Chain-Ladder Bias: Its Reason and Meaning

by Leigh J. Halliwell

ABSTRACT

Over the past twenty years many actuaries have claimed and argued that the chain-ladder method of loss reserving is biased; nonetheless, the chain-ladder method remains the favorite tool of reserving actuaries. Nearly everyone who acknowledges this bias believes it to be upward. Although supporting these claims and beliefs, the author proposes herein to deal with two deeper issues. First, does something inherent in the chain-ladder method dispose it to bias? Is there a diagnostic whereby one can predict how the chain-ladder method will fare with a particular loss triangle? To resolve this issue basic regression theory will suffice, specifically, the much misunderstood concept of regression toward the mean. And second, what lessons can we learn from the phenomenon of bias; in particular, is there a difference between actuarial methods and statistical models? These two issues constitute the reason and meaning of chain-ladder bias.

KEYWORDS

Chain-ladder method, loss development, statistical model, stochastic regressor, regression toward mean, credibility

1. Introduction

The chain-ladder method is a common technique whereby actuaries develop losses from a less mature present to a more mature future. At each stage of development the actuary determines a link ratio or age-to-age factor, namely, the ratio of cumulative losses at the later age to those at the earlier. Immature losses climb toward maturity when multiplied by a concatenation of these ratios, hence the apt description "chain-ladder (CL) method." The origin of the method is obscured in the antiquity of the Casualty Actuarial Society.[1] Actuaries themselves probably borrowed it from underwriters, along with such other stock-in-trade practices as on-leveling and trending. Actuaries and academicians now recognize these practices as rather naive and deterministic, and since the 1990s they have sought to bring modern statistical theory to bear on the problems of loss development, particularly seeking regression-model interpretations of the CL method. However, seldom has this modern theory been unleashed; most of its proponents unwittingly incorporate accidents of the CL method into their modeling attempts. We will attempt to demonstrate here that modern statistical modeling constitutes a revolution against, rather than an improvement of, the CL method. But for maximum effect, our demonstration will take the form of an inside job; we will start with the familiar issue of whether the method, as applied to loss triangles, is biased.

[1] A reviewer suggested that "seeds" of the method come from Thomas F. Tarbell [16]. But the ideas and formulas therein are exposure-related, i.e., IBNR = exposure × frequency × severity, with suitable adjustments. Development from latest loss is suggested only in one paragraph (p. 277) that begins, "It has also been contended that the incurred but not reported reserve may be determined as a function of the reserve for known cases." But even here the dependent variable is case reserves, which excludes paid losses. The method that Tarbell had in mind is not clear; but what he deduced from it is inconsistent with the CL method: "For the major [casualty] lines the reserve [determined as a function of the reserve for known cases] will be too low if the volume of business is increasing and conversely if the volume of business is decreasing the reserve will be too high." As interesting as the paper is, we do not detect in it a primitive CL method.

2. Simulated and anecdotal chain-ladder bias

The issue of chain-ladder bias was raised by James N. Stanard in his 1985 *Proceedings* paper, "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques" [15]. Stanard simulated thousands of $(5 \times 5)^2$ loss rectangles, applied four projection methods (viz., chain-ladder or age-to-age, Bornhuetter-Ferguson, Cape-Cod or Stanard-Bühlmann, and additive) to their upper-left triangles, and compared the projections with their lower-right triangles. He concluded: "The results indicate that the commonly used age-to-age factor approach gives biased estimates and is inferior to the three other methods tested" [15, p. 124] and "The results show that simple age-to-age factors produced biased results" [15, p. 132].[3] Now bias can be either upward or downward, and Stanard never specified its direction. However, in each of his eight exhibits the CL method overpredicted; hence, we are to understand that the bias is upward, i.e., that the CL method predicts too much loss.

Stanard deserves credit for raising the issue of CL bias, and it saddens this author that the paper is no longer on the CAS Examination Syllabus. Nonetheless, Stanard did not prove the CL method to be biased upward, even though he believed that his appended "Analytical Argument" proved that it was biased *in some direction*.[4] Perhaps for some triangles the bias would be upward, for others downward, and overall unbiased. Perhaps even for some triangles the bias would be zero. Then the issue of bias would be how to determine a priori how the method would

[2] Actually the rectangles were (6×5). However, the losses were considered known as of the end of the fifth accident year. Hence, there was no observation of the sixth year (being, as it were, a ratemaking year), and the CL method could project only five years.

[3] Stanard here mentions that he first asserted the bias of the CL method in his 1980 discussion paper, "Experience Rates as Estimators: A Simulation of their Bias and Variance" [14].

[4] In this Appendix Stanard inferred CL bias from the premise that the expectation of a quotient is not equal to the quotient of the expectations. But this premise is not required of CL regression models.

fare with the triangle in question. To borrow the words of Goldilocks, as applied to a particular triangle is the CL method "too hot, too cold, or just right?"

This prompts us to clarify what we mean by "bias." Stanard used "bias" in different senses. In footnote 1 and in Appendix A of his paper "bias" approximated the statistically accepted meaning, i.e., that the expectation of the estimator equals what it purports to estimate. In his footnote 13, he claimed that simple-average factors are "likely to produce substantial additional bias" as compared with weighted-average factors—a claim that mistakes the bias of an estimator for the size of its variance. But most commonly, by "bias" he meant that the method missed the mark. In this sense all four methods are biased, as we see in one of his concluding statements:

> *The additive method 4 and the average-then-adjust method 3 have significantly lower variances than methods 1 and 2 [CL and Bornhuetter-Ferguson], and small biases (if adjusted for inflation). In fact, method 4 may be completely unbiased. [15, p. 135]*

We say "approximated" in the first instance, because even here the bias applies to the total projection. Although the total projection is of greatest importance, how can a method be trusted to be unbiased on the total, if it is biased on the subtotals?[5]

In actuality, Stanard demonstrated only that the CL method was the least accurate of his four methods *as far as his simulated rectangles were concerned*. In footnote 8 he claimed that his findings "are not particularly sensitive to the choice of the underlying loss generation model." We are unconvinced that his computer model, even with an inflation provision, adequately mimics real loss triangles.[6] But more disconcerting than the poor performance of the CL method relative to other methods is the fact that it consistently overpredicted. Was this just an accident of his simulation? There is anecdotal evidence that the CL method overpredicts with real loss triangles more often than it underpredicts; however, "anecdote" may be just a fancy word for "feeling" or "opinion." The author knows of no ex post testing of large numbers of real triangles; even though Barnett and Zehnwirth have modeled hundreds of real triangles, and believe the CL method to be out of step with most datasets, they do not cite statistics of how often or in what circumstances the method will overpredict, underpredict, or be just right.

In fact, we see little value in such testing and tabulation. If something is not good enough generally, time is more profitably spent in searching for something else than in identifying the specific situations in which it works well enough. However, there is heuristic value in seeking to diagnose the behavior of the CL method. We will do this in the next section, arriving at an explanation for CL bias, particularly for overprediction.

3. Diagnosing chain-ladder bias

The triangle in Table 1, taken from Brosius [4, Table 6], illustrates how to diagnose the bias of the chain-ladder method, i.e., to diagnose whether the method in this case will underpredict or overpredict the expected development.[7] From this illustration we will be able to form general conclusions. In Exhibit 1 [see Appendix A] are found premiums, link ratios, their weighted averages, and chain-ladder projections to sixty months.

[5]Furthermore, actuarial (informed) judgment is no antidote to bias, as if actuaries possessed some expertise or intuition to herd or prod methods into correctness. Inevitable failures will only bring opprobrium and discredit upon the profession. Actuaries must not presume to judge what they cannot scientifically model.

[6]Venter, himself no CL advocate, suggests that Stanard's simulation method handicapped the CL method [19, p. 817f.]. Barnett and Zehnwirth [2, p. 297] too argue that simulation method is important, and that a model cannot distinguish the real data for which it was designed from data simulated from the model itself.

[7]Bias has to do with the *expected* outcome. If we conclude that the chain-ladder method will overpredict this triangle (and we will), we should not be lulled into thinking that the actual ultimate losses will (with 100% probability) be less than the chain-ladder ultimates, whether by accident year or by total.

Figure 1A. First link, 12 to 24 months

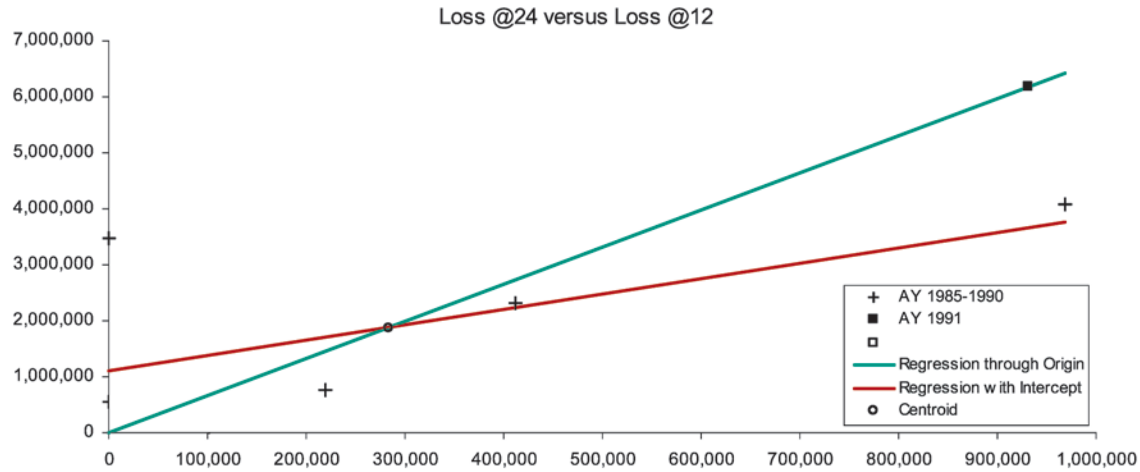


Table 1. Cumulative case-incurred losses

AY	@12	@24	@36	@48	@60
1985	102,000	104,000	209,000	650,000	847,000
1986	0	543,000	1,309,000	2,443,000	3,033,000
1987	412,000	2,310,000	3,083,000	3,358,000	4,099,000
1988	219,000	763,000	1,637,000	1,423,000	
1989	969,000	4,090,000	3,801,000		
1990	0	3,467,000			
1991	932,000				

There are six observations of loss development from 12 to 24 months, which in Figure 1A are marked with the "+" symbol. The x-axis measures the loss at 12 months, the y-axis the loss at 24 months. We would like the observed points to lie close to a one-dimensional curve, viz., $y \approx f(x)$. Then it would be a simple matter to develop AY (accident year) 1991 to 24 months as $f(932{,}000)$. For several reasons, but especially due to the scarcity of observations, we will consider only functions linear in x, i.e., $y \approx \beta_0 + \beta_1 x$. With a random error term, we make the formula exact, and arrive at a so-called "regression" model $\mathbf{y} = \beta_0 + \beta_1 x + \mathbf{e}$. The expression "regression model" is unfortunate; a better name is "linear statistical model." Appendix B details the theory of the linear statistical model.

That the observations in Figure 1A do not line up well means that if the relation between $E[\mathbf{y}]$ and x is linear, it is obscured by an error term of significant variance. Nevertheless, the general line with intercept $\beta_0 = \$1{,}094{,}448$ and slope $\beta_1 = 2.768$ (the red-colored line) must fit the observations better than the line constrained to pass through the origin (the green-colored line). This constrained line, with intercept zero and slope $\gamma = 6.626$ is the "regression-model" equivalent of the chain-ladder method. Appendix C specifies the variance assumptions of the two models (viz., the two-parameter is homoskedastic, the one-parameter is heteroskedastic), and proves that both lines must intersect at the centroid of the observations, i.e., at the average point $(\bar{x}, \bar{y})$. For all loss triangle applications the centroid will lie in the first quadrant of the Cartesian plane. Because the intercept of the general line is positive and the centroid lies in the first quadrant, to the left of the centroid the constrained line is below the general, and to the right of the centroid it is above.

Now our diagnosis rests on the following assumption: *The general line is preferable to the constrained line*. Our rationale is simple: For the purposes of diagnosing CL bias we have limited ourselves to linear functions, and the constrained line is a special case of the general. With the CL method we are, in effect, projecting along the constrained (green-colored) line, whereas we would be better off using the general (red-colored)

Figure 1B. Second link, 24 to 36 months

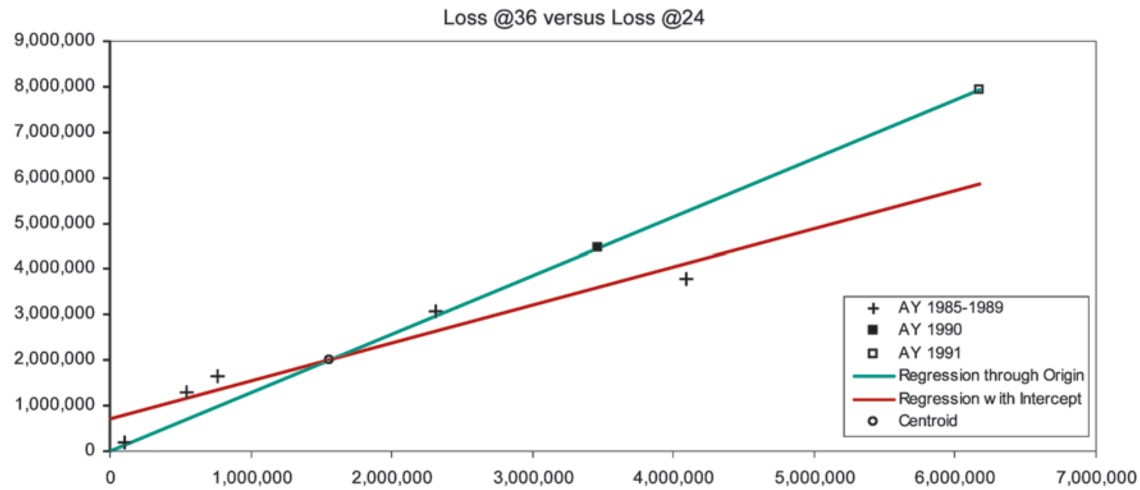
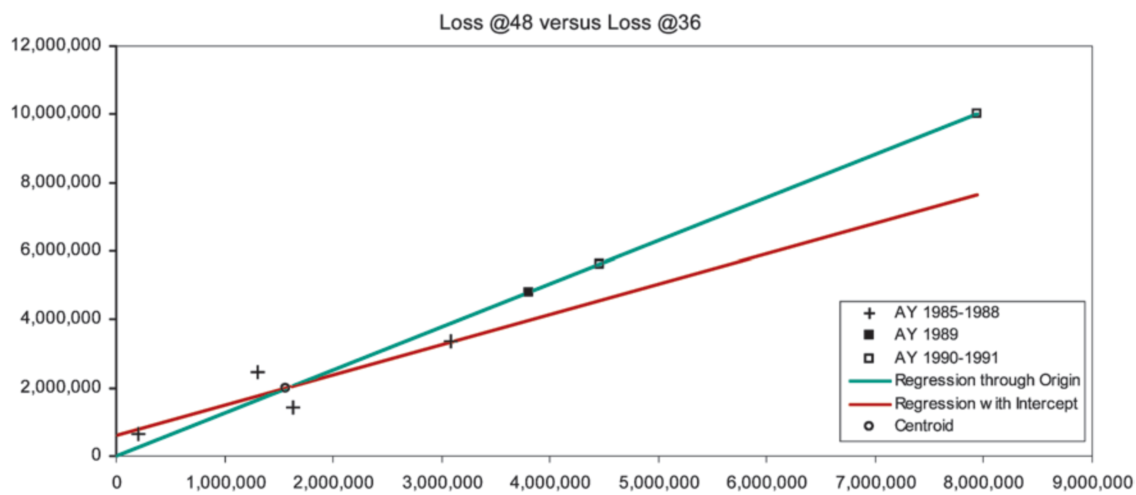


Figure 1C. Third link, 36 to 48 months



line. So we develop AY 1991 as $932,000 $\times$ 6.626 = $6,175,184, which point on the constrained line is marked with the "■" symbol. The point directly underneath on the general line would have a height of $3.7 million. On the basis of our assumptions, the CL method has considerably overpredicted.

To diagnose the next link of the chain, the development from 24 to 36 months, we rely on Figure 1B. Now we are limited to five observations, which the figure shows as "+" signs. The general and the constrained lines are fitted, and their intersection is marked as the centroid. Again we find the intercept of the general line to be positive. Hence, according to our assumption, projections from x values to the left of the centroid will be too small, and projections to the right will be too large. AY 1990 at 24 months lies moderately to the right of the centroid and is moderately overpredicted. AY 1991 at 24 months, now marked with the "□" symbol, lies extremely to the right of the centroid and is much more overpredicted. In fact, the CL method has compounded the overprediction of AY 1991, since it first overpredicted its development from 12 to 24 months.

Figure 1D. Fourth link, 48 to 60 months

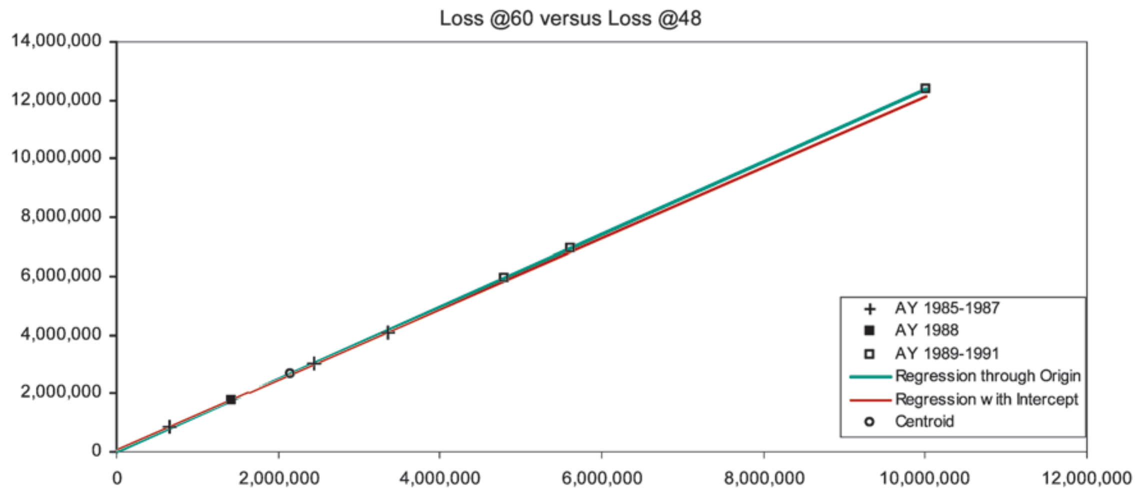


Table 2. Chain-ladder projections by year

AY	@12	@24	@36	@48	@60	Ratio to EarnPrem
1985	102,000	104,000	209,000	650,000	847,000	20%
1986	0	543,000	1,309,000	2,443,000	3,033,000	55%
1987	412,000	2,310,000	3,083,000	3,358,000	4,099,000	53%
1988	219,000	763,000	1,637,000	1,423,000	1,760,055	20%
1989	969,000	4,090,000	3,801,000	4,797,864	5,934,298	56%
1990	0	3,467,000	4,456,493	5,625,269	6,957,684	58%
1991	932,000	6,175,184	7,937,603	10,019,347	12,392,554	96%

We repeat the diagnosis for the development from 36 to 48 months Figure 1C. Although confidence in our lines is lessening, we still find the intercept to be positive; and all predictions are from x values to the right of the centroid. Therefore, we diagnose AY 1989 to be simply overpredicted, and AY 1990–1991 to be multiply overpredicted.

Finally, Figure 1D shows development from 48 to 60 months. At this stage the lines are nearly coincident; the CL method is trustworthy. Our final diagnosis rests on Table 2. AY 1988 at 60 months is fairly estimated as $1.76 million. The later accident years have been overestimated at 60 months, the overestimation compounding as the years advance. AY 1991 is overestimated to an uncharacteristically large 96% loss ratio.

In Exhibits 2–4 we apply the diagnostic to consolidated Schedule-P triangles [3, pp. 199–213]. These exhibits are more quantitative than Figures 1A–1D in that they contain the estimates, standard errors, and t statistics of the intercept and slope of the general lines.[8] Since the t statistic equals the estimate divided by the standard error, the estimate is "t-stat" standard errors away from zero. If the error terms of the regression model are normally distributed, the t-stat will be t-distributed with $n-2$ degrees of freedom, and a probabilistic significance can be assigned to it. Never assuming normally distributed error terms, we will simply abide by such qualitative standards as when the absolute value of the t statistic is greater than one the difference of the estimate from zero is fairly significant, when greater than two quite significant, and when greater than three very significant.

[8]Exhibit 5 provides this quantitative information for the 12-to-24-month regression of the Brosius triangle.

The first two development stages of the paid Workers' Compensation triangle (Exhibit 2) show significant *negative* intercepts. Nevertheless, the CL-predictions are always close to their respective centroids. Development to 60 months is unexceptional. The paid Medical Malpractice triangle (Exhibit 3) displays mildly positive intercepts in the first and fourth development stages. We suspect AY 2001 to be slightly overpredicted. Exhibit 4 is the most interesting. To obtain this case-incurred triangle for Products Liability Occurrence we had to subtract the bulk + IBNR reserves (Schedule P, Part 4) from the incurred triangle (Schedule P, Part 2). The first two development stages have large positive intercepts, and the fourth has a large negative one. AY 2004 at 36 months has been twice overpredicted. Everything else stays close to the centroids, except for one underprediction from 48 to 60 months. This underprediction pertains to AY 2004, so at 60 months this AY has been twice overpredicted and once underpredicted, overall apparently netting a slight overprediction.

At each link or development stage, chain-ladder bias depends on the intercept of the general line. If it is positive, projections from less-than-average x values (to the left of the centroid) will underestimate, projections from greater-than-average x values (to the right of the centroid) will overestimate. The relation is reversed in the case of a negative intercept: projections to the right of the centroid underestimate, those to the left overestimate. CL bias becomes more serious as the intercept moves more significantly away from zero. The method is unbiased, if the regression lines coincide.

From this we venture to explain the anecdotes that the chain-ladder method overpredicts. First, in our experience with loss triangles, we have found significant positive intercepts more often than significant negative ones. Not having kept records, we cannot cite the proportions (significantly positive, significantly negative, not significantly different from zero). Certainly we do encounter intercepts significantly less than zero; here we found them in Exhibits 2 and 4. Nevertheless, not only are positive slopes more frequent than negative; they tend to be more statistically significant, as well.

Because the centroids are in the first quadrant, we can state the first empirical finding: Based on the preponderance of positive intercepts, more often than not the CL method will overpredict, or overdevelop, losses that are greater than average at any given maturity, i.e., losses to the right of the centroid. And second, as with the rest of the economy, it is normal for the insurance business to grow. In this condition exposures increase, and one expects the amounts down any column of a loss triangle to increase. Hence, the CL method commonly projects losses that lie to the right of the centroid. The combination of positive intercepts and business expansion makes for overprediction. The CL method would actually underpredict with the combination of positive intercepts and business contraction. The reverse would hold, if the intercepts were negative. In a triangle of mixed intercepts (i.e., some positive and some negative) conflicting forces would be at work; probably, however, forces at the earlier links would prevail at the overall level.

This explanation discredits appeals to "skewness," i.e., to arguments that the upside of overprediction is unlimited, whereas the downside is limited to zero. For, indeed, the CL method is on average unbiased over the empirical distribution of the observed ordered pairs, and over any bivariate distribution that might be fitted to their moments. The observed (x, y) variables that average to the centroid may each have positive skewness. But the smaller number of large biases to the right of the centroid is offset by the larger number of small biases to the left. Of importance is the relation of the two lines over the domain of the x values that need to be developed.

4. Chain-ladder bias and the regression intercept

One might be tempted from the previous section to "fix" the chain-ladder method by projecting from the general line. At this point we are still within the framework of deterministic methods, i.e., methods that yield point estimates and provide no information about higher moments or probability distributions. Since the 1990s actuaries have advanced from methods to models in order to obtain probabilistic information. And within the modeling framework, this "fix" is tantamount to a movement from the Mack [12, p. 107] model, $E[\mathbf{y} \mid \mathbf{x} = x] = \gamma x$, to the Murphy [13, p. 187] model, $\mathbf{y} = \beta_0 + \beta_1 x + \mathbf{e}$ or $E[\mathbf{y} \mid \mathbf{x} = x] = \beta_0 + \beta_1 x$ (both models expressed in our notation).

But habits and paradigms are hard to identify, and hence to change. This can dispose one to read a model into, rather than out of, the data. Some have not pondered whether their models are truly reasonable and whether they really fit the data. As for a lack of fit, Barnett and Zehnwirth [2, p. 250] faults the Mack model: "It turns out that the assumption that, conditional on $x(i)$, the "average" value of $y(i)$ is $bx(i)$, is rarely true for real loss development arrays." And apart from the question of fitness, the Murphy-like addition of an intercept makes for an unreasonable model. For consider the general model of Figure 1A, whose parameters are estimated in Exhibit 5. The model for the six observations is $\mathbf{y} = \$1{,}094{,}448 + 2.768x + \mathbf{e}$, where the standard deviation of the error term is $\sqrt{2.134\text{E}+12} =$ \$1,460,829. Above all, a model should be reasonable, and we fail to understand why the loss at 24 months should start from a base amount of \$1,094,448. Accordingly, Gregory Alff [1, p. 89] has written, "A constant does nothing to describe the underlying contributory causes of change in the dependent variable." We can even imagine a condition in which an intercept must be rejected, namely, when one derives a negative intercept and projects from such a small x value that the projection itself is negative.

Furthermore, since the centroid is in the first quadrant, if the intercept is positive, the slope of the general line must be less than that of the constrained line. The flattened slope might well not differ significantly from unity.[9] In such a situation it is not the *ratio* of the later loss to the earlier that is important, but the *difference* of the later loss from the earlier. If $y - x$ is fairly constant, then it deserves modeling, not the quotient y/x.

Finally, the exposures of the six accident years range must vary widely. For the growth of premium from \$4.3 to \$12.0 million (and to \$12.9 million in 1991) cannot be attributed to rate increases; exposure must be climbing over these years. If the exposure of one AY were twice that of another, perhaps its intercept should be twice.[10] This reasonable thinking would lead the modeler to replace the constant with an exposure variable ξ: $\mathbf{y} = \beta_1 x + \beta_2 \xi + \mathbf{e}$. However, the earlier loss x and the exposure ξ usually compete, and one can be eliminated without much loss of explanatory power. The key to understanding this competition ("multicollinearity" in statistical parlance) is regression toward the mean, to which we now turn.

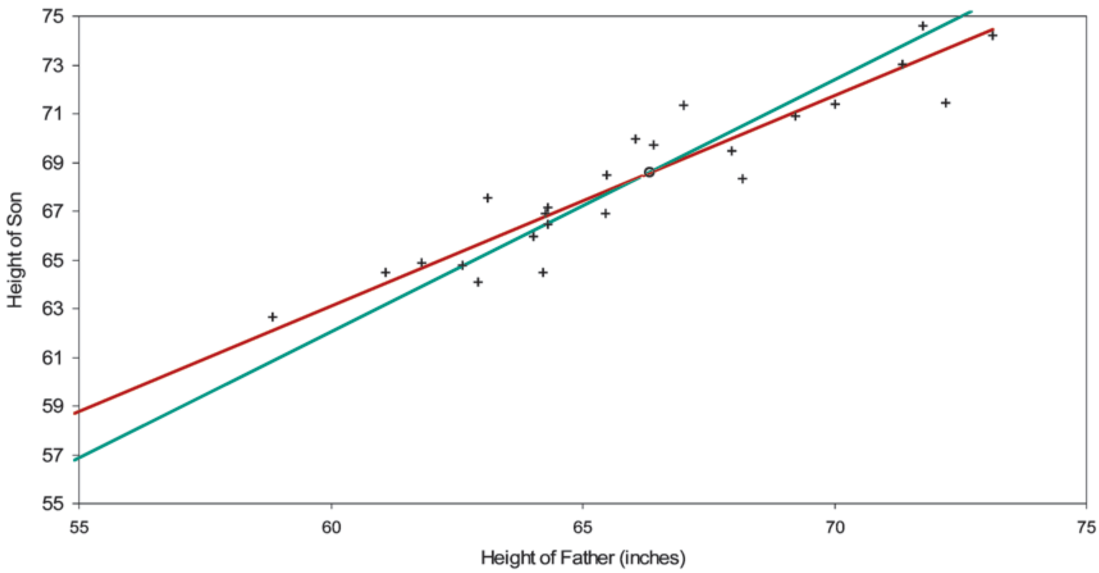
5. Regression toward the mean

The term "regression toward the mean" originated with Sir Francis Galton (1822–1911), who among other biological subjects applied it to the inheritance of height [7]. (See also [21].) In this section we will simplify the excellent discussion in Chapter 12 "Regression and Correlation" of Bulmer [5], and analogize from height inheritance to loss development.

[9] Both Venter [19, p. 815f and 821] and Barnett and Zehnwirth [2, p. 255 and pp. 258–260] note this.

[10] Murphy concedes at one place [13, p. 204]: "An increase in exposure from one accident year to the next will cause an upward parallel shift in the development regression line."

Figure 2A. Empirical heights of fathers and sons



In Figure 2A are graphed twenty-five pairs of heights of fathers and sons, with their centroid and their best-fitting general and constrained lines. Although these points are simulated, they have all the verisimilitude of the data that Galton studied. To reveal up front our simulation mechanism would spoil the joy of repeating Galton's discovery; we will discuss it shortly.

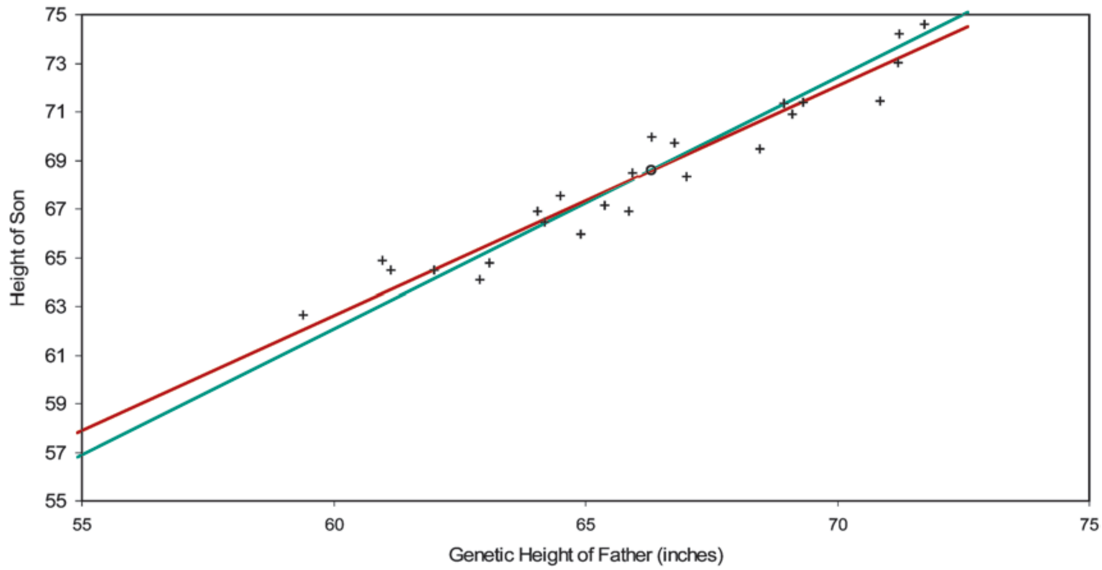
The centroid is (66.34, 68.60) inches. The intercept and the slope of the general line are 11.23 inches and 0.865; the slope of the constrained line is 1.034. We allowed for height "inflation" due to better care and nutrition: sons are on average about two inches taller than their fathers. One would like to regard a son's height as proportional to his father's; but the intercept of the general line with its standard error (11.23 ± 4.32 inches) puts the intercept at a quite significant 2.60 standard errors away from zero. In other words, the data rebels against a line through the origin.

Data that regresses toward the mean is not to be confused with a process that reverts toward the mean. Figure 2A is a picture of regression toward the mean; the best-fitting line, though sharing the centroid of the constrained line, lies between the constrained line and a horizontal line through the centroid. This implies that the intercept of the best line lies between zero and the ordinate of the centroid. A significant difference between the two lines means that somehow the data is defying proportionality.[11]

Galton was convinced that the inheritance of height should be a proportion; there is no reason why nature should begin with 11.23 inches. So he drew the ingenious distinction between genetic height and empirical height. With a tape measure one records empirical height; but one's *empirical* height is the sum of one's *genetic* height and an environmental error term of mean zero. If one could peer into genetic height, one would see that that a son's genetic height is proportional to his father's. If one knew the genetic heights of fathers, the line that best fit the points of the fathers' genetic heights and the sons' empirical heights would pass through the origin. Hence, a best-fitting line regresses toward the mean because the independent variable actually

[11]We know of no term for the opposite case, in which the constrained line lies between the general line and the centroidal horizon. An apt name for it might be "progression from the mean." That no one seems to have studied it indicates that it is deemed, when encountered, as a random accident. See footnote 17.

Figure 2B. Sons' empirical heights versus fathers' genetic heights

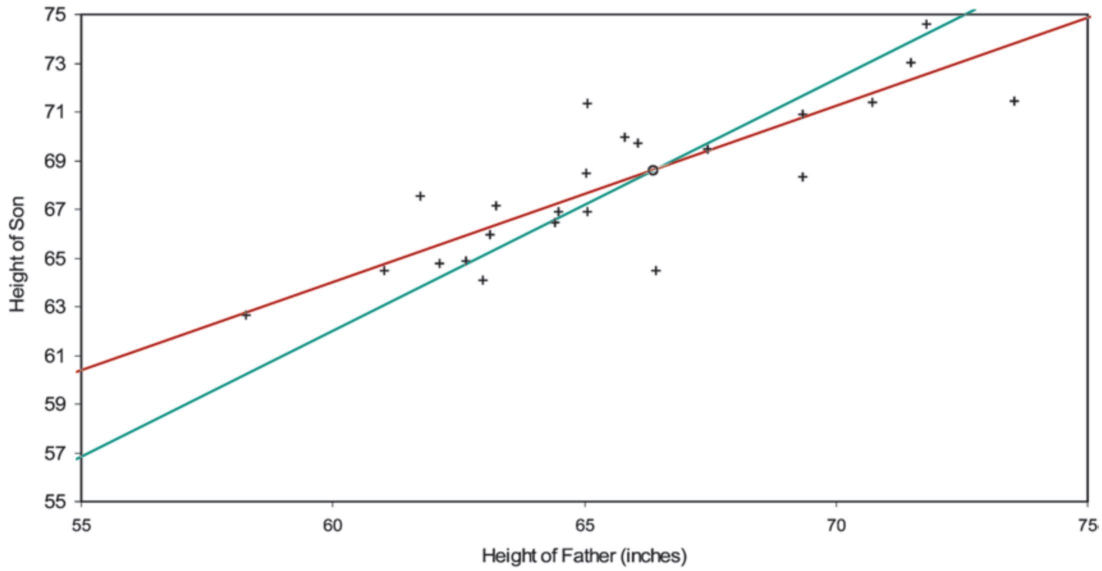


used is a proxy (even though an unbiased proxy) for a better or truer variable.

To a father 58.83 inches tall belongs leftmost point in Figure 2A; to a father 73.13 inches tall the rightmost. Even though the mean of the environmental error term is zero, given that we are looking at a relatively short father, we infer that he is genetically taller than 58.83 inches, and that his error term is negative. So the short father is probably genetically less short than he is empirically. Similarly, the tall father is probably genetically less tall than he is empirically. And the more a father's empirical height differs from the average (66.34 inches), the more his error term is expected to differ from zero. Lifting the veil from the fathers' genetic heights (which we can do, since it's a simulation), we have the result of Figure 2B. The centroid and the constrained line have not changed, but now the general line does not regress as much toward the mean. Its intercept and slope now are 5.96 inches and 0.945, which are not significantly different from the true values 0.00 inches and 1.030 (= 68/66). Switching the independent variable from empirical height to genetic contracts the *x*-axis toward the abscissa of the centroid and pivots counterclockwise the general line toward the constrained line.

Moreover, the better a variable proxies for the true variable, the less the general line will regress toward the mean; the worse it proxies, the more the line regresses. As for our simulation mechanism, the genetic heights of the fathers were normally distributed as 66 ± 4 inches. Then these values were multiplied by 68/66 to simulate the sons' genetic heights. To both sets of heights were added normally distributed environmental terms whose mean and standard deviation were zero and one inch. So here the fathers' genetic heights are relatively discernible; in actuarial parlance, the variance of the hypothetical means is $(4\text{ inches})^2$ and the expectation of the process variance is $(1\text{ inch})^2$. But in Figure 2C, starting with the same genetic heights and environmental errors, we've doubled the fathers' environmental errors. This increases the regression toward the mean, the intercept and slope of the general line (with standard errors) now being 20.59 ± 5.59 inches and 0.724 ± 0.084. Environmental error weighs more heavily on genetic height; actuaries would say that the fathers' heights are less

Figure 2C. Fathers' heights after doubling environmental effect



credible.[12] If the standard deviation of environmental error were many times the standard deviation of genetic height, a father's empirical height would indicate little about his or his son's genetic height, and the general line would be nearly flat—extreme regression toward the mean.[13]

The analogy of Galton's problem to loss development is straightforward. The father's empirical height corresponds to the loss at the earlier stage of development, the son's empirical height to the loss at the later stage. The common phenomenon of regression toward the mean indicates that the loss at the earlier stage is a proxy for something else. Without doubt, zero loss, as in Figure 1A, is a (misleading) proxy for a positive exposure.[14] But this, we believe, is *always* the case. Even if a line that best fits the adjacent columns of a loss triangle passes tolerably close to the origin, it indicates only that the earlier loss is a tolerable proxy. In Figure 1D, losses at 48 months were a tolerable proxy for a variable truly predictive of losses at 60 months; in Figures 1A–C the earlier losses were not good proxies.

To take another example, Medical Malpractice is regarded as a volatile line of business; yet our diagnosis of its paid loss development to 60 months (Exhibit 3) showed that a chain-ladder projection would happen to be reliable. The regression toward the mean in the 12-to-24-month stage is rendered innocuous by the fact that the projection of AY 2004 is close to the centroid. Nevertheless, the earlier losses are still just proxies of something else. Sometimes proxies are good enough; but when they are not, we must expend the effort to recover the true variables.

[12]Some actuaries would prefer to say that the sons' heights, as predicted from this new set of fathers' heights, are less credible. The next section will show that credibility can be interpreted either way. Hence, actuarial tradition has relegated credibility to judging the effectiveness of a proxy; by implication, the true variable is fully credible. Statistical modeling can free credibility for its true and rightful purpose of incorporating prior knowledge.

[13]Econometricians treat regression toward the mean as a theme within the problem of stochastic regressors (footnote 29). Halliwell [8, p. 441] was aware of the general problem in loss-triangle modeling, but did not recognize its specific nature as regression toward the mean. Judge et al. [10, pp. 582–585] and Kennedy [11, pp. 137–140 and p. 149] treat proxy variables as "measurement errors" and "errors in variables." Modeling with proxy variables is biased, even asymptotically with the luxury of limitless observation. For explanations of regression toward the mean in sports and investment (i.e., why winning and losing streaks end) see www.sportsci.org/resource/stats and www.travismorien.com/FAQ/portfolios/meanregression.htm.

[14]The loss incurred from zero exposure must have mean and variance zero. See Appendix C, Figure C1.

In the previous section we mustered reasons why an intercept should make way for an exposure variable. Models that incorporate both earlier loss x and exposure ξ usually suffer from multicollinearity; the earlier loss usually has little of its own to say, merely mimicking exposure. So it will come as no surprise by now that the earlier loss is none other than a proxy of exposure.[15] The loss triangle is a workhorse for the actuary, and will not be put out to pasture in the foreseeable future. But our diagnosis of triangles and our explanation of regression toward the mean together suggest that a column of exposures should be deemed an integral part of every loss triangle. The information really resides in the exposures—a fact obvious to ratemaking, but hardly less important to reserving.[16]

6. Credibility and regression toward the mean

In the previous section we made passing reference to actuarial credibility theory. One actuary, Gary G. Venter [18] related credibility theory to regression toward the mean; another, Eric Brosius [4] related it to loss development. But neither saw how significant regression toward the mean is to loss development, namely, that it indicates proxy variables. Furthermore, they both thought of credibility "forwards," i.e., as applying to the dependent variable. Venter, for example, wrote, "Least squares credibility can be thought of as a least-squares regression estimate in which the dependent variable has not yet been observed." [18, p. 134]. But the "backward" view to be presented here, applying credibility to the independent variable, will lend support to what we've just said about proxy variables.

Consider again the 12-to-24-month stage of the Brosius triangle, as modeled in Exhibit 5. The centroid of the six observations $(\bar{x},\bar{y})$ equals ($283,667, $1,879,500). The slope of the constrained line is $\bar{y}/\bar{x} = \$1{,}879{,}500/\$283{,}667 = 6.626 = \gamma$, and we express this line in the functional form $g(x) = \gamma x = 6.626x$. We express the general line in the form $f(x) = \beta_0 + \beta_1 x = \$1{,}094{,}448 + 2.768x$. Since both lines intersect at the centroid, $\bar{y} = f(\bar{x}) = g(\bar{x})$. Therefore:

$$\begin{aligned} f(x) &= 0 + f(x) \\ &= \bar{y} - f(\bar{x}) + f(x) \\ &= \bar{y} - \beta_0 - \beta_1\bar{x} + \beta_0 + \beta_1 x \\ &= \bar{y} + \beta_1(x - \bar{x}) \\ &= \bar{y} + \frac{\beta_1}{\gamma}(\gamma x - \gamma\bar{x}) \qquad [\Rightarrow \gamma \neq 0] \\ &= \bar{y} + \frac{\beta_1}{\gamma}(g(x) - \bar{y}) \\ &= \left(1 - \frac{\beta_1}{\gamma}\right)\bar{y} + \left(\frac{\beta_1}{\gamma}\right)g(x) \\ &= (1 - Z)\bar{y} + Zg(x). \end{aligned}$$

The general line can be interpreted as a credibility-weighted average of the constrained line and the centroidal horizon. The credibility given to the chain-ladder projection, i.e., to $g(x) = \gamma x$, is $Z = (\beta_1/\gamma) = (\beta_1\bar{x}/\bar{y})$, which in this case is 41.8%. The credibility of the CL projection is the ratio of the slopes, namely that of the general line to that of the constrained.[17] The complement of

[15] In Galton's example the proxy and true variables were of the same scale, i.e., $\mathbf{x} = \xi + \boldsymbol{\eta}$. But for linear modeling, only variable relativities matter, i.e., $\mathbf{x} \propto \xi + \boldsymbol{\eta}$.

[16] A nagging fear of the insurance industry, whether real or imagined, is that the exposures for which actuaries determine rates bear slight relation to the exposures that are underwritten. We see a call for consistency in Halliwell [8, p. 442]: "If there is enough information in the form of a loss triangle to produce development factors, then there must also be substantial knowledge of the underlying exposures. Otherwise, how would the actuary know that the rows of the triangle were commensurate, or that they represented the same process of development?" Feldblum [6, p. 784] specifies manual deviations and schedule credits as the major reasons for the discrepancy.

[17] So long as the intercept of the general line is between zero and $\bar{y}$, credibility will be between zero and one. Provided that $\bar{y}$ is positive, an intercept less than zero implies a credibility greater than one; an intercept greater than $\bar{y}$ (ignorable in practice) would imply a credibility less than zero. Footnote 11 suggested that an intercept less than zero should be deemed a random accident, to which Brosius [4, p. 17] would agree: "[W]hen this [negative intercept] happens we set $Z = 1$ and use a simple link ratio estimate, ignoring the budgeted loss estimate."

the credibility applies to the "prior hypothesis" $\bar{y}$.

Credibility-weighting the CL projection $y = g(x)$ with $\bar{y}$ is what we call the "forward" view of credibility; credibility is applied to the *dependent* variable y. Even more interesting, and revealing of the proxy status of earlier losses, is what we will call the "backward" view of credibility. The backward view is like the switching in the previous section of the *independent* variable x, an attempt to peer behind the proxy into the true variable. We remarked that contracting the x-axis toward the abscissa of the centroid, pivots the general line counterclockwise into coincidence with the constrained line.

Credibility-weight each observed independent variable x with the average value, i.e., with the abscissa of the centroid $\bar{x}$, to form the "contracted" variable $w = (1-Z)\bar{x} + Zx$. This does not disturb the centroid, because $\bar{w} = \bar{x}$. Because g is linear and invariant to Z, for *any* credibility Z:

$$\begin{aligned}(1-Z)\bar{y} + Zg(x) &= (1-Z)g(\bar{x}) + Zg(x)\\ &= g((1-Z)\bar{x} + Zx)\\ &= g(w).\end{aligned}$$

Consequently, credibility-weighting the CL-fitted dependent variables with $\bar{y}$ is equivalent to CL-fitting the credibility-weighted independent variables with $\bar{x}$; in other words, credibility-weighting and CL-fitting (i.e., regressing through the origin) are commutative at *any* credibility.

But consider the regression of y against w according to the general line $y = \gamma_0 + \gamma_1 w$. Using the formula of Appendix C and removing constants from the covariances, we derive:

$$\begin{aligned}\gamma_1 &= \frac{\operatorname{Cov}[w,y]}{\operatorname{Cov}[w,w]} = \frac{\operatorname{Cov}[(1-Z)\bar{x}+Zx,y]}{\operatorname{Cov}[(1-Z)\bar{x}+Zx,(1-Z)\bar{x}+Zx]}\\ &= \frac{\operatorname{Cov}[Zx,y]}{\operatorname{Cov}[Zx,Zx]} = \frac{1}{Z}\frac{\operatorname{Cov}[x,y]}{\operatorname{Cov}[x,x]}\\ &= \frac{1}{Z}\beta_1.\end{aligned}$$

Here β_1 is the slope estimator of the general line modeled on the uncontracted (or fully credible) independent variables. Desiring to unwind the regression toward the mean, we choose Z such that $\gamma_0 = 0$, or equivalently, $\gamma_1 = \bar{y}/\bar{w} = \bar{y}/\bar{x}$. Hence,

$$Z = \frac{1}{\left(\frac{1}{Z}\right)} = \frac{\beta_1}{\left(\frac{1}{Z}\beta_1\right)} = \frac{\beta_1}{\gamma_1} = \frac{\beta_1}{\left(\frac{\bar{y}}{\bar{x}}\right)} = \frac{\beta_1\bar{x}}{\bar{y}},$$

which is the credibility that we derived in the forward view. Though equivalent mathematically, the forward and backward views differ as to interpretation. Whereas the forward view faults the model while passing the data, the backward view faults the data while passing the model. Or, whereas the forward view shores up the outputs of the model, the backward view shores up the inputs. Recognizing this alternative viewpoint should open actuaries to consider earlier losses as exposure proxies. Just as tall fathers are genetically tall, but probably not quite as tall genetically as empirically, so too large losses conceal large relative exposures, but probably not quite as large relatively to the actual losses. Correspondingly, just as short fathers are genetically short, but probably not quite as short genetically as empirically, so too small losses conceal small relative exposures, but probably not quite as small relatively to the actual losses. Progress depends on piercing the veil.[18]

7. Standard models of loss development

The standard models of loss development are all of the same form:

$$\mathbf{y} = (X)\beta + \mathbf{e}, \qquad \operatorname{Var}[\mathbf{e}] = \sigma^2\Phi,$$

$$\mathbf{y}_{ij} = (a_{ij}\xi_i)\beta_j + \mathbf{e}_{ij}, \qquad \operatorname{Var}[e_{ij}] = \sigma^2\phi_{ij}.$$

[18] A tentative step behind the veil is to use earned premium as the independent variable, especially in reinsurance, where most exposure is in the form of "subject premium." However, because of underwriting cycles, earned premium is often a poor proxy of exposure, sometimes even a worse proxy than the developing losses! The next step is to place earned premiums on the same level. But this is still a half measure; the exposures themselves must be captured, as much in reserving as in ratemaking.

The dependent variable $\mathbf{y}_{ij}$ is the *incremental* loss of the ith exposure period at the jth stage of development, i.e., the ijth cell of the loss rectangle.

As an example, Schedule P exhibits, apart from the "Prior" line, have ten accident years (rows) and ten development years (columns), which make for $10 \times 10 = 100$ cells. If the experience is complete, 55 cells are observed (when $i + j \leq 11$) and 45 require prediction ($i + j > 11$). The vector $\mathbf{y}$ is then the 100 cells unraveled as a (100×1) column vector. The design matrix X has 100 rows; the number of its columns, which equals the number of parameters in β, depends on the model. The error term $\mathbf{e}$, like $\mathbf{y}$, is (100×1). The model requires the variance of $\mathbf{e}$, a (100×100) matrix of the covariances of the elements of $\mathbf{e}$.

Variance considerations lead us to incremental formulations. Because variance structures are on the order of n^2, it is desirable to keep them simple. The variances of these standard models are zero off the diagonal, but not necessarily constant (viz., unity) on the diagonal. In terms of the indices of the cells of the original rectangle:

$$\mathrm{Cov}[\mathbf{e}_{\mathrm{ij}}, \mathbf{e}_{\mathrm{kl}}] = \sigma^2 \begin{cases} \phi_{ij} & \text{if} \quad i = k \quad \text{and} \quad j = l \\ 0 & \text{otherwise} \end{cases}.$$

If all the ϕ_{ij} are equal, the variance structure is "homoskedastic"; otherwise it is "heteroskedastic." Homoskedasticity is the sparsest variance structure, on the order of n^0, or 1. Heteroskedasticity is a less sparse variance structure on the order of n; but it too asserts that no two cells of the rectangle covary. Non-covarying incremental losses within an exposure period imply covarying cumulative losses. For example, the covariance of cumulative losses at 48 months with those at 24 months in terms of 12-month incremental error terms is:

$$\begin{aligned} &\mathrm{Cov}[\mathbf{e}_{12} + \mathbf{e}_{24} + \mathbf{e}_{36} + \mathbf{e}_{48}, \mathbf{e}_{12} + \mathbf{e}_{24}] \\ &\quad = \mathrm{Cov}[\mathbf{e}_{12} + \mathbf{e}_{24}, \mathbf{e}_{12} + \mathbf{e}_{24}] \\ &\qquad + \mathrm{Cov}[\mathbf{e}_{36} + \mathbf{e}_{48}, \mathbf{e}_{12} + \mathbf{e}_{24}] \\ &\quad = \mathrm{Cov}[\mathbf{e}_{12} + \mathbf{e}_{24}, \mathbf{e}_{12} + \mathbf{e}_{24}] + 0 \\ &\quad = \mathrm{Var}[\mathbf{e}_{12} + \mathbf{e}_{24}]. \end{aligned}$$

The covariance of two intraperiod cumulative losses equals the variance of the earlier cumulative loss. Devising a variance for incremental losses whose cumulative variance is zero off the diagonal is no more than a mathematical challenge; it is quite unrealistic. To avail ourselves of the simplicity of homo- and heteroskedastic models, we must model incremental losses, not cumulative. Incremental losses may indeed covary; but non-covariance will be our default assumption, an assumption that is testable.[19]

The exposure of the ith exposure period is ξ_i, e.g., car-years, insured value, sales, or payroll. For the examples here, the a_{ij} factors will be unity; however, they allow the modeler to adjust, or to index, the exposures, much as Stanard [15, p. 128] did with Butsic's inflation model. Perhaps the awareness that losses as proxies impound a simple accident-date inflation lent credence to the CL model. However, this benefit does not validate the use of a proxy when the true variable is available; nor is this a benefit in the more complicated situations. Actuaries have done their homework when they explicitly calculate and apply the adjustments—not only to the observed part of the triangle, but also and especially to the future part.

Considerable thought should be given to the variance relativities ϕ_{ij}. Most basically, they should be proportional to exposure. But they should be quadratic to the adjustment factors. Moreover, though our examples will ignore it, models should recognize that, per unit of exposure, some development stages are more volatile than others. If υ_j denotes the variance in stage

[19]The most common test for (auto)correlation is the Durbin-Watson. Halliwell [8] has made a start at autocorrelated variance structures, according to which an error term in one increment correlates (positively or negatively) with the other incremental error terms. Another variance structure, one which considers inflation, provides for covariance between incremental losses of the same calendar periods. Barnett and Zehnwirth [2, p. 294] side with incremental-loss models for an empirical reason: "…for most real loss development arrays, ELRF analysis indicates that the data were generated incrementally."

j per unit of exposure, a reasonable formula is $\phi_{ij} = a_{ij}^2 \xi_i \upsilon_j$. Fortunately, however, less depends on the variance structure than on the design; an inexact or even an improper variance structure detracts from the optimality ("bestness"—the "B" in BLUE[20]) of the predictions, but not from the unbiasedness.

The equation $\mathbf{y}_{ij} = (a_{ij}\xi_i)\beta_j + \mathbf{e}_{ij}$ specifies the additive model, the fourth of Stanard's models. The parameter β_j represents the pure premium of the jth development stage. This would be the raw input to pricing a policy that covered the portion of incurred loss that emerges in the jth development stage. Perhaps not a marketable idea, it has theoretical value inasmuch as an ordinary policy could be synthesized from non-overlapping policies that cover all the time after inception.

The additive method is the most flexible of the standard methods; it presumes the least information about the loss development and hence has the greatest parameter variance. But if one knew what fraction of loss developed at each stage (f_j, which sum to unity),[21] one could express β_j as $f_j\beta$. The model would then be $\mathbf{y}_{ij} = (a_{ij}\xi_i f_j)\beta + \mathbf{e}_{ij}$, and only one parameter, β, the whole pure premium, would be estimated. This statistical model corresponds to the Stanard-Bühlmann, or Cape-Cod, method, the third of Stanard's methods.[22] Presuming more knowledge of the loss development and consequently reducing the number of parameters to just one, this model's parameter variance is less than that of the additive model.

[20]BLUE stands for "Best Linear Unbiased Estimation." See Appendix B.

[21]Such knowledge resides in collateral data sources. For example, many insurance companies, rating bureaus, and consulting firms compile "development factor" databases.

[22]The correspondence is not exact in that the *method* is applied only to the latest diagonal of the cumulative triangle, whereas the *model* applies to whole triangle of incremental losses. The model makes more use of the data than does the method. Feldblum's paper [6] on this method can be read appreciatively and profitably, provided that one recognize its limitations, viz., that it is method-oriented (deterministic) and that it finds no fault with the chain-ladder method and ignores the additive method.

Table 3. Continuum of standard models

$\mathbf{y} = (X)(\beta) + \mathbf{e}$	
$\mathbf{y}_{ij} = (a_{ij}\xi_i)(\beta_j \leftarrow f_j\beta) + \mathbf{e}_{ij}$	Additive
$\mathbf{y}_{ij} = (a_{ij}\xi_i f_j)(\beta) + \mathbf{e}_{ij}$	Stanard-Bühlmann $\sum f_j = 1$
$\mathbf{y}_{ij} = (a_{ij}\xi_i f_j \beta)(1) + \mathbf{e}_{ij}$	Bornhuetter-Ferguson

If one knows not only the relativities of the loss-development pattern (i.e., the f_j factors), but also the magnitude of it (i.e., the whole pure premium β),[23] one reduces to the Bornhuetter-Ferguson model $\mathbf{y}_{ij} = (a_{ij}\xi_i\beta_j)\cdot 1 + \mathbf{e}_{ij} = (a_{ij}\xi_i f_j\beta)\cdot 1 + \mathbf{e}_{ij}$, Stanard's second method. This is a "high-information" model in which no parameters are estimated; hence, its predictions have no parameter variance. At most, one estimates the variance magnitude σ^2, and derives the process variances $\hat{\sigma}^2\phi_{ij}$.

Table 3 shows the continuity of these three models in descending order of parameter complexity. The descending complexity appears in the leftward movement of variables from the parameter vector (β) to the design matrix (X). Finally, we will illustrate these models on the Brosius triangle.

8. Standard models of the Brosius triangle

In Exhibit 6 we have readied the information for the three models. The Brosius triangle has seven accident years and five development stages for a total of 35 cells. The observations form an upper-left trapezoid of 25 cells; the predictions form a lower-right triangle of 10 cells. These two regions are the two boxes of the exhibit, whose rows are indexed by AY and Age. For lack of anything better, earned premium will serve as our exposure ξ_i. Losses $\mathbf{y}_{ij}$ are here incrementalized from the cumulative format of

[23]The left hand of reserving could receive this information from the right hand of ratemaking or underwriting. Many reinsurers reserve some of their liabilities with loss-ratio "picks" and development patterns. This constitutes the Bornhuetter-Ferguson method, whose namesakes were reinsurance actuaries.

Exhibit 1. As mentioned in the previous section, all exposure and variance-relativity adjustments, a_{ij} and υ_j, are unity for the sake of simplicity. The next column contains the adjusted exposures $a_{ij}\xi_i$, which form the design matrix of the additive model. The following column equals $\phi_{ij} = a_{ij}^2\xi_i\upsilon_j$, which we above claimed as a reasonable formula for variance relativity. These relativities will be used in all three models. The rightmost two columns contain the elements of the design matrices of the Stanard-Bühlmann and Bornhuetter-Ferguson models, as per the formulas of Table 3. However, the f_j factors and the overall "pure premium" (more accurately here, "loss ratio") β, were borrowed from the additive solution $\boldsymbol{\beta}$ in Exhibit 7A. Therefore, the predictions from all three models must be the same; only the prediction-error variances will differ.

Exhibit 7A models the 25 observations ($\mathbf{y}$) according to the additive design (X). Because the additive model has five parameters, the additive entries of Exhibit 6 must be slotted into the proper columns, as determined by the Age index (i.e., [0–]12 months slots into the first column, [12–]24 months into the second, etc.). The "Φ" column is the main diagonal of a (25×25) matrix of variance relativities; but in Excel it is just easier to treat it as a column and to multiply element-wise instead of matrix-wise.

Below these matrices are intermediate calculations that lead to the estimator $\hat{\boldsymbol{\beta}} = (X'\Phi^{-1}X)^{-1} \cdot X'\Phi^{-1}\mathbf{y}$. This is the estimator of Appendix B, but it has been modified for heteroskedasticity (i.e., $\Phi \neq I$). Appendix C shows a particular instance of this modification. The formula for "Var[$\boldsymbol{\beta}$]" is $\sigma^2(X'\Phi^{-1}X)^{-1}$, the square root of whose diagonal is "Std[$\boldsymbol{\beta}$]." Thus we estimate probabilistically β_{12} as 0.043 ± 0.029.

Let M be the (25×3) matrix whose columns are $\mathbf{y}$, $X\boldsymbol{\beta}$, and $\mathbf{e}$ (the observed, fitted, and residual values). The "sums-of-squares-and-cross-products" (SSCP) matrix is $M'\Phi^{-1}M$. If a parameter is fit, the diagonal elements of this matrix satisfy the equation $m_{11} = m_{22} + m_{33}$. A rho-square statistic (without intercept) based on this equation allows us to say that this model explains 69.7% of what was observed. Finally,

$$\hat{\boldsymbol{\sigma}}^2 = \frac{m_{33}}{t-k} = \frac{(\mathbf{y}-X\hat{\boldsymbol{\beta}})'\Phi^{-1}(\mathbf{y}-X\hat{\boldsymbol{\beta}})}{t-k}$$

is the estimator for the scale of the variances; it is just the formula of Appendix B adjusted for heteroskedasticity (Φ^{-1}). The formula for the standardized residuals, Std[$\mathbf{e}$],[24] is $\mathbf{e}/\sqrt{\sigma^2\Phi}$. Standardized results outside the range $[-2,2]$ may indicate model deficiencies.

Exhibit 7B applies this solution to the prediction of the 10 cells of the lower-right triangle. The design of the prediction is the (10×5) matrix X_p. The expectation of the prediction is $E[\mathbf{y}_p] = X_p\hat{\boldsymbol{\beta}}$. Again, the "$\Phi$" column is the main diagonal of a (10×10) matrix of variance relativities. The formula for the prediction-error variance, i.e., for $\text{Var}[\mathbf{y}_p - E[\mathbf{y}_p]]$, is $X_p\text{Var}[\hat{\beta}]X_p' + \sigma^2\Phi$, whose two terms respectively are the parameter and the process variance. We summarized the X_p matrix by accident year, since the exhibit more easily accommodates (4×4) matrices than (10×10). All future "Age" values in the summarization are labeled "IBNR."

A powerful feature of the linear statistical model is that the best linear unbiased estimator (BLUE) of a linear combination of a random vector is the linear combination of the BLUE of the vector, i.e., BLUE[$A\mathbf{y}$] = A BLUE[$\mathbf{y}$]. So the formula $X_p\text{Var}[\hat{\beta}]X_p' + \sigma^2\Phi$ holds true, whether X_p and Φ are summarized or not. Both parts of the total prediction-error variance utilize the estimator $\hat{\boldsymbol{\sigma}}^2$. The square root of the diagonal of this variance matrix (the "Std" column) is the prediction-error standard deviation by accident year. Hence, this method probabilistically estimates AY 1991 IBNR as \$5,196,558 ±\$1,954,838. The variance of the AY 1988–1991

[24]To be precise, we should use Std[$\hat{\mathbf{e}}$], whose formula is derived in Halliwell [8, Appendix D].

Table 4. Summary of model predictions (Exhibits 7–9)

		BF Model		SB Model		Additive Model	
AY		$E[\mathbf{y}_p]$	Std[Pred]	$E[\mathbf{y}_p]$	Std[Pred]	$E[\mathbf{y}_p]$	Std[Pred]
1988	IBNR	770,164	601,298	770,164	622,379	770,164	824,468
1989	IBNR	1,582,078	931,518	1,582,078	974,250	1,582,078	1,277,014
1990	IBNR	2,501,198	1,210,602	2,501,198	1,280,547	2,501,198	1,649,700
1991	IBNR	5,196,558	1,448,683	5,196,558	1,635,443	5,196,558	1,954,838
Total	IBNR	10,049,998	2,189,412	10,049,998	2,611,616	10,049,998	3,890,789

total is the sum of all the elements of the total prediction-error variance; hence the total IBNR is $10,049,998 $\pm$ $3,890,789.

Exhibits 8A and 8B solve and predict according to the Stanard-Bühlmann model. This model has only one parameter; hence, the "SB" column of Exhibit 6 *is* the (25×1) design matrix X. The Bornhuetter-Ferguson (BF) model of Exhibits 9A and 9B seems to have one parameter; but in reality, it is set to unity, which is why the betas of Exhibit 9A are not in the bold font indicative of random variables, why the rho-square statistic is not applicable,[25] and why the variance of the parameter is zero. Hence, the BF parameter variance in Exhibit 9B is zero.

Table 4 compares the accident-year IBNR results of the three models. As mentioned above, because the f_j factors and the overall β were borrowed from the additive solution $\boldsymbol{\beta}$, the models yield the same expected results. But the table illustrates the increasing prediction-error variance, primarily due to the progression of parameters from zero, to one, and then to five.[26]

9. Conclusion

Is the chain-ladder method biased? In Section 2 we found the pioneering work of James Stanard suggestive of an answer, but not conclusive. In the next two sections we created a diagnostic tool and learned how the CL method behaves when regression lines refuse to pass through the origin. We discovered that regression toward the mean, coupled with business expansion, biases the CL method to overpredict. Next, in Section 5, Galton showed us that regression toward the mean is symptomatic of proxy variables. The following section reinforced this insight from the actuarial perspective of credibility. From all this we conclude that the chain-ladder method is biased. The bias most commonly takes the form of regression toward the mean, which indicates that earlier losses are serving as proxies for exposure—and the poorer the proxy, the more biased the method. The lesson to be learned from chain-ladder bias, or its meaning for us, is that loss-development models, in addition to being reasonable and empirically tested, should be free of proxy variables as much as possible. In Sections 7 and 8 we reformulated the Bornhuetter-Ferguson, Stanard-Bühlmann, and additive methods as a continuum of exposure-based loss-development models. It was because of this well-chosen base that Stanard found their performance superior to that of the chain-ladder method.

[25] Only our borrowing the additive solution keeps the SSCP matrix invariant; realistically, m_{11} would not equal the sum of m_{22} and m_{33}.

[26] Consequently, the M and SSCP matrices of the three models are identical. But the estimates for σ^2 differ because of the degrees of freedom. So increasing process variance is a secondary reason.

References

[1] Alff, G. N., "A Note Regarding Evaluations of Multiple Regression Models," *Proceedings of the Casualty Actuarial Society* 71, 1984, pp. 84–95.

[2] Barnett, G., and B. Zehnwirth, "Best Estimates for Reserves," *Proceedings of the Casualty Actuarial Society* 87, 2000, pp. 245–303.

[3] *Best's Aggregates and Averages, Property/Casualty*, 2005 edition, Oldwick, NJ, A. M. Best Company, 2005.

[4] Brosius, E., "Loss Development Using Credibility," CAS Study Note, March 1993.

[5] Bulmer, M. G., *Principles of Statistics*, NY: Dover, 1979.

[6] Feldblum, S., "The Stanard-Bühlmann Reserving Procedure—A Practitioner's Guide," Casualty Actuarial Society *Forum*, Fall 2002, pp. 777–822.

[7] Galton, Francis, "Regression Toward Mediocrity in Hereditary Stature," *Journal of the Anthropological Institute* 15, 1886, pp. 246–263.

[8] Halliwell, L. J., "Loss Prediction by Generalized Least Squares," *Proceedings of the Casualty Actuarial Society* 83, 1996, pp. 436–489.

[9] Hayne, R. M., "An Estimate of Statistical Variation in Development Factor Methods," *Proceedings of the Casualty Actuarial Society* 72, 1985, pp. 25–43.

[10] Judge, G. G., R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T.-C. Lee, *Introduction to the Theory and Practice of Econometrics* (2nd edition), New York: Wiley, 1988.

[11] Kennedy, P., *A Guide to Econometrics* (3rd edition), Cambridge: MIT Press, 1992.

[12] Mack, T., "Measuring the Variability of Chain Ladder Reserve Estimates," Casualty Actuarial Society *Forum*, Spring (1) 1994, p. 101.

[13] Murphy, D. M., "Unbiased Loss Development Factors," *Proceedings of the Casualty Actuarial Society* 81, 1994, p. 154.

[14] Stanard, J. N., "Experience Rates as Estimators: A Simulation of Their Bias and Variance," *Casualty Actuarial Society Discussion Paper Program*, May 1980, pp. 485–514, http://www.casact.org/pubs/dpp/dpp80/80dpp485.pdf.

[15] Stanard, J. N., "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *Proceedings of the Casualty Actuarial Society* 72, 1985, pp. 124.

[16] Tarbell, T. F. "Incurred But Not Reported Claim Reserves," *Proceedings of the Casualty Actuarial Society* 20, 1933–1934, pp. 275–280.

[17] Taylor, Greg, "Chain Ladder Bias," *ASTIN Bulletin* 33, 2003, pp. 313–330.

[18] Venter, G. G., "Credibility," Casualty Actuarial Society *Forum*, Fall 1987, p. 79.

[19] Venter, G. G., "Testing the Assumptions of Age-to-Age Factors," *Proceedings of the Casualty Actuarial Society* 85, 1998, pp. 807–847.

[20] Verrall, R. J., "Statistical Methods for the Chain Ladder Technique," Casualty Actuarial Society *Forum*, Spring (1) 1994, p. 393.

[21] Wikipedia contributors, "Regression Toward the Mean," *Wikipedia, The Free Encyclopedia*, http://en.wikipedia.org/w/index.php?title=Regression_toward_the_mean&oldid=161823968 (accessed October 2, 2007).

Appendix A

Exhibit 1. Brosius, "loss development using credibility," from [4], p. 16

			Cumulative Case-Incurred Losses			
AY	@12	@24	@36	@48	@60	EarnPrem
1985	102,000	104,000	209,000	650,000	847,000	4,260,000
1986	0	543,000	1,309,000	2,443,000	3,033,000	5,563,000
1987	412,000	2,310,000	3,083,000	3,358,000	4,099,000	7,777,000
1988	219,000	763,000	1,637,000	1,423,000		8,871,000
1989	969,000	4,090,000	3,801,000			10,645,000
1990	0	3,467,000				11,986,000
1991	932,000					12,873,000

		Link Ratios		
AY	24/12	36/24	48/36	60/48
1985	1.020	2.010	3.110	1.303
1986	#DIV/0!	2.411	1.866	1.242
1987	5.607	1.335	1.089	1.221
1988	3.484	2.145	0.869	
1989	4.221	0.929		
1990	#DIV/0!			
Avg	6.626	1.285	1.262	1.237

			Chain-Ladder Projections		
AY	@12	@24	@36	@48	@60
1985	102,000	104,000	209,000	650,000	847,000
1986	0	543,000	1,309,000	2,443,000	3,033,000
1987	412,000	2,310,000	3,083,000	3,358,000	4,099,000
1988	219,000	763,000	1,637,000	1,423,000	1,760,055
1989	969,000	4,090,000	3,801,000	4,797,864	5,934,298
1990	0	3,467,000	4,456,493	5,625,269	6,957,684
1991	932,000	6,175,184	7,937,603	10,019,347	12,392,554

Exhibit 2. Schedule P–Part D–Workers' Compensation
Cumulative paid net losses and defence and cost containment expenses

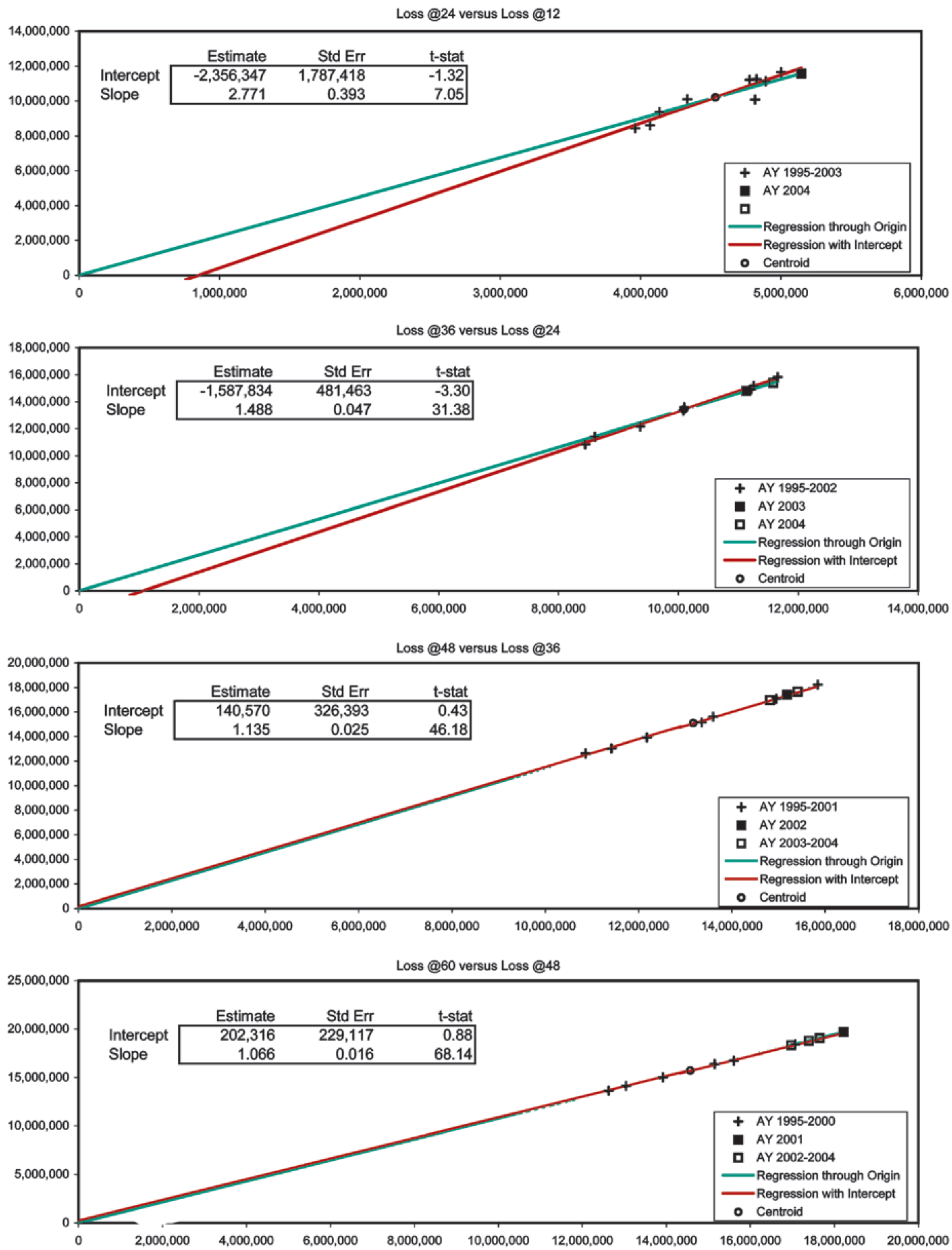
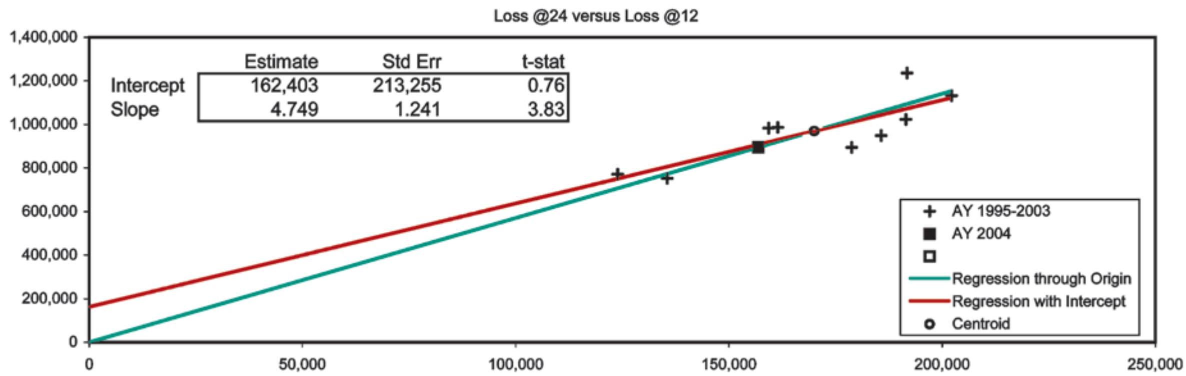
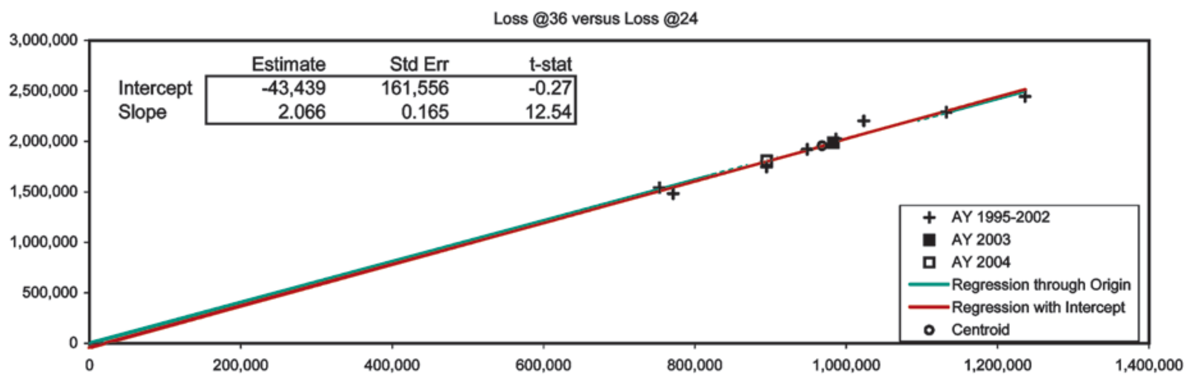
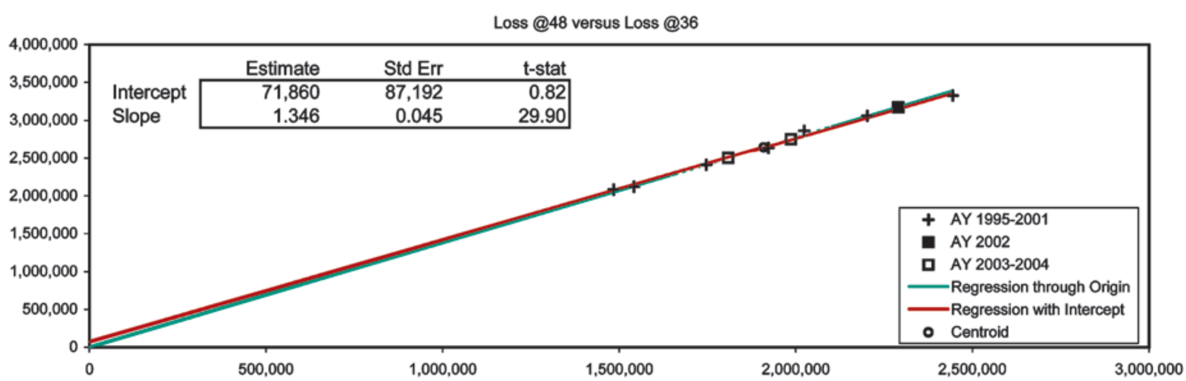


Exhibit 3. Schedule P–Part F–Section 2–Medical Malpractice–Claims-Made
Cumulative paid net losses and defence and cost containment expenses







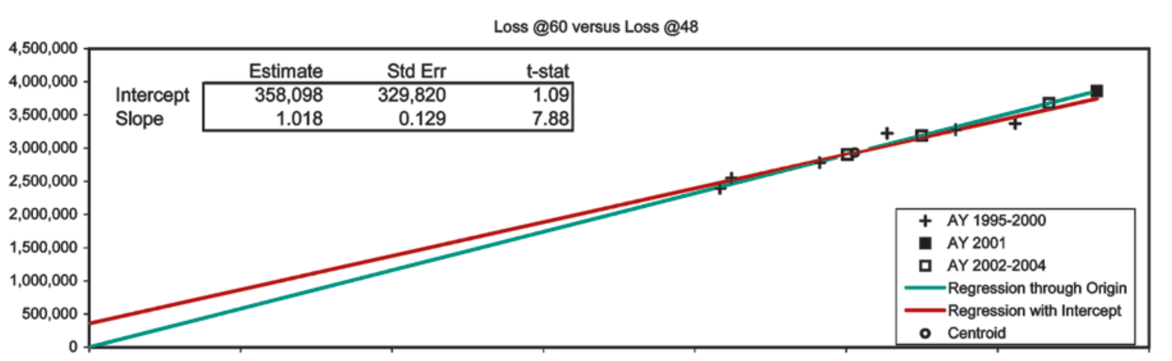


Exhibit 4. Schedule P–Part R–Section 1–Products Liability–Occurrence
Cumulative case-incurred net losses and defence and cost containment expenses

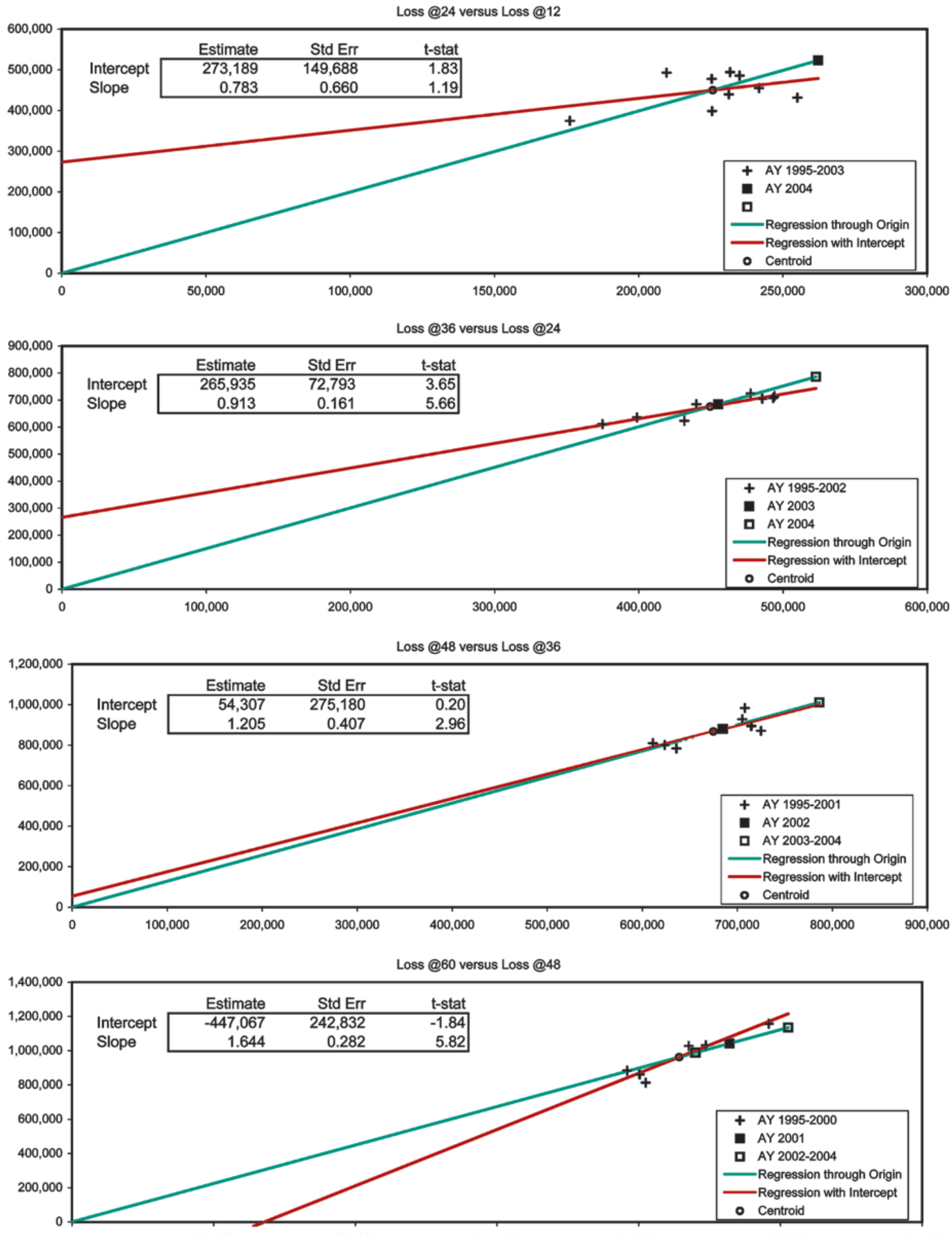
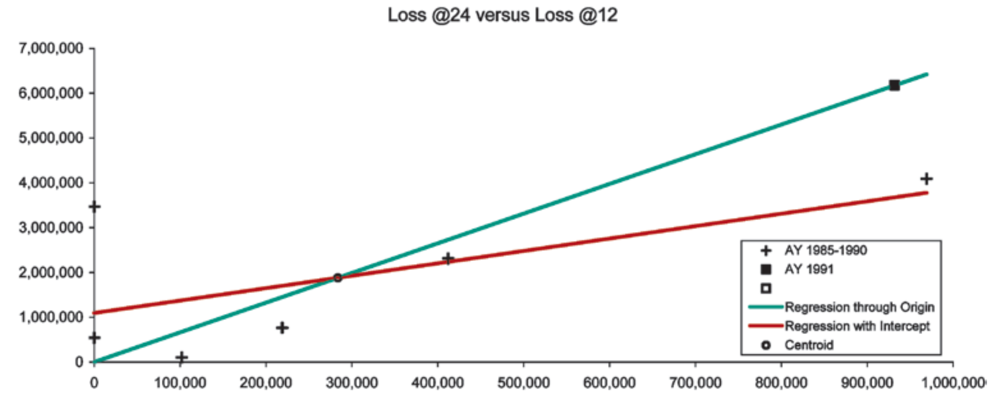


Exhibit 5. 12-24 diagnosis of the Brosius triangle



AY	1	x	y	xx	xy	Residuals
1985	1	102,000	104,000	1.040E+10	1.061E+10	−1,272,735
1986	1		543,000			−551,448
1987	1	412,000	2,310,000	1.697E+11	9.517E+11	75,336
1988	1	219,000	763,000	4.796E+10	1.671E+11	−937,534
1989	1	969,000	4,090,000	9.390E+11	3.963E+12	313,829
1990	1		3,467,000			2,372,552
1991						

Total	6	1,702,000	11,277,000	1.167E+12	5.093E+12	df	4
Avg	1	283,667	1,879,500	1.945E+11	8.488E+11	σ^2	2.134E+12

Two-Parameter Normal Equations

1	283,667	β_0	=	1,879,500
283,667	1.945E+11	β_1		8.488E+11

			Std[β]	t-stat
β_0	=	1,094,448	778,859	1.41
β_1		2.768	1.766	1.57

Inv $X'X$

0.28426185	−4.1455E-07
−4.1455E-07	1.46141E-12

Exhibit 6. Information for the Brosius models

AY i	Age j	Premium ξ_i	Incr Loss $\mathbf{y}_{ij}$	adjust a_{ij}	adjust υ_j	Additive $a_{ij}\xi_i$	ϕ_{ij}	SB $a_{ij}\xi_i f_j$	BF $a_{ij}\xi_i f_j \beta$
1985	12	4,260,000	102,000	1.000	1.000	4,260,000	4,260,000	405,788	181,054
1985	24	4,260,000	2,000	1.000	1.000	4,260,000	4,260,000	1,861,827	830,710
1985	36	4,260,000	105,000	1.000	1.000	4,260,000	4,260,000	573,388	255,834
1985	48	4,260,000	441,000	1.000	1.000	4,260,000	4,260,000	590,082	263,283
1985	60	4,260,000	197,000	1.000	1.000	4,260,000	4,260,000	828,916	369,845
1986	12	5,563,000	0	1.000	1.000	5,563,000	5,563,000	529,905	236,433
1986	24	5,563,000	543,000	1.000	1.000	5,563,000	5,563,000	2,431,301	1,084,797
1986	36	5,563,000	766,000	1.000	1.000	5,563,000	5,563,000	748,769	334,086
1986	48	5,563,000	1,134,000	1.000	1.000	5,563,000	5,563,000	770,570	343,813
1986	60	5,563,000	590,000	1.000	1.000	5,563,000	5,563,000	1,082,455	482,970
1987	12	7,777,000	412,000	1.000	1.000	7,777,000	7,777,000	740,801	330,530
1987	24	7,777,000	1,898,000	1.000	1.000	7,777,000	7,777,000	3,398,926	1,516,532
1987	36	7,777,000	773,000	1.000	1.000	7,777,000	7,777,000	1,046,769	467,047
1987	48	7,777,000	275,000	1.000	1.000	7,777,000	7,777,000	1,077,246	480,646
1987	60	7,777,000	741,000	1.000	1.000	7,777,000	7,777,000	1,513,257	675,185
1988	12	8,871,000	219,000	1.000	1.000	8,871,000	8,871,000	845,010	377,026
1988	24	8,871,000	544,000	1.000	1.000	8,871,000	8,871,000	3,877,057	1,729,865
1988	36	8,871,000	874,000	1.000	1.000	8,871,000	8,871,000	1,194,020	532,748
1988	48	8,871,000	−214,000	1.000	1.000	8,871,000	8,871,000	1,228,784	548,259
1989	12	10,645,000	969,000	1.000	1.000	10,645,000	10,645,000	1,013,993	452,423
1989	24	10,645,000	3,121,000	1.000	1.000	10,645,000	10,645,000	4,652,381	2,075,799
1989	36	10,645,000	−289,000	1.000	1.000	10,645,000	10,645,000	1,432,797	639,285
1990	12	11,986,000	0	1.000	1.000	11,986,000	11,986,000	1,141,730	509,417
1990	24	11,986,000	3,467,000	1.000	1.000	11,986,000	11,986,000	5,238,463	2,337,297
1991	12	12,873,000	932,000	1.000	1.000	12,873,000	12,873,000	1,226,222	547,115
1988	60	8,871,000		1.000	1.000	8,871,000	8,871,000	1,726,129	770,164
1989	48	10,645,000		1.000	1.000	10,645,000	10,645,000	1,474,513	657,898
1989	60	10,645,000		1.000	1.000	10,645,000	10,645,000	2,071,316	924,180
1990	36	11,986,000		1.000	1.000	11,986,000	11,986,000	1,613,293	719,819
1990	48	11,986,000		1.000	1.000	11,986,000	11,986,000	1,660,264	740,777
1990	60	11,986,000		1.000	1.000	11,986,000	11,986,000	2,332,249	1,040,603
1991	24	12,873,000		1.000	1.000	12,873,000	12,873,000	5,626,125	2,510,264
1991	36	12,873,000		1.000	1.000	12,873,000	12,873,000	1,732,681	773,088
1991	48	12,873,000		1.000	1.000	12,873,000	12,873,000	1,783,129	795,596
1991	60	12,873,000		1.000	1.000	12,873,000	12,873,000	2,504,843	1,117,610

SB Information	
f_{12}	0.095
f_{24}	0.437
f_{36}	0.135
f_{48}	0.139
f_{60}	0.195
Total	1.000

BF Information	
β	44.6%

Exhibit 7A. Additive model with solution

AY	Age	$\mathbf{y}$	X					Φ	$\mathbf{y}$	$X\beta$	$\mathbf{e}$	Std[$\mathbf{e}$]
1985	12	102,000	4,260,000	0	0	0	0	4,260,000	102,000	181,054	−79,054	−0.17
1985	24	2,000	0	4,260,000	0	0	0	4,260,000	2,000	830,710	−828,710	−1.78
1985	36	105,000	0	0	4,260,000	0	0	4,260,000	105,000	255,834	−150,834	−0.32
1985	48	441,000	0	0	0	4,260,000	0	4,260,000	441,000	263,283	177,717	0.38
1985	60	197,000	0	0	0	0	4,260,000	4,260,000	197,000	369,845	−172,845	−0.37
1986	12	0	5,563,000	0	0	0	0	5,563,000	0	236,433	−236,433	−0.44
1986	24	543,000	0	5,563,000	0	0	0	5,563,000	543,000	1,084,797	−541,797	−1.02
1986	36	766,000	0	0	5,563,000	0	0	5,563,000	766,000	334,086	431,914	0.81
1986	48	1,134,000	0	0	0	5,563,000	0	5,563,000	1,134,000	343,813	790,187	1.48
1986	60	590,000	0	0	0	0	5,563,000	5,563,000	590,000	482,970	107,030	0.20
1987	12	412,000	7,777,000	0	0	0	0	7,777,000	412,000	330,530	81,470	0.13
1987	24	1,898,000	0	7,777,000	0	0	0	7,777,000	1,898,000	1,516,532	381,468	0.61
1987	36	773,000	0	0	7,777,000	0	0	7,777,000	773,000	467,047	305,953	0.49
1987	48	275,000	0	0	0	7,777,000	0	7,777,000	275,000	480,646	−205,646	−0.33
1987	60	741,000	0	0	0	0	7,777,000	7,777,000	741,000	675,185	65,815	0.10
1988	12	219,000	8,871,000	0	0	0	0	8,871,000	219,000	377,026	−158,026	−0.24
1988	24	544,000	0	8,871,000	0	0	0	8,871,000	544,000	1,729,865	−1,185,865	−1.76
1988	36	874,000	0	0	8,871,000	0	0	8,871,000	874,000	532,748	341,252	0.51
1988	48	−214,000	0	0	0	8,871,000	0	8,871,000	−214,000	548,259	−762,259	−1.13
1989	12	969,000	10,645,000	0	0	0	0	10,645,000	969,000	452,423	516,577	0.70
1989	24	3,121,000	0	10,645,000	0	0	0	10,645,000	3,121,000	2,075,799	1,045,201	1.42
1989	36	−289,000	0	0	10,645,000	0	0	10,645,000	−289,000	639,285	−928,285	−1.26
1990	12	0	11,986,000	0	0	0	0	11,986,000	0	509,417	−509,417	−0.65
1990	24	3,467,000	0	11,986,000	0	0	0	11,986,000	3,467,000	2,337,297	1,129,703	1.45
1991	12	932,000	12,873,000	0	0	0	0	12,873,000	932,000	547,115	384,885	0.48

$X'\Phi^{-1}\mathbf{y}$	$X'\Phi^{-1}X$				
2,634,000	61975000	0	0	0	0
9,575,000	0	49102000	0	0	0
2,229,000	0	0	37116000	0	0
1,636,000	0	0	0	26471000	0
1,528,000	0	0	0	0	17600000

SSCP		
3365661.6	2346725.299	1018936.301
2346725.299	2346725.299	1.09139E-10
1018936.301	2.00089E-11	1018936.301
100.0%	69.7%	30.3%

β	$(X'\Phi^{-1}X)^{-1}$				
0.043	1.61355E-08	0	0	0	0
0.195	0	2.03658E-08	0	0	0
0.060	0	0	2.69426E-08	0	0
0.062	0	0	0	3.77772E-08	0
0.087	0	0	0	0	5.68182E-08

t	25
k	5
df	20
σ^2	50946.81507

Std[β]	Var[β]				
0.029	0.000822054	0	0	0	0
0.032	0	0.001037571	0	0	0
0.037	0	0	0.001372638	0	0
0.044	0	0	0	0.001924628	0
0.054	0	0	0	0	0.002894705

Exhibit 7B. Additive model predictions

AY	Age	$E[\mathbf{y}_p]$	X_p					Φ
1988	60	770,164	0	0	0	0	8,871,000	8,871,000
1989	48	657,898	0	0	0	10,645,000	0	10,645,000
1989	60	924,180	0	0	0	0	10,645,000	10,645,000
1990	36	719,819	0	0	11,986,000	0	0	11,986,000
1990	48	740,777	0	0	0	11,986,000	0	11,986,000
1990	60	1,040,603	0	0	0	0	11,986,000	11,986,000
1991	24	2,510,264	0	12,873,000	0	0	0	12,873,000
1991	36	773,088	0	0	12,873,000	0	0	12,873,000
1991	48	795,596	0	0	0	12,873,000	0	12,873,000
1991	60	1,117,610	0	0	0	0	12,873,000	12,873,000

Summarized by AY

AY								
1988	IBNR	770,164	0	0	0	0	8,871,000	8,871,000
1989	IBNR	1,582,078	0	0	0	10,645,000	10,645,000	21,290,000
1990	IBNR	2,501,198	0	0	11,986,000	11,986,000	11,986,000	35,958,000
1991	IBNR	5,196,558	0	12,873,000	12,873,000	12,873,000	12,873,000	51,492,000
Total	IBNR	10,049,998						

AY		Std
1988	IBNR	824,468
1989	IBNR	1,277,014
1990	IBNR	1,649,700
1991	IBNR	1,954,838
Total	IBNR	3,890,789

Total Prediction-Error Variance

6.797E+11	2.734E+11	3.078E+11	3.306E+11
2.734E+11	1.631E+12	6.149E+11	6.604E+11
3.078E+11	6.149E+11	2.722E+12	9.554E+11
3.306E+11	6.604E+11	9.554E+11	3.821E+12

Parameter Variance

2.278E+11	2.734E+11	3.078E+11	3.306E+11
2.734E+11	5.461E+11	6.149E+11	6.604E+11
3.078E+11	6.149E+11	8.896E+11	9.554E+11
3.306E+11	6.604E+11	9.554E+11	1.198E+12

Process Variance

4.519E+11			
	1.085E+12		
		1.832E+12	
			2.623E+12

Exhibit 8A. Stanard-Bühlmann model with solution

AY	Age	$\mathbf{y}$	X	Φ	$\mathbf{y}$	$X\beta$	$\mathbf{e}$	Std[$\mathbf{e}$]
1985	12	102,000	405,788	4,260,000	102,000	181,054	−79,054	−0.19
1985	24	2,000	1,861,827	4,260,000	2,000	830,710	−828,710	−1.95
1985	36	105,000	573,388	4,260,000	105,000	255,834	−150,834	−0.35
1985	48	441,000	590,082	4,260,000	441,000	263,283	177,717	0.42
1985	60	197,000	828,916	4,260,000	197,000	369,845	−172,845	−0.41
1986	12	0	529,905	5,563,000	0	236,433	−236,433	−0.49
1986	24	543,000	2,431,301	5,563,000	543,000	1,084,797	−541,797	−1.11
1986	36	766,000	748,769	5,563,000	766,000	334,086	431,914	0.89
1986	48	1,134,000	770,570	5,563,000	1,134,000	343,813	790,187	1.63
1986	60	590,000	1,082,455	5,563,000	590,000	482,970	107,030	0.22
1987	12	412,000	740,801	7,777,000	412,000	330,530	81,470	0.14
1987	24	1,898,000	3,398,926	7,777,000	1,898,000	1,516,532	381,468	0.66
1987	36	773,000	1,046,769	7,777,000	773,000	467,047	305,953	0.53
1987	48	275,000	1,077,246	7,777,000	275,000	480,646	−205,646	−0.36
1987	60	741,000	1,513,257	7,777,000	741,000	675,185	65,815	0.11
1988	12	219,000	845,010	8,871,000	219,000	377,026	−158,026	−0.26
1988	24	544,000	3,877,057	8,871,000	544,000	1,729,865	−1,185,865	−1.93
1988	36	874,000	1,194,020	8,871,000	874,000	532,748	341,252	0.56
1988	48	−214,000	1,228,784	8,871,000	−214,000	548,259	−762,259	−1.24
1989	12	969,000	1,013,993	10,645,000	969,000	452,423	516,577	0.77
1989	24	3,121,000	4,652,381	10,645,000	3,121,000	2,075,799	1,045,201	1.55
1989	36	−289,000	1,432,797	10,645,000	−289,000	639,285	−928,285	−1.38
1990	12	0	1,141,730	11,986,000	0	509,417	−509,417	−0.71
1990	24	3,467,000	5,238,463	11,986,000	3,467,000	2,337,297	1,129,703	1.58
1991	12	932,000	1,226,222	12,873,000	932,000	547,115	384,885	0.52

$X'\Phi^{-1}\mathbf{y}$	$X'\Phi^{-1}X$
5,259,595	11788058.91

β	$(X'\Phi^{-1}X)^{-1}$
0.446	8.48316E-08

Std[β]	Var[β]
0.060	0.003601584

SSCP

3365661.6	2346725.299	1018936.301
2346725.299	2346725.299	-9.82254E-11
1018936.301	-1.2551E-10	1018936.301

100.0%	69.7%	30.3%

t	25
k	1
df	24
σ^2	42455.67923

Exhibit 8B. Stanard-Bühlmann predictions

AY	Age	$E[\mathbf{y}_p]$	X_p	Φ
1988	60	770,164	1,726,129	8,871,000
1989	48	657,898	1,474,513	10,645,000
1989	60	924,180	2,071,316	10,645,000
1990	36	719,819	1,613,293	11,986,000
1990	48	740,777	1,660,264	11,986,000
1990	60	1,040,603	2,332,249	11,986,000
1991	24	2,510,264	5,626,125	12,873,000
1991	36	773,088	1,732,681	12,873,000
1991	48	795,596	1,783,129	12,873,000
1991	60	1,117,610	2,504,843	12,873,000

AY			*Summarized by AY*	
1988	IBNR	770,164	1,726,129	8,871,000
1989	IBNR	1,582,078	3,545,829	21,290,000
1990	IBNR	2,501,198	5,605,807	35,958,000
1991	IBNR	5,196,558	11,646,778	51,492,000
Total	IBNR	10,049,998		

AY		Std
1988	IBNR	622,379
1989	IBNR	974,250
1990	IBNR	1,280,547
1991	IBNR	1,635,443
Total	IBNR	2,611,616

Total Prediction-Error Variance

3.874E+11	2.204E+10	3.485E+10	7.241E+10
2.204E+10	9.492E+11	7.159E+10	1.487E+11
3.485E+10	7.159E+10	1.640E+12	2.351E+11
7.241E+10	1.487E+11	2.351E+11	2.675E+12

Parameter Variance

1.073E+10	2.204E+10	3.485E+10	7.241E+10
2.204E+10	4.528E+10	7.159E+10	1.487E+11
3.485E+10	7.159E+10	1.132E+11	2.351E+11
7.241E+10	1.487E+11	2.351E+11	4.885E+11

Process Variance

3.766E+11			
	9.039E+11		
		1.527E+12	
			2.186E+12

Exhibit 9A. Bornhuetter-Ferguson model with solution

AY	Age	$\mathbf{y}$	X	Φ	$\mathbf{y}$	$X\beta$	$\mathbf{e}$	Std[$\mathbf{e}$]
1985	12	102,000	181,054	4,260,000	102,000	181,054	−79,054	−0.19
1985	24	2,000	830,710	4,260,000	2,000	830,710	−828,710	−1.99
1985	36	105,000	255,834	4,260,000	105,000	255,834	−150,834	−0.36
1985	48	441,000	263,283	4,260,000	441,000	263,283	177,717	0.43
1985	60	197,000	369,845	4,260,000	197,000	369,845	−172,845	−0.41
1986	12	0	236,433	5,563,000	0	236,433	−236,433	−0.50
1986	24	543,000	1,084,797	5,563,000	543,000	1,084,797	−541,797	−1.14
1986	36	766,000	334,086	5,563,000	766,000	334,086	431,914	0.91
1986	48	1,134,000	343,813	5,563,000	1,134,000	343,813	790,187	1.66
1986	60	590,000	482,970	5,563,000	590,000	482,970	107,030	0.22
1987	12	412,000	330,530	7,777,000	412,000	330,530	81,470	0.14
1987	24	1,898,000	1,516,532	7,777,000	1,898,000	1,516,532	381,468	0.68
1987	36	773,000	467,047	7,777,000	773,000	467,047	305,953	0.54
1987	48	275,000	480,646	7,777,000	275,000	480,646	−205,646	−0.37
1987	60	741,000	675,185	7,777,000	741,000	675,185	65,815	0.12
1988	12	219,000	377,026	8,871,000	219,000	377,026	−158,026	−0.26
1988	24	544,000	1,729,865	8,871,000	544,000	1,729,865	−1,185,865	−1.97
1988	36	874,000	532,748	8,871,000	874,000	532,748	341,252	0.57
1988	48	−214,000	548,259	8,871,000	−214,000	548,259	−762,259	−1.27
1989	12	969,000	452,423	10,645,000	969,000	452,423	516,577	0.78
1989	24	3,121,000	2,075,799	10,645,000	3,121,000	2,075,799	1,045,201	1.59
1989	36	−289,000	639,285	10,645,000	−289,000	639,285	−928,285	−1.41
1990	12	0	509,417	11,986,000	0	509,417	−509,417	−0.73
1990	24	3,467,000	2,337,297	11,986,000	3,467,000	2,337,297	1,129,703	1.62
1991	12	932,000	547,115	12,873,000	932,000	547,115	384,885	0.53

$X'\Phi^{-1}\mathbf{y}$	$X'\Phi^{-1}X$
2,346,725	2346725.299

β	$(X'\Phi^{-1}X)^{-1}$
1.000	4.26126E-07

Std[β]	Var[β]
0.000	0

SSCP

3365661.6	2346725.299	1018936.301
2346725.299	2346725.299	-9.82254E-11
1018936.301	-1.2551E-10	1018936.301

NA	NA	NA

t	25
k	0
df	25
σ^2	40757.45206

Exhibit 9B. Bornhuetter-Ferguson predictions

AY	Age	$E[\mathbf{y}_p]$	X_p	Φ
1988	60	770,164	770,164	8,871,000
1989	48	657,898	657,898	10,645,000
1989	60	924,180	924,180	10,645,000
1990	36	719,819	719,819	11,986,000
1990	48	740,777	740,777	11,986,000
1990	60	1,040,603	1,040,603	11,986,000
1991	24	2,510,264	2,510,264	12,873,000
1991	36	773,088	773,088	12,873,000
1991	48	795,596	795,596	12,873,000
1991	60	1,117,610	1,117,610	12,873,000

AY			*Summarized by AY*	
1988	IBNR	770,164	770,164	8,871,000
1989	IBNR	1,582,078	1,582,078	21,290,000
1990	IBNR	2,501,198	2,501,198	35,958,000
1991	IBNR	5,196,558	5,196,558	51,492,000
Total	IBNR	10,049,998		

AY		Std
1988	IBNR	601,298
1989	IBNR	931,518
1990	IBNR	1,210,602
1991	IBNR	1,448,683
Total	IBNR	2,189,412

Total Prediction-Error Variance

3.616E+11	0.000E+00	0.000E+00	0.000E+00
0.000E+00	8.677E+11	0.000E+00	0.000E+00
0.000E+00	0.000E+00	1.466E+12	0.000E+00
0.000E+00	0.000E+00	0.000E+00	2.099E+12

Parameter Variance

0.000E+00	0.000E+00	0.000E+00	0.000E+00
0.000E+00	0.000E+00	0.000E+00	0.000E+00
0.000E+00	0.000E+00	0.000E+00	0.000E+00
0.000E+00	0.000E+00	0.000E+00	0.000E+00

Process Variance

3.616E+11			
	8.677E+11		
		1.466E+12	
			2.099E+12

Appendix B. The theory of the linear statistical model[27]

The equation of the linear statistical model is $\mathbf{y} = X\beta + \mathbf{e}$. The $(t \times 1)$ random vector $\mathbf{y}$ is the dependent variable; it contains the t observations. The $(t \times k)$ matrix X is the "design" matrix; each of its k columns is an independent variable. The number of observations must at least equal the number of columns, and the columns must be linearly independent. The $(k \times 1)$ "parameter" vector β contains coefficients of the independent variables, because $X\beta = X_{.1}\beta_1 + \cdots + X_{.k}\beta_k$. The $(t \times 1)$ random vector $\mathbf{e}$ is the source of the randomness of the observations. Its mean is zero, and in this appendix we will assume the simple case of homoskedasticity: $\mathrm{Var}[\mathbf{e}] = \sigma^2 I_{(t\times t)}$. It is important to understand that β is not a random vector. If all the error terms were zero (i.e., $\sigma^2 = 0$), we would have t simultaneous equations in k variables. Since $\mathbf{y}$ is observed, the equations would have to be consistent, and the same β could be determined from any k of the t equations. However, the errors are not zero, thereby obscuring β. So β is a constant; owing to the presence of the randomness in the observations we cannot *solve* for β. Rather, we must *estimate* it, and it is the estimator that is the random variable.

The estimator for β of the linear statistical model is *linear* in $\mathbf{y}$; in symbols, $\tilde{\boldsymbol{\beta}} = W_{(k\times t)}\mathbf{y}$. The goal is to determine a suitable matrix W. First, we want an *unbiased* estimator, i.e, one whose expected value is β. But the expected value of the estimator is:

$$\begin{aligned} E[\tilde{\boldsymbol{\beta}}] &= E[W\mathbf{y}] \\ &= WE[\mathbf{y}] \\ &= WE[X\beta + \mathbf{e}] \\ &= WE[X\beta] \\ &= WX\beta. \end{aligned}$$

Unbiasedness limits our search to matrices W such that $WX = I_{(k\times k)}$, such matrices being known as "left-inverses" of X. When $t > k$ the set of left inverses is infinite, but the simplest left inverse is $(X'X)^{-1}X'$.[28] So we define the particular estimator $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$.

Second, of all unbiased estimators of β, we want the *best* one, or the one with the smallest variance. The following derivation of variance uses the rule $\mathrm{Var}[W\mathbf{e}] = W\mathrm{Var}[\mathbf{e}]W'$, the multivariate equivalent of the familiar scalar rule $\mathrm{Var}[a\mathbf{x}] = a^2\mathrm{Var}[\mathbf{x}]$:

$$\begin{aligned} \mathrm{Var}[\tilde{\boldsymbol{\beta}}] &= E[W\mathbf{y}] \\ &= W\mathrm{Var}[\mathbf{y}]W' \\ &= W\mathrm{Var}[X\beta + \mathbf{e}]W' \\ &= W\mathrm{Var}[\mathbf{e}]W' \\ &= W(\sigma^2 I_{(t\times t)})W' \\ &= \sigma^2 WW'. \end{aligned}$$

Note that this derivation employs the assumption of homoskedastic variance. The variance of the particular estimator is:

$$\begin{aligned} \mathrm{Var}[\hat{\boldsymbol{\beta}}] &= \sigma^2\{(X'X)^{-1}X'\}\{(X'X)^{-1}X'\}' \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}. \end{aligned}$$

We prove that the particular estimator is best (the Gauss-Markov theorem) by expressing the variance of the general estimator as the sum of the variance of the particular estimator and some other variance matrix:

$$\begin{aligned} \mathrm{Var}[\tilde{\boldsymbol{\beta}}] &= \sigma^2 WW' \\ &= \sigma^2(X'X)^{-1} + \sigma^2 WW' - \sigma^2(X'X)^{-1} \\ &= \mathrm{Var}[\hat{\boldsymbol{\beta}}] + \sigma^2(WW' - (X'X)^{-1}) \\ &= \mathrm{Var}[\hat{\boldsymbol{\beta}}] + \sigma^2(WW' - (X'X)^{-1} \\ &\quad - (X'X)^{-1} + (X'X)^{-1}) \end{aligned}$$

[27] To supplement this brief account we recommend the treatments in Chapter 5 and Appendix A of Judge et al. [10], Chapter 3 of Kennedy [11], Appendix A of Halliwell [8], and Appendix A of Barnett and Zehnwirth [2].

[28] The requirement that the columns of X be linearly independent ensures the existence of the inverse of the $(k \times k)$ matrix $X'X$. If X is square (i.e., $t = k$), $(X'X)^{-1}X'$, along with all other left inverses, simplifies to X^{-1}.

$$\begin{aligned}
&= \text{Var}[\hat{\beta}] + \sigma^2(WW' - I_{(k\times k)}(X'X)^{-1} \\
&\quad - (X'X)^{-1}I_{(k\times k)} + (X'X)^{-1}I_{(k\times k)}) \\
&= \text{Var}[\hat{\beta}] + \sigma^2(WW' - WX(X'X)^{-1} \\
&\quad - (X'X)^{-1}X'W' + (X'X)^{-1}X'X(X'X)^{-1}) \\
&= \text{Var}[\hat{\beta}] + \sigma^2\{W - (X'X)^{-1}X'\} \\
&\quad \times \{W' - X(X'X)^{-1}\} \\
&= \text{Var}[\hat{\beta}] + \{W - (X'X)^{-1}X'\} \\
&\quad \times (\sigma^2 I_{(t\times t)})\{W - (X'X)^{-1}X'\}' \\
&= \text{Var}[\hat{\beta}] + \{W - (X'X)^{-1}X'\} \\
&\quad \times \text{Var}[\mathbf{e}]\{W - (X'X)^{-1}X'\}' \\
&= \text{Var}[\hat{\beta}] + \text{Var}[W\mathbf{e} - (X'X)^{-1}X'\mathbf{e}] \\
&= \text{Var}[\hat{\beta}] + \text{Var}[\tilde{\beta} - \hat{\beta}] \\
&\geq \text{Var}[\hat{\beta}].
\end{aligned}$$

An essential element of this proof is the fact that W is a left inverse of X. Hence, we have just proven $\hat{\beta} = (X'X)^{-1}X'\mathbf{y}$ to be the best linear unbiased estimator (BLUE) of β.

Usually, σ^2 is not known, and must be estimated (without bias) as:

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - X\hat{\beta})'(\mathbf{y} - X\hat{\beta})}{t - k}.$$

This is the familiar sum of the squared residuals divided by the degrees of freedom.

The best linear unbiased prediction of $\mathbf{y}_p = X_p\beta + \mathbf{e}_p$, where $\mathbf{e}_p$ does not covary with $\mathbf{e}$, is $X_p\hat{\beta}$. The variance of its prediction error is $X_p\text{Var}[\hat{\beta}]X_p' + \text{Var}[\mathbf{e}_p]$.

Appendix C. General and constrained regression lines

The properties of regression lines are most easily seen when one transforms from the estimator form to the "normal equation" form:

$$\begin{aligned}
X'X\hat{\beta} &= X'X(X'X)^{-1}X'\mathbf{y} \\
&= X'\mathbf{y}.
\end{aligned}$$

Also helpful is to define the fitted values $\hat{\mathbf{y}} = X\hat{\beta}$ and the residuals $\hat{\mathbf{e}} = \mathbf{y} - X\hat{\beta}$. Various reformulations of the normal equation are:

$$\begin{aligned}
X'X\hat{\beta} &= X'\mathbf{y} \\
X'\hat{\mathbf{y}} &= X'\mathbf{y} \\
X'\hat{\mathbf{e}} &= X'(\mathbf{y} - \hat{\mathbf{y}}) = 0.
\end{aligned}$$

Now the design matrix for the general line (the line with two-parameters, the intercept and the slope) is:

$$X_{(t\times 2)} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_t \end{bmatrix}.$$

In scalar form, the ith observation is $\mathbf{y}_i = 1 \cdot \beta_0 + x_i\beta_1 + \mathbf{e}_i$. The matrix form is $\mathbf{y} = X\beta + \mathbf{e}$. The variance of $\mathbf{e}$ is homoskedastic, i.e., $\text{Var}[\mathbf{e}] = \sigma^2 I_{(t\times t)}$.

So the normal equation for this model is:

$$X'X\hat{\beta} = X'\mathbf{y}$$

$$\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_t \end{bmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_t \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_t \end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_t \end{bmatrix}$$

$$\begin{bmatrix} \sum 1 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum \mathbf{y}_i \\ \sum x_i\mathbf{y}_i \end{bmatrix}.$$

Dividing both sides by t will average the summations:

$$\begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \text{Avg}(xx) \end{bmatrix}\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{y}} \\ \text{Avg}(x\mathbf{y}) \end{bmatrix}.$$

According to the first row of this equation, $1 \cdot \hat{\beta}_0 + \bar{x}\hat{\beta}_1 = \bar{\mathbf{y}}$. This proves that the fitted line passes through the centroid $(\bar{x}, \bar{y})$.

Inverting the (2×2) matrix is easily done by formula, and gives the solution:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \frac{1}{\text{Avg(xx)} - \bar{\text{x}}\bar{\text{x}}}\begin{bmatrix} \text{Avg}(xx) & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}\begin{bmatrix} \bar{\mathbf{y}} \\ \text{Avg}(x\mathbf{y}) \end{bmatrix}.$$

The second row of this equation gives the formula for the slope estimator [4, p. 3], to which we add a mnemonic form which will prove useful in Section 6:[29]

$$\hat{\boldsymbol{\beta}}_1 = \frac{\mathrm{Avg}(x\mathbf{y}) - \bar{x}\bar{\mathbf{y}}}{\mathrm{Avg}(xx) - \bar{x}\bar{x}} = \frac{\mathrm{Cov}[x,\mathbf{y}]}{\mathrm{Cov}[x,x]}.$$

Next, the scalar form of the constrained line is $\mathbf{y}_i = x_i\gamma + \mathbf{e}_i$. (Changing the Greek letter for the parameter helps to distinguish the two lines.) Again, the error terms do not covary; however, let the variance matrix be diagonal in the ϕ relativities:

$$\mathrm{Var}[\mathbf{e}] = \boldsymbol{\sigma}^2\Phi = \boldsymbol{\sigma}^2\begin{bmatrix}\phi_1 & & \\ & \ddots & \\ & & \phi_t\end{bmatrix}.$$

This variance structure is heteroskedastic; the random errors are "differently spread." However, dividing each observation by its standard-deviation relativity makes the model homoskedastic in the new error term $\boldsymbol{\eta}$:

$$\frac{\mathbf{y}_\mathrm{i}}{\sqrt{\phi_\mathrm{i}}} = \frac{\mathrm{x_i}}{\sqrt{\phi_\mathrm{i}}}\gamma + \frac{\mathbf{e}_\mathrm{i}}{\sqrt{\phi_\mathrm{i}}} = \frac{\mathrm{x_i}}{\sqrt{\phi_\mathrm{i}}}\gamma + {}_\mathrm{i}, \qquad \mathrm{Var}[{}_\mathrm{i}] = \sigma^2.$$

The normal equation for the transformed model is:

$$\left[x_1/\sqrt{\phi_1}\cdots x_t/\sqrt{\phi_t}\right]\begin{bmatrix}x_1/\sqrt{\phi_1}\\ \vdots \\ x_t/\sqrt{\phi_t}\end{bmatrix}\hat{\gamma}$$

$$= \left[x_1/\sqrt{\phi_1}\cdots x_t/\sqrt{\phi_t}\right]\begin{bmatrix}\mathbf{y}_1/\sqrt{\phi_1}\\ \vdots \\ \mathbf{y}_t/\sqrt{\phi_1}\end{bmatrix}$$

$$\left[\sum\frac{x_i^2}{\phi_i}\right]\hat{\gamma} = \left[\sum\frac{x_i\mathbf{y}_i}{\phi_i}\right].$$

[29]The covariance form is just a memory device, because the independent variable x is not a random variable. Those who express regression in Bayesian terms (e.g., Mack [12] and Murphy [13], whose formulations are cited in Section 4; also Brosius [4, p. 7], and inconsistently Barnett and Zehnwirth [2, p. 252] and Halliwell [8, p. 446]) are to varying degrees unaware of the problem of stochastic regressors (see footnote 13). As the linear statistical model is defined, $E[\mathbf{y} = x\beta + \mathbf{e} \mid x = a]$ is no more meaningful than $E[f(x) \mid x = a] = f(a)$. What is meaningful, however, is $E[\mathbf{y} = x\beta + \mathbf{e} \mid \mathbf{e} = a]$.

If σ^2 must be estimated, the formula in several forms is:

$$\hat{\boldsymbol{\sigma}}^2 = \frac{\sum\left(\frac{\mathbf{y}_i}{\sqrt{\phi_i}} - \frac{x_i}{\sqrt{\phi_i}}\hat{\gamma}\right)^2}{t-1} = \frac{\sum\frac{1}{\phi_i}(\mathbf{y}_i - x_i\hat{\gamma})^2}{t-1}$$

$$= \frac{(\mathbf{y} - X\hat{\gamma})'\Phi^{-1}(\mathbf{y} - X\hat{\gamma})}{t-1}.$$

Barnett and Zehnwirth [2, pp. 249–251] consider the general power expression $\phi_i = x_i^\delta$, especially for $\delta \in \{0,1,2\}$. But if the dependent variable x is like exposure (which is the point of our argument from regression toward the mean), we would expect the exponent to be 1. For exposures are independent units, so the variance of a sum of n units is n times the variance of 1 unit. In this case the normal equation becomes:

$$\left[\sum\frac{x_i^2}{x_i}\right]\hat{\gamma} = \left[\sum\frac{x_i\mathbf{y}_i}{x_i}\right]$$

$$\left[\sum x_i\right]\hat{\gamma} = \left[\sum\mathbf{y}_i\right]$$

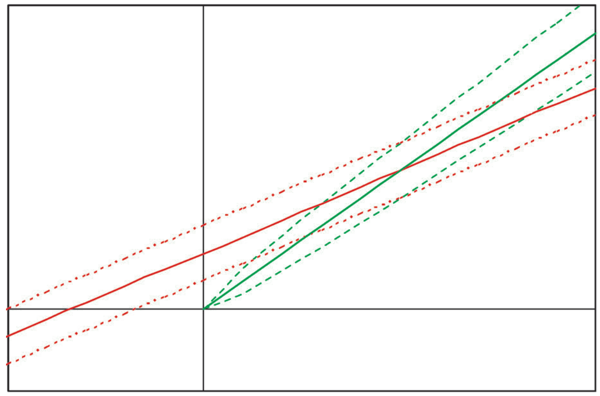
$$\bar{x}\hat{\gamma} = \bar{\mathbf{y}}.$$

The last form proves that the best-fitting constrained line passes through the centroid. But it does so necessarily only on the assumption that the variance of the error term is proportional to the dependent variable.

Figure C.1 graphically compares the two models. The general line with its one-standard-deviation boundaries is thee parallel lines. This homoskedastic model is oblivious to passing from quadrant to quadrant. The constrained line with its one-standard-deviation boundary radiates from the origin and makes sense only within the first and fourth quadrants.[30] We view this het-

[30]With a general variance structure the normal equation is $X'\Phi^{-1}X\hat{\beta} = X'\Phi^{-1}\mathbf{y}$. From this equation one can prove that when an intercept is added to the heteroskedastic model, the best-fitting line still intersects the centroid. Hence, the best-fitting line of a model that radiates from the y-axis shares the intersection of the solid lines of Figure B.1. Those troubled that the constrained line is not a special case of the general due to the different variance structures can redo Sections 3–6 with a heteroskedastic version of the general model, viz.: $\mathbf{y}_i = 1\cdot\beta_0 + x_i\beta_1 + \mathbf{e}_i$, $\mathrm{Var}[\mathbf{e}_i] = \sigma^2 x_i$.

Figure C.1.



eroskedastic, two-quadrant model as the more reasonable; in addition, we prefer it to such log-linear, or lognormal, models as those of Barnett and Zehnwirth [2], Hayne [9], and Verrall [20].[31,32]

[31]To adapt our notation in Section 7, log-linear models presuppose the equation $\mathbf{y}_{ij} = a_{ij}\xi_i\beta_j\mathbf{e}_{ij}$, where $\text{Prob}[\mathbf{e}_{ij} > 0] = 1$. Hence, the log-transformed version is $\ln\mathbf{y}_{ij} = \ln a_{ij} + \ln\xi_i + \ln\beta_j + \ln\mathbf{e}_{ij}$. Treating $\ln a_{ij}$ as a calendar-year parameter (i.e., $\ln a_{ij} = \gamma_{i+j}$ models inflation. Constant inflation simplifies to one parameter: $\ln a_{ij} = (i+j)\gamma$. This replicates most of the Barnett-Zehnwirth model, which fundamentally is a two-way (log-linear) ANOVA model. Which form is better depends empirically on whether error is additive/linear or multiplicative/log-linear. Too simplistically, Barnett and Zehnwirth [2, p. 294] argue, "The data are skewed, so we need to take a transformation." In reply, first, linear error terms do not have to be normal, or non-skewed; it is possible to keep linearity and to work with non-normal error terms (Judge et al. [10, Chapter 22] calls this "robust estimation"). Second, log-linear models concede to exposure insufficient importance; rather than *require* exposure (as ξ_i in the design matrix X), they allow one to *estimate* it (as $\ln\xi_i$ in the parameter vector β). Third, linear models admit observations whose values are zero or negative, whereas such observations must be dropped from log-linear models. And fourth, linear combinations like AY or CY totals derive analytically with linear modeling, but require Monte-Carlo simulation with log-linear modeling (cf. the "BLUE" paragraph of Section 8). Granted, log-linear models allow one to identify trends in exposure and inflation; however, identifying such trends exogenously from insurance and economic data would make them more trustworthy, and would place less estimation burden on the model. As for empirical error-term distributions, we have seen sets of standardized residuals whose skewness and kurtosis were ill-suited to the lognormal; even some whose skewness was negative. Both tails of a realistic linear error-term distribution, whatever its skewness and kurtosis, should be infinite; the lognormal distribution (centered about zero) does not serve well as a linear error-term distribution, because one of its tails is finite.

[32]Late in the editing of this paper we learned of a paper by Greg Taylor [17]. We have not plumbed the formidable mathematics of this paper; but its thesis (p. 313) is that "under rather general assumptions, [the chain ladder forecast] is biased upward." Its apparent agreement with ours is merely formal, forasmuch as its reasons for the bias differ from ours. Mr. Taylor says nothing about exposures, or about losses as exposure proxies.