ACTUARIAL RESEARCH CLEARING HOUSE
1996 VOL. 2

LEAST SQUARES ESTIMATION OF FUTURE COSTS OF ONGOING LARGE CLAIMS

Robert G. Lynch
Wisconsin Physicians Service Insurance Corporation
Madison, Wisconsin

Abstract. This paper develops a method for estimating the future claim costs of known ongoing large medical claims. It applies Least Squares analysis of historical experience to derive two arrays of factors. One array is to be applied as a vector inner product with claim costs already incurred and/or paid to project a total future claim cost. The second array distributes the total future claim cost over the months following the analysis date.

A frequent problem in analyzing the financial position of medical stop-loss reinsurance business is encountered in estimating the liabilities for future claim costs of ongoing claims. Inspection of current claims by health underwriters and projection on an individual basis is a frequently used method for establishing these costs for both experience-rated full-risk coverage and specific (individual) stop-loss coverage issued to self-funded groups. However, using this method for an entire block of business for such purposes as calculating reserves is cumbersome, time-consuming and expensive.

It is desirable, then, to have an alternative method for estimating the future claim costs from currently known, ongoing claims. This method should reflect the following goals in its design:

1) Keep it simple in application. The only variables used should be dollar costs and times of claim incurral/ payment. Including items such as analysis by individual diagnosis codes rapidly increases the complexity and decreases the practicality.

2) Factors generated and applied should reflect real experience and involve as few assumptions as possible.

3) The timing of claims should be taken into account and weighted accordingly. Claims incurred and/or paid recently will probably have greater future costs associated with them than older claims which may have already run their course or have terminated.

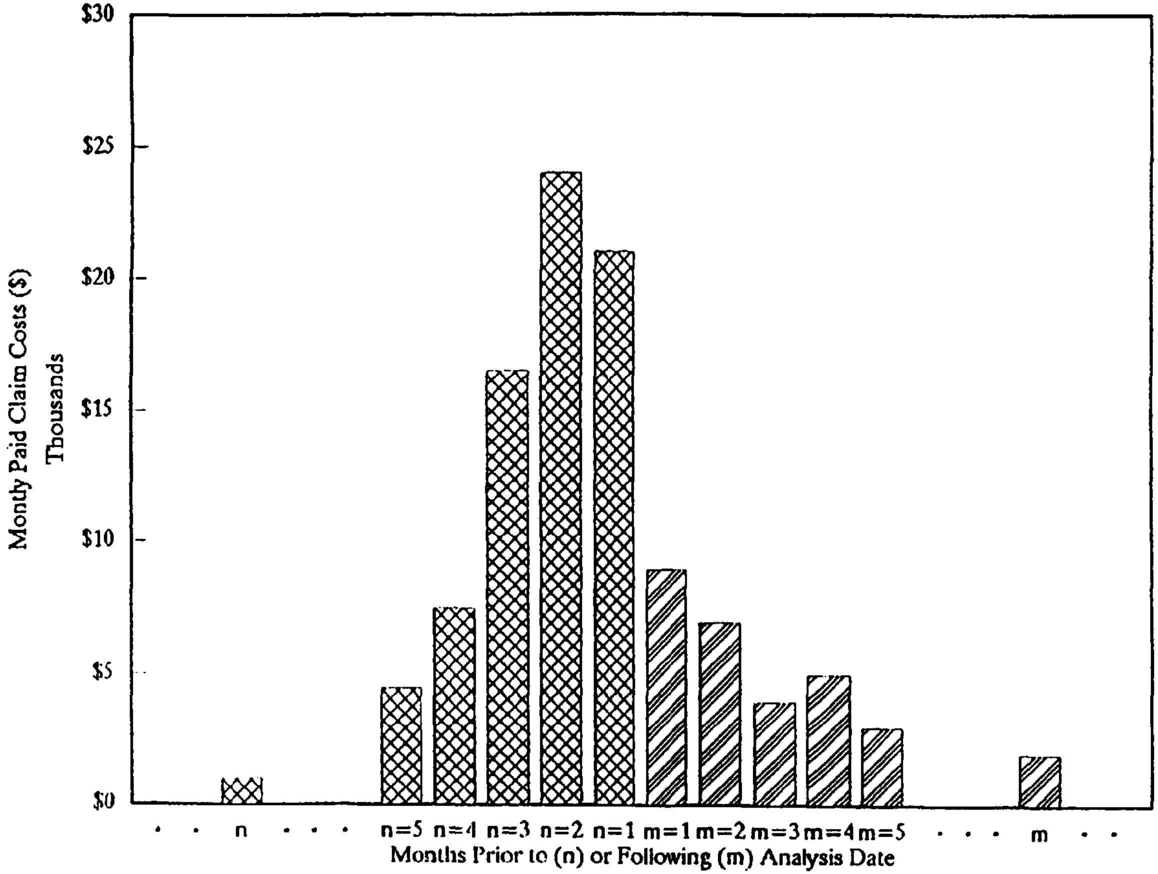
This paper develops methodology for calculating two arrays of factors to be used for projecting the future costs of ongoing claims for a block of business. The first array is weighting factors to be applied to claims incurred and/or paid during the period just before the analysis date. Their purpose is to calculate a projected Total Future Paid Claim Cost for ongoing claims, **P**. (I will discuss the Paid-basis case here, and use **I** for the Total Future Incurred Claim Cost for ongoing claims). These factors will be very similar in concept and application to claim lag completion factors used in calculating IBNR reserves. I will designate this array as $[W^P(n)]$ or $[W^I(n)]$, according to whether the data is on a paid or incurred basis, respectively.

The second array of factors is used to project the distribution of future claim costs beyond the analysis date on a month-by-month basis. These factors multiply the Total Future Paid Claim Cost, **P** (**I**), generated by the first array of factors above to produce an array of Future Paid (Incurred) Claim Costs allocated on a month-by-month basis. I will designate this array as $[F^P(m)]$ or $[F^I(m)]$, according to whether the projection is on a Paid or Incurred basis, respectively.

Let $C^P_i(n)$ be the dollar amount of Claims Paid for the <u>i</u>th covered member in the <u>n</u>th month prior to the analysis date (Figure 1). If the analysis date is December 31, then $C^P_i(1)$ is the claims paid in December, $C^P_i(2)$ is claims paid in November, etc. The Total Paid Claims for the <u>i</u>th member in the previous **N** months is

$$C^P_i(1) + C^P_i(2) + C^P_i(3) + \ . \ . \ . + C^P_i(N)$$

Figure 1 – Monthly Claim Costs of a Typical Large Claim Before and After Analysis Date



Let $W^P(n)$ be the weighting factor to be applied to claims paid in the nth month prior to the analysis date. For example, if the analysis date is December 31, then $W^P(1)$ is the weighting factor applied to claims paid in December, $W^P(2)$ is the weighting for claims paid in November, $W^P(3)$ is for October, etc. (see Figure 2).

Calculate the Total Future Paid Claim Cost for the ith member, $\mathbf{P}_i$, as

$$\mathbf{P}_i = [W^P(n)] \bullet [C^P_i(n)]$$

$$= W^P(1) \cdot C^P_i(1) + W^P(2) \cdot C^P_i(2) + W^P(3) \cdot C^P_i(3) + \ldots$$

Let $F^P(m)$ be the proportion of $\mathbf{P}_i$ which is expected to be paid in the mth month following the analysis date. If the analysis date is December 31, then $F^P(1)$ is the proportion of $\mathbf{P}_i$ paid in January, $F^P(2)$ is the proportion of $\mathbf{P}_i$ paid in February, $F^P(3)$ is for March, etc. (Figure 2). It is required that

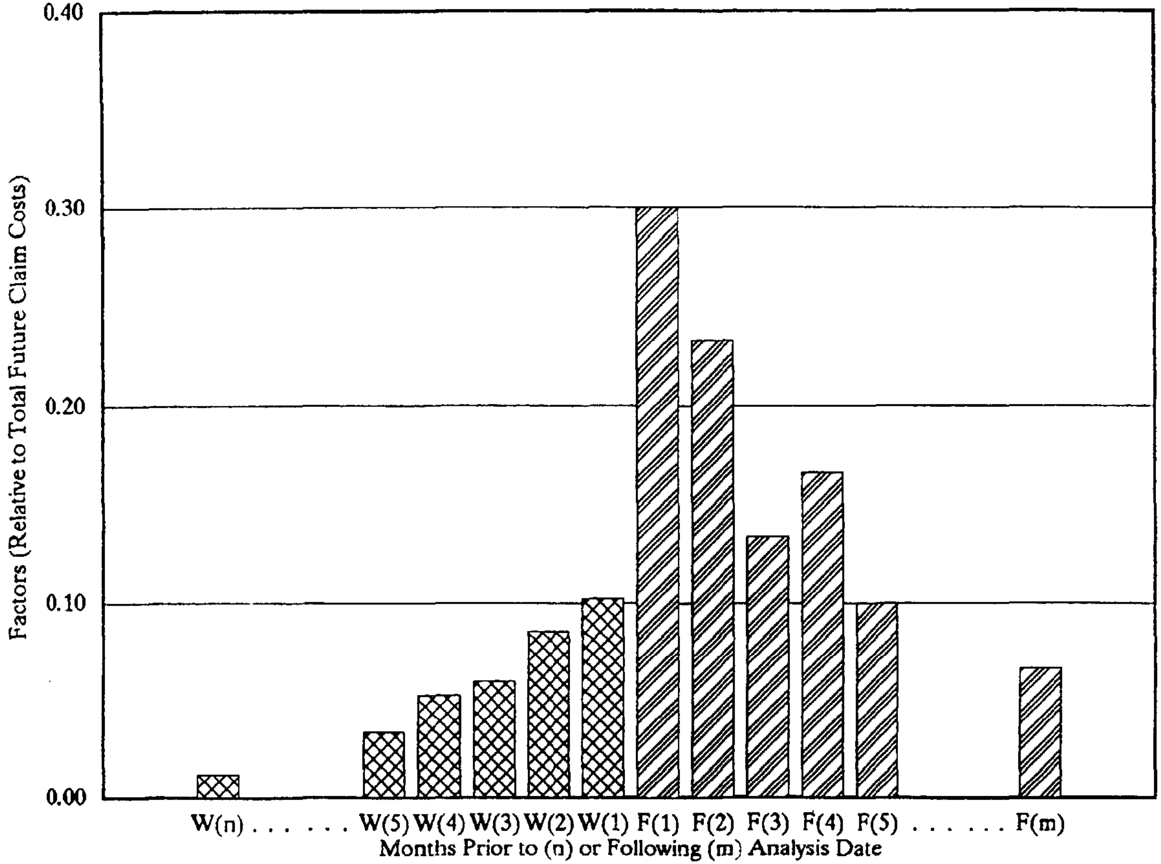
$$F^P(1) + F^P(2) + F^P(3) + \ldots = 1$$

so that

$$\mathbf{P}_i = \mathbf{P}_i \cdot F^P(1) + \mathbf{P}_i \cdot F^P(2) + \mathbf{P}_i \cdot F^P(3) + \ldots$$

If the group has **M** months left in its contracted paid claims accumulation period, then $\mathbf{P}_i(M)$ is the Covered Future Paid Claim Costs from Ongoing Claims for the ith member, where

Figure 2 – Factors for Projection of Future Claim Costs for Ongoing Large Claims



$$\mathbf{P}_i(M) = \mathbf{P}_i \cdot F^P(1) + \mathbf{P}_i \cdot F^P(2) + \mathbf{P}_i \cdot F^P(3) + \ldots + \mathbf{P}_i \cdot F^P(M).$$

The projected Future Specific Paid Claim Costs from Ongoing Claims for the ith member is that portion of $\mathbf{P}_i(M)$ which is in excess of the specific deductible (after allowing for claims already paid during the accumulation period). The total Reserve for Future Specific Paid Claim Costs from Ongoing Claims for the entire block of Paid-basis groups is then the sum of the individual Future Specific Paid Claim Costs from Ongoing Claims for all members in all groups in the block.

The calculation of Future Incurred Claim Costs for Ongoing Claims (in the case of Incurred-basis groups) would be very similar. In the calculation of Future Paid Claim Costs above, we would substitute $W^I(n)$ for $W^P(n)$, $C^I(n)$ for $C^P(n)$, $F^I(m)$ for $F^P(m)$, and $\mathbf{I}$ for $\mathbf{P}$ to indicate that all factors and amounts are on an incurred claim basis rather than on a paid claim basis, respectively. If the group has **M** months left in its contracted claims incurral period, then $\mathbf{I}_i(M)$ is the Covered Future Incurred Claim Cost for Ongoing Claims for the ith member. The projected Future Specific Incurred Claim Costs from Ongoing Claims for the ith member is that portion of $\mathbf{I}_i(M)$ which is in excess of the specific deductible (after allowing for claims already incurred and paid or incurred but not yet paid during the accumulation period). The total Reserve for Future Specific Incurred Claim Costs from Ongoing Claims for the entire block of Incurred-basis groups is then the sum of the individual Future Specific Incurred Claim Costs from Ongoing Claims for all members in all groups in the block.

For incurred-basis groups, the weighting factor $W_i(n)$ can either be calculated on the basis of claims incurred in the nth month prior to the analysis date and paid through the analysis date, or it can be calculated on the basis of estimated incurred claims which have been completed using claim lag factors applied to incurred and paid claims. The former approach has three advantages: (1) It is a simpler calculation, (2) It makes the projection of Future Claim Costs independent from the calculation of IBNR reserves, and (3) The high variance inherent in the completion of claims incurred in the last months of the analysis period can be dealt with more directly.

CALCULATION OF FACTORS FROM EXPERIENCE DATA

Scope - For the sake of simplicity, I will refer here only to the Paid-basis case. The Incurred-basis case is similar.

Simplifying Assumptions - I will assume that for some value of **N**, claims older than **N** months have no predictive value, and so are ignored, and I will also assume that for some value of **M** there are no ongoing claims paid more than **M** months after the analysis date (or at least they can be ignored).

Data - The data to be input to the analysis includes monthly paid claims for the **N** months before and **M** months after the analysis date, for a total of **N + M** month's experience. For the sake of practicality the analysis is restricted to individual members whose claims over the **N** (usually 12) months preceding the analysis date exceed some threshold amount (e.g., $10,000, $25,000, etc.).

Calculation of $[\hat{W}^P(n)]$. The independent variables are $C^P_i(n)$, the Paid Claims for the ith individual in the nth month prior to the analysis date. The dependent variable is $\mathbf{P}_i$, which is the Total Future Claims Paid for the ith individual. We are looking for the coefficients $[\hat{W}^P(n)]$ such that

$$\hat{P}_i = \sum_{n=1}^{N} [\hat{W}^P(n) \cdot C^P_i(n)]$$

is an unbiased estimator.

The Best Linear Unbiased Estimator is given by multivariate least squares regression. We wish to minimize the sum of squares

$$S = \sum_i (P_i - \hat{P}_i)^2$$

$$\text{(I)} \qquad = \sum_i \{P_i - \sum_{k=1}^{N} [\hat{W}^P(k) \cdot C^P_i(k)]\}^2$$

which gives the normal equations

$$\text{(II)} \qquad \frac{\partial S}{\partial \hat{W}^P(n)} = -2 \cdot \sum_i \left[\{P_i - \sum_{k=1}^{N} [\hat{W}^P(k) \cdot C^P_i(k)]\} \cdot C^P_i(n) \right]$$

then setting $\partial S/\partial[\hat{W}^P(n)]$ equal to zero for each n yields the set of independent equations

$$\text{(III)} \qquad \hat{W}^P(n) = \frac{\sum_i [P_i \cdot C^P_i(n)] - \sum_k^{k \neq n} \{\hat{W}^P(k) \cdot \sum_i [C^P_i(n) \cdot C^P_i(k)]\}}{\sum_i [C^P_i(n)]^2}$$

The set of equations in (III) may be resolved for all **N** elements in $[\hat{W}^P(n)]$ by the use of matrix or linear programming methods.

An alternative approach to the problem can be made by imposing the simplifying assumption that the series $[W^P(n)]$ is arithmetic. That is, for a given value of **N**, there is some **X** such that $W^P(1) = \mathbf{N}*\mathbf{X}$, $W^P(2) = (\mathbf{N}-1)*\mathbf{X}$, $W^P(3) = (\mathbf{N}-2)*\mathbf{X}$, etc., and $W^P(n) = 0$ for all n greater than **N** (See Figure 3). Then

$$
\begin{aligned}
P_i &= \sum_{n=1}^{N} [W^P(n) \cdot C^P{}_i(n)] \\
\text{(IV)} \qquad &= \sum_{n=1}^{N} [X \cdot (N+1-n) \cdot C^P{}_i(n)]
\end{aligned}
$$

Applying least squares regression to this model, we seek to minimize

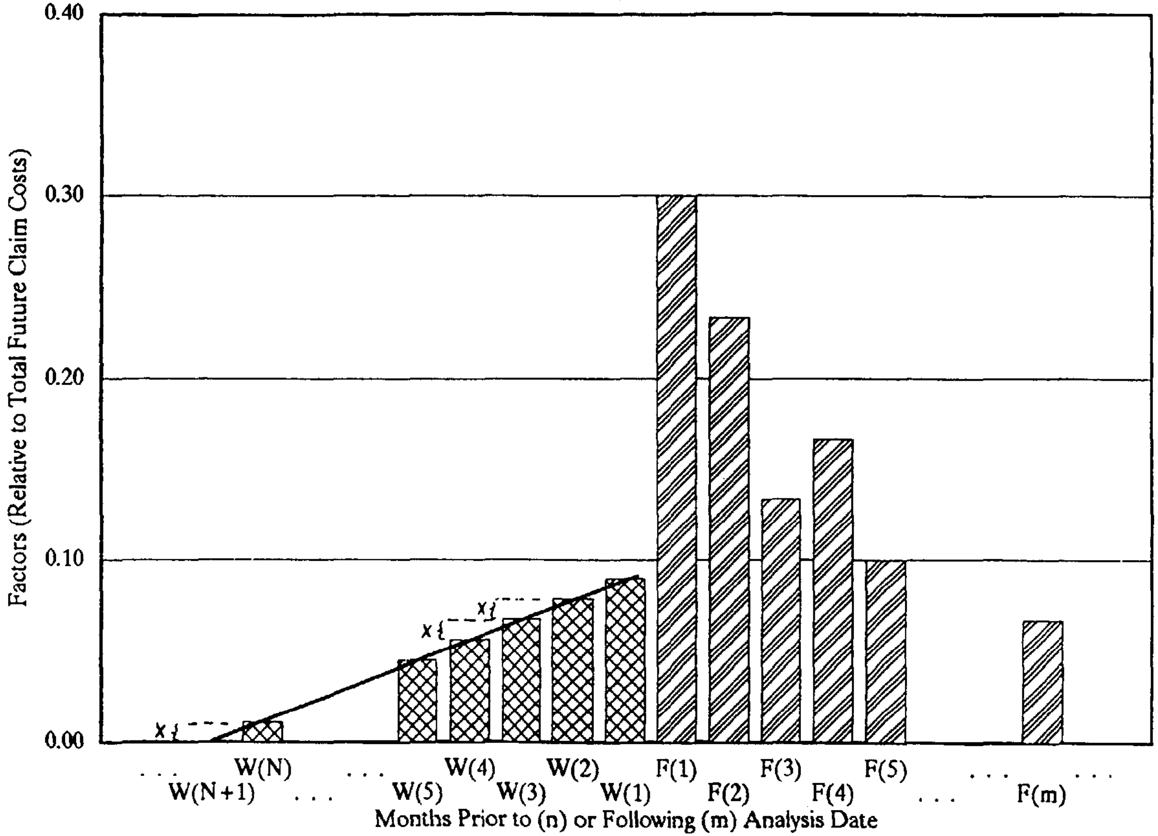
$$
\begin{aligned}
S &= \sum_{i} (P_i - \hat{P}_i)^2 \\
\text{(V)} \qquad &= \sum_{i} \{P_i - \sum_{n=1}^{N} [\hat{X} \cdot (N+1-n) \cdot C^P{}_i(n)]\}^2
\end{aligned}
$$

which gives the normal equation

$$
\text{(VI)} \quad \frac{\partial S}{\partial \hat{X}} = -2 \cdot \sum_{i} \Big[\{P_i - \sum_{n=1}^{N} [\hat{X} \cdot (N+1-n) \cdot C^P{}_i(n)]\} \cdot \sum_{n=1}^{N} \{(N+1-n) \cdot C^P{}_i(n)\} \Big]
$$

Setting $\partial S/\partial \hat{X}$ equal to zero yields

Figure 3 – Factors for Projection of Future Claim Costs for Ongoing Large Claims
Model with Linearity Imposed on [W(n)]



$$\text{(VII)} \qquad \hat{X} = \frac{\sum_{i} \{ P_i \cdot \sum_{n=1}^{N} [(N+1-n) \cdot C^P_i(n)] \}}{\sum_{i} \{ \sum_{n=1}^{N} [(N+1-n) \cdot C^P_i(n)] \}}$$

which can be readily calculated. From these values of $\hat{\mathbf{X}}$

$$\text{(VIII)} \qquad \hat{W}^P(n) = (N+1-n) \cdot \hat{X} \qquad 1 \leq n \leq N$$

There remains the problem, however of determining what is the best value of **N** to use. Since there are only a limited number of integer values of **N** which would be reasonably expected to give the best result using the above approach, a logical approach is to calculate **N** using trial-and-error. Calculate **X** using several values of **N** in the range of what would be reasonably expected to work (probably four to 12 months) and take the one which gives the best fit to the observed $\mathbf{P}_i$'s.

The estimators given by equations (VII) and (VIII) may be adequate for many purposes. However, if better estimators are required, then the results of (VII) and (VIII) may be used as initial values to calculate "better" estimators iteratively using equations (III). The new estimator of $[W^P(n)]$, $[\hat{W}^P(n)']$, would be generated by equations

$$\text{(IX)} \qquad \hat{W}^P(n)' = \frac{\sum_{i} [P_i \cdot C^P_i(n)] - \sum_{k}^{k \neq n} \{ \hat{W}^P(k) \cdot \sum_{i} [C^P_i(n) \cdot C^P_i(k)] \}}{\sum_{i} [C^P_i(n)]^2}$$

where $\hat{W}^P(k)$ is the estimator of $W^P(k)$ from equations (VII) and (VIII). Values of $\hat{W}^P(n)$ can be generated successively for n = 1, 2, ..., N. The values of $[\hat{W}^P(n)]$ generated by (VII) and (VIII) for n>1 are substituted into equation (IX) to generate a new value of $\hat{W}^P(1)$, $\hat{W}^P(1)'$. For values of k, $1<k\leq N$, the estimators $[\hat{W}^P(n)]$ generated by (VII) and (VIII) for n>k are substituted into equations (IX) along with values of $\hat{W}^P(n)'$ already generated by (IX) for n<k to generate a new estimator of $W^P(k)$, $\hat{W}^P(k)'$.

Even closer approximations to $[W^P(n)]$ can be achieved by generating $[\hat{W}^P(n)]''$ using $[W^P(n)]'$ in equations (IX), $[W^P(n)]'''$ using $[\hat{W}^P(n)]''$, etc.

Using equations (IX) to generate new estimators of $[W^P(n)]$ also has the benefit of calculating non-zero estimators of $W^P(n)$ for values of <u>n</u> greater than the **N** used in equations (VII) and (VIII).

Calculation of $[\hat{F}^P(m)]$. The calculation of an estimator for $[F^P(m)]$ given $\mathbf{P_i}$ is much simpler. For the block of data used in the analysis, $\hat{F}^P(m)$ would simply be the proportion of the amount **P** from the experience data which is paid in the month <u>m</u> following the analysis date. That is,

$$\text{(X)} \qquad \hat{F}^P(m) = C^P(m) \div P$$

Incurred Basis Estimators. The calculation of estimators $[\hat{W}^I(n)]$ and $[\hat{F}^I(m)]$ is essentially identical to the paid-basis analysis given above. The only difference is that the data for the months prior to the analysis date would be claims incurred each month and paid through the analysis date, and the data for the months following the analysis date would be claims incurred in each month. Analysis for claims incurred but not yet paid as of the analysis date should be kept separate.

Acknowledgement. Thanks to Antonetta Rodrigues of WPS, Virginia Young of the University of Wisconsin, Steve Millar, and Chuck Fuehrer for reviewing the manuscript and making helpful suggestions.