JIA 114 (1987) 371-392

CONTROL OF UNFUNDED AND PARTIALLY FUNDED SYSTEMS OF PAYMENTS

By G. C. Taylor, B.A., Ph.D., F.I.A., F.I.M.A., F.I.A.A.

SUMMARY

The paper considers systems of payments which are not fully funded, i.e partially funded or fully unfunded. Generally, the objective is to be able to establish a premium formula which is consistent with long term planning as to e.g. a target rate of funding, limited variation in premiums from year to year, etc. The premium formulas considered are those which relate premiums to prior years' premiums, claims experience and accumulated funds. These questions are reviewed in Section 1 which suggests the use of control theory. Section 2 formulates and provides a formal solution to the problem.

Section 3 provides a couple of results which relate premiums to targeting of accumulated funds.

Subsequent sections consider the construction of premium formulas. It is emphasized that the intention is not to supply a definitive set of instructions as to how the premium formula might be constructed, but rather to illustrate some of the principles relevant to it.

In particular, two numerical examples are provided in § 6.2. Premium formulas are constructed which appear to respond reasonably satisfactorily to simulated claims experience.

It is found that accurate prediction of future claims escalation is crucial to the operation of formulas of the linear control theory type (§ 6.2.1). Brief comment on possible further research aimed at dealing with this aspect of the question is given in Section 7.

1. INTRODUCTION

1.1 *Unfunded and partially funded systems of payments*

This paper considers systems which provide for payments to beneficiaries under certain conditions. These payments are supported by premiums, contributions or levies (referred to as premiums in this paper), also payable under defined conditions.

If at any point the total of all premiums paid into the system, together with any investment income earned by them, exceed the total of all payments of benefit, then a surplus will have accumulated. This will be referred to as 'the fund'.

In most of these systems, the fund will be subject to certain objectives. These objectives fall into three categories, as follows.

1.1.1. *Fully funded.* Fully funded systems are those in which the fund is estimated to be sufficient to meet its liabilities in respect of future payments of benefit even in the event that premiums are terminated forthwith.

1.1.2. *Unfunded.* Unfunded systems maintain the fund at approximately zero.

1.1.3. *Partially funded.* Partially funded systems specifically maintain a non-zero fund, though at a level inadequate for full funding.

In the following sections the term partially funded will, for brevity, be used to include both partially funded and unfunded according to the above definitions.

1.2 *Relevance of partially funded systems*

Partially funded systems have existed for many years. For example, in the United Kingdom National Insurance pensions constitute such a system. Their existence is usually found in systems of collection and payment of monies which may be regarded as social insurance.

A decision to adopt partial rather than full funding is, in effect, an endorsement of cross-subsidies between generations of participants in the system. In a fully funded system each generation contributes sufficiently to support its own benefits. In a partially funded sytem this is not necessarily so.

Although full funding of social insurance is by no means sacrosanct, it does provide a neat nexus between the volume of benefit payments and the required volume of premiums in respect of each generation of participants. If the benefits can be estimated, then the required level of premiums flows from them.

This is not the case with partially funded systems. The principles upon which premiums ought to be determined are much less clear.

Worse still is the fact that partially funded systems sometimes arise as systems which, though intended to be fully funded, have lapsed into a partially funded state as a result of government action or political expediency. While it is even possible that the conversion from full to partial funding is an appropriate course, there is a conspicuous absence of long term strategy from such action.

The combined effect of these two factors:

(i) the absence of any clear and simple principles to provide guidance on the level of premiums;
(ii) the tendency for short term expediency and *ad hoc* procedures to hold sway in partially funded systems, especially during transition from fully funded;

results in a considerable danger that partially funded systems under government control will develop in a haphazard manner. If this occurs, then such systems

neither (i) ensure that each generation of beneficiaries contributes sufficiently to meet its own costs;
nor (ii) establishes cross-subsidies between generations on any rational economic basis.

Australia provides an example where these matters have become topical in recent years. Until the mid-1970's workers compensation and third party motor insurance were conducted in the private sector and were (or, at least, were intended to be) fully funded.

In the case of third party motor insurance, Government regulation of premium rates at uneconomic levels led to the withdrawal of almost all private insurers from the field. The underwriting of this class of business then passed by default to the public sector. As this sector was not subject to the same solvency requirements as the private sector, and as regulated premiums remained uneconomic, a decline in the 'fundedness' of liability was permitted to set in. At present, the fact that this class of insurance is only partially funded is explicitly recognized by the state of New South Wales, where the Government has recently decided that full funding will no longer be attempted. There are suggestions that the state of Victoria will follow the same course.

In the workers compensation field, the Victorian Government has recently announced that all underwriting of this class of business will be carried out in the public sector in future. Although it is stated that this business will remain fully funded, the private sector has expressed considerable doubts about this. Most other states of Australia are currently considering, or have recently considered, whether to introduce workers compensation systems involving single insurers administered by government. In any case where this occurs, the question of funding of liabilities under the new system will arise.

1.3 *Control systems*

If a system of collection and payment of monies is to be conducted on a partially funded basis, it appears desirable that it be subject to some medium to long term planning as to:

(i) the rate of increase of premiums in future years;
(ii) the variability of these premiums;
(iii) the magnitude of subsidies between generations.

If a system makes a transition from full to partial funding, it will be possible, at least for a time, to maintain premiums at levels well below those required under full funding. Indeed, such a transition can occur only if premiums are set below the full funding level. However, over the longer term premiums must return to a level approximating that of full funding. Some planning of the future trajectory of premium rates is necessary if it is to be ensured that major disruptions in future premium rates are avoided.

If the fund is maintained at the level required for full funding, transfers between generations are zero. As the fund is permitted to decline from this level, transfers increase. Thus, an examination of the level of intergenerational transfers can be seen as an examination of the future course of the fund. Some planning of this aspect is also desirable.

It may be useful for administrative purposes to establish a formula according to which future premiums will be related to future levels of payments, the future size of the fund, etc.

It is suggested that something of each of the above requirements can be achieved if the premium formula is chosen by means of engineering control theory. The application of this theory to actuarial matters has already been considered in the literature (Balzer and Benjamin, 1980; Balzer, 1982). The theory is concerned essentially with some form of 'black box' whose input-output properties are determined by a number of free parameters. The problem is to choose these parameters in such a way that the output responses to various inputs are of an acceptable nature.

This may be placed in the present context as follows. The claims experience may be thought of as the system input; premiums as the output. The 'black box' is the formula according to which claims experience, fund level, etc. generate future premium rates. The free parameters associated with this black box represent the available choices of premium formula. It may be desirable, in this context, to subject the system output to certain strategic constraints, e.g. on the rate of growth of future premium rates. It will then be necessary to choose the system parameters consistently with these constraints.

1.4 *Outline*

The paper considers the situation in which future premiums are determined as a function of:

(i) claims experience;
(ii) the preceding course of the fund;
(iii) premiums in preceding years.

The response of such a system to changes in the claims experience are examined. Certain features are identified as generating favourable or unfavourable system responses. The analysis indicates:

(i) the general considerations involved in the choice of an efficient premium formula;
(ii) the manner in which the premium formula should be chosen in order to achieve certain specific system performance, e.g. maintenance of some particular target trajectory of the fund.

2. BASIC MODEL

2.1 *Formulation*

The system formulated by Balzer and Benjamin (1980) was reasonably general. It was, however, apparently intended to deal with short term business as no allowance for investment income is included.

Moreover, in the system of Balzer and Benjamin the feedback mechanism in premium rating consisted of awarding premium rebates equal to a certain proportion of the 'accumulated surplus', this latter being equal to the excess of premiums (net of expense loadings) collected over claims.

Whereas the accumulated surplus of Balzer and Benjamin relates premiums to incurred claims (i.e. whether paid or not, whether notified or not), the present paper relates premiums to claim payments, as is usual in an unfunded system. This redefinition of 'accumulated surplus' should be kept in mind in subsequent sections in which the notation of Balzer and Benjamin has been retained.

In the present context, it seems desirable to generalize their model in three directions;

(i) to introduce an allowance for investment income, since partially funded systems of payments may well involve long term business, such as Liability, for which a comparatively large fund may be maintained, and so comparatively large amounts of investment income received;
(ii) to generalize the formula according to which the premiums charged by the system are determined;
(iii) to introduce an explicit allowance for an initial value of the fund, such as would arise in the case of a fully funded system about to become partially funded.

As regards (ii), it would appear that quite a general model could be achieved by assuming that the premiums of any particular year are fixed as a function of:

(i) values of the fund;
(ii) amounts of claim payments;
(iii) premiums;

of prior years.

The third generalization described above will be effected by the introduction of external contributions to the fund.

If the transfer function form of control theory is to remain applicable to this generalized model, it will be necessary that the functional defining premiums be linear in each of the arguments (i), (ii) and (iii). Note however that, subject to this restriction, the premium formula can encompass the entire history of the fund, claims and premiums. Thus, it can include integral and derivative action terms such as are discussed in Balzer (1982).

Strictly, the specification of the values of all three of the fund, claims and premiums as arguments in the premium formula involves redundancy. Only two of these arguments need be specified. It is not difficult to show by an inductive argument that past premiums can be expressed in terms of past claims and values of the fund, and hence can be eliminated from the premium formula. However, from the point of view of mere expression of the premium formula, it may be desirable to retain all three arguments.

Figure 1 is a block diagram which summarizes the system of Balzer and Benjamin, generalized as discussed above. For convenience, the diagram is

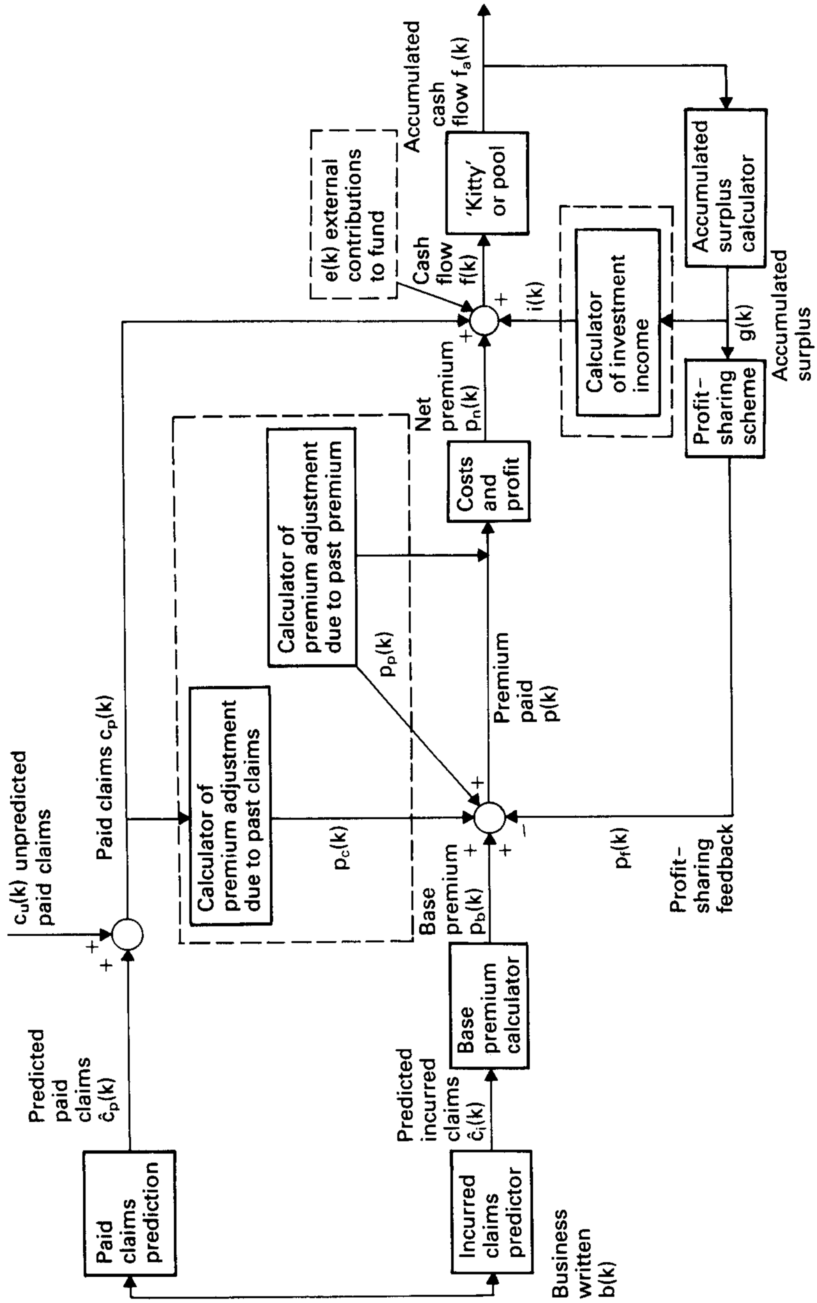
Accumulated cash flow $f_a(k)$
e(k) external contributions to fund
'Kitty' or pool
Cash flow f(k)
Accumulated surplus calculator
i(k)
Calculator of investment income
g(k)
Accumulated surplus
Profit-sharing scheme
Net premium $p_n(k)$
Calculator of premium adjustment due to past premium
Costs and profit
$p_p(k)$
Premium paid p(k)
Paid claims $c_p(k)$
$c_u(k)$ unpredicted paid claims
Calculator of premium adjustment due to past claims
$p_c(k)$
$p_f(k)$
Profit-sharing feedback
Base premium $p_b(k)$
Predicted paid claims $\hat{c}_p(k)$
Base premium calculator
Predicted incurred claims $\hat{c}_i(k)$
Paid claims prediction
Incurred claims predictor
Business written b(k)

Figure 1.

drawn so as to resemble theirs as closely as possible. Also for ease of reference, the notation used here is, wherever possible, the same as theirs.

In Figure 1 the dotted lines mark out the blocks and flows which have been added here to the Balzer-Benjamin system. Again for comparability with the original block diagram (Balzer and Benjamin, 1980), their 'profit-sharing' terminology has been maintained at the foot of the diagram. Strictly, 'accumulated surplus' should be read in that part of the diagram as 'the part of the fund in excess of any target prescribed for it'. 'Profit-sharing' refers to the calculation of a component of premium adjustment related to the history of the size of the fund. As the accumulated surplus is a known quantity, there is no need for its estimation as was the case for Balzer and Benjamin.

Note the above exclusion of the fund target from the definition of accumulated surplus, this target being treated as composed of external contributions to the fund. The investment income generated by this target component is also excluded, being treated as further external contributions to the fund, the reason for which will become apparent in Section 3. It is because of this treatment that the diagram takes 'accumulated surplus' as the pickoff point for calculation of investment income rather than taking the more 'obvious' accumulated cash flow (which does include the effect of the fund target).

With this interpretation of accumulated surplus, the accumulated surplus estimator becomes trivial, i.e. has an identity transfer function, in the case of a zero target fund.

With the introduction of three new blocks to the diagram, three new variables are introduced to the model. These are $i(k)$, $p_c(k)$ and $p_p(k)$, all marked in the diagram. Their meanings are defined in the following list of symbols which, for ease of reference, repeats the nomenclature (as far as necessary) of Balzer and Benjamin.

The following is nomenclature retained from Balzer and Benjamin (1980).

k	integer indicating financial period
$b(k)$	exposure to claims in period k
$c_i(k)$	claims incurred from business written in period k
$c_p(k)$	claims paid in period k
$c_u(k)$	unpredicted claims paid in period k
$f(k)$	cash flow for period k
$f_a(k)$	accumulated cash flow (fund) at end of period k
$g(k)$	accumulated surplus at end of period k ($=f_a(k)-f_0(k)$ with $f_0(k)$ defined below)
k_c	costs and profit factor
$p(k)$	premiums paid in period k
$p_b(k)$	base premiums for period k
$p_n(k)$	net premium income for period k
z	transform parameter (complex variable)
$\hat{x}$	predicted value of variable x

This nomenclature is supplemented by the following.

$p_f(k)$ adjustment to premiums $p(k)$ depending on past values of the fund
$p_c(k)$ adjustment to premiums $p(k)$ depending on past claims experience
$p_p(k)$ adjustment to premiums $p(k)$ depending on past movements of premiums
$f_0(k)$ 'target' value of fund at end of period k
$i(k)$ investment income earned by the fund in period k
$e(k)$ external contribution to the fund in period k
j rate of investment income (assumed constant over time)
J $1+j$.

Transfer functions are also introduced to represent the calculation of:

$p_c(k)$ from $c_p(k-1)$, $c_p(k-2)$, . . . ;
$p_p(k)$ from $p(k-1)$, $p(k-2)$, . . . ;
$p_f(k)$ from $f_a(k-1)$, $f_a(k-2)$, ...

These three transfer functions are represented by blocks in Figure 1. They are denoted by $T_c(z)$, $T_p(z)$ and $T_f(z)$ respectively.

The z-transform of a particular variable is represented by the upper case of the symbol representing that variable, e.g. the z-transform of $p_c(k)$ is $P_c(z)$.

2.2 *Solution*

The solution of this system is most easily obtained by ignoring the existence of $e(k)$ initially. For the moment, it is also assumed that $f_0(k)=0$. Hence $g(k)=f_a(k)$. This restriction will be relaxed in Section 3.

Standard block diagram reduction techniques (Dorf 1970) can then be applied to the diagram of § 2.1 as follows:

(i) elimination of the feedback loop involving T_p;
(ii) moving the point of summation of $c_p(k)$ to the summing point at its left (but not changing the point of summation of $i(k)$);
(iii) moving the summing point involving $i(k)$ to the summing point at its left;
(iv) elimination of the resulting feedback loop involving pickoff of $f_a(k)$ and the summing point at which $p_b(k)$ is input.

This procedure leaves two blocks in cascade, whence the transfer function between paid claims $c_p(k)$ and accumulated cash flow $f_a(k)$ is:

$$T_{ca}(z)=\frac{F_a(z)}{C_p(z)}=\frac{z[k_cT_c-(1-T_p)]}{(z-J)(1-T_p)-k_czT_f}, \qquad (2.2.1)$$

where the argument z has been suppressed on the right, and where $e(k)$ is still omitted from consideration.

The $e(k)$ term is now included. Note that $e(k)$ acts like a deduction from $c_p(k)$ at the summing point where $e(k)$ is introduced. However, it has no influence on the other branch of the diagram along which $c_p(k)$ flows (to the summing point involving $P_b(k)$). Thus, if $c_p(k)$ is set to zero for all k, the transfer function

$T_{ea}(z) = F_a(z)/E(z)$ will be exactly as for $T_{ca}(z)$ but with sign reversed and k_c set to zero in the numerator.

Thus,

$$F_a(z) = T_{ca}(z)C_p(z) + T_{ea}(z)E(z)$$
$$= z[(z-J)(1-T_p) - k_c z T_f]^{-1}\{[k_c T_c - (1-T_p)]C_p + (1-T_p)E\}. \quad (2.2.2)$$

The transform of premium is:

$$P = P_b + P_c + P_p + P_f$$
$$= P_b + T_c C_p + T_p P + T_f F_a,$$

whence

$$P = (1-T_p)^{-1}(P_b + T_c C_p + T_f F_a), \quad (2.2.3)$$

with F_a given by (2.2.2).

The stability properties of the accumulated fund are determined by the roots of the characteristic equation, which by (2.2.2) is:

$$(z-J)(1-T_p(z)) - k_c z T_f(z) = 0. \quad (2.2.4)$$

This equation will be considered further in Section 5. For the moment, however, it may be noted that this equation does not involve the transfer function T_c. In other words, the stability of the system is not affected directly by the manner in which premium adjustments are based upon prior claims experience. Effectively, this reflects the redundancy, pointed out in §2.1, of basing the premium adjustment formula on all three of the arguments $c_p(k)$, $p(k)$ and $f_a(k)$.

3. THE EXISTENCE OF A FUND TARGET

Consider the case in which a fund target $f_0(k)$ exists, e.g.

$$p_f(k) = k_f[f_a(k-1) - f_0(k-1)], \quad (3.1)$$

where k_f is some positive constant.

In fact, use of a formula like (3.1) involves some awkwardness as the presence of the preset value $f_0(k-1)$ introduces a nonlinearity. However, this slight difficulty can be overcome by means of a device.

Recall from §2.1 that

$$g(k) = f_a(k) - f_0(k).$$

Now regard the fund as separated into two funds, A and B. Fund A contains $f_0(k)$ at time k, and Fund B contains $g(k)$. Then Fund B is subject to the linear relation:

$$p_f(k) = k_f g(k-1), \quad (3.2)$$

which is equivalent to (3.1), and where it is assumed that all operations of the fund take place in Fund B except the holding of $f_0(k)$ and the earning of investment income thereon.

With linearity restored in (3.2), all of the discussion of Section 2 becomes applicable to Fund B, with $f_a(k)$ replaced by $g(k)$. The support of Fund A from the operations of Fund B can be dealt with in terms of the external flow $e(k)$.

Note that the amount of $f_0(k-1)$ held in Fund A at time $k-1$ will accumulate to $Jf_0(k-1)$ by time k. Thus, a flow to Fund B will be generated at time k, equal to

$$e(k)=Jf_0(k-1)-f_0(k). \tag{3.3}$$

The premium adjustment (3.1) may be generalized to other forms involving a target fund $f_0(k)$ without any essential change to the above discussion. This leads to two conclusions, as follows:

Proposition 3.1. In cases of premium adjustment formulas involving a target fund, but otherwise linear in f_a, c_p and p, performance of the system is described by the same characteristic equation as that applying to the same system with nil target fund.

This result follows immediately from the facts that:

(i) the existence of a target fund can be dealt with by means of $e(k)$ as defined in (3.3);

(ii) $E(z)$ does not appear in the characteristic equation (2.2.4).

If (3.3) is rewritten as:

$$e(k)=[j-r(k)]f_0(k-1), \tag{3.4}$$

where $r(k)=f_0(k)/f_0(k-1)-1=$ rate of expansion of target fund in period k, then the effect of the target fund on premiums can be seen.

Proposition 3.2. Assume the target fund to be nonnegative. If the rate of investment income exceeds the rate of expansion of the target fund, then required premiums decrease as the target fund increases. Conversely, if the rate of expansion of the target fund exceeds the rate of investment income, then required premiums increase as the target fund increases.

This proposition just enunciates the well known, though often forgotten, fact that, when claims escalation persistently exceeds investment return, partially funded systems require lower premiums over the long term than do fully funded systems. Too often in debates on funding and unfunding, one hears the unqualified assertion that transition from a fully funded to an unfunded system will lead to short term gains but ultimately higher premiums.

It is a simple matter to use the above reasoning to calculate the effect of the target fund on premiums. The earlier diagram of the control process shows that the cash flow $e(k)$ supplements premium receipts net of costs and profit. This flow will therefore be cancelled by an addition to gross premium $p(k)$ of $-e(k)/(1-k_c)$. This result, together with (3.4), shows that if $p(k)$ denotes the premium required when the target fund is $f_0(k)$, and $p^*(k)$ the corresponding required premium for a nil target fund, then:

$$p(k)=p^*(k)-[1-k_c]^{-1}[j-r(k)]f_0(k-1). \tag{3.5}$$

4. RELATIVE STABILITY

It is usual in control theory to regard a system as stable if its response to any finite input is necessarily always finite subsequently.

This concept is not of assistance in the present context. For example, if claims escalation is anticipated at rate r per period, then a system whose premium rate response to an isolated claims input never exceeds order $(1+r)^k$ can be regarded as satisfactory for most purposes.

This leads to the concept of relative stability. A system is said to be relatively stable of order r if its response to a finite input is necessarily bounded by some constant multiple of $(1+r)^k$.

Just as a system is stable if all roots of the characteristic equation are less than unity in modulus, so is it relatively stable of order r if all roots are less than $1+r$ in modulus.

5. SOME SIMPLE PREMIUM FORMULAS

5.1 *Feedback of fund surplus*

The simplest and most 'obvious' way of constructing a premium adjustment formula is to set each adjustment equal to a fixed proportion of the deviation of the fund from some prescribed target, e.g.

$$p_c(k)=p_p(k)=0, \tag{5.1.1}$$

and

$$p_f(k)=k_f[f_a(k-1)-f_0(k-1)]. \tag{3.1}$$

Proposition 3.1 showed that the characteristic equation in this case was exactly as if the target fund $f_0(k)=0$, i.e the characteristic equation is (2.2.4). In fact, (2.2.4) degenerates due to (5.1.1) and becomes:

$$z-J-k_c z T_f(z)=0. \tag{5.1.2}$$

Now according to (3.1), with target fund ignored, $T_f(z)=-k_f z^{-1}$, so that the characteristic equation (5.1.2) reduces to:

$$z-J+K=0, \tag{5.1.3}$$

where K denotes $k_c k_f$.

It is seen that this system is stable provided that $J-K$ is numerically less than unity.

5.2 *Delay in feedback of fund surplus*

Unfortunately, the premium adjustment (3.1) is not a practical possibility in most circumstances. There will inevitably be some delay between promulgation and implementation of premium rates for period k. Prior to this, there will probably be some delay in the completion of the accounts which determine the size of fund. For illustrative purposes, it is assumed here that these delays will total one complete financial period, so that (3.1) is replaced by

$$p_f(k)=k_f[f_a(k-2)-f_0(k-2)], \tag{5.2.1}$$

whence

$$T_f(z)=-k_f z^{-2},$$

and the characteristic equation corresponding to (5.1.3) is:

$$z-J+Kz^{-1}=0, \tag{5.2.2}$$

The roots of this are $\frac{1}{2}J\{1\pm\sqrt{(1-4K/J^2)}\}$ if $4K\leqslant J^2$, or $\frac{1}{2}J\{1\pm i(4K/J^2-1)\}$ if $4K>J^2$.

Note that larger roots of the characteristic equation are more easily obtained in the case of (5.2.2) than (5.1.3). Whereas, the root of (5.1.3) attains the value of $\frac{1}{2}J$ only if $K=\frac{1}{2}J$, the two roots of (5.2.2) become equal to $\frac{1}{2}J$ at $K=(\frac{1}{2}J)^2$. Since $\frac{1}{2}J$ will usually be less than unity, this suggests that larger roots are encountered in the case of (5.2.2) than (5.1.3) for a fixed value of K, the premium feedback control parameter. Or, to put the matter another way, instability or relative instability of the system is encountered at smaller values of the feedback parameter K.

More generally, consider the case:

$$T_f(z)=-k_f z^{-n},$$

which represents a delay of n periods in the recognition of fund surplus in premium.

The characteristic equation of the sytem is:

$$z^n-Jz^{n-1}+K=0. \tag{5.2.3}$$

It is difficult to deal with such an equation in precise numerical terms. However, application of the root locus method (Dorf 1970) demonstrates easily that all n roots of (5.2.3) diverge to infinity as K increases without limit.

Thus, whatever the value of n used, the choice of K needs to be made carefully if system (relative) stability is to be achieved. In fact, consistent with the results found above for $n=1$ and 2, Balzer and Benjamin (1980) showed empirically that, for fixed K, the instability of the system steadily increases with increasing n, i.e. increasing delay in the recognition of fund surplus in premium.

5.3 *Averaging past fund surpluses*

Another simple possibility is the averaging of past fund surpluses. The averaging process may be introduced as a means of smoothing the sequence of premium rates.

The general formula, incorporating such averaging and also the type of delay discussed in § 5.2, is:

$$T_f(z) = k_f z^{-n}(1 + z^{-1} + \ldots + z^{-m}). \tag{5.3.1}$$

The corresponding characteristic equation is:

$$z^{m+n-1}(z - J) + K(z^m + z^{m-1} + \ldots + 1) = 0. \tag{5.3.2}$$

Note that, even in the case $n = 0$, the feedback formula (5.3.1) effectively implies some delay in the recognition of past fund surpluses in premium rates. It may be expected, therefore, that increasing the averaging period m will increase instability of the system in a manner similar to that remarked at the end of § 5.2.

Indeed, application of the root locus method is again simple and demonstrates that, of the $m+n$ roots of (5.3.2), n diverge to infinity as K increases without limit and m do not. It is interesting, therefore, that the case $n = 0$ (no roots diverging to infinity) does not display quite the type of instability which might have been expected.

This very simple analysis suggests that the stability properties of the feedback system (§ 5.3.1) depend upon the delay parameter n in a manner similar to the dependency of the system of § 5.2. Thus, although the process of averaging of past surpluses appears not to have degraded the stability of the system, the same care in the choice of the feedback control parameter K is necessary as in § 5.2.

The system described by (5.3.1) is considered further in Section 6.

5.4 *Comment*

The general conclusions reached in §§ 5.2 and 5.3 are that:

(i) any instability of the system tends to increase with increasing delay between the fund surplus of a particular year and the first recognition of that surplus in premium rates (i.e. with increasing n);

(ii) for this reason, care is necessary in choosing the strength of the feedback of surplus into premium rates, this strength needing to be assigned smaller values for larger values of n;

(iii) the stability of the system is largely unaffected by which fund surpluses, other than the most recent, are recognized in the premium rate formula.

6. MORE GENERAL PREMIUM FORMULAS

6.1. *Theory*

Recall that all of the development of Section 5 was based upon (5.1.1) which specified that $p_p(k) = 0$. The characteristic equation (2.2.4) for the general process considered in Section 2 included a term related to this component of premium rate.

The purpose of the present section is to extend the analysis of the characteristic equation to the general case (2.2.4).

At this point it is perhaps worthwhile referring to the equations governing the z-transform of premium. These are equations (2.2.2) and (2.2.3). It may be seen that the z-transform P can, if necessary, be expressed as a sum of two terms, these terms being transforms of C_p and E respectively. In other words, despite the three feedback components of premium relating to past claims, premiums and accumulated funds, the premium chargeable in any given year is ultimately expressible in terms of just past claims experience and external capital injections. Changing the form of the feedback component related to past premiums, for example, is equivalent to an adjustment of the component relating to past claims experience.

A question may arise, therefore, as to the necessity for the three separate feedback components. Might the premium formula be simplied by being based upon just past claims experience, thus:

$$p(k) = p_b(k) + \sum_{m=1}^{\infty} \alpha_m c_p(k-m)$$

$$\text{(possibly plus terms in the } e(k-m)\text{)}, \qquad (6.1.1)$$

for some suitable choice of the α_m?

While such a procedure might well lead to desirable stability properties of the premium process, the values of α_m used in the feedback mechanism would appear mysterious to the uninitiated. In a case in which the formula (6.1.1) was to be administered by a Government department or public authority, its obscurity might lead to public criticism. The department or authority might prefer to achieve a similar result by means of a more readily accepted formula. The inclusion of all three of the feedback components of premium dealt with above adds some flexibility in this respect.

First consider (2.2.4) expressed in terms of the explicit polynomials representing its transfer function components, $T_p(z)$ and $T_f(z)$. Let

$$T_p(z) = \sum_{m=1}^{\infty} \pi_m z^{-m}, \qquad (6.1.2)$$

$$T_f(z) = -k_f \sum_{m=1}^{\infty} \phi_m z^{-m}, \qquad (6.1.3)$$

where k_f is a constant as in (3.1) and the two transfer functions, written as infinite series, can if necessary be restricted to polynomials by setting the relevant coefficients π_n and ϕ_m to zero.

Substitution of (6.1.2) and (6.1.3) in (2.2.4) yields:

$$z + \sum_{m=0}^{\infty} \gamma_m z^{-m} = 0, \tag{6.1.4}$$

with

$$\gamma_0 = K\phi_1 - \pi_1 - J, \tag{6.1.5}$$

$$\gamma_m = K\phi_{m+1} - \pi_{m+1} + J\pi_m, \quad m = 1, 2, \text{etc.} \tag{6.1.6}$$

The remainder of this section deals with the case in which one or more of the feedback components due to past premiums and claims is present. The intention is not to supply a definitive set of instructions as to how the total premium formula might be constructed, but rather to illustrate some of the principles relevant to it.

First consider what type of feedback components relating to past premiums and claims might be required. They will need to be linear functions of past premium and claims experience if linear control theory of the type discussed above is to remain applicable.

The basic premium formula appearing just prior to (2.2.3) includes a base premium $p_b(k)$, i.e. an initial approximation to the expected claims cost of period k. Suppose that this initial estimate was based upon a particular assumed rate of claims escalation but that, after some periods of experience, it became apparent that claims escalation was actually occurring at a different rate. It would be perfectly reasonable to adjust future premium rates in anticipation of a continuation of the observed rate of escalation.

Such an adjustment cannot be achieved in a straightforward manner by reference to past fund surpluses such as were dealt with in Section 5. Premium adjustments based upon these surpluses provide for the correction of past deviations of experience from expectations but do not anticipate the corresponding future deviations. The required adjustment is most easily obtained by direct reference to the claims experience.

Thus, one may consider a premium adjustment of the form:

$$p_c(k) = k_{c_1}\{3[c_p(k-2) - \hat{c}_p(k-2)] - 2[c_p(k-3) - \hat{c}_p(k-3)]\}, \tag{6.1.7}$$

where k_{c_1} is a positive constant. Note that the expression within the braces in (6.1.7) is a projection of the unpredicted claims paid in period k. It is based upon the experience of only periods $k-2$ and earlier, since the experience of period $k-1$ is assumed to be unavailable at the point of application of the formula (see § 5.2).

If deviations of actual from predicted claims experience are too persistent, it will be necessary to break the control process and reset the base premium using new estimates $\hat{c}_p(k)$. However, in the shorter term (6.1.7) may serve the political purpose of providing a response to such deviations which does not involve "changes to the rules of the game".

The decision to break the control process, and the adopted changes to parameters at the break, constitute an informal feedback loop which undoubtedly occurs in practice, whether fund management is prepared to admit to it or not. Any scientifically based and effective method of making it explicit is desirable. Control theorists refer to this as 'adaptive control', a very active area of study in control theory. Humans in a control loop are very good at it for simple systems and very bad at it in more complex ones. It can also be combined with 'predictive control'.

Consider now the feedback component of premium based on past premiums. It might be thought desirable to avoid unduly abrupt changes in premium between consecutive years. Thus, this component of feedback might take the form:

$$p_p(k) = k_p\,\{2[p(k-1) - p_b(k-1)] - [p(k-2) - p_b(k-2)]\}, \qquad (6.1.8)$$

where k_p is a positive constant and again the expression in braces is a projection of the unpredicted part of the premium (in the sense of unanticipated in the base premium) for period k.

Note that equations (6.1.7) and (6.1.8) do not satisfy the requirement stated above that these premium feedback components should be linear functions of just past claims experience and premiums. However, this is a trifling difficulty as the inadmissible terms, those involving $\hat{c}_p(.)$ and $p_b(.)$ may be mocked up by the 'external flow $e(.)$':

$$p_c(k) = k_{c_1}\,[3c_p(k-2) - 2c_p\,(k-3)], \qquad (6.1.9)$$

$$p_p(k) = k_p\,[2p(k-1) - p(k-2)], \qquad (6.1.10)$$

$$e(k) = k_c p_b(k) - \hat{c}_p(k). \qquad (6.1.11)$$

The system represented in the block diagram of Section 2.1 may be regarded as the superposition of two additive systems. In the first of these, claims experience is exactly as predicted, i.e. $c_u(k) = 0$, whence $p_c(k) = l_p(k) = 0$ by (6.1.7) and (6.1.8) and $f_a(k) = 0$. In the second system predicted claims are zero, i.e. $\hat{c}_p(k) = p_b(k) = 0$ and (6.1.9) and (6.1.10) hold strictly.

In the case of (6.1.10)

$$T_p(z) = k_p\,(2z^{-1} - z^{-2}), \qquad (6.1.12)$$

i.e.

$$\pi_1 = 2k_p, \qquad \pi_2 = -k_p. \qquad (6.1.13)$$

Then (6.1.4) and (6.1.5) yield:

$$\gamma_0 = K\phi_1 - 2k_p - J, \qquad (6.1.14)$$

$$\gamma_1 = K\phi_2 + k_p(2J+1), \tag{6.1.15}$$

$$\gamma_2 = K\phi_3 - k_pJ \tag{6.1.16}$$

$$\gamma_m = K\phi_{m+1},\ m=3,\ 4,\ \text{etc.} \tag{6.1.17}$$

The point of these equations is that, if a value of k_p, say, has been chosen, then values of the ϕ_m must be chosen such that:

(i) they produce a premium feedback component depending on fund surplus which has 'reasonable appearance';
(ii) they produce values of the γ_m consistent with (relative) stability of the system.

In order to simplify the checking of stability of the characteristic equation, the general form of this equation may be chosen in advance of application of formulas (6.1.14) to (6.1.17).

For example, one might choose

$$z + \sum_{m=0}^{\infty} \gamma_m z^{-m} = z(1 + az^{-1})^4. \tag{6.1.18}$$

Combination of this with (6.1.14) to (6.1.17) gives, for the case $\phi_1 = 0$ (as has been assumed above):

$$4a = -(2k_p + J) \tag{6.1.19}$$

$$K\phi_2 = 6a^2 - k_p(2J+1) \tag{6.1.20}$$

$$K\phi_3 = 4a^3 + k_pJ \tag{6.1.21}$$

$$K\phi_4 = a^4 \tag{6.1.22}$$

$$K\phi_m = 0,\ m=5,\ 6,\ \text{etc.} \tag{6.1.23}$$

If the value of k_p has been selected, the last set of equations determines a. In order to satisfy requirement (ii) above, it is necessary that a be numerically less than the rate of expansion of the target fund. The smaller the value of a the more stable the system. In addition requirement (i) above imposes the condition that ϕ_2, ϕ_3 and ϕ_4 appear 'sensible'. It is probably desirable that this sequence be monotone decreasing in absolute value. It is also necessary that ϕ_2 be positive, and that the signs of ϕ_3 and ϕ_4 be such that the sequence of ϕ_m be capable of 'reasonable' explanation. Note that (6.1.19) requires that a be negative.

6.2 *Examples*

6.2.1 *Slow premium feedback.* This example deals with the case investigated in §6.1, but with only a small feedback component relating to past premium rates: $k_p = \cdot 1$. It is assumed that investment return of 10% per annum can be obtained, i.e. $J = 1{\cdot}1$.

Then equations (6.1.19) to (6.1.23) become:

$$\begin{aligned} a &= -\cdot 325 \\ K\phi_2 &= \cdot 314, \\ K\phi_3 &= -\cdot 027, \\ K\phi_4 &= \cdot 011, \\ K\phi_m &= 0,\ m=5,\ 6,\ \text{etc.} \end{aligned}$$

For practical purpose, these values may be rounded:

$$\begin{aligned} K\phi_2 &= 30\%, \\ K\phi_3 &= K\phi_4 = 0, \end{aligned}$$

in which case the roots of the characteristic equation (given by (6.1.4) and (6.1.14) to (6.1.17)) are $\cdot 5$ and $\cdot 4 \pm \cdot 245i$. Thus, the rounded solution of the system retains the desired stability properties, with a dominant characteristic root of $\cdot 5$.

With the numerical values of the control parameters inserted in equations (6.1.3) and (6.1.8), and assuming $k_c = \cdot 8$ (costs and profits) so that $k_f = 1\cdot 25K$, the premium formula for period k becomes:

$$p(k) = p_b(k) - 125\% \times 30\%[f_a(k-2) - f_0(k-2)] \\ + 10\%\ \{2[p(k-1) - p_b(k-1)] - [p(k-2) - p_b(k-2)]\}.$$

Tables 1, 2 and 3 illustrate the development of this premium formula under stochastic variation in paid claims. The predicted claims are $\hat{c}_p(k) = \$100\text{M} \times (1\cdot 12)^k$; correspondingly the base premium is 125% of $\hat{c}_p(k)$. The target fund is $f_0(k) = \$250\text{M} \times (1\cdot 1)^k$ and so, in relative terms, is being eroded.

In Table 1, paid claims are simulated according to:

$$c_p(k) = [\$100\text{M} + \$10\text{M}\ u(k)] \times (1\cdot 12)^k,\ u(k) \sim N(0,\ 1). \qquad (6.2.1.1)$$

That is, mean claims increase in accordance with predicted claims but actual claims include a random error term with 10% coefficient of variation.

Table 1 appears in two parts. The left part displays the response of the system to an isolated shock of \$10M to claims paid in Period 1; the right part shows the response to a sequence of claim payments simulated in accordance with (6.2.1.1.).

The following features of Table 1 may be noted:

(i) the rate of decay of the transient associated with an isolated shock appears consistent with the 50% per period predicted above on the basis of the characteristic equation;

(ii) the complex roots of that equation introduce only innocuous oscillaton;

(iii) premiums increase smoothly despite the erratic fluctuation in claims; in fact, huge increases in claims, as in Periods 14 to 16, produce no particularly remarkable changes in premium;
(iv) the accumulated fund tracks the target fund reasonably efficiently.

As regards (iv) Table 1 suggests that the fund may tend persistently to lie below the target fund. In fact, however, a continuation of the simulation to Period 30 reversed this apparent tendency.

It may be noted from Table 1 that, in relative terms, the premium adjustments $p(k)-p_b(k)$ are reasonably small. This is crucial to the successful operation of the system and depends in turn on accuracy with which claims $c_p(k)$ have been predicted by $\hat{c}_p(k)$.

An alternative simulation was carried out in which claims escalation ran persistently at a higher rate than the 12% p.a. incorporated in $\hat{c}_p(k)$ and $p_b(k)$. All parameters were the same in this second simulation as in the first, except that (6.2.1.1) was replaced by:

$$c_p(k)=[\$100\text{M}+\$10\text{M } u(k)]\times(1\cdot15)^k.$$

Table 1.

	Isolated shock		Simulated claims experience					Increase over previous year	
	Premium adjustment (a)(c) ($M)	Accumulated fund (b) ($M)	Base premium ($M)	Claims paid ($M)	Adjusted premium (c) ($M)	Target fund ($M)	Accumulated fund ($M)	Claims (%)	Premium (%)
1	0	−10·5	140·0	99·5	140·0	275·0	288·1		
2	0	−11·5	156·8	139·8	156·8	302·5	301·9	+41	+12
3	3·9	−9·4	175·6	128·9	170·7	332·8	340·2	−8	+9
4	5·1	−6·0	196·7	169·2	195·9	366·0	361·2	+31	+15
5	4·2	−3·2	220·3	184·1	217·9	402·6	386·9	+9	+11
6	2·6	−1·3	246·7	213·9	248·2	442·9	409·5	+16	+14
7	1·3	−·4	276·3	238·4	282·8	487·2	437·6	+12	+14
8	·5	·0	309·5	262·4	323·2	535·9	477·2	+10	+14
9	·1	·1	346·6	274·3	367·3	589·5	545·5	+5	+14
10	−·0	·1	388·2	370·6	413·0	648·4	557·8	+35	+12
11	−·1	·1	434·8	359·2	454·2	713·3	618·0	−3	+10
12	−·0	·0	487·0	382·0	522·4	784·6	717·5	+6	+15
13	−·0	·0	545·4	391·8	586·3	863·1	870·3	+3	+12
14	−·0	·0	610·9	480·6	640·7	949·4	990·8	+23	+9
15	0	·0	684·2	590·9	683·4	1044·3	1043·6	+23	+7
16	0	·0	766·3	684·4	747·6	1148·7	1057·4	+16	+9
17	0	·0	858·3	669·9	854·9	1263·6	1177·9	−2	+14
18	0	·0	961·3	887·8	996·7	1390·0	1200·8	+33	+17
19	0	·0	1076·6	784·1	1116·2	1529·0	1435·0	−12	+12
20	0	·0	1205·8	1032·2	1281·1	1681·9	1570·9	+32	+15

(a) That is, $p(k)-p_b(k)$.
(b) That part of the accumulated fund attributable to the $10M shock.
(c) Each of the adjustment terms in (6.2.1.1) is set to zero when it involves a period prior to Period 1.

The results are shown in Table 2.

Table 2.

Simulated claims experience

Period	Base premium ($M)	Claims paid ($M)	Adjusted premium ($M)	Target fund ($M)	Accumulated fund ($M)	Increase over previous year: Claims (%)	Increase over previous year: Premium (%)
1	140·0	115·9	140·0	275·0	270·9		
2	156·8	145·4	156·8	302·5	277·1	+26	+12
3	175·6	143·9	177·1	332·8	302·5	−1	+13
4	196·7	168·9	206·5	366·0	328·8	+17	+17
5	220·3	168·3	233·5	402·6	381·0	−0	+13
6	246·7	233·2	262·3	442·9	394·7	+39	+12
7	276·3	250·4	286·2	487·2	411·7	+7	+9
8	309·5	326·0	328·0	535·9	386·1	+30	+15
9	346·6	256·3	377·7	589·5	472·8	−21	+15
10	388·2	371·7	448·8	648·4	506·7	+45	+19
11	434·8	488·1	487·6	713·3	454·6	+31	+9
12	487·0	521·3	544·6	784·6	410·3	+7	+12
13	545·4	634·4	648·7	863·1	330·2	+22	+19
14	610·9	690·3	766·1	949·4	282·1	+9	+18
15	684·2	735·9	904·7	1044·3	297·6	+7	+18
16	766·3	1064·6	1045·1	1148·7	87·7	+45	+16
17	858·3	1234·2	1172·0	1263·6	−214·7	+16	+12
18	961·3	1370·5	1394·0	1390·0	−503·9	+11	+19
19	1076·6	1363·7	1686·1	1529·0	−569·8	−1	+21
20	1205·8	1716·7	1994·6	1681·9	−753·7	+26	+18

It is seen that the control system is still exercising a strong smoothing effect on premiums. However, premium adjustments become larger (65% of the base premium by Period 20) as a result of the divergence of actual from predicted claims paid. The result is that the system does not respond sufficiently rapidly to the escalation of claim costs,and the fund not only falls below its target but goes heavily into debt.

A little further comment on this situation is given in Section 7.

6.2.2. *Rapid premium feedback.* This example is as in § 6.2.1 except that $k_p = {\cdot}45$. Then equations (6.1.19) to (6.1.22) become:

$$a = -{\cdot}5$$
$$K\phi_2 = {\cdot}06$$
$$K\phi_3 = -{\cdot}005$$
$$K\phi_4 = {\cdot}0625$$

Again this solution may be rounded. A possible choice is:

$$K\phi_2 = K\phi_4 = 6\%,$$
$$K\phi_3 = 0,$$

in which case the roots of the characteristic equation are ·5 and ·586 ± ·148i. Again the system is stable, although the dominant roots are complex. The period of oscillation is quite long (about 25 periods).

With the numerical values of the control parameters inserted in equations (6.1.3) and (6.1.8), and assuming $k_f = 1{\cdot}25K$ as in § 6.2.1, the premium formula for Period k becomes:

$$p(k) = p_b(k) - 125\% \times 6\% \ \{[f_a(k-2) - f_0(k-2)] + [f_a(k-4) - f_0(k-4)]\}$$
$$+45\% \ \{2[p(k-1) - p_b(k-1)] - [p(k-2) - p_b(k-2)]\}. \quad (6.2.2.1)$$

This formula was applied to the first simulated claims experience described in § 6.2.1 (Table 3).

Not surprisingly, premium (6.2.2.1) seems to apply stronger smoothing to the claims experience. Presumably, this is the case because of the larger coefficient (45%) associated with the 'premium extrapolation' component of the formula.

Correspondingly, the settling time of the system after a single shock is comparatively long, as the first two columns of Table 3 demonstrate.

Formula (6.2.2.1) was also applied to the second simulation of claims experience described in § 6.2.1. The conclusions to be drawn from that were similar to those of § 6.2.1, and so the numerical detail is not reproduced here.

Table 3.

	Isolated shock		Simulated claims experience						
Period	Premium adjustment ($M)	Accumulated fund ($M)	Base premium ($M)	Claims paid ($M)	Adjusted premium ($M)	Target fund ($M)	Accumulated fund ($M)	Increase over previous year: Claims (%)	Increase over previous year: Premium (%)
1	0	−10·5	140·0	99·5	140·0	275·0	288·1		
2	0	−11·5	156·8	139·8	156·8	302·5	301·9	+41	+12
3	·8	−12·0	175·6	128·9	174·6	332·8	343·5	−8	+11
4	1·6	−11·9	196·7	169·2	195·9	366·0	364·7	+31	+12
5	2·8	−10·8	220·3	184·1	218·2	402·6	391·2	+9	+11
6	3·5	−8·9	246·7	213·9	245·4	442·9	411·8	+16	+13
7	3·7	−6·8	276·3	238·4	276·1	487·2	434·5	+12	+13
8	3·3	−4·7	309·5	262·4	312·3	535·9	464·8	+10	+13
9	2·6	−3·0	346·6	274·3	354·1	589·5	520·7	+5	+13
10	1·9	−1·7	388·2	370·6	401·4	648·4	520·8	+35	+13
11	1·3	−·8	434·8	359·2	452·4	713·3	575·7	−3	+13
12	·8	−·2	487·0	382·0	511·8	784·6	662·1	+6	+13
13	·4	·1	545·4	391·8	575·3	863·1	800·1	+3	+12
14	·2	·2	610·9	480·6	645·4	949·4	917·6	+23	+12
15	·0	·3	684·2	590·9	716·9	1044·3	991·1	+23	+11
16	−·1	·3	766·3	684·4	791·7	1148·7	1036·8	+16	+10
17	−·1	·2	858·3	669·9	875·2	1263·6	1172·1	−2	+11
18	−·1	·1	961·3	887·8	975·8	1390·0	1176·9	+33	+12
19	−·1	·1	1076·6	784·1	1092·9	1529·0	1389·3	−12	+12
20	−·1	·1	1205·8	1032·2	1238·3	1681·9	1484·7	+32	+13

7. FURTHER RESEARCH

It may be remarked that the classical control theory used in this paper, in common with that in the papers of Benjamin and Balzer, is many years old. It is possible that the more powerful modern control theory, developed more recently, could be used to advantage on insurance systems.

Paragraph 6.2.1 mentioned the necessity of accurate prediction of claims trends if the premium feedback mechanisms discussed in this paper are to be made to work. What is required, of course, is long term accuracy. If, for example, underlying claims escalation were to fluctuate widely about a central value of 12% p.a., even with prolonged periods above or below this central value, then presumably a premium formula incorporating a base premium which includes assumed constant escalation of 12% p.a. would operate reasonably efficiently.

What needs to be avoided is the situation in which the long term central value of claims escalation differs from that predicted.

Such a difficulty may be dealt with by *ad hoc* methods. Estimated claims escalation could be monitored, and the premium formula changed when experience appeared to have departed sufficiently from the expectations implicit in the formula.

However, each time such a change in premium formula was effected, it would be necessary to assess not only the future performance of the new formula but also the smoothness of its junction with the old one. Moreover, there might be some difficulty in public justification of the details of the changed formula.

Ad hoc changes could be avoided if the premium formula incorporated a mechanism for estimation of past claims escalation and feedback of this component of claims cost. Such an approach has not been pursued here as it would almost certainly lead to a non-linear premium formula. This would require a departure from the linear control theory which has served as the theoretical background to this paper. Nevertheless, such non-linear formulas might well be of practical use.

REFERENCES

BALZER, L.A. (1982) Control of insurance systems with delayed profit/loss sharing feedback and persisting unpredicted claims. *J.I.A.* **109**, 285–316.

BALZER, L.A. and BENJAMIN, S. (1980) Dynamic response of insurance systems with delayed profit/loss sharing feedback to isolated unpredicted claims. *J.I.A.* **107**, 513–28.

DORF, R.C. (1970) *Modern control systems.* Addison-Wesley Publishing Company, Inc.