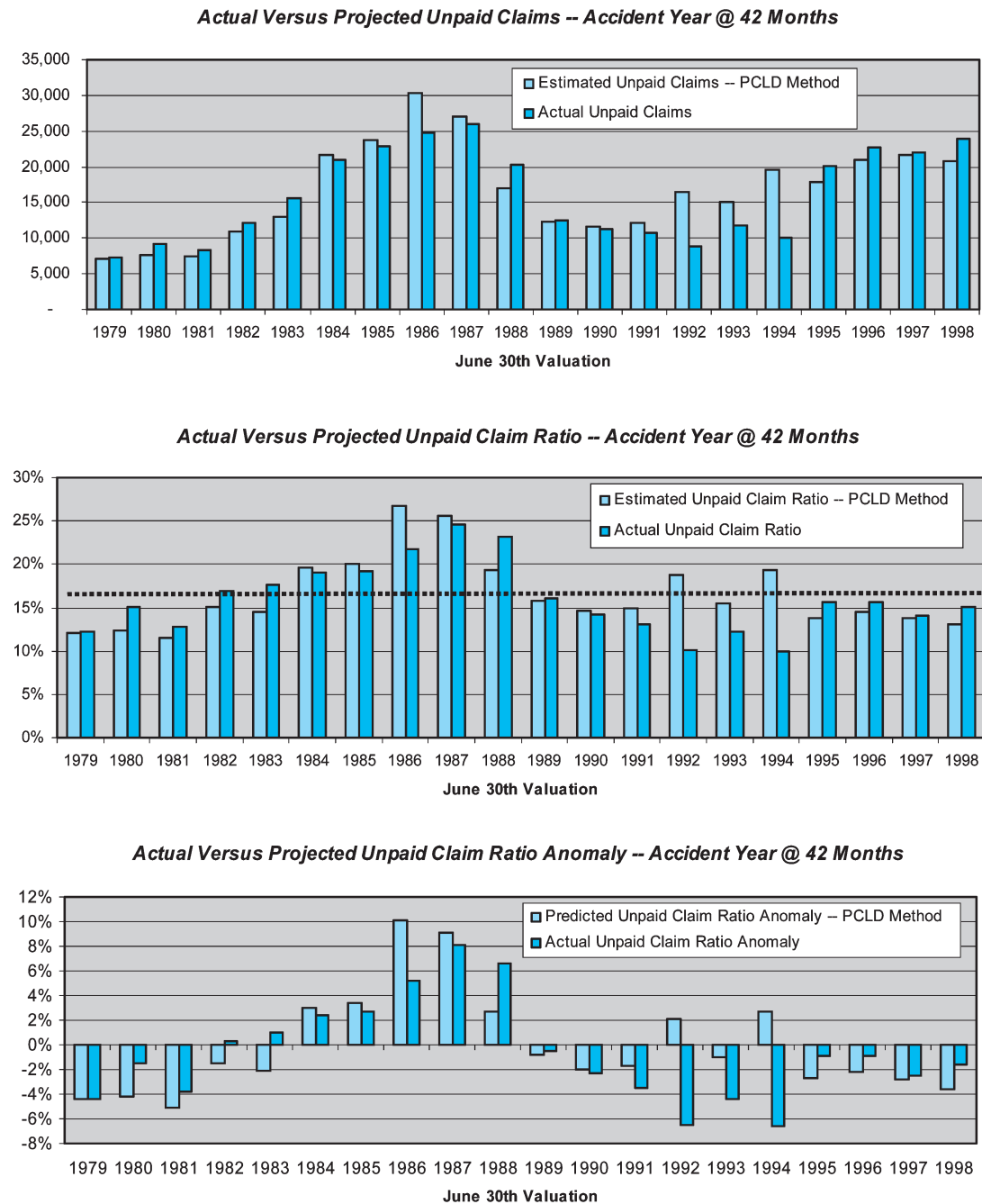
The PCLD method was applied to the available CABI data at each midyear valuation point from June 1979 to June 1998 to obtain estimates of the unpaid claims, by accident year, at each valuation. Application of the PCLD method at the June 1984 valuation is illustrated on sheet 2 of the Appendix.

Many actuaries find it convenient to vary actuarial projection methods by accident year, using one method for mature years, a different method for immature years, and perhaps yet another method for the current year. It will therefore be important to look at performance test results by maturity, to assist the actuary in his or her choices of method by maturity. Rather than focusing initially on the overall performance of the PCLD method, our analysis therefore starts with a single component of the claim liability, the performance results for the accident year at 42 months maturity. (The choice of 42 months is arbitrary.)

Figure 3 displays the performance of the PCLD method as an estimator of the unpaid claims for the accident year at 42 months maturity (i.e., the 1976 accident year at the June 1979 valuation, the 1977 accident year at the June 1980 valuation, etc.) The results in Figure 3 are designed to answer the question “How accurate is the paid chain-ladder development method as an estimator of unpaid claim liabilities when applied to the actual cumulative paid claims at 42 months maturity?”

The uppermost of the three bar graphs in Figure 3 compares the estimated unpaid claim liabilities to the actual unpaid claim liabilities at each valuation point. (Note that, by “actual,” we mean the unpaid claims with the benefit of hindsight through December 2000, consisting of actual payments through December 2000 plus estimated unpaid claims at that point.) One can see that there is some degree of correspondence between the PCLD estimate and the actual liabilities, although at some valuations (1986 and 1992–1994) the correspondence is not terribly strong.

The middle of the three bar graphs converts the predicted dollars of unpaid claim liabilities into unpaid claim ratios, by dividing them by

Figure 3. CABI performance test results—Paid chain-ladder method

the earned premium for the calendar year. This essentially normalizes for variations in the volume of business. The dotted horizontal line is the average unpaid claim ratio, which is approximately 16%. (Note that the premium being used is the combined liability premium for both bodily injury and property damage; these are only the

bodily injury portion of the loss ratio.) The overall picture isn't materially different than the upper bar graph, partially because the company had the misfortune to maximize its volume of business in the late eighties, when loss ratios were highest. The relative heights of the pairs of bars are largely unchanged.

Finally, the lowermost bar graph of Figure 3 simply adjusts the unpaid claim ratios by subtracting the 17% average unpaid claim ratio. The result is to express the projections in terms of their ability to detect deviations from the average. This last presentation is relevant because the skill of a forecasting method is typically defined as

$$Skill_m = 1 - mse_m/msa \quad (4.1)$$

where mse_m is the mean squared error, the average squared difference between the actual unpaid claim ratio and the predicted unpaid claim ratio from the method; and msa is the mean squared anomaly, the average squared difference between the actual unpaid claim ratio for that observation and the overall average actual unpaid claim ratio across the entire test period. The mse_m is specific to the actuarial method, while the msa is an intrinsic property of the experience. The method with the minimum mean squared error will therefore have the maximum skill.

Does the form of the measure of skill look familiar? In a regression context it would be expressed as

$$R^2 = 1 - sse/sst, \quad (4.2)$$

where sse is the squared difference between the actual values and those predicted by the regression equation, and sst is the squared difference between the actual values and their mean. Just as R^2 is a statistical measure of amount of variation captured by the regression equation, $Skill_m$ is a measure of the amount of variation captured by the particular actuarial method.

Those familiar with credibility will also recognize that there is some similarity of the definition of skill to that of credibility.

From the definition of skill, one can recognize that skill is equal to zero when the mse_m is equal to the msa , and negative when the mse_m is greater. In these cases one would be better off multiplying the earned premium by the long-term average unpaid claim ratio to estimate the

unpaid claim liabilities, rather than using the method being tested.⁹ Conversely, when the mse_m is zero, the method perfectly predicts the actual result and skill is 100%.

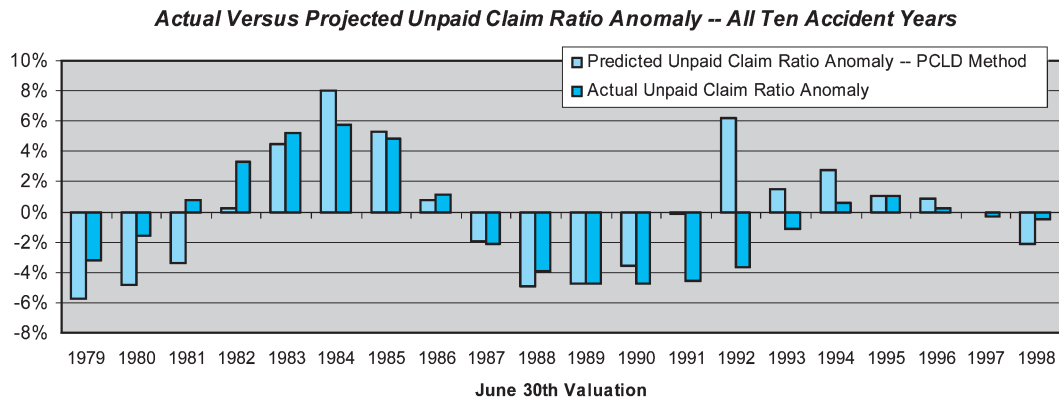
Returning to the lowermost bar graph on Figure 3, one can observe the actual unpaid claim ratio anomalies directly. They reflect the pattern of the underwriting cycle during this era. The actual unpaid claim ratio anomaly is contrasted with the predicted anomaly. The errors are the differences between the pairs of bars, reflecting our ability to predict the direction and magnitude of the anomaly. While it is not a fundamental transformation, removing the mean improves the visual display, allowing one to see the errors more clearly. This lower chart is the typical means of displaying performance test results.

The skill of the paid chain-ladder method in estimating the unpaid claims at 42 months for an accident year is only about 21%. This result is heavily driven by the poor performance of the method at the June 1986, 1992, 1993, and 1994 valuations. At each of these valuations, the PCLD projections look reasonable based on the information available at the time; however, subsequent paid claim development is markedly different than the history up to that point.

The data underlying Figure 3, along with the details of the calculated skill of 21%, are presented on sheet 6 of the Appendix.

Skill can be measured at other maturities, using the same approach as is depicted in Figure 3. It turns out that the PCLD method does not exhibit much skill in this case. Skill is actually at its highest, at 21%, at 42 months maturity. The skill of the PCLD method in estimating an accident year's unpaid claims at both earlier and later maturities is actually negative. For the early maturities, the size of the development factors becomes

⁹In essence, the skill is measured in relation to a naïve paid Bornhuetter-Ferguson method, in which the expected loss ratio and payment pattern are constant long-term averages.

Figure 4. CABI performance test results—Paid chain-ladder method

so large that the estimates become very volatile. For the very mature years the remaining unpaid claim liabilities are small and highly dependent on the numbers and nature of the claims that remain to be settled, so the volume of cumulative paid claims is just not a very good indicator of the remaining unpaid claims. Many readers will recognize these as being typical shortcomings of the PCLD method.

It is easy to extend the skill measure to the overall estimate of unpaid claims across all accident years, by merely taking the weighted average of the individual accident year unpaid claim ratios. Rather than using all accident years, we have elected to use only the latest ten accident years, so that the metric is relatively consistent over time. The results are shown on Figure 4.

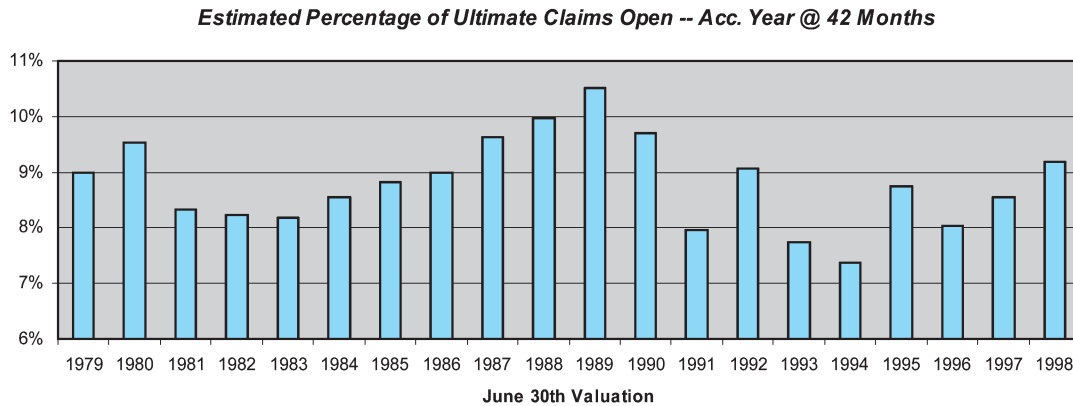
Our performance test results indicate that the overall skill of the paid development method in estimating unpaid claim liabilities for all ten accident years combined is only about 13%. The low skill is driven by the large prediction errors at the 1981, 1991, and 1992 valuations. At each of these valuations the PCLD projections look very reasonable; however, the subsequent development looks markedly different than the history up to that point.

The PCLD method also exhibits some shortcomings against our other BLURS-ICE criteria. Over the twenty-year performance test period,

the method exhibited a positive bias (predicted dollars of claim liabilities above actual liabilities) of about 4%. Given the volatility of the prediction errors, this is likely to simply reflect sampling error. The method is also unresponsive to changing conditions; not surprisingly, it tends to react in a lagged manner to changes in the payment pattern.

One reason for the low skill of the PCLD method in this particular case is the nonstationarity of the claim settlement process over the historical period, as evidenced by the changing percentages of claims open in Figure 5. This figure suggests that an adjusted paid method, reflecting the changing claim settlement rates, should offer improved skill over the simple method we have employed here.

Overall the PCLD performance tests suggest that this method is of little value, with very low skill, due to the changing patterns of claim settlements. The skill of the PCLD on this product line would be higher in more normal circumstances, when settlement patterns are more stable. Performance tests of methods that adjust for changes in settlement patterns could be undertaken. At a minimum, simple diagnostics that detect changing settlement patterns should be introduced into the reserving process so that the actuary can know when the PCLD is not likely to produce a reasonable estimate.

Figure 5. CABI—Percentage of estimated ultimate claims open at 42 months**Table 3. Incurred chain-ladder development method**

The algorithm	<p>The basic chain ladder approach in which the cumulative incurred (paid plus adjusters' individual claim estimates of unpaid) claim and allocated loss adjustment expenses at each historical maturity is divided by the cumulative incurred value at the preceding maturity to obtain a triangle of historical report-to-report development factors. The simple average of the latest four observations is calculated at each maturity. A tail factor to provide for development beyond 150 months (the last available maturity in the triangle) is based on ratios of the current value of reported claim and allocated expenses to the reported value at 150 months for the latest four mature years.</p> <p>A "benchmark" set of factors is used to obtain report-to-report factors to the extent that the observed development data does not extend to 150 months.</p> <p>The selected factors are chained together to produce report-to-ultimate development factors. The development factor from 6 months to 18 months is divided by two to account for the accident half-year. Cumulative incurred claims are subtracted from the projected ultimates to obtain indicated IBNR liabilities, to which are added the current case reserves.</p>
The data	Cumulative incurred claim and allocated loss adjustment expenses, in traditional triangular format, with valuations from 6 months to 150 months. Reported claim and allocated expenses as of the current valuation for accident years beyond 150 months. Case reserves as of the current valuation. No other data is employed.
Intervention points	<p>If any of the calculated simple average report-to-report factors are less than one, the value is set equal to one.</p> <p>Benchmark factors are selected to complete the development when the triangle does not extend to 150 months.</p>

4.3. Incurred chain-ladder development method

We turn next to performance testing of the chain-ladder method applied to incurred (paid plus case reserves) claim development data (ICLD). Once again we start by formally defining the method, as summarized in Table 3.

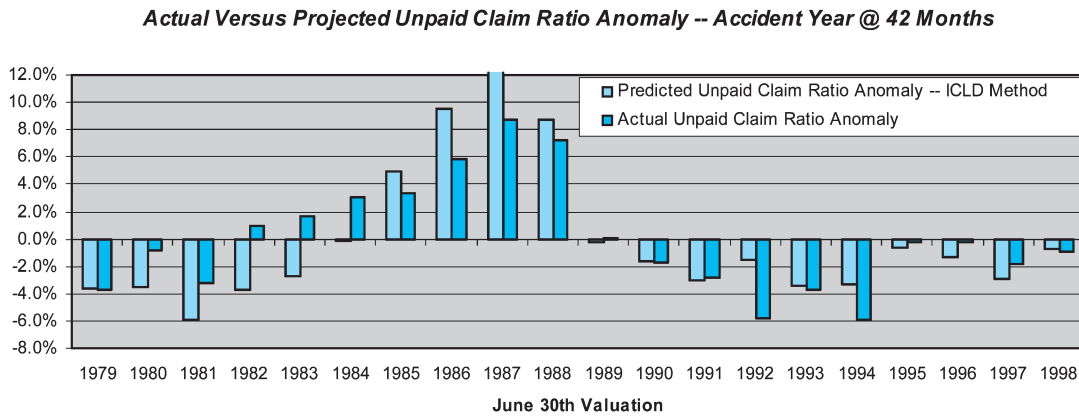
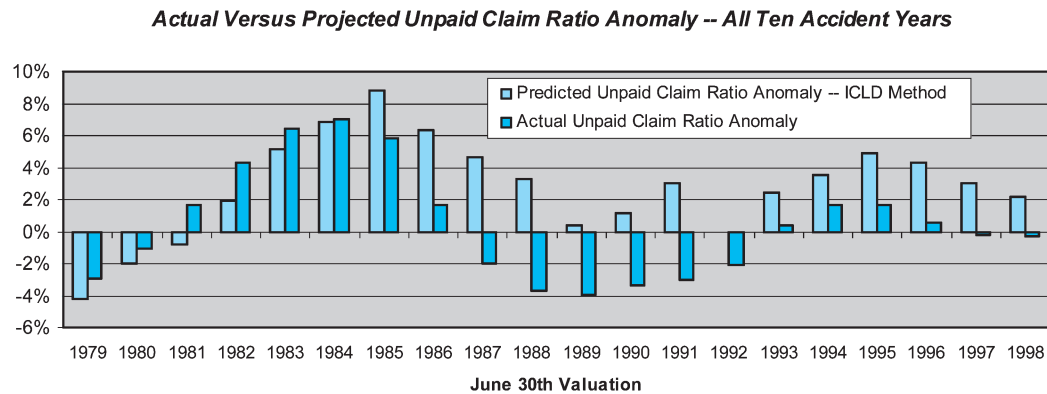
As was the case with the paid method, our approach is insular and largely mechanical, without benefit of judgment or confirming tests based on exogenous data.

We applied the ICLD method to the CABI incurred development data at each midyear valuation point from June 1979 to June 1998 to ob-

tain estimates of unpaid claims at each valuation point. (More specifically, we used the ICLD method to estimate the Incurred But Not Reported [IBNR] liabilities, and then added the case reserves to get total estimated unpaid liabilities.)

Application of the ICLD method at the June 1984 valuation is illustrated on sheet 3 of the Appendix.

The ICLD performance test results for the accident year at 42 months are shown in graphical form in Figure 6. This graphic is analogous to the lower graphic in Figure 3, comparing predicted unpaid claim ratios to actual unpaid claim ratios with the benefit of hindsight.

Figure 6. CABI performance test results—Incurred chain-ladder method**Figure 7. CABI performance test results—Incurred chain-ladder method**

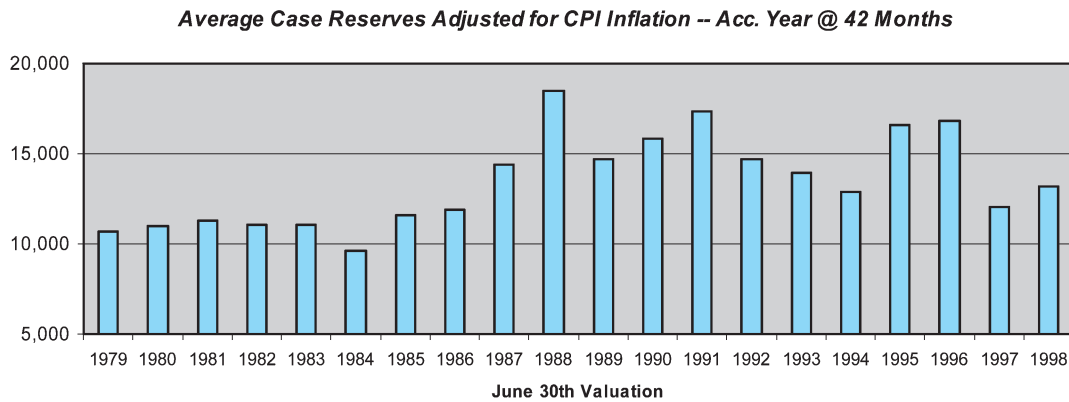
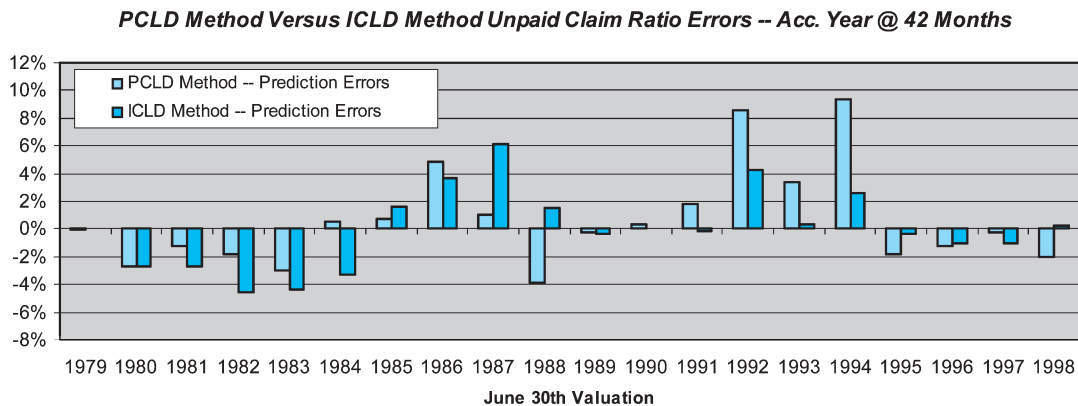
As can be seen in Figure 6, the ICLD method performs poorly in the 1981–1984 period, underestimating the unpaid loss ratio. It then performs poorly in the 1986–1987 period, overestimating the unpaid loss ratio. In 1992–1994 it again overestimates the unpaid loss ratio.

The ICLD method exhibits better skill than the PCLD method, but its absolute skill is still relatively low. The observed skill of the ICLD method at 42 months maturity is 52%, versus 23% for the PCLD method. The skill of the ICLD method is negative for the most recent (18 months) maturity and for the older maturities.

Overall the skill of the ICLD method for the last ten accident years combined is 31%, versus 13% for the paid chain-ladder development method. Figure 7 displays the overall performance test results for the ICLD method.

Over the twenty-year performance test period the ICLD method exhibited a negative bias (predicted dollars of claim liabilities below actual liabilities) of 0.8%. However, this is not likely to be a statistically significant result.

One reason for the poor skill of the ICLD method is the changing levels of case reserve adequacy over the historical period, due to the changes in claim handling procedures mentioned earlier. Figure 8 displays the average case reserve on open claims for the accident year at 42 months, adjusted to a constant-dollar basis using the U.S. Consumer Price Index (CPI). As can be seen, the average case reserve for the accident year at 42 months increased substantially during the period from 1984 to 1988. While somewhat volatile, the average case reserve also appears to have declined from its 1988 peak level during

Figure 8. CABI inflation-adjusted average case reserves**Figure 9. CABI comparison of paid and incurred chain-ladder errors**

the early 1990s. This suggests that an adjusted incurred method, reflecting changes in case reserve adequacy, would offer improved skill over the simple incurred chain ladder method. We test such a method later in this section.

One can compare the PCLD and ICLD methods directly by looking at their prediction errors, as is shown in Figure 9. Inspection of the errors shows that the root cause of the performance difference between the PCLD and the ICLD is the performance of the former in the years 1992 and 1994. These two years contribute significantly to the mean squared error of the paid chain-ladder method. Prior to this period, the company engaged in a significant effort to reduce the inventory of open claims. This manifested itself as higher-than-average paid development, which translated into several diagonals of above-aver-

age paid development factors. As these factors worked their way through the average, the paid chain-ladder forecasts significantly overshot the mark.

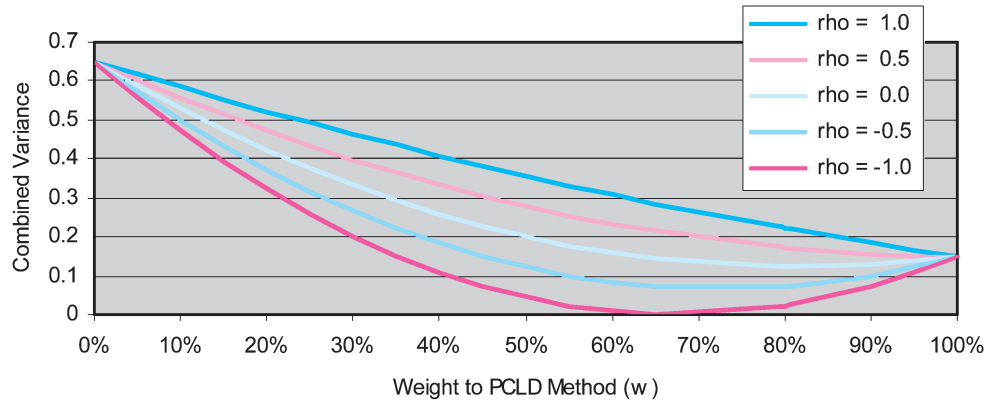
Looking at the errors more broadly, one can see that they are similar during some periods and only occasionally divergent.

Before continuing with performance tests of other methods, we diverge to look at the issue of combining estimates from multiple methods.

4.4. Combining estimates from multiple methods

We now have predictions from two actuarial projection methods, PCLD and ICLD, with measures of skill for each. The next question to be addressed is how to combine the results of the two methods together to create a single actuar-

Figure 10. Illustration of combined variance of two methods



ial central estimate. We want to combine the two estimates according to

$$\hat{L}_C^{(t)} = w \times \hat{L}_I^{(t)} + (1 - w) \times \hat{L}_P^{(t)}. \quad (4.3)$$

We want to choose w so that the variance of the error distribution around $\hat{L}_C^{(t)}$ is minimized. The variance of the error distribution around $\hat{L}_C^{(t)}$ is given by

$$\begin{aligned} V_C &= w^2 \times V_I + 2 \times w \times (1 - w) \\ &\quad \times \text{Cov}(\hat{L}_I^{(t)}, \hat{L}_P^{(t)}) + (1 - w)^2 \times V_P \\ &= w^2 \times V_I + 2 \times w \times (1 - w) \times \rho_{I,P} \\ &\quad \times \sigma_I \times \sigma_P + (1 - w)^2 \times V_P \end{aligned} \quad (4.4)$$

where V_I, V_P, V_C are the variances of the estimation errors for $\hat{L}_I^{(t)}, \hat{L}_P^{(t)}, \hat{L}_C^{(t)}$, respectively, σ_I, σ_P are the standard deviations of the estimation errors for $\hat{L}_I^{(t)}, \hat{L}_P^{(t)}$, respectively, and $\rho_{I,P}$ is the correlation coefficient between $\hat{L}_I^{(t)}, \hat{L}_P^{(t)}$.

To minimize the error distribution around $\hat{L}_C^{(t)}$ we must merely take the derivative of V_C with respect to w , set the derivative equal to zero, and solve for w . Doing so we get

$$w = \frac{V_P - \rho_{I,P} \times \sigma_I \times \sigma_P}{V_I - 2 \times \rho_{I,P} \times \sigma_I \times \sigma_P + V_P}. \quad (4.5)$$

One can see from Equation 4.5 that if the variances of the error terms are equal, then w is equal to 50%. If the variances of the error terms are not

equal, greater weight will be given to the method with the lower error variance, depending on the correlation.

Figure 10 provides additional insight to the relationship between the two variances and the correlation. In this example, $V_P = .65$ and $V_I = .15$. The chart shows how V_C varies with w for a given level of correlation (“rho” in the chart), with the minimum combined error variance being the lowest point on each curve. If there is perfect positive correlation between the estimation errors ($\rho_{I,P} = 1.0$), then the combined error variance is minimized by simply giving 100% weight to the method with the lower error variance. As correlation decreases, one can see that the minimum combined error variance entails giving some small weight to the method with the higher error variance. At the extreme, if the estimation errors for the two methods are negatively correlated, then they can be perfectly offset to produce zero combined error variance by using weights from Equation 4.5—in this case approximately a 67.6% weighting to the lower variance estimate.

In the specific case of our CABI results for the accident year at 42 months, we have the following.

- The *mse* of the PCLD method is 0.1249%.
- The *mse* of the ICLD method is 0.0753%.
- The observed correlation between the errors of the two methods is 60.3%.

Table 4. Case reserve development method

The algorithm	<p>The starting point is a triangle of historical case reserves (i.e., adjusters' individual claim estimates of unpaid claims). From that triangle a triangle of historical case reserve development factors is constructed, for all historical values excluding the latest diagonal. For accident years beyond 150 months maturity, the current values of incurred losses are selected as ultimate. The case reserve development factors are calculated for all earlier valuations of those accident years by subtracting the paid losses at each valuation from the selected ultimate, and dividing by the case reserves at the same valuation. Working successively forward by accident year, the case development factor to apply to the next maturity is calculated by taking the weighted average of the latest ten observations. The projected ultimate for that year is then used to calculate case reserve development factors at all prior maturities.</p> <p>In calculating the case reserve development factors at 6 months maturity, the ultimate losses are divided by two to account for the accident half-year.</p> <p>A "benchmark" set of case reserve development factors, inferred from the benchmark paid and incurred chain ladder factors, is used to the extent that the observed development data does not extend to 150 months, and also to the extent that the case development factors exhibit high volatility.</p>
The data	Case reserves and cumulative paid claim and allocated loss adjustment expenses, in traditional triangular format, with valuations from 6 months to 150 months. Reported claim and allocated expenses as of the current valuation for accident years beyond 150 months. No other data is employed.
Intervention points	Benchmark factors are selected to complete the development when the triangle does not extend to 150 months.

Assuming that these were the only two methods available and we believed that the historical performance test results were reasonable indicators of the current situation, then the indicated minimum-variance weighting would be 79.5% to the estimate from the ICLD method with the balance given to the estimate from the PCLD method. The *mse* of the weighted average is 0.0719%, an improvement over either of the stand-alone methods.

The weighting approach derived in this section could obviously be applied by individual accident year with varying weights, or to the overall estimates from each of the methods using weights based on the variances and correlations between the overall errors of the methods. It is easily extendable to more than two methods, by expansion of the combined variance equation to:

$$V_C = \sum_{i=1}^n w_i^2 \times V_i + 2 \times \sum_{i=1}^n \sum_{j=1, j>i}^n w_i \times w_j \times \rho_{i,j} \times \sigma_i \times \sigma_j \quad (4.6)$$

Again, to minimize the combined variance one must take the derivative of V_C with respect to w_i , subject to the constraint that $\sum_{i=1}^n w_i = 1$.

As the number of actuarial projection methods gets larger, the required number of estimates

certainly grows, including the estimates of the variance of each method and the correlation coefficient between each pair of methods.

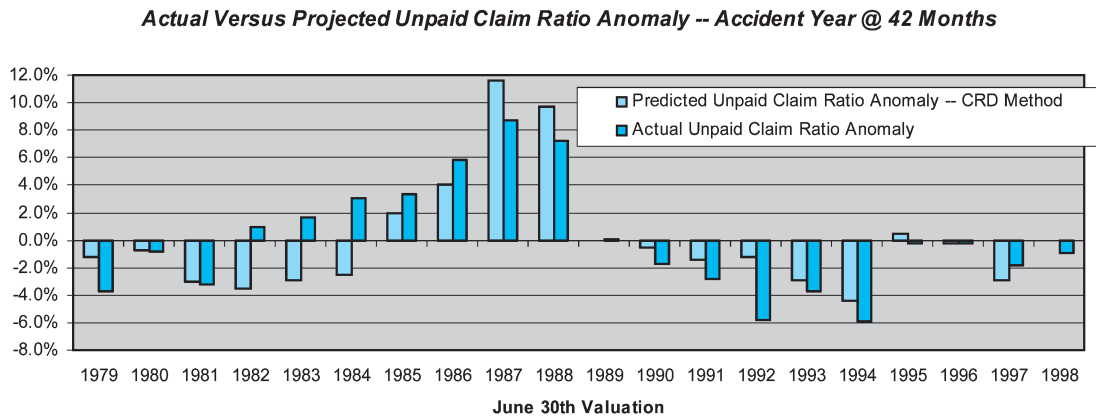
4.5. Case reserve development method

The next actuarial projection method we tested was the case reserve development method (CRD), in which a development factor is applied to the unpaid case-basis claim reserves for each accident year to obtain the estimated unpaid claim liability. The CRD method is a recursive approach, in that the projections from earlier accident years are used to calculate the factor to be applied to the next accident year, successively. The CRD method is formally defined in Table 4.

Application of the CRD method at the June 1988 valuation is illustrated on sheet 4 of the Appendix.

Performance test results for the CRD method for the accident year at 42 months maturity are shown in Figure 11.

The CRD method exhibits better skill than either the paid or incurred development method at older maturities. The observed skill of the CRD method at 42 months is 58%, versus 52% and 23% for the ICLD and PCLD methods, respectively. The observed skill of the CRD method at

Figure 11. CABI performance test results—Case reserve development method**Table 5. Reported claim count chain-ladder development method**

The algorithm	<p>The basic chain ladder approach in which the cumulative reported claim counts at each historical maturity is divided by the cumulative reported value at the preceding maturity to obtain a triangle of historical report-to-report development factors. The simple average of the latest four observations is calculated at each maturity. It is assumed that there is no development beyond 150 months maturity.</p> <p>The selected factors are chained together to produce report-to-ultimate development factors. The development factor from 6 months to 18 months is divided by two to account for the accident half-year.</p> <p>A “benchmark” set of factors is used to obtain report-to-report factors to the extent that the observed development data does not extend to 150 months.</p>
The data	Cumulative reported claim counts, in traditional triangular format, with valuations from 6 months to 150 months. No other data is employed.
Intervention points	Other than selecting the benchmark factors, no interventions are made.

54 months is even higher, at 67%. Similar results have been obtained by the authors in performance test results for other liability lines. It appears that for mature years, in liability lines the case reserves alone are a better predictor than the incurred (i.e., paid plus case reserves) claims. This makes some intuitive sense, as the case reserves relate directly to future claim payments on open claims, whereas the cumulative paid claims relate to claims that are closed which are unrelated to future claim payments. The skill difference is particularly notable when the remaining case reserves are variable, significant in some accident years and less significant in others.

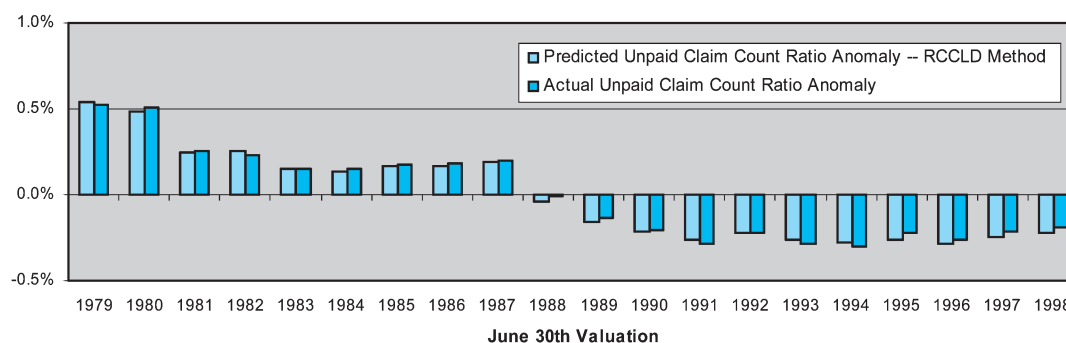
The overall skill of the CRD method across the latest ten accident years is 22%, substantially lower than the 32% skill of the ICLD method, primarily because the CRD method has very little skill for the early maturities.

4.6. Reported claim count chain-ladder development

Since projecting claim counts does not (by itself) yield an estimate of unpaid claim liabilities, it does not technically qualify as an actuarial projection method. Nonetheless, we tested the performance of the chain-ladder applied to reported claim counts (RCCLD). Once again we use a highly mechanical method, as outlined in Table 5 and illustrated at the June 1998 valuation on sheet 1 of the Appendix.

Application of the RCCLD method is illustrated for the June 1984 valuation on sheet 5 of the Appendix.

Figure 12 displays the performance test results for the RCCLD method for the accident year at 42 months. The RCCLD method exhibits remarkably better skill than any of the other meth-

Figure 12. CABI performance test results—Reported count chain-ladder method**Actual Versus Projected Unpaid Claim Count Ratio Anomaly -- Accident Year @ 42 Months**

ods tested, primarily due to the stability in the claim reporting process over the study period. Measured skill at 42 months is 99%. While case reserving policy and settlement practices changed dramatically, the reporting of claims into the claim system did not.

The skill of the RCCLD method is high at all maturities, ranging from a low of 95% at six months to a high of 99% at 42 months and beyond.

The clear implication is that the claim count data has very high information value. Methods that make use of the count data should therefore have enhanced predictive skill over methods that use only the dollar data.

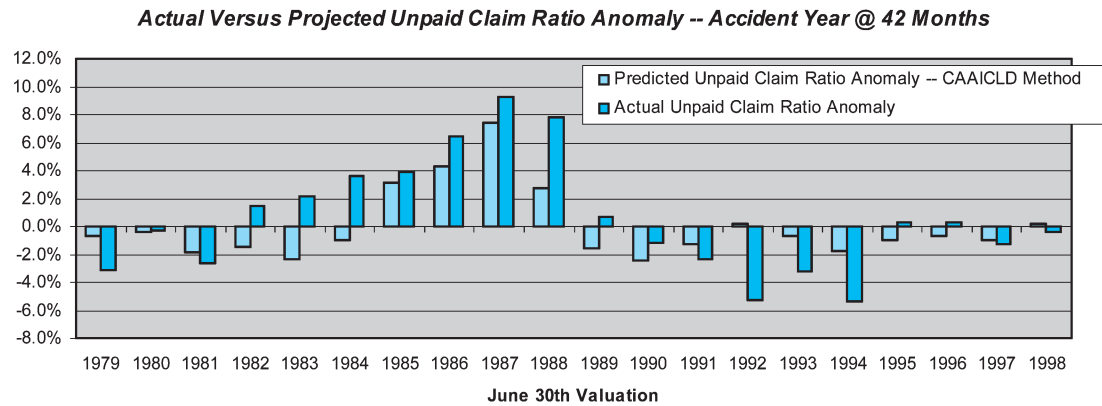
4.7. Incurred chain-ladder with case reserve adequacy adjustment

As was noted earlier, the historical experience exhibits significant changes in case reserve adequacy at various points in time, reflecting changes in case reserving policies at the company. We therefore have tested the performance of a method that measures and adjusts for case reserve adequacy before applying the incurred chain-ladder method. We will refer to the method as case-adequacy adjusted incurred chain-ladder development (CAAICLD). The method is summarized in Table 6, and illustrated at the June 1984 valuation on sheet 5 of the Appendix.

For comparison purposes, we start by analyzing the performance results for the accident year valued at 42 months maturity. Results are displayed in Figure 13.

The skill of the CAAICLD method at 42 months maturity is 52%, exactly the same as the unadjusted ICLD method. However, a comparison of Figure 13 with Figure 6 shows that the estimation errors occur at different points in time. The correlation of the estimation errors is only 37%. At this maturity, the CAAICLD method reduces the estimation errors caused by the changes in case reserve adequacy, but introduces new errors due to the imperfections of the case adequacy adjustment process.

However, the overall skill of the CAAICLD across the latest ten accident years is substantially higher than the unadjusted ICLD method, 52% versus 31%. While skill is not enhanced at 42 months, it is substantially enhanced at 18 and 30 months. The reasons for the overall skill enhancement can be seen in Figure 14. The overall magnitude of the estimation errors is reduced by the case adequacy adjustment. Whereas the unadjusted method produced estimation errors greater than six loss ratio points three times, the adjusted method never produces errors of that magnitude. The adjusted method is far from perfect, as it

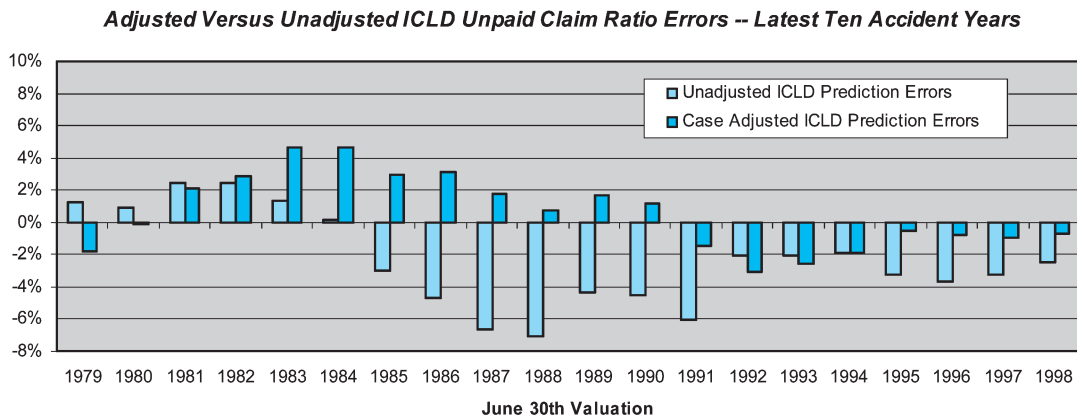
Figure 13. CABI performance test results—Case adequacy incurred chain-ladder**Table 6. Case-adequacy adjusted incurred chain-ladder development method**

The algorithm	<p>Average case reserves are calculated by accident year and maturity for the latest six diagonals of loss development data. A weighted average across all maturities except six months is computed for each diagonal to obtain an overall average case reserve at each valuation point. To prevent the overall averages from being influenced by a changing mix of claims by maturity, the weighted averages are all calculated using the current mix of claims as weights.</p> <p>The annual increases in the weighted average case reserve are calculated and compared to an economic inflation measure. In this case we used the U.S. Consumer Price Index (CPI). For each diagonal of case reserves, the actual case reserves are adjusted by a factor to reflect the extent to which the weighted average case reserve is increasing faster or slower than the inflation rate. The factors are indexed to the current year, so that only the interior values are changed; the latest diagonal of values is unchanged.</p> <p>A similar adjustment is made to the case reserves at six months maturity; the current average case reserve is trended up the column at the inflation rate.</p> <p>The basic chain ladder approach is then applied to the adjusted incurred development data. The simple average of the latest five observations is calculated at each maturity. A tail factor to provide for development beyond 150 months (the last available maturity in the triangle) is based on ratios of the current value of reported claim and allocated expenses to the reported value at 150 months for the latest four mature years.</p> <p>The selected factors are chained together to produce report-to-ultimate development factors. The development factor from six months to 18 months is divided by two to account for the accident half-year.</p> <p>A “benchmark” set of factors is used to obtain report-to-report factors to the extent that the observed development data does not extend to 150 months.</p>
The data	Cumulative paid, case reserves and reported claims and allocated loss adjustment expenses, in traditional triangular format, with valuations from six months to 150 months. Open claim counts in the same format. Reported claim and allocated expenses as of the current valuation for accident years beyond 150 months. No other data is employed.
Intervention points	<p>If any of the calculated simple average report-to-report factors are less than one, the value is set equal to one.</p> <p>Benchmark factors are selected to complete the development when the triangle does not extend to 150 months.</p>

produces errors in excess of four loss ratio points twice in the test period. The results might be further improved by the use of an inflation index that was more relevant to Commercial Auto Bodily Injury claim costs, or by other alternative adjustment techniques than those we employed.

4.8. Summary conclusions from the case study

The case study presented in this section only begins to scratch the surface of performance testing. Test results for many additional methods, and variations on each one, could have been presented—making for a very long paper indeed.

Figure 14. CABI comparison of adjusted and unadjusted ICLD errors

In addition, it is hard to present the test results in a paper format, as much of the real insight comes from poring over the details of the calculations. Our goal was merely to illustrate the methodology and demonstrate the potential insights that can be gained from it, making the case for incorporating it into an ongoing control cycle for the reserving process. The key takeaways we would highlight from the case study are the following:

- The accuracy, or skill, of actuarial methods can be formally tested using the cross-validation approach employed in other predictive sciences. Formal performance testing can help the actuary to understand the strengths and weaknesses of each method, and more realistically assess the potential estimation errors arising from them.
- Performance testing can guide the actuary in the selection of methods, both overall and by maturity. It also provides a methodology for combining the estimates produced by different methods. Correlation between the methods is relevant to the selection of methods, and the weighting scheme used to combine them.
- The accuracy of the basic chain-ladder methods is seriously degraded when there are changes in underlying claim policies and processes. The level of skill indicated by the case study is lower than one would expect absent these changes. These changes are often de-

tectable by analysis of claim settlement rates or average case reserves, which require claim count data. Adjusting for changes in underlying claim policies and processes enhances skill.

- Claim count development is often more predictable than claim dollar development, especially when there are changes in claim procedures and processes. In these situations, methods that make use of the claim count data are likely to improve the accuracy of the estimates.
- Experimentation with less traditional methods may lead to enhancements in skill. Our experimentation with the case reserve development method (including tests not covered in this paper) suggests that it does outperform incurred chain-ladder in many liability lines.

Areas for further exploration would obviously include testing of methods that adjust for changes in claim settlement rates as well as changes in case reserve adequacy. In addition, tests of more complex methods, including regression-based and stochastic methods, would be of interest.

Finally, it would be useful to test whether overall skill is higher when the Bodily Injury Liability claims are combined with the Property Damage Liability claims into a single projection, versus separate projections on each. While homogeneity of the claims is a desirable property—arguing for greater granularity—one of the authors believes

that greater consolidation often leads to greater credibility (i.e., reduced noise in the data) and improved skill. Further performance testing will help to find the optimal trade-off between improved homogeneity and improved credibility.

References

- Barnston, A., and H. van den Dool, "A Degeneracy in Cross-Validated Skill in Regression-Based Forecasts," *Journal of Climate* 6, 1993, pp. 963–977.
- Bellis, C., J. Shepherd, and R. Lyon, *Understanding Actuarial Management: The Actuarial Control Cycle*, Sydney: Southwood Press, 2003.
- Conger, R., and S. Lowe, "Managing Overconfidence," *Emphasis* 3, 2003, pp. 6–9, www.towersperrin.com/emphasis.
- Elsner, J., and C. Schmertmann, "Assessing Forecast Skill through Cross Validation," *Weather and Forecasting* 9, 1994, pp. 619–624.
- Gray, W., and P. Klotzbach, "Extended Range Forecast of Atlantic Seasonal Hurricane Activity and U.S. Landfall Strike Probability for 2008," 2008, <http://typhoon.atmos.colostate.edu/>.
- Goford, J., "The Control Cycle: Financial Control of a Life Assurance Company," *Journal of the Institute of Actuaries Students' Society* 28, 1985, pp. 99–114.
- Hadidi, N., "CAS Board Approves Changes to the Basic Education and Examination Structure," *The Actuarial Review* 35:6, 2008, pp. 14–15.
- Michaelson, J., "Cross-Validation in Statistical Climate Forecast Models," *Journal of Climate and Applied Meteorology* 26, 1987, pp. 1589–1600.
- Wacek, M., "A Test of Clinical Judgment vs. Statistical Prediction in Loss Reserving for Commercial Auto Liability," *Casualty Actuarial Society Forum* Winter 2007, pp. 371–404.

Appendix

ABC INSURANCE COMPANY
 COMMERCIAL AUTO LIABILITY -- BODILY INJURY -- VOLUNTARY
 Loss & ALAE -- Net of Subrogation & Salvage
 Reported Claim Count Development

Appendix
 Sheet 1

Historical Cumulative Reported Claim Count Data														Chain-Ladder Method Reported Claim Counts			
Accident Year	6	18	30	42	54	66	78	90	102	114	126	138	150	Factor to Ultimate	Projected Ultimate Counts	IBNR Counts	Total Open Counts
1972	2,452	8,399	8,964	9,162	9,227	9,257	9,274	9,280	9,286	9,290	9,292	9,293	9,295		9,295		
1973	1,830	6,345	6,749	6,870	6,926	6,949	6,959	6,965	6,969	6,974	6,974	6,975	6,976		6,976		
1974	2,275	8,312	8,907	9,150	9,207	9,236	9,251	9,257	9,261	9,264	9,265	9,266	9,266		9,266		
1975	2,319	8,429	9,126	9,379	9,483	9,514	9,543	9,549	9,555	9,558	9,558	9,560	9,560		9,560		
1976	1,837	6,343	6,889	7,119	7,167	7,187	7,198	7,202	7,204	7,207	7,208	7,208	7,210		7,210		
1977	1,874	6,222	6,690	6,879	6,942	6,966	6,978	6,982	6,988	6,992	6,993	6,993	6,995		6,995		
1978	1,874	5,929	6,268	6,382	6,445	6,464	6,473	6,479	6,482	6,484	6,485	6,485	6,485		6,485		
1979	1,933	6,449	6,823	6,980	7,032	7,052	7,062	7,066	7,070	7,072	7,074	7,074	7,074		7,074		
1980	2,052	7,149	7,618	7,787	7,859	7,886	7,895	7,898	7,898	7,900	7,901	7,904	7,904		7,904		
1981	2,409	8,488	9,097	9,264	9,350	9,388	9,408	9,415	9,418	9,420	9,420	9,423	9,424		9,424		
1982	3,015	9,218	9,756	9,977	10,069	10,108	10,119	10,135	10,138	10,139	10,139	10,139	10,140		10,140		
1983	2,739	8,641	9,161	9,419	9,506	9,546	9,562	9,571	9,575	9,577	9,579	9,580	9,580		9,580		
1984	2,361	7,464	8,127	8,349	8,429	8,459	8,475	8,482	8,486	8,491	8,491	8,491	8,491		8,491		
1985	1,574	4,363	4,726	4,878	4,947	4,968	4,976	4,979	4,981	4,982	4,983	4,984	4,984		4,984		
1986	852	2,851	3,099	3,205	3,247	3,269	3,278	3,280	3,280	3,281	3,283	3,283	3,283	1,0000	3,283	-	1
1987	759	2,581	2,828	2,935	2,973	2,985	2,990	2,992	2,992	2,992	2,992	2,992	2,992	1,0000	2,992	-	3
1988	727	2,646	2,860	2,962	2,994	3,002	3,005	3,006	3,006	3,006	3,006	3,006	3,006	1,0001	3,006	0	1
1989	873	2,929	3,232	3,349	3,389	3,404	3,414	3,416	3,420	3,420	3,420	3,420	3,420	1,0003	3,421	1	4
1990	970	3,170	3,445	3,534	3,559	3,572	3,574	3,576	3,577	3,577	3,577	3,577	3,577	1,0003	3,578	1	4
1991	973	3,334	3,619	3,706	3,740	3,746	3,749	3,750	3,750	3,750	3,750	3,750	3,750	1,0007	3,753	3	14
1992	1,229	4,206	4,585	4,732	4,799	4,816	4,824	4,824	4,824	4,824	4,824	4,824	4,824	1,0011	4,829	5	51
1993	1,412	4,770	5,168	5,331	5,393	5,414	5,414	5,414	5,414	5,414	5,414	5,414	5,414	1,0026	5,428	14	110
1994	1,620	5,391	5,875	6,054	6,129	6,129	6,129	6,129	6,129	6,129	6,129	6,129	6,129	1,0058	6,165	36	252
1995	1,690	5,279	5,740	5,990	5,990	5,990	5,990	5,990	5,990	5,990	5,990	5,990	5,990	1,0177	6,096	106	560
1996	1,744	5,701	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	1,0527	6,534	327	1,171
1997	1,627	5,463	5,990	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	1,1447	6,253	790	2,658
1998	1,756	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	6,207	1,8712	3,286	1,530	2,895
															2,813	7,724	

Calculated Reported Claim Count RTR Factors for Chain-Ladder Method													
	6->18	18->30	30->42	42->54	54->66	66->78	78->90	90->102	102->114	114->126	126->138	138->150	150->Ult
1972	1.7127	1.0673	1.0221	1.0071	1.0033	1.0018	1.0006	1.0006	1.0004	1.0002	1.0001	1.0002	1.0000
1973	1.7341	1.0637	1.0179	1.0082	1.0033	1.0014	1.0009	1.0006	1.0007	1.0000	1.0001	1.0001	1.0000
1974	1.8268	1.0716	1.0273	1.0062	1.0031	1.0016	1.0006	1.0004	1.0003	1.0001	1.0001	1.0000	1.0000
1975	1.8174	1.0827	1.0277	1.0111	1.0033	1.0030	1.0006	1.0006	1.0003	1.0000	1.0002	1.0000	1.0000
1976	1.7265	1.0861	1.0334	1.0067	1.0028	1.0015	1.0006	1.0003	1.0004	1.0001	1.0000	1.0003	1.0000
1977	1.6601	1.0752	1.0283	1.0092	1.0035	1.0017	1.0006	1.0009	1.0006	1.0001	1.0000	1.0003	1.0000
1978	1.5819	1.0572	1.0182	1.0099	1.0029	1.0014	1.0009	1.0005	1.0003	1.0002	1.0000	1.0000	1.0000
1979	1.6681	1.0580	1.0230	1.0074	1.0028	1.0014	1.0006	1.0006	1.0003	1.0003	1.0000	1.0000	1.0000
1980	1.7420	1.0656	1.0222	1.0092	1.0034	1.0011	1.0004	1.0000	1.0003	1.0001	1.0004	1.0000	1.0000
1981	1.7617	1.0717	1.0184	1.0093	1.0041	1.0021	1.0007	1.0003	1.0002	1.0000	1.0003	1.0001	1.0000
1982	1.5287	1.0584	1.0227	1.0092	1.0039	1.0011	1.0016	1.0003	1.0001	1.0000	1.0000	1.0001	1.0000
1983	1.5774	1.0602	1.0282	1.0092	1.0042	1.0017	1.0009	1.0004	1.0002	1.0002	1.0001	1.0000	1.0000
1984	1.5807	1.0888	1.0273	1.0096	1.0036	1.0019	1.0008	1.0005	1.0006	1.0000	1.0000	1.0000	1.0000
1985	1.3860	1.0832	1.0322	1.0141	1.0042	1.0016	1.0006	1.0004	1.0002	1.0002	1.0002	1.0000	1.0000
1986	1.6731	1.0870	1.0342	1.0131	1.0068	1.0028	1.0006	1.0000	1.0003	1.0006	1.0000	1.0000	1.0000
1987	1.7003	1.0957	1.0378	1.0129	1.0040	1.0017	1.0007	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1988	1.8198	1.0809	1.0357	1.0108	1.0027	1.0010	1.0003	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1989	1.6775	1.1034	1.0362	1.0119	1.0044	1.0029	1.0006	1.0012	1.0000	1.0000	1.0000	1.0000	1.0000
1990	1.6340	1.0868	1.0258	1.0071	1.0037	1.0006	1.0006	1.0003	1.0000	1.0000	1.0000	1.0000	1.0000
1991	1.7133	1.0855	1.0240	1.0092	1.0016	1.0008	1.0008	1.0003	1.0000	1.0000	1.0000	1.0000	1.0000
1992	1.7111	1.0901	1.0321	1.0142	1.0035	1.0017	1.0007	1.0003	1.0003	1.0001	1.0000	1.0000	1.0000
1993	1.6891	1.0834	1.0315	1.0116	1.0039	1.0011	1.0007	1.0003	1.0001	1.0000	1.0000	1.0000	1.0000
1994	1.6639	1.0898	1.0305	1.0124	1.0039	1.0011	1.0007	1.0003	1.0001	1.0000	1.0000	1.0000	1.0000
1995	1.5618	1.0873	1.0436	1.0124	1.0039	1.0011	1.0007	1.0003	1.0001	1.0000	1.0000	1.0000	1.0000
1996	1.6345	1.0888	1.0436	1.0124	1.0039	1.0011	1.0007	1.0003	1.0001	1.0000	1.0000	1.0000	1.0000
1997	1.6789	1.0888	1.0436	1.0124	1.0039	1.0011	1.0007	1.0003	1.0001	1.0000	1.0000	1.0000	1.0000
1998	1.6789	1.0888	1.0436	1.0124	1.0039	1.0011	1.0007	1.0003	1.0001	1.0000	1.0000	1.0000	1.0000
S-4 Average	1.6348	1.0873	1.0344	1.0118	1.0032	1.0015	1.0004	1.0004	1.0001	1.0002	1.0001	1.0000	1.0000
Selected RTR	1.6348	1.0873	1.0344	1.0118	1.0032	1.0015	1.0004	1.0004	1.0001	1.0002	1.0001	1.0000	1.0000
Benchmark RTR	1.6784	1.0889	1.0327	1.0117	1.0037	1.0016	1.0007	1.0003	1.0002	1.0001	1.0001	1.0001	1.0000
Credibility	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Cred Wtd RTR	1.6348	1.0873	1.0344	1.0118	1.0032	1.0015	1.0004	1.0004	1.0001	1.0002	1.0001	1.0000	1.0000
RTU	1.8712	1.1447	1.0527	1.0177	1.0058	1.0026	1.0011	1.0007	1.0003	1.0003	1.0001	1.0000	1.0000
Rep'ing Pattern	53.4%	87.4%	95.0%	98.3%	99.4%	99.7%	99.9%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%

Claim Reserving: Performance Testing and the Control Cycle

ABC INSURANCE COMPANY
COMMERCIAL AUTO LIABILITY -- BODILY INJURY -- VOLUNTARY
Loss & ALAE -- Net of Subrogation & Salvage
Paid Loss & ALAE Development

Appendix
Sheet 2[illegible]

Calculated Paid Loss & ALAE RTR Factors for Chain-Ladder Method													
	6->18	18->30	30->42	42->54	54->66	66->78	78->90	90->102	102->114	114->126	126->138	138->150	150->Ult
1972	6.2871	2.1694	1.3973	1.1799	1.0907	1.0381	1.0249	1.0102	1.0046	1.0017	1.0009	1.0014	
1973	6.8196	2.0538	1.3969	1.2160	1.0919	1.0423	1.0165	1.0091	1.0063	1.0045	1.0016		
1974	5.8193	2.1045	1.3582	1.2089	1.0902	1.0497	1.0267	1.0205	1.0044	1.0012			
1975	7.9748	2.1455	1.3063	1.1771	1.0852	1.0507	1.0162	1.0142	1.0077				
1976	6.1824	1.7726	1.2884	1.1991	1.1037	1.0409	1.0132	1.0113					
1977	5.7223	2.0617	1.4198	1.2046	1.1361	1.0391	1.0249						
1978	5.4734	2.1304	1.4074	1.2243	1.0955	1.0429							
1979	7.4955	2.4446	1.4691	1.2319	1.1053								
1980	6.0761	2.0433	1.5444	1.2531									
1981	8.5937	2.1237	1.4706										
1982	8.6842	2.3092											
1983	7.6202												
1984													
1985													
1986													
1987													
1988													
1989													
1990													
1991													
1992													
1993													
1994													
1995													
1996													
1997													
1998													
S-4 Average	7.7435	2.2302	1.4729	1.2285	1.1102	1.0434	1.0203	1.0138	1.0058	1.0025	1.0012	1.0014	
Selected RTR	7.7435	2.2302	1.4729	1.2285	1.1102	1.0434	1.0203	1.0138	1.0058	1.0025	1.0012	1.0014	
Benchmark RTR	6.9642	2.1315	1.4067	1.2106	1.1032	1.0453	1.0232	1.0134	1.0082	1.0022	1.0014	1.0009	1.0016
Credibility	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.50	0.25	0.00
Cred Wtd RTR	7.7435	2.2302	1.4729	1.2285	1.1102	1.0434	1.0203	1.0138	1.0058	1.0024	1.0013	1.0011	1.0016
RTU	37.8929	4.8935	2.1942	1.4897	1.2127	1.0923	1.0469	1.0261	1.0122	1.0064	1.0040	1.0026	1.0016
Percent Unpaid	97.4%	79.6%	54.4%	32.9%	17.5%	8.5%	4.5%	2.5%	1.2%	0.6%	0.4%	0.3%	0.2%

ABC INSURANCE COMPANY
COMMERCIAL AUTO LIABILITY -- BODILY INJURY -- VOLUNTARY
Loss & ALAE -- Net of Subrogation & Salvage
Incurred Loss & ALAE Development

Appendix
Sheet 3[illegible]

	Calculated Incurred Loss & ALAE RTR Factors for Chain-Ladder Method												
	6->18	18->30	30->42	42->54	54->66	66->78	78->90	90->102	102->114	114->126	126->138	138->150	150->Ult
1972	1.9203	1.2008	1.1301	1.0267	1.0277	0.9975	0.9962	0.9794	0.9991	0.9984	0.9986	1.0006	
1973	1.7267	1.1021	1.0364	1.0606	1.0015	1.0000	0.9980	0.9948	0.9992	1.0000	1.0013		
1974	1.9080	1.2181	1.0463	0.9968	1.0075	1.0086	1.0012	0.9694	1.0018	0.9987			
1975	2.4508	1.1431	1.0507	1.0120	1.0143	1.0043	1.0032	1.0066	1.0017				
1976	1.6142	1.1122	1.0999	1.0237	1.0181	1.0123	0.9986	0.9981					
1977	1.9900	1.2086	1.0851	1.0339	1.0426	0.9995	1.1207						
1978	1.7697	1.2040	1.0738	1.0331	1.0327	1.0202							
1979	1.8940	1.2246	1.1307	1.0967	1.0185								
1980	2.0867	1.2958	1.2012	1.0803									
1981	1.8009	1.2704	1.1539										
1982	1.8545	1.2133											
1983	1.8504												
1984													
1985													
1986													
1987													
1988													
1989													
1990													
1991													
1992													
1993													
1994													
1995													
1996													
1997													
1998													
S-4 Average	1.8981	1.2510	1.1399	1.0610	1.0280	1.0091	1.0309	0.9972	1.0004	0.9990	1.0000	1.0006	
Selected RTR	1.8981	1.2510	1.1399	1.0610	1.0280	1.0091	1.0309	1.0000	1.0004	1.0000	1.0000	1.0006	
Benchmark RTR	1.9145	1.1992	1.1008	1.0475	1.0285	1.0150	1.0040	1.0028	1.0018	1.0005	1.0004	1.0002	1.0002
Credibility	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.50	0.25	0.00
Cred Wtd RTR	1.8981	1.2510	1.1399	1.0610	1.0280	1.0091	1.0309	1.0000	1.0004	1.0001	1.0002	1.0003	1.0002
RTU	3.0748	1.6199	1.2949	1.1360	1.0707	1.0415	1.0321	1.0012	1.0012	1.0007	1.0006	1.0005	1.0002
Percent Unrpr'td	67.5%	38.3%	22.8%	12.0%	6.6%	4.0%	3.1%	0.1%	0.1%	0.1%	0.1%	0.0%	0.0%
Percent Unpaid	97.4%	79.6%	54.4%	32.9%	17.5%	8.5%	4.5%	2.5%	1.2%	0.6%	0.4%	0.3%	0.2%
Percent in Case	29.9%	41.3%	31.7%	20.9%	10.9%	4.5%	1.4%	2.4%	1.1%	0.6%	0.3%	0.2%	0.1%
Implied Case RTU	3.2580	1.9267	1.7194	1.5725	1.6034	1.8917	3.2772	1.0486	1.1088	1.1337	1.1905	1.2106	1.1048

Claim Reserving: Performance Testing and the Control Cycle

ABC INSURANCE COMPANY
COMMERCIAL AUTO LIABILITY -- BODILY INJURY -- VOLUNTARY
Loss & ALAE -- Net of Subrogation & Salvage
Case Reserve Loss & ALAE Development

Appendix
Sheet 4

Historical Case-Basis Reserve Loss & ALAE Data														Case Reserve Development Method			
Accident Year	6	18	30	42	54	66	78	90	102	114	126	138	150	Case Devel. Factor	Projected Ultimate Losses	IBNR Losses	Total Unpaid Losses
1972	4,269	12,243	8,910	6,603	4,002	2,766	1,811	1,109	321	182	99	42	22		25,311		
1973	5,180	13,682	9,718	5,900	3,746	1,975	1,068	653	325	162	58	52	31		23,274		
1974	5,004	14,722	12,165	8,457	4,486	2,661	1,677	1,021	188	118	52	7	-		27,108		
1975	4,615	17,130	11,636	7,878	4,306	2,516	1,209	825	600	415	387	139	60		31,447		
1976	5,101	11,133	7,614	5,962	3,211	1,578	974	641	335	270	218	203	159	1,1500	23,696	24	183
1977	4,074	12,303	9,754	6,446	3,668	1,846	893	3,391	890	120	86	68		1,1750	26,852	12	80
1978	4,505	12,041	9,154	5,730	2,612	1,368	870	567	346	90	76			1,2000	25,679	15	91
1979	5,500	15,826	11,206	7,127	4,562	2,074	1,061	700	564	140				1,1254	36,214	18	158
1980	4,884	14,614	12,369	8,696	4,602	3,299	1,615	840	456					1,0320	43,370	15	471
1981	9,436	22,819	16,932	10,034	6,089	2,938	2,831	1,155						0.9967	65,226	(4)	1,151
1982	10,525	27,249	16,636	14,189	8,162	6,323	2,971							1,3105	71,187	922	3,893
1983	11,501	28,877	23,102	14,540	7,296	4,881								1,3340	85,135	1,630	6,511
1984	9,481	26,882	23,386	17,637	9,453									1,5353	79,257	5,060	14,513
1985	7,687	19,434	19,059	13,593										1,6204	55,250	8,433	22,026
1986	5,033	13,296	20,859											1,8755	54,905	18,261	39,120
1987	4,376	16,124												2,2583	42,415	20,289	36,413
1988	3,962													3,7252	15,037	10,797	14,759
1989																	
1990																	
1991																	
1992																	
1993																	
1994																	
1995																	
1996																	
1997																	
1998																	
															65,472	139,369	

Calculated Case-Basis Reserve Factors to Ultimate Outstanding																	
	6->Ult	18->Ult	30->Ult	42->Ult	54->Ult	66->Ult	78->Ult	90->Ult	102->Ult	114->Ult	126->Ult	138->Ult	150->Ult				
1972	2.8531	1.5790	1.3850	1.0884	0.9810	0.7180	0.6057	0.4454	0.7508	0.6923	0.8485	1.4762	1.1818				
1973	2.1668	1.2694	1.2045	1.2056	0.9672	0.9205	0.8521	0.8300	1.0338	1.1852	1.5172	1.0000	0.9677				
1974	2.5969	1.3994	1.1028	1.0064	1.0314	0.9767	0.8235	0.6787	0.8032	0.2712	0.0000	0.0000	1.0000				
1975	3.2994	1.3730	1.2410	1.1716	1.2299	1.2198	1.3483	1.3916	1.1985	1.1593	0.7367	0.7705	1.1517				
1976	2.2082	1.4798	1.4310	1.2085	1.2211	1.1856	1.0040	1.0575	1.2443	1.3180	1.4305	1.3490					
1977	3.1669	1.6951	1.4853	1.4423	1.5553	1.5308	2.1119	0.3727	0.6920	1.0408	1.3128						
1978	2.7331	1.6535	1.4627	1.4629	1.7237	1.8123	1.7060	1.6229	1.3705	1.6244							
1979	3.2109	1.8648	1.7698	1.7044	1.4396	1.6613	1.8648	1.5079	1.1340								
1980	4.2919	2.3665	2.0549	1.7991	1.8745	1.5237	1.5378	1.1257									
1981	3.3691	2.2393	2.0802	2.1030	1.9358	1.6461	0.9511										
1982	3.2998	2.0624	2.1985	1.6288	1.4758	1.0545											
1983	3.5981	2.3223	1.9821	1.7815	1.6644												
1984	4.0414	2.4050	1.9433	1.5400													
1985	3.4523	2.3121	1.7202														
1986	5.3890	3.5964															
1987	4.7661																
1988																	
1989																	
1990																	
1991																	
1992																	
1993																	
1994																	
1995																	
1996																	
1997																	
1998																	
W-10 Avg. CRDF	3.7252	2.2583	1.8755	1.6204	1.5353	1.3340	1.3105	0.8396	1.0320	1.1254							
CV of CRDF	0.2352	0.2656	0.1934	0.2015	0.2214	0.2685	0.3872	0.5412	0.2442	0.3917							
Benchmark CRDF	3.2243	1.9359	1.6840	1.4730	1.4403	1.3250	1.3000	1.2750	1.2500	1.2250	1.2000	1.1750	1.1500				
Credibility	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.64	1.00	1.00	0.00	0.00	0.00				
Cred Wtd CRDF	3.7252	2.2583	1.8755	1.6204	1.5353	1.3340	1.3105	0.9967	1.0320	1.1254	1.2000	1.1750	1.1500				
Selected CRDF	3.7252	2.2583	1.8755	1.6204	1.5353	1.3340	1.3105	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000				

ABC INSURANCE COMPANY
COMMERCIAL AUTO LIABILITY -- BODILY INJURY -- VOLUNTARY
Loss & ALAE -- Net of Subrogation & Salvage
Adjusted Avg Case Reserve -- Reported Loss & ALAE Development

Appendix
Sheet 5

Historical Average Case Reserve														Avg 18+ Vals	Percent Increase	Selected Trend	Adj. Factors
Valuation Date	6	18	30	42	54	66	78	90	102	114	126	138	150				
Jun-79	4,283	6,656	10,477	10,685	11,361	14,949	19,415	41,058	19,494	19,494	19,494	19,494	19,494	9,082			0.9761
Jun-80	3,509	7,842	11,204	11,720	11,981	12,394	16,441	21,061	29,215	18,219	18,219	18,219	18,219	9,804	7.9%	11.7%	1.0103
Jun-81	5,796	6,691	12,062	13,112	12,476	10,808	11,406	17,603	17,080	22,704	15,920	15,920	15,920	9,434	-3.8%	10.1%	1.1564
Jun-82	4,886	8,657	12,686	14,575	11,357	13,574	12,987	13,306	8,545	16,184	24,714	15,147	15,147	10,751	14.0%	6.6%	1.0820
Jun-83	5,975	9,544	13,666	16,408	18,175	10,523	13,132	16,436	15,000	9,833	14,461	21,183	15,383	11,977	11.4%	5.1%	1.0208
Jun-84	5,911	10,321	13,384	15,557	16,204	13,827	13,385	82,707	18,611	17,292	26,000	13,107	21,899	12,770	6.6%	4.5%	1.0000

Historical Open Claim Counts														
Valuation	6	18	30	42	54	66	78	90	102	114	126	138	150	
Jun-84	1,604	2,798	1,243	645	284	150	65	41	18	24	2	4	1	

Adjusted Incurred Loss & ALAE Development														Adjusted Case Reserves Incurred Chain Ladder Method			
Accident Year	6	18	30	42	54	66	78	90	102	114	126	138	150	Factor to Ultimate	Projected Ultimate Losses	IBNR Losses	Total Unpaid Losses
1972	-	-	-	-	-	-	-	25,900	25,394	25,395	25,334	25,292	25,307	1.0002	25,311	4	26
1973	-	-	-	-	-	-	23,407	23,392	23,314	23,257	23,245	23,274		1.0004	23,284	10	62
1974	-	-	-	-	-	27,106	27,421	27,596	27,160	27,196	27,160			1.0006	27,177	17	69
1975	-	-	-	-	30,354	30,920	31,215	31,192	31,340	31,381				1.0007	31,404	23	438
1976	-	-	-	22,311	23,019	23,650	23,772	23,672	23,614					1.0008	23,633	19	354
1977	-	-	21,885	24,068	25,389	26,023	25,878	28,979						1.0008	29,002	23	3,414
1978	-	17,523	21,539	23,923	24,003	24,596	25,065							1.0249	25,690	625	1,495
1979	5,716	22,691	29,340	31,777	34,303	34,842								1.0331	35,995	1,153	3,227
1980	7,105	25,686	31,335	36,602	39,345									1.0561	41,553	2,208	6,810
1981	9,043	38,817	47,288	54,159										1.1078	59,995	5,836	15,870
1982	12,460	42,804	51,249											1.2422	63,662	12,413	29,049
1983	12,079	46,952												1.5298	71,828	24,876	53,753
1984	10,793													2.9372	31,701	20,908	30,389
1985																	
1986																	
1987																	
1988																	
1989																	
1990																	
1991																	
1992																	
1993																	
1994																	
1995																	
1996																	
1997																	
1998																	

68,117 144,958

Calculated Adjusted Reported Loss & ALAE RTR Factors for Chain-Ladder Method														Simulated	
	6->18	18->30	30->42	42->54	54->66	66->78	78->90	90->102	102->114	114->126	126->138	138->150	150->Ult		
1972								0.9805	1.0000	0.9976	0.9983	1.0006			
1973							0.9994	0.9967	0.9976	0.9995	1.0012				
1974						1.0116	1.0064	0.9842	1.0013	0.9987					
1975					1.0186	1.0095	0.9992	1.0048	1.0013						
1976				1.0318	1.0274	1.0052	0.9958	0.9975							
1977			1.0997	1.0549	1.0250	0.9944	1.1199								
1978		1.2292	1.1107	1.0033	1.0247	1.0191									
1979	1.9848	1.2930	1.0831	1.0795	1.0157										
1980	1.8076	1.2199	1.1681	1.0750											
1981	2.1462	1.2182	1.1453												
1982	1.7176	1.1973													
1983	1.9435														
1984															
1985															
1986															
1987															
1988															
1989															
1990															
1991															
1992															
1993															
1994															
1995															
1996															
1997															
1998															
S-5 Average	1.9199	1.2315	1.1214	1.0489	1.0223	1.0080	1.0241	0.9927	1.0001	0.9986	0.9998	1.0006			
Selected RTR	1.9199	1.2315	1.1214	1.0489	1.0223	1.0080	1.0241	1.0000	1.0001	1.0000	1.0000	1.0006			
Benchmark RTR	1.9145	1.1992	1.1008	1.0475	1.0285	1.0150	1.0040	1.0028	1.0018	1.0005	1.0004	1.0002	1.0002		
Credibility	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.50	0.25	0.00		
Cred Wtd RTR	1.9199	1.2315	1.1214	1.0489	1.0223	1.0080	1.0241	1.0000	1.0001	1.0001	1.0002	1.0003	1.0002		

ABC INSURANCE COMPANY

COMMERCIAL AUTO LIABILITY -- BODILY INJURY -- VOLUNTARY

Loss & ALAE -- Net of Subrogation & Salvage

Calculation of PCLD Method Skill at 42 Months

Appendix

Sheet 6

Valuation Date	Accident Year	Earned Premium	PCLD Projected Unpaid	Ratio	"Actual" Unpaid	Ratio	Percent Paid	Actual Anomaly	Predicted Anomaly	Error
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Jun-79	1976	59,040	7,173	12.1%	7,211	12.2%	97%	-3.73%	-3.79%	-0.06%
Jun-80	1977	61,516	7,598	12.4%	9,268	15.1%	100%	-0.88%	-3.59%	-2.71%
Jun-81	1978	65,088	7,506	11.5%	8,295	12.7%	100%	-3.20%	-4.41%	-1.21%
Jun-82	1979	72,444	10,910	15.1%	12,214	16.9%	99%	0.92%	-0.88%	-1.80%
Jun-83	1980	88,861	12,917	14.5%	15,613	17.6%	100%	1.63%	-1.41%	-3.03%
Jun-84	1981	110,447	21,609	19.6%	20,989	19.0%	100%	3.06%	3.62%	0.56%
Jun-85	1982	118,920	23,766	20.0%	22,899	19.3%	100%	3.31%	4.04%	0.73%
Jun-86	1983	113,763	30,334	26.7%	24,796	21.8%	100%	5.85%	10.72%	4.87%
Jun-87	1984	105,594	27,098	25.7%	26,040	24.7%	100%	8.72%	9.72%	1.00%
Jun-88	1985	87,695	16,932	19.3%	20,342	23.2%	100%	7.25%	3.37%	-3.89%
Jun-89	1986	77,969	12,301	15.8%	12,504	16.0%	99%	0.09%	-0.17%	-0.26%
Jun-90	1987	79,207	11,550	14.6%	11,289	14.3%	100%	-1.69%	-1.36%	0.33%
Jun-91	1988	81,725	12,157	14.9%	10,683	13.1%	100%	-2.87%	-1.07%	1.80%
Jun-92	1989	87,874	16,463	18.7%	8,905	10.1%	100%	-5.81%	2.79%	8.60%
Jun-93	1990	96,657	15,031	15.6%	11,794	12.2%	100%	-3.74%	-0.39%	3.35%
Jun-94	1991	101,162	19,541	19.3%	10,108	10.0%	100%	-5.95%	3.37%	9.32%
Jun-95	1992	128,231	17,767	13.9%	20,104	15.7%	97%	-0.27%	-2.09%	-1.82%
Jun-96	1993	145,262	20,967	14.4%	22,764	15.7%	94%	-0.27%	-1.51%	-1.24%
Jun-97	1994	156,751	21,633	13.8%	22,086	14.1%	93%	-1.85%	-2.14%	-0.29%
Jun-98	1995	159,211	20,733	13.0%	23,924	15.0%	88%	-0.92%	-2.92%	-2.00%

Weighted Average Ratios

16.6%

15.9%

Bias

0.6%

msa = 0.1573%
mse = 0.1249%
Skill = 21%

Notes

- (2) Accident year that is at 42 months maturity as of valuation date
- (3) Earned premium for that accident year
- (4) Projected unpaid claims using PCLD method as of the valuation date
- (5) Projected unpaid claims as ratio to earned premium
- (6) "Actual" unpaid claims as of valuation date, with benefit of hindsight payments through 2000 and projections beyond
- (7) "Actual" unpaid claims as ratio to earned premium
- (8) Percent of "actual" claims that are actually paid as of December 2000; used as weight in calculating averages
- (9) Actual ratio in column (7), less weighted average ratio in column (7)
- (10) Predicted ratio in column (5), less weighted average ratio in column (7)
- (11) Difference between actual and predicted anomalies = (10) - (9)

msa is weighted average of anomalies, using column (8) as weights

mse is weighted average of errors, using column (8) as weights