[D6]
MEASURING THE VARIABILITY OF CHAIN LADDER RESERVE ESTIMATES

Contributed by T Mack

Abstract

The variability of chain ladder reserve estimates is quantified without assuming any specific claims amount distribution function. This is done by establishing a formula for the so-called standard error which is an estimate for the standard deviation of the outstanding claims reserve. The information necessary for this purpose is extracted only from the usual chain ladder formulae. With the standard error as a tool it is shown how a confidence interval for the outstanding claims reserve and for the ultimate claims amount can be constructed. Moreover, the analysis of the information extracted and of its implications shows when it may be appropriate to apply the chain ladder method and when it may not be.

Note

The original version of this paper was submitted to the prize paper competition "Variability of Loss Reserves" held by the Casualty Actuarial Society and was awarded a joint second prize. The present text differs from that paper in a few changes to the text and a changed and more thorough test procedure in Appendix H. This paper is included in the Claims Reserving Manual with the specific permission of the Casualty Actuarial Society, which otherwise retains ownership and all rights to continue to publish and disseminate this paper anywhere.

Contents

1. Introduction and Overview

The chain ladder method is probably the most popular method for estimating outstanding claims reserves. The main reason for this is its simplicity and the fact that it is distribution-free, that is, it seems to be based on almost no assumptions. In this paper, it will be seen that this impression is wrong and that the chain ladder algorithm has far-reaching implications. These implications also allow it to measure the variability of chain ladder reserve estimates. With the help of this measure it is possible to construct a confidence interval for the estimated ultimate claims amount and for the estimated reserves.

Such a confidence interval is of great interest for the practitioner because the estimated ultimate claims amount can never be an exact forecast of the true ultimate claims amount and therefore a confidence interval is of much greater information value. A confidence interval also allows one to consider business strategy in conjunction with the claims reserving process, using specific confidence probabilities. Moreover, there are many other claims reserving procedures and the results of all these procedures can vary widely. With the help of a confidence interval it can be seen whether the difference between the results of the chain ladder method and any other method is significant or not.

The structure of the paper is as follows. In section 2 a first basic assumption underlying the chain ladder method is derived from the formula used to estimate the ultimate claims amount. In section 3, the comparison of the age-to-age factor formula used by the chain ladder method with other possibilities leads to a second underlying assumption regarding the variance of the claims amounts. Using both of these derived assumptions and a third assumption on the independence of the accident years, it is possible to calculate the so-called standard error of the estimated ultimate claims amount. This is done in section 4, where it is also shown that this standard error is the appropriate measure of variability for the construction of a confidence interval. Section 5 illustrates how any given run-off triangle can be checked using some plots to ascertain whether the assumptions mentioned can be considered to be met. If these plots show that the assumptions do not seem to be met, the chain ladder method should not be applied without adaptation. In section 6 the formulae and two statistical tests (set out in Appendices G and H) are applied to a numerical example. For the sake of comparison, the reserves and standard errors according to a well-known claims reserving software package are also quoted. Complete and detailed proofs of all results and formulae are given in the Appendices A–F.

The proofs are quite long and take up about one fifth of the paper. However, the resulting formula for the standard error is very simple and can be applied directly after reading the basic notations in the first two paragraphs of section 2. In the numerical example, too, the formula for the standard error could be applied immediately to the run-off triangle. Instead, an analysis of whether the chain ladder assumptions are met in this particular case is made first. Because this analysis comprises many tables and plots, the example takes up another two fifths of the paper (including the tests in Appendices G and H).

2. Notation and First Analysis of the Chain Ladder Method

Let C_{ik} denote the accumulated total claims amount of accident year i, $1 \le i \le I$, either paid or incurred up to development year k, $1 \le k \le I$. The values of C_{ik} for $i + k \le I + 1$ are known to us (run-off triangle) and we want to estimate the values of C_{ik} for $i + k > I + 1$, in particular the ultimate claims amount C_{iI} of each accident year i = 2, ..., I. Then

$$R_i = C_{iI} - C_{i,I+1-i}$$

is the outstanding claims reserve of accident year i as $C_{i,I+1-i}$ has already been paid or incurred up to now.

The chain ladder method consists of estimating the ultimate claims amounts C_{iI} by

(1) $$\mathbf{C_{iI}} = C_{i,I+1-i} \cdot \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}}, \quad 2 \le i \le I$$

where

(2) $$\mathbf{f_k} = \sum_{j=1}^{I-k} C_{j,k+1} / \sum_{j=1}^{I-k} C_{jk}, \quad 1 \le k \le I-1$$

are the so-called age-to-age factors.

This manner of projecting the known claims amount $C_{i,I+1-i}$ to the ultimate claims amount $\mathbf{C_{iI}}$ uses for all accident years $i \ge I + 1 - k$ the same factor $\mathbf{f_k}$ for the increase of the claims amount from development year k to development year k+1, although the observed individual development factors $C_{i,k+1} / C_{ik}$ of the accident years $i \le I - k$ are usually different from one another and from $\mathbf{f_k}$. This means that each increase from C_{ik} to $C_{i,k+1}$ is considered a random disturbance of an expected increase from C_{ik} to $C_{ik} f_k$ where f_k is an unknown 'true' factor of increase which is the same for all accident years and which is estimated from the available data by $\mathbf{f_k}$.

Consequently, if we imagine to be at the end of development year k we have to consider $C_{i,k+1}, \ldots, C_{iI}$ as random variables whereas the realizations of $C_{i1}, \ldots, C_{ik}$ are known to us and are therefore no longer random variables but scalars. This means that for the purposes of analysis every C_{ik} can be a random variable or a scalar, depending on the development year at the end of which we imagine to be but independently of whether C_{ik} belongs to the known part $i + k \le I + 1$ of the run-off triangle or not. When taking expected values or variances we therefore must always also state the development year at the end of which we imagine to be. This will be done by explicitly indicating those variables C_{ik} whose values are assumed to be known. If nothing is indicated all C_{ik} are assumed to be unknown.

What we said above regarding the increase from C_{ik} to $C_{i,k+1}$ can now be formulated in stochastic terms as follows. The chain ladder method assumes the existence of accident-year-independent factors $f_1, \ldots, f_{I-1}$ such that, given the development $C_{i1}, \ldots, C_{ik}$, the realization of $C_{i,k+1}$ is 'close' to $C_{ik}f_k$, the latter being the expected value of $C_{i,k+1}$ in its mathematical meaning, that is

$$(3) \quad E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik}f_k, \quad 1 \le i \le I, \ 1 \le k \le I-1$$

Here to the right of the '$|$' those C_{ik} are listed which are assumed to be known. Mathematically speaking, (3) is a conditional expected value which is just the exact mathematical formulation of the fact that we already know $C_{i1}, \ldots, C_{ik}$, but do not know $C_{i,k+1}$. The same notation is also used for variances since they are specific expectations. The reader who is not familiar with conditional expectations should not refrain from further reading because this terminology is easily understandable and the usual rules for the calculation with expected values also apply to conditional expected values. Any special rule will be indicated wherever it is used.

We want to point out again that the equations (3) constitute an assumption which is not imposed by us but rather implicitly underlies the chain ladder method. This is based on two aspects of the basic chain ladder equation (1). One is the fact that (1) uses the same age-to-age factor $\mathbf{f}_k$ for different accident years $i = I + 1 - k, \ldots, I$. Therefore equations (3) also postulate age-to-age parameters f_k which are the same for all accident years. The other is the fact that (1) uses only the most recent observed value $C_{i,I+1-i}$ as basis for the projection to ultimate ignoring on the one hand all amounts $C_{i1}, \ldots, C_{i,I-i}$ observed earlier and on the other hand the fact that $C_{i,I+1-i}$ could substantially deviate from its expected value.

Note that it would easily be possible to also project to ultimate the amounts $C_{i1}, \ldots, C_{i,I-i}$ of the earlier development years with the help of the age-to-age factors $\mathbf{f}_1, \ldots, \mathbf{f}_{I-1}$ and to combine all these projected amounts together with $C_{i,I+1-i}\mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1}$ into a common estimator for C_{iI}. Moreover, it would also easily be possible to use the values $C_{j,I+1-i}$ of the earlier accident years $j < i$ as additional estimators for $E(C_{i,I+1-i})$ by translating them into accident year i with the help of a measure of volume for each accident year.

These possibilities are all ignored by the chain ladder method which uses $C_{i,I+1-i}$ as the only basis for the projection to ultimate. This means that the chain ladder method implicitly must use an assumption which states that the information contained in $C_{i,I+1-i}$ cannot be augmented by additionally using $C_{i1}, \ldots, C_{i,I-i}$ or $C_{1,I+1-i}, \ldots, C_{i-1,I+1-i}$. This is very well reflected by the equations (3) which state that, given $C_{i1}, \ldots, C_{ik}$, the expected value of $C_{i,k+1}$ only depends on C_{ik}.

Having now formulated this first assumption underlying the chain ladder method we want to emphasize that this is a rather strong assumption which has important consequences and which cannot be taken as met for every run-off triangle. Thus the widespread impression that the chain ladder method would work with almost no assumptions is not justified. In section 5 we will elaborate on the linearity constraint contained in assumption (3). But here we want to point out another consequence of formula (3). We can rewrite (3) in the form

$$E(C_{i,k+1} / C_{ik} \mid C_{i1}, \ldots, C_{ik}) = f_k$$

because C_{ik} is a scalar under the condition that we know $C_{i1}, \ldots, C_{ik}$. This form of (3) shows that the expected value of the individual development factor $C_{i,k+1} / C_{ik}$ equals f_k irrespective of the prior development $C_{i1}, \ldots, C_{ik}$ and especially of the foregoing development factor $C_{ik} / C_{i,k-1}$.

As is shown in Appendix G, this implies that subsequent development factors $C_{ik} / C_{i,k-1}$ and $C_{i,k+1} / C_{ik}$ are uncorrelated. This means that after a rather high value of $C_{ik} / C_{i,k-1}$ the expected size of the next development factor $C_{i,k+1} / C_{ik}$ is the same as after a rather low value of $C_{ik} / C_{i,k-1}$.

We therefore should not apply the chain ladder method to a business where we usually observe a rather small increase $C_{i,k+1} / C_{ik}$ if $C_{ik} / C_{i,k-1}$ is higher than in most other accident years, and vice versa. Appendix G also contains a test procedure to check this for a given run-off triangle.

3. Analysis of the Age-to-Age Factor Formula: the Key to Measuring the Variability

Because of the randomness of all realizations C_{ik} we can not infer the true values of the increase factors $f_1, ..., f_{I-1}$ from the data. They only can be estimated and the chain ladder method calculates estimators $\mathbf{f}_1, ..., \mathbf{f}_{I-1}$ according to formula (2). Among the properties which a good estimator should have, one prominent property is that the estimator should be unbiased, that is its expected value $E(\mathbf{f}_k)$ (under the assumption that the whole run-off triangle is not yet known) is equal to the true value f_k, in other words $E(\mathbf{f}_k) = f_k$. Indeed, this is the case here as is shown in Appendix A under the additional assumption that

(4) the variables $\{C_{i1}, ..., C_{iI}\}$ and $\{C_{j1}, ..., C_{jI}\}$ of different accident years $i \neq j$ are independent

Because the chain ladder method neither in (1) nor in (2) takes into account any dependency between the accident years we can conclude that the independence of the accident years is also an implicit assumption of the chain ladder method. We will therefore assume (4) for all further calculations. Assumption (4), too, cannot be taken as being met for every run-off triangle because certain calendar year effects (such as a major change in claims handling or in case reserving or greater changes in the inflation rate) can affect several accident years in the same way and can thus distort the independence. How such a situation can be recognized is shown in Appendix H.

A closer look at formula (2) reveals that

$$\mathbf{f}_k = \frac{\sum_{j=1}^{I-k} C_{j,k+1}}{\sum_{j=1}^{I-k} C_{jk}} = \sum_{j=1}^{I-k} \frac{C_{jk}}{\sum_{j=1}^{I-k} C_{jk}} \cdot \frac{C_{j,k+1}}{C_{jk}}$$

is a weighted average of the observed individual development factors $C_{j,k+1} / C_{jk}$, for $1 \leq j \leq I-k$, where the weights are proportional to C_{jk}. Like $\mathbf{f}_k$ every individual development factor $C_{j,k+1} / C_{jk}$, $1 \leq j \leq I-k$, is also an unbiased estimator of f_k because

$$\begin{aligned}
E(C_{j,k+1} / C_{jk}) &= E(E(C_{j,k+1} / C_{jk} \mid C_{j1}, ..., C_{jk})) && \text{(a)} \\
&= E(E(C_{j,k+1} \mid C_{j1}, ..., C_{jk}) / C_{jk}) && \text{(b)} \\
&= E(C_{jk} f_k / C_{jk}) && \text{(c)} \\
&= E(f_k) && \\
&= f_k && \text{(d)}
\end{aligned}$$

Here equality (a) holds due to the iterative rule E(X) = E(E(X|Y)) for expectations, (b) holds because, given C_{j1} to C_{jk}, C_{jk} is a scalar, (c) holds due to assumption (3) and

(d) holds because f_k is a scalar. (When applying expectations iteratively, e.g. E(E(X|Y)), one first takes the conditional expectation E(X|Y) assuming Y being known and then averages over all possible realizations of Y.)

Therefore the question arises as to why the chain ladder method uses just $\mathbf{f_k}$ as estimator for f_k and not the simple average

$$\frac{1}{I-k}\sum_{j=1}^{I-k} C_{j,k+1}/C_{jk}$$

of the observed development factors which also would be an unbiased estimator as is the case with any weighted average

$$g_k = \sum_{j=1}^{I-k} w_{jk}C_{j,k+1}/C_{jk} \quad \text{with} \sum_{j=1}^{I-k} w_{jk} = 1$$

of the observed development factors. (Here, w_{jk} must be a scalar if $C_{j1},\ldots,C_{jk}$ are known.)

Here we recall one of the principles of the theory of point estimation which states that among several unbiased estimators preference should be given to the one with the smallest variance, a principle which is easy to understand. We therefore should choose the weights w_{jk} in such a way that the variance of g_k is minimal. In Appendix B it is shown that this is the case if and only if (for fixed k and all j)

w_{jk} is inversely proportional to $\mathrm{Var}(C_{j,k+1}/C_{jk} \mid C_{j1},\ldots,C_{jk})$

The fact that the chain ladder estimator $\mathbf{f_k}$ uses weights which are proportional to C_{jk} therefore means that C_{jk} is assumed to be inversely proportional to $\mathrm{Var}(C_{j,k+1}/C_{jk} \mid C_{j1},\ldots,C_{jk})$, or stated the other way around, that

$$\mathrm{Var}(C_{j,k+1}/C_{jk} \mid C_{j1},\ldots,C_{jk}) = \alpha_k^2/C_{jk}$$

with a proportionality constant α_k^2 which may depend on k but not on j and which must be non-negative because variances are always non-negative.

Since here C_{jk} is a scalar and because generally $\mathrm{Var}(X/c) = \mathrm{Var}(X)/c^2$ for any scalar c, we can state the above proportionality condition also in the form

(5) $\mathrm{Var}(C_{j,k+1} \mid C_{j1},\ldots,C_{jk}) = C_{jk}\alpha_k^2,\ 1 \le j \le I,\ 1 \le k \le I-1$

with unknown proportionality constants α_k^2, $1 \le k \le I-1$.

As with assumptions (3) and (4), assumption (5) also has to be considered a basic condition implicitly underlying the chain ladder method. Again, condition (5) cannot a priori be assumed to be met for every run-off triangle. In section 5 we will show how to check a given triangle to see whether (5) can be considered met or not. But before doing so we turn to the most important consequence of (5): together with (3) and (4) it enables us to quantify the uncertainty in the estimation of C_{iI} by $\mathbf{C}_{iI}$.

4. Quantifying the Variability of the Ultimate Claims Amount

The aim of the chain ladder method and of every claims reserving method is the estimation of the ultimate claims amount C_{iI} for the accident years i = 2, ..., I. The chain ladder method does this by formula (1), that is

$$\mathbf{C}_{iI} = C_{i,I+1-i} \cdot \mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1}$$

This formula yields only a point estimate for C_{iI} which will normally turn out to be more or less wrong, that is there is only a very small probability for C_{iI} being equal to $\mathbf{C}_{iI}$. This probability is even zero if C_{iI} is considered to be a continuous variable. We therefore want to know in addition if the estimator $\mathbf{C}_{iI}$ is at least on average equal to the mean of C_{iI} and how large on average the error is. Precisely speaking we first would like to have the expected values $E(\mathbf{C}_{iI})$ and $E(C_{iI})$, $2 \le i \le I$, being equal. In Appendix C it is shown that this is indeed the case as a consequence of assumptions (3) and (4).

The second thing we want to know is the average distance between the forecast $\mathbf{C}_{iI}$ and the future realization C_{iI}. In Mathematical Statistics it is common to measure such distances by the square of the ordinary Euclidean distance ('quadratic loss function'). This means that one is interested in the size of the so-called mean squared error

$$\text{mse}(\mathbf{C}_{iI}) = E((C_{iI} - \mathbf{C}_{iI})^2 \mid D)$$

where $D = \{C_{ik} \mid i + k \le I + 1\}$ is the set of all data observed so far. It is important to realize that we have to calculate the mean squared error on the condition of knowing all data observed so far because we want to know the error due to <u>future</u> randomness only. If we calculated the unconditional error $E(C_{iI} - \mathbf{C}_{iI})^2$, which due to the iterative rule for expectations is equal to the mean value $E(E((C_{iI} - \mathbf{C}_{iI})^2 \mid D))$ of the conditional mse over all possible data sets D, we also would include all deviations from the data observed so far which obviously makes no sense if we want to establish a confidence interval for C_{iI} on the basis of the given particular run-off triangle D.

The mean squared error is exactly the same concept which also underlies the notion of the variance

$$\text{Var}(X) = E(X - E(X))^2$$

of any random variable X. Var(X) measures the average distance of X from its mean value E(X).

Due to the general rule $E(X-c)^2 = \mathrm{Var}(X) + (E(X)-c)^2$ for any scalar c we have

$$\mathrm{mse}(\mathbf{C}_{iI}) = \mathrm{Var}(C_{iI} \mid D) + (E(C_{iI} \mid D) - \mathbf{C}_{iI})^2$$

because $\mathbf{C}_{iI}$ is a scalar under the condition that all data D are known. This equation shows that the mse is the sum of the pure future random error $\mathrm{Var}(C_{iI} \mid D)$ and of the estimation error which is measured by the squared deviation of the estimate $\mathbf{C}_{iI}$ from its target $E(C_{iI} \mid D)$. On the other hand, the mse does not take into account any future changes in the underlying model, that is future deviations from the assumptions (3), (4) and (5), an extreme example of which was the emergence of asbestos. Modelling such deviations is beyond the scope of this paper.

As is to be expected and can be seen in Appendix D, $\mathrm{mse}(\mathbf{C}_{iI})$ depends on the unknown model parameters f_k and $\alpha_k{}^2$. We therefore must develop an estimator for $\mathrm{mse}(\mathbf{C}_{iI})$ which can be calculated from the known data D only. The square root of such an estimator is usually called 'standard error' because it is an estimate of the standard deviation of C_{iI} in cases in which we have to estimate the mean value, too. The standard error $\mathrm{s.e.}(\mathbf{C}_{iI})$ of $\mathbf{C}_{iI}$ is at the same time the standard error $\mathrm{s.e.}(\mathbf{R}_i)$ of the reserve estimate

$$\mathbf{R}_i = \mathbf{C}_{iI} - C_{i,I+1-i}$$

of the outstanding claims reserve

$$R_i = C_{iI} - C_{i,I+1-i}$$

because

$$\mathrm{mse}(\mathbf{R}_i) = E((\mathbf{R}_i - R_i)^2 \mid D) = E((\mathbf{C}_{iI} - C_{iI})^2 \mid D) = \mathrm{mse}(\mathbf{C}_{iI})$$

and because the equality of the mean squared errors also implies the equality of the standard errors. This means that

$$(6) \quad \mathrm{s.e.}(\mathbf{R}_i) = \mathrm{s.e.}(\mathbf{C}_{iI})$$

The derivation of a formula for the standard error $\mathrm{s.e.}(\mathbf{C}_{iI})$ of $\mathbf{C}_{iI}$ turns out to be the most difficult part of this paper; it is done in Appendix D. Fortunately, the resulting formula is simple

$$(7) \quad (\mathrm{s.e.}(\mathbf{C}_{iI}))^2 = \mathbf{C}_{iI}{}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_k{}^2}{\mathbf{f}_k{}^2} \cdot \left(\frac{1}{\mathbf{C}_{ik}} + \frac{1}{\sum\limits_{j=1}^{I-k} C_{jk}} \right)$$

where

$$(8) \quad \alpha_k^{\,2} = \frac{1}{I-k-1}\sum_{j=1}^{I-k} C_{jk}\left(\frac{C_{j,k+1}}{C_{jk}} - \mathbf{f}_k\right)^2, \; 1 \le k \le I-2$$

is an unbiased estimator of $\alpha_k^{\,2}$ (the unbiasedness being shown in Appendix E) and

$$\mathbf{C}_{ik} = C_{i,I+1-i} \cdot \mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{k-1}, \; k > I+1-i$$

are the amounts which are automatically obtained if the run-off triangle is completed step by step according to the chain ladder method. In (7), for notational convenience we have also set

$$\mathbf{C}_{i,I+1-i} = C_{i,I+1-i}$$

Formula (8) does not yield an estimator for α_{I-1} because it is not possible to estimate the two parameters f_{I-1} and α_{I-1} from the single observation $C_{1,I} / C_{1,I-1}$ between development years I – 1 and I. If $\mathbf{f}_{I-1} = 1$ and if the claims development is believed to be finished after I – 1 years we can put $\alpha_{I-1} = 0$. If not, we extrapolate the usually decreasing series $\alpha_1, \alpha_2, \ldots, \alpha_{I-3}, \alpha_{I-2}$ by one additional member, for instance by means of loglinear regression (see the example in section 6) or more simply by requiring that

$$\alpha_{I-3} / \alpha_{I-2} = \alpha_{I-2} / \alpha_{I-1}$$

holds at least as long as $\alpha_{I-3} > \alpha_{I-2}$.

This last possibility leads to

$$(9) \quad \alpha_{I-1}^2 = \min(\alpha_{I-2}^4 / \alpha_{I-3}^2, \min(\alpha_{I-3}^2, \alpha_{I-2}^2))$$

We now want to establish a confidence interval for our target variables C_{iI} and R_i. Because of the equation

$$C_{iI} = C_{i,I+1-i} + R_i$$

the ultimate claims amount C_{iI} consists of a known part $C_{i,I+1-i}$ and an unknown part R_i. This means that the probability distribution function of C_{iI} (given the observations D which include $C_{i,I+1-i}$) is completely determined by that of R_i. We therefore need to establish a confidence interval for R_i only and can then simply shift it to a confidence interval for C_{iI}.

For this purpose we need to know the distribution function of R_i. Up to now we only have estimates $\mathbf{R_i}$ and $\text{s.e.}(\mathbf{R_i})$ for the mean and the standard deviation of this distribution. If the volume of the outstanding claims is large enough we can, due to the central limit theorem, assume that this distribution function is a Normal distribution with an expected value equal to the point estimate given by $\mathbf{R_i}$ and a standard deviation equal to the standard error $\text{s.e.}(\mathbf{R_i})$. A symmetric 95%-confidence interval for R_i is then given by

$$(\mathbf{R_i} - 2 \cdot \text{s.e.}(\mathbf{R_i}), \mathbf{R_i} + 2 \cdot \text{s.e.}(\mathbf{R_i}))$$

But the symmetric Normal distribution may not be a good approximation to the true distribution of R_i if this latter distribution is rather skewed. This will especially be the case if $\text{s.e.}(\mathbf{R_i})$ is greater than 50 % of $\mathbf{R_i}$. This can also be seen at the above Normal distribution confidence interval whose lower limit then becomes negative even if a negative reserve is not possible.

In this case it is recommended to use an approach based on the Lognormal distribution. For this purpose we approximate the unknown distribution of R_i by a Lognormal distribution with parameters μ_i and σ_i^2 such that mean values as well as variances of both distributions are equal, so that

$$\exp(\mu_i + \sigma_i^2/2) = \mathbf{R_i}$$

$$\exp(2\mu_i + \sigma_i^2)(\exp(\sigma_i^2) - 1) = (\text{s.e.}(\mathbf{R_i}))^2$$

This leads to

$$(10) \quad \sigma_i^2 = \ln(1 + (\text{s.e.}(\mathbf{R_i}))^2 / \mathbf{R_i}^2)$$

$$\mu_i = \ln(\mathbf{R_i}) - \sigma_i^2/2$$

Now, if we want to estimate the 90th percentile of R_i, for example, we proceed as follows. First we take the 90th percentile of the Standard Normal distribution which is 1.28. Then $\exp(\mu_i + 1.28\sigma_i)$ with μ_i and σ_i^2 according to (10) is the 90th percentile of the Lognormal distribution and therefore also approximately of the distribution of R_i.

For instance, if $\text{s.e.}(\mathbf{R_i})/\mathbf{R_i} = 1$ then $\sigma_i^2 = \ln(2)$ and the 90th percentile is $\exp(\mu_i + 1.28\sigma_i) = \mathbf{R_i}\exp(1.28\sigma_i - \sigma_i^2/2) = \mathbf{R_i}\exp(.719) = 2.05 \cdot \mathbf{R_i}$. If we had assumed that R_i has approximately a Normal distribution, we would have obtained in this case $\mathbf{R_i} + 1.28 \cdot \text{s.e.}(\mathbf{R_i}) = 2.28 \cdot \mathbf{R_i}$ as 90th percentile.

This may come as a surprise since we might have expected that the 90th percentile of a Lognormal distribution always must be higher than that of a Normal distribution with same mean and variance. But there is no general rule, it depends on the percentile chosen and on the size of the ratio $\text{s.e.}(\mathbf{R}_i)/\mathbf{R}_i$. The Lognormal approximation only prevents a negative lower confidence limit. In order to set a specific lower confidence limit we choose a suitable percentile, for instance 10%, and proceed analogously as with the 90% before. The question of which confidence probability to choose has to be decided from a business policy point of view. The value of 80% = 90% – 10% taken here must be regarded merely as an example.

We have now shown how to establish confidence limits for every R_i and therefore also for every $C_{iI} = C_{i,I+1-i} + R_i$. We may also be interested in having confidence limits for the overall reserve

$$R = R_2 + \ldots + R_I$$

and the question is whether, in order to estimate the variance of R, we can simply add the squares $(\text{s.e.}(\mathbf{R}_i))^2$ of the individual standard errors as would be the case with standard deviations of independent variables. But unfortunately, whereas the R_i's themselves are independent, the estimators $\mathbf{R}_i$ are not because they are all influenced by the same age-to-age factors $\mathbf{f}_k$, that is the $\mathbf{R}_i$'s are positively correlated. In Appendix F it is shown that the square of the standard error of the overall reserve estimator

$$\mathbf{R} = \mathbf{R}_2 + \ldots + \mathbf{R}_I$$

is given by

$$(11)\quad (\text{s.e.}(\mathbf{R}))^2 = \sum_{i=2}^{I}\left\{(\text{s.e.}(\mathbf{R}_i))^2 + \mathbf{C}_{iI}\left(\sum_{j=i+1}^{I}\mathbf{C}_{jI}\right)\sum_{k=I+1-i}^{I-1}\frac{2\alpha_k^2/\mathbf{f}_k^2}{\sum_{n=1}^{I-k}C_{nk}}\right\}$$

Formula (11) can be used to establish a confidence interval for the overall reserve amount R in quite the same way as it was done before for R_i. Before giving a full example of the calculation of the standard error, we will deal in the next section with the problem of how to decide for a given run-off triangle whether the chain ladder assumptions (3) and (5) are met or not.

5. Checking the Chain Ladder Assumptions Against the Data

As has been pointed out, the three basic implicit chain ladder assumptions

(3) $E(C_{i,k+1} | C_{i1},\ldots,C_{ik}) = C_{ik} f_k$

(4) Independence of accident years

(5) $\mathrm{Var}(C_{i,k+1} | C_{i1},\ldots,C_{ik}) = C_{ik} \alpha_k^2$

are not met in every case. In this section we will indicate how these assumptions can be checked for a given run-off triangle. We have already mentioned in section 3 that Appendix H develops a test for calendar year influences which may violate (4). We can therefore concentrate in the following on assumptions (3) and (5).

First, we look at the equations (3) for an arbitrary but fixed k and for i = 1, ..., I. There, the values of C_{ik}, $1 \le i \le I$, are to be considered as given non-random values and equations (3) can be interpreted as an ordinary regression model of the type

$$Y_i = c + x_i b + \varepsilon_i, \quad 1 \le i \le I$$

where c and b are the regression coefficients and ε_i the error term with $E(\varepsilon_i) = 0$, that is $E(Y_i) = c + x_i b$. In our special case, we have c = 0, b = f_k and we have observations of the dependent variable $Y_i = C_{i,k+1}$ at the points $x_i = C_{ik}$ for i = 1, ..., I – k. Therefore, we can estimate the regression coefficient b = f_k by the usual least squares method

$$\sum_{i=1}^{I-k} (C_{i,k+1} - C_{ik} f_k)^2 = \text{minimum}$$

If the derivative of the left hand side with respect to f_k is set to 0 we obtain for the minimizing parameter f_k the solution

$$(12) \quad f_{k0} = \sum_{i=1}^{I-k} C_{ik} C_{i,k+1} \Big/ \sum_{i=1}^{I-k} C_{ik}^2$$

This is not the same estimator for f_k as according to the chain ladder formula (2). We therefore have used an additional index '0' at this new estimator for f_k. We can rewrite f_{k0} as

$$f_{k0} = \sum_{i=1}^{I-k} \frac{C_{ik}^2}{\sum_{i=1}^{I-k} C_{ik}^2} \cdot \frac{C_{i,k+1}}{C_{ik}}$$

which shows that f_{k0} is the C_{ik}^2-weighted average of the individual development factors $C_{i,k+1} / C_{ik}$, whereas the chain ladder estimator $\mathbf{f_k}$ is the C_{ik}-weighted average. In section 3 we saw that these weights are inversely proportional to the underlying variances $\mathrm{Var}(C_{i,k+1}/C_{ik} \mid C_{i1},\ldots,C_{ik})$.

Correspondingly, the estimator f_{k0} assumes

$$\mathrm{Var}(C_{i,k+1}/C_{ik} \mid C_{i1},\ldots,C_{ik}) \text{ being proportional to } 1/C_{ik}^2$$

or equivalently

$$\mathrm{Var}(C_{i,k+1} \mid C_{i1},\ldots,C_{ik}) \text{ being proportional to } 1$$

which means that $\mathrm{Var}(C_{i,k+1} \mid C_{i1},\ldots,C_{ik})$ is the same for all observations $i = 1, \ldots, I-k$. This is not in agreement with the chain ladder assumption (5).

Here we remember that indeed the least squares method implicitly assumes equal variances $\mathrm{Var}(Y_i) = \mathrm{Var}(\varepsilon_i) = \sigma^2$ for all i. If this assumption is not met, that is if the variances $\mathrm{Var}(Y_i) = \mathrm{Var}(\varepsilon_i)$ depend on i, one should use a weighted least squares approach which consists of minimizing the weighted sum of squares

$$\sum_{i=1}^{I} w_i (Y_i - c - x_i b)^2$$

where the weights w_i are in inverse proportion to $\mathrm{Var}(Y_i)$.

Therefore, in order to be in agreement with the chain ladder variance assumption (5), we should use regression weights w_i which are proportional to $1/C_{ik}$ (more precisely to $1/(C_{ik}\alpha_k^2)$, but α_k^2 can be amalgamated with the proportionality constant because k is fixed).

Then minimizing

$$\sum_{i=1}^{I-k} (C_{i,k+1} - C_{ik} f_k)^2 / C_{ik}$$

with respect to f_k yields indeed

$$f_{k1} = \sum_{i=1}^{I-k} C_{i,k+1} \Big/ \sum_{i=1}^{I=k} C_{ik}$$

which is identical to the usual chain ladder age-to-age factor $\mathbf{f_k}$.

It is tempting to try another set of weights, namely $1/C_{ik}^{\ 2}$ because then the weighted sum of squares becomes

$$\sum_{i=1}^{I-k} (C_{i,k+1} - C_{ik} f_k)^2 / C_{ik}^{\ 2} = \sum_{i=1}^{I-k} \left(\frac{C_{i,k+1}}{C_{ik}} - f_k \right)^2$$

Here the minimizing procedure yields

$$(13) \quad f_{k2} = \frac{1}{I-k} \sum_{i=1}^{I-k} \frac{C_{i,k+1}}{C_{ik}}$$

which is the ordinary unweighted average of the development factors. The variance assumption corresponding to the weights used is

$$\mathrm{Var}(C_{i,k+1} \,|\, C_{i1}, \ldots, C_{ik}) \text{ being proportional to } C_{ik}^{\ 2}$$

or equivalently

$$\mathrm{Var}(C_{i,k+1} / C_{ik} \,|\, C_{i1}, \ldots, C_{ik}) \text{ being proportional to } 1$$

The benefit of transforming the estimation of the age-to-age factors into the regression framework is the fact that the usual regression analysis instruments are now available to check the underlying assumptions, especially the linearity and the variance assumption. This check is usually done by carefully inspecting plots of the data and of the residuals, as described below.

First, we plot $C_{i,k+1}$ against C_{ik}, $i = 1, \ldots, I-k$, in order to see if we really have an approximately linear relationship around a straight line through the origin with slope $\mathbf{f_k} = f_{k1}$. Second, if linearity seems acceptable, we plot the weighted residuals

$$(C_{i,k+1} - C_{ik}\mathbf{f_k}) / \sqrt{C_{ik}}, \quad 1 \le i \le I-k,$$

(whose squares have been minimized) against C_{ik} in order to see if the employed variance assumption really leads to a plot in which the residuals do not show any specific trend but appear purely random. It is recommended to compare all three residual plots (for $i = 1, \ldots, I-k$)

Plot 0: $C_{i,k+1} - C_{ik} f_{k0}$ against C_{ik}

Plot 1: $(C_{i,k+1} - C_{ik} f_{k1}) / \sqrt{C_{ik}}$ against C_{ik}

Plot 2: $(C_{i,k+1} - C_{ik} f_{k2}) / C_{ik}$ against C_{ik}

and to find out which one shows the most random behaviour. All this should be done for every development year k for which we have sufficient data points, say at least 6, that is for $k \leq I - 6$.

Some experience with least squares residual plots is useful, especially because in our case we have only very few data points. Consequently, it is not always easy to decide whether a pattern in the residuals is systematic or random. However, if Plot 1 exhibits a non-random pattern, and either Plot 0 or Plot 2 does not, and if this holds true for several values of k, we should seriously consider replacing the chain ladder age-to-age factors $f_{k1} = \mathbf{f}_k$ with f_{k0} or f_{k2} respectively.

The following numerical example will clarify the situation a bit more.

6. Numerical Example

The data for the following example are taken from the 'Historical Loss Development Study', 1991 Edition, published by the Reinsurance Association of America (RAA). There, we find on page 96 the following run-off triangle of Automatic Facultative business in General Liability (excluding Asbestos & Environmental):

	C_{i1}	C_{i2}	C_{i3}	C_{i4}	C_{i5}	C_{i6}	C_{i7}	C_{i8}	C_{i9}	C_{i10}
i=1	5012	8269	10907	11805	13539	16181	18009	18608	18662	18834
i=2	106	4285	5396	10666	13782	15599	15496	16169	16704	
i=3	3410	8992	13873	16141	18735	22214	22863	23466		
i=4	5655	11555	15766	21266	23425	26083	27067			
i=5	1092	9565	15836	22169	25955	26180				
i=6	1513	6445	11702	12935	15852					
i=7	557	4020	10946	12314						
i=8	1351	6947	13112							
i=9	3133	5395								
i=10	2063									

The above figures are cumulative incurred case losses in $1000. We have taken the accident years from 1981 (i=1) to 1990 (i=10) which is enough for the sake of example but does not mean that we believe to have reached the ultimate claims amount after 10 years of development.

We first calculate the age-to-age factors $\mathbf{f}_k = f_{k1}$ according to formula (2). The result is shown in the following table together with the alternative factors f_{k0} according to (12) and f_{k2} according to (13)

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
f_{k0}	2.217	1.569	1.261	1.162	1.100	1.041	1.032	1.016	1.009
f_{k1}	2.999	1.624	1.271	1.172	1.113	1.042	1.033	1.017	1.009
f_{k2}	8.206	1.696	1.315	1.183	1.127	1.043	1.034	1.018	1.009

If one has the run-off triangle on a personal computer it is very easy to produce the plots recommended in section 5 because most spreadsheet programs have the facility of plotting X–Y graphs. For every $k = 1, ..., 8$ we make a plot of the amounts $C_{i,k+1}$ (y-axis) of development year k+1 against the amounts C_{ik} (x-axis) of development year k for $i = 1, ..., 10 - k$, and draw a straight line through the origin with slope f_{k1}.

The plots for k = 1 to 8 are shown in the upper graphs of Figures 1 to 8, respectively. (All figures are to be found at the end of the paper after the appendices.) The number above each point mark indicates the corresponding accident year. (Note that the point mark at the upper or right hand border line of each graph does not belong to the plotted points $(C_{ik}, C_{i,k+1})$, it has only been used to draw the regression line.) In the lower graph of each of the Figures 1 to 8 the corresponding weighted residuals $(C_{i,k+1} - C_{ik}) / \sqrt{C_{ik}}$ are plotted against C_{ik} for $i = 1, \ldots, 10 - k$.

The two plots for k = 1 (Figure 1) clearly show that the regression line does not capture the direction of the data points very well. The line should preferably have a positive intercept on the y-axis and a flatter slope. However, even then we would have a high dispersion. Using the line through the origin we will probably underestimate any future C_{i2} if C_{i1} is less than 2000 and will overestimate it if C_{i1} is more than 4000. Fortunately, in the one relevant case i = 10 we have $C_{i1} = 2063$ which means that the resulting forecast $C_{10,2} = C_{10,1}\mathbf{f}_2 = 2063 \cdot 2.999 = 6187$ is within the bulk of the data points plotted. In any case, Figure 1 shows that any forecast of $C_{10,2}$ is associated with a high uncertainty of about ±3000 or almost ±50% of an average-sized C_{i2} which is subsequently even larger when extrapolating to ultimate. If in a future accident year we have a value C_{i1} outside the interval (2000, 4000) it is reasonable to introduce an additional parameter by fitting a regression line with positive intercept to the data and using it for the projection to C_{i2}. Such a procedure of employing an additional parameter is acceptable between the first two development years in which we have the highest number of data points of all years.

The two plots for k = 2 (Figure 2) are more satisfactory. The data show a clear trend along the regression line and quite random residuals. The same holds for the two plots for k = 4 (Figure 4). In addition, for both k = 2 and k = 4 a weighted linear regression including a parameter for intercept would yield a value of the intercept which is not significantly different from zero. The plots for k = 3 (Figure 3) seem to show a curvature to the left but because of the few data points we can hope that this is incidental. Moreover, the plots for k = 5 have a certain curvature to the right such that we can hope that the two curvatures offset each other. The plots for k = 6, 7 and 8 are quite satisfactory. The trends in the residuals for k = 7 and 8 have no significance in view of the very few data points.

We need not look at the regression lines with slopes f_{k0} or f_{k2} as these slopes are very close to $\mathbf{f}_k$ (except for k=1). But we should look at the corresponding plots of weighted residuals in order to see whether they appear more satisfactory than the previous ones. (Note that due to the different weights the residuals will be different even if the slopes are equal.) The residual plots for f_{k0} and k = 1 to 4 are shown in Figures 9 and 10. Those for f_{k2} and k = 1 to 4 are shown in Figures 11 and 12. In the residual plot for $f_{1,0}$ (Figure 9, upper graph) the point furthest to the left is not an outlier as it is in the plots for $f_{1,1} = \mathbf{f}_1$ (Figure 1, lower graph) and $f_{1,2}$ (Figure 11, upper graph).

But with all three residual plots for k=1 the main problem is the missing intercept of the regression line which leads to a decreasing trend in the residuals. Therefore the improvement of the outlier is of secondary importance. For k = 2 the three residuals plots do not show any major differences between each other. The same holds for k = 3 and 4. The residual plots for k = 5 to 8 are not important because of the small number of data points. Altogether, we decide to keep the usual chain ladder method, that is the age-to-age factors $\mathbf{f_k} = f_{k,1}$, because the alternatives $f_{k,0}$ or $f_{k,2}$ do not lead to a clear improvement.

Next, we can carry through the tests for calendar year influences (see Appendix H) and for correlations between subsequent development factors (see Appendix G). For our example neither test leads to a rejection of the underlying assumption as is shown in the appendices mentioned.

Having now finished all preliminary analyses we calculate the estimated ultimate claims amounts $\mathbf{C_{iI}}$ according to formula (1), the reserves $\mathbf{R_i} = \mathbf{C_{iI}} - C_{i,I+1-i}$ and its standard errors (7). For the standard errors we need the estimated values of $\boldsymbol{\alpha}_k^2$ which according to formula (8) are given by

k	1	2	3	4	5	6	7	8	9
$\boldsymbol{\alpha}_k^2$	27883	1109	691	61.2	119	40.8	1.34	7.88	

A plot of $\ln(\boldsymbol{\alpha}_k^2)$ against k is given in Figure 13 and shows that there indeed seems to be a linear relationship which can be used to extrapolate $\ln(\boldsymbol{\alpha}_9^2)$. This yields $\boldsymbol{\alpha}_9^2 = \exp(-.44) = .64$. But we use formula (9) which is more easily programmable and in the present case is a bit more on the safe side: it leads to $\boldsymbol{\alpha}_9^2 = 1.34$. Using formula (11) for s.e.(**R**) as well we finally obtain

	$\mathbf{C_{i,10}}$	$\mathbf{R_i}$	$\text{s.e.}(\mathbf{C_{i,10}}) = \text{s.e.}(\mathbf{R_i})$	$\text{s.e.}(\mathbf{R_i})/\mathbf{R_i}$
i=2	16858	154	206	134%
i=3	24083	617	623	101%
i=4	28703	1636	747	46%
i=5	28927	2747	1469	53%
i=6	19501	3649	2002	55%
i=7	17749	5435	2209	41%
i=8	24019	10907	5358	49%
i=9	16045	10650	6333	59%
i=10	18402	16339	24566	150%
Overall		52135	26909	52%

(The numbers in the 'Overall'-row are **R**, s.e.(**R**) and s.e.(**R**)/**R**.) For i = 2, 3 and 10 the percentage standard error (last column) is more than 100% of the estimated reserve $\mathbf{R_i}$. For i = 2 and 3 this is due to the small amount of the corresponding reserve and is not important because the absolute amounts of the standard errors are rather small. But the standard error of 150% for the most recent accident year i = 10 might lead to some concern in practice. The main reason for this high standard error is the high uncertainty of forecasting next year's value $C_{10,2}$ as was seen when examining the plot of C_{i2} against C_{i1}. Thus, one year later we will very likely be able to give a much more precise forecast of $C_{10,10}$.

Because all standard errors are close to or above 50% we use the Lognormal distribution in all years for the calculation of confidence intervals. We first calculate the upper 90%-confidence limit (or with any other chosen percentage) for the overall outstanding claims reserve R. Denoting by μ and σ^2 the parameters of the Lognormal distribution approximating the distribution of R and using s.e.(**R**)/**R** = .52 we have σ^2 = .236 (cf. (10)) and, in the same way as in section 4, the 90th percentile is $\exp(\mu + 1.28\sigma) = \mathbf{R}\cdot\exp(1.28\sigma - \sigma^2/2) = 1.655\cdot\mathbf{R} = 86298$.

Now we allocate this overall amount to the accident years i = 2,..., 10 in such a way that we reach the same level of confidence for every accident year. Each level of confidence corresponds to a certain percentile t of the Standard Normal distribution and — according to section 4 — the corresponding percentile of the distribution of R_i is $\mathbf{R}_i \exp(t\sigma_i - \sigma_i^2/2)$ with $\sigma_i^2 = \ln(1 + (\text{s.e.}(\mathbf{R}_i))^2/\mathbf{R}_i^2)$. We therefore only have to choose t in such a way that

$$\sum_{i=2}^{I} \mathbf{R}_i.\exp(t\sigma_i - \sigma_i^2/2) = 86298$$

This can easily be solved with the help of spreadsheet software (for example. by trial and error, or by using a "Solver") and yields t = 1.13208 which corresponds to the 87th percentile per accident year and leads to the following distribution of the overall amount 86298:

	$\mathbf{R}_i$	$\text{s.e.}(\mathbf{R}_i)/\mathbf{R}_i$	σ_i^2	upper confidence limit $\mathbf{R}_i \exp(t\sigma_i - \sigma_i^2/2)$
i=2	154	1.34	1.028	290
i=3	617	1.01	.703	1122
i=4	1636	.46	.189	2436
i=5	2747	.53	.252	4274
i=6	3649	.55	.263	5718
i=7	5435	.41	.153	7839
i=8	10907	.49	.216	16571
i=9	10650	.59	.303	17066
i=10	16339	1.50	1.182	30981
Total	52135			86298

In order to arrive at the lower confidence limits we proceed completely analogously. The 10th percentile, for instance, of the total outstanding claims amount is $\mathbf{R}\cdot\exp(-1.28\sigma - \sigma^2/2) = .477\cdot\mathbf{R} = 24871$. The distribution of this amount over the individual accident years is made as before and leads to a value of t = –.8211 which corresponds to the 21st percentile. This means that a 87% – 21% = 66% confidence interval for each accident year leads to a 90% – 10% = 80% confidence interval for the overall reserve amount. In the following table, the confidence intervals thus obtained for R_i are already shifted (by adding $C_{i,I+1-i}$) to confidence intervals for the ultimate claims amounts C_{iI} (for instance, the upper limit 16994 for i=2 has been obtained by adding $C_{2,9}$ = 16704 and 290 from the preceding table):

	$C_{i,10}$	confidence intervals for 80% prob. overall	empirical limits
i=2	16858	(16744, 16994)	(16858, 16858)
i=3	24083	(23684, 24588)	(23751, 24466)
i=4	28703	(28108, 29503)	(28118, 29446)
i=5	28927	(27784, 30454)	(27017, 31699)
i=6	19501	(17952, 21570)	(16501, 22939)
i=7	17749	(15966, 20153)	(14119, 23025)
i=8	24019	(19795, 29683)	(16272, 48462)
i=9	16045	(11221, 22461)	(8431, 54294)
i=10	18402	(5769, 33044)	(5319, 839271)

The column "empirical limits" contains the minimum and maximum size of the ultimate claims amount resulting if, in formula (1), each age-to-age factor $\mathbf{f_k}$ is replaced with the minimum (or maximum) individual development factor observed so far. These factors are defined by

$$f_{k,min} = \min\{C_{i,k+1}/C_{ik} \mid 1 \le i \le I-k\}$$
$$f_{k,max} = \max\{C_{i,k+1}/C_{i,k} \mid 1 \le i \le I-k\}$$

and can be taken from the table of all development factors which can be found in Appendices G and H. They are

	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9
$f_{k,min}$	1.650	1.259	1.082	1.102	1.009	0.993	1.026	1.003	1.009
$f_{k,max}$	40.425	2.723	1.977	1.292	1.195	1.113	1.043	1.033	1.009

In comparison with the confidence intervals, these empirical limits are narrower in the earlier accident years i ≤ 4 and wider in the more recent accident years i ≥ 5. This was to be expected because the small number of development factors observed between the late development years only leads to a rather small variation between the minimum and maximum factors. Therefore these empirical limits correspond to a confidence probability which is rather small in the early accident years and becomes larger and larger towards the recent accident years. Thus, this empirical approach to establishing confidence limits does not seem to be reasonable.

If we used the Normal distribution instead of the Lognormal we would obtain a 90th percentile of $\mathbf{R} + 1.28\cdot\mathbf{R}\cdot(\text{s.e.}(\mathbf{R})/\mathbf{R}) = 1.661\cdot\mathbf{R}$ (which is almost the same as the $1.655\cdot\mathbf{R}$ with the Lognormal) and a 10th percentile of $\mathbf{R} - 1.28\cdot\mathbf{R}\cdot(\text{s.e.}(\mathbf{R})/\mathbf{R}) = .34\cdot\mathbf{R}$ (which is lower than the $.477\cdot\mathbf{R}$ with the Lognormal). Also, the allocation to the accident years would be different.

Finally, we compare the standard errors obtained to the output of the claims reserving software package ICRFS by Ben Zehnwirth.

This package is a modelling framework in which the user can specify his own model within a large class of models. But it also contains some predefined models, inter alia also a 'chain ladder model'. But this is not the usual chain ladder method, instead, it is a log-linearized approximation of it. This is very similar to the model described in the paper, Regression Model Based on Log-Incremental Payments by S.Christofides, see Section D5, Volume 2 of the Claims Reserving Manual.

The slight difference in the results is due to a different estimator for the variance, σ^2. Therefore, the estimates of the outstanding claims amounts differ from those obtained here with the usual chain ladder method. Moreover, it works with the logarithms of the incremental amounts $C_{i,k+1} - C_{ik}$ and one must therefore eliminate the negative increment $C_{2,7} - C_{2,6}$. In addition, $C_{2,1}$ was identified as an outlier and was eliminated. Then the ICRFS results were quite similar to the chain ladder results as can be seen in the following table

	est. outst. claims amount $\mathbf{R}_i$		standard error	
	chain ladder	ICRFS	chain ladder	ICRFS
i=2	154	387	206	528
i=3	617	674	623	624
i=4	1636	1993	747	1435
i=5	2747	2602	1469	1688
i=6	3649	4097	2002	2476
i=7	5435	5188	2209	3156
i=8	10907	12174	5358	7685
i=9	10650	15343	6333	11158
i=10	16339	27575	24566	28333
Overall	52135	70032	26909	33637

Even though the reserves $\mathbf{R}_i$ for i=9 and i=10 as well as the overall reserve $\mathbf{R}$ differ considerably they are all within one standard error and therefore not significantly different. But it should be remarked that this manner of using ICRFS is not intended by Zehnwirth because any initial model should be further adjusted according to the indications and plots given by the program. In this particular case there were strong indications for developing the model further but then one would have to give up the 'chain ladder model'.

7. Final Remarks

This paper develops a complete methodology of how to attack the claims reserving task in a statistically sound manner on the basis of the well-known and simple chain ladder method. However, the well-known weak points of the chain ladder method should not be concealed. These are the fact that the estimators of the last two or three factors f_I, f_{I-1}, f_{I-2} rely on very few observations and the fact that the known claims amount C_{I1} of the last accident year (sometimes $C_{I-1,2}$, too) forms a very uncertain basis for the projection to ultimate.

This is most clearly seen if C_{I1} happens to be 0: Then we have $\mathbf{C}_{iI} = 0$, $\mathbf{R}_I = 0$ and $\text{s.e.}(\mathbf{R}_I) = 0$ which obviously makes no sense. (Note that this weakness can often be overcome by translating and mixing the amounts C_{i1} of earlier accident years $i < I$ into accident year I with the help of a measure of volume for each accident year.)

Thus, even if the statistical instruments developed do not reject the applicability of the chain ladder method, the result must be judged by an actuary and/or underwriter who knows the business under consideration. Even then, unexpected future changes can make all estimations obsolete. But for the many normal cases it is good to have a sound and simple method. Simple methods have the disadvantage of not capturing all aspects of reality but have the advantage that the user is in a position to know exactly how the method works and where its weaknesses are. Moreover, a simple method can be explained to non-actuaries in more detail. These are important advantages of simple models over more sophisticated ones.

Appendix A: Unbiasedness of Age-to-Age Factors

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I, 1 \le k \le I-1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \neq j$ are independent

the age-to-age factors $\mathbf{f}_1, \ldots, \mathbf{f}_{I-1}$ defined by

$$(2) \quad \mathbf{f}_k = \sum_{j=1}^{I-k} C_{j,k+1} / \sum_{j=1}^{I-k} C_{jk}, \qquad 1 \le k \le I-1$$

are unbiased, that is we have $E(\mathbf{f}_k) = f_k, \quad 1 \le k \le I-1$

Proof: Because of the iterative rule for expectations we have

$$(A1) \quad E(\mathbf{f}_k) = E(E(\mathbf{f}_k \mid B_k))$$

for any set B_k of variables C_{ij} assumed to be known. We take

$$B_k = \{C_{ij} \mid i + j \le I + 1, j \le k\}, \qquad 1 \le k \le I\text{-}1$$

According to the definition (2) of $\mathbf{f}_k$ and because C_{jk}, $1 \le j \le I-k$, is contained in B_k and therefore has to be treated as scalar, we have

$$(A2) \quad E(\mathbf{f}_k \mid B_k) = \sum_{j=1}^{I-k} E(C_{j,k+1} \mid B_k) / \sum_{j=1}^{I-k} C_{jk}$$

Because of the independence assumption (4) conditions relating to accident years other than that of $C_{j,k+1}$ can be omitted, that is we get

$$(A3) \quad E(C_{j,k+1} \mid B_k) = E(C_{j,k+1} \mid C_{j1}, \ldots, C_{jk}) = C_{jk} f_k$$

using assumption (3) as well.

Inserting (A3) into (A2) yields

$$\text{(A4)}\quad E(\mathbf{f}_k \mid B_k) = \sum_{j=1}^{I-k} C_{jk} f_k \Big/ \sum_{j=1}^{I-k} C_{jk} = f_k$$

Finally, (A1) and (A4) yield $E(\mathbf{f}_k) = E(f_k) = f_k$ because f_k is a scalar.

Appendix B: Minimizing the Variance of Independent Estimators

Proposition: Let $T_1, \ldots, T_I$ be independent unbiased estimators of a parameter t, that is with

$$E(T_i) = t, \qquad 1 \le i \le I$$

then the variance of a linear combination

$$T = \sum_{i=1}^{I} w_i T_i$$

under the constraint

$$\text{(B1)} \quad \sum_{i=1}^{I} w_i = 1$$

(which guarantees $E(T) = t$) is minimal iff the coefficients w_i are inversely proportional to $Var(T_i)$, that is iff

$$w_i = c / Var(T_i), \qquad 1 \le i \le I$$

Proof: We have to minimize

$$Var(T) = \sum_{i=1}^{I} w_i^2 Var(T_i)$$

(due to the independence of $T_1, \ldots, T_I$) with respect to w_i under the constraint (B1).

A necessary condition for an extremum is that the derivatives of the Lagrangian are zero, that is

$$\text{(B2)} \quad \frac{\partial}{\partial w_i}\left(\sum_{i=1}^{I} w_i^2 \ Var(T_i) + \lambda\left(1 - \sum_{i=1}^{I} w_i\right)\right) = 0, \qquad 1 \le i \le I$$

with a constant multiplier λ whose value can be determined by additionally using (B1).

(B2) yields

$$2w_i \mathrm{Var}(T_i) - \lambda = 0$$

or

$$w_i = \lambda / (2 \cdot \mathrm{Var}(T_i))$$

These weights w_i indeed lead to a minimum as can be seen by calculating the extremal value of Var(T) and applying Schwarz's inequality.

Corollary: In the chain ladder case we have estimators $T_i = C_{i,k+1} / C_{ik}$, $1 \le i \le I - k$, for f_k where the variables of the set

$$A_k = \bigcup_{i=1}^{I-k} \{C_{i1}, \ldots, C_{ik}\}$$

of the corresponding accident years i = 1, ..., I – k up to development year k are considered to be given. We therefore want to minimize the conditional variance

$$\mathrm{Var}\left(\sum_{i=1}^{I-k} w_i T_i \mid A_k \right)$$

From the above proof it is clear that the minimizing weights should be inversely proportional to $\mathrm{Var}(T_i \mid A_k)$. Because of the independence (4) of the accident years, conditions relating to accident years other than that of $T_i = C_{i,k+1} / C_{ik}$ can be omitted. We therefore have

$$\mathrm{Var}(T_i \mid A_k) = \mathrm{Var}(C_{i,k+1} / C_{ik} \mid C_{i1}, \ldots, C_{ik})$$

and arrive at the result that the minimizing weights should be inversely proportional to

$$\mathrm{Var}(C_{i,k+1} / C_{ik} \mid C_{i1}, \ldots, C_{ik})$$

Appendix C: Unbiasedness of the Estimated Ultimate Claims Amount

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I,\ 1 \le k \le I\text{-}1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \neq j$ are independent

the expected values of the estimator

(1) $\mathbf{C}_{iI} = C_{i,I+1-i}\mathbf{f}_{i+1-i}\cdot\ldots\cdot\mathbf{f}_{I-1}$

for the ultimate claims amount and of the true ultimate claims amount C_{iI} are equal, that is we have $E(\mathbf{C}_{iI}) = E(C_{iI})$, $2 \le i \le I$.

Proof: We first show that the age-to-age factors $\mathbf{f}_k$ are uncorrelated. With the same set

$$B_k = \{ C_{ij} \mid i + j \le I + 1, j \le k \}, \qquad 1 \le k \le I - 1$$

of variables assumed to be known as in Appendix A we have for $j < k$

$$\begin{aligned}
E(\mathbf{f}_j\mathbf{f}_k) &= E(E(\mathbf{f}_j\mathbf{f}_k \mid B_k)) && \text{(a)}\\
&= E(\mathbf{f}_j E(\mathbf{f}_k \mid B_k)) && \text{(b)}\\
&= E(\mathbf{f}_j f_k) && \text{(c)}\\
&= E(\mathbf{f}_j) f_k && \text{(d)}\\
&= f_j f_k && \text{(e)}
\end{aligned}$$

Here (a) holds because of the iterative rule for expectations, (b) holds because $\mathbf{f}_j$ is a scalar for B_k given and for $j < k$, (c) holds due to (A4), (d) holds because f_k is a scalar and (e) was shown in Appendix A.

This result can easily be extended to arbitrary products of different $\mathbf{f}_k$'s, that is we have

(C1) $E(\mathbf{f}_{I+1-i}\cdot\ldots\cdot\mathbf{f}_{I-1}) = f_{i+1-i}\cdot\ldots\cdot f_{I-1}$

This yields

$$
\begin{aligned}
E(\mathbf{C}_{iI}) &= E(E(\mathbf{C}_{iI} \mid C_{i1}, \dots, C_{i,I+1-i})) && \text{(a)} \\
&= E(E(C_{i,I+1-i}\mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1} \mid C_{i1}, \dots, C_{i,I+1-i})) && \text{(b)} \\
&= E(C_{i,I+1-i}E(\mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1} \mid C_{i1}, \dots, C_{i,I+1-i})) && \text{(c)} \\
&= E(C_{i,I+1-i}E(\mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1})) && \text{(d)} \\
&= E(C_{i,I+1-i}) \cdot E(\mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1}) && \text{(e)} \\
&= E(C_{i,I+1-i}) \cdot f_{I+1-i} \cdot \ldots \cdot f_{I-1} && \text{(f)}
\end{aligned}
$$

Here (a) holds because of the iterative rule for expectations, (b) holds because of the definition (1) of $\mathbf{C}_{iI}$, (c) holds because $C_{i,I+1-i}$ is a scalar under the stated condition, (d) holds because conditions which are independent from the conditioned variable $\mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1}$ can be omitted (observe assumption (4) and the fact that $\mathbf{f}_{I+1-i}, \dots, \mathbf{f}_{I-1}$ only depend on variables of accident years $\leq$ i), (e) holds because $E(\mathbf{f}_{I+1-i}, \dots, \mathbf{f}_{I-1})$ is a scalar and (f) holds because of (C1).

Finally, repeated application of the iterative rule for expectations and of assumption (3) yields for the expected value of the true reserve C_{iI}

$$
\begin{aligned}
E(C_{iI}) &= E(E(C_{iI} \mid C_{i1}, \dots, C_{i,I-1})) \\
&= E(C_{i,I-1} f_{I-1}) \\
&= E(C_{i,I-1}) f_{I-1} \\
&= E(E(C_{i,I-1} \mid C_{i1}, \dots, C_{I-2})) f_{I-1} \\
&= E(C_{i,I-2} f_{I-2}) f_{I-1} \\
&= E(C_{i,I-2}) f_{I-2} f_{I-1} \\
&= \text{and so on} \\
&= E(C_{i,I+1-i}) f_{I+1-i} \cdot \ldots \cdot f_{I-1} \\
&= E(\mathbf{C}_{iI})
\end{aligned}
$$

Appendix D: Calculation of the Standard Error of $\mathbf{C}_{iI}$

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \ne j$ are independent

(5) There are unknown constants $\alpha_1, \ldots, \alpha_{I-1}$ with

$$\mathrm{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} \alpha_k^2, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

the standard error s.e.$(\mathbf{C}_{iI})$ of the estimated ultimate claims amount $\mathbf{C}_{iI} = C_{i,I+1-i} \mathbf{f}_{I+1-i} \cdot \ldots \cdot \mathbf{f}_{I-1}$ is given by the formula

$$(\mathrm{s.e.}(\mathbf{C}_{iI}))^2 = \mathbf{C}_{iI}^{\,2} \sum_{k=I+1-i}^{I-1} \left(\frac{\boldsymbol{\alpha}_k^{\,2}}{\mathbf{f}_k^{\,2}} \left(\frac{1}{\mathbf{C}_{ik}} + \frac{1}{\sum_{j=1}^{I-k} C_{jk}} \right) \right)$$

where $\mathbf{C}_{ik} = C_{i,I+1-i} \mathbf{f}_{I+1-i} \ldots \mathbf{f}_{k-1}$, $k > I+1-i$, are the estimated values of the future $C_{i,k}$ and $\mathbf{C}_{i,I+1-i} = C_{i,I+1-i}$.

Proof: As stated in section 4, the standard error is the square root of an estimator of $\mathrm{mse}(\mathbf{C}_{iI})$ and we have also seen that

(D1) $\mathrm{mse}(\mathbf{C}_{iI}) = \mathrm{Var}(C_{iI} \mid D) + (E(C_{iI} \mid D) - \mathbf{C}_{iI})^2$

In the following, we use the abbreviations

$$E_i(X) = E(X \mid C_{i1}, \ldots, C_{i,I+1-i})$$
$$\mathrm{Var}_i(X) = \mathrm{Var}(X \mid C_{i1}, \ldots, C_{i,I+1-i})$$

Because of the independence of the accident years we can omit in (D1) that part of the condition $D = \{C_{ik} \mid i + k \le I + 1\}$ which is independent from C_{iI}, that is we can write

(D2) $\mathrm{mse}(\mathbf{C}_{iI}) = \mathrm{Var}_i(C_{iI}) + (E_i(C_{iI}) - \mathbf{C}_{iI})^2$

We first consider $\mathrm{Var}_i(C_{iI})$. Because of the general rule $\mathrm{Var}(X) = E(X^2) - (E(X))^2$ we have

(D3) $\mathrm{Var}_i(C_{iI}) = E_i(C_{iI}{}^2) - (E_i(C_{iI}))^2$

For the calculation of $E_i(C_{iI})$ we use the fact that for $k \geq I+1-i$

(D4)
$$\begin{aligned} E_i(C_{i,k+1}) &= E_i(E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik})) \\ &= E_i(C_{ik} f_k) \\ &= E_i(C_{ik}) f_k \end{aligned}$$

Here, we have used the iterative rule for expectations in its general form $E(X|Z) = (E(X|Y)|Z)$ for $\{Y\} \supset \{Z\}$ (mostly $\{Z\}$ is the empty set). By successively applying (D4) we obtain for $k \geq I+1-i$

(D5)
$$\begin{aligned} E_i(C_{i,k+1}) &= E_i(C_{i,I+1-i}) f_{I+1-i} \cdot \ldots \cdot f_k \\ &= C_{i,I+1-i} f_{I+1-i} \cdot \ldots \cdot f_k \end{aligned}$$

because $C_{i,I+1-i}$ is a scalar under the condition 'i'.

For the calculation of the first term $E_i(C_{iI}{}^2)$ of (D3) we use the fact that for $k \geq I+1-i$

(D6)
$$\begin{aligned} E_i(C_{i,k+1}{}^2) &= E_i(E(C_{i,k+1}{}^2) \mid C_{i1}, \ldots, C_{ik}) && \text{(a)} \\ &= E_i(\mathrm{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) + (E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}))^2) && \text{(b)} \\ &= E_i(C_{ik}\alpha_k{}^2 + (C_{ik} f_k)^2) && \text{(c)} \\ &= E_i(C_{ik})\alpha_k{}^2 + E_i(C_{ik}{}^2) f_k{}^2 \end{aligned}$$

Here, (a) holds due to the iterative rule for expectations, (b) due to the rule $E(X^2) = \mathrm{Var}(X) + (E(X))^2$ and (c) holds due to (3) and (5). Now, we apply (D6) and (D5) successively to get

(D7)
$$\begin{aligned} E_i(C_{iI}{}^2) &= E_i(C_{i,I-1})\alpha_{I-1}{}^2 + E_i(C_{i,I-1}{}^2) f_{I-1}{}^2 && \text{(D6)} \\ &= C_{i,I+1-i} f_{I+1-i} \cdots f_{I-2} \alpha_{I-1}{}^2 + && \text{(D5)} \\ &\quad + E_i(C_{i,I-2})\alpha_{I-2}{}^2 f_{I-1}{}^2 + && \text{(D6)} \\ &\quad + E_i(C_{i,I-2}{}^2) f_{I-2}{}^2 f_{I-1}{}^2 \end{aligned}$$

$$= C_{i,I+1-i} f_{I+1-i} \cdots f_{I-2} \alpha_{I-1}^2 +$$
$$+ C_{i,I+1-i} f_{I+1-i} \cdots f_{I-3} \alpha_{I-2}^2 f_{I-1}^2 + \qquad \text{(D5)}$$
$$+ E_i(C_{i,I-3}) \alpha_{I-3}^2 f_{I-2}^2 f_{I-1}^2 + \qquad \text{(D6)}$$
$$+ E_i(C_{i,I-3}^2) f_{I-3}^2 f_{I-2}^2 f_{I-1}^2$$

= and so on

$$= C_{i,I+1-i} \sum_{k=I+1-i}^{I-1} f_{I+1-i} \cdots f_{k-1} \alpha_k^2 f_{k+1}^2 \cdots f_{I-1}^2$$

$$+ C_{i,I+1-i}^2 f_{I+1-i}^2 \cdot \ldots \cdot f_{I-1}^2$$

where in the last step we have used $E_i(C_{i,I+1-i}) = C_{i,I+1-i}$ and $E_i(C_{i,I+1-i}^2) = C_{i,I+1-i}^2$ because under the condition 'i' $C_{i,I+1-i}$ is a scalar. Due to (D5) we have

(D8) $(E_i(C_{iI}))^2 = C_{i,I+1-i}^2 f_{I+1-i}^2 \cdot \ldots \cdot f_{I-1}^2$

Inserting (D7) and (D8) into (D3) yields

(D9) $\mathrm{Var}_i(C_{iI}) = C_{i,I+1-i} \sum_{k=I+1-i}^{I-1} f_{I+1-i} \cdots f_{k-1} \alpha_k^2 f_{k+1}^2 \cdots f_{I-1}^2$

We estimate this first summand of $\mathrm{mse}(\mathbf{C_{iI}})$ by replacing the unknown parameters f_k, α_k^2 with their unbiased estimators $\mathbf{f_k}$ and $\boldsymbol{\alpha}_\mathbf{k}^2$, that is by

(D10) $C_{i,I+1-i} \sum_{k=I+1-i}^{I-1} \mathbf{f_{I+1-i}} \cdots \mathbf{f_{k-1}} \cdot \boldsymbol{\alpha}_\mathbf{k}^2 \cdot \mathbf{f_{k+1}}^2 \cdots \mathbf{f_{I-1}}^2$

$$= C_{i,I+1-i}^2 \mathbf{f_{I+1-i}}^2 \cdots \mathbf{f_{I-1}}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_\mathbf{k}^2 / \mathbf{f_k}^2}{C_{i,I+1-i} \mathbf{f_{I+1-i}} \cdots \mathbf{f_{k-1}}}$$

$$= \mathbf{C_{iI}}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha}_\mathbf{k}^2 / \mathbf{f_k}^2}{\mathbf{C_{ik}}}$$

where we have used the notation $\mathbf{C_{ik}}$ introduced in the proposition for the estimated amounts of the future $C_{i,k}$, $k > I + 1 - i$, including $\mathbf{C_{i,I+1-i}} = C_{i,I+1-i}$.

We now turn to the second summand of the expression (D2) for $\mathrm{mse}(\mathbf{C}_{iI})$. Because of (D5) we have

$$E_i(C_{iI}) = C_{i,I+1-i}f_{I+1-i}\cdot\ldots\cdot f_{I-1}$$

and therefore

$$\text{(D11)}\quad (E_i(C_{iI}) - \mathbf{C}_{iI})^2 = C_{i,I+1-i}{}^2(f_{I+1-i}\cdot\ldots\cdot f_{I-1} - \mathbf{f}_{I+1-i}\cdot\ldots\cdot\mathbf{f}_{I-1})^2$$

This expression cannot simply be estimated by replacing f_k with $\mathbf{f}_k$ because this would yield 0 which is not a good estimator because $\mathbf{f}_{I+1-i}\cdot\ldots\cdot\mathbf{f}_{I-1}$ generally will be different from $f_{I+1-i}\cdot\ldots\cdot f_{I-1}$ and therefore the squared difference will be positive. We therefore must take a different approach. We use the algebraic identity

$$\begin{aligned} F &= f_{I+1-i}\cdot\ldots\cdot f_{I-1} - \mathbf{f}_{I+1-i}\cdot\ldots\cdot\mathbf{f}_{I-1} \\ &= S_{I+1-i} + \ldots + S_{I-1} \end{aligned}$$

with

$$\begin{aligned} S_k &= \mathbf{f}_{I+1-i}\cdot\ldots\cdot\mathbf{f}_{k-1}f_k f_{k+1}\cdot\ldots\cdot f_{I-1} - \\ &\qquad - \mathbf{f}_{I+1-i}\cdot\ldots\cdot\mathbf{f}_{k-1}\mathbf{f}_k f_{k+1}\cdot\ldots\cdot f_{I-1} \\ &= \mathbf{f}_{I+1-i}\cdot\ldots\cdot\mathbf{f}_{k-1}(f_k - \mathbf{f}_k)f_{k+1}\cdot\ldots\cdot f_{I-1} \end{aligned}$$

This yields

$$\begin{aligned} F^2 &= (S_{I+1-i} + \ldots + S_{I-1})^2 \\ &= \sum_{k=I+1-i}^{I-1} S_k{}^2 + 2\sum_{j<k} S_j S_k \end{aligned}$$

where in the last summation j and k run from $I + 1 - i$ to $I - 1$. Now we replace $S_k{}^2$ with $E(S_k{}^2 \mid B_k)$ and S_jS_k, $j < k$, with $E(S_jS_k \mid B_k)$. This means that we approximate $S_k{}^2$ and S_jS_k by varying and averaging as little data as possible so that as many values C_{ik} as possible from data observed are kept fixed. Due to (A4) we have $E(f_k - \mathbf{f}_k \mid B_k) = 0$ and therefore $E(S_jS_k \mid B_k) = 0$ for $j < k$ because all $\mathbf{f}_r$, $r < k$, are scalars under B_k.

Because of

$$\text{(D12)}\quad E((f_k - \mathbf{f_k})^2 \mid B_k) = \mathrm{Var}(\mathbf{f_k} \mid B_k)$$

$$= \sum_{j=1}^{I-k} \mathrm{Var}(C_{j,k+1} \mid B_k) \Big/ \left(\sum_{j=1}^{I-k} C_{jk} \right)^2$$

$$= \sum_{j=1}^{I-k} \mathrm{Var}(C_{j,k+1} \mid C_{j1}, \dots, C_{jk}) \Big/ \left(\sum_{j=1}^{I-k} C_{jk} \right)^2$$

$$= \sum_{j=1}^{I-k} C_{jk} {\alpha_k}^2 \Big/ \left(\sum_{j=1}^{I-k} C_{jk} \right)^2$$

$$= {\alpha_k}^2 \Big/ \sum_{j=1}^{I-k} C_{jk}$$

we obtain

$$E({S_k}^2 \mid B_k) = \mathbf{f}_{\mathbf{I+1-i}}^2 \cdots \mathbf{f}_{\mathbf{k-1}}^2 \alpha_k^2 f_{k+1}^2 \cdots f_{I-1}^2 \Big/ \sum_{j=1}^{I-k} C_{jk}$$

Taken together, we have replaced $F^2 = (\sum S_k)^2$ with $\sum_k E({S_k}^2 \mid B_k)$ and because all terms of this sum are positive we can replace all unknown parameters f_k, ${\alpha_k}^2$ with their unbiased estimators $\mathbf{f_k}, \boldsymbol{\alpha_k}^2$. Altogether, we estimate
$F^2 = (f_{I+1-i} \cdot \ldots \cdot f_{I-1} - \mathbf{f_{I+1-i}} \cdot \ldots \cdot \mathbf{f_{I-1}})^2$ by

$$\sum_{k=I+1-i}^{I-1} \left(\mathbf{f}_{\mathbf{I+1-i}}^2 \cdots \mathbf{f}_{\mathbf{k-1}}^2 \cdot \boldsymbol{\alpha}_{\mathbf{k}}^2 \cdot \mathbf{f}_{\mathbf{k+1}}^2 \cdots \mathbf{f}_{\mathbf{I-1}}^2 \Big/ \sum_{j=1}^{I-k} C_{jk} \right) = \mathbf{f}_{\mathbf{I+1-i}}^2 \cdot \ldots \cdot \mathbf{f}_{\mathbf{I-1}}^2 \sum_{k=I+1-i}^{I-1} \frac{\boldsymbol{\alpha_k}^2 / \mathbf{f_k}^2}{\sum_{j=1}^{I-k} C_{jk}}$$

Using (D11), this means that we estimate $(E_i(C_{iI}) - \mathbf{C}_{iI})^2$ by

$$\text{(D13)}\quad C^2_{i,I+1-i}\mathbf{f}^2_{I+1-i}\cdot\ldots\cdot\mathbf{f}^2_{I-1}\sum_{k=I+1-i}^{I-1}\frac{\boldsymbol{\alpha}_k{}^2/\mathbf{f}_k{}^2}{\sum_{j=1}^{I-k}C_{jk}}$$

$$= \mathbf{C}^2_{iI}\sum_{k=I+1-i}^{I-1}\frac{\boldsymbol{\alpha}_k{}^2/\mathbf{f}_k{}^2}{\sum_{j=1}^{I-k}C_{jk}}$$

From (D2), (D10) and (D13) we finally obtain the estimator $(\text{s.e.}(\mathbf{C}_{iI}))^2$ for $\text{mse}(\mathbf{C}_{iI})$ as stated in the proposition.

Appendix E: Unbiasedness of the Estimator $\boldsymbol{\alpha}_k^2$

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \ne j$ are independent.

(5) There are unknown constants $\alpha_1, \ldots, \alpha_{I-1}$ with

$$\mathrm{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} \alpha_k^2, \qquad 1 \le i \le I,\ 1 \le k \le I-1$$

the estimators

$$\boldsymbol{\alpha}_k^2 = \frac{1}{I-k-1} \sum_{j=1}^{I-k} C_{jk} \left(\frac{C_{j,k+1}}{C_{jk}} - \mathbf{f}_k \right)^2, \qquad 1 \le k \le I-2$$

of α_k^2 are unbiased, that is we have

$$E(\boldsymbol{\alpha}_k^2) = \alpha_k^2, \qquad 1 \le k \le I-2$$

Proof: In this proof all summations are over the index j from j = 1 to j = I – k. The definition of $\boldsymbol{\alpha}_k^2$ can be rewritten as

$$\text{(E1)}\quad (I-k-1)\boldsymbol{\alpha}_k^2 = \Sigma(C_{j,k+1}^2 / C_{jk} - 2 \cdot C_{j,k+1} \mathbf{f}_k + C_{jk} \mathbf{f}_k^2)$$
$$= \Sigma(C_{j,k+1}^2 / C_{jk}) - \Sigma(C_{jk} \mathbf{f}_k^2)$$

using $\Sigma C_{j,k+1} = \mathbf{f}_k \Sigma C_{jk}$ according to the definition of $\mathbf{f}_k$. Using again the set

$$B_k = \{ C_{ij} \mid i+j \le I+1, j \le k \}$$

of variables C_{ij} assumed to be known, (E1) yields

$$\text{(E2)}\quad E((I-k-1)\boldsymbol{\alpha}_k^2 \mid B_k) = \Sigma E(C_{j,k+1}^2 \mid B_k) / C_{jk} - \Sigma C_{jk} E(\mathbf{f}_k^2 \mid B_k)$$

because C_{jk} is a scalar under the condition of B_k being known.

Due to the independence (4) of the accident years, conditions which are independent from the conditioned variable can be omitted in $E(C_{j,k+1}{}^2 \mid B_k)$, that is

$$\text{(E3)}\quad \begin{aligned} E(C_{j,k+1}{}^2 \mid B_k) &= E(C_{j,k+1}{}^2 \mid C_{j1},\ldots,C_{jk}) \\ &= Var(C_{j,k+1} \mid C_{j1},\ldots,C_{jk}) + (E(C_{j,k+1} \mid C_{j1},\ldots,C_{jk}))^2 \\ &= C_{jk}\alpha_k{}^2 + (C_{jk}f_k)^2 \end{aligned}$$

where the rule $E(X^2) = Var(X) + (E(X))^2$ and the assumptions (5) and (3) have also been used.

From (D12) and (A4) we gather

$$\text{(E4)}\quad \begin{aligned} E(\mathbf{f}_k{}^2 \mid B_k) &= Var(\mathbf{f}_k{}^2 \mid B_k) + (E(\mathbf{f}_k \mid B_k))^2 \\ &= \alpha_k{}^2 / \Sigma C_{jk} + f_k{}^2 \end{aligned}$$

Inserting (E3) and (E4) into (E2) we obtain

$$\begin{aligned} E((I-k-1)\boldsymbol{\alpha}_k{}^2 \mid B_k) &= \sum_{j=1}^{I-k}(\alpha_k{}^2 + C_{jk}f_k{}^2) - \sum_{j=1}^{I-k}\left(C_{jk}\alpha_k{}^2 / \sum_{j=1}^{I-k} C_{jk} + C_{jk}f_k{}^2\right) \\ &= (I-k)\alpha_k{}^2 - \alpha_k{}^2 \\ &= (I-k-1)\alpha_k{}^2 \end{aligned}$$

From this we immediately obtain $E(\boldsymbol{\alpha}_k{}^2 \mid B_k) = \alpha_k{}^2$

Finally, the iterative rule for expectations yields

$$E(\boldsymbol{\alpha}_k{}^2) = E(E(\boldsymbol{\alpha}_k{}^2 \mid B_k)) = E(\alpha_k{}^2) = \alpha_k{}^2$$

Appendix F: The Standard Error of the Overall Reserve Estimate

Proposition: Under the assumptions

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \quad 1 \le i \le I, 1 \le k \le I-1$$

(4) The variables $\{C_{i1}, \ldots, C_{iI}\}$ and $\{C_{j1}, \ldots, C_{jI}\}$ of different accident years $i \neq j$ are independent.

(5) There are unknown constants $\alpha_1, \ldots, \alpha_{I-1}$ with

$$\mathrm{Var}(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik}\alpha_k^2, \quad 1 \le i \le I, 1 \le k \le I-1$$

the standard error s.e.($\mathbf{R}$) of the overall reserve estimate

$$\mathbf{R} = \mathbf{R}_2 + \ldots + \mathbf{R}_I$$

is given by

$$(\text{s.e.}(\mathbf{R}))^2 = \sum_{i=2}^{I} \left\{ (\text{se.}(\mathbf{R}_i))^2 + \mathbf{C}_{iI} \left(\sum_{j=i+1}^{I} \mathbf{C}_{jI} \right) \sum_{k=I+1-i}^{I-1} \frac{2\alpha_k^2 / \mathbf{f}_k^2}{\sum_{n=1}^{I-k} C_{nk}} \right\}$$

Proof: This proof is analogous to that in Appendix D. The comments will therefore be brief. We first must determine the mean squared error mse($\mathbf{R}$) of $\mathbf{R}$. Using again D=$\{ C_{ik} \mid i + k \le I + 1\}$ we have

$$\text{(F1)} \quad \mathrm{mse}\left(\sum_{i=2}^{I} \mathbf{R}_i\right) = E\left(\left(\sum_{i=2}^{I} \mathbf{R}_i - \sum_{i=2}^{I} R_i\right)^2 \mid D\right)$$

$$= E\left(\left(\sum_{i=2}^{I} \mathbf{C}_{iI} - \sum_{i=2}^{I} C_{iI}\right)^2 \mid D\right)$$

$$= \mathrm{Var}\left(\sum_{i=2}^{I} C_{iI} \mid D\right) + \left(E\left(\sum_{i=2}^{I} C_{iI} \mid D\right) - \sum_{i=2}^{I} \mathbf{C}_{iI}\right)^2$$

The independence of the accident years yields

$$\text{(F2)}\quad \mathrm{Var}\left(\sum_{i=2}^{I} C_{iI} \mid D\right) = \sum_{i=2}^{I} \mathrm{Var}(C_{iI} \mid C_{i1}, \ldots, C_{i,I+1-i})$$

whose summands have been calculated in Appendix D, see (D9). Furthermore

$$\text{(F3)}\quad \left(E\left(\sum_{i=2}^{I} C_{iI} \mid D\right) - \sum_{i=2}^{I} \mathbf{C_{iI}}\right)^2 = \left(\sum_{I=2}^{I}(E(C_{iI} \mid D) - \mathbf{C_{iI}})\right)^2$$

$$= \sum_{2\le i,j\le I}(E(C_{iI} \mid D) - \mathbf{C_{iI}}) \cdot (E(C_{jI} \mid D) - \mathbf{C_{jI}})$$

$$= \sum_{2\le i,j\le I} C_{i,I+1-i} C_{j,I+1-j} F_i F_j$$

$$= \sum_{i=2}^{I}(C_{i,I+1-i}F_i)^2 + 2\sum_{i<j} C_{i,I+1-i} C_{j,I+1-j} F_i F_j$$

with (as for (D11))

$$F_i = f_{I+1-i} \cdots f_{I-1} - \mathbf{f_{I+1-i}} \cdots \mathbf{f_{I-1}}$$

which is identical to F of Appendix D but here we have to carry the index i, too. In Appendix D we have shown (cf. (D2) and (D11)) that

$$\mathrm{mse}(\mathbf{R_i}) = \mathrm{Var}(C_{iI} \mid C_{i1}, \ldots, C_{i,I+1-i}) + (C_{i,I+1-i}F_i)^2$$

Comparing this with (F1), (F2) and (F3) we see that

$$\text{(F4)}\quad \mathrm{mse}\left(\sum_{i=2}^{I} \mathbf{R_i}\right) = \sum_{i=2}^{I} \mathrm{mse}(\mathbf{R_i}) + \sum_{2\le i<j\le I} 2 \cdot C_{i,I+1-i} C_{j,I+1-j} F_i F_j$$

We therefore need only develop an estimator for $F_i F_j$. A procedure completely analogous to that for F^2 in the proof of Appendix D yields for $F_i F_j$, $i < j$, the estimator

$$\sum_{k=I+1-i}^{I-1} \mathbf{f_{I+1-j}} \ldots \mathbf{f_{I-i}} \mathbf{f_{I+1-i}^2} \ldots \mathbf{f_{k-1}^2} \alpha_k^2 \mathbf{f_{k+1}^2} \ldots \mathbf{f_{I-1}^2} \Big/ \sum_{n=1}^{I-k} C_{nk}$$

which immediately leads to the result stated in the proposition.

Appendix G: Testing for Correlations between Subsequent Development Factors

In this appendix we first prove that the basic assumption (3) of the chain ladder method implies that subsequent development factors $C_{ik} / C_{i,k-1}$ and $C_{i,k+1} / C_{ik}$ are not correlated. Then we show how we can test if this uncorrelatedness is met for a given run-off triangle. Finally, we apply this test procedure to the numerical example of section 6.

Proposition: Under the assumption

(3) There are unknown constants $f_1, \ldots, f_{I-1}$ with

$$E(C_{i,k+1} | C_{i1}, \ldots, C_{ik}) = C_{ik} f_k, \quad 1 \le i \le I, \; 1 \le k \le I-1$$

subsequent development factors $C_{ik} / C_{i,k-1}$ and $C_{i,k+1} / C_{ik}$ are uncorrelated, that is we have (for $1 \le i \le I$, $2 \le k \le I-1$)

$$E\left(\frac{C_{ik}}{C_{i,k-1}} \cdot \frac{C_{i,k+1}}{C_{ik}}\right) = E\left(\frac{C_{ik}}{C_{i,k-1}}\right) \cdot E\left(\frac{C_{i,k+1}}{C_{ik}}\right)$$

Proof: For $j \le k$ we have

$$
\begin{aligned}
\text{(G1)}\quad E(C_{i,k+1} / C_{ij}) &= E(E(C_{i,k+1} / C_{ij} | C_{i1}, \ldots, C_{ik})) && \text{(a)}\\
&= E(E(C_{i,k+1} | C_{i1}, \ldots, C_{ik}) / C_{ij}) && \text{(b)}\\
&= E(C_{ik} f_k / C_{ij}) && \text{(c)}\\
&= E(C_{ik} / C_{ij}) f_k && \text{(d)}
\end{aligned}
$$

Here equation (a) holds due to the iterative rule $E(X) = E(E(X|Y))$ for expectations, (b) holds because, given $C_{i1}, \ldots, C_{ik}$, C_{ij} is a scalar for $j \le k$, (c) holds due to (3) and (d) holds because f_k is a scalar.

From (G1) we obtain through the special case $j = k$

$$\text{(G2)}\quad E(C_{i,k+1} / C_{ik}) = E(C_{ik} / C_{ik}) f_k = f_k$$

and through $j = k - 1$

$$
\begin{aligned}
\text{(G3)}\quad E\left(\frac{C_{ik}}{C_{i,k-1}} \cdot \frac{C_{i,k+1}}{C_{ik}}\right) &= E\left(\frac{C_{i,k+1}}{C_{i,k-1}}\right) && \text{(G1)}\\
&= E\left(\frac{C_{ik}}{C_{i,k-1}}\right) f_k
\end{aligned}
$$

Inserting (G2) into (G3) completes the proof.

Designing the test procedure

The usual test for uncorrelatedness requires that we have identically distributed pairs of observations which come from a Normal distribution. Both conditions are usually not fulfilled for adjacent columns of development factors. (Note that due to (G2) the development factors $C_{i,k+1} / C_{ik}$, $1 \le i \le I - k$, have the same expectation but assumption (5) implies that they have different variances.) We therefore use the test with Spearman's rank correlation coefficient because this test is distribution-free and because by using ranks the differences in the variances of $C_{i,k+1} / C_{ik}$, $1 \le i \le I - k$, become less important. Even if these differences are negligible the test will only be of an approximate nature because, strictly speaking, it is a test for independence rather than for uncorrelatedness. But we will take this into account when fixing the critical value of the test statistic.

For the application of Spearman's test we consider a fixed development year k and rank the development factors $C_{i,k+1} / C_{ik}$ observed so far according to their size starting with the smallest one on rank one and so on. Let r_{ik}, $1 \le i \le I - k$, denote the rank of $C_{i,k+1} / C_{ik}$ obtained in this way, $1 \le r_{ik} \le I - k$. Then we do the same with the preceding development factors $C_{ik} / C_{i,k-1}$, $1 \le i \le I - k$, leaving out $C_{I+1-k,k} / C_{I+1-k,k-1}$ for which the subsequent development factor has not yet been observed. Let s_{ik}, $1 \le i \le I - k$, be the ranks obtained in this way, $1 \le s_{ik} \le I - k$. Now, Spearman's rank correlation coefficient T_k is defined to be

$$\text{(G4)} \quad T_k = 1 - 6\sum_{i=1}^{I-k}(r_{ik} - s_{ik})^2 / ((I-k)^3 - I + k)$$

It can be seen that

$$-1 \le T_k \le +1$$

and, under the null-hypothesis,

$$E(T_k) = 0$$
$$Var(T_k) = 1/(I - k - 1)$$

A value of T_k close to 0 indicates that the development factors between development years $k - 1$ and k and those between years k and $k + 1$ are not correlated. Any other value of T_k indicates that the factors are (positively or negatively) correlated.

For a formal test we do not want to consider every pair of columns of adjacent development years separately in order to avoid an accumulation of the error probabilities. We therefore consider the triangle as a whole. This also is preferable from a practical point of view because it is more important to know whether correlations globally prevail than to find a small part of the triangle with correlations. We therefore combine all values T_2, T_3, ..., T_{I-2} obtained in the same way like T_k. (There is no T_1 because there are no development factors before development year k=1 and similarly there is also no T_I; even T_{I-1} is not included because there is only one rank and therefore no randomness.)

According to Appendix B we should not form an unweighted average of $T_2, ..., T_{I-2}$ but rather use weights which are inversely proportional to $Var(T_k) = 1/(I-k-1)$. This leads to weights which are just equal to one less than the number of pairs (r_{ik}, s_{ik}) taken into account by T_k which seems very reasonable.

We thus calculate

$$\text{(G5)} \qquad T = \sum_{k=2}^{I-2}(I-k-1)T_k \Big/ \sum_{k=2}^{I-2}(I-k-1)$$

$$= \sum_{k=2}^{I-2} \frac{I-k-1}{(I-2)(I-3)/2} T_k$$

$$E(T) = \sum_{k=2}^{I-2} E(T_k) = 0$$

$$\text{(G6)} \quad Var(T) = \sum_{k=2}^{I-2}(I-k-1)^2\, Var(T_k) \Big/ \left(\sum_{k=2}^{I-2}(I-k-1)\right)^2$$

$$= \sum_{k=2}^{I-2}(I-k-1) \Big/ \left(\sum_{k=2}^{I-2}(I-k-1)\right)^2$$

$$= \frac{1}{(I-2)(I-3)/2}$$

where for the calculation of Var(T) we used the fact that under the null-hypothesis subsequent development factors and therefore also different T_k's are uncorrelated.

Because the distribution of a single T_k with $I-k \geq 10$ is Normal in good approximation and because T is the aggregation of several uncorrelated T_k's (which all are symmetrically distributed around their mean 0) we can assume that T has approximately a Normal distribution and use this to design a significance test. Usually, when applying a significance test one rejects the null-hypothesis if it is very unlikely to hold, e.g. if the value of the test statistic is outside its 95% confidence interval. But in our case we propose to use only a 50% confidence interval because the test is only of an approximate nature and because we want to detect correlations already in a substantial part of the run-off triangle. Therefore, as the probability for a

Standard Normal variate lying in the interval (−.67, .67) is 50% we do not reject the null-hypothesis of having uncorrelated development factors if

$$-\frac{0.67}{\sqrt{((I-2)(I-3)/2)}} \leq T \leq +\frac{0.67}{\sqrt{((I-2)(I-3)/2)}}$$

If T is outside this interval we should be reluctant with the application of the chain ladder method and analyze the correlations in more detail. In such a case, an autoregressive model of an order > 1 is probably more appropriate, for example by replacing the fundamental chain ladder assumption (3) with

$$E(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}) = C_{ik} f_k + C_{i,k-1} g_k$$

Application to the example of section 6:

We start with the table of all development factors:

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
i=1	1.6	1.32	1.08	1.15	1.20	1.11	1.033	1.00	1.01
i=2	40.4	1.26	1.98	1.29	1.13	0.99	1.043	1.03	
i=3	2.6	1.54	1.16	1.16	1.19	1.03	1.026		
i=4	2.0	1.36	1.35	1.10	1.11	1.04			
i=5	8.8	1.66	1.40	1.17	1.01				
I=6	4.3	1.82	1.11	1.23					
I=7	7.2	2.72	1.12						
I=8	5.1	1.89							
I=9	1.7								

As described above we first rank column F_1 according to the size of the factors, then leave out the last element and rank the column again. Then we do the same with columns F_2 to F_8. This yields the following table:

r_{i1}	s_{i2}	r_{i2}	s_{i3}	r_{i3}	s_{i4}	r_{i4}	s_{i5}	r_{i5}	s_{i6}	r_{i6}	s_{i7}	r_{i7}	s_{i8}	r_{i8}
1	1	2	2	1	1	2	2	5	4	4	3	2	1	1
9	8	1	1	7	6	6	5	3	2	1	1	3	2	2
4	3	4	4	4	3	3	3	4	3	2	2	1		
3	2	3	3	5	4	1	1	2	1	3				
8	7	5	5	6	5	4	4	1						
5	4	6	6	2	2	5								
7	6	8	7	3										
6	5	7												
2														

We now add the squared differences between adjacent rank columns of equal length, that is we add $(s_{ik} - r_{ik})^2$ over i for every k, $2 \le k \le 8$. This yields 68, 74, 20, 24, 6, 6 and 0. (Remember that we have to leave out k = 1 because there is no s_{i1}, and k = 9 because there is only one pair of ranks and therefore no randomness.) From these figures we obtain Spearman's rank correlation coefficients T_k according to formula (G4):

k	2	3	4	5	6	7	8
T_k	4/21	−9/28	3/7	−1/5	2/5	−1/2	1
$I-k-1$	7	6	5	4	3	2	1

The (I − k − 1)-weighted average of the T_k's is T = .070 (see formula (G5)). Because of Var(T) = 1/28 (see (G6)) the 50% confidence limits for T are $\pm .67/\sqrt{28} = \pm .127$. Thus, T is within its 50%-interval and the hypothesis of having uncorrelated development factors is not rejected.

Appendix H: Testing for Calendar Year Effects

One of the three basic assumptions underlying the chain ladder method was seen to be assumption (4) of the independence of the accident years. The main reason why this independence can be violated in practice is the fact that we can have certain calendar year effects such as major changes in claims handling or in case reserving or external influences such as substantial changes in court decisions or inflation. Note that a constant rate of inflation which has not been removed from the data is extrapolated into the future by the chain ladder method. In the following, we first generally describe a procedure to test for such calendar year influences and then apply it to our example.

Designing the test procedure:

A calendar year influence affects one of the diagonals

$$D_j = \{C_{j1}, C_{j-1,2}, \ldots, C_{2,j-1}, C_{1j}\}, \qquad 1 \le j \le I$$

and therefore also influences the adjacent development factors

$$A_j = \{C_{j2} / C_{j1}, C_{j-1,3} / C_{j-1,2}, \ldots, C_{1,j+1} / C_{1j}\}$$

and

$$A_{j-1} = \{C_{j-1,2} / C_{j-1,1}, C_{j-2,3} / C_{j-2,2}, \ldots, C_{1j} / C_{1,j-1}\}$$

where the elements of D_j form either the denominator or the numerator. Thus, if due to a calendar year influence the elements of D_j are larger (smaller) than usual, then the elements of A_{j-1} are also larger (smaller) than usual and the elements of A_j are smaller (larger) than usual.

Therefore, in order to check for such calendar year influences we only have to subdivide all development factors into 'smaller' and 'larger' ones and then to examine whether there are diagonals where the small development factors or the large ones clearly prevail. For this purpose, we order for every k, $1 \le k \le I - 1$, the elements of the set

$$F_k = \{C_{i,k+1} / C_{ik} \mid 1 \le i \le I - k\}$$

that is of the column of all development factors observed between development years k and k + 1, according to their size and subdivide them into one part LF_k of larger factors being greater than the median of F_k and into a second part SF_k of smaller factors below the median of F_k. (The median of a set of real numbers is defined to be a number which divides the set into two parts with the same number of elements.) If the number $I - k$ of elements of F_k is odd there is one element of F_k which is equal to the median and therefore assigned to neither of the sets LF_k and SF_k; this element is eliminated from all further considerations.

Having done this procedure for each set F_k, $1 \le k \le I-1$, every development factor observed is

- either eliminated (like e.g. the only element of F_{I-1})
- or assigned to the set $L = LF_1 + ... + LF_{I-2}$ of larger factors
- or assigned to the set $S = SF_1 + ... + SF_{I-2}$ of smaller factors

In this way, every development factor which is not eliminated has a 50% chance of belonging to either L or S.

Now we count for every diagonal A_j, $1 \le j \le I-1$, of development factors the number L_j of large factors, that is elements of L, and the number S_j of small factors, that is elements of S. Intuitively, if there is no specific change from calendar year j to calendar year j + 1, A_j should have about the same number of small factors as of large factors, that is L_j and S_j should be of approximately the same size apart from pure random fluctuations. But if L_j is significantly larger or smaller than S_j or, equivalently, if

$$Z_j = \min(L_j, S_j)$$

that is the smaller of the two figures, is significantly smaller than $(L_j + S_j)/2$, then there is some reason for a specific calendar year influence.

In order to design a formal test we need the probability distribution of Z_j under the null-hypothesis that each development factor has a 50 % probability of belonging to either L or S. This distribution can easily be established. We give an example for the case where $L_j + S_j = 5$, that is where the set A_j contains 5 development factors without counting any eliminated factor. Then the number L_j has a Binomial distribution with n = 5 and p = .5, that is

$$\text{prob}(L_j = m) = \binom{n}{m}\frac{1}{2^n} = \binom{5}{m}\frac{1}{2^5}, \; m = 0, 1, ..., 5$$

Therefore

$$\begin{aligned}
\text{prob}(S_j = 5) &= \text{prob}(L_j = 0) = 1/32 \\
\text{prob}(S_j = 4) &= \text{prob}(L_j = 1) = 5/32 \\
\text{prob}(S_j = 3) &= \text{prob}(L_j = 2) = 10/32 \\
\text{prob}(S_j = 2) &= \text{prob}(L_j = 3) = 10/32 \\
\text{prob}(S_j = 1) &= \text{prob}(L_j = 4) = 5/32
\end{aligned}$$

$$\text{prob}(S_j = 0) = \text{prob}(L_j = 5) = 1/32$$

This yields

$$\text{prob}(Z_j = 0) = \text{prob}(L_j = 0) + \text{prob}(S_j = 0) = 2/32$$
$$\text{prob}(Z_j = 1) = \text{prob}(L_j = 1) + \text{prob}(S_j = 1) = 10/32$$
$$\text{prob}(Z_j = 2) = \text{prob}(L_j = 2) + \text{prob}(S_j = 2) = 20/32$$

In this way we obtain very easily the following table for the cumulative probability distribution function of Z_j:

n	$\text{prob}(Z_j \le 0)$	$\text{prob}(Z_j \le 1)$	$\text{prob}(Z_j \le 2)$ ...
≤4	>10%	>10%	>10%
5	6.25%	>10%	>10%
6	3.1%	>10%	>10%
7	1.6%	>10%	>10%
8	0.8%	7.0%	>10%
9	0.4%	3.9%	>10%
10	0.2%	2.1%	>10%
11	0.1%	1.2%	6.5%
...			

Now, we use this table in the following way: Any realization $Z_j = z_j$ with a cumulative probability $\text{prob}(Z_j \le z_j) \le 10\,\%$ indicates that the corresponding set $A_j = \{C_{j2}/C_{j1}, C_{j-1,3}/C_{j-1,2}, \ldots\}$ contains either significantly many "larger" or significantly many "smaller" development factors. Then, the factors of the predominant type (either the larger or the smaller factors of A_j) are assumed to be influenced by a specific calendar year effect and are viewed to be outliers. Therefore, it seems to be advisable to reduce their weight when calculating the age-to-age factors $\mathbf{f}_k$.

Specifically, it is proposed to reduce the weight of each of these outlying development factors to 50 % of its original weight, that is to calculate

$$\mathbf{f}_k = \sum_{i=1}^{I-k} w_{ik} C_{i,k+1} \Big/ \sum_{i=1}^{I-k} w_{ik} C_{ik}$$

with $w_{ik} = .5$ if $C_{i,k+1}/C_{ik}$ belongs to the factors of the predominant type (either larger or smaller) of its corresponding set A_{i+k-1} and if A_{i+k-1} shows a significant calendar year effect, that is if $\text{prob}(Z_{i+k-1} \le z_{i+k-1}) \le 10\,\%$. In all other cases we put $w_{ik} = 1$ as usual. Strictly speaking, the formulae for $\boldsymbol{\alpha}_k^2$ and for the standard error must be changed analogously, if some $w_{ik} < 1$ are used.

As with every test procedure which is applied several times there is an accumulation of the error probabilities, that is the danger increases that we find a significant case which in reality is not extraordinary. But here, this will not cause any essential disadvantage as we only change weights and do not discard anything entirely.

Application to the example of section 6:

We start with the triangle of all development factors observed:

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9
i=1	1.6	1.32	1.08	1.15	1.20	1.11	1.033	1.00	1.01
i=2	40.4	1.26	1.98	1.29	1.13	0.99	1.043	1.03	
i=3	2.6	1.54	1.16	1.16	1.19	1.03	1.026		
i=4	2.0	1.36	1.35	1.10	1.11	1.04			
i=5	8.8	1.66	1.40	1.17	1.01				
i=6	4.3	1.82	1.11	1.23					
i=7	7.2	2.72	1.12						
i=8	5.1	1.89							
i=9	1.7								

We have to subdivide each column F_k into the subset SF_k of 'smaller' factors below the median of F_k and into the subset LF_k of 'larger' factors above the median. This can be done very easily with the help of the rank columns r_{ik} established in Appendix G: The half of factors with small ranks belongs to SF_k, those with large ranks to LF_k and if the total number is odd we have to eliminate the mean rank. Replacing a small rank with 'S', a large rank with 'L' and a mean rank with '*' we obtain the following picture:

	j=1	j=2	j=3	j=4	j=5	j=6	j=7	j=8	j=9
j=1	S	S	S	S	L	L	*	S	*
j=2	L	S	L	L	*	S	L	L	
j=3	S	S	*	S	L	S	S		
j=4	S	S	L	S	S	L			
j=5	L	L	L	L	S				
j=6	*	L	S	L					
j=7	L	L	S						
j=8	L	L							
j=9	S								

We now count for every diagonal A_j, $2 \leq j \leq 9$, the number L_j of L's and the number S_j of S's. We have left out A_1 because it contains at most one element which is not eliminated, and therefore Z_1 is not a random variable but always = 0. With the notations s_j, l_j, z_j for the realizations of the random variables S_j, L_j, Z_j and with $n = s_j + l_j$ as above, we obtain the following table:

j	s_j	l_j	z_j	n	$\text{prob}(Z_j \le z_j)$
2	1	1	1	2	>10%
3	3	0	0	3	>10%
4	3	1	1	4	>10%
5	1	3	1	4	>10%
6	1	3	1	4	>10%
7	2	4	2	6	>10%
8	4	4	4	8	>10%
9	4	4	4	8	>10%

According to the probabilities $\text{prob}(Z_j \le z_j)$ there does not seem to be any calendar year effect. Therefore, there is no reason to change any weight in the calculation of the age-to-age factors.

As a final check for calendar year effects we can plot all standardized residuals

$$(C_{i,k+1} / C_{ik} - f_k)\sqrt{C_{ik}} / \alpha_k, \qquad 2 \le i + k \le I$$

against the calendar years $j = i + k$. For the data of our example, the resulting plot is shown in Figure 14. There does not seem to be any specific trend or irregularity in the pattern of these residuals. The fact that only positive residuals are absolutely larger than 1.6 hints at a positive skewness of the distribution of the development factors.

◇

Figure 1: Regression and Residuals
Ci2 against Ci1

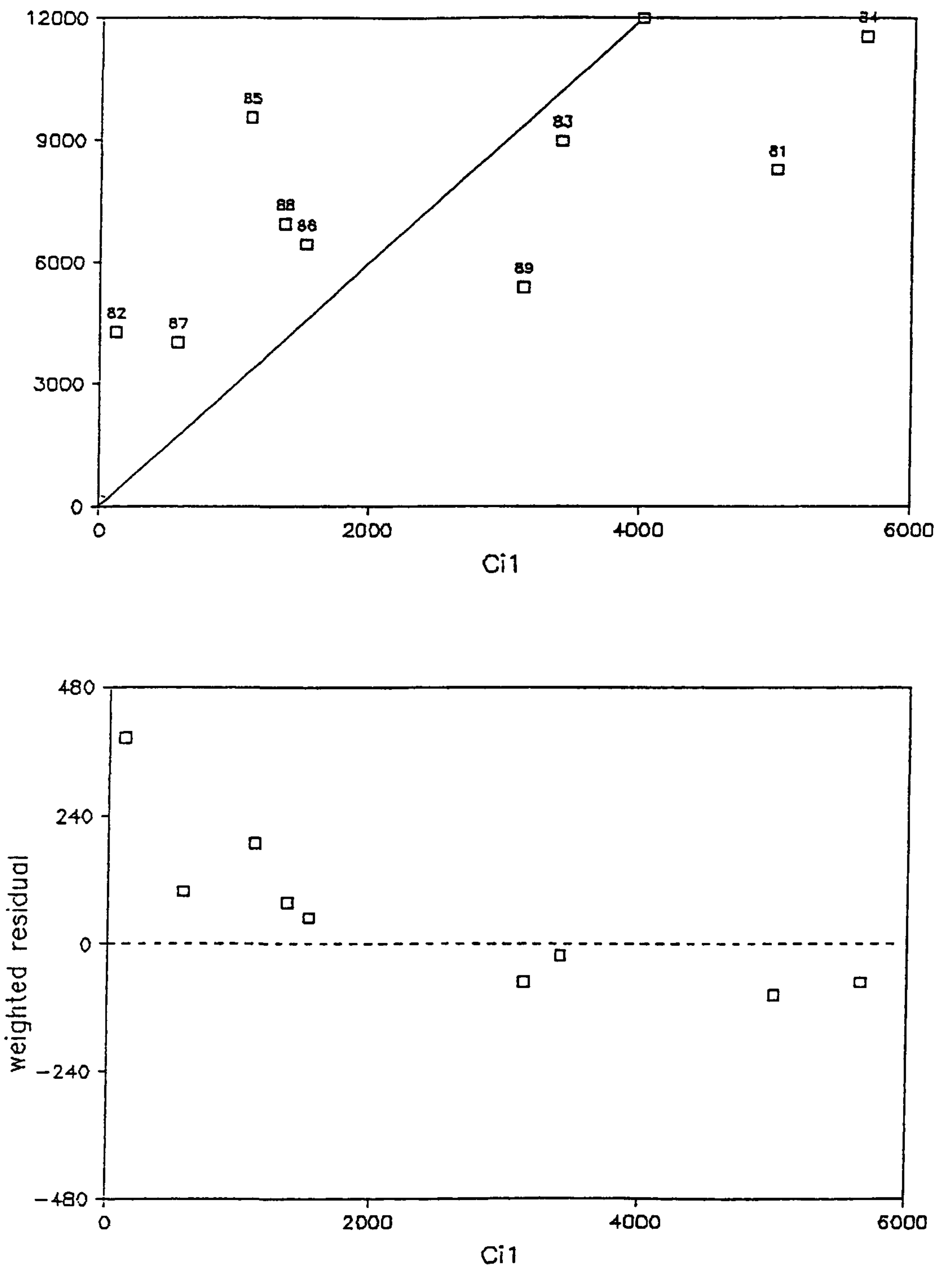


Figure 2: Regression and Residuals

Ci3 against Ci2

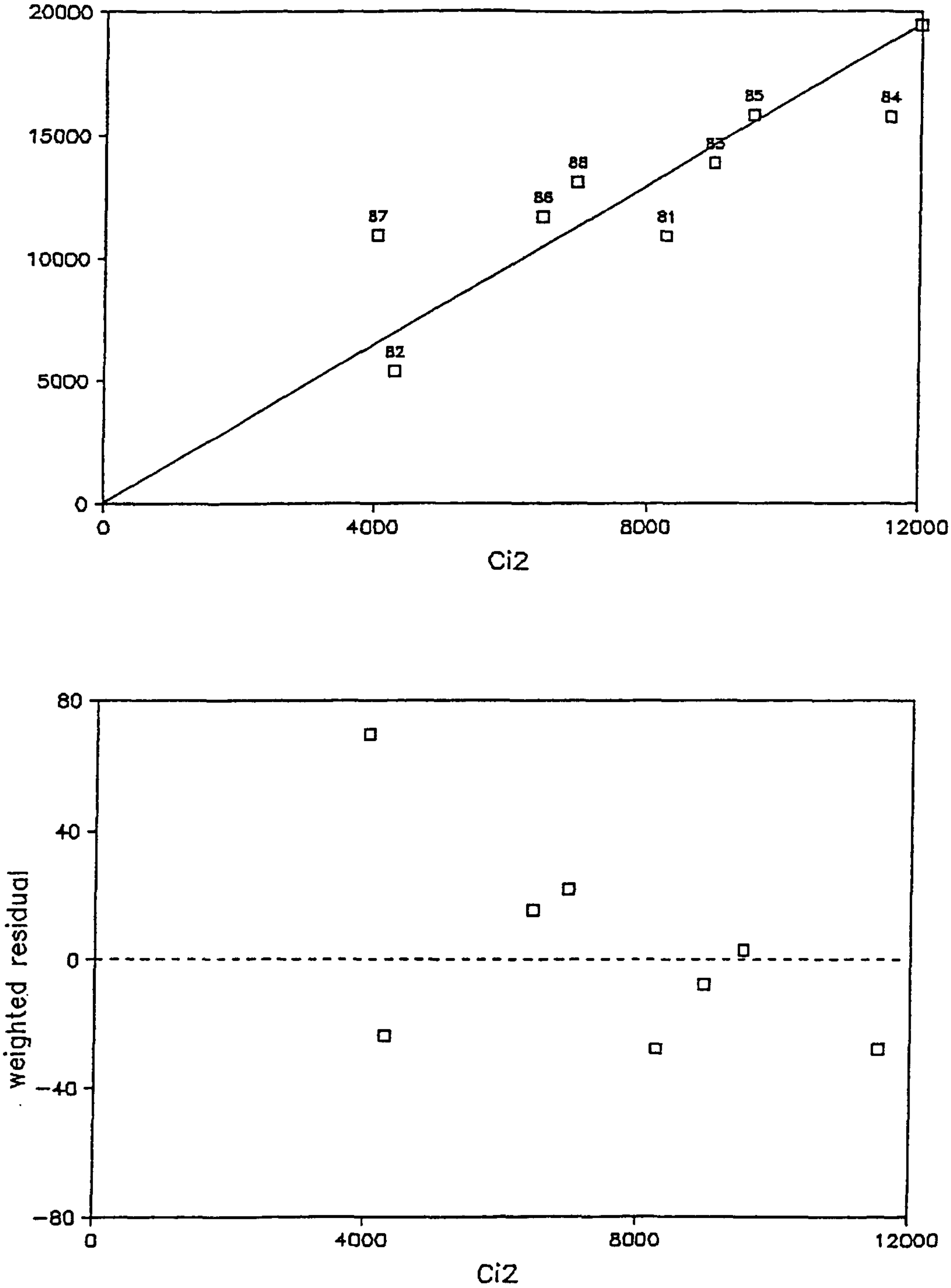
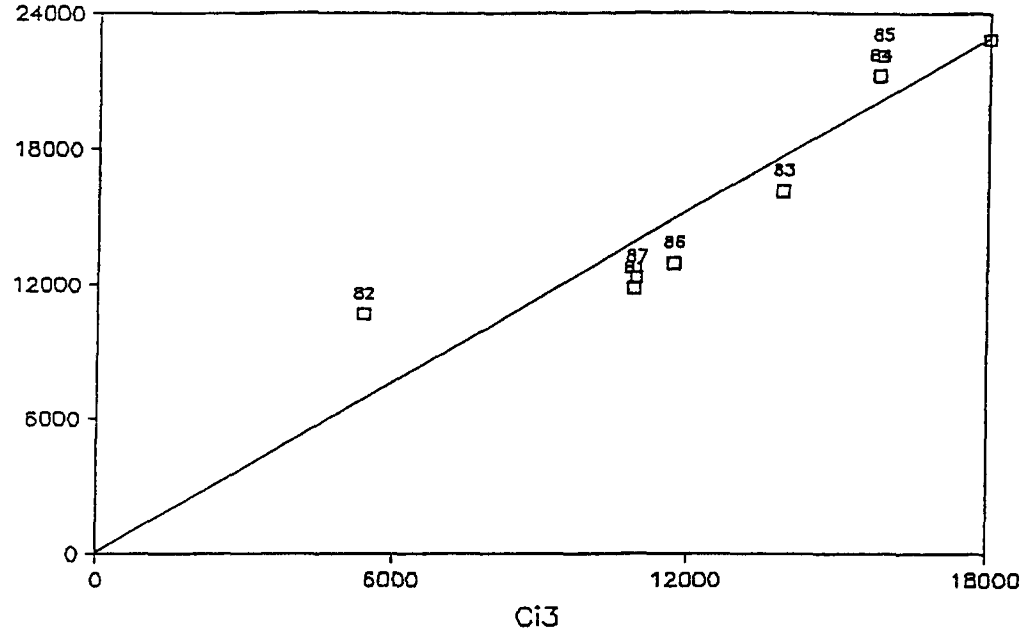


Figure 3: Regression and Residuals

Ci4 against Ci3



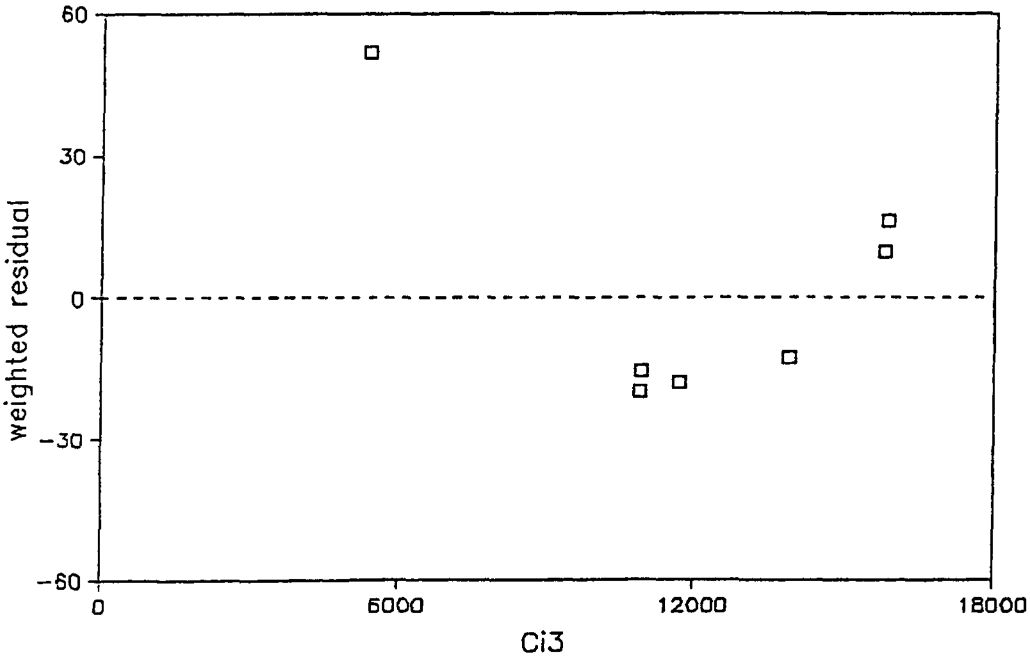


Figure 4: Regression and Residuals

Ci5 against Ci4

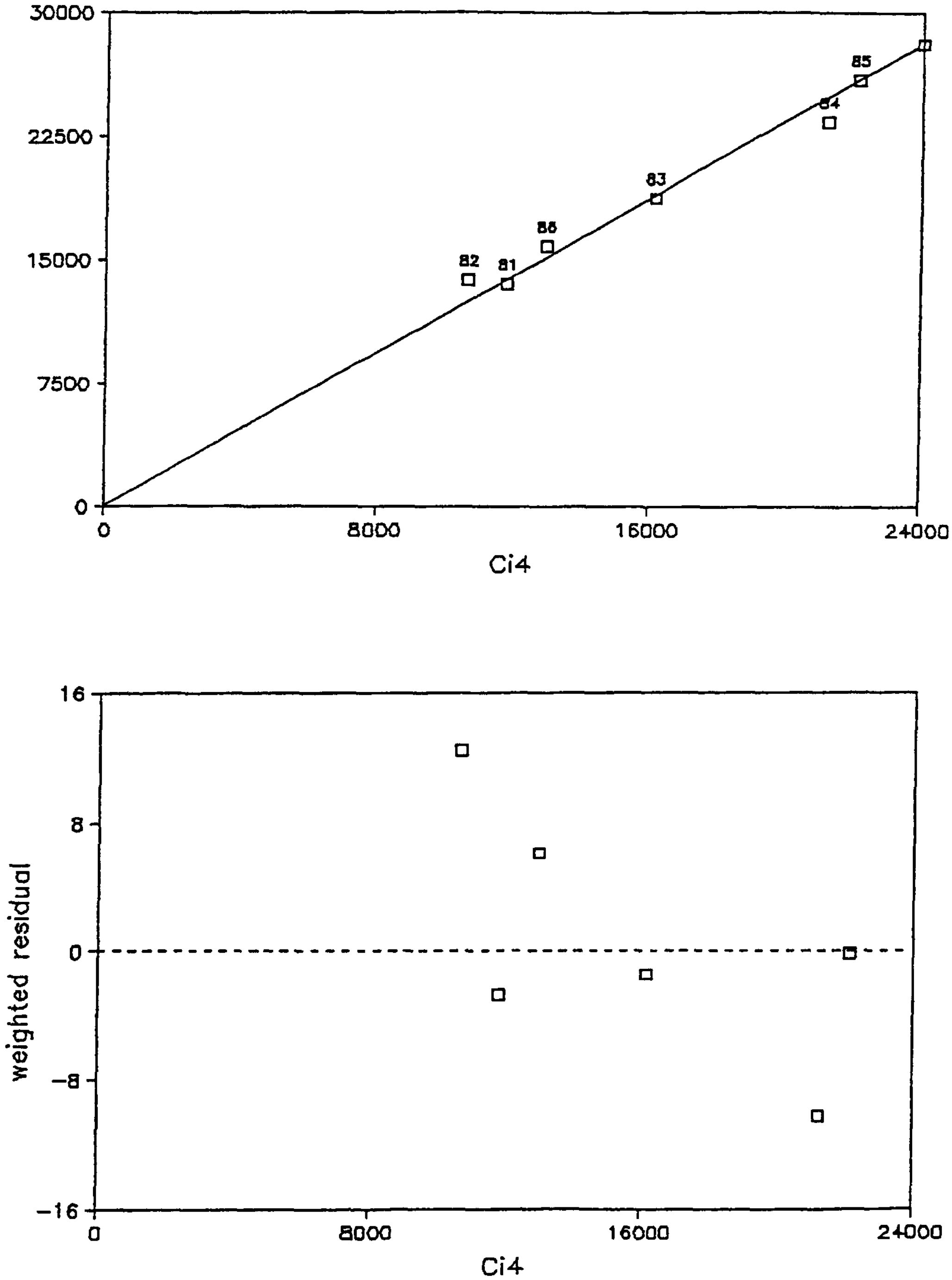


Figure 5: Regression and Residuals

Ci6 against Ci5

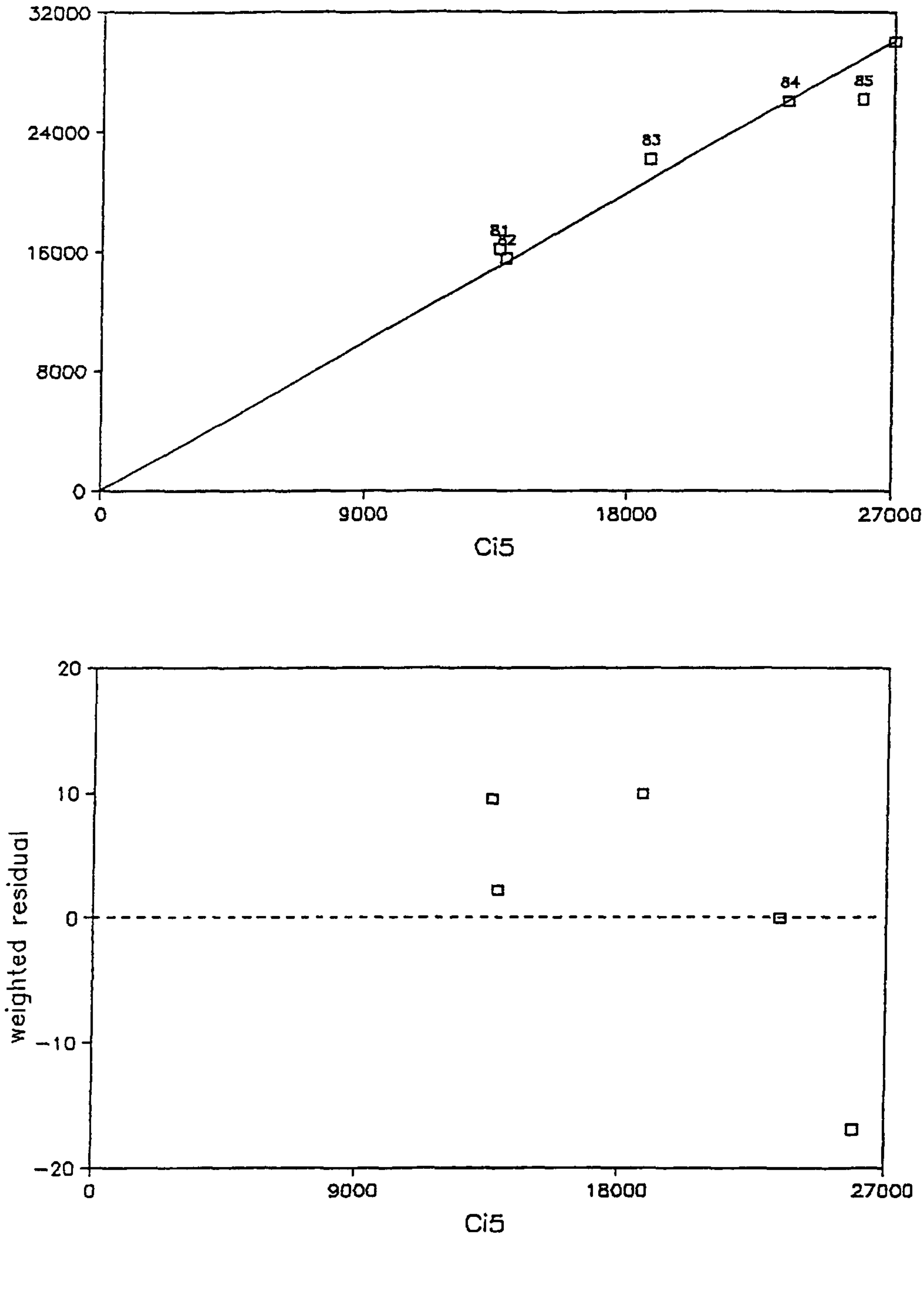


Figure 6: Regression and Residuals

Ci7 against Ci6

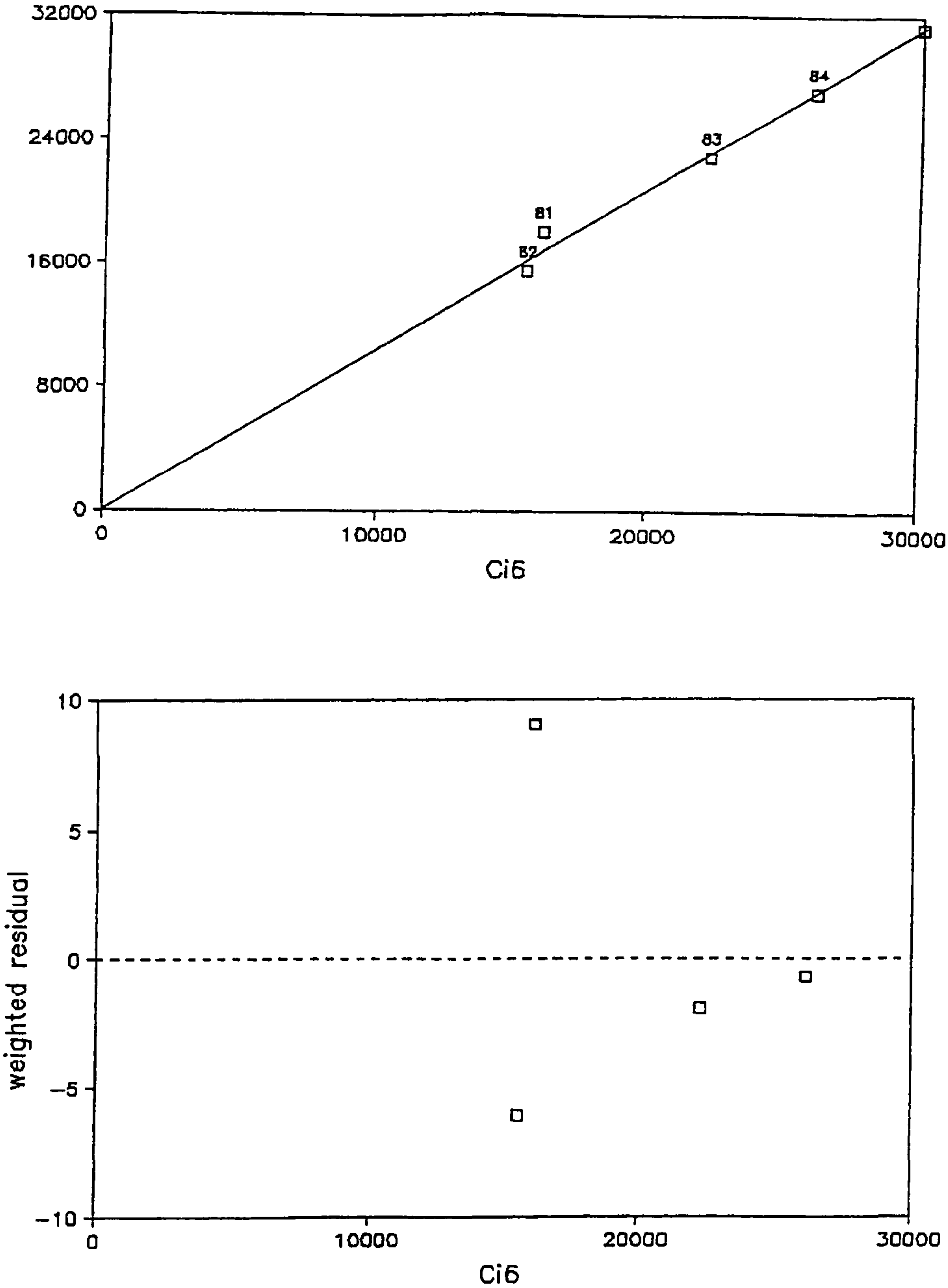


Figure 7: Regression and Residuals

Ci8 against Ci7

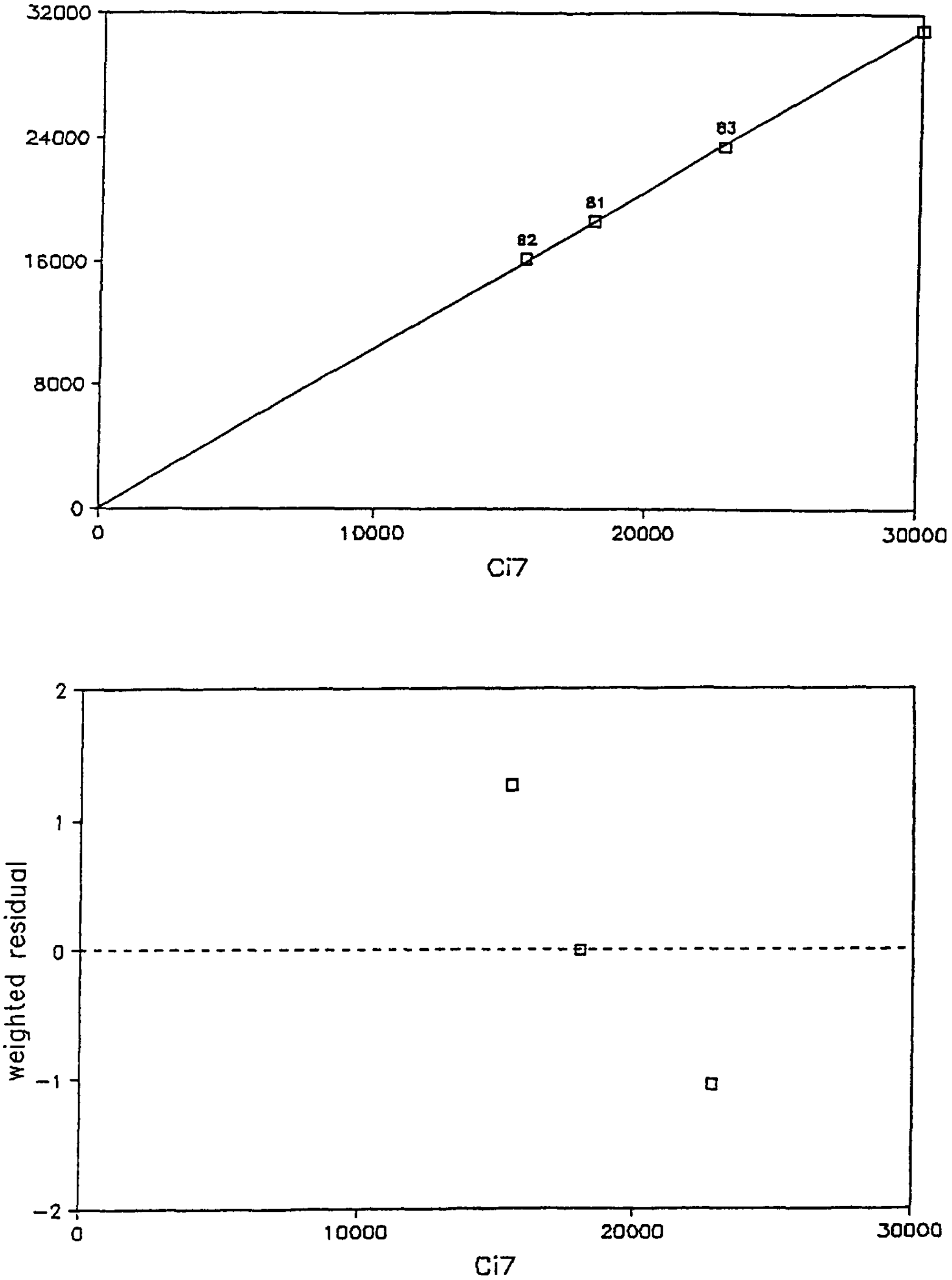


Figure 8: Regression and Residuals
Ci9 against Ci8

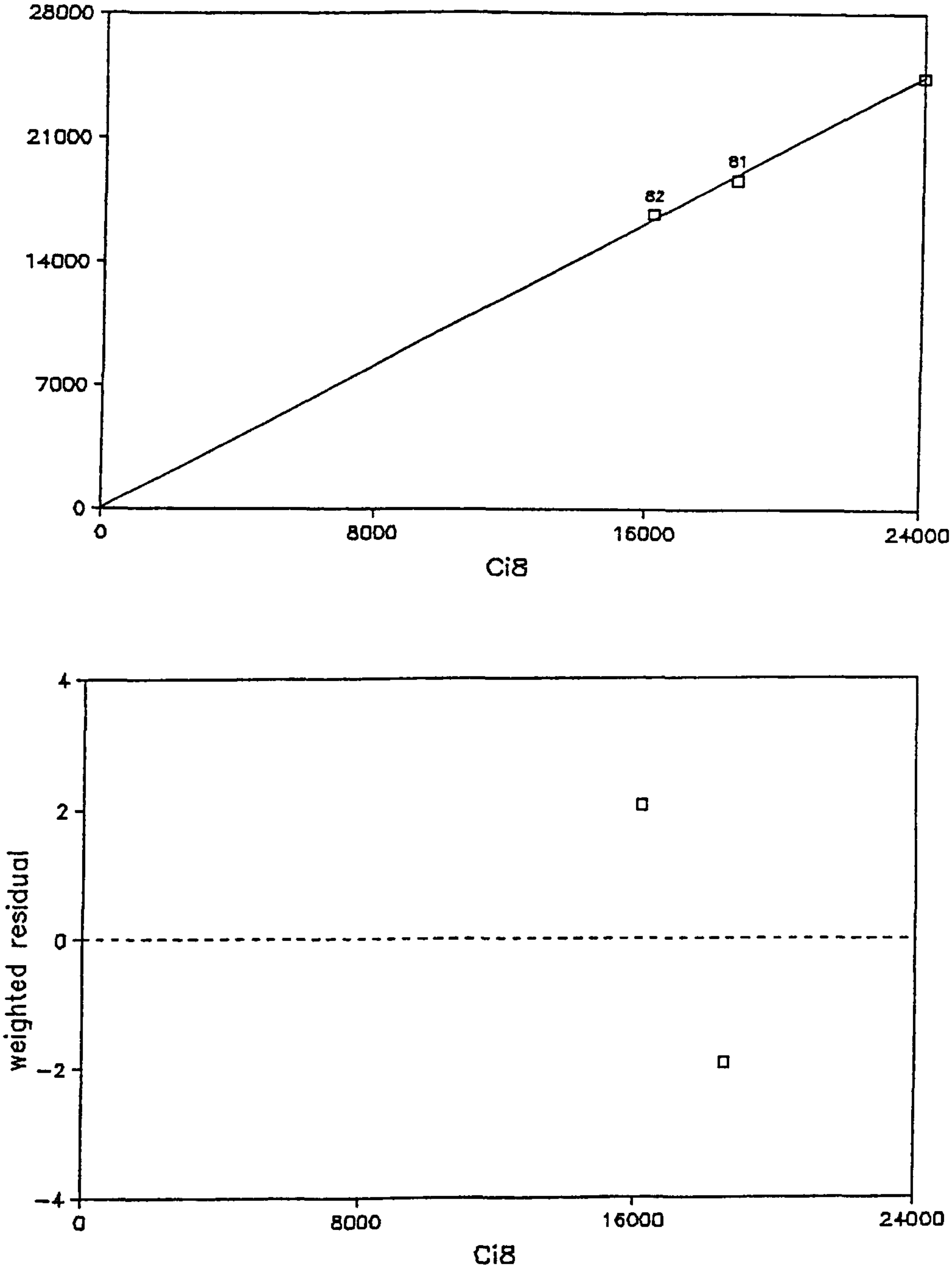


Figure 9: Residual Plots for fk0

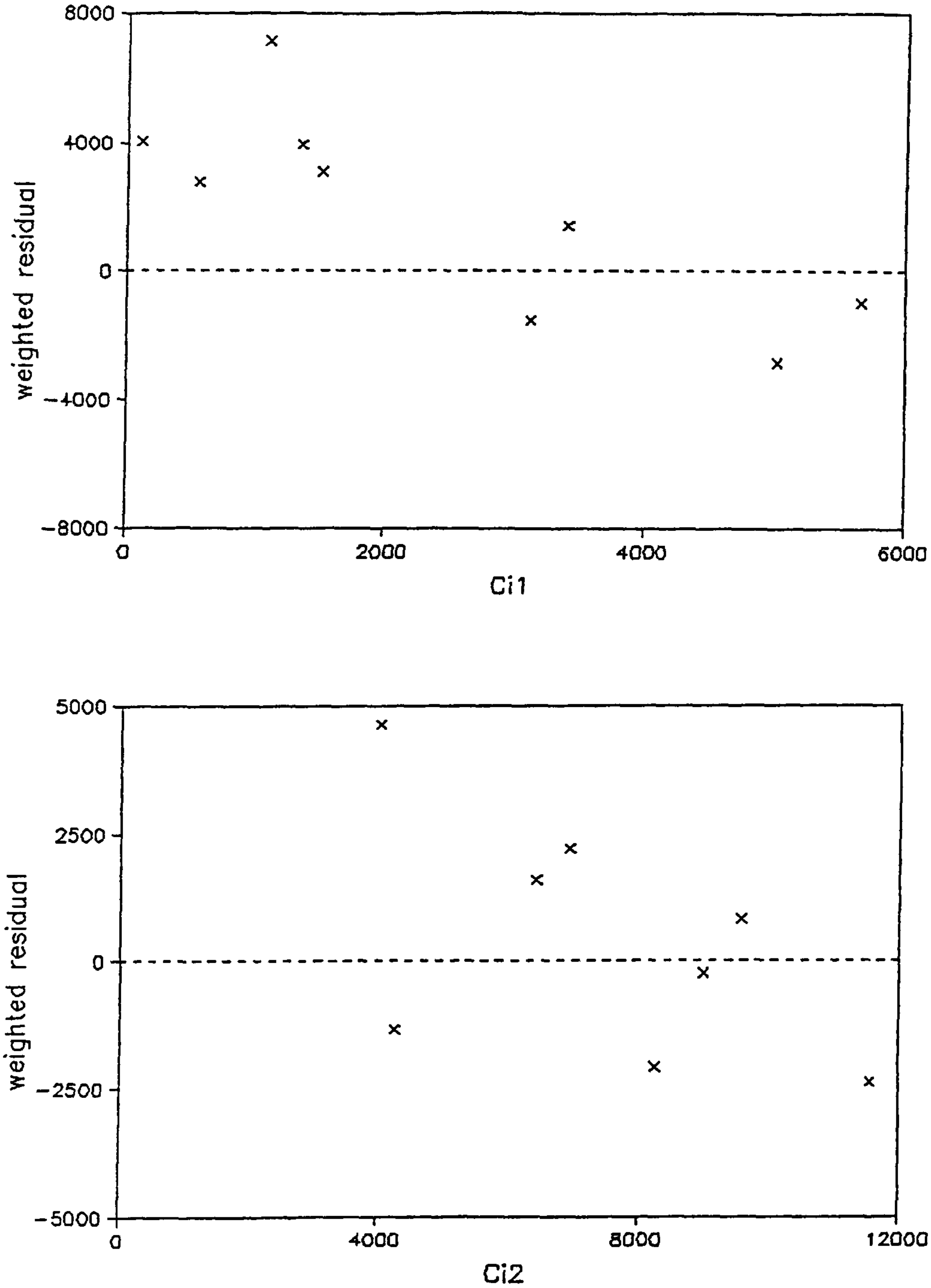
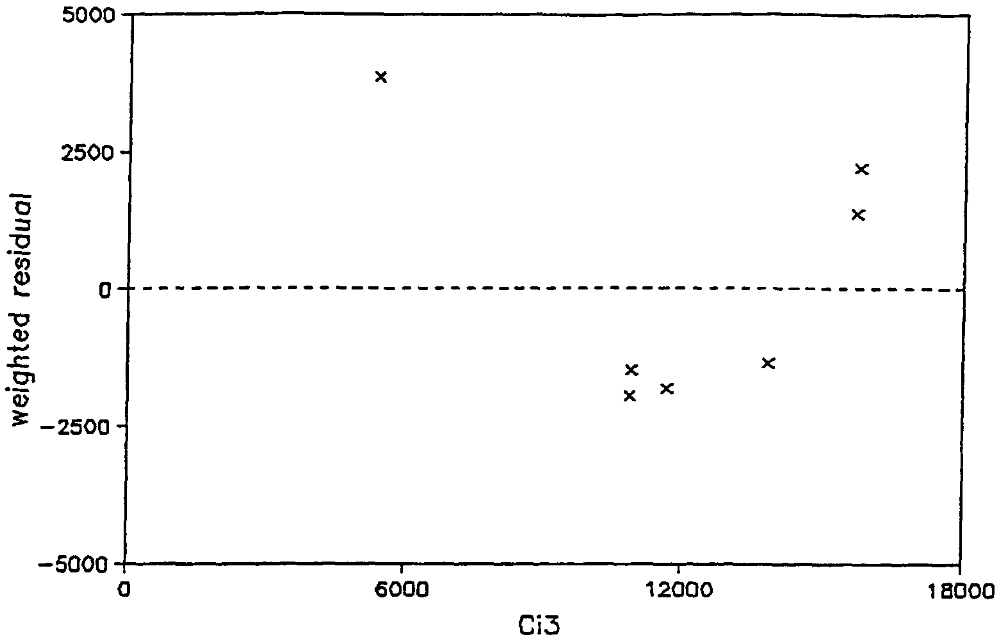


Figure 10: Residual Plots for fk0



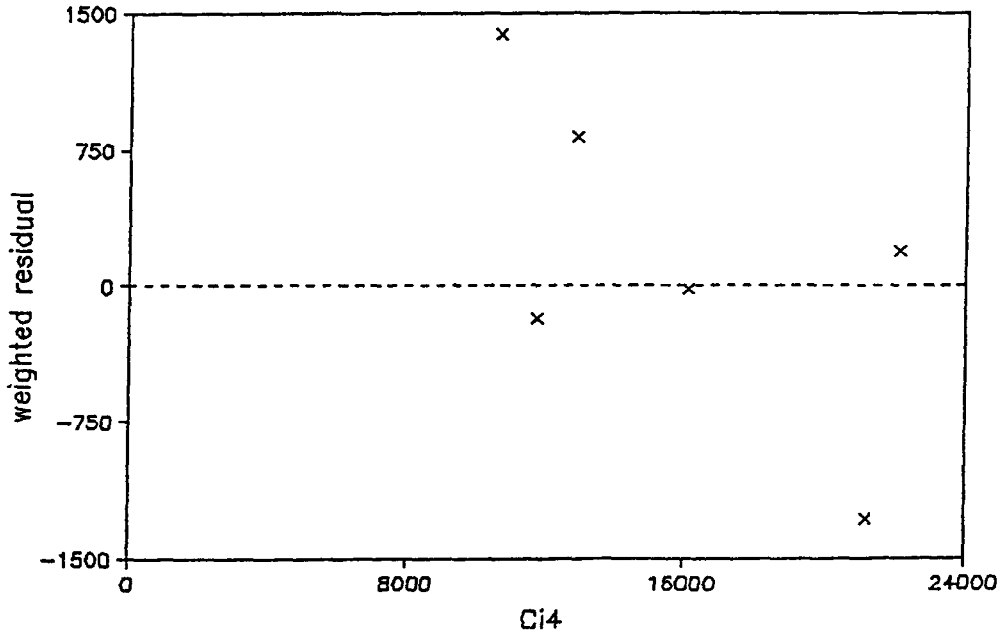


Figure 11: Residual Plots for fk2

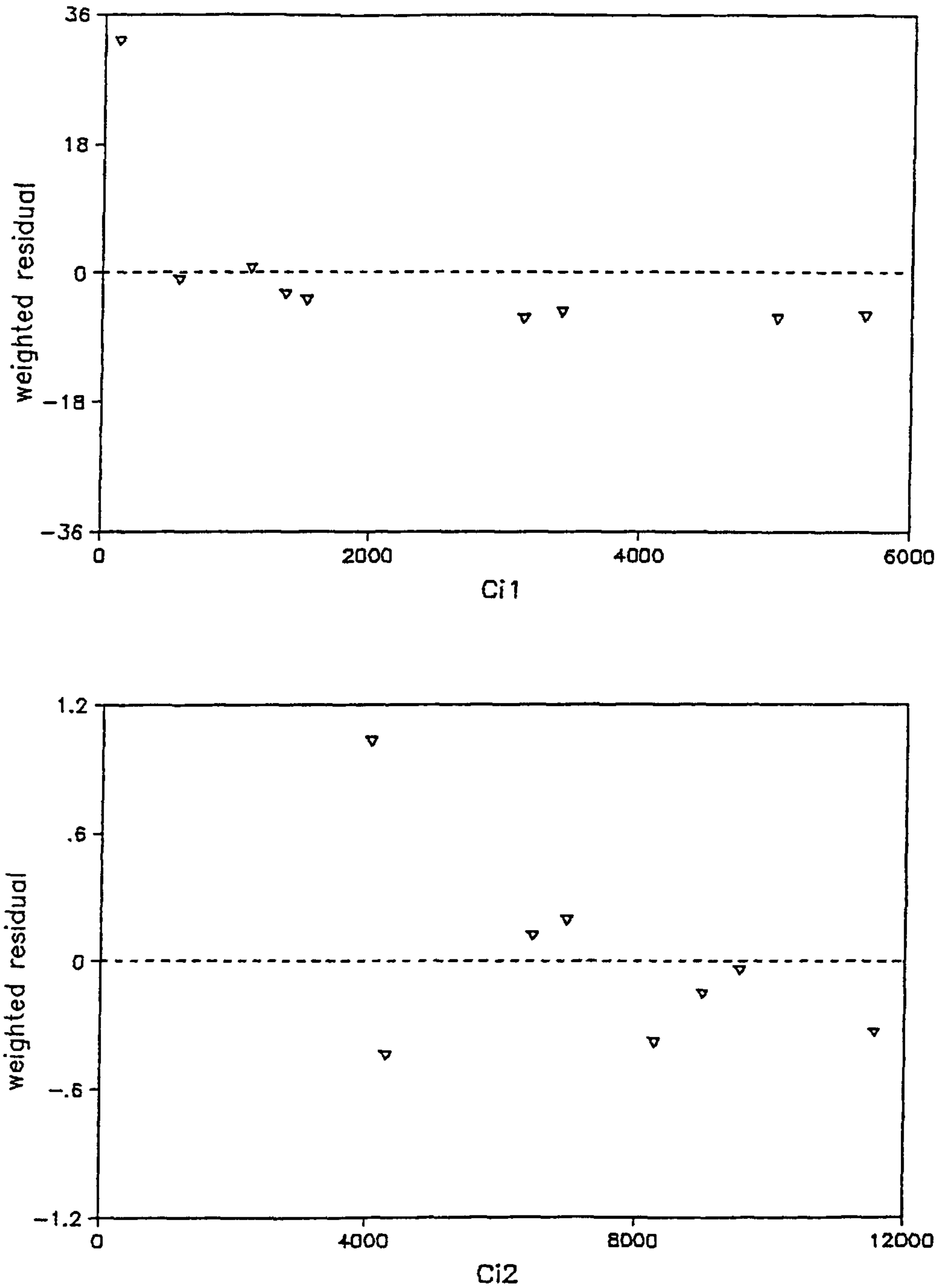
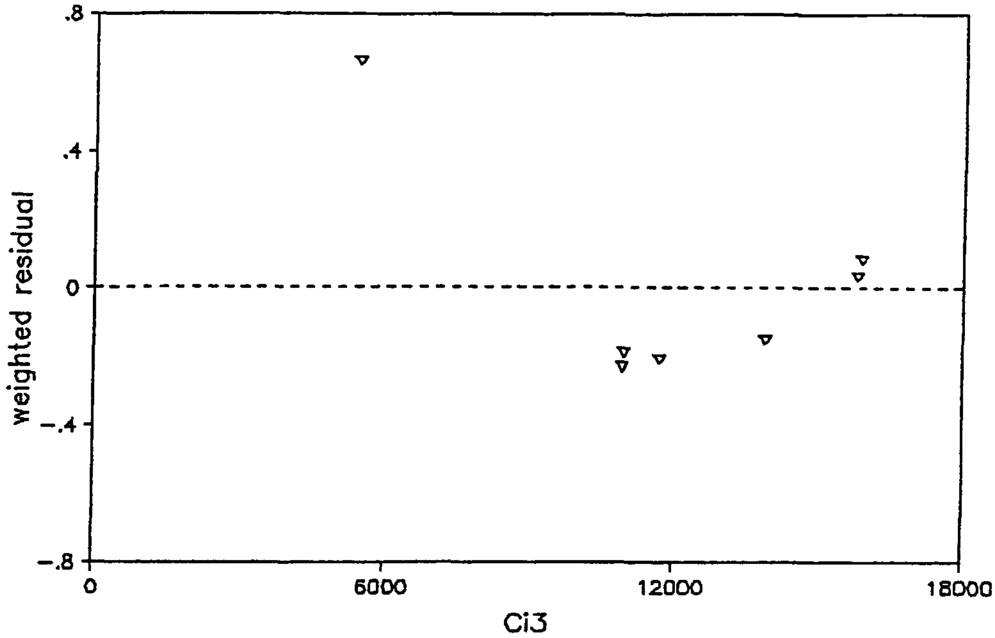


Figure 12: Residual Plots for fk2



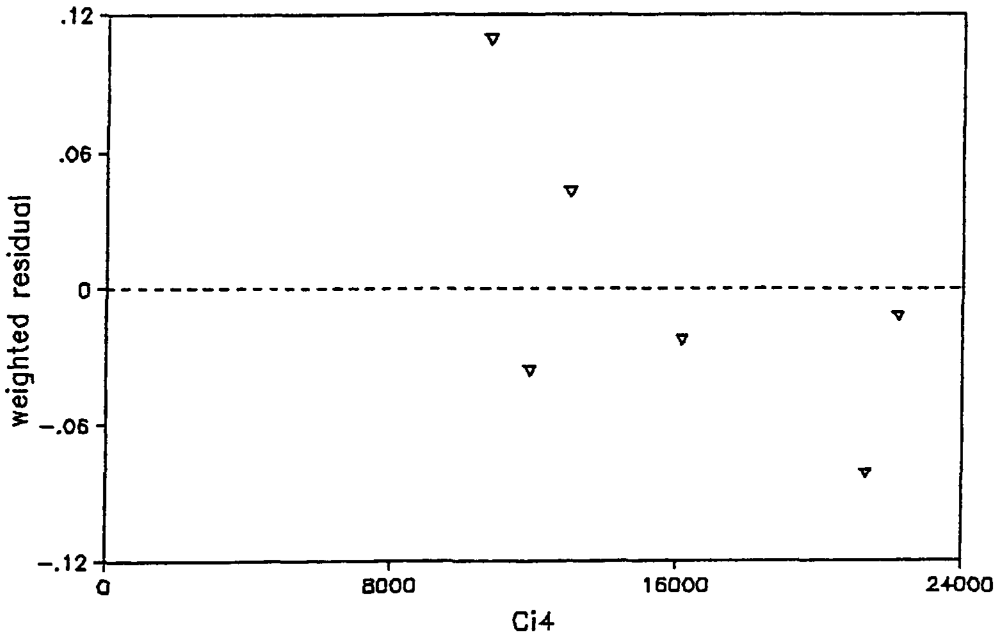
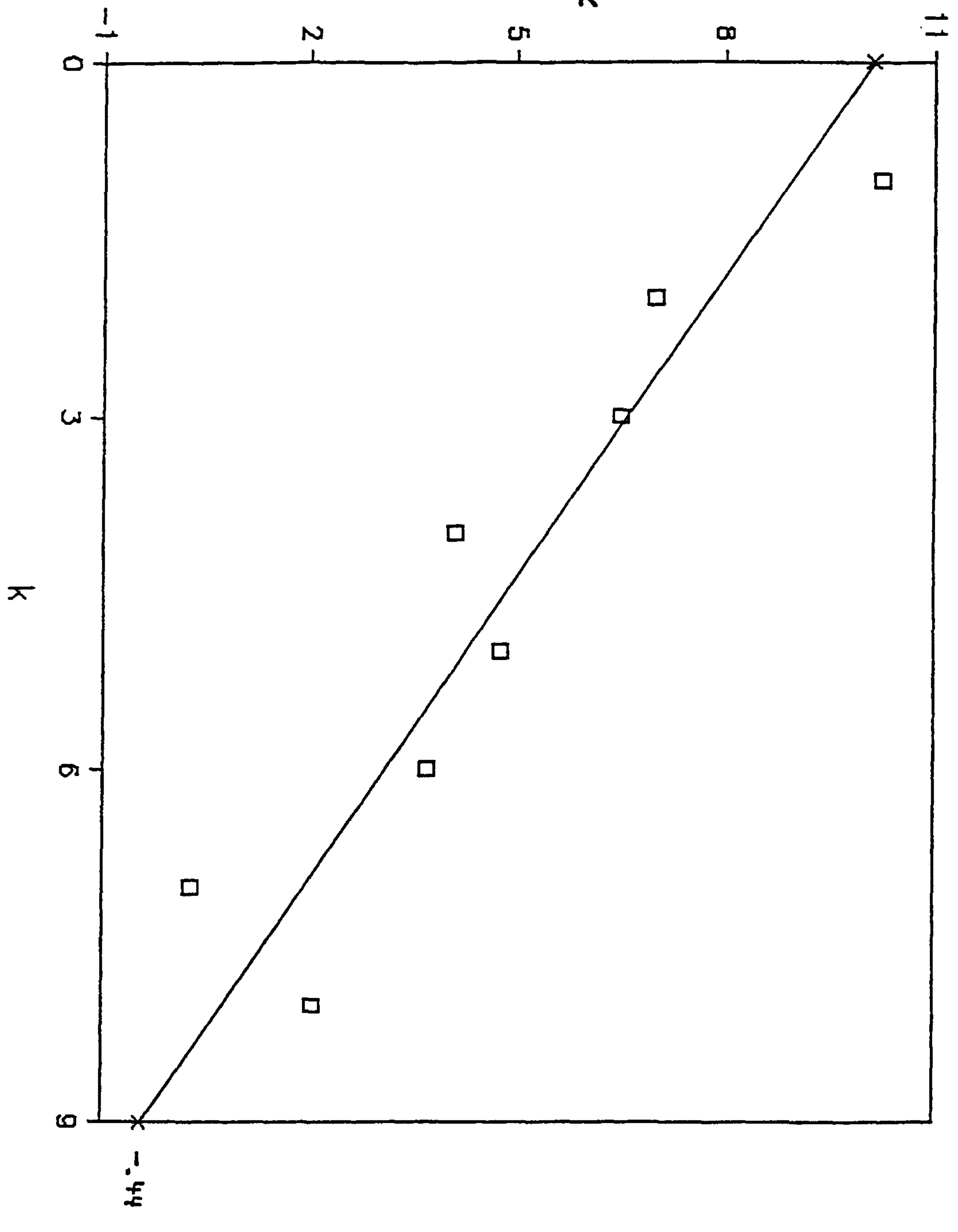


Figure 13: Plot of $\ln(\alpha_k^2)$ against k



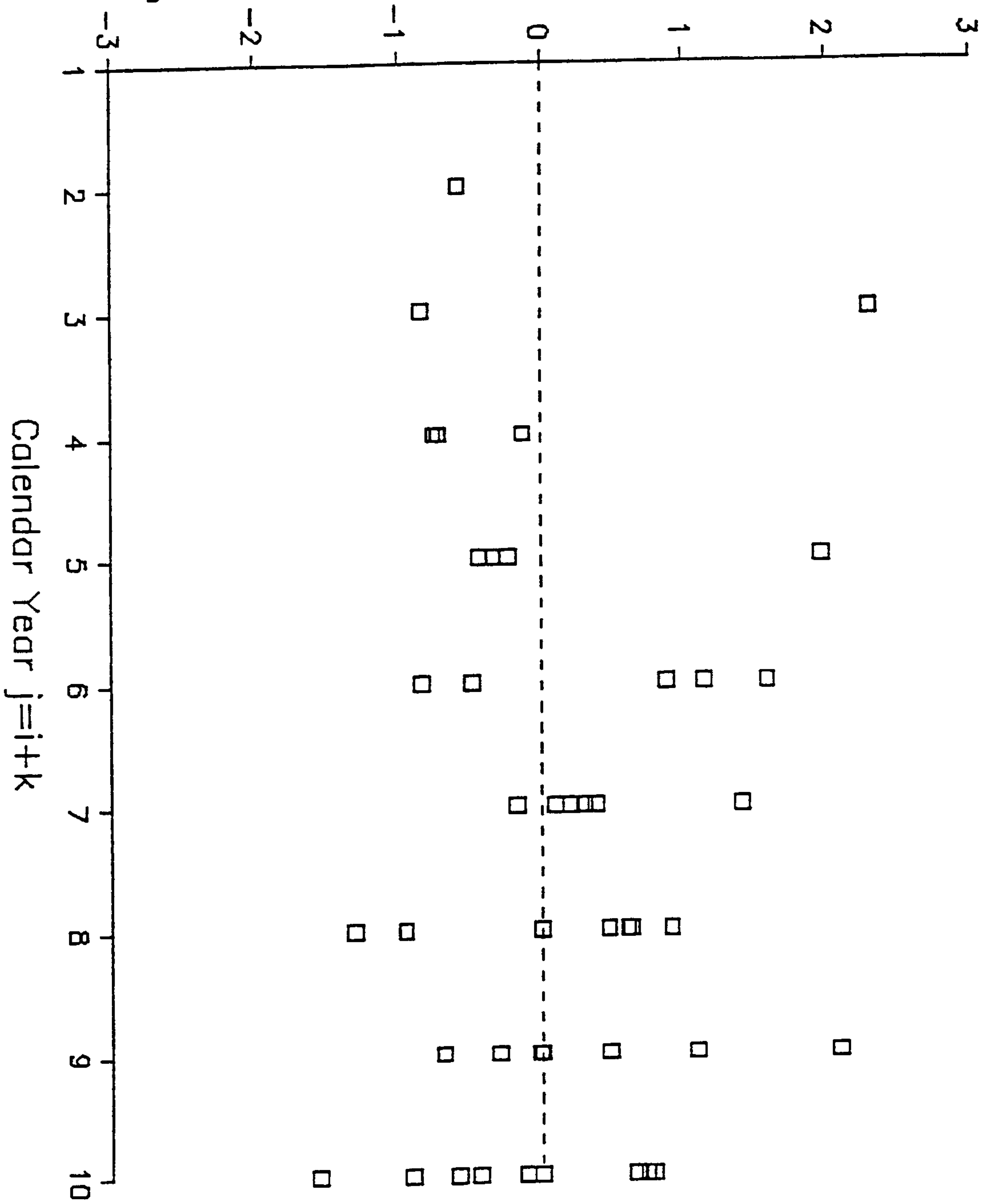


Figure 14: Std.Residuals vs. Calendar Year