JIA 109 (1982) 285-316

CONTROL OF INSURANCE SYSTEMS WITH DELAYED PROFIT/LOSS SHARING FEEDBACK AND PERSISTING UNPREDICTED CLAIMS

By L. A. Balzer, B.E., B.Sc., Ph.D., A.F.A.I.M., A.F.I.M.A.

WITH AN INTRODUCTION
By S. Benjamin, M.A., F.I.A., F.I.S., A.S.A., F.B.C.S.

A PERSONAL JOURNEY

In 1966, I presented a paper to the Institute entitled, "Putting Computers on to Actuarial Work".[1] The order of the words in the title was significant. What I considered to be the aspect of the subject which contained the most potential, paragraphs 25 to 29, received little mention in the discussion. One speaker was not sure if that section of the paper was "profound" or "obvious".

At that stage, I had already used the machine *Atlas* for some years; it was as large and powerful as many of the modern computers. Programs for routine actuarial calculations and software to write programs was well-established in the office and my colleague L. M. Eagles and I had moved heavily into the area of simulation work covering a range of subjects from non-life stop-loss to complicated reversions.[2]

One day Mr Eagles brought back in the taxi (our data-link to *Atlas*) a particularly thick bundle of output from a simulation program, which we then laid out along the corridor so that we could walk along it in order to survey the results. A worry had already started to formulate in my mind and the realization that we could easily fill, and walk along, miles of corridors confirmed my worry. The point is that simulations produce a set of future histories. We might simulate very many future possible experiences of a life office. Suppose in each case we carry out annual valuations and distributions of surplus. In practice, we shall live through only one realization of those simulations and we shall determine the valuation basis from time to time according as the single experience unfolds. To carry out lots of simulations of future experience without changing the valuation in each simulation according to its own development will not illustrate our control of surplus in any useful way at all—to ourselves or anyone else.

The consequence of this point is enormous. The whole nature of the problem and the focus of our attention changes. We could invite an experienced actuary to look at the state of the simulated company each 'year' and the simulated experience in order to tell us the valuation basis to use. To carry out a thousand simulations of a 10-year projection would not be practicable. We would need to formalize the process by which the experience was used to determine the valuation basis in order to program it into the simulation. Our attention is now focused on the problem of formalizing and modelling that process. Without inserting that decision process into the simulations themselves, there is little point in simulating.

From the point of view of the working actuary, there is little point in simulating a situation over which he has no control. In practice, a company can control many items such as volume of new business, mix of new business, investment matching, bonus declarations, etc. In the simpler example, the actuary can control the emergence of surplus through his valuation basis.

At the time of the 1966 paper my own thoughts had not progressed much further. The book by Jay W. Forrester, *Industrial Dynamics*,[3] unsettled me and confirmed my worries but did not take me any further.

At about that time, Colin Stewart at the G.A.D. was producing demographic projections for the United Kingdom for several decades ahead. I asked him what his basis for the period, e.g. 25–50 years ahead, would be if the experience for the first 25 years were to follow his basis and he were looking back from that point to determine his basis; would it be the same as the basis he was using for the second 25-year period? He replied that he would probably look at the American experience at that time because it tended to precede that of the U.K! His answer 'threw' me.

The first concrete example of what I had in mind came in a paper by Hilary Seal[4] which I had the good fortune to hear him deliver to a conference on simulation organized by the Research Committee of the Society of Actuaries. His example was only part of his paper and almost a subsidiary part at that. I was much more excited than anyone else around—including Hilary himself, who seemed almost embarrassed by my enthusiasm.

He had set up a very simple model of the total annual claims of a motor insurer. He then set up several 'rules' for setting the premium rates based on recent experience. All of them were intuitively reasonable—like taking the average of recent years (in the days before inflation). Some of the rules had been suggested by other actuaries. The simulated insurance companies went broke.

This, to me, was the first actuarial paper which explored the decision procedure itself, i.e. the procedure which used the experience to determine and change the control variable. The several 'rules' were examples of decision procedures—decision algorithms incorporated into simulations. Even John Ryder,[5] who wrote about adaptive control in actuarial work, seemed to miss the significance of Hilary Seal's paper when he criticized it.

In the last few years, part of my work has involved giving advice on the control of a certain type of non-life portfolio. As is usual in non-life insurance, the work had to be started with inadequate historical data. My own thoughts were to control the emergence of surplus by a mechanism which resembled the algebra of profit-sharing in a group-life scheme, and a great deal of simulation work was carried out in order to find a simple working solution. One element built into the model was the effect of incorrectly forecasting next year's rate of inflation.

During the course of a long irregular private correspondence with Dr Harold Bohman of Sweden, I learned that he was thinking along similar lines in order to control a whole non-life office. He also published a paper[6] similar (from my point of view) to Hilary Seal's paper. He postulated, and simulated, two insurance companies which had the same claims experience. One company knew the type of claims distribution but not the value of the parameters; the other did not know the claims distribution. Hence, their methods of analysis of the experience, as it emerged, differed, and hence their premium rates differed.

However, there was one aspect of the data which Harold Bohman was ignoring, but which was giving me difficulties even in the most simplified model of my problem. It was the delay in the information which was inherent in the situation. As Brian Hey puts it cogently, "it is all very well trying to forecast the next few years, but we don't even know what has happened in the last few years; we're still having to forecast that".

The difficulty is explained in the joint paper by Les Balzer and myself.[7] Very briefly:

Let P_t^x = basic premium for year t
P_t = premium actually charged for year t
C_t = claims incurred for year t
G_t = accumulated surplus to date (or solvency margin)
$= G_{t-1}+P_t-C_t$.

Now suppose we control surplus by the simple adjustment to the basic premium as follows:

$$P_t = P_t^x - h\,G_{t-1}$$

where h is a constant. We can explore this model. However, in practice, at the time we quote P_t we shall not know C_{t-1} and, hence, not know G_{t-1}; we may not even know C_{t-2}; but if we substitute hG_{t-2} or hG_{t-3} the nature of the system seems to undergo a change; in particular, it induces oscillations into the system. Even though one can see how the algebra leads to oscillation, it is puzzling and makes it much harder to interpret the results or to know what to do with them.

Without realizing it, I lacked a formal conceptual background against which to formulate and explore my ideas and problems. So when the occasion arose, by chance, to be presented with just such a conceptual framework, I was primed ready for it.

During a visit back to Cambridge, I took part in a guided tour of the Cambridge University Computer Control Laboratory. As I started to ask more and more questions about the various demonstrations, I found that someone standing behind me was starting to answer them and we gradually drifted off into a corner together. I discovered he was not a visitor; he was a control engineer.

During the next two hours—until my wife tracked me down—we talked, steadily establishing intellectual contact across our separate disciplines. To me, it was as if a search-light was being shone on my problems from a direction I did not know existed.

"Yes," he said, "with delay, you will induce oscillations. Indeed, if you increase the delay the system will go unstable. It sounds to me", he said, "as if you are using a proportional control system. The 'symptoms' you are describing", he said, "can be deduced by algebraic inspection of the transfer function you are using. The analysis will tell you how the system will respond to different types of input signal and we would expect to improve the properties of the system by changing the transfer function, using recognized procedures."

During the conversation, I made an important (to me) mental leap. Normally, if we thought in those terms at all, we would think of premiums as input and claims as output. If I reversed this, and thought of the claims as input signals and the premiums as output from the system, then I could achieve a direct analogy with an engineering system and start to relate the other concepts across the analogy.

Because Les Balzer was due to return to Australia in a few months we worked fast to establish at least a first clear stage of re-stating the insurance problem in terms of control theory and of commenting on it, using orthodox control theory. By the time his early drafts were being written up I had covered the reading he had recommended and had described in some detail. It was sufficient to check his work passively and to comment where the correspondence with actuarial or insurance concepts seemed faulty. A problem for the ordinary reader is that most of the engineering textbooks deal with continuous systems via Laplace transforms.

The work was full of surprises. The intriguing idea that the old actuarial finite difference operate E could be replaced by z, and then usefully treated as a complex variable: The idea that the feedback control could be isolated from the larger insurance context: The idea of a transfer function which was separate from the input and from the output: The idea that a transfer function could be designed: The idea that rules for designing transfer functions existed: The idea of instability—my familiarity with 'hunting' in a radar aerial had been useful in early discussion: The idea of standard test input signals to examine the response: The idea that the separate properties of the response could be described usefully: The idea that some properties were mutually exclusive: The ideas of a transient response and an ultimate steady state response: The idea that the method I was using was well-known as a proportional controller: The idea that one would expect the addition of a derivative and/or integral controller to be an improvement. These all came at the time of the first joint paper.

In this his second (solo) paper, and in the private correspondence between us, the extra surprises were: The analysis showing that the first design led to a steady state position which was not zero. The introduction in the second design of an item—the accumulated accumulated surplus—to which there was no corresponding financial concept. Also, there was the cold shock of being told that a lot of my own 'workplay' during a long lull in our correspondence was not the first direction in which to progress even though one would think so at first sight.

Whilst Les Balzer was absorbing various concepts such as the difference between paid claims and incurred claims, I was absorbing the difference between an estimate made in year t of claims in year $t+1$ and an estimate made in year $t-1$ of claims in year $t+1$ when setting up an equation using the z-transform. Having fumbled my way through that on my own, I started to analyse the properties of different estimators. Les Balzer said that was not the best direction in which to start.

In his second paper, Les Balzer looks at the response to a 'step' input signal. In his paper to the Students' Society,[8] A. S. Clarke looked at the response of a simple reversionary bonus control system to a step change in the earned rate of interest. The earned rate of interest was the input signal; the changing bonus rate was the output signal; the valuation was the transfer function. Like the control engineers he examined the response as a change from a steady state initial position.

When we value a pension fund we take the input signals from the recent experience, pass them through the valuation as a transfer function to produce the output signal, the recommended contribution rate. Control theory tends to examine multiple input/output systems by matrix methods which, on his advice, are not used at all in Les Balzer's two notes. A simple examination of a pension fund control system can be started by using a change in the real rate of interest as the (single) input signal. The transfer function would be the way in which the real rate of interest was used to set the valuation rate of interest, together with the method of funding used. In a simplified model, the recommended contribution rate would be the single output signal. Even in a simple example, the transfer function seems to be non-linear and hence harder to deal with. (Brealey & Hodges, in the backing papers to their Appendix to the recent Scott Report[9]—*J.I.A.* paper by Edward Johnston,[10] discussed by S. Benjamin and David Wilkie—incorporate an algebraic rule for deriving market expectations about the future from the data of the past. I have not examined whether their model is reasonable or crucial to their results.)

Economists have entered the field, describing economic systems with control theory methods and concepts, mostly using matrix methods. Much of the work appears to be enjoyable mathematics which is difficult to follow. The first book on the subject (I believe) is in the Institute Library. It is *The Mechanism of Economic Systems* by Arnold Tustin, Professor of Electrical Engineering, Birmingham University, subtitled *An Approach to the Problem of Economic Stabilization from the Point of View of Control-system Engineering*.[11] The book is not very mathematical.

The standard control theory books mostly deal with test input signals of four types; the names are self-explanatory: spike, step, ramp and sine wave. There is a gap for the ordinary reader in the treatment of random input or noise. Random input must be described by its statistical properties, and the autocovariance function is the main means of description. At the moment of writing I have just succeeded in reproducing my numerical simulation, the algebraic result which describes, very succinctly, the relationship between the input and the output in such situations.

In July 1981, Dr Anders Martin-Lof kindly sent me a shortened version in English of a report he had written on "Dynamic Control of an Insurance Business". I believe he is a control engineer and was invited to write this report by the Swedish insurance industry. I also believe I can see the influence of Harold Bohman in the paper. The model follows closely the 1-year accounting system of an insurance company. It tackles the problem of a minimum solvency margin but it seems to ignore the effect of the delay in the claims information. At one point the author remarks (but not in these words) that the use of earned premiums (instead of written premiums) causes the oscillatory nature of the output. I understand the paper is for publication in the *Scandinavian Actuarial Journal*.

Exposure to control theory changes one's way of thinking. It is an example of the type of scientific revolution described by Bill Jewell in his paper[12] to the 1980 International Congress which he presented again at the Institute in May 1981, at a special meeting organized by the Research Committee.

I do not know if we can systematize the design of our actuarial control systems for financial institutions, but the idea, set in a context of fully developed analogous examples, seems to place us in a different conceptual universe.

REFERENCES

(1) Benjamin, S. (1966) Putting Computers on to Actuarial Work. *J.I.A.* **92,** 134.

(2) Benjamin, S. (1963) Computers and Actuarial Science. *J.S.S.* **17,** 93.

(3) Forrester, J. W. (1961) *Industrial Dynamics*. M.I.T. Press.

(4) Seal, H. L. (1970) Simulation of the Ruin Potential of Non-life Insurance Companies. *T.S.A.* **21,** 563.

(5) Ryder, J. (1976) Subjectivism: a Reply in Defence of Classic Actuarial Methods. *J.I.A.* **103,** 59.

(6) Bohman, H. (1979) Insurance Economics. *Scand. Act. J.* pp. 57–74.

(7) Balzer, L. A. & Benjamin, S. (1980) Dynamic Response of Insurance Systems with Delayed Profit/Loss Feedback to Isolated Unpredicted Claims. *J.I.A.* **107,** 513.

(8) Clarke, A. S. (1961) The Rhythm of Life Office Surplus and Bonus Distribution Under Modern Conditions. *J.S.S.* **16,** 325.

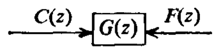
(9) Brealey, R. A. & Hodges, S. D. (1981) An Approach to Allowing for Uncertainty in Valuing Pension Schemes. Appendix 7, in *Inquiry into the Value of Pensions: Report*. (Chairman: Sir Bernard Scott.) Cmnd. 8147. H.M.S.O.

(10) JOHNSTON, E. *J.I.A.* **109**, 16 and 18.
(11) TUSTIN, A. (1953) *The Mechanism of Economic Systems: An Approach to the Problem of Economic Stabilization from the Point of View of Control-system Engineering*. Heinemann.
(12) JEWELL, W. (1980) Topic 1. Generalized Models of the Insurance Business (Life and Non-life Insurance). *Trans. I.C.A. (21st)* **S**, 87.

APPENDIX

The following explanatory appendix should have been incorporated in our first paper. The technique of 'block diagram reduction' is used in several places—first mentioned on p. 523, line 1 of the first paper. For our readers, the following brief explanation would be very helpful.

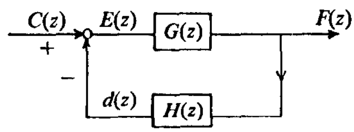
Suppose we have a system where the input $C(z)$ is transformed to output $F(z)$



i.e.

$$F(z) = G(z)C(z)$$

and we introduce a feedback loop as follows



We can reduce this to the earlier form. Let $e(z)$ and $d(z)$ be as indicated. Then

$$D(z) = H(z)F(z)$$
$$E(z) = C(z) - D(z)$$
$$F(z) = G(z)E(z)$$

Eliminating $d(z)$ and $e(z)$ we have

$$F(z) = \left[\frac{G(z)}{1+G(z)H(z)}\right] C(z)$$

Hence

$$G'(z) = \frac{G(z)}{1+G(z)H(z)}$$

is the equivalent transfer function.

SUMMARY

The insurance system with delayed profit/loss sharing feedback introduced by Balzer & Benjamin (1980) is subjected to further analyses which give greater insight into its dynamic behaviour. The steady state response of accumulated cash flow to a demanding disturbance consisting of a persisting stream of unpredicted claims is shown to be non-zero. This means that a persisting surplus or deficit, which cannot be distributed or recovered, occurs. Following a discussion of the stability of the system, the transient response is investigated using the control theoretic technique of root locus. Based on this analysis, a

profit sharing distribution of 31·25% (for a 20% cost and profit margin) is recommended in place of the intuitively appealing figure of 50%. The concept of integral action is introduced and shown to eliminate the steady state surplus/deficit noted above. A value of 2·82% for the amount of integral action is shown to produce the fastest possible response with no oscillation or overshoot. Finally the addition of so-called derivative action is shown to degrade rather than to improve the dynamic response of the system.

1. INTRODUCTION

In an earlier introductory paper (*J.I.A.* **107,** 513), Balzer and Benjamin (1980) presented a model of an insurance system with delayed profit/loss sharing feedback from insurer to insured. The system was considered from a general control and dynamic systems theory viewpoint. Certain fundamental concepts were introduced and the dynamic responses of cash flow $f(k)$ and accumulated cash flow $f_a(k)$ to an isolated group of unpredicted claims presented. This paper takes the analysis further in a number of important ways.

The general structure and details of the model of the insurance system can be found in the earlier paper.

2. NOMENCLATURE

The nomenclature used here is that used by Balzer and Benjamin (1980) in the previous paper together with the following additions:

$c_1, c_2, \ldots$	Constants
j	$\sqrt{-1}$
k_d	Amount of derivative action
k_i	Amount of integral action
N	Number of paths to infinity
n_p	Number of finite poles
n_z	Number of finite zeros
$p_1, p_2, \ldots$	Poles of a transfer function
$P(z)$	Function of interest; see equation (17)
s	Laplace transform parameter (complex variable)
T_p	Period of oscillation
T_{s2}	2% settling time
ϕ_a	Angle between asymptote and real axis
σ	Real part of s
σ_a	Asymptote centroid
ω	Imaginary part of s
$\lvert\ \rvert$	Magnitude of complex function or variable
$\measuredangle$	Angle of complex function or variable
$^\circ$	Degrees

3. STEADY STATE RESPONSE TO PERSISTING UNPREDICTED CLAIMS

In the previous paper the response to an isolated group of unpredicted claims occurring at the initial time instant $k=0$ was considered. It was shown that for $l=2$ (a delay of two periods) or more, the response was oscillatory. For $l=5$ or greater the system became unstable with the accumulated cash flow diverging to plus or minus infinity.

An obvious question relates to what happens when a certain level of unpredicted claims persists over a longer period of time. In other words, what happens if the predicted claims are consistently under- or over-estimated? Diagrammatically, the unpredicted claims $c_u(k)$ might appear as in Figure 1. The control systems theorist would describe this as a unit step input at time zero and represent it mathematically by

$$c_u(k) = \begin{cases} 0, k < 0 \\ 1, k \geqslant 0 \end{cases}. \tag{1}$$

The unit of measurement can be any convenient quantity, say thousands or millions of pounds.

In the previous paper the z-transform of a sequence of numbers was introduced and was shown to be a powerful tool for dynamic systems analysis. The z-transform, $C_u(z)$, of the input sequence, $c_u(k)$, of unpredicted claims is

$$C_u(z) = \frac{z}{z-1}. \tag{2}$$

Also the transfer function relating the z-transform, $F(z)$, of the cash flow, $f(k)$, was shown by Balzer & Benjamin (1980) to be

$$\frac{F(z)}{C_u(z)} = -\frac{z^{l-1}(z-1)}{z^l - z^{l-1} + k_c k_p}. \tag{3}$$

Consequently, from equations (2) and (3)

$$F(z) = -\frac{z^{l-1}(z-1)}{z^l - z^{l-1} + k_c k_p} \times \frac{z}{z-1}$$

$$= -\frac{z^l}{z^l - z^{l-1} + k_c k_p}. \tag{4}$$

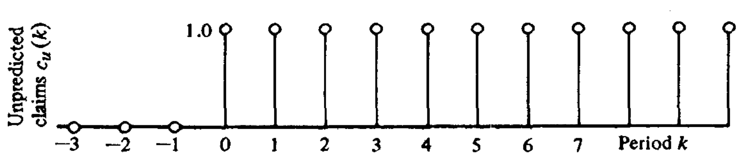


Figure 1. Persisting stream of unpredicted claims.

Similarly, the transfer function relating the z-transform, $F_a(z)$, of the accumulated cash flow, $f_a(k)$, to $C_u(z)$ is

$$\frac{F_a(z)}{C_u(z)} = \frac{-z^l}{z^l - z^{l-1} + k_c k_p} \tag{5}$$

Hence

$$F_a(z) = -\frac{z^{l+1}}{(z^l - z^{l-1} + k_c k_p)(z-1)}.$$

Important conclusions about the steady state values, that is the values which result in the long term as the period k approaches infinity, can now be drawn. Provided that the system is stable (discussed later), the Final Value Theorem leads to steady state values

$$f(\infty) = \lim_{k \to \infty} f(k) = \lim_{z \to 1} (z-1) F(z) = 0 \tag{6}$$

and

$$f_a(\infty) = \lim_{k \to \infty} f_a(k) = \lim_{z \to 1} (z-1) F_a(z) = \frac{-1}{k_c k_p}. \tag{7}$$

Equation (6) implies that, provided the system is stable, the cash flow in any individual period will eventually settle down to zero, which is obviously desirable. However, equation (7) indicates that for a stable system, the accumulated cash flow will not settle down to zero but will reach a steady state value of $-1/k_c k_p$. For the realistic values $k_c = \cdot 8$ and $k_p = \cdot 5$ used in the previous paper, $f_a(\infty) = -2{\cdot}5$. Hence if unpredicted claims amount to £1 million per period over a long length of time, the accumulated cash flow will settle down to a steady state deficit of £2·5 million. This seems highly undesirable.

Two remedies are possible. First, on seeing this situation establishing itself, management would probably begin to adjust both the Paid Claims Predictor and the Incurred Claims Predictor discussed by Balzer & Benjamin (1980). Whilst this *ad hoc* adjustment is a perfectly natural response, the control theorist would claim that a more sophisticated profit sharing scheme would eliminate the need for any such *ad hoc* action. This second control theoretic approach has a number of virtues. It is automatic in that a management decision is not required to activate it. Also, it will begin taking action immediately the estimated accumulated surplus, $\hat{g}(k)$, starts to drift away from zero. To be more specific, the control theorist would state that the addition of integral action (see later) is required. Perhaps the ideal solution is a combination of both approaches.

4. TRANSIENT RESPONSE TO PERSISTING UNPREDICTED CLAIMS

In the preceding section, the steady state response to a persisting or continuing stream of unpredicted claims was presented. Attention is now directed to the transient portion of the response. Balzer & Benjamin (1980) used the techniques of inverse z-transformation and direct division to obtain closed form analytic expressions and numerical results for dynamic transient responses. In this paper an alternative method, which is particularly convenient for digital computer or programmable pocket calculator use, is presented.

From the transfer function of equation (3)

$$(z^l - z^{l-1} + k_c k_p)\, F(z) = -z^{l-1}(z-1)\, C_u(z).$$

Dividing through by z^l gives

$$(1 - z^{-1} + k_c k_p z^{-l})\, F(z) = -(1 - z^{-1})\, C_u(z).$$

Remembering that the backward shift operator z^{-1} is equivalent to a time delay of one period, inverse z-transformation leads to the difference equation

$$f(k) - f(k-1) + k_c k_p f(k-l) = -c_u(k) + c_u(k-1).$$

A recurrence relation which can be used to calculate successive values of $f(k)$ is then

$$f(k) = f(k-1) - k_c k_p f(k-l) - c_u(k) + c_u(k-1). \tag{8}$$

Consider an example of the use of equation (8). Let $k_c = {\cdot}8$ and $k_p = {\cdot}5$, as in the earlier paper, and let the unpredicted claims be a unit step as in equation (1). For a delay of $l = 2$ time periods

$$\begin{aligned} f(0) &= f(-1) - {\cdot}4f(-2) - c_u(0) + c_u(-1) \\ &= 0 - 0 - 1 + 0 = -1 \\ f(1) &= f(0) - {\cdot}4f(-1) - c_u(1) + c_u(0) \\ &= -1 - 0 - 1 + 1 = -1 \\ f(2) &= f(1) - {\cdot}4f(0) - c_u(2) + c_u(1) \\ &= -1 + {\cdot}4 - 1 + 1 = -{\cdot}6 \end{aligned}$$

and so on for larger values of k.

Similarly, the accumulated cash flow is governed by

$$(z^l - z^{l-1} + k_c k_p)\, F_a(z) = -z^l C_u(z)$$

whence

$$(1 - z^{-1} + k_c k_p\, z^{-l})\, F_a(z) = -C_u(z)$$

which leads to the recurrence relationship

$$f_a(k) = f_a(k-1) - k_c k_p f_a(k-l) - c_u(k). \tag{9}$$

The transient responses for $k_c = {\cdot}8$, $k_p = {\cdot}5$ and delay periods of $l = 1$, 2 and 5 are shown in Figures 2 and 3.

In Figure 3, the responses for $l = 1$, 2 and 5 are superimposed for convenience. The broken lines have been added for the sole purpose of making the resulting figure more intelligible. They do *not* indicate the behaviour of $f_a(k)$ between any

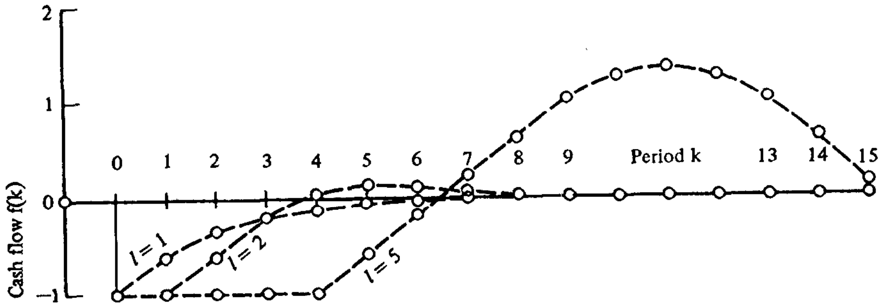


Figure 2. Effect of persisting unpredicted claims on cash flow.

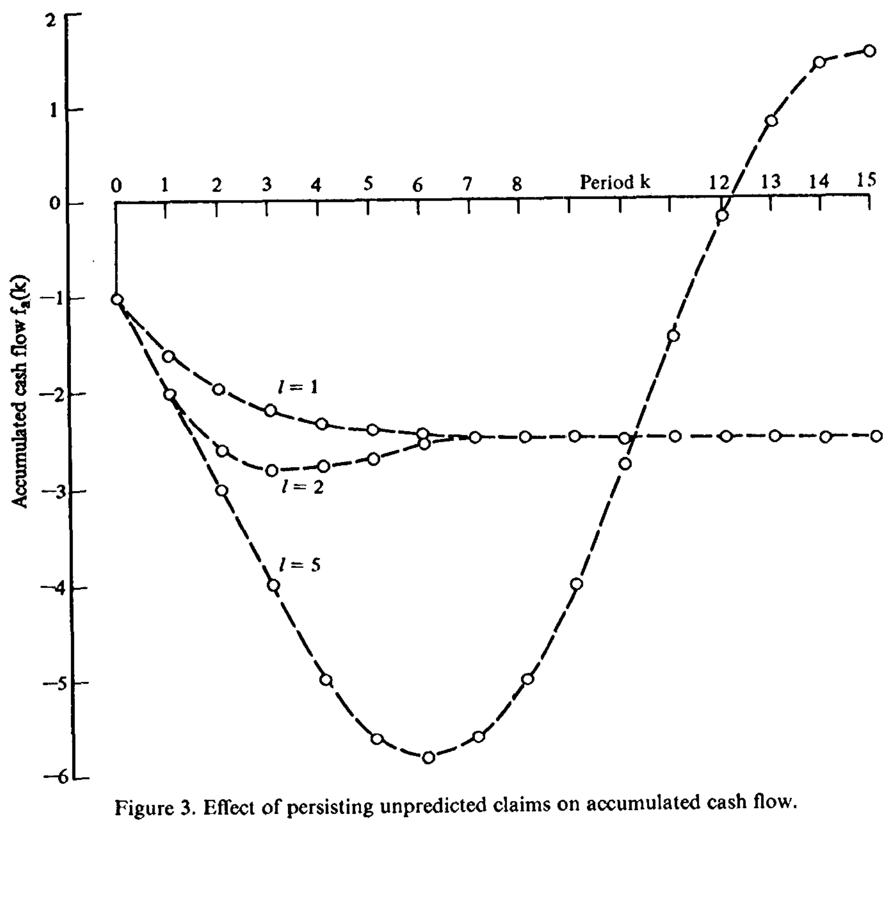


Figure 3. Effect of persisting unpredicted claims on accumulated cash flow.

two time instants. Such behaviour remains totally undescribed by the discrete-time model used here.

As expected from the preceding section, the cash flow in any period k approaches zero as k increases for the cases $l=1$ and 2. The approach is monotonic for a delay of one period and oscillatory for a delay of two periods, which is qualitatively similar to the results of the earlier paper. Whether the rate of decay to zero is fast enough is a matter for subjective judgement. (The control theorist might expect the speed of response to be improved by the addition of derivative action to the profit sharing scheme. This will be taken up in detail later.) For the case of $l=5$, the response is oscillatory and becomes unbounded as time increases. In short, the system is unstable.

For delays of one and two periods, the accumulated cash flow settles down to a finite but undesirably non-zero value. Again, the responses are monotonically convergent and oscillatory for $l=1$ and 2 respectively.

For the case of $l=5$, the accumulated cash flow becomes unbounded and does so in a dramatic manner.

5. STABILITY OF DISCRETE-TIME SYSTEMS

Unbounded responses are entirely unsatisfactory, hence it is timely to discuss the stability of the system. A system is said to be stable if its output remains bounded in response to any bounded input signal. Stability is thus a property of the system itself and independent of the sequence of values for the input variable.

Consider a discrete-time system with input sequence $\{x(k)\}$ and output sequence $\{y(k)\}$. The transfer function (see Balzer & Benjamin (1980)), $H(z)$, is defined as the ratio of the z-transform of the output sequence to that of the input sequence, $H(z)=Y(z)/X(z)$. Consequently

$$y(z) = H(z)X(z) = \frac{N(z)}{D(z)}X(z) \text{ (say)} \tag{10}$$

where $N(z)$ and $D(z)$ are polynomials in the transform parameter z. The zeros of the denominator polynomial $D(z)$, that is the roots or solutions of $D(z)=0$, are termed the poles of the transfer function. (The magnitude of the transfer function becomes infinite at a pole.) If the degree of $D(z)$ is n, there will be n poles and the system is termed nth order. The poles may not all be distinct; some may be repeated. Also, they may not all be real. If complex poles are present they occur in complex conjugate pairs, since the roots of $D(z)=0$ must so occur.

For the present, let the poles of $H(z)$ be distinct and be represented by $p_1, p_2, \ldots, p_n$. It can be shown (Cadzow, 1973) that the transient portion of the response $y(k)$ is given by

$$c_1p_1^k + c_2p_2^k + \cdots + c_np_n^k$$

where the c_i are constant coefficients. Clearly for this expression to decay to zero, the magnitude of each pole must be less than unity. A more rigorous analysis shows that a necessary and sufficient condition for a system to be stable is simply that the poles of its transfer function have magnitude less than unity, that is

$$|p_i| < 1, \text{ for all } i = 1, \cdots, n. \tag{11}$$

Plotted in the complex plane, all poles must lie within the unit circle for the system to be stable.

It can also be shown that the dynamic nature of each component of the total response is dictated by the location of its pole in the complex plane. For real p_i, the relationship is:

$p_i < -1 \Rightarrow$ oscillation diverging to infinity
$p_i = -1 \Rightarrow$ sustained oscillation
$-1 < p_i < 0 \Rightarrow$ oscillation decaying to zero
$0 < p_i < 1 \Rightarrow$ monotonic decay to zero
$p_i = 1 \Rightarrow$ constant
$1 < p_i \Rightarrow$ monotonic divergence to infinity.

Complex poles lead to oscillatory behaviour. Furthermore, within the unit circle, the closer a pole is to zero, the faster its contribution to the transient response dies away.

The equation formed by setting the denominator of the transfer function equal to zero, namely

$$D(z) = 0 \tag{12}$$

is known as the characteristic equation. The poles of the transfer function are thus the roots of the characteristic equation. For the present insurance system, the characteristic equation is

$$z^l - z^{l-1} + k_c k_p = 0. \tag{13}$$

For $l=1$ equation (13) becomes

$$z - 1 + k_c k_p = 0$$

and the single pole is located at $z = 1 - k_c k_p$. Since k_c and k_p are always positive, $|z|$ will be less than unity, when

$$k_c k_p < 2 \tag{14}$$

(which is certainly the case for $k_c = \cdot 8$ and $k_p = \cdot 5$). Realistic values of k_c and k_p are never likely to exceed unity and hence, *in practice*, the system with a delay of one period should always be stable and non-oscillatory.

For $l=2$ equation (13) becomes

$$z^2 - z + k_c k_p = 0$$

the roots of which are $z = \frac{1}{2} \pm \frac{1}{2}(1 - 4k_c k_p)^{\frac{1}{2}}$. Hence the poles will lie within the unit circle provided that

$$0 < k_c k_p < 1. \tag{15}$$

The system is stable with a delay of two periods, provided that k_c and k_p are realistically chosen to satisfy equation (15). The response will be oscillatory if the poles form a complex conjugate pair, namely when $k_c k_p > \frac{1}{4}$. For typical values of k_c and k_p this will often be the case.

For $l = 5$, equation (13) becomes

$$z^5 - z^4 + k_c k_p = 0. \tag{16}$$

Finding an analytic expression for the roots of this quintic polynomial would be difficult if not impossible. Given any particular numerical values for k_c and k_p, it is possible to determine the real root without too much difficulty. Determination of the two remaining pairs of complex conjugate roots is a little more complicated. More importantly, however, the way in which they vary as the product $k_c k_p$ is changed would only be available in a piecemeal manner. Control theorists have developed a technique termed the 'root locus method', which is systematic, simple and sheds considerable light on the behaviour of the system as $k_c k_p$ is varied. This technique is presented and applied in the following section.

6. ROOT LOCUS METHOD

Dorf (1980) and Ogata (1970) discuss the root locus method for continuous-time systems. Its application to discrete time systems and to the present case with $l = 5$ is presented without proof†, in order to introduce the reader to its power (through a problem of direct interest to actuaries). In essence the root locus method is a graphical technique which shows clearly how the roots of the characteristic equation $D(z) = 0$, that is the poles of the transfer function, move in the complex z-plane as some particular parameter of interest ($k_c k_p$ here) is changed.

The characteristic equation is first rearranged into the form

$$1 + KP(z) = 0 \tag{17}$$

where K is the parameter of interest. In the present problem, the parameter of interest is the product $k_c k_p$, so that equation (13) is rearranged into the form

$$1 + \frac{k_c k_p}{z^{l-1}(z-1)} = 0$$

where

$$P(z) = \frac{1}{z^{l-1}(z-1)} \tag{18}$$

$$= \frac{1}{z^4(z-1)}, \text{ for } l = 5.$$

We shall now develop a sketch of how the poles of the transfer functions in equations (3) and (5), that is the roots of the characteristic equation (13), move as

† See Appendix B for a justification of some of the rules.

$k_c k_p$ is increased from zero to infinity. The path so traced out in the complex z-plane is termed the root locus. Fortunately, this root locus can be sketched with reasonable accuracy from a few simple calculations involving $P(z)$ alone.

It can be shown that the paths of the root locus begin at the poles of $P(z)$ and end at the zeros of $P(z)$. For $l=5$, there are five finite poles of $P(z)$, namely four at $z=0$ and one at $z=1$. These poles are denoted by a cross (x) on Figure 4. Here $P(z)$ possesses no finite zeros (values of z which make $P(z)=0$), so that all of the paths or branches of the root locus will diverge to infinity.

The portion of the root locus which lies on the real axis can only lie to the left of an odd number of finite poles plus finite zeros. Consequently, the portion of the root locus shown as a thick line on part (a) of Figure 4 can now be drawn. The arrowheads denote the direction of increasing $k_c k_p$ and the Roman IV has been used to indicate a quadruple pole of $P(z)$ at $z=0$.

The number, N, of paths diverging to infinity is

$$N = n_p - n_z = 5-0 = 5$$

where n_p is the number of finite poles of $P(z)$ and n_z is the number of finite zeros of $P(z)$. The paths to infinity become asymptotic to straight lines as $k_c k_p$ becomes large. These asymptotes meet at a common point σ_a, called the asymptote centroid, on the real axis given by

$$\sigma_a = \frac{\Sigma\text{ Finite poles} - \Sigma\text{ Finite zeros}}{N} = \frac{0+0+0+0+1}{5} = 0{\cdot}2 \text{ (here).}$$

The angles ϕ_a, which they make with the real axis are given by

$$\phi_a = \frac{(2i+1)}{N}180°,\ i = 0,\ \pm 1,\ \pm 2, \cdots$$

$$= 36°,\ 108°,\ 180°,\ -36°,\ -108° \text{ (here).}$$

These asymptotes are shown as broken lines in part (b) of Figure 4.

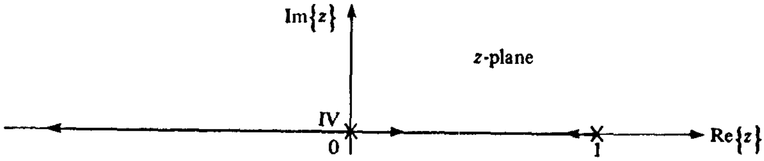
Clearly the root locus has to break away from the real axis at one or more points in the interval [0,1]. Break-away and break-in occur when the parameter of interest has a maximum or minimum value, or sometimes a point of horizontal inflection, with respect to z. Such points are usually found by the simple calculus procedure of setting

$$\frac{dK}{dz} = \frac{d}{dz}\left[\frac{-1}{P(z)}\right] = 0$$

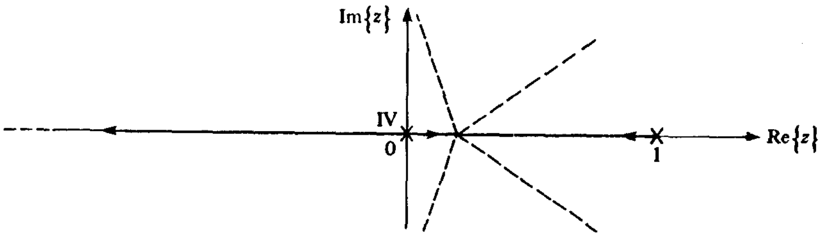
and solving for z. Any solutions lying outside the expected range are rejected as inadmissible. Here

$$\frac{d}{dz}(-z^4(z-1)) = -5z^4+4z^3 = 0.$$

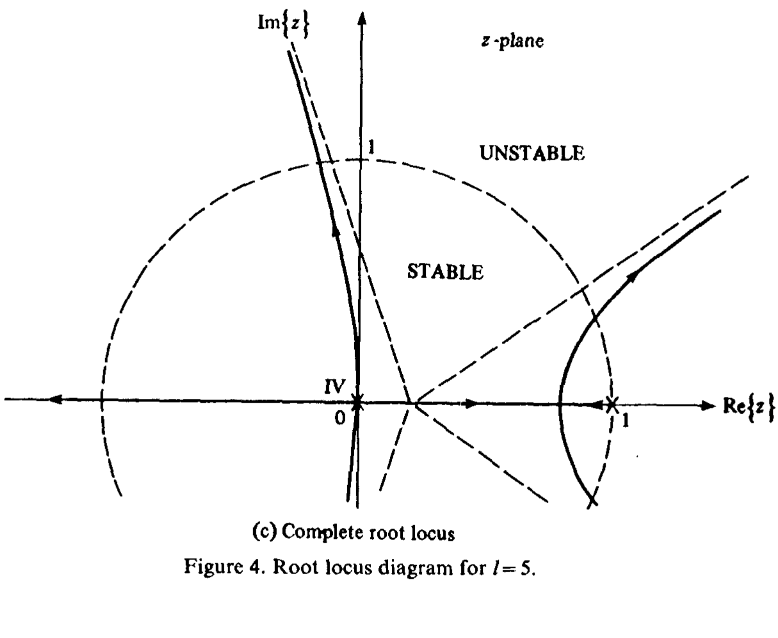
Hence $z=0$ or $\cdot 8$, both of which are admissible as break-away points.



(a) Section on real axis



(b) Asymptotes added



(c) Complete root locus

Figure 4. Root locus diagram for $l = 5$.

At each break-away point, two paths leave the real axis. Furthermore, the tangents to these paths must be equally spaced in 360°, while the root locus must be symmetric about the real axis. Consequently the tangents must be vertical. The nature of the complete root locus is now defined and can be sketched approximately to give Figure 4(c).

The only additional points which might be required accurately are those where the paths cross the unit circle, since these are the limiting points for stability. These points are often termed the crossover points. From an approximate sketch of the root locus paths, one can conclude that one crossover point lies at exactly -1, while the other four lie somewhere near $\cdot 92 \pm \cdot 40j$ and $-\cdot 15 \pm \cdot 99j$ where j indicates the square root of minus one. (These guesses for the crossover points were made prior to drawing the root locus accurately.) Determination of the exact crossover points is now done by rapidly convergent trial and error.

First, consider the crossover point near $\cdot 92 + \cdot 40j$. The simplest procedure is to vary the imaginary part and calculate the corresponding real part $(= [1 - (\mathrm{Im}\{z\}^2]^{\frac{1}{2}})$ necessary to stay on the unit circle. This point is then tested to see if it lies on the root locus using the simple 'angle criterion'. The angle criterion states that if the point z_1 lies on the root locus, the angle of $P(z_1)$ is an odd multiple of 180°.

$$\measuredangle P(z_1) = (2i+1)\,180°,\; i = 0, \pm 1, \pm 2, \cdots. \tag{20}$$

Here, straightforward complex variable theory leads to

$$\begin{aligned} \measuredangle P(z) &= \measuredangle 1 - (\measuredangle z^4 + \measuredangle (z-1)) \\ &= 0 - 4\measuredangle z - \measuredangle (z-1) \\ &= -4\tan^{-1}\left(\frac{b}{a}\right) - \tan^{-1}\left(\frac{b}{a-1}\right) \end{aligned} \tag{21}$$

where $b = \mathrm{Im}\{z\}$ and $a = \mathrm{Re}\{z\}$. The important point to note about the angle criterion is that the value of $k_c k_p$ is not required. This is both normal and essential.

Equation (21) can be used repeatedly to test successive estimates of the crossover points. For example,

$$\measuredangle P(\cdot 92 + \cdot 40j) = -195\cdot 3°.$$

Using a programmable pocket calculator, one crossover point was found (in 2 or 3 minutes) to be $\cdot 9397 + \cdot 3420j$. By symmetry, another crossover point must occur at $\cdot 9397 - \cdot 3420j$. The remaining crossover points occur at $-\cdot 1736 \pm \cdot 9848j$.

The values of the parameter $k_c k_p$ at the crossover points can now be determined from the 'magnitude criterion', which states that the magnitude of $KP(z)$ is unity for any point z_1 on the root locus. Hence

$$K = 1/|P(z_1)|. \tag{22}$$

Here

$$\begin{aligned} k_c k_p &= |z^4||z-1| \\ &= (a^2+b^2)^2((a-1)^2+b^2)^{\cdot 5} \end{aligned} \tag{23}$$

where

$$z_1 = a+jb.$$

At the crossover point $z = \cdot 9397 \pm \cdot 3420j$, hence

$$\begin{aligned} k_c k_p &= (1)^2 \times ((-\cdot 0603)^2 + (\cdot 3420)^2)^{\cdot 5} \\ &= \cdot 3473. \end{aligned}$$

Similarly at $z = -\cdot 1736 \pm \cdot 9848j$, $k_c k_p = 1 \cdot 5320$ and at $z = -1$, $k_c k_p = 2$. In summary, if the product $k_c k_p$ is greater than ·35 then two or more of the roots of the characteristic equation (poles of the transfer function) will lie outside the unit circle and the system will be unstable. In the numerical example studied in this and the earlier paper, $k_c k_p$ was equal to ·4 and hence the system was unstable.

The response, to any unpredicted claims, will always be oscillatory for any positive non-zero choice of $k_c k_p$. This can be seen immediately from even part (a) of Figure 4, since one of the roots lies on the negative real axis for $k_c k_p > 0$. This indicates an oscillatory component in the response.

7. ADDITION OF INTEGRAL ACTION

In § 3, an undesirable non-zero steady state response to a persisting stream of unpredicted claims was demonstrated. This is not an uncommon situation for control strategies which include only 'proportional action'. The present system is of this type because a proportion of the estimated accumulated surplus $\hat{g}(k)$ is fed back by the Profit Sharing Scheme. It is widely known to control theorists that the addition of so called 'integral control action' will generally alleviate this problem. Integral action for a continuous-time system involves using a control component based on the integral (with respect to time) of the difference between the desired and the actual values of the controlled variable. If our model had been developed in continuous time, the integral component would be based on

$$\int_0^t \hat{g}(t)dt$$

since the desired value of $\hat{g}(t)$ is zero. In discrete time, the integral is replaced by a summation over consecutive time instants. A proportional plus integral $(P+I)$ action profit sharing strategy which replaces equation (19) of Balzer & Benjamin (1980) is then

$$p_f(k) = k_p \hat{g}(k) + k_i T \sum_{i=0}^{k} \hat{g}(i) \tag{24}$$

where k_i is a constant and T is the length of the financial period. Taking z-transform of equation (24) leads to

$$P_f(z) = k_p\hat{G}(z) + k_iT\frac{z}{z-1}\hat{G}(z)$$

and hence the transfer function of the new $P+I$ Profit Sharing Scheme becomes

$$\begin{aligned}\frac{P_f(z)}{\hat{G}(z)} &= k_p + k_iT\frac{z}{z-1}\\ &= \frac{(k_p+k_iT)z - k_p}{z-1}\end{aligned} \tag{25}$$

In passing, it is interesting to look at the difference equation which corresponds to equation (25). Dividing through by z and cross-multiplying gives

$$(1-z^{-1})\,P_f(z) = [(k_p+k_iT) - k_pz^{-1}]\hat{G}(z).$$

Taking inverse z-transforms

$$p_f(k) = p_f(k-1) + (k_p+k_iT)\,\hat{g}(k) - k_p\hat{g}(k-1). \tag{26}$$

Equation (26) clearly demonstrates that some 'memory' is associated with the new strategy, in that the present feedback is equal to the previous value plus an updated correction based on the past and present estimated surpluses.

Block diagrams showing the effect of unpredicted claims $c_u(k)$ on cash flow $f(k)$ and on accumulated cash flow $f_a(k)$ are shown in Figure 5. By straightforward block diagram reduction, the transfer functions become

$$\frac{F(z)}{C_u(z)} = -\frac{(z-1)^2z^{l-1}}{(z-1)^2z^{l-1} + k_c(k_p+k_iT)z - k_ck_p} \tag{27}$$

and

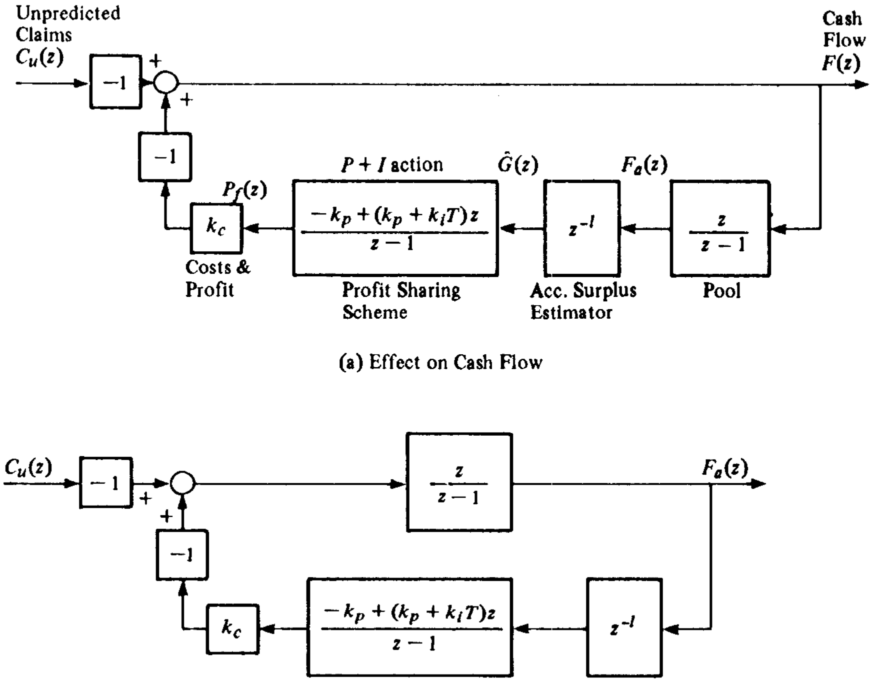
$$\frac{F_a(z)}{C_u(z)} = -\frac{(z-1)z^l}{(z-1)^2z^{l-1} + k_c(k_p+k_iT)z - k_ck_p}. \tag{28}$$

Using the Final Value Theorem it is now possible to demonstrate that even in the presence of a persisting stream of unpredicted claims, both the cash flow and the accumulated cash flow approach steady state values of zero as time progresses. From equation (2), $C_u(z) = z/(z-1)$, hence

$$\begin{aligned}f(\infty) &= \lim_{z\to 1}(z-1)\,F(z)\\ &= \lim_{z\to 1}(z-1)\times\frac{-(z-1)^2z^l}{(z-1)^2z^l + k_c(k_p+k_iT)z^2 - k_ck_pz}\times\frac{z}{z-1}\\ &= 0\end{aligned}$$

and

$$\begin{aligned}f_a(\infty) &= \lim_{z\to 1}(z-1)\,F_a(z)\\ &= \lim_{z\to 1}(z-1)\times\frac{-(z-1)z^l}{(z-1)^2z^l + k_c(k_p+k_iT)z^2 - k_ck_pz}\times\frac{z}{z-1}\\ &= 0\end{aligned}$$



(a) Effect on Cash Flow

(b) Effect on Accumulated cash flow

Figure 5. Effect of unpredicted claims.

Hence, provided that the delay l is small enough for the system to be stable, both the cash flow and the accumulated cash flow will settle down to a steady state value of zero in the presence of a persisting stream of unpredicted claims. Clearly this is a highly desirable feature for any profit sharing strategy. As discussed earlier, it has definite advantages over *ad hoc* adjustment of base premiums.

8. SELECTION OF PARAMETER VALUES

The root locus method will now be used to select numerical values for the parameters k_p and k_i in the profit sharing strategy. Up until now the intuitively appealing value of $k_p = \cdot 5$ (50% sharing) has been used. It will now be shown that a smaller value has advantages to both parties. Following this, a compromise value for the amount k_i of integral action is determined.

It was argued in the earlier paper that a delay of $l = 2$ periods was realistic. This value was based on the time and accounting/actuarial effort involved in collecting information and also on the nature of the time history of paid claims. Hence the analysis is pursued for the case of $l=2$. Also, without loss of generality, the length of the financial time period T can be set to one time unit (which might be anything from one day to one year).

From equation (27) or (28), the characteristic equation is

$$(z-1)^2 z + k_c(k_p + k_i)z - k_c k_p = 0$$

or

$$z^3 - 2z^2 + (1 + k_c k_p + k_c k_i)z - k_c k_p = 0 \tag{29}$$

which for $k_i = 0$ can be arranged after some pole-zero cancellation into the form

$$1 + \frac{k_c k_p}{z(z-1)} = 0. \tag{30}$$

Hence the function of interest is

$$P(z) = \frac{k_c k_p}{z(z-1)}.$$

It is simple to show that the root locus for the system with $l=2$, $T=1$ and $k_i=0$ is as shown in Figure 6.

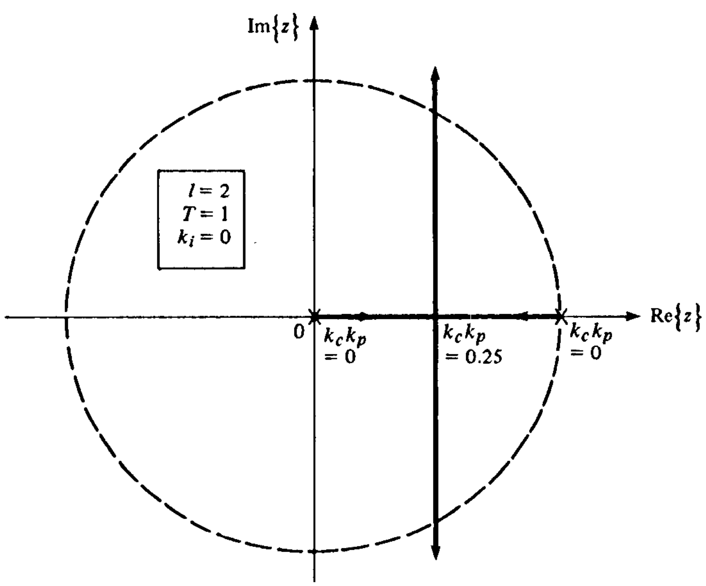


Figure 6. Root locus for proportional action.

It is now claimed that the fastest response is obtained when both roots are coincident at the breakaway point $z = \cdot 5$ (where $k_c k_p = \cdot 25$). To amplify this last statement we need to define settling time and show how it varies within the z-plane. The settling time is the time required for the output of the system to reach its steady state value within plus or minus a certain percentage following a step change in input. Normally ± 2 or $\pm 5\%$ settling times are used. In the z-plane, lines of constant settling time are circles centred on the origin. The origin of the z-plane corresponds to a zero settling time, while the unit circle corresponds to an infinite settling time. The variation between these two extremes is exponential, as shown in Figure 7, where the lines of constant settling time are labelled with a figure indicative of the ratio of the 2% settling time, T_{s2}, to the sampling period, T. The development of this ratio is given in Appendix A.

When more than one root is present, the one furthest away from the origin will dominate the response. This happens because those roots which are closer to $z = 0$ produce transient components which die away more quickly, thus leaving the more slowly decaying component to dominate. From Figure 6, it is clear that the root furthest from the origin can get no closer than the breakaway point, $z = \frac{1}{2}$, where both roots are coincident. At this point, $k_c k_p = \cdot 25$. Thus the fastest response possible is achieved when $k_c k_p = \cdot 25$. If a margin of 20% for costs and profits is allowed, then $k_c = \cdot 8$ and the fastest response occurs for $k_p = \cdot 3125$. Thus for these conditions, a 31·25% rather than 50% sharing arrangement is

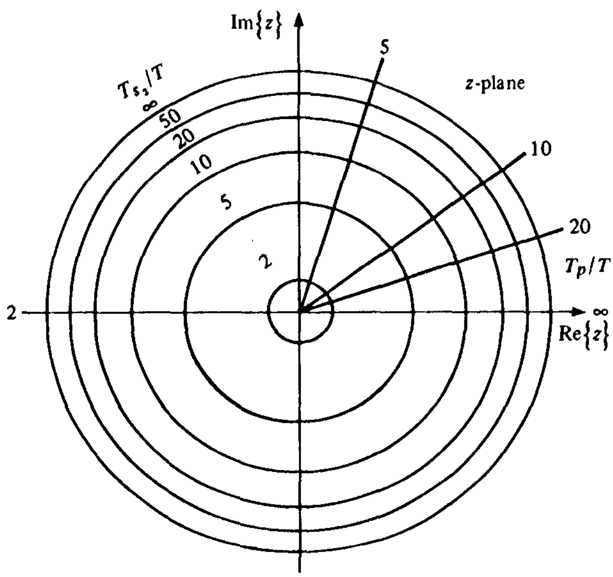


Figure 7. Indicative settling and periodic times.

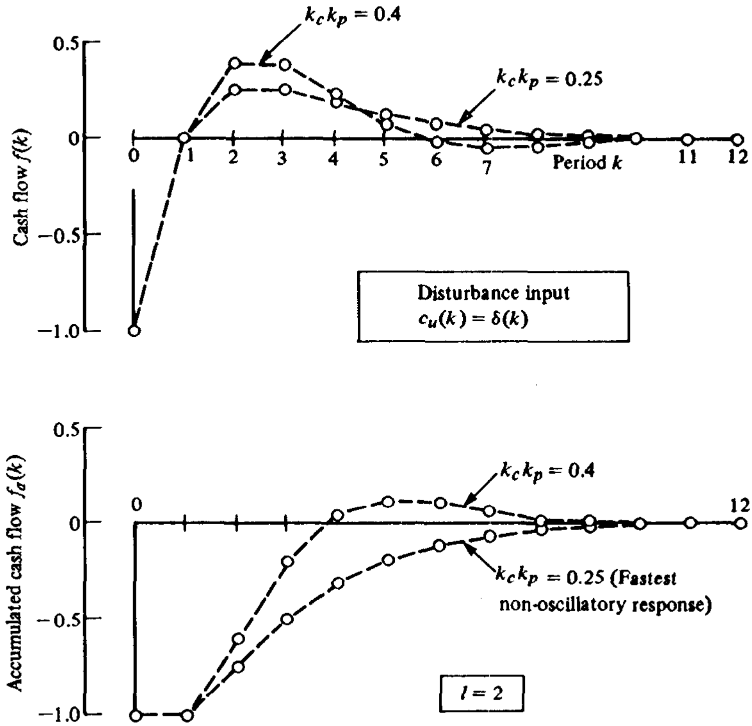


Figure 8. Improved proportional profit sharing.

recommended for the fastest possible response with no 'overshoot' or oscillation. Comparative responses to an isolated group of unpredicted claims are shown in Figure 8.

The following analysis assumes that the proportional constant k_p has been chosen such that $k_c k_p = \cdot 25$.

Consider now the case when the integral constant k_i is non-zero. The characteristic equation (29) is arranged, after a pole zero cancellation at the origin, into the form

$$1 + \frac{k_c k_i z}{z^3 - 2z^2 + 1\cdot 25z - \cdot 25} = 0. \tag{31}$$

The function of interest is then

$$P(z) = \frac{k_c k_i z}{z^3 - 2z^2 + 1\cdot 25z - \cdot 25} \tag{32}$$

which leads to the root locus diagram shown in Figure 9. The Roman II denotes a double pole of $P(z)$ at $z = \cdot 5$. Again, a minimum settling time exists and corresponds to the new breakaway point, at which $k_c k_i = \cdot 02254$.

The response of the accumulated cash flow to a persisting stream of unpredicted claims, when proportional plus integral action profit sharing is used,

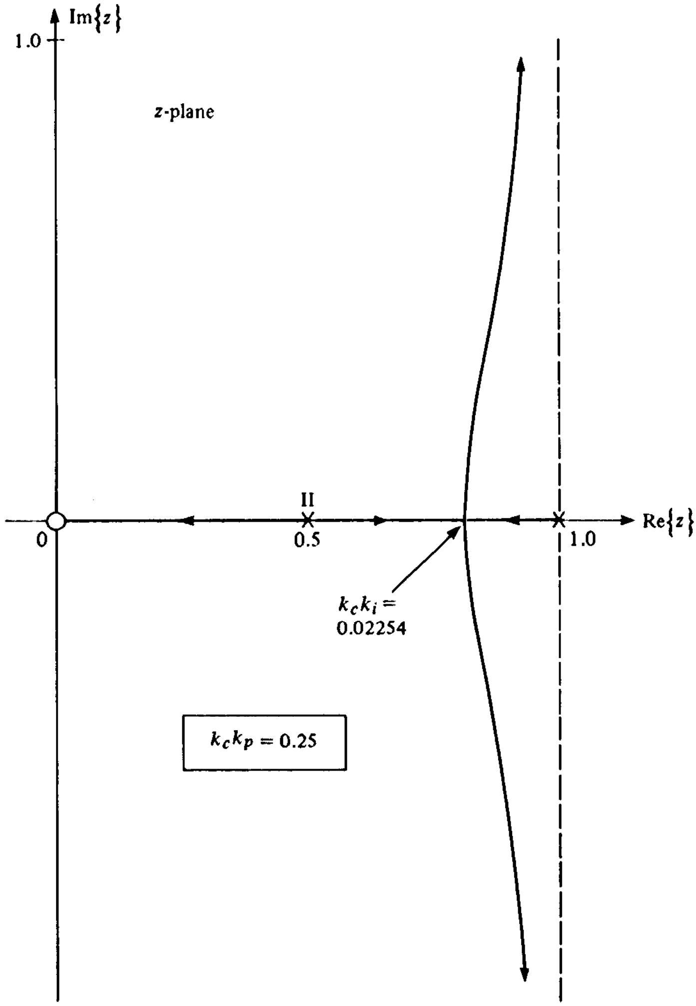


Figure 9. Root locus for k_i varying.

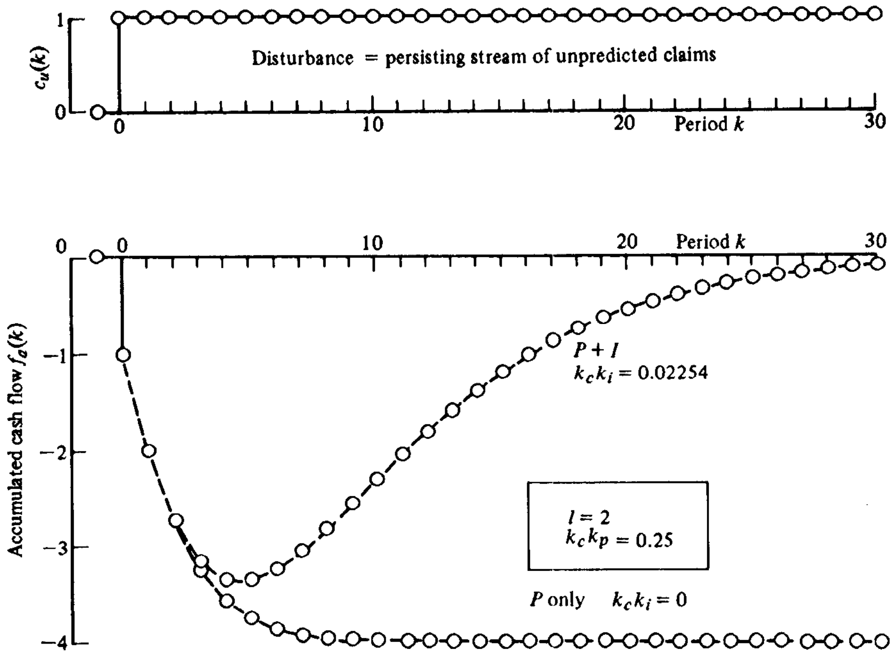


Figure 10. Effect of integral profit sharing on response.

is shown in Figure 10 for $k_c k_p = \cdot 25$ and $k_c k_i = \cdot 02254$. The response for P action only is also shown for comparison. The eventual return of the accumulated cash flow to zero is both desirable and reassuring. It represents a considerable improvement in approach over any *ad hoc* adjustment of the Base Premium Calculator to compensate for the stream of unpredicted claims. The only problem is that the settling time is relatively long.

If $k_c = \cdot 8$ then $k_c k_i = \cdot 02254$ leads to $k_i = \cdot 0282$ or 2·82%.

9. DERIVATIVE ACTION

Control theorists have long known that the addition of derivative action to a feedback control system usually improves its speed of response. Derivative action makes use of information regarding the rate of change of the variable, upon which the feedback is based. The usefulness of such information is obvious in the following familiar situation. Assume that you are driving a motor car in a stream of traffic and that a 5 m gap between your vehicle and the one in front is to be maintained. If the gap opens up to 10 m then under proportional control you should accelerate. If, however, the vehicle in front is stationary and yours is travelling at 60 km/h, then you should be braking fairly sharply even though the gap is larger than desired. The difference in information is simply the rate of change of the gap. Its importance to improving the speed of response of an automatic control system is obvious.

In a continuous-time system, the derivative component of the control action is based on the time rate of change, $de(t)/dt$, of the error, $e(t)$, between desired and actual values of the controlled variable. In a discrete-time system the derivative is replaced by $(e(k)-e(k-1))/T$, so that a proportional plus integral plus derivative (PID) control strategy for the present profit sharing scheme is

$$p_f(k) = k_p\hat{g}(k)+k_iT\sum_{i=0}^{k}\hat{g}(i)+\frac{k_d}{T}(\hat{g}(k)-\hat{g}(k-1)). \tag{33}$$

Taking z-transforms

$$\begin{aligned} P_f(z) &= k_p\hat{G}(z)+\frac{k_iTz}{z-1}\hat{G}(z)+\frac{k_d}{T}(1-z^{-1})\hat{G}(z) \\ &= \frac{(k_p+k_iT+k_d/T)z^2-(k_p+2k_d/T)z+k_d/T}{z(z-1)}\hat{G}(z) \end{aligned} \tag{34}$$

where k_d is a constant to be selected.

For $l=2$ it can be shown that the transfer functions relating the cash flow and accumulated cash flow to unpredicted claims are

$$\frac{F(z)}{C_u(z)} = \frac{-z^2(z-1)^2}{z^4-2z^3+(1+k_ck_p+k_ck_iT+k_ck_d/T)z^2-(k_ck_p+2k_ck_d/T)z+k_ck_d/T}$$

and

$$\frac{F_a(z)}{C_u(z)} = \frac{-z^3(z-1)}{z^4-2z^3+(1+k_ck_p+k_ck_iT+k_ck_d/T)z^2-(k_ck_p+2k_ck_d/T)z+k_ck_d/T} \tag{35, 36}$$

The root locus diagram for PID profit sharing with $l=2$, $T=1$, $k_ck_p=\cdot25$ and $k_ck_i=\cdot02254$ is shown in Figure 11. The diagram might surprise many practitioners of automatic control who have been led to believe that the addition of derivative action always improves the speed of response of a system. Here it does not. Certainly as k_d and hence the amount of derivative action is increased from zero, the root at $\cdot382$ moves towards the origin until it meets the root moving out from the origin. They meet at the breakaway point at $z=\cdot1776$. This would represent a speeding up of the system but for the effects of the other two roots. The two roots which start at $z=\cdot809$ for $k_d=0$, move away from the origin as k_d is increased. Moreover these roots dominate the response, since their effects remain long after those of the other two are negligible.

Consequently, the addition of any derivative action at all leads to a slowing of

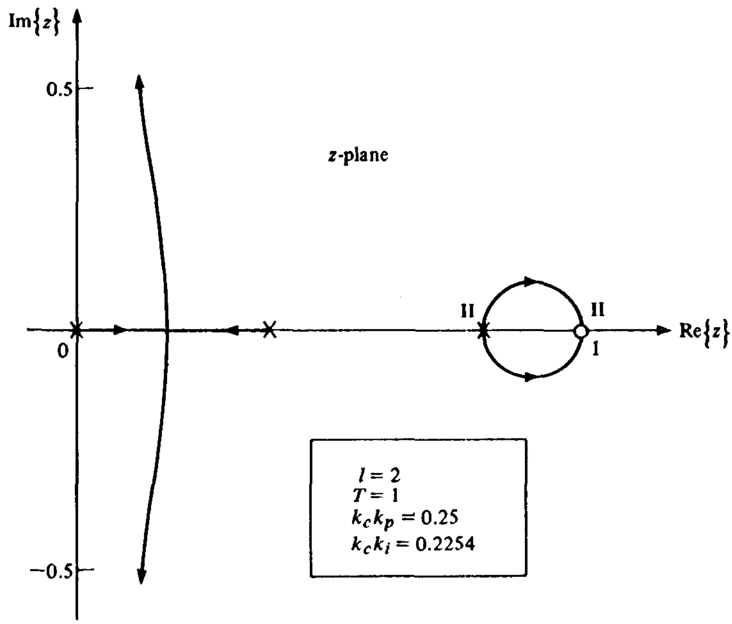


Figure 11. Root locus for PID profit sharing.

the response. The gross amount of movement of the dominant roots may not look great, but it should be remembered that the point $z = 1$ corresponds to an infinite settling time. Since settling time varies exponentially with distance from the origin, even a small movement can have a large effect. Also, since the dominant roots break away from the real axis immediately k_d becomes non-zero, any amount of derivative action will introduce oscillatory components to the response.

For this system derivative action is undesirable. Why should this be so? The key to the answer lies in the time delay of l periods incurred in producing an estimate $\hat{g}(k)$. Consider again the previous example of driving a motor vehicle. Think of the adverse effect of introducing a time delay between the measurement of the distance between vehicles (or its rate of change) and the driver acting upon that information. Most drivers do not even rely on instantaneous information but make use of preview information by looking several cars ahead and predicting what the car in front of them will do before it actually happens. The dramatic improvements achievable using preview control are demonstrated and quantified in a different context by Balzer (1981). However, driving behind a large truck, for example, removes the preview information and reduces the quality of control markedly. The effect of a time delay is even more deleterious than the removal of preview information. It is not surprising then that derivative action is not effective in the present system.

10. CONCLUSIONS

The profit/loss sharing scheme introduced by Balzer & Benjamin (1980) has been subjected to further analyses, which give greater insight into its dynamic behaviour. Under the more demanding disturbance of a persisting stream of unpredicted claims, a significant non-zero accumulated cash flow is found to occur after steady state conditions are reached. The dynamic behaviour was then investigated using the root locus technique and improved. The addition of integral action was seen to drive the steady state value of the accumulated cash flow to the desirable value of zero. Finally derivative action was shown conclusively to offer no improvements due to the time delay present in the system.

ACKNOWLEDGEMENT

The author is indebted to Sidney Benjamin for his perceptive questioning, enlightened suggestions and, not least, for his general enthusiasm.

REFERENCES

BALZER, L. A. (1981) Optimal Control with Partial Preview of Disturbances and Rate Penalties. *Int. J. Control* **33** (2), 323.

BALZER, L. A. & BENJAMIN, S. (1980) Dynamic Response of Insurance Systems with Delayed Profit/Loss Sharing Feedback to Isolated Unpredicted Claims. *J.I.A.* **107**, 513.

CADZOW, J. A. (1973) *Discrete-time Systems—An Introduction with Interdisciplinary Applications*. Prentice-Hall.

DORF, R. C. (1980) *Modern Control Systems*. Addison-Wesley.

OGATA, K. (1970) *Modern Control Engineering*. Prentice-Hall.

APPENDIX A

INDICATIVE RATIO FOR SETTLING TIME

Texts dealing with the root locus method for continuous-time systems are widely available (for example, Dorf, 1980; Ogata, 1970). The same is not true for discrete-time systems. Hence this appendix addresses one aspect which does not appear to be covered adequately elsewhere. Familiarity with or ready access to the root locus method for continuous systems is assumed.

For continuous-time systems the root locus is plotted in the complex s-plane. On this plane a line of constant settling time is a vertical line with a constant real part. More specifically, if T_{s2} is the $\pm 2\%$ settling time and if

$$s = \sigma + j\omega \tag{A1}$$

then

$$T_{s2} = -\frac{4}{\sigma}. \tag{A2}$$

The relationship between the s- and z-planes is the conformal mapping

$$\begin{aligned} z &= e^{sT} \\ &= e^{(\sigma + j\omega)T} = e^{\sigma} e^{j\omega T}. \end{aligned} \tag{A3}$$

The magnitude of the complex variable z is then

$$|z| = e^{\sigma} = \exp(-4T/T_{s2}) \tag{A4}$$

where T_{s2} is the $\pm 2\%$ settling time of the continuous-time response of a continuous-time system of the same order and having the same response at the discrete sampling instants as the discrete system under study. Because the discrete response only exists at discrete instants in time, the following results can only be classed as indicative of the discrete settling times. For sampling periods which are short compared to the settling time, the error is negligible. For constant T_{s2}, $|z|$ is constant. Hence a contour of constant T_{s2} is a circle with centre $z=0$, as shown in Figure 7, where the corresponding T_{s2}/T ratios are based on equation (A4).

In the s-plane, the frequency of oscillation of an oscillatory response is entirely dependent upon the imaginary part ω of the root. Lines of constant frequency of oscillation are then horizontal lines in the s-plane. In the z-plane, equation (A3) implies that the angle of z is

$$\measuredangle z = \omega T \text{ (radians)} \tag{A5}$$

Consequently in the z-plane, lines of constant frequency of an oscillatory response become straight lines or rays radiating from the origin $z=0$. Such lines are also shown in Figure 7. Rather than labelling them with the dimensionless parameter ωT, the ratio of T_p/T, where T_p is the period of oscillation, is used.

Clearly

$$\frac{T_p}{T} = \frac{2\pi}{\omega T}$$

hence equation (A5) becomes

$$\measuredangle z = \frac{2\pi}{T_p/T}\,(\text{radians}) = \frac{360}{T_p/\mathrm{T}}\,(\text{degrees}). \tag{A6}$$

The values of T_p/T shown in Figure 7 are based on equation (A6).

A long period of oscillation (which means a low frequency of oscillation) is desirable since the response will oscillate fewer times before reaching its steady state value.

APPENDIX B

JUSTIFICATION FOR SELECTED ROOT LOCUS RULES

This appendix has been added, at the suggestion of Sidney Benjamin, to assist those who do not have ready access to the suggested textbooks but who wish to see a justification for some of the rules used in constructing the root locus diagrams. I am also grateful for a novel justification of one of the rules.

Rearranging the characteristic equation (17) leads to

$$KP(z) = -1. \tag{B1}$$

Magnitude and angle criteria. In general z will take complex values and hence $P(z)$ will also. Taking the magnitude of each term in equation (B1) for a particular value z_1 of z, leads to

$$K\,|\,P(z_1)\,| = 1 \tag{B2}$$

or

$$K = 1/\,|\,P(z_1)\,|$$

which is the Magnitude Criterion for any point z_1 on the root locus. By taking the phase angle of each term in equation (B1)

$$\measuredangle K + \measuredangle P(z_1) = \measuredangle(-1)$$

hence

$$\measuredangle P(z_1) = (2i+1)\,180^\circ \tag{B3}$$

which is the Angle Criterion for a point z_1 on the root locus.

Terminal points. $P(z)$ can be expanded into the quotient of two products involving its zeros and poles, so that the characteristic equation (17) can be rewritten as

$$1 + K\frac{\prod_{i=1}^{n_z}(z-z_i)}{\prod_{i=1}^{n_p}(z-p_i)} = 0$$

where z_i are the finite zeros and p_i the poles of $P(z)$. Consequently

$$\prod_{i=1}^{n_p}(z-p_i) + K\prod_{i=1}^{n_z}(z-z_i) = 0. \tag{B4}$$

When $K=0$, $z=p_i$ and hence the roots lie at the poles of $P(z)$. Clearly the root locus begins at the poles of $P(z)$. Similarly, when $K=\infty$, $z=z_i$ or $z=\infty$ and the root locus ends at the finite zeros of $P(z)$ or at infinity.

Portion on real axis. That the portion(s) of the root locus lying on the real axis must lie to the left of an odd number of poles plus finite zeros, follows immediately from the angle criterion.

Symmetry about real axis. Complex roots of a polynomial can only appear in complex conjugate pairs. Consequently the root locus must be symmetric about the real axis.

Paths to infinity. The number of infinite zeros of $P(z)$ and hence the number N of paths to infinity on the root locus, is clearly equal to the number of poles minus the number of finite zeros.

Asymptotes. At a very distant point on one of the paths to infinity, the angles from each pole and finite zero are essentially equal. Call this angle ϕ_a. The angle criterion then implies that at this point.

$$N\phi_a = (2i+1)\,180°, i = 0, \pm 1, \pm 2, \cdots$$

Hence

$$\phi_a = \frac{(2i+1)}{N}\,180°, i = 0, \pm 1, \pm 2, \cdots$$

Derivation of the value for σ_a is not sufficiently brief to be included here.

Breakaway points. Breakaway points obviously occur where there are multiple roots of the characteristic equation. Denoting the characteristic polynomial by $f(z)$, the characteristic equation can be written as

$$f(z) = 1+KP(z) = 0$$

which will have multiple roots, when

$$\frac{df(z)}{\mathrm{dz}} = 0.$$

Noting that, on the root locus, z is a function of K

$$\frac{df(z)}{dz} = 0+\mathrm{K}\frac{dP(z)}{dK}\frac{dK}{dz} = K\frac{d}{dK}\left(\frac{-1}{K}\right)\frac{dK}{dz}$$
$$= \frac{1}{K}\frac{dK}{dz}.$$

Consequently, at a breakaway point $dK/dz=0$. (The possible multiplicity of roots at $K=\infty$ is of no consequence.)

Tangents at breakaway points. These tangents are equally spaced over 360°; for example, two roots imply tangents perpendicular to the real axis and four roots, an angle of 45°, etc. This result follows by taking a small departure from the real axis and noting that the angle criterion is only satisfied under the above circumstances.

Other rules. Several other rules exist but will not be covered here.