MIDDLESEX UNIVERSITY

THE USE AND MEANING OF ALL SOLUTIONS (INTEREST RATES) TO THE TIME VALUE OF MONEY EQUATION

A DISSERTATION SUBMITTED TO THE MIDDLESEX UNIVERSITY BUSINESS SCHOOL IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY BY PUBLIC WORKS

BY MICHAEL JAMES OSBORNE

November 2010
ABSTRACT

The time value of money (TVM) equation is a key equation in economics and finance. It takes the form of an $n^{th}$ order polynomial having $n$ roots. It is usual to calculate and use only one root (interest rate). The remaining $(n-1)$ roots are mostly complex or negative, and are usually discarded. In this thesis it is shown that the unorthodox roots (interest rates) should not be ignored because they have utility and meaning.

New expressions are developed for existing concepts in economics and finance. The concepts include duration in bond mathematics, net present value (NPV) in capital budgeting, the value of a stock of capital in capital theory, and the total charge for credit in loan analysis. The new expressions for these concepts employ all possible interest rates as components.

The new expressions provide solutions to puzzles. In bond mathematics, the new equation for duration delivers what previous formulas for duration fail to provide: an accurate measure of the impact of a change in interest rate on the price of a bond. In capital budgeting, the new equation for NPV offers a resolution to the debate about the relative merits of NPV and internal rate of return (IRR) as investment criteria. In economics, a solution is proposed to the reswitching debate in the Cambridge capital controversies. In credit analysis, a new relationship is developed between the total charge for credit and the orthodox measure of the cost of a loan, the annual percentage rate (APR). The result implies that the complicated APR need no longer be the focus of consumer credit legislation; the total charge for credit and its variants suffice.

The new analysis not only employs all interest rates, it also endows them with meaning. The suggested interpretation of a complete cluster of interest rates sheds new light on the meaning of orthodox rates such as IRR and APR.
CONTENTS

ACKNOWLEDGEMENTS 4

REGULATIONS: CONTEXT STATEMENT 5

THE CONTEXT STATEMENT

PREAMBLE 6

CHAPTER 1: Preliminaries 7

CHAPTER 2: The multiple interest rate literature 18

CHAPTER 3: Chronological overview of Works 1 to 10 34

CHAPTER 4: Further work and open questions 44

CHAPTER 5: Conclusions 59

APPENDIX A: Research methodology 62

APPENDIX B: A perspective on the research 65

APPENDIX C: On the calculation of roots 70

APPENDIX D: An example of chaotic behaviour 73

REFERENCES 74

THE WORKS See overleaf
THE WORKS

Work 1

Work 2

Work 3

Work 4

Work 5
Osborne, M. & Osborne, M. 2003. A polynomial that is a statistical prism, *Mathematical Gazette*, 87(508)

Work 6

Work 7

Work 8

Work 9

Work 10
ACKNOWLEDGEMENTS

This research project has lasted many years. Numerous people deserve thanks. Each work contains its own acknowledgements therefore a small set remains to be acknowledged here.

Thanks go to my wife, Susan, who gave support throughout, and to Ephraim Clark and Yacine Belghitar, who steered me through the final stages, giving advice on content, structure and process.
7. **The Context Statement**

7.1 For the PhD by public works, the general purpose of the context statement is to supplement work of intrinsic doctoral standard in such a way as to create, overall, a PhD-equivalent submission. 

7.2 The context statement should be used creatively to supplement, or make explicit, or fill in the gaps left by the public works to ensure a PhD-equivalent submission. Thus, the content of the context statement will vary from case to case dependent on the character and content of the public works submitted. It is, therefore, not possible to put forward a template for the context statement or to prescribe definitively its content or character.

However, for the award of a research degree, the academic apparatus of the research must be clearly demonstrated. Thus, the following elements must be made explicit in the context statement where these are not already evident in the public work itself: an extensive reviewing of the existing literature or work in the field, an account and critique of the research methodologies used, evidence and exemplification of claims made that the selected work is equivalent to a conventional (thesis) route PhD ..., a description of the limitations of the research, a clear argument as to how the works relate to each other (in all cases where more than one work is submitted), and a full bibliography.

These elements are not the only ones that supervisors may require as part of a context statement. The idea is to use the context statement in a flexible way so that the qualities and elements present in a research degree by thesis are also present in a research degree by public works. In this way, the submission, taken as a whole, (public works and context statement) is equivalent to a research degree by thesis.

Extract from:
Research Degrees, Public Works, Guidance Notes, Middlesex University
RD/PW/GD/JUN05
Preamble

In the context of a PhD by Public Works, the relationship between supervisor and supervised is shorter than the norm. Consequently some tacit knowledge of the research process is missing because the ideas are developed, and the bulk of the works written, before the relationship begins. Examples of such tacit knowledge include:

- An awareness of false starts and mistakes
- Provision of critiques of style and content through the many drafts of documents
- Witnessing verbal presentations and contributions to discussions

To address the issue, this context statement contains not only an account of the literature, the works and their core ideas, but also material not always included in a regular thesis. This material includes:

- A detailed account of the origin of the project incorporated into the chronological overview of works (Chapter 3)
- A more detailed account of the research methodology than is usually offered (Appendix A)
- A perspective on why ‘multiple interest rate’ research developed as and when it did (Appendix B)
Chapter 1

Preliminaries: motivation, summary of the results
and an annotated list of the submitted works

Motivation

The time value of money (TVM) equation is a key equation in finance. It takes the form of an \( n \)th order polynomial having \( n \) roots. In finance it is usual to calculate and use only one root (interest rate). The remaining \( (n-1) \) roots are mostly complex or negative, and are usually discarded. The works in this thesis contain analysis of the discarded, unorthodox roots and evidence that they should not be ignored. The roots have financial use and meaning.

The original motivation to study all roots, including the unorthodox, lay in econometric advice that data is valuable and should not be discarded lightly. This advice usually applies to the treatment of outliers in a data set. A recent example is Zellner (2007). Outliers, unusual or exceptional observations, are often few in number compared with the total data set, a rule of thumb being less than 1%. Though scarce, their inclusion, or not, can make a large difference to econometric results. They can be significant facts demanding explanation, possibly leading to revisions in the theory. For this reason Zellner writes: ‘In view of the potential importance of unusual and surprising data, it is troubling to see how often outlying observations are discarded without thought or averaged with usual observations by means of “robust” techniques.’

In this thesis, the situation is different from the econometric in several ways:

- The ‘unusual or exceptional’ data is output, not input;
- Complex solutions can comprise the bulk of solutions to a financial polynomial. In some circumstances there are thousands of roots and only a few are real.

The profusion of complex solutions to the TVM equation makes their exclusion from most economic and financial analyses both ‘troubling’ and difficult to comprehend. Some thoughts about the historical and technical
reasons for their exclusion are discussed in Appendix B. The possibility that they might usefully be included provided initial motivation for this research.

In the beginning the objective of the project was vague to the extent that no clear picture existed of what use or meaning would emerge. It was not certain that there would be any results at all.

Below are three summaries of what emerged. The first summary is an overview from a mathematical perspective. The second provides a more focused overview from an economic and financial perspective. The third describes results obtained in the context of one topic in finance: capital budgeting. Together, the summaries convey the main ideas of the thesis. They are written with the benefit of hindsight, constituting an orderly, backward look at a twenty-year research project.

A mathematical perspective

The TVM equation is a polynomial. Eq. (1) is a simple example. The cash flows, $c_i$, are treated as parameters. The variable $p$ is present value; it is a function of the interest rate, $r$.

$$p = \sum \frac{c_i}{(1 + r)^t}$$ (1)

Aleksandrov et al. (1969) summarize a well-known result about factorization of polynomials.

'If we accept without proof the so-called fundamental theorem of algebra that every equation $f(x)=0$, where

$$f(x) = x^n + a_1x^{n-1} + a_2x^{n-2} + ... + a_n$$

is a polynomial in $x$ of given degree $n$ and the coefficients $a_1, a_2, ..., a_n$ are given real or complex numbers, has at least one real or complex root, and take into consideration that all computations with complex numbers are carried out with the same rules as with rational numbers, then it is easy to show that the polynomial $f(x)$ can be represented (and in only one way) as a product of first-degree factors $f(x) = (x - a)(x - b)...(x - l)$ where $a, b, ..., l$ are real or complex numbers.'
Furthermore:

‘Multiplying out the expression \((x-a)(x-b)(x-c)...(x-l)\) and comparing the coefficients of the same powers of \(x\), we see immediately that

\[\begin{align*}
-a_1 &= a + b + c + ... + l, \\
a_2 &= ab + ac + ... + kl, \\
-a_3 &= abc + abd + ...,
\end{align*}\]

\[\vdots
\]

\[\pm a_n = abc...l
\]

which are Viete’s formulas.\(^1\) Aleksandrov et al. (1969)

These results, the factorization of a polynomial, and Viete’s formulas linking its parameters with its zeros, enable a transformation of the TVM equation. In Eq. (1), present value, \(p\), is expressed as a function of the interest rate and the parameters. After transformation, present value is expressed as a function of the parameters and all possible interest rates that solve Eq. (1). Eq. (2) is an example of the transformed Eq. (1).\(^2\) Since this is a summary, the transformed equation is stated without proof.

\[
p = \frac{\sum c_i}{1 + \prod r_i}
\]

(2)

Another transformation is possible. Assume \(r = 0\) and the resulting value of \(p\) is \(p^* = \sum c_i\). Substitute \(p^*\) into Eq. (2). After some manipulation, Eq. (3) results.

---

\(^1\) Viete is sometimes written Vieta. Viete’s formulas are also known as Newton’s relations; see, for example, Weisstein (2003). There is discussion in the mathematical literature about priority of discovery for the formulas. See Funkhouser (1930) who attributes the idea to Viete (1540-1603) and Girard (1595-1632).

\(^2\) The reader is alerted to a notational issue. For each set of cash flows \(c_i\) for \(i = 1\) to \(n\) in Eq. (1), there are \(n\) possible values of \(r\) that solve for that set. Therefore each of the following equations is true:

\[
p = \sum \frac{c_i}{(1 + r_i)^j}; \quad p = \sum \frac{c_i}{(1 + r_j)^i}; \quad \ldots \quad p = \sum \frac{c_j}{(1 + r_i)^j}.
\]

In Eqs. (2) and (3) the product contains no subscript; it is understood that the product is of all possible values of \(r\) satisfying Eq. (1). In subsequent equations this convention is dropped: the subscript \(i\) is also used as a counter referring to the \(i^{th}\) value of \(r\). It is understood that the two things being counted are conceptually different - the \(i^{th}\) value of \(r\) is not paired with the \(i^{th}\) value of \(c\) - although there are \(n\) values of both because, in both cases, the size of \(n\) reflects the order of the polynomial.
\[
\frac{\Delta p}{p} = \prod |r|
\]  

(3)

Eq. (3) states that the proportionate change in the value of \( p \) when the interest rate falls to zero is the product of the differences between every possible value of the original interest rate and the new value of zero.

Viete’s formulas demonstrate that an equation defined by its parameters, Eq. (1), can be transformed into one defined by its zeros, Eq. (3). That all possible interest rates can be incorporated into a financial equation is the first significant idea.

An economic and financial perspective

Perhaps the most interesting question that can be asked of Eq. (1) is what happens to present value, \( p \), when the interest rate, \( r \), changes. A simple example, when the interest rate falls to zero, is described in the last section. The more general question is what happens when the interest rate shifts to an arbitrary new position, not necessarily zero.

A way forward is to apply differential calculus; one of the first examples is Hicks (1939). A difficulty with this approach is that the relationship between \( p \) and \( r \) in Eq. (1) is non-linear, therefore the first differential of \( p \) with respect to \( r \) provides a linear approximation to the true relationship. The larger the actual change in \( r \), the greater the error. An approach is needed to take the analysis ‘around the bend’.

The orthodox approach is to incorporate the second differential into the analysis. This provides a closer approximation. Even closer approximations are obtained with the addition of the third, fourth and higher order terms in a Taylor series expansion of Eq. (1). Unfortunately increasing accuracy is provided at the expense of increasing complication and opacity. The extensive literature about duration and convexity in bond mathematics provides a good example of the difficulties encountered when applying calculus to the question of interest elasticity of present value. Works 3, 4 and 7 contain references to this literature.
There is an alternative approach. Eq. (1) is a levels equation. Eq. (4) is a second equation in levels that shows a shift from \( r \) to \( R \) producing a change in present value from \( p \) to \( P \).

\[
P = \sum \frac{c_i}{(1 + R)^i}
\]  

(4)

A comparison of (1) with (4) opens the possibility of deriving a difference equation in which the change in present value \((P - p) = \Delta p\) is expressed as a function of the change in the interest rate \((R - r) = \Delta r\). Such a derivation is possible. Eq. (5) is derived from Eqs. (1) and (4) using the factorisation theorem and some algebraic manipulation. It states that the relative shift in present value is the discounted product of all incremental shifts in the interest rate.

\[
\left| \frac{\Delta p}{p} \right| = \prod \left| \frac{R - r_i}{(1 + R)^n} \right|
\]  

(5)

This result can be developed further. The relationship between the new rate of interest \( R \) and the old rate, \( r \), can be expressed as \((1 + r) = (1 + R)(1 + m)\) in which \( m \) is the interest rate that marks up or marks down the old rate to the new. Assuming a single new rate and \( n \) possible old rates, there must be \( n \) multiplicative mark-ups (or mark-downs) of \( R \) relative to the \( n \) values of \( r \) that solve Eq. (1), i.e., \((1 + r_i) = (1 + R)(1 + m_i)\) for all \( i \) from 1 to \( n \). This relationship can be restated as \( \prod \left| \frac{R - r_i}{(1 + R)^n} \right| = \prod |m_i| \). If the last result is applied to Eq. (5) then the latter becomes Eq. (6).

\[
\left| \frac{\Delta p}{p} \right| = \prod |m_i|
\]  

(6)
Eq. (6) states that the proportionate change in present value is the product of all multiplicative shifts in the interest rate. It is a generalisation of Eq. (3) when $R$ is set to zero in the expression $(1 + r) = (1 + R)(1 + m_i)$.

Eqs. (5) and (6) are combinations of the information in Eqs. (1) and (4) from which the parameters, $c_i$, have been eliminated. The equations are expressed entirely in terms of incremental or multiplicative shifts in interest rates. When employed simultaneously, the entire cluster of interest rates (implied by the zeros of Eq. (1)) contains implicitly all information about the cash flows (parameters of Eq. (1)). As observed earlier, Viete’s formulas imply that the two sets of information are equivalent. Thus, the construction of a difference equation requires the inclusion of differences between the new interest rate and all initial interest rates. One outcome of this exercise is the realization that the independent variable is more complicated than it appears. The orthodox shift in the interest rate, $\Delta r$, is, by itself, not sufficient to explain the change in present value, $\Delta p$. All possible values of $\Delta r$ are needed.

The switch from levels to differences described above is the second significant idea. The combination of the various components – factorisation, Viete’s formulas, levels to differences – produces a fresh perspective on the impact of a shift in the interest rate on present value. The perspective leads to additional ideas; below is a selection from one example of the several examples offered in this thesis.

A specific example: capital budgeting

The nature of the relationship between NPV and IRR is reviewed and the meaning of the two concepts reconsidered.

First, a new equation is produced for NPV per dollar invested. The orthodox IRR criterion is the mark-up of the single orthodox IRR over the cost of capital. In this thesis it is shown that NPV per dollar invested consists of all possible mark-ups of IRR over the cost of capital. The academic argument in
favour of NPV is supported because NPV carries more information than IRR alone.

Secondly, in a recent paper on capital budgeting, Magni (2010) writes about IRR that ‘... three issues remain unsolved:

- Complex-valued return rates and complex valued capitals are devoid of meaning;
- Project ranking with IRR is not compatible with NPV ranking;
- The IRR may be applied only if the cost of capital is constant.’

The three issues correspond to IRR pitfalls 2, 3 and 4 in the classic text on corporate finance by Brealey et al. (2009). The analysis in this thesis shows that these pitfalls do not exist.3

Eqs. (2), (5) and (6) are unfamiliar mathematical expressions that determine familiar financial concepts. At this stage they are difficult to interpret because most of the elements on the right-hand side lack financial meaning. Another result in this thesis is the attribution of meaning to these unorthodox elements: the multiple differences between interest rates. The meaning attributed to the mark-ups of the multiple, unorthodox IRRs over the cost of capital leads to a fresh interpretation of the meaning of an orthodox IRR relative to the cost of capital. The reinterpretation implies that the mark-up of the orthodox IRR over the cost of capital cannot be an investment criterion. Instead, the mark-up is interpreted as a unit of measurement of value that is a component of an investment criterion (NPV). Given this result, the NPV versus IRR debate dissolves.

If the results described above are accepted, then the chapters on capital budgeting in finance and economics texts require revision. Furthermore, the academic arguments against practitioners using IRR as an investment criterion are buttressed, and the academic case for using NPV is enhanced.

The three summaries above give the flavour of the thesis. An annotated list of the submitted works follows.

---

3 Pitfall 1 is not addressed in this thesis. The pitfall is that, by itself, the IRR does not distinguish between borrowing and lending. In the opinion of this author the pitfall can be dismissed. No practitioner takes investment decisions based on a single statistic. Whether a project involves borrowing or lending is obvious from the entire project proposal.
An annotated list of the submitted works

1 Osborne, M. 1990. Financial chaos,  
*Management Accounting*, 68(10)

2 Osborne, M. 1993. Chaos and the internal rate of return,  
*Computers in Higher Education Economics Review*, 7(19)

Works 1 and 2 are preliminary papers that expose a practical issue in calculating interest rates. When a cash flow is such that multiple real interest rates exist, the first guess in the search procedure is critical in determining which of the multiple real results is output. The link between guess and output is deterministic, but chaotic, making caution necessary when calculating an interest rate in the presence of multiple real solutions. The analysis in the two papers is confined to orthodox interest rates on the real number line (in one dimension). However, the nature of the link between inputs and outputs is exposed using Newton’s method to locate the complex roots (in two dimensions). This initial contact with the complex roots led to the project’s conception. Work 1 is written for accountants and Work 2 for economists.

3 Osborne, M. 2000. Visualising financial concepts in the complex plane,  
*Computers in Higher Education Economics Review*, 14(1),  
http://ideas.repec.org/a/che/chepap/v14y2000i1p4-8.html

Modified duration is an estimate of the interest elasticity of the price of a bond, a risk measure in fixed income analysis. In Work 3 the equation for modified duration is recast as a combination of distances between all roots of the TVM equation. To the author’s knowledge the result is only the second conversion of a financial concept into a form that employs all rates, including the unorthodox rates. (The first is Dorfman (JoF 1981) which contains a growth equation employing all rates.)

Modified duration yields approximations to the true value of the price elasticity of a bond. Since Macaulay (1938), analysts have searched for expressions that yield better results. Work 4 extends Work 3 to provide, for the first time, an expression that is accurate. When the interest rate does not change, the new expression is equivalent to modified duration. When the interest rate changes, modified duration requires higher order corrections while the new expression continues to provide accurate results.


Work 5 demonstrates a link between the locations of the roots and the statistical properties of the coefficients in the TVM equation. The base, or reference, position is that of a par bond. The financial implications of this information have yet to be fully explored.


Work 6 employs several new ideas. First, it employs multiplicative rather than additive shifts in the interest rate. Secondly, it introduces the twin notions of the special form of a polynomial, and the special relationship between the coefficients and roots of the special form. These ideas permit three developments: simpler and shorter proofs of the core equations; simpler final equations that make possible an attribution of meaning to the unorthodox
solutions; an argument that the key insights are robust in the face of non-parallel shifts of non-flat yield curves.\(^4\)


Work 7 gathers together and expresses more clearly the insights of Works 2 and 3. It also incorporates answers to the computational issues first posed in Osborne (2001b).


Work 8 offers a resolution to a long-standing debate in capital budgeting, that is, which investment criterion is best, NPV or IRR. The work exposes a hitherto unknown expression linking NPV and IRR, namely that NPV per dollar invested is the product of the mark-ups of all IRRs over the cost of capital. The expression shows that NPV contains more information than the orthodox IRR alone, supporting the academic conviction that NPV is the better investment criterion. The resolution is amicable because the additional information consists of all the unorthodox IRRs, therefore the concept of IRR remains fundamental as a component of NPV and cannot be discarded.


Work 9 applies the new methodology to a long-standing debate in economic theory. Reswitching is a part of the Cambridge controversies in the theory of capital. In the context of the Sraffa-Pasinetti reswitching example (see Pasinetti, 1966), all \(n\) possible solutions for the interest rate are used to produce a

\(^4\) Work 6, lodged in SSRN in September 2004, was not the first work to introduce these ideas. Osborne (2004a), lodged in June 2004, was the first, but it did so in the context of the NPV-IRR debate. The latter paper is not included in the works submitted for this thesis because most of its results were published in Work 8 – Osborne (2010a) – that is included.
new equation in which reswitching does not occur.


Work 10 suggests an answer to the question posed in Work 8 concerning the meaning of multiple interest rates, including a reinterpretation of the meaning of an orthodox interest rate. The reinterpretation implies that the orthodox IRR is not an investment criterion.

The research continues.

• Working papers await revision and publication (Works 6, 9 & 10);
• Existing works await logical sequels, e.g., Work 7;
• Results need writing up (on retail banking);
• Sub-projects await development, e.g., an extension to Work 5, and a project to make the analysis stochastic;
• There are connections to be made with the work of others, including Dorfman (1981), Bosch et al. (2007) and Barney and Danielson (2004).

This list of ideas about the future development of the research is amplified in Chapter 4, Further Work and Open Questions.
Chapter 2
The multiple interest rate literature

Introduction

The essence of this thesis is a reinterpretation of discounted cash flow (DCF) analysis. DCF techniques find application in numerous topics in economics and finance. The submitted works and this context statement reflect this fact: the topics studied include capital budgeting, bond mathematics, capital theory, and retail and wholesale credit. Each work contains its own list of references appropriate to the topic under investigation. At no time, however, does any work trace the history of ‘multiple interest rate’ analysis. This shortcoming is now remedied in the context of the literature on capital budgeting.

The beginning

Irving Fisher was one of the earliest researchers to write extensively about the rate of interest. Fisher makes much use of the time value of money equation in his classic book, *The Rate of Interest* (Fisher, 1907). In the book he proposes three decision criteria for the selection of one investment project from among several. Fisher saw them as equivalent.

‘(1) Out of all options that one is selected which has the maximum present value, reckoned at the market rate of interest; (2) Out of all options that one is selected of which the advantages over any other outweigh, in present value, its disadvantages, when both are discounted at the same rate of interest; (3) Out of all options that one is selected which, compared with any other option, yields a rate of return on sacrifice greater than the rate of interest.’ (Fisher, 1907)

The last criterion is restated as follows. The cash flow of project A can be compared with the cash flow of project B by subtracting flow B from flow A to form a third cash flow. The internal rate of return (IRR) of this third flow of differences, A minus B, is Fisher’s rate of return on sacrifice. The rate should exceed the cost of capital if project A is to be preferred to project B. Pairwise comparisons applied to a portfolio of projects produce a project ranking and determine best project.
In *The Rate of Interest* Fisher does not state explicitly that solving for the ‘rate of return on sacrifice’ yields multiple solutions. It is likely he was aware of them but chose not to analyse them.\(^5\)

*The Theory of Interest* (Fisher, 1930) is a considerably amended and re-titled version of the earlier book. It is probable that, here too, Fisher chose not to analyse all possible solutions. This assertion is made for two reasons.

First, many of Fisher’s analyses are of two-period cash flows, or of multi-period cash flows divided into many two-period elements. A different interest rate is calculated for each element. Under this procedure all solutions to the two-period rates must be real. The only possible ‘unusual’ solutions in this circumstance are interest rates that could be zero or negative.\(^6\) This ‘divide and rule’ approach was possibly a strategy to avoid the difficulties of a full multi-period analysis, although he did give other reasons why he thought it best to theorise this way (see Fisher (1907), pp. 383-385).

Secondly, after venturing into the geometric analysis of incomes spread over three years in the appendix to chapter 10, he states the following:

“To proceed beyond three years would take us into the fourth dimension and beyond. Such a representation cannot be fully visualized, and therefore has little meaning except to mathematicians.” (Fisher, 1930)

In this way Fisher’s work lays the foundation for much of the research to come. First, there is no attempt to explore the multiple solutions to a multi-period problem in the manner described in this thesis. The most likely reasons

---

\(^5\) Hirschleifer (1958) writes that Fisher was aware of the ‘multiple interest rate’ issue at an early stage and refers to a footnote on page 155 of *The Rate of Interest*. In Hirschleifer’s words: ‘Where more than single-sign alternation takes place, he [Fisher] suggests the use of the present-value method rather than attempting to compute ‘the rate of return on sacrifice.’’ There is no such footnote in my facsimile of the first edition of Fisher’s book and I can find no such reference elsewhere in the book. It is likely that Hirschleifer is referring to a later edition than the original 1907 version.

\(^6\) *The Theory of Interest*, part III, chapter XI, Second Approximation in Geometric Terms, section 9, Can Interest Disappear, and also chapter XIII, Second Approximation in Terms of Formulas, section 10, Zero or Negative Rates of Interest.
for this are explored in Appendix B. Secondly, Fisher adopts a strategy that recurs in the literature, namely, coping with multi-period cash flows by breaking them into their many two-period elements, each of which has one solution for the interest rate.

In a paper describing *The Theory of a Single Investment*, Boulding (1935) also glosses over the problem of multi-period analysis and its multiple interest rates. Early in his exposition he sets up an equation to demonstrate the calculation of the internal rate of return and writes ‘... the solution is not mathematically a simple one, but it is clear on general grounds that in most practical cases a single solution can be found.’

In (1936a) Boulding publishes again on the time value of money equation and the internal rate of return, and he attempts to introduce a concept called ‘time spread’. In an appendix, he refers to the difficulty of calculating the internal rate of return and describes a procedure for finding a solution.

Wright (1936) replied to Boulding’s article and provided several arguments against the concept of ‘time spread’. Wright’s reply is interesting because, possibly for the first time in the academic literature, a researcher points out the existence of all $n$ solutions in unequivocal terms. He writes:

‘The third argument [against time spread] depends upon what is commonly called "the fundamental theorem of algebra"; the theorem asserts that any algebraic equation of degree $n$ has $n$ solutions. Applied to Mr. Boulding's definition of the interest rate, it means that several different and reasonable values may be obtained for $i$ …’ Wright (1936).

The mention of the fundamental theorem of algebra shows that Wright must have been aware of all solutions, including the complex. In the first sentence of the quote above, Wright acknowledges all solutions. In the second sentence, on the assumption that ‘real’ equals ‘reasonable’, he implicitly

---

7 Time spread ‘... is the length of time for which a quantity equal to the sum of the positive payments would have to grow (at the rate of return) in order to equal the sum of the negative payments’, Boulding (1936). The concept is not important for the current analysis therefore its definition is confined to this footnote.
dismisses all but the real solutions: ‘it means that several different and reasonable values may be obtained for i...’

In his reply to Wright, Boulding takes a stance that will - with a few exceptions - characterise most research on the subject for the next 70 years.

‘Now it is true that an equation of the nth degree has n roots of one sort or another, and that therefore the general equation for the definition of a rate of interest can also have n solutions, where n is the number of ”years” concerned. Indeed, if we adopt continuous compounding, as in strict theory we should, the theoretical number of solutions is infinite! Nevertheless, in the type of payments series with which we are most likely to be concerned, it is extremely probable that all but one of these roots will be either negative or imaginary, in which case they will have no economic significance.’ Boulding (1936b).

Samuelson (1937) discusses both papers by Boulding. When discussing the 1936 papers he refers to the possibility of ‘a multiplicity of solutions’ but does not elaborate on the nature of the multiplicity.

Thus, the literature first displays a full awareness of the issue of multiple solutions to the time value of money equation during the 1930s. From the beginning, most of the multiple solutions, especially the complex, were dismissed as having no economic significance; therefore the attention of researchers focused on the real solutions.

*The development of the multiple-IRR literature in the twentieth century*

There is a pattern to the development of ideas about multiple rates. A brief history is as follows:

• 1930s: In the first stage, as noted above, researchers acknowledged the existence of n rates, although most of the solutions, the complex solutions, were dismissed as having no economic significance. Early researchers focused on the real rates, negative and positive.

• 1950-1999: In the second stage the focus progressively narrows. This trend has two features. First, there are attempts to limit the range of ‘legitimate’
rates found along the real number line, and attempts to isolate the most relevant, real, positive rate from the range. Secondly, there is literature showing how to restructure cash flows so as to force the production of only one positive, real interest rate.

- 2000 and after: In the third stage the trend reverses and there emerges an interest in the entire range of rates.

This brief history is described in more detail in the remainder of this chapter and the next.

Magni (2010) summarises the literature on multiple rates of return in capital budgeting. Table A shows a chronological history of the literature grouped into logical categories. The categories are based on the discussion in Magni (2010) with additions by the author. The following paragraphs discuss the content of Table A.

Column 1 of Table A lists articles published between the 1930s and the 1960s. They examine possible solutions for the roots in the restricted range from zero to plus infinity along the real number line, implying a range of interest rates from minus 100% to plus infinity. Some authors identify the restrictions on the structure of the cash flows necessary to guarantee a single, real-valued rate of return. From the beginning, the story is of restriction.

It is likely that the lower bound of the ‘permissible’ range of rates was imposed for the same reason that modern financial calculators and spreadsheets have their output from IRR calculations limited to solutions greater than minus 100%. The search for a root of a TVM equation is the search for the solutions of \((1+r)\). At the point \((1+r) = 0\), any discounted value goes to infinity. Such a point is a barrier. Anything on the left or negative side of the barrier has traditionally been ignored. The search is simpler if it is restricted to the right side of the barrier, i.e., \((1+r) > 0\) which implies \(r > -1\) (r greater than minus 100%).
Beginning in the 1950s, a second series of papers appear that further restrict the range of ‘permissible’ solutions for the root \((1+r)\) to between plus one and plus infinity, i.e., they rule out negative interest rates (column 2 of Table A).

The seminal paper of Lorie and Savage (1955) belongs to this group. They were first to point out the possibility of inconsistent ranking of investment projects by the NPV and the IRR criteria. Lorie and Savage also introduced the oil-pump problem. The problem produces two real interest rates, both of which are feasible. This situation prompts and facilitates discussion about which rate is most relevant, and why. For this reason the oil-pump problem is often quoted in the literature.

Also in this group, Hirschleifer (1958) and Bailey (1959) follow in the tradition of Fisher by recommending analysis of the many individual two-period returns in a multi-period investment thereby sidestepping the issue of multiple solutions.

Two of the papers in this group explicitly mention the possibility of complex solutions: Hirschleifer (1958) and Feldstein et al. (1964). They point out that wholly positive or wholly negative cash flows, with no change of sign at all, give rise to imaginary solutions. Hirschleifer concludes that the idea that IRR ‘represents a growth rate in any simple sense cannot be true.’ Feldstein et al. observe only that ‘the examples given are rather peculiar.’

---

8 Their reference to imaginary solutions is not strictly correct. The examples they provide have complex solutions that contain both real and imaginary components.
Table A: A history of multiple-rate analysis seen through the literature on capital budgeting (based on Magni (2010) with additions by the author).

<table>
<thead>
<tr>
<th>Decade</th>
<th>1 Types of project having a real-valued rate between $(-1, \infty)$</th>
<th>2 Types of project having a real-valued rate between $(0, \infty)$</th>
<th>3 Project truncation to obtain one real-valued IRR</th>
<th>4 MIRR based on different lending &amp; borrowing rates</th>
<th>5 Choosing the relevant IRR from real alternatives</th>
<th>6 All IRRs included in the discussion</th>
<th>7 All IRRs employed as components of other concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1930-39</td>
<td>Fisher (1930)</td>
<td>Wright (1936)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wright (1936)*</td>
<td>Boulding (1936b)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Samelson (1937)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940-49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950-59</td>
<td>Pitchford et al. (1958)</td>
<td>Lorie et al. (1955)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Solomon (1956)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hirschleifer (1958)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bailey (1959)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960-69</td>
<td>Kaplan (1965)</td>
<td>Feldstein et al. (1964)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teichroew et al. (1965a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teichroew et al. (1965b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Jean (1968)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>De Faro (1978)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bernhard (1977, 1979)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pratt et al. (1979)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gronchi (1986)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990-99</td>
<td></td>
<td>Promislows et al. (1996)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- The papers bearing an asterisk (*) are examples of early papers that mention the possibility of complex roots but dismiss them as economically irrelevant.
- The literature is not easily divided into neat categories. Whatever the criteria, some papers straddle several categories. For example, Lorie and Savage (1955) anticipate in a footnote the work of Cannaday et al. (1986); Gronchi (1986) discusses both project truncation and different lending and borrowing rates; Teichroew et al. (1965) assume different lending and borrowing rates but do not go so far as to recommend MIRR.
- Osborne (2010)** was first deposited in SSRN in 2004, followed by the unpublished paper Osborne (2004a). Core insights in the two works were first described in Osborne (2000, 2001a, 2005). These earlier papers are about bond mathematics therefore they are excluded from this table about capital budgeting.
- Bosch et al. (2007) is a paper that does not fit easily into the table but it merits discussion and is described in the text.
Several authors, including Ramsey (1970), examine multi-period cash flows that possess several changes of sign and begin the practice of invoking Descartes’ law of signs to determine the number of real rates.\(^9\)

In the 1950s the search intensified for a unique rate of return. In 1959, the first of a series of articles is published that invokes project truncation as a way to control the structure of the cash flow such that a unique, positive, real solution is guaranteed (column 3 of Table A).

Karmel (1959), in a response to Pitchford and Haggar (1958), shows …

‘… that, if a project is terminable at any stage during its lifetime and provided that the scrap value is always non-negative, the marginal efficiency of the truncated project expected to have the highest marginal efficiency will be a unique value.’

Soper (1959) explains the same idea in different words.

‘If the investor is aiming at a maximum rate of profit on his investment – a maximum \( r \) – then this will always cause him to discriminate between different “lengths of life” for the investment, choosing that length which includes the maximum number of consecutive yields which are still consistent with a discounting equation with one, and only one, positive root.’

A famous paper in this series is by Arrow and Levhari (1969) who revise the conclusion of the earlier papers. They argue that choosing the truncation period in order to maximise net present value is superior to choosing the truncation period in order to maximise the rate of return. If, ‘with a given constant rate of discount, we choose the truncation period so as to maximise the present value of the project, then the internal rate of return of the truncated project is unique.’

\(^9\) Descartes’ law of signs states that the number of changes of sign in the coefficients of a polynomial is greater than or equal to the number of real, positive roots (see Weisstein, 2003). The law also states that if the signs are reversed on all the coefficients attached to odd powers, then the number of changes of sign in the coefficients is greater than or equal to the number of real, negative roots. A negative root in a financial polynomial implies the existence of an interest rate less than minus 100%. As indicated in the text, most financial calculators and spreadsheets will not calculate such interest rates and they are usually ignored.
In the light of the ideas in this thesis, truncation is not an answer to the issue of multiple interest rates, only to the issue of multiple real, positive solutions. While an equation incorporating a suitably truncated cash flow might yield only one positive, real interest rate, there are still $n$ solutions to every equation, and the questions about their use and meaning remain.

The next stage in the search for a unique rate of return was the creation of the concept of MIRR – modified internal rate of return (see column 4 of Table A). The MIRR literature could be said to begin with Teichroew et al. (1965a,b) who suggest the problem of multiple real rates associated with ‘non-simple’ projects can be eliminated if different rates for project investment and project finance are used. Teichrow et al. did not introduce the concept of MIRR, but they did introduce the idea of separate borrowing and lending rates according to whether the project balance is positive or negative.

Lin (1976) and Athanasopoulos (1978) begin the MIRR literature proper. They assert that the IRR technique assumes all returned cash flows are reinvested at the IRR, and that the NPV technique assumes the returned cash flows are reinvested at the cost of capital. Researchers who favour the use of MIRR assert that neither assumption is realistic. They suggest that if all project inflows are compounded forward to the terminal date at some reinvestment rate (preferably more realistic than the IRR), and all outflows are discounted back to the start date at some finance rate (possibly the cost of capital), then the terminal value relative to the start value is a monotonic function possessing one positive, real interest rate: MIRR.

A number of commentators object to this approach: see Lohmann (1988), Keef and Roush (2001) and Eagle et al. (2008). They argue that the assumptions about reinvestment are ‘fallacious’, that neither the NPV nor the IRR criteria assume anything about how the incoming funds are reinvested. In the two latter articles the authors trace the assumption back to a confusing discussion in Solomon (1956). Despite these objections to MIRR, articles continue to appear in support of it, e.g., Chang & Swales (1999) and Kierulff (2008).
Like the truncation technique, the development of MIRR is an attempt to reconfigure the problem to give only one positive, real interest rate as output. Even if the objections of Keef and Roush and Eagle et al. are set aside, from the perspective of the current research the concept of MIRR is not helpful. There might be only one positive, real MIRR, but it is still possible to calculate $n$ values for MIRR. Redefining how the ‘relevant’ rate of interest is calculated does not remove the question about the use and meaning of every possible interest rate.

In the 1980s several papers appear that advocate criteria for choosing the ‘correct’ rate of return from among multiple real solutions (column 5 in Table A). Cannaday et al. (1986) suggest the following criterion: if the net future value function has a negative derivative at a relevant root, and the associated IRR is greater than minus one, then the IRR is appropriate. They point out that the criterion suffers from a crucial drawback, namely that, for some projects, it is still possible that more than one real positive IRR could satisfy the criterion. Hajdasinski (1987) comments that the analysis of Cannaday et al. needs adjustment according to whether the project involves lending or borrowing, but nevertheless goes along with the criterion, drawback included.

Continuing with the theme of identifying the single, relevant rate of return, Zhang (2005) proposes a simple technique for accepting or rejecting a project using IRR when multiple real IRRs exist. The technique involves counting the number of real rates of return greater than the cost of capital; if the number is even then reject the project (because NPV must be negative) and if the number is odd then accept the project (because NPV must be positive). Zhang concludes that ‘while the proposed method is not computationally easier than the NPV method, it serves as a simple way of retaining the use of the IRR without having conflicts between the two methods.’

Zhang’s technique can also be applied to pairwise comparisons of projects within a portfolio. This is done by evaluating the viability of the combined cash flows (one cash flow minus the other) in the style of Fisher. In this way the technique ranks the projects in the portfolio.
Zhang goes on to suggest how the ‘relevant’ IRR can be identified. The relevant IRR is - in the case of project acceptance - the IRR above and closest to the cost of capital. In the case of project rejection it is the IRR below and closest to the cost of capital.

Counting the number of IRRs that exceed the cost of capital is an arbitrary method. It is justified only on the grounds that any investment decision based on it is supported by the NPV criterion. If the NPV criterion is to be the ultimate arbiter of investment superiority then why try to salvage IRR as an investment criterion. Zhang wonders why the search for the rate of return has persisted over the years and concludes that it is because practitioners find a single rate of return intuitive and easy to use. Zhang’s decision criterion, however, is based on the number of IRRs exceeding the cost of capital, therefore, in general, it is not a criterion based on a single rate of return.

*The twenty-first century: the inclusion of complex solutions*

With one exception, Dorfman (1981), the research prior to 2000 is characterised by gradual restriction. As detailed above, the restriction is either on the structure of the cash flow in order to force one, real, positive rate of interest, or on the range of permitted interest rates with an emphasis on the choice of one of them: the ‘relevant’ rate.

A new focus appears in the ‘multiple interest rate’ literature after the millennium. The research opens up to consider all possible cash flows (no truncation) and all possible solutions, including the complex. Some reflections on why the reversal took place at this particular time are in Appendix B.

Dorfman (1981), Hazen (2003), Hartman and Schafrick (2004), Pierru (2010), Magni (2010) and the series of works that form this thesis, contain explicit discussions about the complex valued results (columns 6 and 7 in Table A). Each author takes a different approach. Some (Dorfman, Hazen and Pierru) take such
rates seriously by employing them in some way. Others (Hartman and Schafrick, and Magni) discuss them but find ways to dismiss them.

The ideas in Dorfman (1981) are relevant to this thesis for many reasons. He is the first researcher to employ all solutions to the TVM equation. However, there are crucial differences and similarities between his work and the works in this thesis. The latter are not described until the next chapter; therefore discussion of Dorfman’s article is postponed to Chapter 4.

Hazen (2003) sees IRR as an investment criterion and tries to find such usage for each individual IRR. He describes the procedure as ‘roundabout’, although it appears to work for any IRR. He writes ‘there is no need to discard “unreasonable” or “extreme” internal rates – all are equally valid’. All IRRs are used to convert the original cash flow for a project into $n$ alternative cash flows called ‘investment streams’, each stream employing one of the $n$ IRRs. He then determines the net present value of each investment stream by valuing it at the cost of capital. The original project is judged to be profitable, or not, according to whether an investment stream is profitable, or not. If the investment stream is a net investment (borrowing) and IRR is above (below) the cost of capital, then the investment stream is profitable and, therefore, the original project is profitable.

The important point is that Hazen makes explicit use of all IRRs, including the complex valued. He considers only the real part of complex valued IRRs and only the real part of complex valued investment streams. In doing so he reaches the same conclusions as he does using the real valued IRRs. Moreover, it does not appear to matter which of the many IRRs (and its associated investment stream) is used to produce a decision, therefore the method needs to be applied only once; and a real-valued IRR will serve. It follows that, in Hazen’s analysis, complex valued results appear to be superfluous. Moreover, since the NPV criterion is simpler anyway, Hazen urges use of NPV. He shows that the IRR criterion can give invest/not-invest decisions for single projects that are consistent with the NPV criterion. He also shows that …
‘... the problem of multiple or non-existent [complex] internal rates of return – universally regarded as a fatal flaw for the IRR method – is not really a flaw at all, and can easily be dealt with conceptually and procedurally.’

The IRR criterion suggested by Hazen does present problems. First, when comparing mutually exclusive projects, the rank order produced by Hazen’s method can still conflict with the rank order from the NPV criterion; this is the well-known IRR pitfall introduced by Lorie and Savage (1955) that is much discussed in the literature; see Brealey et al. (2009). Secondly, he considers only the real components of complex valued IRRs, ignoring any information in the imaginary components. Thirdly, the method does not provide an interpretation of complex rates. In Hazen’s own words:

‘We are currently unaware of an economic interpretation of complex-valued rates of return or complex-valued investment streams, and without such an interpretation, it would be hard to justify any economic recommendation without resort to performance measures such as present value.’

Hazen (2003).

Hartmann and Schafrick (2004) adapt the procedure from Cannaday et al. (1986) outlined in the previous section. They partition the cash flow into lending and borrowing periods according to whether the first derivative of the present worth function with respect to the interest rate is positive or negative. The divisions between partitions occur where the first derivatives of the present worth function are equal to zero. The relevant IRR is found in the partition containing the cost of capital. The usual comparison of the IRR with the cost of capital is made. They explicitly acknowledge the existence of complex roots but do not employ them. They avoid the issue. In their words, ‘in our partitioning scheme, they [the complex roots] are removed from the analysis’. They do this removal by determining which partitions have complex roots assigned to them and collapsing together partitions such that a real root is assigned to every partition.

Despite removing the complex solutions from the analysis, Hartman and Schafrick wonder about them:

*In our partitioning scheme, they [the complex internal rates of return] are removed from the analysis and we assume that a project’s status (loaning or borrowing) does not change in the new partition. ... Unfortunately, while our
method of collapsing partitions allows for correct analysis in the presence of complex roots, it muddles our definition of a project being loaning or borrowing according to the slope of the present worth. This might signal that complex roots do have meaning, although we do not have an interpretation at this time. (Hartman and Schafrick, 2004)

Hartman and Schafrick’s methodology is arbitrary. They do not clarify why each complex solution for the interest rate should have a partition of the present worth function assigned to it. What is the reasoning behind ‘assignment’? Moreover, why should partitions associated with complex roots be collapsed into the surrounding partitions until a real rate can be assigned to the combined partition (other than because the researcher does not know what to do with the complex roots)?

Pierru (2010) examines complex interest rates in the context of a portfolio of two assets. ‘When a project involves the joint production of two outputs whose markets are subject to different risks, our approach allows the project’s cash flows to be discounted at a single (but complex) rate.’ The single complex rate is interpreted to represent several different real rates at the same time. A difficulty is that the interpretation is confined to a narrow range of applications. Pierru acknowledges this when he writes ‘we are aware of the apparently limited practical interest of the interpretations proposed …’

Magni (2010) also discusses complex rates explicitly, but adopts yet another approach to them. He circumvents the possibility of awkward, i.e., complex rates, by finding an average of many specially defined two-period rates of return embedded in the investment. In Magni’s words, this average internal rate of return (AIRR) approach ‘wipes out complex valued numbers.’

Magni follows in the footsteps of researchers such as Fisher, Hirschleifer and Bailey who also recommend breaking up one long investment into separate short investments, each of which has its own rate of return. As discussed earlier, the ‘many-short-periods’ approach sidesteps the issue addressed in this thesis: that of multiple rates of return, each one of which is valid over the whole life of a project.
Magni comments that ‘unfortunately, the venerable internal rate of return ... is not a reliable profitability index because it may not exist, multiple roots may arise, and, in general, is incompatible with the NPV.’ As will be seen, the analysis in this thesis suggests that Magni is correct about the orthodox IRR not being a reliable profitability index, but for reasons entirely different from those he suggests.

At this point we turn to the works that comprise this thesis. Elucidation of their approach is in the next chapter and in the works themselves. The main features of the approach are anticipated here. They are gathered below in summary form.

No restrictions

No restrictions are placed on the length and structure of the cash flows or on the number of ‘permissible’ solutions. All cash flows are permitted; all possible solutions for the interest rate are calculated.

All interest rates have utility

The solutions for all the multiple interest rates are not used as rates per se; rather, they are employed as elements in another economic concept. This change in usage is the difference between columns 6 and 7 in Table A. The new usage means that each rate is as important as any other rate and no rate should be discarded.

Real numbers, not complex

The complex solutions to the TVM equation are not employed in their raw, complex form. Rather the entities employed are absolute differences between complex solutions, or between complex solutions and other salient points. Thus, the equations are expressed in terms of differences between interest rates and these differences are real numbers.

All interest rates have meaning different from that attributed by orthodox analysis

Under the usual interpretation, the rate of interest measures financial value. An example is when IRR serves as an investment criterion. Another is when APR measures the cost of a loan. As we have seen, in the literature to date (with the
exception of Dorfman (1981)), the interpretation of unorthodox interest rates echoes the interpretation of orthodox rates, i.e., each unorthodox rate is considered singly as a measure of financial value.

Under the interpretation offered in this thesis, all \( n \) solutions for the rate of interest are employed simultaneously. One of the \( n \) rates, the ‘orthodox’, is interpreted as the unit of value in which the total value of an investment is measured. The total value of an investment is the product of this unit of value and the number of such units. The number of units is the product of the remaining \((n-1)\) unorthodox interest rates.

Under this interpretation, a rate of interest such as the internal rate of return, or the annual percentage rate, cannot be a measure of total financial value. It is merely one component - the unit in which total value is measured. If the objective is a measure of total value then each rate cannot be considered alone. All rates are determined simultaneously; and all rates considered together, as a cluster, have meaning.

Having anticipated the results of the research, Chapter 3 contains a more detailed account of the works. They are addressed in chronological order, beginning with an explanation of the project’s origins.

The claims made above are justified in the works themselves.
Chapter 3
A chronological overview of Works 1 to 10

Origins – multiple real roots from cash flows having multiple sign changes

The origin of the research lies in two events, both occurring in 1988. The first event was the publication by James Gleick of a popular science book on the emerging subject of chaos theory (Gleick, 1988). Gleick describes the research of John Hubbard, a mathematician investigating the workings of Newton’s method for calculating the roots of a polynomial equation. The method involves (a) a first guess for the solution and (b) an algorithm to move from first guess towards a better solution which forms the next guess, and so on iteratively, until the process settles down to a stable solution.

Hubbard researched the connection between the location of the first guess and the location of the eventual solution. Given many solutions, the expectation was that the method would converge on the solution closest to the first guess. The expectation was based on the assumption that there is a clean boundary between ‘basins of descent’. The expectation is only partially true. For most guesses the procedure does converge on the nearest solution. However, for guesses occurring roughly midway between two solutions the boundary is not clean. Instead, there is an intricate braid of possibilities in which the first guess converges on either of the two solutions, but, surprisingly, not necessarily the closest. Even more surprising, it can converge on a third solution far away. The phenomenon is an example of chaotic behaviour.

The second event, coincident with the first, was the provision by the author of a course in financial mathematics for bankers. The course focused on the time value of money equation and its application in retail and corporate banking. The financial calculator for the course was the Hewlett Packard 12C, the standard calculator at the time for bankers and traders. The manual for the calculator gave an example in an appendix of how the algorithm for the internal rate of return could

---

10 A recent account of Hubbard’s research is Hubbard et al. (2001). Hubbard (2008) is a memoir containing, amongst other things, an account of the origins of the work mentioned in Gleick (1988).
sometimes give unexpected results. The example is in the table below. No guidance was given in the manual as to the nature of the ‘unexpectedness’ except that it was somehow dependent on the first guess and the variation in the cash flows.

### The cash flows from the HP12C manual

<table>
<thead>
<tr>
<th>Period</th>
<th>Number of periods</th>
<th>Cash flows</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>-180,000</td>
</tr>
<tr>
<td>1-5</td>
<td>5</td>
<td>100,000</td>
</tr>
<tr>
<td>6-10</td>
<td>5</td>
<td>-100,000</td>
</tr>
<tr>
<td>11-19</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>200,000</td>
</tr>
<tr>
<td></td>
<td><strong>Total=21</strong></td>
<td></td>
</tr>
</tbody>
</table>

When Hubbard’s ideas are applied to this example, they lead to a more structured description of the way the results emerge relative to first guess. There are three sign changes, therefore the cash flow yields three real interest rates as solutions: 1.86%; 14.35%; 29.02%. Given guesses in a section of the number line roughly midway between any two of these solutions, the algorithm can home in on any of the three real interest rates, not just the two on either side of the guess. The eventual solution is highly sensitive to small differences in the initial guess. This behaviour is puzzling when analysis is confined to orthodox interest rates on the real number line. The nature of the link between inputs and outputs, however, is readily apparent from Newton’s method to locate the complex roots. The ‘unexpected’ in one dimension can be understood when viewed in the two dimensions of the complex plane. See Appendix D for an example.

*Chaotic behaviour from the time value of money equation – Works 1 and 2*

The observations described in the previous section led to the first two papers. While fluctuating cash flows of the kind that yield multiple real solutions are not

---

11 The past tense, ‘gave’, is used because the manual no longer includes the example. The author wrote to the company about their example and suggested the possible reason for the unexpected results. Coincidentally or not, the example was removed from the next edition of the manual.
common in capital budgeting, they do occur. In certain kinds of banking such cash flows are common. Works 1 and 2 demonstrate the possibility of unexpected outputs from such cash flows. Two pieces of advice are offered about what to do when multiple (2 or more) changes of sign exist in the cash flow. The first is not to accept the default value for the initial IRR built into the software (usually 10%); rather to try several different guesses to check for stability of output. The second is to graph the NPV-interest rate profile at several different interest rates. The purpose is to identify regions where the profile is flat, the most likely regions in which the phenomenon may occur.

Work 1 is written for accountants and incorporates the cash flow from the HP manual. Work 2 is written for economists and demonstrates that, as a result of the phenomenon, different hardware, different brands of software, and even different versions of the same brand of software, can result in different IRR outputs from the same cash flow input.

One consequence of writing these works was a heightened awareness of the many roots of the TVM polynomial. The works focus on the real solutions along the real number line; however, the Hubbard analysis highlights the fact that most solutions to any financial polynomial are likely to lie off the real number line in the complex plane. This leads to questions about the usefulness and meaning of all the multiple solutions to a financial polynomial, including the complex roots.

In this way, the core research programme of this thesis began in 1988 and the first publications appeared in 1990 and 1993. The search for the next step continued until 1999. During the interim period much was learned about analysis in the complex plane. Most importantly, the research methodology emerged during this time. Appendix A contains an account of the methodology and its development.

---

12 Work 1 was published in *Management Accounting*, the house journal of the Chartered Institute of Management Accountants. The journal changed its name from *Management Accounting* to *Financial Management* in 1998.
An alternative formula for modified duration – Work 3

The next significant step in the research came in the autumn of 1999 when analysing a simple TVM equation for the price of a bond. Using the trial and error technique described in Appendix A, the known value of modified duration from the orthodox calculation was reproduced via a combination of distances between all the roots of the equation. The equality is to 12 decimal places; this level of precision means coincidence is unlikely. The new formula is a fundamentally different way of expressing a well-known financial concept. Further calculation using many different values for the cash flows established that this ‘complex twin’ for modified duration seems always to produce results equal to those from the orthodox equation; the stability of the result increases the likelihood that the new formula is not a fluke. Initially, the link was established numerically.

The next step was to find an algebraic path between the orthodox equation and the new equation, and so prove the general equivalence between the two. A review of various theorems about polynomials in mathematical texts led to the theorem required for the proof: Cotes’ theorem, first published in 1722 and described in Stillwell (1989), Needham (1997) and Nahin (1998).

The above results, the empirical observation and its algebraic proof, were first published in Work 3 - Osborne (2000).

A new formula for duration that is accurate – Work 4

The next significant step is that the complex twin for modified duration can be transformed into another ‘complex’ formula having no orthodox equivalent. This general formula gives an accurate version of duration under any shift in the interest rate, a formula sought since Macauley (1938). The new formula encompasses the old formula for modified duration as a special case.

There is irony in this result. Modified duration is an attempt to establish the impact on the price of a bond caused by a change in the interest rate (the interest elasticity of the bond price). The only time that modified duration is equivalent to the new formula that works is when the interest rate does not change. The general formula was first published in Work 4 - Osborne (2001).
A statistical prism - Work 5

When calculating the locations of the zeros of many different financial polynomials a pattern emerges between the magnitude and structure of the coefficients on the one hand and the locations of the zeros on the other.

The base position is that the zeros of a polynomial capturing the cash flows of a par bond are all located around the unit circle with the exception of the one real, positive solution. The exception is located just to the right of the unit circle on the real number line at \((1+r)\) where \(r\) is the yield to maturity (also equal to the coupon because it is a par bond).\(^{13}\)

When the bond is not a par bond the coefficients of the polynomial are not regular in size or format. When the coefficients are not regular the zeros depart from the unit circle. It becomes apparent that the greater the variation of the coefficients around their mean value, or, what amounts to the same thing, the greater the departure from the structure of a par bond, then the greater the variation in the location of the zeros around the unit circle.

Many months were spent trying to establish the precise relationship between the coefficients of the polynomial and the locations of its zeros. My collaborator on this work succeeded in perceiving and proving the precise link. The results were published as Osborne and Osborne (2003). The formulas for the mean and standard deviation of the coefficients in the special form of the TVM polynomial have ‘complex twins’. These ‘twins’ are formulas containing combinations of the distances of the zeros from the cyclotomic roots on the unit circle.

At the time of writing this result stands apart from the rest of the research because it is not clear how it applies to any financial problem, apart from the fact that the point of departure (the reference structure) is that of a par bond. The formula is so neat, however, that the author is convinced that some application will be found. It is another open question. In the absence of a financial application the

\(^{13}\) A trivial implication being that the equation for a par bond with zero coupon has the cyclotomic roots for its zeros.
results were published in a mathematics journal with a readership perhaps interested in the tidiness of the result.

**A better way of expressing the complex results – Work 6**

The methodology established up to this stage of the research shows two things. First, it shows the possibility that a formula for an orthodox financial concept based on the TVM polynomial can have a ‘complex twin’ expressed in terms of distances between salient points in the plane, including all the roots of the TVM polynomial. Secondly, it shows that a variation on the ‘complex twin’ can exist, a variation having no orthodox equivalent and solving a previously unsolved problem, e.g., a version of duration that is accurate.

An important next step was the discovery of a different way of expressing the new formulas - Osborne (2004b), Work 6. Initially, the new formulas were couched in terms of the difference between two interest rates, or an increment to an existing rate, i.e., in the move from \( r \) to \( R \) there is an increment ‘\( a \)’, i.e., \((1+r+a) = (1+R)\). Subsequently, it became apparent that the formulas can be restated in simpler, more useful forms. The new formulas employ the multiplicative mark-up of one interest rate on another. In the move from \( r \) to \( R \) there is a mark-up ‘\( m \)’, i.e., \((1+r)(1+m)=(1+R)\). The multiplicative mark-up is an interest rate in its own right.

The revised formulas incorporate all multiplicative mark-ups (\( m_i \)) instead of all increments to existing rates (\( a_i \)).

The change from the incremental to the multiplicative approach is significant for three reasons. First, it fits neatly with the development of the twin ideas of the special form of a polynomial and the special relationship between its coefficients and roots. These ideas are introduced in the text of Work 6 and are in the appendices of most subsequent works. The multiplicative approach combined with these two ideas allows shorter and simpler proofs of all the main results. In addition, the main results themselves become neater. Secondly, the multiplicative approach is useful because it permits application of the new analysis to non-parallel shifts in a non-flat yield curve. Finally, the equations containing multiplicative shifts in interest rates more readily allow meaning to be attributed to the unorthodox rates. This last idea is developed in Work 10.
Consolidating ideas in bond analysis and confronting calculation issues – Work 7

Work 7 contains answers to questions posed in Osborne (2001b). The questions are about the most efficient way to compute the roots of a typical financial polynomial. The answers are consolidated with the results from Works 2 and 3 in order to present the whole more clearly.

Osborne (2001b) is not included in this submission because it is a conference paper containing a summary of the research to that date by way of a briefing for the experts in numerical computation attending the conference. The questions and expert answers are summarised in Appendix C.

The application of all solutions to a new topic: capital budgeting – Work 8

The first results to emerge from the research are in bond mathematics. This was happenstance. Once the methodology was established it could be applied to concepts in other areas of finance and economics employing the TVM equation. It quickly became clear that capital budgeting is a fertile subject for treatment.

Work 8 was conceived, written and posted in a working paper archive in 2004. Published in a refereed journal in 2010, it sheds light on the long-standing debate about the relative merits of NPV and IRR as investment criteria. The application of the multiple interest rate approach shows that NPV per dollar invested is equal to the product of the multiplicative mark-ups of all possible IRRs over the cost of capital. As noted in the summary of results in the first chapter, this finding has a number of implications.

First, NPV must be a superior investment criterion to IRR because it carries more information. This is because the IRR criterion employs the single mark-up of the orthodox IRR over the cost of capital, whereas NPV per dollar invested employs all possible mark-ups of IRRs over the cost of capital.

A second implication is that the existence of multiple IRRs is no longer a pitfall. On the contrary, all possible IRRs, along with the cost of capital, are the
fundamental building blocks of NPV per dollar invested. The unorthodox rates play a part.

A third implication involves a second pitfall of IRR: that NPV and IRR can give inconsistent rankings when investment projects are compared. The textbook chart in capital budgeting has NPV on the vertical axis and the interest rate on the horizontal axis. The analysis in Work 8 shows that this chart no longer makes sense. NPV per dollar invested is associated with all possible mark-ups of the IRR over the cost of capital, in which case inconsistent rankings do not arise and the pitfall does not exist.

A fourth implication involves yet another pitfall about IRR. The pitfall is that the IRR approach cannot be ‘finessed’ to take account of a variable yield curve. Osborne (2004a) shows that the effect on value of non-parallel shifts in a non-flat yield curve can be analysed using the new approach in which IRR plays a fundamental part. Therefore this pitfall, too, does not exist. The analysis is developed further in Chapter 4 of this context statement.

Third round of applications: reswitching – Work 9

When two techniques of production are compared, reswitching is the possibility that one technique can be cheapest at a low interest rate, switch to being more expensive at a higher rate, and reswitch to being cheapest at even higher rates. For some, this inconsistency undermines the foundations of neoclassical economic theory.

The debate about reswitching has lasted more than 100 years, peaking in the 1960s when many eminent economists contributed to a symposium on the subject. The reswitching debate and allied controversies were summarised in Harcourt (1972).

---

The solution to reswitching in Work 9 is a clarification of the changes occurring when the orthodox interest rate shifts. The independent variable is more complicated than the orthodox analysis suggests. When the entire independent variable is included, reswitching no longer occurs.

Researching the reswitching paper led to the realisation that, over the last century, research in economics and finance has exposed various difficulties with the TVM equation. At their core, all the difficulties are about one question: how changes in interest rates impact value. The reswitching puzzle in economics, the pursuit of a definitive version of duration in bond mathematics, and the NPV versus IRR debate in capital budgeting, are all long-standing issues providing evidence that an answer has not been readily available. In particular, it has not been obvious that the topics have a common core and are intimately connected. The many issues can be seen as anomalies in the Kuhnian sense (Kuhn, 1962).

The meaning of all solutions to the TVM equation – Work 10

As its title makes clear, Work 10 is a follow-on paper to Work 8. The earlier paper produces a mathematical equation that is simple but not obvious. A drawback is that most of the elements on the right-hand side, the multiplicative mark-ups, are unfamiliar and lack a financial meaning.

Work 10 proposes meaning for all the interest rates that solve the TVM equation. It does this via analogies with simpler financial concepts. The notion of standard value structure (SVS) is introduced, whereby the difference in value relative to initial value is equal to the product of the unit of value and the number of such units. The SVS is used to compare values of two things at the same moment in time, and values of the same thing at two different moments in time. Crucially, the SVS is then applied to comparisons between two different outputs from a TVM equation when two different interest rates are input.

When applied to the new equations emerging from this thesis, the notion of SVS implies that one interest rate, such as the mark-up of the orthodox IRR over the

---

15 There is research into combining duration analysis with DCF techniques in capital budgeting. See the literature mentioned in the list of topics for future research in Chapter 4 of this context statement.
cost of capital in capital budgeting, is the unit of value. The product of all the other
mark-ups, previously ignored or discarded, is the number of units of value. In the
light of this interpretation the NPV versus IRR debate dissolves because IRR is not
an investment criterion. Instead, IRR is a component of a criterion, albeit a
fundamental component.

Work 10 signals the end of a sequence of papers that gradually develops use
and meaning for all solutions to the TVM equation. It is possible to see the
attribution of meaning as the insertion of the final piece of a jigsaw puzzle. In fact
this is not the case; the research programme is just beginning. Many other puzzles
remain to be solved. The next chapter contains a summary of next steps, both
certain and speculative.
Chapter 4
Further work and open questions

In this chapter suggestions for further work and the open questions noted earlier are summarised together with additional items not yet noted. In some instances the questions and their answers are developed to the point where the outline of a future publication is indicated. In other instances, the questions remain open, therefore the general thrust, but not the detail, of a research programme is outlined.

Incorporating the yield curve

Work 6 (Osborne, 2004b) and Osborne (2004a) are about duration-convexity and NPV-IRR respectively. The articles show, amongst other things, that the core formulas are robust under non-parallel shifts of non-flat yield curves. The observation that the yield curve can be accommodated has yet to be published in a refereed journal.\footnote{Although the point about accommodating a varied yield curve remains unpublished, other observations in Osborne (2004a) are published in Work 8 and, therefore, are included in this thesis. Osborne (2004b), though unpublished, is submitted in this thesis as Work 6. This is for two reasons. First, its inclusion provides some continuity in the story by including initial thoughts on how the yield curve can be accommodated. Secondly, the work passed a refereeing process for a conference and is permanently available on a website as part of the proceedings.} The methodology is to shrink the size of the unit of value, and therefore the unit of time, to the point where a varied yield curve, and non-parallel shifts in the yield curve, can be accommodated. The logic likely to be incorporated into a future publication is outlined below.

The application of the ‘standard value structure’ in Work 10 demonstrates that any time period can be chosen when calculating the interest rate from a TVM equation with a specified cash flow. The interest rates in the output from the calculation vary such that their product is always equal to the relative shift in value.

This logic can be reversed. However small, or large, a unit of value (interest rate) is chosen when analysing a TVM equation with a specified cash flow, the number of periods that results from the calculation varies appropriately such that the product of all interest rates is equal to the relative shift in value.
A simple example is as follows. If \((1+r)^n=1.21\) then \(n\) can be chosen and \(r\) is the result. For example, if \(n=2\) then \(r=0.1\) or 10\%. Alternatively \(r\) can be chosen and \(n\) is the result. Assume a small unit of value, \(q\), and set \(q\) equal to one basis point, or one hundredth of one per cent. The equation is re-expressed as \((1+q)^n=(1.0001)^n=1.21\) therefore \(n = 953\) rounded to the nearest integer.

The ideas in the previous paragraphs can be used to advantage. Assume a simple cash flow with a varied yield curve, e.g., a value for NPV resulting from an investment \(I_0\), four different cash flows, and a different spot rate of interest applied to each period.

\[
NPV = -I_0 + \frac{C_1}{(1+r_1)} + \frac{C_2}{(1+r_2)^2} + \frac{C_3}{(1+r_3)^3} + \frac{C_4}{(1+r_4)^4}
\]  

(7)

\[R\] is the IRR that sets NPV=0.

\[0 = -I_0 + \frac{C_1}{(1+R)} + \frac{C_2}{(1+R)^2} + \frac{C_3}{(1+R)^3} + \frac{C_4}{(1+R)^4}
\]  

(8)

Assume a series of mark-ups of the form \((1+q)^i\) where \(i\) is a single number appropriate to the task of raising the \(i^{th}\) cost of capital to the IRR, i.e., 

\[(1 + r_i)(1 + q)^i = (1 + R)\]. Therefore Eq. (8) becomes (9).

\[0 = -I_0 + \frac{C_1}{(1+r_1)(1+q)^{i_1}} + \frac{C_2}{(1+r_2)^2(1+q)^{i_2}} + \frac{C_3}{(1+r_3)^3(1+q)^{i_3}} + \frac{C_4}{(1+r_4)^4(1+q)^{i_4}}
\]  

(9)

This last equation is manipulated into the special form.

\[
0 = 1 + \frac{C_1/I_0}{(1 + q)^{i_1}} + \frac{C_2/I_0}{(1 + r_1)^2} + \frac{C_3/I_0}{(1 + r_2)^3} + \frac{C_4/I_0}{(1 + r_3)^4} - 1
\]  

(10)
Eq. (7) is rearranged and absolute values taken to produce Eq. (11).

\[
\frac{C_1/I_0}{(1+r_1)} + \frac{C_2/I_0}{(1+r_2)^2} + \frac{C_3/I_0}{(1+r_3)^3} + \frac{C_4/I_0}{(1+r_4)^4} - 1 = \prod_{j=3}^{14}|q_j|
\]  

(Eq. 10)

Eqs (10) and (11) combine to make (12).

\[
\frac{NPV}{I_0} = \frac{C_1/I_0}{(1+r_1)} + \frac{C_2/I_0}{(1+r_2)^2} + \frac{C_3/I_0}{(1+r_3)^3} + \frac{C_4/I_0}{(1+r_4)^4} - 1
\]  

(Eq. 11)

\[
\frac{NPV}{I_0} = \prod_{j=1}^{16}|q_j|
\]  

(Eq. 12)

The element \(q_j\) in Eq. (12) is one basis point. Given the proposal about the meaning of all interest rates described in Work 10, the product of the \((t4-I)\) remaining interest rates, i.e., \(\prod_{j=2}^{14}|q_j|\), is the number of basis points comprising NPV per dollar invested.\(^{17}\)

In this way, the margins of a single orthodox IRR over the various interest rates along the yield curve are forced to combine in a manner that produces a formula for NPV per dollar invested. Thus, another of the four pitfalls of IRR listed in Brealey et al. (2009) is demonstrated not to exist.

The structure of Eq. (12) is similar to the structure of the equations developed elsewhere in this thesis. There is a difference, however. The structure of the early formula for NPV per dollar is the product of the margins of all IRRs over

\(^{17}\) The notion of very small increments in interest rates may appear impracticable; however, one prominent practitioner in the financial markets, Richard Olsen, has suggested second-by-second interest rates in the context of intra-day trading in the foreign exchange and money markets. Olsen (2010) provides persuasive financial reasons why it should happen and argues that it is technologically possible.
the cost of capital. In this new equation NPV per dollar invested is the product of the margins of a single value of IRR over all interest rates along the yield curve.

Continuous time versus discrete time

The methodology applied in this thesis requires discrete rather than continuous time. The process of shrinking the unit of value (and time) described in the previous section is limited. Quite apart from the computational issues, if the process goes far enough, practically, it becomes continuous time. Earlier, the words of Boulding (1936b) were noted. The words are appropriate here: ‘… if we adopt continuous compounding, as in strict theory we should, the theoretical number of solutions is infinite!’ If the number of solutions is infinite, then the methodology becomes intractable. If the problems discussed here can only be solved using this discrete approach, the implication is that there might exist a class of financial problems soluble only in discrete time. The existence of this class of problems, and the boundaries of the class, are open questions.

An additional point is that most theoretical models analysed in continuous time are, in practice, solved using discrete numerical methods.

Mean and variance of cash flows

Work 5 shows a link between the locations of the roots and the statistical properties of the coefficients in the TVM equation. The mean and variance of the coefficients in the special form of a polynomial are linked to the locations of all roots relative to the cyclotomic roots. As noted earlier, there is no clear application of this finding to economic or financial analysis. An obvious research project would be to find such an application. A pertinent, possibly significant, observation is that the base structure, around which variation takes place, is that of a par bond.

Making the analysis stochastic

The research in this thesis shows that non-stochastic formulas in orthodox finance can have complex twins. This raises a question. Do stochastic formulas also have complex twins? The question can be more focused and more provocative: does the Black-Scholes formula have a complex twin? If yes, can the complex twin
be developed into an improved formula, in the way that the complex twin for modified duration was developed into an improved formula?

Putting the last questions to one side, two simpler, prior questions are asked, using one of the new expressions for NPV per dollar as an example.

\[ \frac{NPV}{I_0} = \frac{\prod |R - r_i|}{(1 + R)^n} \]

Assume the cash flows (parameters) in the orthodox equation for NPV are known. This assumption means the zeros of the equation, and therefore the many values of \( r_i \), are given. If so, only \( R \) can be stochastic in the equation above. The interesting question is how NPV per dollar behaves as \( R \) moves. For example, if \( R \) is a normally distributed variable, how is NPV per dollar distributed?

The second question is more difficult to answer. Capital budgeting is a forward-looking process. In reality, the cash flows in the orthodox NPV equation are not given; they are forecasts. If the cash flows are random variables, it follows that the zeros of the orthodox NPV equation, and the values of \( r_i \), are also random variables. The points in the plane move randomly. How can this be modelled?

A suggestion is made at this stage. Physicists have modelled similar processes for some years. There is a large literature concerning random polynomials in quantum dynamics. An example is Bogomolny et al. (1996). They examine how the zeros of a polynomial are distributed in the plane when the coefficients of the polynomial are normally distributed random variables. It may be possible to adapt their methods.

Another class of applications: retail financial products and APR

The UK Office of Fair Trading (OFT) issues a booklet called ‘Credit Charges and APR’ that details ‘how to calculate the total charge for credit and the annual percentage rate’ on a retail loan. The booklet is based on the Consumer Credit Act 1974 and the accompanying legislation, The Consumer Credit (Total Charge for Credit) Regulations 1980. There have been many amendments to these
two Acts - they are detailed in the OFT booklet - but the core of the documents remains unchanged. The 1980 Act in particular is an extraordinary legal document. Most of the document consists of an explanation of a mathematical equation and a financial concept: the TVM equation and the APR. Neither the equation nor the concept is easy. The OFT’s explanatory booklet is similar to the 1980 Act in that 48 pages out of 64 (75%) are devoted to the explanation of the mathematics of retail loans.

The EU Consumer Credit Directive of 2008 is a less forbidding legal document than the two UK Acts; nevertheless Annex 1 of the Directive is equally technically difficult because, like the UK legislation, it sets out the TVM equation in detail in an effort to explain exactly what is meant by APR.

There is a considerable academic literature lamenting the lack of understanding and transparency demonstrated by all parties to consumer credit agreements. The difficulties and opacity have their origins in the complex legislation that binds all parties. Alternative, simpler arrangements are often proposed. Buch et al. (2002) is an example that also contains references to some of the literature.

Before unpacking the retail legislation in the light of the analysis of this thesis, two observations are made.

First, the legislation attaches much importance to the concept of the total charge for credit (TCC); hence the inclusion of the phrase in the title of the UK’s 1980 Act. The primary item in the list of charges for inclusion in the total charge for credit is the interest paid, because it is usually the largest single item; this primacy explains why a detailed mathematical description of the APR immediately follows the list.

Secondly, the legislation refers to the concept of flat rate, or the simple rate of interest, often used by sales people to explain the cost of retail credit to consumers. The legislation does not forbid quoting the flat rate to consumers. Flat rate is included in examples in the OFT booklet. But the legislation does insist that
the TCC and the APR are given prominence over the flat rate in advertisements and documentation. The OFT’s booklet goes to some length to compare and contrast the flat rate and the APR, and to explain the supremacy of the latter.

One aspect of a loan that APR illustrates is the ‘use’ the borrower gets of the money loaned. In an agreement where the borrower repays the loan in regular instalments, the credit available to the borrower can be regarded as falling from the full amount borrowed at the start of the loan period to nothing by the end. Looked at in this way, on average over the lifetime of the loan, the amount of credit available to the borrower is about half that they originally borrowed. Generally therefore, with this type of agreement the APR will be about twice the annual flat rate charged by the lender. On the other hand, when a loan is repaid by a single repayment, the borrower has the ‘use’ of the amount of credit originally borrowed for the whole loan period, and the APR is therefore much closer to the annual flat rate (although the APR will be lower for longer loans).

The main ideas in this thesis are now employed to analyse the APR calculation detailed in the legislation. The most recent and comprehensive official account is that found in Annex 1 of the EU Consumer Credit Directive, 2008. The part containing the core equation is quoted below.

‘The basic equation, which establishes the annual percentage rate of charge (APR), equates, on an annual basis, the total present value of drawdowns on the one hand and the total present value of repayments and payments of charges on the other hand, i.e.,

$$\sum_{k=1}^{m} C_k (1 + X)^{-t_k} = \sum_{l=1}^{m^*} D_l (1 + X)^{-s_l}$$

where
- \(X\) is the APR,
- \(m\) is the last drawdown,
- \(k\) is the number of a drawdown, thus \(1 \leq k \leq m\),
- \(C_k\) is the amount of drawdown \(k\),
- \(t_k\) is the interval, expressed in years and fractions of a year, between the date of the first drawdown and the date of each subsequent drawdown, \(t_1 = 0\),
- \(m^*\) is the number of the last payment or payment of charges,
- \(l\) is the number of a repayment or payment of charges,
- \(D_l\) is the amount of a repayment or payment of charges,
- \(s_l\) is the interval, expressed in years and fractions of a year, between the date of the first drawdown and the date of each repayment or payment of charges.’
In order to analyse the APR equation some simplification is made. Assume a typical retail loan in which there is one drawdown at the beginning, $C$, and then a series of equal, regular payments, $D$, that include any charges, until the loan is repaid in full. The APR equation given above is modified to reflect these assumptions. The notation is simplified slightly: the initial drawdown is in period 0 and the subsequent payments are in periods $l = 1$ to $n$.

$$C_0 = \sum_{l=0}^{n} \frac{D}{(1 + X)^l}$$  \hspace{1cm} (13)

The total charge for credit in this situation is the total amount repaid less the total borrowing, i.e., $TCC = nD - C_0$. The flat rate of interest per period is the total charge for credit divided by the amount borrowed, $C_0$, and further divided by the number of periods, $n$, i.e., $F = \frac{nD - C_0}{nC_0}$. Another way of looking at this last result is that $n$ flat rates are equal to the total charge for credit relative to the amount borrowed, i.e., $nF = \frac{nD - C_0}{C_0}$.

‘Multiple interest rate’ analysis is now applied. Manipulate the APR equation (13) into the special form.

$$-1 + \frac{D/C_0}{(1 + X)} + \frac{D/C_0}{(1 + X)^2} + \ldots + \frac{D/C_0}{(1 + X)^n} - \frac{1}{(1 + X)^n} + \frac{1}{(1 + X)^n} = 0$$  \hspace{1cm} (14)

Extract the special relationship from the special form.

$$\left| \frac{nD}{C_0} - 1 \right| = \prod_{l=1}^{n} |X_l|$$  \hspace{1cm} (15)

This last result connects neatly with the definition of the flat rate.

$$|nF| = \prod_{l=1}^{n} |X_l|$$  \hspace{1cm} (16)
The interpretation of the product of all interest rates in Work 10 is applied to Eq. (16). \( X_1 \) is the orthodox APR and the unit of measurement of the cost of the loan. The product \( \prod_{i=2}^{n} X_i \) is the number of units of measurement. As a result of this interpretation Eq. (16) can be expressed in the following words:

\[
\text{[Number of flat rates x flat rate per period]} = \text{[Number of APRs x APR per period]}
\]

The entities on both sides of the equals sign are identical to the total charge for credit relative to the initial borrowing.

This last result is not obvious. An interpretation of the result is as follows.

In the same way that IRR does not convey all there is to say about the return on an investment, the orthodox APR does not convey all there is to say about the cost of a loan. The APR is the mark-up applied to each dollar in each period. The entire cost of a particular loan not only depends on the mark-up per dollar but also on how many dollars are marked up, and when. The entire cost is encapsulated in the product of all possible APRs. The product of all APRs is identical to the product of the number of periods and the flat rate per period.

The result brings into question the emphasis on the orthodox APR in consumer credit legislation; it undermines the law’s insistence that salespeople tell consumers the APR is the measure of the relative cost of a loan.

Sales-people sometimes break the law by explaining the flat rate and not explaining the APR. The former is much easier to explain. In the interest of public understanding, it would seem that an explanation of the flat rate alone is sufficient. The existing focus on the total charge for credit could be maintained, and emphasis on the technically difficult APR lowered.
Connections with other works: item 1: Dorfman (1981)

Dorfman (1981) is a seminal article. Dorfman was the first researcher to analyse all internal rates of return in preference to the subset of real rates. Moreover he did not employ them in the traditional way, individually, as investment criteria. He employed them simultaneously, as components in another financial concept, as in this thesis. Furthermore he produced the work nearly a decade before mathematical software became widely available to calculate the zeros easily and explore their locations in the complex plane. In all these respects Dorfman (1981) is a pioneering work.

There are two differences between Dorfman’s work and the analysis in this thesis. The first difference is that Dorfman employs the complex solutions in their raw form, i.e., in the form $a + bi = a + b\sqrt{-1}$. The works in this thesis use the absolute values of differences between complex solutions, which are real numbers.

The second difference concerns the structure of the equations. Dorfman builds on the work of Fisher to argue that the NPV criterion is based on particular assumptions.

‘It assumes that the purpose of investing is to be able to afford the greatest possible amount of consumption. (In the presence of the perfect financial market assumption, the greatest amount of consumption is well defined as the consumption pattern with greatest possible present value). If the purpose of investment is anything other than enhancing the ability to consume, then some other criterion may well be appropriate. ... One particularly appealing alternative objective is growth: growth of the enterprise or growth of the economy as the case may be. In the sequel, we shall develop the implications of the maximum growth objective, and shall see that in conjunction with some assumptions about reinvestment opportunities, it entails an internal rate of return criterion for selecting investments.’ Dorfman (1981)

Dorfman goes on to assume that the money earned from a given investment project is serially reinvested in identical projects *ad infinitum* and asks what determines the long term growth rate of a dollar invested in the extended project. He produces an expression for $Y(t)$, the total amount of gross investment in calendar year $t$. Below is the simplest version of his equation.
\[ Y(t) = \sum_{i=1}^{T} c_i (1 + r)^t \]  

(17)

‘... where \( T \) is the length of the individual project’s life, ... the \( c_i \) are constants which are determined by any \( T \) values of \( Y(t) \) along the growth path, usually the first \( T \), ...’ and \((1+r_i)\) represents all possible roots of the project’s internal rate of return equation.

There is an analytical bridge to be built between Dorfman’s equation and the analysis in this thesis. The nature of the bridge is an open question. An obstacle is the structure of the equations. Dorfman’s equation contains a \textit{summation} of roots, including complex roots in their raw form, each being one-plus-an-interest-rate. The core equations in this thesis employ \textit{products} of interest rates, or \textit{products} of increments to or mark-ups on interest rates, all in the form of real numbers.

\textit{Connections with other works: item 2: Bosch et al. (2007)}

Bosch et al. (207) is an article on the NPV-IRR debate that does not mention multiple IRRs directly. The authors refer to conventional and unconventional projects in passing; by ‘conventional’ they mean ‘projects that present only one change of sign in their cash flow sequence.’ Therefore the article does not appear in Table A. Nevertheless it receives separate mention here because the authors derive a ‘functional relation between NPV and IRR’ that bears a resemblance to the new equations in this thesis.

To present the relation, their notation is adapted to that of the various works presented here. The relation is built directly from the standard equations for NPV and IRR.

\[ NPV = -I_0 + \sum c_i \frac{(1 + r)^t}{(1 + r)} \]  

(18)

\[ 0 = -I_0 + \sum c_i \frac{(1 + R)^t}{(1 + R)} \]  

(19)
NPV is net present value; $I_0$ is the initial investment; $c_i$ are the cash flows in periods 1 to $n$; $r$ is the cost of capital; and $R$ the internal rate of return.

In their words (with adapted notation): ‘We can assign a non-negative proportion $\alpha_i$ of the initial capital outlay $I_0$ to any free cash flow $c_i$, in such a way that the free cash flow $c_i$ pays back the capital assigned $\alpha_i I_0$ and compensates it at the IRR:

$$\alpha_i I_0 (1 + R)^i = c_i$$  \hspace{1cm} (20)

The authors label the component $\alpha_i$ the payback coefficient of period $i$. In addition they note the following relationship $(1 + r)(1 + m) = (1 + R)$ from which is deduced $m = \frac{(R-r)}{(1+r)}$.

From the above ingredients they produce their ‘functional relation between NPV and IRR.’

$$\frac{NPV}{I_0} = \left[ -1 + \sum \alpha_i (1 + m)^i \right]$$  \hspace{1cm} (21)

NPV per dollar invested is a function of the payback coefficients and the multiplicative mark-up of the orthodox IRR over the cost of capital. Compare expression (21) with the expression for NPV per dollar derived in Work 7 that involves the multiplicative mark-ups of every possible IRR over the cost of capital.

$$\left| \frac{NPV}{I_0} \right| = \prod |m_i|$$  \hspace{1cm} (22)

The link between Eqs. (21) and (22) needs to be established, the objective being a comment on their paper.

Connections with other works: item 3: Barney and Danielson (2004)

The present value of a cash flow depends on the size and timing of the elements in the flow, and the interest rate (or rates) employed to discount or compound them. The size and timing of the elements of a cash flow can be
encapsulated in a single statistic: duration. The concept is commonly employed in bond mathematics. Duration is the size-weighted average time of the elements of the flow. It follows that the first sentence in this paragraph can be rewritten: the present value of a cash flow depends on its duration and the interest rate (or rates) employed to discount or compound it.

At this point it is worth quoting Barney and Danielson (2004) who summarise the issues:

‘While some of the relationships between NPV, IRR, and cash-flow timing seem intuitive — indeed, they are discussed in popular introductory textbooks — until now a precise mathematical link has not been established. Nor has there been adequate formal development of the role of duration in explaining ranking conflicts between projects. The literature on duration’s uses in capital budgeting has focused on duration as an alternative to the payback period [Durand (1974), Finch and Payne (1996), Karsak, (1998)], and on duration’s role in determining the impact of changes in the discount rate [Barney and White (2003), Blocher and Stickney (1979), Brown and Kulkarni (1993), Cornell (1999)]. By way of contrast, this paper examines the mathematical ties between NPV, IRR, and duration.’

Note: this author’s emphasis.

Barney and Danielson’s ‘mathematical ties’ are built around their concept of ‘return duration’: ‘the effective number of years the initial investment in a project will earn a compounded annual return equal to the project’s IRR.’ For continuity, the following two equations use the notation established earlier in this context statement rather than that of Barney and Danielson. The equations are stated here (without proof) to demonstrate Barney and Danielson’s methodology in order to contrast it with the results of this thesis. The first equation defines return duration.

\[
\frac{NPV}{I_0} = -1 + (1 + m)^\tau
\] (23)

\(NPV\) and \(I_0\) take on their usual meanings. As before, \(m\) is defined by the relation \((1 + r)(1 + m) = (1 + R)\) in which \(r\) is the cost of capital and \(R\) is the internal rate of return. Finally, \(\tau\) (tau) is ‘return duration’. Return duration is equal to the number of periods in the project if there are only two cash flows, one out \((I_0)\) and one in, and it is less than the number of periods in the project where there are more than two flows. It is calculated by inverting Eq. (23).
Barney and Danielson then derive a second equation that links orthodox Macaulay duration with return duration.

\[
D = \tau - (1 + r) \ln(1 + m) \frac{dm}{d\tau}
\]  

(24)

They use the concept of return duration to reconcile the ranking conflicts between NPV and IRR. In their words:

‘Return duration provides the conceptual link between a project’s internal rate of return and its net present value. Having a single equation relating duration, IRR, and NPV aids in understanding how cash flow timing differences can create ranking conflicts. Using return duration, a project having a higher IRR and a longer duration will necessarily have a higher NPV when compared to a lower-IRR, shorter-duration project. This intuitively appealing result surprisingly does not always hold with Macaulay duration.’ Barney and Danielson (2004)

They acknowledge a drawback of their analysis: that return duration requires the existence of only one real IRR (or requires the value can be selected if more than one real value exists).

The analysis in this thesis is now applied to the question of a mathematical link between NPV, IRR and duration. In bond mathematics, the long search for an accurate formula for duration is effectively a search for the element that stands between the proportionate change in the value of a bond and the change in the interest rate that causes it. Given the work in this thesis, such an element is stated thus:

\[
\left| \frac{AB}{B} \right| = \prod |m_i| - \prod \frac{|R_i - r|}{(1 + r)^n} = \prod_{2}^{n} \frac{|R_i - r|}{(1 + r)^n} |R_i - r|
\]  

(25)

The interest rate changes from the orthodox rate \(R_i\) to \(r\) therefore the bond’s value changes from \(B\) to \(B^*\). The vital element ‘duration’ is the discounted product of the shifts of \(r\) away from the \((n-1)\) unorthodox values of \(R\) that solved the original bond equation.
Switching to capital budgeting, Work 8 shows that NPV per dollar invested is the product of all multiplicative mark-ups of every IRR over the cost of capital.

\[ \frac{\Delta B}{B} = D |R_i - r| \]  

(26)

\[ \frac{NPV}{I_0} = \prod m_i = \prod \frac{|R_i - r|}{(1+r)^n} = \prod \frac{1}{2} (1+r)^n |R_i - r| \]  

(27)

The shift in interest rate from the orthodox internal rate of return to the cost of capital raises net present value from zero to \( NPV/I_0 \). The analogy with the previous equation is clear. Duration is the element that stands between the two changes. It is the discounted product of the shifts of the interest rate from all the original, unorthodox values of the internal rate of return to the cost of capital.

\[ \frac{NPV}{I_0} = D |R_i - r| \]  

(28)

Eq. (28) may be the most concise mathematical link possible between NPV, IRR and duration.
Chapter 5

Conclusion

The original ideas in the works comprising this thesis are summarised by focusing on their practical applications.

Capital budgeting

Work 8 contains numerous references describing capital budgeting practice around the world. The references themselves contain further references documenting capital budgeting practice in past years. Studies of capital budgeting practice have proved fascinating to academic analysts and industry participants alike. The general conclusion is that DCF techniques have steadily become more popular than less sophisticated techniques such as payback period. Moreover, within DCF practice, the NPV criterion is gradually outpacing IRR in popularity. Nevertheless, in some areas of industry and commerce, around half of practitioners continue to employ IRR.

The traditional academic arguments for NPV rest partly on the ‘pitfalls’ of the alternative criterion, IRR. In this thesis it is argued that the pitfalls do not exist. NPV is superior to IRR for reasons entirely different from the pitfalls; NPV is superior because it carries more information than any individual IRR. All IRRs are components of NPV.

Furthermore, in Work 10, it is argued that orthodox IRR should not be used as an investment criterion because, by itself, an IRR is not a measure of overall value. The orthodox IRR is a unit of value employed to measure overall value, and the unorthodox IRRs together measure the quantity of such units. In this sense, the resolution to the NPV-IRR debate is amicable. The judgment of many practitioners, that the IRR concept is useful, is not misplaced. IRR has a fundamental role providing the many elements that comprise NPV.

As noted in the brief summary in Chapter 1, on the assumption that these arguments about capital budgeting are correct, some chapters in finance textbooks require revision. If, via the textbooks, the arguments prove persuasive to the finance
profession at large, then capital budgeting practice may shift further towards the use of NPV.

**Bond mathematics**

Duration is the percentage change in the value of a bond in response to a change in the interest rate. It can be established empirically using market data (effective duration) or theoretically using the bond pricing formula (Macaulay and modified duration). Whichever approach is employed, the target is a stable coefficient, standing between the two percentage changes, to act as a guide to action or decision.

The ideas developed in Works 3, 4, 6 and 7 show that duration in the theoretical sense is a chimera. The ideas show the percentage change in the price of a bond can be expressed as the product of the percentage changes between the new rate and all old ones. There is nothing else to the analysis. There is no stable coefficient to stand between the two orthodox percentage changes. The ‘coefficient’, the element that stands between the two orthodox percentage changes, is part of the independent variable. The new equation is more of an identity than a behavioural equation.

Most financial theory now employs a stochastic approach. As discussed in the previous chapter, the complex analysis outlined in the thesis has yet to be made stochastic. Therefore, whether and how the analysis is absorbed into the fixed income literature is for discussion; but it is asserted here that academics and practitioners should be aware of the analysis.

**Reswitching**

The current value of a stock of capital can be seen as the present value of its past costs of production. Or it can be seen as the present value of the future stream of income likely to accrue to the stock. Either way an interest rate is necessary to compound or discount. When performing ‘what-if’ calculations at a moment in time, different assumptions about the interest rate result in different values of the capital stock. The ideas in Work 9 demonstrate that the analysis of such shifts in
value is not straightforward. As with duration in bond mathematics, the independent variable is not what it seems.

Some researchers see reswitching as a sign of inconsistency in neoclassical economics. When viewed from the perspective of ‘multiple interest rate’ analysis, the inconsistency disappears. Nevertheless, difficulties remain. The new view of the independent variable requires exploration. Moreover, the mode of analysis is comparative statics; time is not passing. There are questions about how to incorporate the passage of time into the analysis and make it dynamic. Finally, there is the question about how to incorporate a future stream of income that is stochastic.

*Retail and wholesale credit*

Current retail credit legislation is complicated and subject to criticism for being opaque. In part the complication and opacity are because the legislation emphasizes APR as the true measure of the cost of a loan, and the calculation of APR is not an everyday task for most people. The analysis in Chapter 4 of this thesis suggests that the emphasis on the orthodox APR may be misplaced. The total charge for credit and the flat rate of interest are simpler concepts. It is shown that the concepts are adequate to the task of measuring the cost of a loan precisely because they are built of all possible APRs. It follows that consumer credit legislation could be simplified by downplaying the importance of APR in points of contact with the consumer such as advertisements and agreements. This could be done without loss of information to the consumer, and without denigrating the vital role of APR as a fundamental component of the two simpler concepts.

*Final thoughts*

To the author there is both irony and surprise in the analysis offered in this thesis. Irony, because complex numbers and analyses in the plane are, without question, abstract. Yet their employment in economics and finance permits solutions to several long-standing problems, solutions that are simple to the extent that the ‘complex twin’ equations have simpler structures than their orthodox equivalents.

Surprise because such abstract analysis leads to practical conclusions.
Appendix A

Research methodology

At the beginning of the research project, little was known beyond two facts. First, there is the fundamental theorem of algebra due to Gauss: every polynomial has $n$ roots, either real or complex (see Stillwell, 1989). Secondly, the complex roots come in complex conjugate pairs if the parameters of the polynomial are all real (see Erdos & Turan, 1950).

Two unknowns were the locations of the roots in the complex plane for a typical financial polynomial and the links between all the roots of a financial polynomial and any financial or economic concept.

The locations of the roots

The locations of the roots associated with financial polynomials were established using three strategies. The first two strategies involved calculation and the third involved research in the mathematics literature.

The initial strategy was as follows. In the absence of mathematical software to calculate the roots from given parameters, values were assumed for the roots and the polynomials were reconstructed. The values of the resulting parameters were studied to see if they made financial or economic sense. If they did, then the locations of the roots and the values of the parameters were compared. This was a slow process, abandoned when availability of mathematical software made possible the second strategy, in which the procedure is put into reverse.18

With the second strategy, the locations of the roots are calculated directly from financial polynomials with known parameters (cash flows) having credible magnitudes; this process involves considerably less time and effort. It makes possible a large increase in the number of comparisons between root locations on the one hand and parameter magnitudes and patterns on the other.

18 The research began in 1988. MathCad was first introduced in 1986. Mathematica was first released in 1988. Maple was developed in 1980 but it only became widely available when sold commercially, also in 1988. First versions of the software were expensive and not always easy to obtain.
The third strategy was to identify and consult some of the classic books on polynomials. In the early years these included Marsden (1949) and Polya & Szego (1976). These works yielded a fund of background knowledge about polynomials and the locations of their roots but did not yield any direct help to the research. Works consulted later in the research were Farahmand (1998), Nahin (1998) and Needham (1997). These later works, especially the latter two, were useful.

From the beginning, work focused on fourth degree polynomials. Four roots is the minimum number necessary to obtain roots of every type: two real roots (one positive and one negative) plus a complex conjugate pair. Less than four and one of the types is missing. More than four and calculation difficulty rises rapidly. A polynomial with four roots offers sufficient complexity with tractability.

A link between the locations of the roots and financial concepts

A strategy was needed to bridge a gap separating two kinds of knowledge, one known and the other conjectured. On one side of the gap were known values of key financial statistics calculated from typical financial polynomials. Examples of financial statistics include duration in fixed income mathematics and NPV in capital budgeting. On the other side of the gap was knowledge of how all roots, complex as well as real, could be combined or connected into new expressions that act as alternative equations for the key financial statistics. It was a conjecture that useful, meaningful expressions exist.

The strategy employed to bridge the gap between known equations and their hypothetical equivalents was trial and error. It involved two steps.

First, simple, whole number values were assumed for the parameters (cash flows) of a fourth-order TVM polynomial. The orthodox values of key statistics were calculated based on the parameters. The key statistics included duration and convexity (assuming the cash flows describe a bond) and NPV (assuming the cash flows describe a project). Although the inputs were whole numbers, the key statistics were calculated to many decimal places, usually twelve.
Secondly, the roots were located to twelve decimal places. These roots were employed in various calculations of a speculative kind. The calculations include combinations of roots, or combinations of distances between roots, or distances between roots and other salient points in the plane such as (0,0) or (1,0). The objective was to see if any of the guessed combinations were equal, or nearly equal, to any values of the key financial statistics.

In the first instance, many months were spent calculating combinations of the complex roots themselves in the belief that the raw complex numbers, numbers of the form $x+iy$, might have some financial meaning. Nothing came of this. The meaningful results come from combining distances between roots, and between roots and other salient points such as (0,0) or (1,0), as described in the works. Such distances are absolute values in complex space therefore they are real numbers. The complex numbers per se do not seem to have financial meaning.\(^{19}\)

Other researchers have pondered the significance of the complex roots as complex numbers per se, see particularly Dorfman (1981) and Hazen (2003). None, to the author’s knowledge, have employed the absolute values of the differences between interest rates, as done in the current research.

The trial and error method described above led to the first breakthrough described in the text. The breakthrough, in turn, led to awareness of Cotes’ theorem (see Work 3). The combination of the breakthrough and Cotes’ theorem led to the twin concepts of the special form of a polynomial and the special relationship between the roots and the coefficients of the special form. From then on the guessing methodology was no longer required because these latter concepts provide the starting point for most subsequent analyses.

\(^{19}\) An email exchange between the author and Dr Peter Carr is documented in the submitted works. Dr Carr expresses the belief that such meaning exists. His proposal is found in Osborne (2010a). It is not clear to the author how Dr Carr’s proposal can be usefully employed in finance. This is not to say that it cannot be done, merely that this author cannot see how. Dr Carr is currently Global Head of Quantitative Research at Morgan Stanley.
Appendix B

A perspective: why the twenty-first century and not the twentieth?

The author has speculated about why the analysis described in this thesis did not happen earlier. As documented in the text, most research into ‘multiple interest rates’ during the twentieth century considers real solutions to the TVM equations. Few researchers study the complex solutions. Yet, every equation has $n$ solutions and the complex solutions considerably outnumber the real. Why, given the profusion of complex results, has there been a dearth of research about what they could do or mean? Some possible explanations are offered below.

The (non)existence of algebraic solutions

Abel’s Impossibility Theorem is a long-standing result in mathematics (Niels Abel, 1802-1829). It states there is no algebraic formula for the roots of a polynomial of order five or above (see Stillwell, 1989). The general algebraic formulas for the roots of polynomials of order three and four exist, but they are complicated. For example, when modern symbolic mathematics software is employed to find the algebraic solutions to a fourth order polynomial, the output describing the equation for each root is long (many sides of A4 paper). The algebraic solutions to polynomials of order one and two are easy to find. The former is trivial and the latter is well known to students of elementary mathematics in the form $x = (-b \pm \sqrt{b^2 - 4ac})/2a$. Such low order polynomials only arise in the context of extremely simple economic or financial problems, e.g., cash flows having only two or three periods.

To summarise, it is impossible or extremely difficult to find algebraic solutions for the roots of most polynomials. The very simple polynomials that are open to algebraic analysis are associated with trivial economic and financial problems. The only alternative to algebraic analysis is to solve for the roots numerically.

Technology and the numerical solutions

Although numerical solutions are obtainable the process is not always straightforward. Early researchers (pre-1980s) were hampered by the fact that
numerical solutions for high order polynomials only became easy and cheap to obtain with the invention of affordable computers and the development of appropriate mathematical software. As documented earlier, suitable software became commercially available in the late 1980s at about the same time as the research in this thesis began. Appendix C contains discussion of the practical calculation of high order polynomials.

Unrealism

Another reason for avoiding the complex (and sometimes the highly negative) roots of the time value of money equation is that they seem unreal. Complex interest rates containing the square root of minus one, or highly negative rates, have typically been dismissed.

Boulding’s early judgement is noted in the text that ‘all but one of these roots will be either negative or imaginary, in which case they will have no economic significance’ (Boulding, 1936b). This judgement is an a priori judgement that was repeated from time to time during the interim period. For example, Soper (1959) writes, with no supporting argument, that ‘some of these roots can be ignored as irrelevant; those which are less than zero or are complex.’

During the twentieth century, Dorfman (1981) was alone in being willing, on theoretical grounds, to consider all solutions. In the current century, Hazen (2003) employs all solutions, although he employs only the real part of the complex solutions, ignoring the information locked up in the imaginary elements. Pierru (2010) attempts use and interpretation, but for a highly restricted financial situation.

These three researchers aside, Boulding’s a priori judgement seems to have been the common view. Several observations are offered on this state of affairs.

The first observation is the crucial conceptual difference noted in the text between, on the one hand, the solutions for \((1+r)\) that are complex and, on the other hand, the absolute values of the differences between solutions for \((1+r)\) that are real numbers. To date, all researchers who have applied themselves to this problem have thought in terms of the raw complex numbers. The new equations offered in this
thesis are couched in terms of the absolute differences between rates. These differences are real numbers. Thus, one of the findings of the current research is that one can begin with a real problem, end with a real solution, and complex numbers act as a catalyst on the way. This finding illustrates ‘a famous saying attributed to the French mathematician Jacques Hadamard (1865-1963): The shortest path between two truths in the real domain passes through the complex domain’. (Nahin, 1998)

The second observation is that charges of ‘unrealism’ call to mind Friedman’s essay on The Methodology of Positive Economics (Friedman, 1966). The essay is famous for advocating that realism of assumptions in economic theory is not important; ‘... the only relevant test of the validity of a hypothesis is comparison of its predictions with experience.’ Friedman’s essay has its critics; see Boland (1979) for an early account. Nevertheless the essay remains influential. It is suggested in one recent biography that Friedman saw his advocacy of positive economics as his greatest contribution to the subject (Ebenstein, 2007).

The time value of money equation does not capture a theory about human behaviour in the sense that Friedman meant. It is an equation that encapsulates market practice (bond mathematics) or the law (retail loans and the APR) or sound financial judgement (capital budgeting). Nevertheless there are parallels to be drawn between Friedman’s advice and the situation here.

Bond mathematics provides a particularly good example. The history of duration is essentially a long search for a simple, accurate expression to stand between the percentage change in the price of a bond and the causal change in the interest rate. History shows a series of attempts to derive a suitable expression, all of which are approximations (see Bierwag et al., 1983, for an early account and de Grandville, 2001, for a more recent one). The works in this thesis demonstrate that bringing all interest rates into play makes possible a simple, highly accurate expression. It is ironic that a journey into the complex plane is necessary to answer what many might judge to be an elementary question – what is the interest elasticity of the price of a bond? The ultimate justification for the journey is that it works.
The third observation about unrealism is drawn from physics. In an address to the Prussian Academy of Sciences, Einstein (1921) makes some comments on the use of mathematics in scientific theory. Ideas about how the world works are often expressed using mathematics of the intuitively acceptable variety. Statistical testing of the package of idea-plus-mathematics is a test of the idea, not of the mathematics’ suitability.

Einstein points out that mathematicians have long used the ‘axiomatic’ approach to their subject, resulting in mathematics not necessarily connected with the real world in an obvious way. Such mathematics might not be intuitively acceptable as a medium for expressing ideas about how the world works. Nevertheless it can be done. If such a package of idea-plus-mathematics is tested statistically against reality, then the test is not of the idea alone, but also of the suitability of the mathematics. The pragmatic approach to the test is that if the package of idea-plus-mathematics better fits real world observations than any alternative then it should be accepted. Intuition about the suitability of the mathematics is not necessarily a reliable guide to its successful use. The critical question is whether it works.

This thesis contains several demonstrations that the ‘multiple interest rate’ approach works. As mentioned above, in the context of bond mathematics, a new expression for duration is produced that employs the complex solutions and produces more accurate results than those of any prior expression. In the context of capital budgeting, there is a new expression for net present value per dollar invested that is not obvious but is illuminating. Also in the context of capital budgeting, new meaning is proposed for all solutions for IRR that, if correct, suggests a new interpretation of an orthodox interest rate. In the context of economic theory, a solution is suggested to the reswitching puzzle in the Cambridge capital controversies; the solution employs all possible rates of profit. Thus, new expressions incorporating all solutions to the time value of money equation enable more accurate calculations and produce illuminating results in a number of different fields. The critical question is whether the cumulative weight of these calculations and results, and their interpretation, can overcome pre-conceived ideas about the
A fourth observation about unrealism concerns the practical use of complex numbers. The need for complex solutions to equations has been apparent to mathematicians since the seventeenth century. However, there were no practical applications of complex numbers until the second half, indeed the last quarter, of the twentieth century. The core equations of quantum and relativity theories incorporate complex numbers; their practical applications include satellite navigation systems, imaging techniques such as MRI and PET, and circuit design in electronic devices such as computers, to mention just a few. A figure often quoted is that one third of US GDP is now dependent on applications of quantum theory.

Thus, the practical applications of physical theories employing complex numbers emerged comparatively recently. Therefore it is understandable that, in economics and finance, the complex solutions for the interest rate have been overlooked.
Appendix C

On the calculation of roots

A 30-year bond with semi-annual payments, e.g., a US Treasury bond, has 60 coupon payments. It follows that, on the day of issue, the polynomial employed to evaluate the bond has 60 solutions for the yield to maturity. Typically, given the simple structure of cash flows for a bond, only two solutions are real: the orthodox positive solution and a highly negative solution. The latter is usually discarded. The other 58, in the form of complex, conjugate pairs, are also discarded.

More extreme situations can be imagined. Integer numbers of sub-periods, identical in length, are necessary to analyse such polynomials. On the day after the issue of the same bond, the sub-periods of time into which the life of the bond is broken become irregular. This is because the first period is now one day less than six months. To make the sub-periods equal in length, there is no alternative to breaking the life of the bond into periods of one day each. In 30 years there are approximately 10,950 days. Therefore the order of the polynomial, and the potential number of solutions for the interest rate, rises to around 10,950. Most (99.9%) of these rates lie in the complex plane and are discarded.

Aspects of calculation technique

As observed in the text, and in Appendix B, a probable reason why the unorthodox roots have traditionally been discarded is that the calculation of thousands of solutions presents a computing challenge. Osborne (2001b) is a paper presented to experts in numerical computing at the conference Advanced Computing in Financial Markets (ACFM, 2001) to solicit their advice. The work is not submitted as part of this thesis because the contents replicate the findings of the works to that date. The questions asked at the conference were as follows:

The resulting polynomial is of order approximately 10,950, and its coefficients sparse. What are the limits to a computation like this? What order of polynomial can be factorised in a reasonable period of time on a good machine? And what degree of accuracy results? How far is it possible to go into the realm of practical bond calculations, given the required degree of accuracy?
The expert answers were:

- Use Matlab because it is designed to manipulate sparse matrices; the algorithm inside the program is important;

- The form in which the instructions are coded is important. The formulas involve the multiplication of distances between roots. The distances vary between approximately 2 and almost zero. If extremely small distances are multiplied together, then the interim products can be so small as to cause difficulty. Either the interim products go so close to zero that they are registered as such, in which case the final product is reported as zero (when logic suggests otherwise), or the software gives an error message. If, via the coding, the software is forced to pair small distances with large distances then the resulting multiplications complete successfully.

- As might be expected, the power of the computer also matters. The size of the random access memory, the speed of the processor and the number of processors affect how quickly solutions are output;

- If multiple processors exist, the software has to be able to take advantage of them and it cannot always be assumed that this is the case.

*The upper bound*

What is the upper bound of the problem (the largest order of polynomial required by the analysis)? In the previous section, past research was quoted suggesting an upper bound of around 10,000 roots. Since that research was done, reasons have developed to suggest the figure may be higher.

For the reasons outlined in the text, assume a ‘base’ interest rate of one basis point, i.e., \( q=0.0001 \). Also assume a historically high interest rate of 20% over the entire life of a long investment such as a 30-year US Treasury bond. An estimate of the upper bound is \( n \) in the equation \((1.0001)^n=(1.2)^{30}\). The figure is approximately 55,000.
**Computing time**

In Osborne (2005), Work 7, estimates are given of times to calculate the roots of random polynomials of different orders. The estimates are in the table below. More recent estimates using a different computer are also in the table. The software is Matlab in both cases.

<table>
<thead>
<tr>
<th>Number of roots (order)</th>
<th>Time (seconds)*</th>
<th>Revised times**</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>100</td>
<td>0.020</td>
<td>0.024</td>
</tr>
<tr>
<td>200</td>
<td>0.120</td>
<td>0.157</td>
</tr>
<tr>
<td>500</td>
<td>3.946</td>
<td>1.196</td>
</tr>
<tr>
<td>1,000</td>
<td>29.90</td>
<td>9.266</td>
</tr>
<tr>
<td>2,000</td>
<td>310.4</td>
<td>76.03</td>
</tr>
<tr>
<td>4,000</td>
<td>§</td>
<td>614.1</td>
</tr>
<tr>
<td>5,000</td>
<td>§</td>
<td>1,188</td>
</tr>
<tr>
<td>10,000</td>
<td>§</td>
<td>§</td>
</tr>
</tbody>
</table>

§ Error: insufficient memory

**Equipment employed to calculate the times in the table above**

<table>
<thead>
<tr>
<th></th>
<th>*</th>
<th>**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>2004</td>
<td>2010</td>
</tr>
<tr>
<td>Machine</td>
<td>PC</td>
<td>Apple</td>
</tr>
<tr>
<td>Processor</td>
<td>Intel Pentium 4</td>
<td>Intel Core 2 Duo</td>
</tr>
<tr>
<td>Processor speed</td>
<td>2.53 GHz</td>
<td>2 GHz</td>
</tr>
<tr>
<td>Memory</td>
<td>256 MB</td>
<td>2 GB</td>
</tr>
</tbody>
</table>

One final remark concerns the need for such calculations. For some applications, calculation of the complex solutions is not necessary; the theoretical insights are enough. In such situations, questions about the time and computing power to calculate the solutions are moot.
Appendix D
An example of chaotic behaviour: using Newton’s method to find the roots of a polynomial.
The polynomial illustrated is $f(z) = z^3 - 2z + 2$. The locations of the roots are indicated by the white crosses at $-1.7693$ and $0.8846 \pm 0.5897i$. The origin is the white dot at the centre of the figure. ‘Colors indicate to which of three roots a given starting point converges; black indicates starting points which converge to no root …’ [The iterations go into an endless cycle.]
Source: Hubbard et al. (2001)
References


Athanasopoulos, P. 1978. A note on the modified internal rate of return and investment criterion, The Engineering Economist, 23(2)


Bernhard, R. 1979. A more general sufficient condition for a unique internal rate of return, Journal of Financial and Quantitative Analysis, 14(2)


Boland, L. 1979. A Critique of Friedman's Critics, Journal of Economic Literature, 17(2)


Boulding, K. 1935. The theory of the single investment, Quarterly Journal of Economics, 49(3)

Boulding, K. 1936a. Time and investment, Economica, 3 (10)


Eagle, D., Kiefer, D. & Grinder, B. 2008, MIRR vs. IRR: Exploring the logic of the incremental reinvestment assumption, Journal of International Finance and Economics, 8(4)

Eatwell, J. 1975. A note on the truncation theorem, Kyklos, 28(4)

Ebenstein, L. 2007. Milton Friedman, Palgrave Macmillan


Kaplan, S. 1965. A note on a method for precisely determining the uniqueness or non- uniqueness of the internal rate of return for a proposed investment, *The Journal of Industrial Engineering*, 16(Jan/Feb)


Osborne, M. & Osborne, M. 2003. A polynomial that is a statistical prism, *Mathematical Gazette*, 87(508)


Osborne, M. 2005. On the computation of a formula for the duration of a bond that yields precise results, Quarterly Review of Economics and Finance, 45(1)


Pasinetti, L. 1966. Changes in the rate of profit and switches of techniques, Quarterly Journal of Economics, 80(4)

Pierru, A. 2010. The simple meaning of complex rates of return, The Engineering Economist, 55(2)


Samuelson, P. 1937. Some aspects of the pure theory of capital, Quarterly Journal of Economics, 51(2)

Samuelson, P. (1966) A summing up, Quarterly Journal of Economics, 80(4)


Work 1
Osborne, M.
Financial chaos
*Management Accounting*, 68 (10), 1990
FINANCIAL CHAOS

Michael J. Osborne BA, MA, MPhil, PGCE, Head of the Banking Studies Diploma Programme at Bahrain Institute of Banking and Finance, points out a new pitfall in applying the IRR.

The concept of the internal rate of return (IRR) is widely used in investment analysis. From one end of a portfolio to the other, from the appraisal of direct investments in physical projects to the assessment of the latest piece of financial engineering in the money market, the IRR may be used as one of a number of possible decision criteria.

The concept is not without its problems. It is the purpose of this note to point out yet another drawback to its use. The next section, by way of introduction to the subject, outlines the problems that are familiar to most users of the technique. The third section provides an explanation of the new pitfall to look for when applying the IRR. The pitfall emerges as yet another application of a relatively new branch of mathematics: chaos theory. The final section provides a summary.

The IRR and its drawbacks

Equation (1) describes in a precise way the relationship between the four components of an IRR calculation:

\[ \text{NPV} = \sum_{t=0}^{n} \frac{\text{CF}_i}{(1 + r)^t} \]

where:
- \( \text{NPV} \) = net present value of an investment
- \( \text{CF}_i \) = the cashflows at time \( t \) from the investment
- \( n \) = the number of time periods (time 0 is the present)
- \( r \) = the discount rate (shown as a decimal, not a percentage)

Given the cashflows (\( \text{CF}_i \)) and the number of time periods (\( n \)), the IRR is the value of the discount rate (\( r \)) which reduces the net present value (NPV) to zero.

The IRR has drawbacks familiar to anybody who has studied the concept. The calculation has the inbuilt assumption that all cashflows are reinvested or discounted at the same rate. In the real world interest rates vary through time, and opportunities for placing money do not appear in a steady homogeneous stream; therefore it is most unlikely that positive cashflows received during the life of an investment will be reinvested at a constant rate.

Sometimes the IRR can be difficult to calculate; except for the simplest cashflows there is no analytical way of calculating an IRR. Numerical methods have to be used which, described crudely, involve guessing the answer until one is found that works. The guessing procedure can sometimes produce strange results on a spreadsheet or financial calculator. These results are often attributed to the (usually unknown) algor-
in sign, therefore there are two real IRRs (−8.7 and 14.98 per cent). Figure 1 shows the relationship between the NPV and the IRR for this cashflow.

The reaction to this kind of ambiguity varies. Some dismiss the concept as misleading and look for other criteria. Others see the first positive real interest rate as an acceptable result. The latter reaction has its difficulties. If the first positive real rate is sought then the logical answer to the request for a seed or first guess is 0 per cent. With the cashflow in Table 1 and a seed of 0 per cent, the outcome is an IRR of −8.7 per cent, not the first positive rate of 14.98 per cent. Clearly the choice of first guess is an important part of the process and the obvious choice may not yield the ‘required’ result.

All of the problems outlined so far are to be found in the literature on the IRR and are reasonably well known. However there is another difficulty that is not so well known and it is the subject of the remainder of this note.

Financial chaos

Consider the cashflow in Table 2. This cashflow is an example taken from the handbook to a well known financial calculator, the Hewlett Packard HP12C (edition 1 dated July 1987). The figures are described in the handbook as ‘unconventional’. They can be thought of as the result of an initially profitable direct investment of 180,000; the first five months yield positive net cashflows. Then things go wrong: losses appear over a five-month period and improve only slightly to a zero return for nine months.

Eventually the project is terminated, which results in a single positive cashflow in the final period.

<table>
<thead>
<tr>
<th>Period</th>
<th>No of months</th>
<th>Cashflows</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>−180,000</td>
</tr>
<tr>
<td>1 to 9</td>
<td>5</td>
<td>100,000</td>
</tr>
<tr>
<td>10 to 19</td>
<td>5</td>
<td>−100,000</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>200,000</td>
</tr>
</tbody>
</table>

The three changes of sign cause the cashflow to yield three real IRRs: 1.86, 14.35 and 29 per cent. The choice of a seed of first guess has already been shown to be an important determinant of the outcome in an IRR calculation. Figure 2 shows the relationship between the NPV and the IRR for the cashflows in Table 2. Study of the graph may suggest that seeds between 0 and the point A (approximately the low point of the trough) will converge on 1.86 per cent; seeds between A and B (approximately the peak of the function) will converge on 14.35 per cent while seeds to the right of B will converge on 29 per cent.

The precise values of A and B may not be clear; they may be the trough and peak respectively but it is not certain. However, intuition suggests that proximity of the value of a seed to the eventual outcome may well be important. A few calculations are enough to test the hypothesis.

In Table 3 the first two columns show the IRRs which result from the seeds 0 to 10 per cent moving in steps of 1 per cent. While most of the results support the hypothesis — small values lead to 1.86 per cent and large values lead to 14.35 per cent — there is one very odd result. At about the lowest point of the trough, at a value of the seed of 6 per cent, the IRR jumps to the high solution of 29 per cent. This result runs counter to intuition. Furthermore the jump does not even take place at the interface between 1.86 and 14.35 per cent but occurs slightly before the interface. The IRR jumps up to 29 per cent from 1.86 per cent then down again before the switch to 14.35 per cent takes place.

Curiosity aroused we can go on to investigate the jump from 1.86 to 29 per cent more closely. Columns 3 and 4 in Table 3 show the result of using seeds between 5 and 6 per cent in steps of 0.1 per cent. They show that the actual jump occurs between 5.9 and 6 per cent. Columns 5 and 6 take the process a stage further by showing the outcomes for the IRR with seeds between 5.9 and 6 per cent, the step being further reduced to 0.01 per cent. The result is even more curious. Between the two outcomes of 1.86 and 29 per cent (using the seeds 5.93 and 5.98 per cent respectively) the value of 14.35 per cent appears three times.

Not only that, but 1.86 per cent also appears jammed between the jump from 14.35 to 29 per cent at a seed value of 5.97 per cent.

A pattern is beginning to emerge: every time the outcome jumps between two of the three IRRs the third IRR inserts itself between them. This result is not immediately apparent, rather it appears as the values of the seed are taken to finer degrees of accuracy. An investigation of the interface between the IRRs of 1.86 and 14.35 per cent in column 6 at initial guesses of 5.93 and 5.94 per cent is shown in columns 7 and 8. This time the range of seeds is from 5.930 to 5.940 per cent in steps of 0.001 per cent. The pattern is confirmed: at the interface between 1.86 and 14.35 per cent the third IRR of 29 per cent appears at the seed value 5.939 per cent.

Columns 9 and 10 reconfirm the pattern for another interface at the next level of accuracy. Indeed at every interface between two of the three IRRs an ‘island’ can be found which leads to the third outcome.

Note also that the area in which these results appear is in the region of the low point A in Figure 2, a turning point in the relationship between the NPV and the IRR. This was also the area in which the errors occurred in the previous example (the high point in Figure 1). It should therefore come as no surprise to learn that a similar area containing this behaviour can be found in the region of the point B in Figure 2, i.e. an IRR of around 21 per cent.

What is happening? The remarkable behaviour being exhibited is the result of the non-linear character of the equation linking the NPV to the IRR. That such behaviour is typical of non-linear equations, even the simplest of such equations, was not fully realised until the 1970s when the biologist Robert May stumbled upon the result. The phenomenon comes under the heading of what has come to be called ‘chaos theory’.

At the heart of the theory is the finding that deterministic equations may display chaotic behaviour or behaviour with a pattern so deep or subtle that it is difficult to discern. Excellent introductory accounts of the subject can be found in books by Gleick1 or Stewart.2

Summary

The outcome of an IRR calculation is remarkably sensitive to the choice of seed. It is quite possible that, in a complex cashflow, a small positive whole number chosen as the seed will not necessarily lead to the first real rate above zero; moreover the seed could be on one of the ‘islands’ of values which lead to a result far from that suggested by intuition.

The regions from which seeds ought not to be selected if the problem is to be avoided, the regions in which the ‘islands’ occur, are the turning points in the relationship between the NPV and the IRR. These turning points are most easily determined from a chart, like Figure 2, which shows the behaviour of the NPV and IRR.

1 There are actually seven solutions to the IRR problem posed by the cashflows in Table 1 but four of them are complex, i.e. they involve 1 = V1, and they are therefore dismissed as financially irrelevant.
Work 2

Osborne, M.

Chaos and the internal rate of return

*Computers in Higher Education Economics Review, 7 (19), 1993*
Chaos and the Internal Rate of Return

Most spreadsheets and financial calculators have a routine built into them to enable the user to calculate the internal rate of return (IRR) of a cash flow. Most textbooks on finance contain a section on the use of the IRR as a tool for investment appraisal or financial analysis. Both spreadsheet manuals and textbooks usually contain some warnings about using the IRR. The pitfalls listed in Brealey and Myers [1] are an example of this. The main idea in this article is the elaboration of a difficulty posed by the existence of one such pitfall, that of multiple solutions.

Not all cash flows pose this difficulty, but where they do, the algorithms used by spreadsheets and calculators to estimate the IRR are shown to display another problem not discussed in the literature. The problem is not only that multiple solutions exist but also that the algorithms may lead to the various solutions in a confusing and apparently haphazard way. The problem is explained with the help of the new mathematical theory of chaos and it is shown that the confusion is more apparent then real. It is a subsidiary idea in the article to explain the particular form of chaos that arises for the benefit of those who seek applications of the theory to finance. The literature on the applications of chaos theory to finance, typified by Peters [2], has so far not mentioned the issue discussed here. In teaching the IRR and its pitfalls to students it has been my experience that the problem, and its elucidation using chaos theory, is intriguing enough to capture more than the usual amount of attention.

The results quoted in the article are, for the most part, those obtained from Lotus 1-2-3 release 2.3. It will be shown that it is in the nature of the problem that results from other spreadsheets and calculators may be different. Some results from another release of 1-2-3 (r3.1+), other programs (Excel r4.0) and machines (HP 19b and Casio FC 1000) are given to illustrate the point. A DOS version of Lotus 1-2-3 has been chosen as the default program because surveys by the computer press suggest that it is still the most commonly used spreadsheet in use in the business world.
Consider the simple cash flow given in Table A. It can be thought of as the cash flow for an investment project. It is not a forecast or textbook cash flow of the kind used by accountants when doing an appraisal of some future project (such cash flows are usually simple and well-behaved mathematically speaking). Instead, it could be a look back at the history of an ongoing project from the point of view of, say, a banker who wishes to know the financial yield of the project so far. The numbers could refer to millions or thousands of units of currency -- it is not important. It is the pattern of the flow that is critical.

For the first three periods money is invested in a project and in the fourth period the project starts to pay back. This continues until period eight when the project falters and starts to run into trouble. Recovery eventually comes in period sixteen but in period twenty the net cash flow starts to falter again. If the project should be terminated what would be its yield to date?

<table>
<thead>
<tr>
<th>Period</th>
<th>Cash flow</th>
<th>Period</th>
<th>Cash flow</th>
<th>Period</th>
<th>Cash flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-2.0</td>
<td>7</td>
<td>3.7</td>
<td>14</td>
<td>-6.7</td>
</tr>
<tr>
<td>1</td>
<td>-1.3</td>
<td>8</td>
<td>-1.0</td>
<td>15</td>
<td>-3.0</td>
</tr>
<tr>
<td>2</td>
<td>-1.3</td>
<td>9</td>
<td>-5.5</td>
<td>16</td>
<td>2.1</td>
</tr>
<tr>
<td>3</td>
<td>3.2</td>
<td>10</td>
<td>-8.2</td>
<td>17</td>
<td>6.2</td>
</tr>
<tr>
<td>4</td>
<td>4.7</td>
<td>11</td>
<td>-9.4</td>
<td>18</td>
<td>8.3</td>
</tr>
<tr>
<td>5</td>
<td>7.3</td>
<td>12</td>
<td>-9.9</td>
<td>19</td>
<td>11.1</td>
</tr>
<tr>
<td>6</td>
<td>5.8</td>
<td>13</td>
<td>-9.7</td>
<td>20</td>
<td>8.5</td>
</tr>
</tbody>
</table>

Estimating the yield on the cash flow is, at first sight, easy. The function @IRR is applied to the cash flow. The only extra piece of information supplied by the analyst is a first guess for the yield so that the algorithm built into the spreadsheet can proceed to do its work. For example, a first guess of 0% gives a result of 2.90%. If the cash flows are, say, quarterly, then the implied return is 11.60% on an annualized basis.
What exactly is the IRR? The @IRR function is finding the rate of interest, r, which causes the net present value (NPV) of the project's cash flows to be zero. Mathematically this is expressed very succinctly:

\[ NPV = \sum_{i=0}^{n} \frac{C_i}{(1+r)^i} \]

where \( C_i \) is the cash flow in the \( i \)th period, \( n \) is the number of periods and \( r \) is expressed as a decimal, i.e., the interest rate divided by 100.

The equation can be written in a longer but more readily understood way using the data from Table A:

\[ NPV = \frac{-2.0}{(1+r)^0} + \frac{-1.3}{(1+r)^1} + \frac{-1.3}{(1+r)^2} + \ldots + \frac{11.1}{(1+r)^{19}} + \frac{8.5}{(1+r)^{20}} \]

Solving for \( r \) when \( NPV = 0 \) in equations like these is not easy. When the value of \( n \) is five or more, as it is here, there is no direct, analytical way of finding \( r \). Instead the value must be estimated using a search procedure until one is found that fits, which is why the program requires a first guess. As a matter of interest, what difference does it make if the first guess is not 0%? Table B shows the IRRs for guesses from 0% to 10% in increments of 1%. For most of the guesses the internal rate of return is 2.9%. But with a first guess of 8% the IRR given by the built-in function is 42.17%. With guesses of 9% and 10% the IRR that results is 12.68%. This is the problem of multiple solutions. Which one is correct?

The answer is that they are all correct. With a cash flow of twenty-one periods there are twenty possible answers to the question of what rate of interest satisfies the requirement that the NPV in equation 1 (or 2) is reduced to zero?
Table B
[All numbers are percentages]

<table>
<thead>
<tr>
<th>IRR</th>
<th>First Guess</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.90</td>
<td>0</td>
</tr>
<tr>
<td>2.90</td>
<td>1</td>
</tr>
<tr>
<td>2.90</td>
<td>2</td>
</tr>
<tr>
<td>2.90</td>
<td>3</td>
</tr>
<tr>
<td>2.90</td>
<td>4</td>
</tr>
<tr>
<td>2.90</td>
<td>5</td>
</tr>
<tr>
<td>2.90</td>
<td>6</td>
</tr>
<tr>
<td>2.90</td>
<td>7</td>
</tr>
<tr>
<td>42.17</td>
<td>8</td>
</tr>
<tr>
<td>12.68</td>
<td>9</td>
</tr>
<tr>
<td>12.68</td>
<td>10</td>
</tr>
</tbody>
</table>

Technically speaking there are twenty roots to a twentieth degree polynomial. With a program like Mathematica [3] it is possible to identify all twenty roots; something that spreadsheets cannot do. An even number of the roots, in this particular case sixteen of them, are complex roots which come in the form of complex conjugate pairs and involve the square root of minus one, an imaginary concept. The remaining roots, in this example four of them, are real roots. They imply interest rates of 2.90%, 12.68%, 42.17% and, oddly, -196.39%. Like many a mathematical answer to an apparently straightforward question there are more answers than are necessary. Only some of the answers are interesting from the point of view of the applied analyst. What is to be done?

The normal practice is to throw away any answers that make no sense in the context of the question. Clearly this means all the complex answers and the negative number; which leaves the three positive, real solutions. Then what? The usual procedure is to choose the first positive number above zero that takes us back to the figure of 2.90%. Since this solution resulted from a first guess of zero all that has to be done to get the "right" answer when using the @IRR function is to choose as the first guess the number zero. Correct? Surely a first guess of zero, or between the values of 0% and 10%, is bound to lead to the first positive number above zero. In fact this is not always so. Such a procedure would not guarantee the "right" answer.
To see why not, go back to Table B and modify it a little. Table B shows eleven first guesses from 0% to 10% in steps of 1% and the associated IRRs. Table C shows the result of homing in on one of the points of transition in Table B. The points of transition are highlighted. In particular, first guesses from 7% to 8% in steps of 0.1% are chosen and the resulting IRRs displayed. This shows the transition from an IRR of 2.90% to 42.17%. The transition occurs between 7.8% and 8.00%. The strange thing is that the intermediate IRR of 12.68% is found inserted between the two as a result of the first guess of 7.9%.

The next columns of Table C illustrate what happens when more transition points are highlighted at increasingly refined first guesses. An important observation to be made is that where a change in the value of the IRR occurs, as the result of a different first guess, the change does not take place cleanly. Instead another of the possible IRRs is always to be found inserted between the two. The pattern is to be found no matter how deep the analysis, no matter where the switch in values takes place, no matter how small the increment in the first guess. The result obtained in an IRR calculation is critically dependent on the choice of first guess. Furthermore the switching between IRRs seems to follow a pattern.

Note that an ERR (error) appears which shows simply that 1-2-3 was not able to find one of the four possible real solutions within the default number of iterations. It does not mean that an answer does not exist.

Note also that whether or not an ERR appears depends upon the release of 1-2-3 being used. If release 3.1+ is used then the solution -196.39% does not appear in the table; instead more ERRs appear.

Even more odd is the fact that the identical file can be picked up by Excel (or fed in by hand if you don't trust the translate function) and the pattern that emerges is different again. Table D shows the results. The numbers are restricted to the four possible solutions as we might have expected but they are not linked to the first guesses in exactly the same way.
### Table C -- Searching for the IRR using Lotus 1-2-3 r2.3

<table>
<thead>
<tr>
<th>Increment</th>
<th>Increment</th>
<th>Increment</th>
<th>Increment</th>
<th>Increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.01</td>
<td>0.001</td>
<td>0.0001</td>
</tr>
<tr>
<td>IRR</td>
<td>Guess</td>
<td>IRR</td>
<td>Guess</td>
<td>IRR</td>
</tr>
<tr>
<td>2.90</td>
<td>0</td>
<td>2.90</td>
<td>7</td>
<td>12.68</td>
</tr>
<tr>
<td>2.90</td>
<td>1</td>
<td>2.90</td>
<td>7.1</td>
<td>12.68</td>
</tr>
<tr>
<td>2.90</td>
<td>2</td>
<td>2.90</td>
<td>7.2</td>
<td>12.68</td>
</tr>
<tr>
<td>2.90</td>
<td>3</td>
<td>2.90</td>
<td>7.3</td>
<td>12.68</td>
</tr>
<tr>
<td>2.90</td>
<td>4</td>
<td>2.90</td>
<td>7.4</td>
<td>12.68</td>
</tr>
<tr>
<td>2.90</td>
<td>5</td>
<td>2.90</td>
<td>7.5</td>
<td>12.68</td>
</tr>
<tr>
<td>2.90</td>
<td>6</td>
<td>2.90</td>
<td>7.6</td>
<td>12.68</td>
</tr>
<tr>
<td>2.90</td>
<td>7</td>
<td>2.90</td>
<td>7.7</td>
<td>12.68</td>
</tr>
<tr>
<td>42.17</td>
<td>8</td>
<td>2.90</td>
<td>7.8</td>
<td>-196.39</td>
</tr>
<tr>
<td>12.68</td>
<td>9</td>
<td>12.68</td>
<td>7.9</td>
<td>42.17</td>
</tr>
<tr>
<td>12.68</td>
<td>10</td>
<td>42.17</td>
<td>8</td>
<td>42.17</td>
</tr>
</tbody>
</table>

### Table D -- Searching for the IRR using Excel r4.0

<table>
<thead>
<tr>
<th>Increment</th>
<th>Increment</th>
<th>Increment</th>
<th>Increment</th>
<th>Increment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.01</td>
<td>0.001</td>
<td>0.0001</td>
</tr>
<tr>
<td>IRR</td>
<td>Guess</td>
<td>IRR</td>
<td>Guess</td>
<td>IRR</td>
</tr>
<tr>
<td>2.90</td>
<td>0</td>
<td>2.90</td>
<td>6</td>
<td>12.68</td>
</tr>
<tr>
<td>2.90</td>
<td>1</td>
<td>2.90</td>
<td>6.1</td>
<td>42.17</td>
</tr>
<tr>
<td>2.90</td>
<td>2</td>
<td>#NUM!</td>
<td>6.2</td>
<td>42.17</td>
</tr>
<tr>
<td>2.90</td>
<td>3</td>
<td>#NUM!</td>
<td>6.3</td>
<td>42.17</td>
</tr>
<tr>
<td>2.90</td>
<td>4</td>
<td>#NUM!</td>
<td>6.4</td>
<td>42.17</td>
</tr>
<tr>
<td>2.90</td>
<td>5</td>
<td>12.68</td>
<td>6.5</td>
<td>42.17</td>
</tr>
<tr>
<td>2.90</td>
<td>6</td>
<td>42.17</td>
<td>6.6</td>
<td>#NUM!</td>
</tr>
<tr>
<td>12.68</td>
<td>7</td>
<td>42.17</td>
<td>6.7</td>
<td>#NUM!</td>
</tr>
<tr>
<td>12.68</td>
<td>8</td>
<td>#NUM!</td>
<td>6.8</td>
<td>#NUM!</td>
</tr>
<tr>
<td>12.68</td>
<td>9</td>
<td>12.68</td>
<td>6.9</td>
<td>42.17</td>
</tr>
<tr>
<td>12.68</td>
<td>10</td>
<td>12.68</td>
<td>7</td>
<td>42.17</td>
</tr>
</tbody>
</table>
With the HP 19b and Casio FC 1000 calculators the first column of the Table is different again. The switch in values between 2.9% and 12.68% occurs between first guesses of 7% and 8%. This results in yet another pattern of outcomes as the guesses are refined.

This phenomenon, that guesses drawn from some areas of the real number line yield consistently the same result, while guesses from other areas yield the whole range of possible results in an apparently chaotic and software-specific fashion, is typical of the numerical procedure embodied in the software. Lotus 1-2-3 and Excel, along with most other spreadsheets and calculators with built in financial functions, use a very efficient estimation procedure to obtain their solutions -- an algorithm known as Newton-Raphson, or simply Newton's Method. The algorithm is efficient to the extent that it is extremely fast. But, as is apparent from the examples, it is also subject to this intriguing ambiguity.

The difficulty with Newton-Raphson was first explored systematically by an American mathematician called John Hubbard in 1976. If complex, as well as real, numbers are included in the analysis then the apparent chaos in the appearance of solutions becomes a beautiful pattern. An account of Hubbard's discovery and a computer graphic to illustrate the point is contained in Gleick [4]. At the heart of Hubbard's analysis was the discovery that the solutions found by Newton's method are extremely sensitive to initial conditions. The term chaos is a misnomer because there is in fact pattern in the numbers that results from different first guesses. The pattern is difficult to see because the spreadsheet is limited to real numbers and because the whole range of solutions is not normally explored. Wegner and Peterson [5] provide routines to create the computer graphics based on Newton's method.

Is there anything significant about the area where this phenomenon occurs? Why does one range of first guesses result in the switches between IRRs and another range cause the IRR function to yield consistently a particular IRR? Chart A provides some clues. The chart shows the relationship between the rate of interest, \( r \), and the Net Present Value (NPV) of the cash flows in Table A. The points at which the curve crosses the horizontal axis are the rates of interest where the NPV is equal to zero, i.e., the internal rates of return.
The rates of interest chosen as first guesses that give rise to the strange behavior occur between 7% and 8%. This range is where the curve is experiencing a turning point, where the slope changes from negative to positive, where the curve is approximately horizontal. When the Newton-Raphson algorithm starts its search from a value in this region it does not necessarily home in on the nearest IRR (root); any of the real solutions are possible. In contrast first guesses away from the turning points do result in stable results; the algorithm homes in on the nearest solution. To see why this happens a closer look at the Newton-Raphson algorithm is necessary. Equation 3 gives the algebra.

\[ r' = r - \frac{f(r)}{f'(r)} \]

where \( f(r) \) is the function [1] or [2] linking the NPV with the rate of interest and \( f'(r) \) is the first derivative of this function.

A guess of \( r \) is fed into the right hand side. The new value of \( r \), called \( r' \), is then put back into the right hand side of [3] to give a new value, \( r'' \), and this, in turn, is fed back into the equation. The iteration continues until the output differs from the input by less than a small specified amount (0.00001 % in both Excel and 1-2-3). The resulting value of \( r \) is one of the solutions.
Normally this procedure is extremely efficient in the sense that a solution is found in only a few iterations. However, if the first guess is chosen from an area where the first derivative is close to zero then the second element on the right hand side of [3] becomes very large indeed. The addition or subtraction of this element from the first guess results in the next iterated value being very far away. It may be so far away that it causes the iteration to settle down, not to the closest of the solutions on either side of the first guess, but to a solution some distance away. There is more. During the iteration, one or more of the values of r that are calculated may be within one of the regions of low slope, and so involve values of the first derivative close to zero. The result is the same. The iteration may bounce around for quite a while before coming to rest. There is a default number of iterations in a spreadsheet (20 in Excel and 30 in 1-2-3) and if a solution does not appear within that default number then an error will be reported as we have seen in Tables C and D. In short, a first guess of the value of r taken from an area where the curve has a slope of almost zero may result in any one of the possible solutions. The nearest solution is not guaranteed.

Differences in the level of precision in the software, or the machine, can affect the outcome because the same first guess can lead to different subsequent values because of a different degree of rounding. Even the minute alterations caused by rounding can, when felt in areas where the first derivative is close to zero, result in the iteration taking a very different path through the figures. This is the most probable cause of the different structure of results in Tables C and D for 1-2-3 and Excel respectively.

There is a lesson here: if ambiguity is to be avoided in a cash flow where multiple IRRs are possible then the NPV-interest rate relationship ought to be graphed so that the turning point(s) of the curve can be established and guesses from these regions avoided. The region to be avoided is not small. For example, readers may surmise, correctly, that a similar effect is to be found in the region of the second turning point. First guesses of between 33% and 35% in 1-2-3 r2.3 also result in the algorithm homing in on all the possible IRRs in a chaotic manner. For Excel r4.0 the region is 20% to 25%.
At last, it is possible to give an answer to the question posed earlier of why a first guess of zero, or within the recommended range of 0% to 10%, may not be the ideal choice. It may be that the cash flow is such that the turning point in the NPV-interest rate relationship occurs at, or close to, an interest rate of zero. The first guess of zero, the most likely to be made, will be drawn from a region that leads to the possibility of ambiguity. The possibility will not be recognized unless a visual inspection of the NPV-interest rate relationship takes place. Without the visual inspection an unsuspecting analyst could punch in a first guess, accept the resulting IRR without question if it is feasible, not realizing that he/she is operating in a difficult area.

In some spreadsheets it is either mandatory, or there is an option, to leave out the first guess from the @IRR command and allow the software to choose it. The option exists in Excel. The program starts with a guess between 0.1 % and 10% and the resulting IRR in the example is 12.68%. Given the way in which the outcome is so sensitive to the initial conditions it is dangerous to leave the choice of first guess to the software. Programs where the user is denied any choice of first guess are particularly awkward because calculation of the "correct" IRR (if such a thing can be said to exist in these circumstances) can only be made from the @NPV function or the chart showing the NPV-interest rate relationship.

Another method of calculating the IRR is to use BSOLVER (1-2-3) or Goalseeker (Excel) on the @NPV function. In the example the result of asking for the rate of interest that gives an NPV of zero is 2.9% in both cases, which is the first rate above zero. It is not clear that BSOLVER or Goalseeker will always be so obliging because they too use the Newton-Raphson method. In a previous issue of CHEER, Lorenz [6] gives a macro for implementing Newton-Raphson in versions of 1-2-3 that do not have BSOLVER. Lorenz calls the macro Goalseek. He notes that "if ... a ... formula has more than one root the result ... will depend on the ... start." This whole article can be seen as an extended explanation of the considerable difficulty involved in that harmless looking sentence.

How likely is the phenomenon to arise? Simple cash flows with only one change of sign are not
likely to cause problems precisely because of their simplicity. Accountants most often generate such cash flows when performing project appraisals. The data is normally describing expectations about the future. Complicated cash flows are not envisaged and would be regarded with suspicion. To such users this phenomenon would indeed be a mere curiosity.

The problem is most likely to arise when there are two or more sign changes in the cash flow. This yields the possibility that there is more than one real answer to the question of what is the IRR, and, what is more, that two of the solutions could occur in a feasible region. It is particularly acute when there are three or more sign changes because it is even more likely that multiple solutions are in a feasible region. Cash flows of this kind are more likely in banking and finance. Corporate lending can involve complex structures involving drawdowns and paybacks throughout the life of a loan agreement. Also, actual cash flows, seen in retrospect, are messier than ideal, forward-looking cash flows. Calculation of the yield on such complex, actual loan structures implies use of the IRR concept and exposes the user to the pitfall described here. The pitfall is that a first guess may lead to any one of the real solutions, not necessarily the one closest to the first guess, and perhaps not the IRR that would be chosen if the full NPV-interest rate profile were known.

Mike Osborne
Bahrain Institute of Banking and Finance

_I am grateful to Ian Mcveigh of the University of Southern Queensland, Australia, and Eggert Peterson of the University of Bahrain for helpful comments on the subject._


References


Work 3
Osborne, M.
Visualising financial concepts in the complex plane
http://ideas.repec.org/a/che/chepap/v14y2000i1p4-8.html
Visualising financial concepts in the complex plane

Michael J. Osborne

The shortest path between two truths in the real domain passes through the complex domain.
Jacques Hadamard (1865-1963)

Introduction

Any student of economics and finance is likely to meet complex numbers. For example, they will encounter them when studying the stability of difference equations used in business cycle analysis (see Turner [1993]).

This article shows how the complex plane can be put to a new purpose - to give a novel and visual interpretation of two important financial measures, duration and the internal rate of return. Only the most cursory of descriptions of these two concepts is given here because good, orthodox descriptions can be found in any of the many texts on finance or financial economics (for example, see Fabozzi [1996] and Cuthbertson [1996]).

The financial algebra

Consider a series of numbers, $a_i$. Place these numbers in the particular setting of equation [1], i.e., embed them in a polynomial. Equation [1] is in the form most often found in books on finance, it is the value of money equation, while [2] is the same thing recast in the more general form found in maths books.

\[ -1 + \frac{a_1}{(1+r)} + \frac{a_2}{(1+r)^2} + \ldots + \frac{a_{(n-1)}}{(1+r)^{n-1}} + \frac{a_n}{(1+r)^n} + \frac{1}{(1+r)^n} = 0 \]  

\[ x^n - a_nx^{(n-1)} - a_{(n-1)}x^{(n-2)} - \ldots - a_2x + a_1 = 0 \quad \text{where} \quad x = (1+r) \quad \text{[2]} \]

According to equation [1], we pay 1 out, subsequently receive a stream of returns, the $a_i$, and get 1 back alongside the last return. Any stream of numbers can be accommodated in this structure but for the moment we focus on the fixed income bond market. Equations [3] and [3a] show the algebra of the bond pricing formula in the familiar notation of a financial calculator. Equation [3] is in the same format as [1].

\[ \frac{\text{PMT}}{\text{PV}} + \frac{\text{PMT}}{(1+r)^2} + \ldots + \frac{\text{PMT}}{(1+r)^{n-1}} + \frac{\text{PV} - \text{FV}}{(1+r)^n} = 0 \]

Equation [3] is a rearrangement of [3a] the more familiar version of the bond pricing equation.

\[ \frac{\text{PV}}{(1+r)} + \frac{\text{PV}}{(1+r)^2} + \ldots + \frac{\text{PV}}{(1+r)^{n-1}} + \frac{\text{FV}}{(1+r)^n} = 0 \quad \text{[3a]} \]

Setting $PV - FV = 1$ returns the equations [3] and [3a] to the ones of a par bond.

We normally think of a bond as having a stream of equal returns in which all the $a_i$ are equal but [3] shows that we can reframe the equation in the form of [1] in which the stream of numbers is constant for $(n-1)$ periods, but different in the final period, $a_n$. This single difference means that we have a variable stream.

The heart of this article is the more general analysis of a variable stream in which the $a_i$ can take any values. In particular we examine the way in which the $a_i$ play a part in determining the internal rate of return and duration via the complex plane.

The internal rate of return is that value of $r$ that satisfies equations [3] or [3a] given values for all the other variables. It is a measure of the return per period on the stream of numbers, or the yield on a bond.

Duration is the first derivative of the PV (or price) with respect to $r$ in [3a] divided by PV. It shows the interest elasticity of the price of a bond and is therefore a measure of risk. Duration has a number of valuable properties, for example, the duration of a weighted portfolio of assets is equal to the weighted average of their durations. It is useful to know when hedging risk.

Some foundations

Equation [1] is an $n$th degree polynomial therefore it has $n$ roots. Usually the first real root above zero, $(1+r)$, is the only solution in which we are interested during any
financial analysis. This number is variously interpreted as 1 plus the internal rate of return (or the real rate, the annual percentage rate, the marginal efficiency of capital, etc., depending on the interpretation of the \( a_i \)) A simple but subtle question arises at this point, a question that is not usually asked. What do the other \((n-1)\) roots mean?

Reference to any maths text on the subject shows that any \(n^{th}\) degree polynomial with real coefficients will have \(n\) roots that will either be real or occur in complex conjugate pairs (we assume none of the coefficients are complex). The fact that the complex results have an imaginary dimension means that their interpretation in a financial context is not obvious. However, the vast amount of meaning that exists for complex numbers in scientific and engineering theory is an encouragement to look deeper (see Riley et al [1997] for a text, or Nahin [1998] for a book aimed at the lay-person).

From here on, a four period example is used to illustrate most results. This simple case is chosen because it is particularly easy to visualise the results when there are only four roots. The results are easily generalised to \(n\) periods with \(n\) roots. Where it is easy to do so, the \(n\) period case is given.

The simplest case is where the \(a_i\) are zero, ie, FV and PV are equal to 1 (there is no capital gain/loss) and all the returns are zero. Then equation [1] reduces to [4].

\[
\frac{1}{(1+r)^4} = 0 \tag{4}
\]

The four roots of equation [4] are evenly dispersed around the unit circle in the complex plane (Diagram A shows the relevant Argand diagram).ii

![Diagram A](image)

This result easily generalises to the \(n\) period case where there are \(n\) roots evenly dispersed around the unit circle with equal angles (\(2\pi/n\) radians or \(360/n\) degrees) between them. The general version of [4] with \(n\) roots is known as the cyclotomic equation [4a]. If there were 100 periods there would be 100 roots distributed around the unit circle with angles of \(2\pi/100\) radians, or 3.6 degrees, between them.

\[
x^n-1 \text{ where } x=(1+r) \tag{4a}
\]

For the next case assume that the \(a_i\) are put into the equation and further assume, for the moment, that they are equal. The equation is that of the par bond. PV and FV are both equal to 1, and the \(a_i\) are equal to the coupon. In the four period case, three \((n-1)\) of the roots are in exactly the same position as for the cyclotomic equation, while the fourth, the one positive real root, now takes the value \((1+r)\) instead of unity. The result of adding in a steady stream of returns of value \(a\) is simply to move the positive real root a distance \(r\) to the right along the real axis, ie, \((1+r)=(1+a)=(1+\text{coupon})\). The result is depicted in diagram B.

![Diagram B](image)

This result also generalises very easily. As the value of the \(a_i\) or the coupon, rises and falls so the real root moves to the right or the left along the real axis by equal amounts. The remaining roots, no matter how many exist, are once again distributed evenly around the unit circle.

Now allow the \(a_i\) to vary in any way. For example, we can imagine the non-par bond case with all coefficients remaining constant except for the last. The last coefficient changes to reflect PV departing from the value of 1 (see equation [3]). When the \(a_i\) vary, the \((n-1)\) roots depart from their positions on the unit circle. They become points on the ends of tendrils attached to the cyclotomic roots. These tendrils can be considered measures of dispersion from the even-flow case. In diagram C there are three such tendrils (the \(t_i\)). The mathematical literature, see Erdos and Turan [1950], suggests that the roots of a polynomial with random coefficients are dispersed around specific points on the unit circle.iii

![Diagram C](image)

At the same time the \(n^{th}\) root, the first positive real root, remains at its customary position on the real axis showing the rate of return on the overall financial flow (the 'middle' of the \(a_i\)). It is the output from the IRR button on a financial calculator. It is a kind of average. Indeed, below, it will be shown formally, in a surprising way, that the rate of return, \(r\), is a particular kind of average of the \(a\).

Initially it is difficult to discern patterns in the positions of the roots. Professional mathematicians researching into the subject of the roots of polynomials with varying coefficients speak of its capacity to surprise. This is partly because of the difficulty of visualising what is happening in the complex plane and partly because of the complexity of the calculations. For example Farahmand [1998] says that 'random polynomials, so simple and innocent at first sight, are among the most fascinating and mysterious in mathematics. Although some of their unexpected and amusing behaviour has been known for as long as a century, yet they still reveal their secrets so that many of those both closely and indirectly involved with the subject are aroused'.

Fortunately, there exists a theorem in complex analysis that helps. It allows duration and the internal rate of return to be given simple, intuitive, geometric interpretations involving distances in the complex plane.

---

*CHEER* Volume 14 • Issue 1 Page 5
Cotes theorem

The interpretation relies on a theorem by Roger Cotes who was a mathematician at Cambridge about 300 years ago (see Nahin [1998] or Needham [1998]). Consider diagram D. Place a point, $x$, on the real axis inside the unit circle. The point is at a distance $g$ from the origin. The distances of $x$ from the cyclotomic roots are labelled $d_i$ where $i$ ranges from 1 to 4. Cotes' theorem shows the following:

$$\prod_{i=1}^{4} d_i = d_1 \cdot d_2 \cdot d_3 \cdot d_4 = R^4 - g^4$$  \hspace{1cm} (5)  

where the radius $R = 1$ and $g < R$

or

$$\prod_{i=1}^{4} d_i = d_1 \cdot d_2 \cdot d_3 = g^4 - R^4$$  \hspace{1cm} (5a)  

This immediately gives rise to a conjecture. Is the formula in the left hand side of (5a) also equal to modified duration when the are $a_i$ allowed to vary? For example, is it still true in the case of non-par bonds when the last coefficient is different from the rest, or, indeed, when we consider any stream of numbers in which each of the $a_i$ is different? The answer is yes. The proof is worth giving in full here because of its elegant simplicity and the insight it affords into the concept of duration.

The general equation for duration

Write down equation [1] $n$ times in $n$ rows to form a matrix, reverse the sign of each element, divide throughout by $(1+r)$ and call it equation [7]. In our particular example there are four rows:

$$+ \frac{1}{(1+r)^2} \cdot \frac{a_1}{(1+r)^3} - \frac{2a_2}{(1+r)^3} - \frac{3a_3}{(1+r)^3} - \frac{4a_4}{(1+r)^3} \cdot \frac{1}{(1+r)^3} = 0$$ \hspace{1cm} (7)$$

The matrix is the null matrix because each row sums to zero. Partition the matrix diagonally to include the whole of the first column, the top three elements of the second column, the top two elements of the third column and the top element of the fourth column. The partition to the top left we call $P_1$. The partition to the lower right we call $P_2$. The elements of the matrix sum to zero, therefore $P_1 = -P_2$.

Consider the two partitions in turn. First $P_1$ because it is easier and it has the more familiar interpretation. Write down the elements of $P_1$ in the form of an equation, slightly rearrange the final term, and it is recognisable as modified duration [8]. This could be for a par or a non-par bond, because the $a_i$ have not been specified, they can take any values.

$$- \frac{a_1}{(1+r)^2} - \frac{2a_2}{(1+r)^3} - \frac{3a_3}{(1+r)^3} - \frac{4(a_4+1)}{(1+r)^3} = D$$ \hspace{1cm} (8)$$

We know that $P_1 = -P_2$, therefore $P_1$ must also be equal to the negative of modified duration. The extra meaning of the partition $P_1$ is not so obvious. We have to show it is equal to the left hand side of equation [6] - the product of the $(n-1)$ distances between the $(n-1)$ roots and the real $n^{th}$ root, divided by $(1+r)^4$.

We follow the line of reasoning set out in Nahin [1998]. Take the universal bond pricing equation [1], reverse all the signs, multiply throughout by $(1+r)^4$ and set $(1+r)^4 = z$. This gives equation [9]. The notation ‘$z$’ is used because of its universal interpretation as any complex number, and
here we are recognising that the various values of the root
\((1+r)^2 = z\) can be any number, complex as well as real.

\[ z^2 - a_1z^2 - a_2z^2 - a_3z - (a_4 + 1) = 0 \]  

[9]

We know that the equation [9] can be factorised to give
equation [10].

\[ z^2 - a_1z^2 - a_2z^2 - a_3z - (a_4 + 1) = 0 = (z - z_1)(z - z_2)(z - z_3)(z - z_4) \]  

[10]

The value \( z \) is the point \( z \) in diagram D. It lies wherever we choose to set it. The \( z_i \) are the roots. The absolute value of the product on the right hand side of [10] is equal to the product of the absolute values, ie,

\[ |z - z_1|(z - z_2)(z - z_3)(z - z_4)| = |z - z_1||z - z_2||z - z_3||z - z_4| \]  

[11]

The absolute values of differences between complex numbers like \( z_i - z \) is the distance between the two points \( z \) and \( z_i \) in the plane. If we let \( z - z_i = (1 + r) \), ie, they are equal to the first positive real root, then equations [10] and [11] are equal to zero - because the very last element of both is zero. However the other three elements of the right hand side of [10] and [11] are not equal to zero. The product of these three elements is the top element in the left hand side of [6], ie, \( d_1d_2d_3 \). Therefore we have to find a way to evaluate it.

Evaluating the last product is straightforward. The element \( x_i - z \) has to be eliminated from equation [10]. To do so we divide [10] throughout by \( (z - z_i) \) to give [12].

\[ 2z_1 + (x - a_1)z^2 + (a_2 - a_3)z + (a_4 - a_2)z^2 - a_3z - a_4(z - z_i)(z - z_2)(z - z_3)(z - z_4) \]  

[12]

Only after this division do we set \( z_i \) equal to \( z_i \) by replacing \( z_i \) by \( (1 + r) \) in [12] and also replace \( z \) by \( (1 + r) \). Divide the result throughout by \( (1 + r)^3 \) and find the absolute value of both sides. This gives the expression [13].

\[ \frac{4}{(1 + r)^2} \cdot \frac{a_1}{(1 + r)^3} - \frac{a_2}{(1 + r)^2} - \frac{a_3}{(1 + r)^3} - \frac{a_4}{(1 + r)^4} = d_1d_2d_3 \]  

This is the required proof, because the left hand side of [13] is the partition \( P_1 \) of the matrix, which is already known to be the negative of modified duration.

Equation [6] is now known to be true no matter what values the \( a_i \) take. This can be generalised to the \( n \) period case [6a].

\[ \prod_{i=1}^{n} \frac{a_i}{(1 + r)^i} = D \]  

[6a]

where the \( a_i \) are the \( (n-1) \) distances in the complex plane from the chosen, positive, real, \( n^{th} \) root to the other \( (n-1) \) roots.

**The internal rate of return**

We have stated, but not proved, that the value \( r \) - the IRR - a component of the first positive real root, \( (1 + r) \), is a simple kind of average of the stream of numbers, \( a_i \). Simple is not the usual adjective applied to the IRR because it is usually seen through the complicated algebra of the time value of money equation. Solving for \( r \) in equation [1] when \( n = 1 \) or 2 is easy. Solving for \( r \) when \( n = 3 \) or \( n > 4 \) is tedious. When \( n > 4 \) it is impossible to solve analytically. It has to be done using numerical methods - sophisticated guessing techniques; hence the need for a first guess in most spreadsheets and calculators, and the message 'running, running, running' when finding the IRR in the professional analyst's principal tool, the HP12C financial calculator.

However, in the complex plane it is easy to see what it is, even if its calculation remains difficult. To see its simplicity, look at diagram F. The three rods linking the four roots in diagram E have been replaced by four rods linking all four roots to another point in the plane, the point \((1,0)\) on the real axis. This is the same thing as setting \( x = 1 \) in equation [10] (instead of setting it equal to \( x \) as we did in the previous section). The result is equation [14] and diagram F.

\[ [a_1 + a_2 + a_3 + a_4] - [1 - z_1] + [1 - z_2] + [1 - z_3] + [1 - z_4] \]  

[14]

On the left hand side of [14] we have equation [9] with \( z = 1 \). On the right hand side we have it's factorisation with \( z = 1 \). Taking absolute values on both sides means that the sum of the coefficients, the \( a_i \), is equal to the product of the lengths of the rods (the distances, \( n \) in number, from the roots to the point \((1,0)\) in the plane). One of these distances is \( |z_i - z_1| \), therefore it is a short step to solve for \( r \). Equation [15] shows that \( r \) is the sum of the \( a_i \) divided by the \( (n-1) \) roots (the \( w_i \) with \( i = 1 \) to 3 in diagram F). Thus, the IRR is a kind of weighted average of the coefficients in the polynomial, the weights being distances in the complex plane. Like duration, the IRR has a simple interpretation in the complex plane and a new formula. Seen in this light it also can be interpreted as a measure of central tendency.

\[ \sum_{i=1}^{n-1} \frac{a_i}{(n-1)} = r = IRR \]  

[15]

\[ \prod_{i=1}^{n-1} w_i \]

where the \( w_i \) are the \( (n-1) \) distances of the \( (n-1) \) roots (other than the positive real root) from the point \((1,0)\).

**Diagram F**

**A summary of the results**

The results can be summarised for the four period model with variable flow, \( a_i \). The four roots all lie on or around the unit circle. The real root, \((1+r)\), lies close to the point \((1,0)\) because the internal rate of return, \( r \), reflects the average magnitude of the flow - it is a measure of central tendency. The other roots lie off the unit circle if the \( a_i \) are varied, or on the unit circle if they are smooth. The

**CHEER** Volume 14 • Issue 1  Page 7
distances of the roots from the cyclotomic points on the unit circle (the tendrils \( t_1, t_2 \) and \( t_3 \) in diagram C) in some way reflect the dispersion of the \( a_i \) from the constant flow case. The product of the distances of the \((n-1)\) roots from the \(n^{th}\) real root \((d_1, d_2, d_3)\) divided by \((1+r)^4\) gives modified duration.

The observations concerning the IRR and duration have been proved true no matter what the coefficients of the polynomial. Thus, duration for a bond, par or non-par, is merely a special case of a more general model. It is the special case when \( n \) coefficients in the polynomial are the same (par bond), or \((n-1)\) of them are the same and the is \( n^{th} \) different (non-par bond). In the general case, the new equation for duration holds true no matter what values the \( a_i \) take and no matter how many there are.

Viewing the time value of money equation through the lens of its roots in the complex plane, yields novel and elegant views of two important financial concepts.

**References**


**Endnotes**

\(^1\) Mike Osborne is VP and Human Resource Development Manager for Gulf International Bank (GIB) in Bahrain.

The views expressed in this article are the author’s and not necessarily those of the bank. Thanks go to Mark Osborne of St. John’s College, University of Cambridge, for the right comments at the right moments, Eggert Peterson of the University of Bahrain for constructive conversations, colleagues at Gulf International Bank for a congenial working environment as well as helpful comments. Special thanks go to Abdul Bassett Al-Mahmoud of the Bahrain Institute of Banking and Finance for never doubting that the complex plane could be a source of insight into financial matters. Finally, thanks go to an anonymous referee for valuable suggestions about both form and content. The author may be contacted on email at mo@ocean.com.bh or mi.ne.osborne@gib.com.bh.

\(^2\) Any program like Mathcad, Mathematica, Maple and Matlab can be used to manipulate the algebra and graph the results. The calculations used to generate the ideas in this article were made using Mathcad. The charts in the article have been produced with the aid of a drawing program to exaggerate and emphasise the interesting bits.

\(^3\) For readers without the appropriate software to visualise the translation of the coefficients of a polynomial into roots, but who do have access to the internet, visit www.cecm.sfu.ca/organics/papers/odlyzko/support/polyform.html. This site contains an online calculator and graphing utility that allows the viewer to see the roots of a polynomial in the complex plane. The limitation is that the coefficients have to take the values 0 or 1, however a ‘feel’ for the process can be obtained from the site.

\(^4\) It is the negative of modified duration because duration is itself a negative number. The right hand side of equation [6] must be positive to agree with the left hand side. The LHS is always positive because the top line consists of absolute values in the complex plane and the bottom line is obviously positive in most conceivable circumstances. Modified duration is so called because it is a modified version of the original concept developed by Macaulay (1938).
Work 4
Osborne, M.

Three extensions to the visualisation of financial concepts in the complex plane


http://ideas.repec.org/a/che/chepap/v14y2001i2p16-20.html
Three Extensions to the Visualisation of Financial Concepts in the Complex Plane

Michael J. Osborne
Human Resource Development Manager, Gulf International Bank

Introduction

Osborne (2000) contains new expressions for two financial concepts from the bond markets: modified duration and the yield to maturity. The concepts are given in terms of the distances between the roots of the time value of money equation and other salient points in the complex plane. This paper offers three extensions.

First, another new expression is offered, this time a simple and elegant equation for the price of a bond.

Secondly, since its introduction by Macaulay (1938), the concept of duration has suffered from several shortcomings. One of these shortcomings is the fact that the traditional formula for duration, Macaulay or modified, give estimates of the interest elasticity of the bond price that do not allow for the curvature of the link between price and the interest rate. The orthodox formula can be made more accurate by supplementing it with convexity; however, the result is still an approximation. The problem can be found illustrated in any finance text; see, for example, Fabozzi (1996).

It is shown here that the new approach to duration outlined in Osborne (2000) can be adjusted to give a formula that yields precise results for the change in the price of a bond in response to a change in the interest rate. This result, and the methodology surrounding it, is important because it demonstrates that the new perspective from the complex plane not only gives an alternative view of an existing concept, but also improves on it. In addition, it explains why the orthodox formula does not provide precise results, and why it cannot be adjusted to do so.

Thirdly, in order to make the theory operational in the bond markets, a question must be answered. How can the theory be adjusted to cope with bonds priced part way through a coupon period? The required level of detail in time in the bond markets is down to the individual day, and the required level of accuracy down to $1 in a million. In the context of these requirements, the new approach poses questions about computation. The questions are stated here, but not answered.

The heart of the matter

The essence of the approach is that the time value of money equation is a polynomial. A polynomial can be factorised into elements of the form \((z - z_i)\), where \(z\) can be anywhere in the plane we care to place it, and the \(z_i\) are the roots of the polynomial. The mathematics of complex numbers shows that the absolute values of these elements are the distances in the complex plane between \(z\) and the \(z_i\). It follows that the time value of money, and any formula derived from it, can be seen as the product of distances in the plane. As a result, understanding the financial concept becomes less a matter of algebra and more a matter of geometry.

Some preliminary analysis

The bond pricing equation (1) is a particular example of the time value of money equation with four periods. It is shown in the notation of a financial calculator. In a series of steps, the equation can be adjusted through (1a) and (1b) to the more general version, (1c). Note that the redemption or final value, \(F\), is set equal to 1. This assumption does not affect the results in any way; it only serves to simplify the analysis. Note also the care taken to preserve the structure of the constant term - 1 plus the coefficient. Finally, the left-hand side of (1c) can be factorised and absolute values taken, (1d). This last stage is only equal to zero if \(z\) is set equal to one of the roots, or zeros, of the equation.

\[
-pv + \frac{\text{port}}{(1+r)^2} + \frac{\text{port}}{(1+r)^3} + \frac{\text{port}}{(1+r)^4} + \frac{hv}{(1+r)^4} = 0 \tag{1}
\]

\[
-1 + \frac{\text{port} + pv}{(1+r)} + \frac{\text{port} + pv}{(1+r)^2} + \frac{\text{port} + 3 + pv}{(1+r)^3} + \frac{1}{(1+r)^4} = 0 \tag{1a}
\]

\[
-1 + \frac{\text{port} + pv}{(1+r)} + \frac{\text{port} + pv}{(1+r)^2} + \frac{\text{port} + 3 + pv}{(1+r)^3} + \frac{1}{(1+r)^4} = 0 \tag{1b}
\]

\[
-1 + \frac{\text{port} + pv}{(1+r)} + \frac{\text{port} + pv}{(1+r)^2} + \frac{\text{port} + 3 + pv}{(1+r)^3} + \frac{1}{(1+r)^4} = 0 \tag{1c}
\]

\[
-1 + \frac{\text{port} + pv}{(1+r)} + \frac{\text{port} + pv}{(1+r)^2} + \frac{\text{port} + 3 + pv}{(1+r)^3} + \frac{1}{(1+r)^4} = 0 \tag{1d}
\]
\[
\begin{align*}
-1 + \frac{a_1}{(1+r)} + \frac{a_2}{(1+r)^2} + \frac{a_3}{(1+r)^3} + \frac{1}{(1+r)^4} &= 0 \quad (1b) \\
\sum_{i=1}^{n} z_i - a_1 z_i - a_2 z_i^2 - a_3 z_i^3 - (a_4 + 1) z_i^4 &= 0 \text{ where } z_i(1+r) \quad (1c) \\
\sum_{i=1}^{n} |z_i - z_1| &= \prod_{i=1}^{n} |z_i - z_1| = D \quad (1d)
\end{align*}
\]

All of the results found below, and in Osborne (2000), are obtained by allowing \( z \) in equation (1d) to approach various critical values, and manipulating the resulting output.

**An interim result**

If, in equation (1d), \( z \) approaches 1 or, more precisely, the point \((1,0)\) in the plane, then the following useful result holds.

\[
\sum_{i=1}^{n} z_i = \prod_{i=1}^{n} a_i, \text{ where } a_i = |1 - z_i| \quad (2)
\]

The absolute value of the sum of all the coefficients is equal to the product of all the interest rates.\(^5\) The absolute value of 1 minus a root is interpreted as an interest rate because each root is equal to \((1+r)\). In the orthodox analysis, there is only one interest rate because only one root is used. The remaining roots are discarded. In the new analysis, all roots are used explicitly. All four interest rates and roots are depicted distributed around the unit circle in Figure 1, and the distances that represent the interest rates are indicated.

**A new equation for a bond price**

If equations (1a) and (1b) are compared, and the appropriate coefficients are fed into (2), the result can be manipulated into a new bond-pricing equation, (3). The new equation says the bond price is 1 plus the sum of all the coupons divided by 1 plus the product of all the interest rates.

\[
p_n = \frac{\sum_{i=1}^{n} \text{prof } + 1}{\prod_{i=1}^{n} a_i} \quad (3)
\]

This result is an elegant and memorable insight into the structure of a bond price. However, it seems to have no practical application because in order to find all the interest rates you need all the roots, and in order to find all the roots you need the price of the bond.

**Adjusting the equation for modified duration**

In Osborne (2000) an alternative and totally new expression for modified duration was derived: it is equal to the product of the distances between the positive real root and the other \((n-1)\) roots, divided by the positive real root to the \(n\)th power. The formula for the four-period case is equation (4).

\[
\prod_{i=1}^{n} |z_i - z_1| = \prod_{i=1}^{n} \frac{|z_i|}{|z_1|} = -D \quad (4)
\]

where \( |z_1| \) is the distance from the origin to \( z_1 = (1+r_1) \).

The proof of the formula is not repeated here suffice it to say that it is obtained by allowing \( z \) in equation (1d) to converge on the principal root \( z_1 \).

In contrast to the traditional equation, equation (4) can be amended to provide a version of the interest elasticity of the price of a bond that is precise.

The essence of the amendment is that only the change in the one real interest rate, \( r_1 \), is allowed to affect the result. In other words, only one change of root is allowed, from \( z_1 \) to a new position, call it \( z'_1 \). This change affects both the top and bottom lines of (4). The next few paragraphs amplify this point.

The bottom line of equation (4) is changed very simply: replace \( z_1 \) by \( z'_1 \).

The top line consists of the product of all the distances between the \((n-1)\) roots at \( z_2 \), \( z_3 \) and \( z_4 \) and the positive real root at \( z_1 \). When a new interest rate is fed into the bond-pricing equation the effect is not only on price, but also on the positions of all the other roots in the equation. Therefore the distances \( |z_1 - z_i| \) change by movement at both ends. In the new expression for duration, only the one real root \( z_1 \) is allowed to move to \( z'_1 \) and so affect the distances. The other roots are held in place.

In other words, equation (4) must be modified slightly because in its current form both it and the traditional formula use all the roots from the same moment in time. This is incorrect. The correct formulation of the interest elasticity of the bond price uses\( n-1 \) roots from the original equation before the interest rate change, and the new \( n\)th root that reflects the interest rate change. This provides an accurate bridge between the two bond prices.

The resulting formula is equation (5).

\[
\prod_{i=1}^{n} |z_i - z_1| = \prod_{i=1}^{n} \frac{|z_i|}{|z_1|} = -D \quad (5)
\]

where the \( z_2 \) are the \((n-1)\) roots from the starting equation and \( z'_1 \) is the new principal root, i.e. 1 plus the new interest rate. See the appendix for a proof.

The new expression for duration yields an insight into why the traditional formula, obtained by differentiation of the bond price, was not accurate. How many of us have puzzled over which interest rate to enter into the traditional formula for modified duration – the old one, the
new one, some kind of average of the two? Clearly, differentiation of the time value of money equation with respect to the interest rate is not a simple procedure because there are multiple interest rates. Which one is being changed? It gives an interesting new slant on the famous phrase used by economists – *ceteris paribus*.

The new approach also answers the question about the wisdom of throwing away all the roots apart from the single real root that is traditionally calculated. All the interest rates can be used, wrapped up in their roots.

Finally, the new approach demonstrates why the orthodox formula is incapable of yielding precise results. The reason for the deficiency usually supplied in the textbooks is that duration is a straight-line approximation of a non-linear function, which is true. However, the reason supplied here is deeper than that; it is that the symbol for the interest rate contained in the orthodox equation represents simultaneously all possible interest rates. A true, *ceteris paribus* change cannot be made in that context. Such a change requires a formula that shows all rates explicitly.5

### Part coupon periods and day counts

The next issue is best illustrated with a numerical example. Consider a US Treasury bond with a coupon of 6% and exactly two years to run to maturity. The market rate of interest today, 3 April 2000, is 10%. The bond matures on 3 April 2002. The market convention for US T-bonds is that they pay coupons semi-annually on an actual/actual basis. If the face value is 1, then the current price is 0.929080989917. The situation is depicted in equation (6) and Figure 2.

\[
0.929080989917 = \frac{0.03}{(1+0.05)^2} + \frac{0.03}{(1+0.05)^1} + \frac{1+0.03}{(1+0.05)^0}
\]  

(6)

Now assume that time passes. The date is now 3 August 2000, which is exactly two thirds of the way through the first coupon period. The new price is 0.959797851371. The situation is no longer one of four periods of even length, i.e. n is no longer an integer. The number of periods is now 3 1/3. It is not possible to find the roots of such a polynomial.

The official method for pricing a US T-bond when a part-coupon period exists is shown in equation (7). Steiner (1998) and Fabozzi (1996) give details. The structure of the equation is that future payments are discounted back to the date of the next coupon payment 3.10.2000 and then the value on that coupon date is discounted on a fractional basis back to the present 3.8.2000. The fraction is the number of days between the valuation date and the next coupon date divided by the number of days in a whole coupon period (y).

\[
-p_r\left[ \frac{pm_1}{(1+r)^3} + \frac{pm_1}{(1+r)^2} + \frac{pm_1}{(1+r)^1} \right] \frac{1}{(1+r)^3} = 0
\]  

(7)

where \( r \) is the market interest rate per period.

One way forward is to crack the whole period into ten periods of two months each, the current date being time 0 and the redemption date being time 10. Coupon payments occur at times 1, 4, 7 and 10. The new approach is clearly able to handle the uneven series of payments because the \( a_k \) in the original polynomials, (1b) to (1d), can take any values, including zero.6

The result of cracking the cash flow into smaller time periods is a new polynomial of greater order than before:

\[
-p_r\left[ \frac{pm_1}{(1+r)^7} + \frac{pm_1}{(1+r)^5} + \frac{pm_1}{(1+r)^1} \right] = 0
\]  

(7a)

where the interest rate, \( r^* \), is found from \((1+r^*) = (1+r)^{0.7} = (1+0.05)^{0.7}\).

Finding the roots of equation (7a) is easily done using one of the standard mathematical programs like Mathcad, Mathematica, Maple or Matlab. A precise value for duration can be calculated using only a few lines of instructions to establish the distances between the roots. In this example, if the market interest rate is assumed to increase by 1% on an annual basis on 3 August 2000, then the periodic (two-month) rate increases from 1.639853568149% to 1.800713028126% and the value of duration, \( \frac{\Delta r}{\Delta r_{\text{price}}} \), is 9.23171323986. Some straightforward arithmetic using these figures gives the change in the price of the bond as 1.8701987561%. The actual value using the original bond pricing formula is 1.87002006541%, a difference of 0.0000001298, or 13 cents in $1 million. One rule of thumb in the industry is that calculations like this should yield results accurate to one hundredth of one basis point, where a basis point is one hundredth of 1%. Thus, the required degree of accuracy is $1 in $1 million. The answer in the example passes this test.7

However, the new approach does face a computational problem. The example works out beautifully because it was carefully chosen to illustrate an application of the theory using a polynomial of low degree. But consider the following extreme situation. We have a 30-year US Treasury bond and it is one day after issue. Time will need to be cracked into periods as short as one day, and the interest rate to become as small as the daily rate. The resulting polynomial is of the order of 10,950, and its coefficients are sparse. What are the limits to a computation like this? What order of polynomial can be factorised in a reasonable period of time on a good machine? And what degree of accuracy results? How far is it possible to go into the realm of practical bond calculations, given the required degree of accuracy?
In addition, the ‘period cracking’ route using high-order polynomials proves to be impracticable, do any other routes exist to utilise the theoretical insight provided by the new approach? For example, if the periods are cracked into the maximum feasible number (to a day close to the actual day rather than on it) and this maximum falls short of the optimum needed for precise results, is the resulting approximation, nevertheless, a better approximation than that yielded by the orthodox calculation?

Conclusion
The new approach to duration is a particular example of the insight that working in the complex plane gives to any financial concept that incorporates the time value of money. The addition of the imaginary dimension is useful. It shows why the orthodox approach is inaccurate, and what can be done to correct it. In particular, it shows why differentiating the time value of money with respect to the interest rate is not straightforward. This last observation is generalisation does not apply to the derivation of duration alone. Any attempt to identify a sensitivity to an interest rate change in the context of the time value of money may be subject to the problem.

However, the value of the theoretical insight is limited if it cannot be made operational. Nearly all bond traders and portfolio managers in the world have calculators built into the programs on their computers, and algorithms for duration are standard components of these calculators. The orthodox equations for duration yield only approximate results. The new approach outlined above is attractive because it suggests that precise results are possible, but only if the limits to computation have reasonable bounds. The bounds have yet to be identified. Readers with more skill than the author in these matters may be able to offer guidance or solutions.

Appendix:
The proof of equation (5) – the new expression for modified duration
Consider equation (1) from the text with \( PV \) set equal to 1. This is the equation that shows the price of a bond using the notation of a financial calculator simplified to the four-period model used throughout the paper.

\[
\frac{1-\text{PV}}{(1+r)^t} + \frac{\text{PMT}}{(1+r)^t} - \frac{\text{PV}}{(1+r)^t} = 0
\]

Equation (1)

If the interest rate changes from \( r \) to \( r' \) to \( PV \) then becomes \( PV' \). Equation (A1) is the result.

\[
-\frac{1-\text{PV'}}{(1+r')^t} + \frac{\text{PMT}}{(1+r')^t} - \frac{\text{PV'}}{(1+r')^t} = 0
\]

(A1)

Convert both equations (1) and (A1) to the form of the polynomial having the particular structure described in the text - the four-period version of (1a). Call the two versions (A2) and (A3).

\[
-1 + \frac{\text{PV} \cdot PV'}{(1+r)^t} + \frac{\text{PMT} \cdot PV'}{(1+r)^t} + \frac{\text{PV} \cdot PV'}{(1+r)^t} + \frac{1}{(1+r)^t} = 0
\]

(A2)

Keep these equations in mind. Now jump to equation (1d) from the text, the general version of the polynomial with \( n = 4 \) and \( z = (1+r) \). Note that the \( a_i \) on the left-hand side of (1d) correspond to the middle four parameters on the top line of (A2).

\[
[(a_1 + 1)z - z_i]x_1 - [a_2 + 1]z - z_i[1 - z_i]
\]

(1d)

In the text of the article it is pointed out that \( z \) can take any value. If it happens to take the value of any one of the roots, e.g., \( z_1 \), then both sides of the equation are equal to zero because the first element on the right-hand side is zero. In this particular case, let \( z \) converge on the new root containing the new interest rate, \( (1+r') = z_1' \). Substitute the left- or right-hand side of this last expression, as appropriate, into equation (1d), divide throughout by the same elements raised to the fourth power, and take absolute values on both sides. This gives (A4).

\[
\frac{1 + a_1}{(1+r')} + \frac{a_2}{(1+r')^2} + \frac{a_3}{(1+r')^3} + \frac{a_4}{(1+r')^4} + \frac{1}{(1+r')^5} = \frac{1}{|z_1'|}
\]

(A4)

Notice two things about (A4). First, on the right-hand side, the first of the four elements in the product on the top line is the change in the interest rate from \((1+r)\) to \((1+r')\), i.e., \( |z_1' - z_1| = \Delta r \). If this first element is taken outside, then the remainder is the new expression for duration, \( -D \). Therefore, the right-hand side of (A4) is \( -D \Delta r \). (Note that \( D \) itself is intrinsically negative, therefore \( -D \) is positive.)

The second thing to notice about (A4) is the structure of the left-hand side. It is the same structure as the left-hand side of (A2) or (A3). Yet it cannot be equal to zero. This is because the \( a_i \) are from equation (A2), i.e. they contain the original price, \( PV \), while the bottom line contains the new interest rate \( r' \) from equation (A3) rather than \( r \). Thus, it is a mixture of the two equations. It is the old parameters being evaluated at the new interest rate. Whatever this value is, it is not zero, therefore we set it equal to \( X \).

\[
-1 + \frac{a_1}{(1+r')} + \frac{a_2}{(1+r')^2} + \frac{a_3}{(1+r')^3} + \frac{a_4}{(1+r')^4} + \frac{1}{(1+r')^5} = X
\]

Or

\[
1 - \frac{\text{PV} \cdot PV'}{(1+r)^t} - \frac{\text{PMT} \cdot PV'}{(1+r)^t} + \frac{\text{PV} \cdot PV'}{(1+r)^t} + \frac{1}{(1+r)^t} = X
\]

Subtract X from both sides and divide throughout by \((1+X)\).

\[
1 + \frac{\text{PV} \cdot \text{PV}(1+X)}{(1+r)^t} + \frac{\text{PMT} \cdot \text{PV}(1+X)}{(1+r)^t} + \frac{\text{PV} \cdot \text{PV}(1+X)}{(1+r)^t} + \frac{1}{(1+r)^t} = 0
\]

Compare the coefficients in this last result with those of equation (A3). It is clear that \( PV' = PV(1+X) \). This means that \( X \), the left-hand side of (A4), is the change in the price of the bond relative to its original price, i.e.

\[
X = \frac{PV' - PV}{PV} = \frac{60P}{PV}
\]
Put together the results from both sides of (A-4) to give equation (5) in the text.

\[ \frac{\Delta p}{pv} = -\Delta t \text{ or } \frac{\Delta p}{pv} + \frac{1}{\Delta t} = \sum_{i=1}^{n} \frac{1}{|c_i|} (t_i - 1) \]

Notes

1. Mike Osborne is VP and Human Resource Development Manager for Gulf International Bank (GIB) in Bahrain. The views expressed in this article are the author's and not necessarily those of the bank. Thanks for advice go to Eggert Peterson of the University of Bahrain and colleagues at Gulf International Bank, especially Nabil Abdul Aal. The author may be contacted via email at mosborne@boredco.com.bh or mike.osborne@gibah.com.

2. The word 'precise' is used rather than 'exact'. This is because the new formula uses the roots of a polynomial as inputs. Except in special cases, there are no closed-form solutions for polynomials of third order greater than four; numerical methods have to be used. Therefore, despite the fact that the formula has the appearance of a closed-form solution, it relies on numerical methods. However, the root-finding methods employed in most mathematical programs can give high levels of precision, usually greater than that needed by most financial problems, though the latter part of this paper does raise a question about the issue.

3. In Osborne (2000) the notation for these distances is \( w_i \); in this article the notation used is \( r_i \) in order to indicate that they are 'interest rates'.

4. The entry for the 'root dragging theorem' in the Encyclopedia of Mathematics (Weinstein 1998) reads: 'If any of the roots of a polynomial are increased then all of the critical points increase.'

5. It follows that any financial equation incorporating the time value of money that is differentiated to obtain the interest sensitivity is subject to the same problem. One obvious example is the interest sensitivity of the value of a portfolio of fixed rate loans in the balance sheet of a financial institution.

6. The method can be extended to find the duration of a portfolio of bonds. The orthodox approach is to calculate the duration for each bond and then find the weighted average of the durations. The new method means that the cash flows of bonds in a portfolio can be summed vertically to produce a composite bond with an uneven cash flow and irregular time periods. The lowest common denominator of time can be imposed to produce a new, higher-order polynomial with varied coefficients and integer powers. One single calculation for duration can then be made.

7. The actual bond prices, and the actual changes, were calculated to the maximum possible 13 places using the bond-pricing functions built into the HP10B financial calculator. The calculations of duration from the distances in the complex plane, and the predicted bond price change, were made using Mathcad with the precision set to 11 decimal places.

References


Work 5
Osborne, M. & Osborne, M.
A polynomial that is a statistical prism

NOTICE: this is the author’s version of a work that was accepted for publication in *The Mathematical Gazette*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *The Mathematical Gazette*, VOL 87, No. 508, (Mar., 2003), pp. 76-85

http://www.jstor.org/stable/3620567
A polynomial that is a statistical prism
The Mathematical Gazette, Vol. 87, March 2003
Mark J. Osborne and Michael J. Osborne

Introduction
The roots of a polynomial can be represented as points in the complex plane. The time value of money (TVM) equation that is commonly used in finance is a polynomial. (See the appendix for a short description of the TVM equation and an example of its use in finance.) In [1] and [2] it is shown that concepts from financial mathematics can be obtained from the pattern of the roots of the TVM equation. The concepts are given in terms of distances between the roots and other salient points in the plane. This note shows that this particular polynomial, and the technique, can be applied more generally. When a series of data is fed into the coefficients of the polynomial, the mean and standard deviation of the data are seen in the complex plane as combinations of distances between the roots and other salient points. The results are aesthetically pleasing as well as mathematically interesting.

The basic equation - a particular polynomial
Eq. (1) is a particular polynomial where the coefficients are a series of numbers, \( a_i \), that can take any real values. Note the -1 at the front end, and the 'discounted' +1 at the far end, of (1). This structure, with its financial flavour, is an important part of the analysis.

\[
-1 + \frac{a_1}{(1 + R)} + \frac{a_2}{(1 + R)^2} + \ldots + \frac{a_{n-1}}{(1 + R)^{n-1}} + \frac{a_n}{(1 + R)^n} + \frac{1}{(1 + R)^n} = 0
\]  

(1)

Eq. (2) is the same thing recast in the more general form likely to be found in any mathematical publication. The element \((1 + R)\) is set equal to \(z\) rather than \(x\) to indicate that the variable can take on complex values as well as real.

\[
z^n - a_1z^{n-1} - \ldots - a_{n-1}z - a_n = 1 \quad \text{where} \quad z = (1 + R)
\]  

(2)

The fact that (1) is a version of the TVM equation makes it tempting for anyone steeped in finance to interpret the \(a_i\) as money payments and \(R\) as a yield or rate of interest. This is a deliberate interpretation in the two papers cited earlier. In this article the analysis is perfectly general to the extent that each of the \(a_i\) can take any real value. They are treated as an ordered series of numbers.

Interpreting the roots in the complex plane.
Eq. (1) is an \(n^{th}\) degree polynomial, therefore it has \(n\) roots. The \(n\) roots of a polynomial with real coefficients are either real or occur in complex, conjugate pairs (it is assumed that none of the coefficients are ever complex). Because the roots can be complex their interpretation is not obvious.
From this point onwards, a fourth order polynomial is used to illustrate the results, as it is particularly easy to visualize results when there are only four roots. The generalization to the $n^{th}$ order with $n$ roots is not always stated or proved, though in most cases it is intuitively clear.

The simplest case is where all the $a_i$ are zero. Then Eq. (2), with $n=4$, reduces to

$$z^4 = 1$$

This equation is a cyclotomic equation therefore its four roots are evenly dispersed around the unit circle in the complex plane at points $z_1, z_2, z_3$ and $z_4$ in Fig. 1.

The result generalizes to the $n^{th}$ order case, $z^n = 1$, where there are $n$ roots evenly dispersed around the unit circle with equal angles ($2\pi / n$ radians or $360/n$ degrees) between them.

Next, allow the $a_i$ in Eq. (2) to be non-zero but assume that they all take the same real value $a$. In the fourth order case, three of the roots are now in exactly the same position as for the cyclotomic equation, while the fourth, the one positive real root, is now $(1 + a)$. This result is evident if $z^4 - az^3 - az^2 - az - a - 1$ is divided by the factor $(z - 1 - a)$ giving the cyclotomic polynomial $z^3 + z^2 + z + 1$, the roots of which are $-1, i$ and $-i$. The result of adding in a stream of numbers of constant size $a$ is simply to move the one positive, real root the distance $a$ along the real axis. The move is to the right if $a$ is positive and to the left of it is negative. A positive value for $a$ is depicted in Fig. 1 as the move from $z_1$ to $z_1^*$.
This result also generalizes very easily to the $n^{th}$ order case.

If the $a_i$ are allowed to vary, the other $(n-1)$ roots depart from their positions on the unit circle. They become points on the ends of tendrils attached to the cyclotomic roots. In Fig. 1 they are shown as $z_2^*$, $z_3^*$ and $z_4^*$. The mathematical literature, see [3], suggests that the roots of a polynomial with random coefficients are scattered around the neighborhood of the unit circle.\footnote{Fig. 1 lays a foundation for the basic ideas contained in the remainder of this paper. The roots of the polynomial are clustered around the unit circle. The cyclotomic roots are equally spaced around the unit circle at $z_1, z_2, z_3$ and $z_4$. The position of one root, $z_1^*$, relative to its cyclotomic counterpart, in some way reflects the mid-point of the data, while the positions of the other roots, $z_2^*, z_3^*$ and $z_4^*$, relative to their cyclotomic counterparts, in some way reflect the variability of the data. These vague statements now require sharpening.

The crucial idea at the heart of the analysis is that standard statistical concepts can be seen as combinations of distances between roots in the complex plane. The following results are based on variations of Cotes’ Theorem. See [4] or [5] for an account of it.

The mean

Take Eq. (2) again and once more set $n=4$. This polynomial can be factorized to give (3).

$$z^4 - a_1 z^3 - a_2 z^2 - a_3 z - (a_4 + 1) = (z-z_1^*).(z-z_2^*).(z-z_3^*).(z-z_4^*)$$ (3)

The $z_i^*$ are the roots of the polynomial. In general, they lie off the unit circle if the $a_i$ are not zero and not all the same.

Take the modulus of Eq. (3) to give (4).

$$|z^4 - a_1 z^3 - a_2 z^2 - a_3 z - (a_4 + 1)| = |z - z_1^*||z - z_2^*||z - z_3^*||z - z_4^*|$$ (4)

The absolute value of the difference between two complex numbers, i.e., $|z - z_i^*|$, is the distance between the two points $z$ and $z_i^*$ in the plane.

Now assume all the $a_i$ in (4) are zero. This is the case where all the roots lie on the circle at the positions given by $z_1, z_2, z_3, z_4$ (the roots minus their asterisks). In addition, $z = z_1$ is substituted into the equation. The result is Eq. (5), which is equal to zero because the first element on the right hand side becomes zero. Setting $z$ equal to any of its roots would have this effect.
\[ z_1^4 - 1 = |z_1 - z_1| |z_1 - z_2| |z_1 - z_3| |z_1 - z_4| = 0 \quad (5) \]

However, the other three elements on the right hand side of (5) are not equal to zero. They represent the distances between three of the cyclotomic roots and the fourth cyclotomic root across the unit circle. The right hand side of (5) contains the product of these distances spanning the unit circle. What is the value of this product?

The following equation is obtained by setting the \( a_i \) equal to zero in Eq. (3).

\[ z^4 - 1 = (z - z_1)(z - z_2)(z - z_3)(z - z_4) \]

Eliminate the element \((z - z_1)\) by dividing the equation throughout by \((z - z_1)\).

\[ z^3 + z_1z^2 + z_1^2z + z_1^3 = (z - z_2)(z - z_3)(z - z_4) \]

Since \( z_1 \) is the positive, real, cyclotomic root of a polynomial of fourth degree it is equal to 1. Let \( z = z_1 = 1 \). Finally, take absolute values on both sides. The result is (6) or, in more general terms, (6a). The fourth order case is illustrated in Fig. 2.

\[ 4 = |z_1 - z_2| |z_1 - z_3| |z_1 - z_4| = b_2b_3b_4 = \prod_{i=2}^{l=4} b_l \quad (6) \]

\[ n = \prod_{i=2}^{l=n} |z_1 - z_i| \quad (6a) \]

The product of the \((n-1)\) distances across the unit circle between one of the cyclotomic roots and the other cyclotomic roots is \( n \), the order of the polynomial. In Eq. (6) and Fig. 2 the \((n-1)\) distances are labeled the \( b_i \) from \( i = 2 \) to \( n \).

![Figure 2](image-url)
The equation for the arithmetic mean of the \( a_i \) is \( a = \frac{a_1 + a_2 + a_3 + a_4}{4} \) and this can now be rewritten:

\[
-\frac{a}{a} = \frac{a_1 + a_2 + a_3 + a_4}{b_2 \cdot b_3 \cdot b_4} \quad \text{or, in general,} \quad -\frac{a}{a} = \sum_{i=2}^{i=n} \frac{a_i}{b_i}
\]  

(7)

This appears to be a trivial adjustment to the denominator in the formula for the arithmetic mean. However, the analysis is now extended to the numerator to show a more substantive result.

Return to Eq. (4) and set \( z = 1 \). Leave the asterisks on the roots to indicate that it is the general case, that the \( a_i \) are varied and the roots can be anywhere in the plane. The result is (8).

\[
|a_1 + a_2 + a_3 + a_4| = |-z_1^*||-z_2^*||-z_3^*||-z_4^*|
\]  

(8)

On the left hand side is the absolute value of the sum of the coefficients from the original polynomial. On the right hand side are the distances between the roots and the point (1,0) in the plane (the \( r_i \) in Fig. 3). Since each root is one of the possible values that \( 1 + R \) in Eq. (1) can take, it follows that the absolute value of a root minus 1 is an absolute value of \( R \). Therefore the right hand side of (8) is the product of all the \( r_i \) (= \( |R_i| \)). The general form of Eq. (8) is (9). This is an unexpected and satisfying result.

\[
\left| \sum_{i=2}^{i=n} a_i \right| - \prod_{i=1}^{i=n} r_i = \prod_{i=1}^{i=n} |R_i|
\]  

(9)

It follows from (7) and (9) that the mean can be expressed in a totally new way - Eq. (10). Figs. 2 and 3 illustrate the components of (10).

\[
-\frac{a}{a} = \frac{\prod_{i=1}^{i=n} r_i}{\prod_{i=2}^{i=n} b_i}
\]  

(10)

We are also interested in measures of the dispersion of the \( a_i \), to which we now turn.
Expressions for the standard deviation or variance

Return to Eq. (4). Instead of letting $z$ converge on 1 (the first of the four cyclotomic roots around the unit circle), let $z$ converge on the three other cyclotomic roots ($i, -1, -i$). This produces three more equations: (11), (12) and (13). Set these equations alongside Eq. (8), making four equations in all. Each equation represents the product of the distances between each root in the plane and a particular cyclotomic root, as depicted in Figs. 3, 4, 5 and 6. The figures are kept separate to avoid too much clutter in a single figure. Fig. 3 is repeated for completeness.

\[ |a_1 + a_2 + a_3 + a_4| = |1 - z_2^*| |1 - z_3^*| |1 - z_4^*| = r_1 r_2 r_3 r_4 \]  
(8)

\[ |a_1 i + a_2 - a_3 i - a_4| = |i - z_1^*| |i - z_2^*| |i - z_3^*| |i - z_4^*| = s_1 s_2 s_3 s_4 \]  
(11)

\[ |a_1 - a_2 + a_3 - a_4| = |-1 - z_1^*| |-1 - z_2^*| |-1 - z_3^*| |-1 - z_4^*| = t_1 t_2 t_3 t_4 \]  
(12)

\[ |-a_1 i + a_2 + a_3 i - a_4| = |i - z_1^*| |i - z_2^*| |i - z_3^*| |i - z_4^*| = u_1 u_2 u_3 u_4 \]  
(13)

The next step is to square each of these equations and add them together. We obtain Eq. (14) or, in more general terms, (14a).

\[ 4\left(a_1^2 + a_2^2 + a_3^2 + a_4^2\right) = (v_1 r_2 r_3 r_4)^2 + (s_1 s_2 s_3 s_4)^2 + (t_1 t_2 t_3 t_4)^2 + (u_1 u_2 u_3 u_4)^2 \]  
(14)

\[ n \sum_{i=1}^{\infty} a_i^2 = \left(\sum_{i=1}^{\infty} \left|z_j - z_i^*\right|^2\right)^2 \]  
(14a)

Put this result to one side for a moment and consider the following equation for the population variance of the series, $\sigma_n^2$. It is written in two different ways. The version on the right side is of most interest.
It can be rearranged to give an expression for the element on the left-hand side of (14a).

\[ n \sum_{i=1}^{n} a_i^2 = n^2 \sigma_a^2 + (\sum_{i=1}^{n} a_i)^2 \]

This, combined with (14a) implies (15).

\[ n^2 \sigma_a^2 + (\sum_{i=1}^{n} a_i)^2 = (r_1, r_2, r_3, r_4)^2 + (s_1, s_2, s_3, s_4)^2 + (t_1, t_2, t_3, t_4)^2 + (u_1, u_2, u_3, u_4)^2 \quad (15) \]

It is known from Eq. (9) that the modulus of the sum of the \( a_i \) is equal to the product of the \( r_j \).

Therefore, given that it is a squared term, the second element on the left-hand side of (15) is equal to the first element on the right-hand side. They cancel. Rearranging the remainder, and inserting what is known about the value of \( n \), gives an elegant expression for the standard deviation (Eq. (16)) or the variance (Eq. (16a)).
Some comments on the view from the complex plane

There are many things to be said about this unfamiliar view of a familiar object provided by (16).

This new expression for the standard deviation clarifies the intuition, discussed earlier in the paper, that the positions of the \((n-1)\) roots relative to their cyclotomic equivalents in some way measure dispersion.

The standard deviation, like the mean, is invariant to the order of a particular set of \(a_i\). The denominator of (16) does not depend at all on the order of the \(a_i\). It follows that the complicated expression in the numerator of (16) must be invariant. As the \(a_i\) change order so movements in the roots must cause precise compensations in the components of the expression.

The polynomial (1) acts as a prism for any set of data. Put the data through the prism and out come simple and elegant relationships between the roots that characterize the summary statistics for the data. The unit circle provides the base positions for the roots. The manner in which the roots depart from their base positions on the unit circle is determined by the different structures of the \(a_i\). The relative positions of the roots of the polynomial allow a neat partition of the data into the mean and standard deviation. This is shown most clearly by the following variant of Eq. (15).

\[
\frac{-1}{a} + \sigma_a^2 = \frac{(r_1 r_2 r_3 r_4)^2}{(b_2 b_3 b_4)^2} + \frac{\left( s_1 s_2 s_3 s_4 \right)^2}{(b_2 b_3 b_4)^2} + \frac{\left( t_1 t_2 t_3 t_4 \right)^2}{(b_2 b_3 b_4)^2} + \frac{\left( u_1 u_2 u_3 u_4 \right)^2}{(b_2 b_3 b_4)^2}
\]

Another measure of dispersion is the normalized standard deviation, the ratio of the standard deviation to the mean. Division of Eq. (16) by (10) sees the \(b_j\) cancel, leaving an expression for normalized standard deviation that also looks rather different from the norm.

\[
\frac{\sigma_a}{a} = \sqrt{\frac{\left( s_1 s_2 s_3 s_4 \right)^2}{(r_1 r_2 r_3 r_4)^2} + \left( t_1 t_2 t_3 t_4 \right)^2 + \left( u_1 u_2 u_3 u_4 \right)^2}
\]

This equation, when squared and inverted, gives a yet more elegant expression, Eq. (17). A study of Eq. (17) alongside Figs. 3, 4, 5 and 6 reveals an unexpectedly beautiful structure.
\[
\left( \frac{\bar{a}}{\sigma_a} \right)^2 = \frac{\left( r_1, r_2, r_3, r_4 \right)^2}{(s_1, s_2, s_3, s_4)^2 + (t_1, t_2, t_3, t_4)^2 + (u_1, u_2, u_3, u_4)^2} \tag{17}
\]

In statistics, when the square root of (17) is adjusted by \(\sqrt{n}\), it is the statistic for the test that the mean of the set of data is significantly different from zero.

**Concluding thoughts**

Two further thoughts occurred to the authors as this note was prepared.

- What other statistical objects can be seen in the plane by means of the particular polynomial? The covariance and correlation coefficient between two sets of data spring to mind. Clearly two sets of roots from two polynomials would be involved.

- How does this analysis link into the mathematical and statistical literature, particularly that surrounding the normal distribution? Searches by the authors have so far yielded no analysis of this specific polynomial outside of the financial.

In the meantime it can be said that the complex plane is able to give an uncommon view of some commonplace statistical objects.
The time value of money equation

The TVM equation is the ‘workhorse’ of finance. It is the means by which most financial products are priced and analysed (the most notable exception being the financial option, the product used by Mr. Leeson to sink Barings). Any financial calculator contains the equation pre-programmed. For example, the Hewlett-Packard 12C is the calculator much favoured by investment analysts and traders. It has five buttons labeled $n$, $i$, $PV$, $PMT$ and $FV$. Sadly, there is no generally agreed notation in finance. It varies from calculator to calculator, publication to publication and, in this article, from text to appendix. In the HP12C the labels stand for the number of time periods, the interest rate per time period, the present value of an investment, the periodic payment of an investment and the future value of an investment respectively. They are linked through the following equation:

$$PV = \frac{PMT}{(1 + i)} + \frac{PMT}{(1 + i)^2} + \ldots + \frac{PMT}{(1 + i)^{n-1}} + \frac{PMT}{(1 + i)^n} + \frac{FV}{(1 + i)^n}$$

A simple example of its use is the pricing of a consumer loan. Assume the amount borrowed today (the present value) is $10,000. This figure is entered with a plus sign because the money is being received. The loan will be paid back over 48 months (the number of time periods), the future value will be zero because the loan will be fully repaid at the end of the contract, and the monthly payment is -$250, entered with a minus sign because the money is being paid away. Enter the wrong signs and the calculator gives an error message for obvious reasons. Pressing the button for the rate of interest per period will show on the screen a figure of 0.7701472488 which means $i = 0.007701472488$. This is the interest rate per month. Rates are usually quoted on an annual basis so the rate is annualised to what is often referred to as the Annual Percentage Rate (APR). In the USA this is done by multiplying by twelve to obtain 9.2%. In the UK the legally correct procedure is to compound the monthly rate using the formula $APR = \left((1 + i)^{12} - 1\right) \times 100$. In the example the result is 9.6%. It is usual in consumer finance to round to one decimal place. In wholesale banking the demand is for calculations accurate to within $1 in $1 million, therefore more decimal places are needed inside the calculation, and rates are usually quoted to 4 decimal places. Some calculators give the APR automatically. If they do then the user has to be careful about which annualisation procedure is in the machine.

There are several excellent books on financial calculations, covering both wholesale and retail products, of which perhaps the most comprehensive is [6].

________________________________________________________________________

Appendix
Acknowledgement
The authors are indebted to the referee for many useful suggestions about both style and content that have significantly improved the paper.

References

Mark Osborne is doing postgraduate research at the Cavendish Laboratory and St. John's College, University of Cambridge.
email: mjo24@cam.ac.uk.
Mike Osborne works for Gulf International Bank (GIB) in Bahrain.
email: mosborne@batelco.com.bh or mike.osborne@gibbah.com.

\footnote{Visit www.cecm.sfu.ca/organics/papers/odlyzko/support/polyform.html. This site contains an online calculator and graphing utility that allows the viewer to see the roots of a polynomial in the complex plane. The limitations are that the polynomial is not quite the same as the TVM equation, and the coefficients have to take the values 0, 1 or -1; however, a 'feel' for the process can be obtained from the site.}
Work 6
Osborne, M.

A simple, accurate formula for the duration of a portfolio of bonds under a non-parallel shift of a non-flat yield curve


www.econ.ku.dk/Fru/Conference, International Conference on Finance

Finance Research Unit, University of Copenhagen, 2-5 September, 2005
A SIMPLE, ACCURATE FORMULA FOR THE DURATION OF A PORTFOLIO OF BONDS UNDER A NON-PARALLEL SHIFT OF A NON-FLAT YIELD CURVE

Michael J. Osborne *, Gulf International Bank **

First draft, September 2004; second draft, January 2005; third draft April 2005.

Abstract
It is well known that the various formulas for the duration of a vanilla bond give inaccurate results. Their accuracy can be improved by the addition of extra elements, such as convexity or duration vectors. But the results remain inaccurate. A recent paper proposed a new formula for the duration of a portfolio of vanilla bonds. The formula gives a precise, accurate value for any parallel shift in a flat yield curve, without the need for auxiliary concepts. The analysis is performed in the complex plane, and uses all possible interest rates that solve the time value of money equation. In this paper, the analysis is reworked to produce a second, ‘complex’ formula that is more general. It copes with any non-flat yield curve and any non-parallel shift in the curve, and it is simpler and easier to prove. Some insights and puzzles presented by the new analysis are discussed.

Key words: Duration, convexity, fixed income, bond, complex plane

JEL classification codes: C60, G10, G11, G12

* The author wishes to thank Hani Redha and Jaffar Hussain of National Commercial Bank, and Duncan Larraz of UBS, for valuable conversations on the subject; Peter Carr, Head of Quantitative Financial Research at Bloomberg, for his thoughts and permission to quote from some correspondence; and Jay Marchand of Mercer University for his comments.

** Gulf International Bank, Al-Dowali Building, 3 Palace Avenue, P.O. Box 1017, Manama, Kingdom of Bahrain. Telephone 00 973 17 522589, fax 00 973 17 522643, email mike.osborne@gibbah.com.

This paper is deposited in the Social Science Research Network working paper archive. http://ssrn.com/abstract=587242
A simple, accurate formula for the duration of a portfolio of bonds under a non-parallel shift of a non-flat yield curve

1 Introduction

The fixed income bond market is one of the largest financial markets in the world. The interest rate swap market, intimately linked with the bond market, is one of the largest derivatives markets. It follows that the price sensitivity of a bond or swap to a change in the rate of interest is of great importance to market participants.

One of the earliest writers to analyse this sensitivity was Macauley (1938). He noticed that the weighted, average time to maturity of a bond’s cash flows provided an approximation to the elasticity of its price with respect to its yield. This connection, between the elasticity and the timing of the cash flows, is the source of the term ‘duration’.

Later, Hicks (1939) and Samuelson (1945) used calculus to derive ‘modified duration’. This measure is more accurate, and has a more logical derivation, than Macauley duration.

In the 1970s, a more general interest in the concept began. The onset of floating exchange rates generated volatility in financial markets and a greater awareness of financial risk. As a result, this period saw an increased application of the measure, a greater awareness of its shortcomings and much research to overcome them. By the 1980s, duration had a compilation of contributions (Hawawini, 1982), a textbook (Bierwag, 1987) and several surveys (Bierwag, 1983 and 1988).

The shortcomings of duration include the following:

1. It is a linear approximation to the curvilinear relationship between the price of a bond and its yield. This makes it an inaccurate measure of elasticity, and the larger the change in yield, the greater the inaccuracy.
2. It assumes a flat yield curve.
3. It assumes parallel shifts in the yield curve.
4. If the bond is callable or puttable (an option is embedded in it) then the measure fails.
The issue of inaccuracy was tackled in a variety of ways. It was partly answered by the introduction of convexity. Duration uses the first term in a Taylor series expansion of the equation for the bond price. Convexity uses the second. When put together, accuracy is increased. If the other terms of the expansion are not included, however, full accuracy is still not obtained. The Taylor series approach involves a trade-off between simplicity and accuracy. D’Antonio and Cook (2004) contains a recent analysis of convexity. Livingston and Zhou (forthcoming) follow another route to greater accuracy, ‘exponential duration’, which uses the logarithms of variables.

Various approaches to incorporating the term structure of interest rates were initiated by Fisher and Weil (1971). Chambers et al (1988) pursued the same objective with their duration vectors. Ho (1992) introduced the concept of key-rate duration, which focuses on the price elasticity with respect to various key points along the yield curve. Yet another way of coping with shifts in the yield curve is the M-vector approach, originally proposed by Fong and Vasicek (1984) and Fong and Fabozzi (1985), and further developed more recently by Nawalkha and Chambers (1997) and Nawalkha et al (2003). The technique employs weighted averages of the ‘maturities’ between each cash flow and a planning horizon.

The issue of optionality has been partly answered by the concept of ‘effective duration’ (Fabozzi, 1996), in which the interest rate is perturbed up and down, and the average of the two effects on the bond price is calculated.

These various adjustments and additions to duration all give better results, but at the expense of greater complication in the formulas. These difficulties, combined with the development of bond pricing using stochastic analysis, saw a diminution of the importance of modified duration in the theoretical literature during the 1990s. Vasicek (1977) and Heath et al (1992) are seminal papers in the development of the stochastic approach to bond pricing. Buetow et al (2001) and Jeffrey (2000) contain summaries of the stochastic literature.

Practitioners continue to employ duration in its various forms. Pugachevsky (2003) contains an example in the context of credit default swaps, and Rendaleman (1999) in the
context of hedging interest rate risk. For this reason, duration is widely taught in fixed income courses, with the necessary caveats and cautions about its shortcomings. Recently, there has been a revival of interest; see the contemporary papers cited above, and the text by de la Grandville (2001) that introduces ‘directional duration’.

During this revival, the issue of inaccuracy was fully addressed for the first time. It is possible to restate the orthodox equation for modified duration. All interest rates that are solutions to the bond equation are used in the restatement, rather than the single, orthodox rate of interest (Osborne, 2000a and 2005). In its restated form, the equation can be adjusted to make it a precise and accurate measure of rate sensitivity, able to cope with any sized parallel shift in a flat yield curve (Osborne, 2000b and 2005). The new equation removes the need for auxiliary devices, such as convexity or duration vectors. The fundamental difference between the old and the new analysis is a move from the single dimension of the real number line, into the two dimensions of the complex plane. The multiple solutions for the yield, previously perceived to be a problem, are put to work.

The restated formula for modified duration can cope with any cash flow, and any parallel shift in a flat yield curve, but not with a non-flat yield curve or a non-parallel shift. The contribution of this article is to rework the new, ‘complex’ analysis to produce a second equation that copes with any shaped yield curve and any shaped shift in the curve. The equation is simpler and easier to prove. The new approach involves a change in the way a shift in the interest rate is defined, from the orthodox, additive shift (or spread) to a multiplicative shift. A solution to the fourth shortcoming, the need to make the approach stochastic to account for optionality, is briefly discussed below, but an answer is left for future research.

2 From ‘real duration’ to ‘complex duration’

The derivation of modified duration in its orthodox form begins with the differentiation of the formula for the price of a bond with respect to the interest rate. Equation (1) is a formula for the price of a bond with four coupons remaining. It is priced on a coupon date, to avoid the problem of pricing between coupon dates. (The assumption is
made for simplicity. Both the orthodox and the new, complex formulas cope with the complication.)

\[
p = \frac{c}{(1 + r)} + \frac{c}{(1 + r)^2} + \frac{c}{(1 + r)^3} + \frac{c + f}{(1 + r)^4}
\]  

(1)

In (1), \(c\) is the coupon, \(p\) is the bond price, \(f\) is the final or face value of the bond and \(r\) is the yield to maturity.

The differentiation of (1), and its division throughout by the bond price, yields the formula for modified duration (\(D_m\)) that is equation (2).

\[
D_m = \frac{dp}{dr} \cdot \frac{1}{p} = -\left[ \frac{c/p}{(1 + r)^2} + \frac{2c/p}{(1 + r)^3} + \frac{3c/p}{(1 + r)^4} + \frac{(4c + 4f)/p}{(1 + r)^5} \right]
\]  

(2)

In closed form, for \(n\) time periods, the equation can be reworked to (2a).

\[
D_m = -\left[ \frac{c(1 + r)^{n+1} + fnr^2 - cr(n + 1) - c}{r(1 + r)[c(1 + r)^n + fr - c]} \right]
\]  

(2a)

The equation represents a straight-line approximation to the curved relationship between the bond price and the yield and, as such, is inaccurate. Attempts to improve accuracy have involved auxiliary devices such as convexity, or the higher terms in a Taylor series expansion. Some residual error always remains.

The analysis of the ‘complex’ version of modified duration employs a different approach in order to overcome the inaccuracy problem. The approach makes use of the fact that there is not just one solution to the time value of money equation; there is a constellation of solutions. There are always as many solutions as there are time periods; four in the current example. In Figure 1, the values for \((1 + r) = z\) that are solutions to the bond equation (1) are plotted in the complex plane. The unit circle is shown to provide scale, and because solutions to polynomials like the time value of money equation are always clustered around the unit
circle (see Bogolmolny et al (1996) and Hughes and Nikeghbali (2004)). The formula for modified duration can be restated in terms of distances between the constellation of solutions and other salient points, such as the origin and the point \((l, 0)\). Equation (3) is the new formula for the orthodox concept. Mathematically, (2) and (3) are identical. See Osborne (2000a and 2005) for accounts of the analysis.

\[
|D_m| = \left| \frac{\Delta p}{\Delta r} \right| = \left| \frac{z_1 - z_2, |z_1 - z_3|, |z_1 - z_4|}{|z_1|^4} \right| = \prod_{i=2}^{\infty} \left| \frac{z_1 - z_i}{|z_1|^4} \right| = \frac{d_2 d_3 d_4}{(1 + r_1)^4}
\]  

(3)

In equation (3), each \(z_i\) is a possible solution for \((l + r)\) in the bond equation (1). The term \(|z_n - z_m|\) indicates the distance between two points \(z_n\) and \(z_m\), i.e., the difference between two interest rates. In Figure 1, these distances are represented by \(|z_1 - z_i| = d_i\). If \(z_m = 0\), then it is the center of the circle and \(|z_n - 0| = |z_n|\) is the distance from the circle’s center to \(z_n\). In the expression \(|z_i| = |1 + r_i|\), the element \(r_i\) is the orthodox yield to maturity. The element \(\prod_{i=a}^{b} \left| z_i \right|\) stands for the product of the \(z_i\) from \(i = a\) to \(b\). The term \(\Delta\) indicates that change is discrete rather than continuous. The equation can be generalized to the \(n^{th}\) order.

[Fig. 1 approximately here]

The ‘complex’ formula for modified duration defines how the four solutions to the bond equation (1), for \(z = (1 + r)\), should be employed. The three distances between the orthodox solution and the other three solutions are determined, multiplied together, and divided by the orthodox solution for \((1 + r)\) taken to the fourth power.

The value of this new analysis is that equation (3) can be modified into a new formula for duration that yields precise, accurate results. This new formula has no orthodox equivalent. The modification is small. The element \(z_i = (1 + r_i)\), representing the original, ‘principal’ interest rate, is replaced by the new interest rate, \(z_i^* = (1 + r_i^*)\).
Figure 2 illustrates equation (4) in which $D_c$ stands for ‘complex duration’. The distances, formerly labelled $d_i$, are relabelled $a_i$ to denote the change in meaning. The element $z_i^*$ is now the locus of all the distances (differences in interest rates) instead of $z_i$.

$$|D_c| = \frac{\Delta p}{\Delta r_p} = \left| \frac{z_1^* - z_2^*}{|z_1^*|^4} \left| \frac{z_2^* - z_3^*}{|z_2^*|^4} \left| \frac{z_3^* - z_4^*}{|z_3^*|^4} \right| \prod_{i=2}^{4} \frac{|z_i^* - z_i|}{|z_i^*|^4} = \frac{a_2a_3a_4}{(1 + r^*)^4} \right.$$ (4)

where $|z_i^* - z_i| = a_i$

The equation gives precise, accurate results. No auxiliary devices, such as convexity, are necessary. This solution, to the quest for greater accuracy in duration, is only possible within the context of the complex plane (see Osborne, 2000b and 2005).

[Fig. 2 approximately here]

3 A reinterpretation of the new ‘complex’ equation for duration

The current analysis goes beyond that given in Osborne (op cit.), in order to demonstrate more of the underlying pattern in the new formula. In the equations below, the original expression for the bond price, (1), is shown alongside (5). In the latter equation, the term ‘$a$’ has been added to the interest rate to make the price fall to $p_a$. ¹ If all inputs to equation (5), apart from $a$, are known, then it can be solved for the four values of $a$ that satisfy it. The additive term should more properly be written as $a_i$, from $i = 1$ to $n$, in which one of the values, $a_1$, is the orthodox value, the one that first springs to mind when thinking about solving for $a$ in (5).

¹ The coupon, $c$, is redefined to be $c_i$, from $i = 1$ to $n$, to indicate that the new approach copes with any cash flow, although that is not material to the current argument.
\[ p = \frac{c_1}{(1+r)} + \frac{c_2}{(1+r)^2} + \frac{c_3}{(1+r)^3} + \frac{c_4 + f}{(1+r)^4} \]  

(1)

\[ p_a = \frac{c_1}{(1+r+a)} + \frac{c_2}{(1+r+a)^2} + \frac{c_3}{(1+r+a)^3} + \frac{c_4 + f}{(1+r+a)^4} \]

(5)

The equation for complex duration can now be restated.

\[ |D_c| = \left| \frac{\Delta p}{\Delta r} \right| = \left| \frac{\Delta p}{a_1 p} \right| = \frac{a_2 a_3 a_4}{(1+r+a)^4} \]

In the above equation \( p_a - p = \Delta p \), and \( |z_i^* - z_i| = a_i \), which means that \( \Delta r = a_1 \).

Therefore

\[ \frac{\Delta p}{p} = \frac{a_1 a_2 a_3 a_4}{(1+r+a)^4} = \prod_{i=1}^{4} a_i \]

or, in its general form,

\[ \frac{\Delta p}{p} = \frac{\prod_{i=1}^{n} a_i}{(1+r+a_i)^n} \]

Equation (6) states that a precise, accurate value for duration, in the sense of a percentage change in the price of a bond, is given by the present value of the product of all possible solutions for the additive term in equation (5). Each additive term is a difference between one of the solutions for \( (1+r) \) in (1) and the new, orthodox rate, \( (1+r+a_i) \). In short, plot in the complex plane all the solutions for \( (1+r) \) from equation (1). Connect them with straight lines to the new interest rate, and multiply these distances to get the numerator in (6). Finally, discount the result to the present value at the new interest rate. Equation (6) is a neater representation of the concept than (4). Figure 3 illustrates the equation assuming a fall in the rate of interest.

[Fig. 3 approximately here]
4 An observation on the new interpretation

In the process of calculating the old and the new equations, a theoretical insight emerges concerning the issue of multiple solutions for the interest rate, an issue that Brealey and Myers (2003) refer to as a ‘pitfall’ in their discussion of the internal rate of return in capital budgeting.

The procedure in Osborne (2005) organizes the financial inputs to the time value of money equation into a polynomial (call it poly), and calls up the relevant command for calculating the roots of a polynomial in whatever language is being used (for example, roots). The result of implementing roots (poly), is a list of roots (or zeros) that contains all the possible values for \((1 + r)\) that solve the equation. The list appears as \(\{z_1, z_2, ..., z_n\}\), or something similar, depending on the software used.

When programming formula (3), to calculate the orthodox value of modified duration, there is a difficulty. The locus in Figure 1 is not the new interest rate; it is one particular value of \(z_i\) from the list. The particular value is the ‘principal’ or orthodox value, the value that would normally appear in a financial calculator or spreadsheet, and it has to be identified from the list. In (3), the ‘principal’ value is referred to as \(z_1\). It is not always possible to predict, however, where in the list any particular value of \(z\) will occur. Osborne (2005) refers to the issue, but side steps it by identifying \(z_1\) manually.

Any attempt to identify the principal root from within the program can run into difficulties. A first step is to focus on the real roots, and ignore the complex. A next step is to identify the first root on the real number line above the point \((1, 0)\). For a well-behaved cash flow, with only one change of sign, this works. But if there is more than one change of sign in the cash flow, then there are likely to be multiple, real roots to the right of the point \((1, 0)\). There is a choice. Any automatic procedure to identify the principal root is likely to consist of a series of steps, a sieve, in which some arbitrary decisions have to be made. This issue, of choice within the programming, echoes the problem of choice between multiple interest rates often described in the financial literature (Brealey and Myers, op. cit.).
When computing the new formula for duration, the issue does not arise. In Figures 2 and 3, the locus is not one of the values from the list of roots. Instead, the new interest rate is the locus for the calculation. In equation (6), the absolute values of the differences between the new interest rate, \( z_i^* = (1 + r + a_i) \), and all the interest rates from the list (the \( z_i \)) are calculated and multiplied together. No preference is shown for any particular value of \( z_i \).

There is no need to choose between interest rates, because all of them are used. Not only does the multiple-rates pitfall disappear, but also the multiple rates turn out to be essential elements in the solution to the quest for an accurate version of duration. Multiple rates are not a nuisance; they are useful.

5 From additive to multiplicative spreads

The change in the interest rate in equation (5) is additive. An additive component, the ‘spread’, is commonly found in the empirical and theoretical finance literature. See, for example, Davies and Pugachevsky (2003) and Pugachevsky (2003), who describe various kinds of spreads used in the analysis of credit default swaps. Various versions of such spreads are widely quoted on financial news services. In this paper, the assumption is altered.

The change in the interest rate is made multiplicative. This alteration leads to the generalization of the new formula to the point where it can cope with any shaped yield curve and any shaped shift in the yield curve.

The transition from additive to multiplicative shifts is shown in the following equations, culminating in equation (7). A fall in the rate of interest is assumed.

Let the original rate \( r \) fall to \( R \) such that \( (1 + r) = (1 + R + a) \), and \( r = R + a \). This is the additive expression for \( r \).

If the same change is made using a multiplicative term, then
\[
(1 + r) = (1 + R)(1 + m) = (1 + R + m + Rm)
\]
After cancellation, \( r = (R + m + Rm) \). This is the multiplicative expression for \( r \).
Combining the additive result with the multiplicative result gives
\[ R + a = R + m + Rm, \text{ therefore } a = m + Rm = m(1 + R) \text{ and } \frac{a}{1 + R} = m. \]

Assume, for the moment, that there is one value of \( R \) (the new rate), but \( n \) possible values of \( a \) and \( m \). The last expression becomes \( \frac{a_i}{1 + R} = m_i \). Finally, absolute values are taken, and all possible values are multiplied together to give equation (7). This expression provides a simple link between all possible additive changes to the interest rate in (1) and all possible multiplicative changes.

\[
\prod_{i=1}^{n} |a_i| \over (1 + R)^n = \prod_{i=1}^{n} |m_i| \tag{7}
\]

The expression on the left-hand side of (7) is the right-hand side of equation (6). It is the new expression for complex duration, i.e., the product of all the additive differences discounted at the new rate of interest. It follows that the percentage change in the price of a bond can be restated in terms of the multiplicative changes to the interest rate – equation (8).

\[
\left| \frac{\Delta p}{p} \right| = \prod_{i=1}^{n} |m_i| \tag{8}
\]

Equation (8) is the core equation of the current analysis. It shows that the percentage change in the price of a bond is the product of all the possible multiplicative changes to the interest rate, \( m \), that satisfy equations like (9), below.

\[
p = \frac{c_1}{(1 + R)(1 + m)} + \frac{c_2}{(1 + R)^2(1 + m)^2} + \frac{c_3}{(1 + R)^3(1 + m)^3} + \frac{c_4 + f}{(1 + R)^4(1 + m)^4} \tag{9}
\]
Equation (9) differs from the original bond pricing equation (1), in that the interest rate \((1+r)\) is split into the two factors, \((1+R)\) and \((1+m)\). Equation (10) shows the new bond price \(p_m\) after the interest rate falls from \((1+r)\) to \((1+R)\) by the factor \((1+m)\).

\[
p_m = \frac{c_1}{(1+R)} + \frac{c_2}{(1+R)^2} + \frac{c_3}{(1+R)^3} + \frac{c_4 + f}{(1+R)^4}
\]  

Equation (8) holds true under all circumstances that are non-stochastic. It is general, to the extent that it copes with a non-flat yield curve and any shaped shift in the curve. The next section describes an alternative route to the proof of (8). This new proof is employed to demonstrate the generality of the equation.

6 Another derivation of the new, complex equation for duration

Any polynomial can be transformed into the special form of (11).

\[
-1 + \sum_{i=1}^{n} \frac{A_i}{(1 + x)^i} + \frac{1}{(1 + x)^n} = 0
\]

It is possible to prove, via an analysis of the roots of this special form, that its coefficients and interest rates are linked in the particular way shown by equation (12), below. See Osborne (op. cit.).

\[
|\sum A_i| = \prod |x_i|
\]

Equation (12) shows that the absolute value of the sum of the coefficients is equal to the product of the absolute values of the ‘interest rates’ (the \(x_i\)).

This allows the core equation (8) to be derived from the bond pricing equation (9). The sequence of reasoning is as follows. Equation (9) is manipulated into the special form of equation (11). It becomes equation (9a), below.
\[ -1 + \frac{c_1}{p(1 + R)} + \frac{c_2}{(1 + m)^2} + \frac{c_3}{(1 + m)^3} + \frac{c_4 + f}{p(1 + R)^4} - 1 = 0 \quad (9a) \]

Equation (10) is divided throughout by \( p \), and unity is subtracted from both sides.

\[ \frac{p_m - 1}{p} = \frac{c_1}{p(1 + R)} + \frac{c_2}{(1 + R)^2} + \frac{c_3}{(1 + R)^3} + \frac{(c_4 + f)}{p(1 + R)^4} - 1 \]

When the particular relationship (12), between the coefficients and interest rates of (11), is applied to the two previous equations, the result is the fourth order version of the core equation (8) shown below. This derivation will be used several times in the next section.

\[ \left| \frac{\Delta p}{p} \right| = \prod_{i=1}^{4} |m_i| \]

The multiplicative interest rates, \( m_i \), are rays in the complex plane that have the point \((1, 0)\) as their locus. This is illustrated in Figure 4. The example in the figure assumes an increase in the rate of interest.

[Fig. 4 approximately here]

7 The increasingly general application of the new, ‘complex’ equation

‘It always seems odd to me that the fundamental laws of physics, when discovered, can appear in so many different forms that are not apparently identical at first, but, with a little mathematical fiddling you can show the relationship. … [These] theories … may be equivalent in all their predictions and are hence scientifically indistinguishable. However, they are not psychologically identical when trying to move from that base into the unknown. For different views suggest different kinds of modifications which might be made and hence are not equivalent in the hypotheses one generates from them.’ (Feynman, 1966).
Feynman’s observation is applied twice here. First, in Section 2, the orthodox equation for modified duration was transformed into its complex twin. This transformed equation was then adjusted to make the formula precise and accurate. The adjustment cannot be done to the orthodox equation, because not all solutions to the bond pricing equation are exposed. The term, \( r \), in the orthodox equation denotes all rates simultaneously, but is habitually interpreted as the orthodox rate. In the new equation, all possible rates are distinguished and the relationship between them is clarified. To achieve the new result, the orthodox, principal rate is shifted, while the remaining, unorthodox rates are kept constant.

Secondly, the change from the additive to the multiplicative spread, combined with its analysis in the complex plane, leads to the realization that complex duration can be generalized to cope with non-flat yield curves, and any shaped changes to the curves. The manner in which equation (8) copes with these two issues is described in Sections 8 and 9.

8 Non-flat yield curves

In the following sequence of equations, the yield curve is not assumed to be flat. There are many values for \( r \) that, together, represent a spot yield curve. The values for \( R_i \) are calculated according to the particular value of \( m \) chosen for the multiplicative change in the interest rate. All these values are known in advance, therefore, in the sequence below, the values of \( R_i \) eventually become part of the coefficients.

\[
p = \frac{c_1}{(1+r_1)} + \frac{c_2}{(1+r_2)^2} + \frac{c_3}{(1+r_3)^3} + \frac{c_4 + f}{(1+r_4)^4}
\]

If \( (1 + r_i) = (1 + R_i)(1 + m) \) then

\[
p = \frac{c_1}{(1+R_i)(1+m)} + \frac{c_2}{(1 + R_2)^2 (1 + m)^2} + \frac{c_3}{(1 + R_3)^3 (1 + m)^3} + \frac{c_4 + f}{(1 + R_4)^4 (1 + m)^4}
\]

therefore
\[ p_m = \frac{c_1}{(1 + R_1)} + \frac{c_2}{(1 + R_2)^2} + \frac{c_3}{(1 + R_3)^3} + \frac{c_4 + f}{(1 + R_4)^4} \]

and

\[ \frac{p_m - 1}{p} = \frac{\Delta p}{p} = \frac{c_1}{p} \frac{1}{(1 + R_1)} + \frac{c_2}{p} \frac{1}{(1 + R_2)^2} + \frac{c_3}{p} \frac{1}{(1 + R_3)^3} + \frac{(c_4 + f)}{p} \frac{1}{(1 + R_4)^4} - 1 \]

The second of the equations in this section can be transformed into (9b), which is in the special form of (11) in Section 6.

\[ -1 + \frac{c_1}{(1 + m)(1 + R_1)} + \frac{c_2}{(1 + m)^2} + \frac{c_3}{(1 + m)^3} + \frac{(c_4 + f)}{(1 + m)^4} - 1 \]

\[ = 0 \quad (9b) \]

If the relationship between the coefficients and interest rates of the special form is applied to the last two equations, a fourth order version of equation (8) results.

\[ \left| \frac{\Delta p}{p} \right| = \prod_{i=1}^{4} |m_i| \]

This completes the proof that equation (8) copes with a non-flat yield curve in the context of this fourth-order example. This method of incorporating the yield curve into the analysis was first used in the context of capital budgeting (see Osborne, 2004).

9 Non-parallel shifts in the yield curve

The same reasoning applies to a similar sequence of equations (below) that demonstrates how the equation copes with any shaped shift in the yield curve. This time, the move from the first bond pricing equation to the next equation incorporates indices on \((1+m)\) that can take any values. The value of \(m\) and the values of the indices are chosen to fit with
expectations about how the yield curve shifts. In the example below, the factor \((1+m)\) is
given indices that produce large shifts at the near end of the yield curve, and small shifts at
the far end. The indices need not be smoothly descending or ascending; they can take any
pattern. If the highest indice on \((1+m)\) exceeds that on \((1+R_i)\), then gaps in the sequence of
indices appear, and the order of the polynomial is raised. The gaps have to be padded with
zeros. In equation (9c), below, two such gaps appear, because the highest power on \((1+m)\) is
six, two higher than that on \((1+R_i)\). In contrast, if the highest indice on \((1+m)\) is smaller than
that on \((1+R_i)\), then the order of the polynomial is lowered and some coefficients are
collapsed together.

\[
p = \frac{c_1}{(1 + r_1)} + \frac{c_2}{(1 + r_2)^2} + \frac{c_3}{(1 + r_3)^3} + \frac{c_4 + f}{(1 + r_4)^4}
\]

\[
p = \frac{c_1}{(1 + R_1)(1 + m)^6} + \frac{c_2}{(1 + R_2)^2(1 + m)^4} + \frac{c_3}{(1 + R_3)^3(1 + m)^3} + \frac{c_4 + f}{(1 + R_4)^4(1 + m)^4}
\]

therefore

\[
p_m = \frac{c_1}{(1 + R_1)} + \frac{c_2}{(1 + R_2)^2} + \frac{c_3}{(1 + R_3)^3} + \frac{c_4 + f}{(1 + R_4)^4}
\]

and

\[
\frac{p_m - 1}{p} = \frac{\Delta p}{p} = \frac{\frac{c_1}{p}}{(1 + R_1)} + \frac{\frac{c_2}{p}}{(1 + R_2)^2} + \frac{\frac{c_3}{p}}{(1 + R_3)^3} + \frac{\frac{c_4 + f}{p}}{(1 + R_4)^4} - 1
\]

The second of the equations in this section can be transformed into (9c), which has the
special form of (11) in Section 6.
The relationship between the coefficients and interest rates in the special form is applied to the last two equations. The result, once again, is a fourth order version of the general equation (8) for the percentage change in the price of the bond. The equation still holds, even though the order of the polynomial has changed from four to six (there are now six interest rates). This completes the proof, in the context of a four-period example, that equation (8) copes with any shaped shift in the yield curve.

\[
\frac{\Delta p}{p} = \prod_{i=1}^{6} |m_i| \text{ where the } m_i \text{ are the absolute values of all the solutions from (9c).}
\]

In any non-stochastic situation, the new equation describes the interest elasticity of all vanilla bonds and portfolios of vanilla bonds; it copes with payments between coupon dates, any sequence of cash flows, any shaped yield curve and any shaped shift in the curve. It is a general equation.

10 Two observations on the ‘complex approach’ to duration

A new formula for duration is not a necessity if the objective is to calculate the exact price sensitivity of a portfolio of bonds. The most direct route is to apply the bond pricing equation twice to each bond, once for the original yield curve and once for the new one, then calculate the weighted average of the results. Modern computing power is such that the tedious nature of such calculations is not an obstacle, although the programming is messy.\(^2\)

Moreover, the sum of the coefficients in the equations (9a), (9b) and (9c) -- the coefficients that are counterparts to the \(A_i\) in (11) -- gives duration immediately. It does not

\(^2\) Programming the complex approach is never messy. If the portfolio of bonds is being analysed using a single yield to maturity, i.e., the yield curve is flat at the yield to maturity, and the shift in the yield curve is parallel, then the additive approach embodied in equation (6) is applicable. The cash flows for all bonds can be summed vertically into one large polynomial, the roots of this polynomial identified and connected to the new yield to maturity. Equation (6) can then be applied in a few lines of code to obtain complex duration. The programming is short and not tedious. The process does require suitable math software; spreadsheets cannot do it. Osborne (2005) contains an example of such a calculation. In the same context of a flat yield curve, the multiplicative approach embodied in (8) is also easy to program. A single line of code suffices to calculate complex duration for a vanilla bond, or an entire portfolio of bonds, in one fell swoop. For example, if the vector of roots for the bond equation (1) is labelled \(V\), and \(R\) is the new yield, then, in the language of the math program Derive, \(\text{product(abs}(V/(1+R)-1))=|D_r|\).
make sense to calculate the roots of these equations to find duration. The work has already been done.

So why attempt, yet again, to improve the formula for duration? Two reasons are suggested. First, duration, as a concept, enters into other analyses and becomes a part of other formulas used in empirical finance. Examples are the analyses of Pugachevsky (2003) and Rendleman (1999) cited earlier, in the contexts of credit default swap spreads and hedging respectively. These analyses could be reworked using the new definitions.

Secondly, better formulas lead to better understanding. A formula for a concept may yield results that almost, but not quite, fit reality. The discrepancies may be reduced by means of ad hoc adjustments. But while the discrepancies exist, and the ad hoc adjustments are needed, the original formula cannot be said to have captured the essence of the concept, however useful the formula is for day-to-day, practical purposes. To increase understanding, an entirely new interpretation may have to be imposed on the problem, an interpretation that is both simpler and more accurate. The complex approach to duration is offered in this spirit.

The last point leads to the second observation on the complex approach to duration. It concerns the nature of convexity. In the context of a vanilla bond, or portfolio of such bonds, convexity is traditionally portrayed as a correction for the fact that duration employs a linear approximation to a curvilinear function. Convexity is a measure of the curvature of the bond price – yield relationship.

The ‘complex approach’ sheds new light on the concept. Figure 1 shows that the orthodox approach to the change in the price of a bond joins the (n-1) unorthodox old yields, as well as the new yield, to the one orthodox old yield ($z_i$). Figure 2 shows that the new approach joins all old yields to the one orthodox new yield ($z_i^*$). It follows that inaccurate, modified duration can be converted to accurate, complex duration by one simple move, which is to shift the old, orthodox yield to its new location while holding all other yields constant. Therefore, a correct accounting for convexity requires clarity about what is meant by a change in yield, about which yield from the constellation of yields is shifting. The orthodox approach cannot achieve the necessary clarity, because the formula is locked into
the single dimension of the real number line and does not expose all possible yields in the
constellation. Under the orthodox approach, when ‘the yield’ shifts, all yields shift. The
approach does not discriminate. And there, in this failure to discriminate, lies the source of
convexity.

11 Some puzzles

What is the precise nature of all the interest rates labelled \( m \)? The value, \( m_1 \), is
usually clear enough; it is the orthodox value that is chosen as the interest rate shift, or the
value that results from solving for \( m \) in equation (9) using a calculator or spreadsheet. The
meaning of the \((n-1)\) unorthodox values for \( m \) is not so clear. The fact that the product,
\[ \prod |m_i| \]

is a tangible concept – the percentage change in the price of a bond or a portfolio –
dicates that they have some meaning.

When seen as roots, \((1+m_i)\), rather than rates, \(|m_i|\), each value is a complex number of
the form \( a + b\sqrt{-1} \); it is a point, rather than a distance, in the plane. Any compounding
(discounting) done with these complex rates of interest still works, although it is difficult to
give a financial meaning to the interim results of a calculation. The initial problem and the
final result, however, are both real enough.

One way forward is to see the unorthodox rates as catalysts, in the spirit of the quote
by the mathematician, Hadamard: ‘The shortest path between two truths in the real domain
passes through the complex domain’.

A diversion into unfamiliar territory is worthwhile if it
is rewarded by accurate results. The end justifies the means.

Another way forward is to search for some intrinsic meaning in an imaginary interest
rate. The following thought experiment is due to Peter Carr.

‘Suppose an investor invests a dollar at year zero and it turns into four dollars at the
end of two years. The gross return over the two years is four. The annualized gross return is
two since \( 1 \times 2 \times 2 = 4 \). Now suppose the investor invests a dollar at year zero and it turns
into negative one dollar at the end of two years. In other words, the investor owes another
dollar. This can't happen with a limited liability asset but it can happen with a forward contract or a leveraged stock position. What is the annualized gross return for our unfortunate investor? It has to be the imaginary numbers i and −i since 1x i x i = - 1 and 1 x (-i) x (-i) = -1. Complex interest rates allow us to annualize gross returns even if they are negative. ... Thus, I can be interested directly in complex interest rates.4

A final comment concerns the fourth shortcoming of duration. The new, complex version has yet to be made stochastic in order to cope with optionality. Work in physics suggests that this might be possible. Bogomolny et al (1996) is an example. They study the location of roots of polynomials with random coefficients. The implications of their work for finance are not pursued here, but left for future research.

12 Conclusion

The financial analysis described above takes place in the complex plane. It employs all solutions for the interest rate from the time value of money equation, instead of the one, orthodox rate. The result is a simple equation for duration that yields accurate solutions under any yield curve and any shift in the curve. The core equation shows that an increment in value is the product of all the possible multiplicative increments in the interest rate that cause it. The new approach provides insights, including a cogent reason why the orthodox formulas for duration fail, and a demonstration that multiplicative spreads can be more useful than their additive equivalents. The approach also offers puzzles, particularly surrounding the intrinsic meaning of a complex interest rate.

Finally, duration is but one application of the new methodology. The time value of money equation is ubiquitous in finance. When applied to other financial matters, it is possible that the new approach could yield additional insights.

3 Quoted in Nahin (1998)
4 Quoted, with permission, from correspondence with the author.
References


Figure 1

A picture of equation (3), the ‘complex’ representation of orthodox, modified duration when the interest rate rises. It illustrates the four roots of the bond equation (1) in the complex plane (the \( z_i \)), and all the distances between the roots that comprise modified duration, i.e.,

\[ d_i = |z_i - z_j| \]. It follows that \( |z_i| = |z_i - 0| = |1 + r| \) and \( |\Delta r| = |z_i^* - z_i| \). The locus of the rays is \( z_1 \) (each ray being a difference between two interest rates).
Figure 2

A picture of equation (4) for complex duration when the interest rate rises. It illustrates the four roots of the bond equation (1) in the complex plane (the \( z_i \)), and all the distances between the roots that comprise complex duration, i.e., \( a_i = \left| z_i^* - z_i \right| \). It follows that

\[
\left| z_1^* \right| = \left| z_1^* - 0 \right| = \left| (1 + r^*) \right| = \left| (1 + r + a_1) \right|
\]

and

\[
\left| \Delta r \right| = \left| z_1^* - z_1 \right| = \left| (1 + r^*) - (1 + r) \right| = \left| (1 + r + a_1) - (1 + r) \right| = a_1.
\]

The difference between this figure and Figure 1 is that the locus of the rays is now \( z_1^* \) rather than \( z_1 \).
Figure 3
A picture of equation (6) for complex duration when the interest rate falls. It illustrates the four roots of the bond equation (1) in the complex plane (the $z_i$), and all the distances between the roots that comprise complex duration, i.e., $a_i = |z_i^* - z_i|$. It follows that $|z_i^*| = |z_i^* - 0| = |1 + r_i^*|$ and $|\Delta r| = |z_i^* - z_i| = a_i$. As in Figure 2, the locus of the rays is the new interest rate, $z_i^*$, rather than the original rate, $z_i$. 
Figure 4
A picture of equation (8) for complex duration given a multiplicative increase in the interest rate. The figure illustrates the four roots of equations (9a), (9b) or (9c) in the complex plane, and all the distances that comprise complex duration (the $m_i$). The locus of the rays is the point $(I, 0)$ (each ray being a multiplicative interest rate).
Work 7
Osborne, M.

On the computation of a formula for the duration of a bond that yields precise results

NOTICE: this is the author’s version of a work that was accepted for publication in *The Quarterly Review of Economics and Finance*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *The Quarterly Review of Economics and Finance*, VOL 45, ISSUE 1, (2005), DOI:10.1016/j.qref.2004.07.003
ON THE COMPUTATION OF A FORMULA FOR THE DURATION OF A BOND THAT YIELDS PRECISE RESULTS

Michael J. Osborne *, Gulf International Bank **

Abstract

In fixed income analysis it is known that the various measures of interest rate sensitivity (duration) yield approximate results. Even with the addition of concepts like convexity, the results remain approximations. This paper summarizes a new approach based on the fact that the time value of money equation is a polynomial, and a polynomial has more than one root. The result of taking the multiple roots into account is a solution to the problem of inaccuracy. A new equation for duration is given that provides precise results. The paper contains a summary of previous work, describes the computational issues presented by the new approach, and suggests ways to deal with them.

Key words: Duration, convexity, fixed income, bond, complex plane

JEL classification codes: C60, G10, G11, G12

* The author wishes to thank the following for advice and assistance at various times: Robert Durand of the University of Western Australia; Chris Deeley of Charles Sturt University; Sunil Dattani, P.K. Nambiar and Jaffar Hussain of Gulf International Bank; Chris Brooks of the ISMA Centre at Reading University; Petros Geroulanos of Epigon; Carter Daniel of Rutgers University; Peter Carr of Bloomberg and NYU’s Courant School of Applied Mathematics; Zoran Kozic, Patrick Hagan, Gene Cohler and Leo Tilman of Bear Stearns; and Jay Marchand of Mercer University. Guy Judge, editor of Computers in Higher Education Economics Review, graciously gave permission to use material, notably some proofs, from previous publications. Thanks go to the participants in the conference (Computational Intelligence: Methods and Applications) held at the University of Wales in Bangor, UK, during June 2001, who answered my questions patiently and pointed me to the methods and software to implement the new theory.

** Gulf International Bank, Al-Dowali Building, 3 Palace Avenue, P.O. Box 1017, Manama, Kingdom of Bahrain. Telephone 00 973 17 522589, fax 00 973 17 522643, email mike.osborne@gibbah.com.
On the Computation of a Formula for the Duration of a Bond that yields Precise Results

1 Introduction

This article is about the time value of money (TVM) equation and one of its applications in fixed income analysis.

The TVM equation is a general-purpose analytical tool in economics, banking, accounting and finance. Various assumptions about the variables and parameters of the equation, or operations on the equation such as its differentiation, can lead to a host of formulas for different financial concepts. The concepts range from pricing formulas for loans, bonds and equity, through many of the derivatives of these products, to tools used to analyze the concepts, such as measures of their interest rate sensitivity.

The TVM equation plays a smaller role in the analysis of other concepts, e.g., options and portfolio management. Statistical ideas, such as mean and variance, and the normal distribution, are additional parts of the toolkit for these concepts. Yet, even here, the TVM equation is important.

It is the thesis of this article that an exhaustive analysis of the TVM equation has yet to be completed. Additional analysis is made available by means of a single act which is to recognize that the equation is a polynomial, and a polynomial has a far richer range of analysis attached to it than is currently acknowledged in the financial literature. In particular, when solved for the rate of interest, the equation gives multiple solutions. All but one of these solutions are currently ignored. It is recognition of this fact that opens up the research program.
The article focuses on one example of the TVM equation, that for pricing a bond, and an equation derived from it, that for modified duration (the duration of callable or puttable bonds is not discussed). Other financial applications exist already but are not explored here. It is possible that many other areas of finance could be analyzed using the new technique, but they are not discussed either.

The structure of the article is as follows:

- Some basic ideas are outlined in Section 2, including the required mathematical concepts and the notion that the TVM equation has multiple solutions for the rate of interest.
- Sections 3, 4 and 5 show how the core idea is applied to bond pricing and duration. Most especially, a solution to the long-standing problem of the inaccuracy of duration is described in Section 5. A precise version of the measure of risk is found by means of the new analysis, a version that has no need of the ancillary concepts, such as convexity.
- Section 6 contains a worked example that shows how the new approach can be programmed, easily and intuitively, using commercially available software. The example is the calculation of the interest rate sensitivity of a small portfolio of two assets that, together, have an uneven and irregular cash flow. It is shown that the new method copes with the problem simply and accurately, in a way that the orthodox approach cannot.
- Section 7 deals with implementation; a computational question is posed by the new approach and this section provides two different possible answers.
- Section 8 contains a brief discussion of the practical uses of the new formula for duration in the contexts of reporting, hedging and trading strategy.
- The final section is the conclusion.
Sections 2 to 5, providing the theoretical foundation of the new approach, are based on prior work (Osborne, 2000a, 2000b). Sections 6 and 7 build on these theoretical foundations by exploring the practical issues of implementation (first raised but not fully explored in Osborne, 2001).

The mathematics in the text has been kept to a minimum; for example, mathematical proofs of assertions made in the text are relegated to three appendices. However the point at issue is the mathematical interpretation of the TVM equation, and therefore some equations in the text are unavoidable.

2 Not one interest rate, but many

In this section the basic ideas are explained using very simple examples that are spelt out in some detail.

Consider equation (1). It is a general version of the TVM equation -- 'general' to the extent that the cash flows can take any value (the flow need not be uniform) and the number of cash flows is not determined.

\[
p = \sum_{i=1}^{n} \frac{c_i}{(1+r)^i} + \frac{f}{(1+r)^n}
\]  
(1)

In equation (1), \(p\) is the price or present value of the cash flows \(c_i\) (paid or received at times \(i = 1\) to \(n\)) and the cash flow \(f\) (some final or future cash flow occurring at time \(n\)); \(r\) is the internal rate of return, the yield to maturity, the annual percentage rate, or whatever name is appropriate to the context.

A simple version of (1), that for a plain vanilla bond, has the following characteristics:
\{p, c, f, n, r\}=\{0.913223140496, 0.05, 1, 2, r\}.

Twelve places of decimals are used for inputs because the precision allows clearer verification of the results. The price is 0.913223140496, the coupon is 0.05 in each period, the face value is one and the number of periods is two years. The yield is the unknown quantity. The equation is (1a):

\[
0.913223140496 = \frac{0.05}{(1 + r)} + \frac{1.05}{(1 + r)^2}
\]  

(1a)

Solving for the yield, using a financial calculator or spreadsheet, the answer is 0.1 or 10%. This result could be calculated manually by rearranging equation (1a) into the mathematical presentation (1b) and applying the familiar formula for the solution of a quadratic:

\[
0.913223140496z^2 - 0.05z - 1.05 = 0 \quad \text{where} \quad z = (1 + r)
\]

(1b)

If \(az^2 + bz + c = 0\) then \(z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\)

The manual calculation produces two values for \(z = (1 + r)\), namely 1.1 and -1.045249, which implies the yield is 0.1 or 10%, which is known already, and -2.045249 or -204.5%. Both values are mathematically correct, but since only the first makes financial sense, it is usual to discard the negative value, which financial calculators and
spreadsheets are not even programmed to calculate. It is, however, a thesis of this article that negative values are of use and should not be discarded.

The two solutions are depicted on the real number line in Fig. 1. This simple figure is included for reasons that may not be clear now, but which will become apparent.

[ Fig. 1 approximately here ]

The story can become more complicated. The example is changed by increasing the number of periods to three, thereby creating three solutions, and assuming another price. A price is assumed that, as in the previous example, makes the yield come out to the convenient value of 10%; i.e., the bond has the following characteristics:

\[ \{p, c, f, n, r\} = \{0.875657400451, 0.05, 1, 3, r\} \]

The resulting equation is seen in its financial form as (1c) or in its mathematical form as (1d):

\[ 0.875657400451 = \frac{0.05}{1 + r} + \frac{0.05}{(1 + r)^2} + \frac{1.05}{(1 + r)^3} \]  

\[ 0.875657400451z^3 - 0.05z^2 - 0.05z - 1.05 = 0 \quad \text{where} \quad z = (1 + r) \]

The formula for finding the values of \(z\) when the equation is a cubic \((n=3)\) is rarely used, although a formula does exist. The solutions to this example were calculated using numerical methods in a dedicated math program. Any of the popular programs will do, e.g.,
Mathcad, Matlab, Mathematica, Maple, Derive. In this case it was Matlab. The three results are listed below.

\[ z = 1.1, \quad z = -0.52145 + 0.904533\sqrt{-1} \quad \text{and} \quad z = -0.52145 - 0.904533\sqrt{-1} \]

The puzzle is the meaning of a value for \( z = (l+r) \) that has two components, one of which contains the element \( \sqrt{-1} \). Despite the fact that its intrinsic meaning is unclear, the element \( \sqrt{-1} \) is known as \( i \) and has been used successfully to solve problems in electrical engineering and the sciences for many years. Numbers like \( z \) are given the name complex numbers because they contain two components.\(^1\) The second component is described as imaginary, an adjective that is unfortunate because it suggests such numbers do not exist. Mathematicians (see Stillwell, 2002 or Nahin, 1998) see them as real and useful as any other kind of number.

Complex numbers are usually interpreted geometrically, with the real axis depicted on a horizontal line, and the imaginary axis depicted as a vertical line, measured in units of \( i \). The result is the complex plane. The two solutions found earlier are depicted in Fig. 2, along with the single real, orthodox solution that has the value 1.1. A circle with radius one, the unit circle, is drawn for reference, because all solutions for a polynomial with real coefficients of the magnitude usually encountered in the TVM equation are clustered on, or around, the unit circle.\(^2\)

---

\(^1\) The letter \( z \), rather than \( x \), is used throughout this article to indicate a solution to the TVM equation. This is because \( z \) is orthodox mathematical notation wherever a variable can take complex values.

\(^2\) Visit the website www.cecm.sfu.ca/organics/papers/odlyzko/support/polyform.html. This site contains an online calculator and graphing utility that allows the viewer to see the roots of a polynomial in the complex
The financial meaning of a complex solution that has an imaginary component is not clear. Perhaps there is none. But it does not matter in the current context because an intrinsic meaning is not needed for the complex solutions to be useful. There exist real mathematical questions that are sometimes very difficult, or even impossible, to solve, except by acknowledging the existence of complex numbers. The numbers are used as a catalyst to a solution, not as an end in themselves. The French mathematician Jacques Hadamard put it pithily: 'The shortest path between two truths in the real domain passes through the complex domain' (quoted in Nahin, 1998).

The remaining sections show how both the complex and negative solutions to the TVM equation are used to solve a real financial problem: the inaccuracy of the various versions of duration in fixed income analysis. The problem has persisted in the literature since the time of Macauley (1938) because financial analysis has been constrained to the real domain. I show that a version of duration yielding precise results can be obtained only by acknowledging the other, complex, solutions to the TVM equation. This new approach uses complex solutions as a catalyst on the route between the problem and its solution. The various 'patches' to duration, such as convexity, and the other higher order terms of a Taylor's series, are unnecessary complications in the search for an accurate formula for an interest elasticity.

This section concludes with a visual explanation of the three interest rates in the current example. At the moment the three roots, or values of \((1+r)\) that satisfy equation (1c)
are plotted in the complex plane. The three interest rates are the three values of \( r \) obtained by subtracting unity from each root. In order to do this in the complex plane, absolute values are taken, i.e., \( |z - 1| = |(1 + r) - 1| = |r| \). Since the absolute value of a complex number is a distance in the complex plane, the results are three distances in the plane radiating from the point \((1,0)\) to the three roots. They are depicted in Fig. 3. The rate \( r_i \) is the orthodox rate, which, in this example, is 0.1 or 10%. In general, if there are \( n \) time periods in the TVM equation, then there are \( n \) roots that satisfy the equation, and, therefore, \( n \) interest rates that can be depicted in the complex plane as \( n \) distances radiating from the point \((1,0)\) to the roots.

[Fig. 3 approximately here]

3  A new bond pricing equation

Before giving the new, precise equation for duration, a small detour is taken. In the next few paragraphs a new bond pricing equation is derived.

First, note that any TVM equation can be rearranged to have the structure of equation (3). The equation below (3) is an example of this -- it is a rearranged version of the bond pricing equation (1) with \( n = 4 \).

\[
-1 + \frac{a_1}{(1 + r)} + \frac{a_2}{(1 + r)^2} + \frac{a_3}{(1 + r)^3} + \frac{a_4}{(1 + r)^4} + \frac{1}{(1 + r)^4} = 0 \quad (3)
\]

\[
-1 + \frac{c/p}{(1 + r)} + \frac{c/p}{(1 + r)^2} + \frac{c/p}{(1 + r)^3} + \frac{(c + f)/p - 1}{(1 + r)^4} + \frac{1}{(1 + r)^4} = 0
\]
It is stated here (and proved in Appendix A) that the coefficients and the roots of (3) are linked in a special way. The absolute value of the sum of the coefficients is equal to the product of the absolute values of the interest rates.

\[ |\sum a_i| = \prod |r_i| \]

This result is applied to the equation immediately below equation (3), that is the rearranged version of the bond pricing equation. The result is itself rearranged and the end product is a new equation for the price of a bond -- equation (4) -- that says the price of a bond is equal to the sum of all the cash flows divided by one plus the product of all the interest rates. It is not obvious that equations (1) and (4) are identical.

\[ p = \frac{\sum c + f}{1 + \prod |r|} \quad (4) \]

At this point the focus shifts to the analysis of duration. In this context it is shown that the new approach has practical value.

4 Duration

First, consider the following numerical example. The example contains cash flows over four periods of one year each \((n=4)\). The coupon remains at 5\% \((c=0.05)\), the face value is still \$1 \((f=1)\), while the price is 0.841506727683.

\[ \{p, c, f, n, r\} = \{0.841506727683, 0.05, 1, 4, r\} \]
\[
0.841506727683 = \frac{0.05}{(1 + r)} + \frac{0.05}{(1 + r)^2} + \frac{0.05}{(1 + r)^3} + \frac{1.05}{(1 + r)^4}
\]

The values for \(z=(1+r)\) are:

\(1.1; \ -1.042913; \ 0.001165 + 1.042906i; \ 0.001165 - 1.042906i.\)

These values are illustrated in Fig. 4.

[Fig. 4 approximately here]

What is the interest sensitivity of the price of this bond, i.e., what is the value for its duration? The answer depends in part on which definition of duration is used.

One way to calculate the duration of a plain vanilla bond is to calculate the weighted average of the times of the cash flows, each weight being the ratio of the present value of the cash flow to the price of the bond. This gives ‘Macauley duration’ (Macauley, 1938). A second way is to start with the bond pricing equation, then differentiate price with respect to yield and construct the price elasticity. This gives ‘modified duration’ (Hicks, 1939, Samuelson, 1945). A third method is to perturb, up and down, the yield in the bond pricing equation and calculate the average effect of the perturbations. This gives ‘effective duration’ (see Fabozzi, 1996). All the methods have their advantages and disadvantages -- elegance versus awkwardness, simplicity versus complexity, ease versus difficulty of application to more complex callable and puttable bonds. Yet all are wrong, because they all give approximate results.

The fact that none of the methods yields precise results has given rise to a large literature describing auxiliary procedures to increase the precision. The procedures include
the addition of convexity and other, higher order terms of a Taylor's series, or creating a duration vector, whereby each cash flow in the series has its own duration. Chambers et al. (1988) make an important contribution to this literature and also summarize the early contributions. A more recent example of the many attempts to improve accuracy within the orthodox framework is Youngsoo Choi and Jinwoo Park (2002). In the last few years, attention has turned to the interest rate sensitivity of the risk measures themselves, i.e., how duration and convexity are affected by changes in the structure of the yield curve (see Crack and Nawalkha, 2000). Unfortunately, all the procedures obtain their more accurate (though not precise) results at the expense of greatly increased complexity in the calculations.

From this point onwards the focus is on modified duration (which encompasses Macauley duration). Effective duration, the version useful in the context of callable or puttable bonds, is not considered at all.

The orthodox procedure used to calculate modified duration is as follows. First, the equation for the bond price is written down. The equation below is equation (1) with four coupons remaining. It is priced on a coupon date so that the problem of pricing between coupon dates is avoided, a simplifying assumption that does not affect the analysis in any way.

\[
p = \frac{c}{(1+r)} + \frac{c}{(1+r)^2} + \frac{c}{(1+r)^3} + \frac{c}{(1+r)^4} + \frac{f}{(1+r)^4}
\]

Then the expression is differentiated with respect to yield, and the elasticity known as modified duration is constructed, equation (5).
\[ D_4 = \frac{dp}{dr} \cdot \frac{1}{p} = \left[ \frac{c/p}{(1+r)^2} + \frac{2c/p}{(1+r)^3} + \frac{3c/p}{(1+r)^4} + \frac{4c/p}{(1+r)^5} + \frac{4f/p}{(1+r)^5} \right] \tag{5} \]

The general result expressed in closed form is (5a).

\[ D_n = -\left[ \frac{c(1+r)^{n-1} + fnr^2 - cr(n+1) - c}{r(1+r)c(1+r)^n + fr - c} \right] \tag{5a} \]

5 Some new formulas for duration

Two new alternative formulas for duration are now stated. They are embedded in the complex plane and make use of all the solutions to the TVM equation. First, equation (6) looks very different from the orthodox but is, in fact, identical to (5) and (5a). A proof of this equivalence is given in Appendix B. Fig. 5 illustrates equation (6) in the context of the four period numerical example given earlier.

\[ D_4 = -\frac{\left| z_1 - z_2 \right| \left| z_1 - z_3 \right| \left| z_1 - z_4 \right|}{\left| z_1 \right|^4} = -\prod_{i=2}^{i=4} \left| z_1 - z_i \right| = \frac{d_2d_3d_4}{(1+r)^4} \tag{6} \]

In equation (6), \( \left| z_n - z_m \right| \) indicates the distance between two points \( z_n \) and \( z_m \). If \( z_m = 0 \), i.e., it is the center of the circle, then \( \left| z_n - 0 \right| = \left| z_n \right| \) is the distance from the circle’s center to \( z_n \). Finally, \( \prod_{i=1}^{n} z_i \) stands for the product of the \( z_i \) from \( i = 1 \) to \( n \).

[Fig. 5 approximately here]
The new version of modified duration states that, given four solutions for \( z = (1 + r) \) to the TVM equation, the three distances between the orthodox solution and the other three solutions are determined, multiplied together, and divided by the orthodox solution for \( (1 + r) \) taken to the fourth power.

The second equation, \( (6a) \), is a slightly modified version of \( (6) \). It yields precise results and has no orthodox equivalent. The proof of this is given in Appendix C. Fig. 6 illustrates equation \( (6a) \), once again in the context of the four period numerical example. The modification made to equation \( (6) \) is a change that cannot be made to the orthodox formulas. The point \( z_1 \) is shifted along the number line by the amount of the interest rate change, \( \Delta r \), shifting outwards from the origin if the yield goes up and inwards if it goes down. In Fig. 6 the rightward move of \( z_1 \) to \( z_1^* \) indicates a rise in the yield from 10% to 11%. Most importantly, the other roots, or \( z_i \), are left where they are. This procedure is equivalent to changing one yield, the orthodox yield, while leaving all the other yields unchanged. Now duration is recalculated using the new method. The straight lines \( d_i = |z_i - z_1| \) in Fig. 5 are now stretched to \( |z_i^* - z_1| \) in Fig. 6.

\[
\text{precise } D_4 = \frac{|z_1^* - z_2| |z_2^* - z_3| |z_3^* - z_4|}{|z_1^*|^4} = \frac{\prod_{i=2}^4 |z_i^* - z_1|}{|z_1^*|^4} = -\frac{d_2^* d_3^* d_4^*}{(1 + r^*)^4}
\]

(6a)

The only difference between equations \( (6) \) and \( (6a) \) is that \( z_1 \) moves to \( z_1^* \), showing the new higher yield ( \( r \) to \( r^* \) ). All other elements remain the same. The new formula is using one new yield and \( (n-1) \) old yields, something the old formula could never do.
It is worth stressing this critical point. The precise formula for duration can be achieved only via the complex plane because it is necessary to move one of the roots (or interest rates), and only one of the roots, to achieve a precise result. The new formula exposes all the roots and facilitates this action. The orthodox formulas can never yield a precise solution because the term $z = (1 + r)$ stands for all the roots (or interest rates) simultaneously. In the straitjacket of the 'real formula' it is impossible to move the primary root without shifting all the others. To date, these 'others' have existed behind the scenes, and their effect on the analysis has been ignored.

To complete the picture for the four-period numerical example of this section, the various values for modified duration are given. Using the orthodox formula, or its new equivalent (6), the answer is -3.35918332712. The precise value for duration obtained from the new equation (6a) is -3.28618514247. This assumes an increase in yield of 1% (which makes it comparable with the assumption of a 1% change that is implicit in the usual way the orthodox formula is quoted). The precise value obtained from repeated application of the bond pricing formula, assuming an increase in yield of 1%, is -3.28618514259. Thus, the new approach is accurate to the ninth decimal place, while the orthodox approach is not accurate even to one decimal place.

The general formula for ‘n’ periods is given in equation (6b).

$$\text{precise } D_n = - \frac{\prod_{i=2}^{n} |z_i^* - z_i|}{|z_1^*|^n}$$

(6b)
This general equation holds true not only for the analysis of a single cash flow, but also for multiple cash flows of all kinds: coupon bearing bonds; bonds between coupon dates where the cash flow is not regular; portfolios of bonds where the cash flow is both varied and irregular. An example of the calculation of precise duration for a small portfolio, in which the cash flow is varied and irregular, is given in the next section. Where the cash flow is multiple and varied in both amount and timing, identifying the locations of the roots (the \( z_i \)) can be more difficult than in the situation described here. A discussion of this issue is in Section 7.

6 \hspace{1em} \textbf{A worked example: a two-asset portfolio with uneven and irregular cash flows}

Once again, we ratchet up the level of complexity. The date is 1 August 2002. Two plain vanilla bonds are purchased, both paying semi-annual, actual-actual. The first bond matures on 1 April 2004, pays a 6% coupon and costs 0.959797851371 per unit value. As before, twelve places of decimals are used in order to emphasize the precision available from the new approach. The second bond matures on 1 August 2005, pays an 8% coupon and costs 0.949243079328 per unit value (it is bought on a coupon day, ex-coupon). The yield for both bonds, once again by construction and for convenience only, is 10\% pa. The cash flows for both bonds and the resulting portfolio are depicted in Fig. 7.

The unit of time is two months, chosen because all the powers in the TVM equation must be whole numbers. The unit of time should be the shortest unit that captures the cash flows in a polynomial with integers as indices. The shorter the unit of time, the higher the order of the polynomial. At the limit, for a very complex cash flow in a large portfolio, the unit of time could be one day. The computational difficulties that arise from short units of time, and the resulting high order polynomials, are addressed in the next section.
The third cash flow in Fig. 7 is the vertical summation of the first and second flows. The essential point about this example is that the third cash flow for the portfolio is both irregular and uneven and therefore it represents a challenge for the new approach.

Under the orthodox approach, finding the modified duration of the portfolio involves calculating the modified durations of the two bonds separately, and then finding a weighted average of the two results. Since one of the bonds is partway through a coupon period, an adjusted version of the equation for modified duration is needed for that bond (an equation that has, itself, been the subject of some debate in the literature -- see, for example, Cole and Young, 1995). The resulting duration for the portfolio is an approximation because the individual durations are approximations. A more accurate yield elasticity requires further calculations for convexity, and yet the result is still not precise.

Contrast the foregoing situation with the following application of the new formula for precise duration using standard commands in Matlab. The commands are given in Exhibit 1. It is a measure of the intuitive nature of the new approach that a few standard commands from an off-the-shelf program are all that is needed to implement it and produce a precise result.

The first line in Exhibit 1 makes sure that all previous activity in Matlab is cleared. The second line calls up many decimal places to show accuracy. The third line sets up the polynomial using the cash flows of the portfolio as coefficients. The fourth line finds the roots -- the values of \( z = (1 + r) \) in complex space.

[Fig. 7 approximately here]
The fifth line re-orders the roots so that the first positive root above zero comes first. One of the slight problems when using any math program to find the roots of a polynomial is that the output is not always in the preferred order. When performing such calculations it is worth checking the output before continuing, or inserting a line of code that reorders the results automatically. This reordering, to bring the primary root to the top of the list, is not essential, but it does make the critical command in the seventh line easier to code.

The sixth line calculates the change in the interest rate on a two-monthly basis. In this example the annualized change is an increase of 1%, but it could be any change. It is assumed to be 1% in order to compare it with orthodox duration (which is nearly always quoted on the assumption of a 1% change in the annual rate). A change of 1% on an annual rate of 10% means that the semi-annual rates of 5% and 5.5% are taken to the power 'one third' in order to convert the change of rate to a two-monthly basis.

The seventh line implements the new formula.

[Exhibit 1 approximately here]

The last two elements in the last line of code may need some explanation. The value for duration in the example has been calculated using units of two months. It needs to be adjusted to a result based on units of one year to compare with the orthodox method of quoting it.

\[
\frac{dp}{dr} = \frac{1}{p} \quad \text{therefore} \quad \frac{dp}{p} * 100 = D * dr * 100
\]
Multiplying both sides of the equation for duration (expressed in units of two months) by the change in the yield on a two monthly basis \((dr=a)\), restores the result to the required quotation for duration.

The intermediate outputs from the code on Exhibit 1 are not given here for the sake of brevity. However, the final output that matters, the value for new duration from the seventh line, is -2.01401758390146. This precise value can be checked by the direct method of calculating the value of the portfolio at the two interest rates of 10% and 11% using a standard financial calculator. The accuracy is to $1 in $100bn. The new approach eliminates the need to use convexity, the other higher order terms of a Taylor's series, or devices such as duration vectors.

Using a math program to solve the simple problems outlined in the earlier sections would be inappropriate. But in the context of this last example, with its more complicated cash flow, it is appropriate to use a math program. The intuitive nature of the new approach allows standard commands in Matlab to cut through the complexity of the calculations. Precise results are produced in a few lines of code.

7 A computational issue

The new method requires the location of the roots of the TVM equation to be identified with some precision. Formulas for finding these locations do not exist for \(n>4\). Numerical methods have to be used. Osborne (2001) raised the issue of the scale of the problem. Is it tractable or not?

A fixed-rate retail loan over four years with monthly repayments requires the calculation of 48 solutions (monthly interest rates). A 25-year fixed rate retail mortgage with
monthly repayments results in 300 solutions. In the bond markets a 30-year bond with semi-annual coupons on the day of issue has 60 solutions.

The level of difficulty is raised when the calculations are done between coupon or repayment dates. The unit of time used in the calculation has to be such that the polynomial has integers for all powers. In the example in Section 6 the portfolio extended over three years, but the irregular coupon dates required the unit of time to be two months in order to have whole numbers for all the indices. The resulting calculation involved finding the roots of a polynomial of order eighteen.

In the worst case, the day after the issue of a 30-year bond that is included in a complicated portfolio, the unit of time may have to fall to one day. The order of the polynomial would then be raised to 30 years x 365 days = 10,950. This is an extreme situation representing the upper bound of the problem. Many conservative portfolios have durations of just a few years, with their longest dated components of duration less than 10 years. If the cash flow is such that a week could be the unit of time, then the problem is of order 52 weeks x 10 years = 520.

What are the limits to a computation like this? What order of polynomial can be analyzed in a reasonable period of time on a good machine? How far is it possible to make practical bond calculations with a reasonable degree of accuracy, say, one dollar in a million?

There are two ways forward. One way is to confront the problem directly and do the calculations using numerical methods -- the brute force approach. Another way is to get around the issue by redefining the problem. If the cash flow is broken into a series of zero coupons the computational problems are greatly eased.

The brute force approach is examined first. While most financial calculators and spreadsheets have numerical methods built into them, most cannot handle complex numbers,
and thus we must resort to math programs. Fortunately many off-the-shelf programs are available commercially, as mentioned earlier.

Table 1 shows the time to compute the locations of the roots of a series of polynomials of increasing order. As in Section 6, the computation was done using standard commands in Matlab. The coefficients of the polynomials were generated randomly from within the software. The timing was also done from within the program. The computer was a standard PC currently available commercially. It had a Pentium 4 chip running at 2.53 GHz and 256 MB of RAM.

[Table 1 approximately here]

It remains to be seen how a better computer could speed up the process to bring the upper bound of computation within range (a computer with a higher clock speed, larger RAM and possibly multiple processors running in parallel). It also remains to be seen what unit of time is actually needed for the typical portfolio, i.e., to what extent the upper bound can be lowered.

The second way forward is to simplify the problem. Computation is easier if a portfolio is broken into its individual cash flows, a series of zero coupons, because the roots of a zero coupon are much easier to locate. They are regularly distributed around a circle of radius \((1 + r)\), where the yield is the relevant market yield.

For example, in the portfolio analyzed in Section 6, the tenth cash flow is 1.03 and the yield is 1.639625681\% or 1.05^{(1/3)}. The ten roots for this zero coupon are located on a circle of radius 1.01639635681, starting with the point 1.01639635681 on the real number line, and thereafter at angular intervals of 36 degrees. The situation is illustrated in Fig. 8, with all
roots shown on the circumference, and all interest rates shown radiating from the point (1,0).
The distances in the figure are exaggerated to show the pattern clearly.

The orthodox modified duration for this zero coupon can be calculated using the
distances between the principal root $z_i$ and the other roots according to the new formula.

Application of the precise formula for duration requires that the principal root is
moved and that the distances are calculated between its new location and the locations of the
other roots (that have not moved). This situation is depicted in Fig. 9 with the principal root
now labeled $z_i^*$ to indicate that it is at its new location.

[Fig. 8 approximately here]

If a precise value for duration is calculated for each cash flow, then the overall
duration can be calculated from the duration vector as the weighted average of the individual
durations. The weights are the present values of the cash flows relative to the value of the
portfolio. The calculation is similar to the orthodox calculations done to produce duration
vectors, or key rate and partial rate durations. Such a procedure has the added advantage that
the different yields along the yield curve, as well as the different movements in yields along
the yield curve, can be taken into account. The only differences are in the new method used
to calculate the individual durations, and the precision of the results.

[Fig. 9 approximately here]
8  Practical uses: reporting, hedging and trading strategy

There are three main areas where duration is of practical use and where more accurate figures may be welcome.

The first area is the reporting process. It is normal for any portfolio to be reviewed regularly to determine if it is to be restructured or left as it is. Such decisions are based on various decision tools. Nowadays, tools like Value at Risk (VAR) are used; however, duration figures are still included on many management reports because they are familiar and readily understood by most market participants. More accurate figures, even if they are limited to a standard assumption about the causal change in yield of, say, one per cent, are likely to lead, at the margin, to better decision making. It is difficult to quantify this effect. It may come down to something as simple as the psychological issue of whether the duration of a portfolio is above or below some critical figure in the minds of the decision-makers.

The second area is hedging. Many hedging decisions require the calculation of hedge ratios that incorporate duration figures. The precise number of contracts calculated as necessary to hedge a position may change with a fresh and more accurate input. Since only a whole number of contracts can be bought or sold, it is possible that more, rather than less, accurate figures for duration may mean one more, or one fewer, contract is bought or sold. More accurate hedging may result.

Finally, there is strategy formulation by traders and portfolio managers. The decision as to what strategy to follow will often depend on 'what-if' calculations, many of which incorporate duration figures. A more accurate figure may tip a strategy over a critical mark and so cause a trade to take place when it may otherwise not have occurred, and vice versa.
9 Conclusion

The focus of this article has been on fixed income analysis, specifically duration. Duration (and its allied notion of convexity) is found in almost every syllabus and financial text. It cannot be said, however, that the concept is at the cutting edge of finance. Risk management in fixed income markets is currently carried out using more modern techniques, such as Value at Risk. Despite this caveat, it has been shown that the new approach does yield insights into a genuine financial problem concerning the TVM equation. The point has been made already that the TVM equation is ubiquitous in finance. It follows that duration may not be the only concept whose analysis could benefit from the approach. Many areas of finance remain unexamined in the light of the complex solutions to the equation. Further useful and interesting results might be available.

One research prospect is to introduce a stochastic element into the analysis, most particularly into the coefficients (or cash flows). The analysis of polynomials having random coefficients via the distribution of their roots in the complex plane is not new – the physicists have been doing it for some years. Bogomolny et al. (1996) is an example.

Finally, while the preceding analysis has shown that taking finance into the complex plane may provide some practical benefits, it is surely not only the practical benefits that should be considered. Any practitioner or academic aspiring to fixed income expertise may wish to know about these results. To date, the explanation of why the various versions of duration do not give accurate results has been limited to the observation that they incorporate a linear approximation to a curvilinear function (that between the yield and the price of a bond). The paper provides a better theoretical understanding of this classic measure of risk, which is, perhaps, a desirable thing in itself.
Appendix A

A proof of the link between the coefficients and roots of the TVM equation.

In the text, it is noted that any TVM equation can be rearranged to have the structure of equation (3) in the text. Equation (3) is repeated below for convenience. Its general form is equation (A1).

\[-1 + \frac{a_1}{(1 + r)} + \frac{a_2}{(1 + r)^2} + \frac{a_3}{(1 + r)^3} + \frac{a_4}{(1 + r)^4} + \frac{1}{(1 + r)^4} = 0\]  \hspace{1cm} (3)

\[-1 + \sum_{i=1}^{n} \frac{a_i}{(1 + r)^i} + \frac{1}{(1 + r)^n} = 0\]  \hspace{1cm} (A1)

It was stated that the coefficients and the roots of (A1) are linked in a special way. The absolute value of the sum of the coefficients is equal to the product of the absolute values of the interest rates.

\[|\sum a_i| = \prod |r_i|\]

A proof of the statement for the general equation (A1) is as follows. First, set \(z = (1 + r)\) and multiply throughout by \(z^n\), then factorize the equation and take absolute values. The result is (A2).
\[ |z^n - a_1 z^{n-1} - \ldots - a_{n-1} z - a_n - 1| = |z - z_1||z - z_2|\ldots|z - z_{n-1}||z - z_n| \quad (A2) \]

Both sides of (A2) are equal to zero if \( z \) is set equal to one of the roots (the \( z_i \)). Instead of making this assumption, rather set \( z = 1 \). The left-hand side of (A2) becomes the sum of the coefficients while the right hand side becomes the product of the \( r_i \), since

\[ |r_i| = |1 - z_i| \]. This completes the proof.
Appendix B

A proof of the equivalence between the old and the new formulas for modified duration.

The fourth order version of equation (1) used in the text is written below as equation (B1).

\[ p = \frac{c}{(1+r)} + \frac{c}{(1+r)^2} + \frac{c}{(1+r)^3} + \frac{c}{(1+r)^4} + \frac{f}{(1+r)^4} \]  \hspace{1cm} (B1)

In the text, and in Appendix A, it was suggested that any TVM equation could be expressed in the form of equation (3). The rearranged version of (B1) is written below (3) to illustrate the point.

\[ -1 + \frac{a_1}{(1+r)} + \frac{a_2}{(1+r)^2} + \frac{a_3}{(1+r)^3} + \frac{a_4}{(1+r)^4} + \frac{1}{(1+r)^4} = 0 \]  \hspace{1cm} (3)

\[ -1 + \frac{c/p}{(1+r)} + \frac{c/p}{(1+r)^2} + \frac{c/p}{(1+r)^3} + \frac{(c+f)/p - 1}{(1+r)^4} + \frac{1}{(1+r)^4} = 0 \]

Next, equation (3) is written down \( n \) times in \( n \) rows to form a matrix, the sign of each element is reversed, the result is divided throughout by \((1+r)\) and the whole matrix is named (B2). In this particular example there are four rows.
\[
\begin{align*}
&+ \frac{1}{(1+r)^2} - \frac{a_1}{(1+r)^2} - \frac{a_2}{(1+r)^3} - \frac{a_3}{(1+r)^4} - \frac{a_4}{(1+r)^5} - \frac{1}{(1+r)^3} = 0 \\
&+ \frac{1}{(1+r)^2} - \frac{a_1}{(1+r)^2} - \frac{a_2}{(1+r)^3} - \frac{a_3}{(1+r)^4} - \frac{a_4}{(1+r)^5} - \frac{1}{(1+r)^3} = 0 \\
&+ \frac{1}{(1+r)^2} - \frac{a_1}{(1+r)^2} - \frac{a_2}{(1+r)^3} - \frac{a_3}{(1+r)^4} - \frac{a_4}{(1+r)^5} - \frac{1}{(1+r)^3} = 0 \\
&+ \frac{1}{(1+r)^2} - \frac{a_1}{(1+r)^2} - \frac{a_2}{(1+r)^3} - \frac{a_3}{(1+r)^4} - \frac{a_4}{(1+r)^5} - \frac{1}{(1+r)^3} = 0
\end{align*}
\] (B2)

The whole matrix sums to zero because each row sums to zero. Partition the matrix diagonally to include the whole of the first column, the top three elements of the second column, the top two elements of the third column and the top element of the fourth column. The partition to the top left is named \(P_1\). The partition to the lower right is named \(P_2\). The elements of the matrix sum to zero, therefore \(P_1 = -P_2\).

Consider the two partitions in turn. First \(P_2\) because it is easier and it has the more familiar interpretation. Write down the elements of \(P_2\) in the form of an equation, and slightly rearrange the final term. The result, equation (B3), is recognizable as modified duration. This could be for a par or a non-par bond, because the \(a_i\) have not been specified, they can take any values. To see this, compare (B3) with the formula for modified duration given in the text, equation (5), and note the meaning of the \(a_i\) from the substitution of coefficients earlier in this appendix (and the assumption that \(f = 1\)).

\[
- \frac{a_1}{(1+r)^2} - \frac{2.a_2}{(1+r)^3} - \frac{3.a_3}{(1+r)^4} - \frac{4.(a_4+1)}{(1+r)^5} = D
\] (B3)

It is known that \(P_1 = -P_2\), therefore \(P_1\) must also be equal to (the negative of) modified duration. The extra meaning of the partition \(P_1\) is not so obvious. It must be shown that it is
equal to the new formula for modified duration, equation (6) in the text, i.e., the product of
the distance between the \((n-1)\) roots and the real \(n\)th root, divided by \((1 + r)^4\).

The line of reasoning is that of Cotes' Theorem described in Nahin (1998). Take the
bond pricing equation (1) with \(n = 4\), reverse all the signs, multiply throughout by \((1 + r)^4\)
and set \((1 + r) = z\). This gives equation (B4).

\[
z^4 - a_1z^3 - a_2z^2 - a_3z - (a_4 + 1) = 0
\]  

(B4)

Equation (B4) can be factorized to give equation (B5).

\[
z^4 - a_1z^3 - a_2z^2 - a_3z - (a_4 + 1) = (z - z_1)(z - z_2)(z - z_3)(z - z_4)
\]  

(B5)

The absolute value of the product on the right hand side of (B5) is equal to the
product of the absolute values, i.e.,

\[
|z - z_1)(z - z_2)(z - z_3)(z - z_4)| = |z - z_1||z - z_2||z - z_3||z - z_4|
\]  

(B6)

The absolute value of the difference between two complex numbers, an entity like
\(|z - z_i|\), is the distance between the two points \(z\) and \(z_i\) in the plane. Equation (B6) is equal
to zero if the value of \(z\) is equal to one of the four roots. In this particular case, if
\(z = z_1 = (1 + r)\), i.e., \(z\) is set equal to the first positive real root, then the very first element on
the right hand side of the equation is equal to zero. However the other three elements of the
right hand side of (B6) are not equal to zero, and it is these three elements that are of interest.
If \( z = z_i \) then the product of these other three elements is the top element in the right hand side of equation (6), i.e., \( d_3d_3d_4 \). A way has to be found to evaluate the product on the right hand side of (B6) in isolation from the element that is zero.

Evaluating the product is done through the elimination of the element \((z - z_i)\) from equation (B5), by dividing (B5) throughout by \((z - z_i)\).

\[
z^3 + (z_i - a_i)z^2 + (z_i - a_i, z_1 - a_2)z + (z_i^3 - a_i, z_1^2 - a_2, z_i - a_3) = (z - z_2), (z - z_3), (z - z_4)
\]

Only after this division has been done is \( z \) set equal to \( z_i \), by replacing \( z \) by \( z_i \) throughout. Now \( z = z_i = (1 + r) \), therefore replace \( z_i \) on the left-hand side by \((1 + r)\), gather terms and divide the result throughout by \((1 + r)^4\). Finally we take absolute values on both sides. This gives expression (B7).

\[
\frac{4}{(1 + r)} - \frac{3a_i}{(1 + r)^3} - \frac{2a_2}{(1 + r)^2} - \frac{a_3}{(1 + r)^4} = \frac{|(z_i - z_2)| |(z_i - z_3)| |(z_i - z_4)|}{|z_i|^4} = \frac{d_2, d_3, d_4}{(1 + r)^4} \quad \text{(B7)}
\]

This is the required proof, because the left hand side of (B7) is the partition \( P_1 \) of the matrix, which is already known to be the negative of modified duration, and the right hand elements are the same as those of equation (6) in the text.

This result for modified duration can be generalized to the \( n \) period case.

\[
\frac{\prod_{i=2}^{n} |z_i - z_i|}{|z_i|^n} = \frac{\prod_{i=2}^{n} d_i}{(1 + r)^n} = -D
\]
Appendix C

The proof that equation (6a), the new expression for duration, gives precise results.

Once again, consider the equation for the price of a bond with \( f = 1 \) and \( n = 4 \).

\[
p = \frac{c}{(1 + r)} + \frac{c}{(1 + r)^2} + \frac{c}{(1 + r)^3} + \frac{c + 1}{(1 + r)^4} \quad (C1)
\]

If the interest rate changes from \( r \) to \( r' \), then \( p \) becomes \( p' \). Equation (C2) is the result.

\[
p' = \frac{c}{(1 + r')} + \frac{c}{(1 + r')^2} + \frac{c}{(1 + r')^3} + \frac{c + 1}{(1 + r')^4} \quad (C2)
\]

Convert both equations to the form of the polynomial having the particular structure described in the text and Appendix A. Call the two versions (C3) and (C4).

\[
-1 + \frac{c/p}{(1 + r)} + \frac{c/p}{(1 + r)^2} + \frac{c/p}{(1 + r)^3} + \frac{(c + 1)/p - 1}{(1 + r)^4} + \frac{1}{(1 + r)^4} = 0 \quad (C3)
\]

\[
-1 + \frac{c/p'}{(1 + r')} + \frac{c/p'}{(1 + r')^2} + \frac{c/p'}{(1 + r')^3} + \frac{(c + 1)/p' - 1}{(1 + r')^4} + \frac{1}{(1 + r')^4} = 0 \quad (C4)
\]

Keep these equations in mind. Now jump to equation (B5) from Appendix B, the general version of the polynomial with \( n = 4 \) and \( z = (1 + r) \). It is reproduced below for
convenience. Note that the $a_i$ on the left-hand side of (B5) correspond to the middle four parameters on the top line of (C3).

$$z^4 - a_1 z^3 - a_2 z^2 - a_3 z - (a_4 + 1) = (z - z_1)(z - z_2)(z - z_3)(z - z_4)$$

(B5)

In Appendix B it is pointed out that $z$ can take any value. If it happens to take the value of any one of the roots, e.g., $z_i$, then both sides of the equation are equal to zero because the first element on the right hand side is zero. In this proof, let $z$ become, not the root $(1 + r) = z_1$, but the new root containing the new interest rate, $(1 + r) = z_i$. Substitute the left or right hand side of this last expression, as appropriate, into equation (B5), divide throughout by the same elements raised to the fourth power, and take absolute values on both sides. This gives (C5).

$$\left| -1 + \frac{a_1}{(1 + r)} + \frac{a_2}{(1 + r)^2} + \frac{a_3}{(1 + r)^3} + \frac{a_4}{(1 + r)^4} + \frac{1}{(1 + r^i)^4} \right| = \prod_{i=1}^{i=n} \left| z_i - z_i \right|$$

(C5)

There are two things to notice about (C5). First, on the right hand side, the product on the top line has four elements. The first of the four elements is the change in the interest rate from $(1 + r)$ to $(1 + r)^i$, i.e., $|z_i - z_i| = \Delta r$. If this first element is taken outside of the product, then the remainder is the new expression for duration, $-D$. Therefore, the right hand side of (C5) is $- D \cdot \Delta r$. (Note that $D$ itself is intrinsically negative therefore $-D$ is positive.)

The second thing to notice about (C5) is the structure of the left-hand side. It is the same structure as the left-hand side of (C3) or (C4). Yet it cannot be equal to zero. This is because the $a_i$ are from equation (C3), i.e., they contain the original price, $p$, while the bottom
line contains the new interest rate \( r' \) from equation (C4) rather than \( r \). Thus, it is a mixture of the two equations. It is the old parameters being evaluated at the new interest rate. Whatever this value is, it is not zero, therefore it is set equal to \( X \).

\[
-1 + \frac{a_1}{(1+r')} + \frac{a_2}{(1+r')^2} + \frac{a_3}{(1+r')^3} + \frac{a_4}{(1+r')^4} + \frac{1}{(1+r')^4} = X
\]

or

\[
-1 + \frac{c/p}{(1+r')} + \frac{c/p}{(1+r')^2} + \frac{c/p}{(1+r')^3} + \frac{(c+1)/p-1}{(1+r')^4} + \frac{1}{(1+r')^4} = X
\]

Subtract \( X \) from both sides, divide throughout by \( (1+r') \), and rearrange to give:

\[
-1 + \frac{c/p(1+X)}{(1+r')} + \frac{c/p(1+X)}{(1+r')^2} + \frac{c/p(1+X)}{(1+r')^3} + \frac{(c+1)/p(1+X)-1}{(1+r')^4} + \frac{1}{(1+r')^4} = 0
\]

Compare the coefficients in this last result with those of equation (C4). It is clear that \( p' = p(1+X) \). This means that \( X \), known to be the left-hand side of (C5), is the change in the price of the bond relative to its original price, i.e.,

\[
X = \frac{p' - p}{p} = \frac{\Delta p}{p}.
\]

Put together the two results from both sides of (C5) to give equation (6a) in the text.
\[
\frac{\Delta p}{p} = -D\Delta r \quad \text{or} \quad \Delta p \cdot \frac{1}{p} = -D = \frac{\prod_{j=2}^{i=4} |z^*_i - z_i|}{|z^*_1|^4}
\] 

(6a)

Like the alternative expression for modified duration, this new formula for precise duration can be generalized to the n\textsuperscript{th} order (equation (6b) of the text).

\[
\text{precise } D_n = -\frac{\prod_{i=2}^{i=n} |z^*_i - z_i|}{|z^*_1|^n}
\] 

(6b)
References


Fig. 1. The roots of a two-period TVM equation, shown on the real number line.
Fig. 2. The roots of a three-period TVM equation, shown in the complex plane
Fig. 3. The interest rates of a three-period TVM equation, shown in the complex plane.
Fig. 4. The roots of a four-period TVM equation, plotted in the complex plane.
Fig. 5. Distances in the complex plane that are components of the orthodox formula for modified duration.
Fig. 6. Distances in the complex plane that are components of the new formula for precise duration.
Fig. 7. A depiction of the cash flows for two different bonds plus the irregular and uneven cash flow that results from combining them into a portfolio.
Fig. 8. The distances in the complex plane that are the interest rates of a zero-coupon bond.
Fig. 9. The distances in the complex plane that are components of the precise duration of a zero-coupon bond.
Table 1. The time to calculate the roots of a random polynomial

<table>
<thead>
<tr>
<th>Number of roots</th>
<th>Time (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.000</td>
</tr>
<tr>
<td>100</td>
<td>0.020</td>
</tr>
<tr>
<td>200</td>
<td>0.120</td>
</tr>
<tr>
<td>500</td>
<td>3.946</td>
</tr>
<tr>
<td>1000</td>
<td>29.903</td>
</tr>
<tr>
<td>2000</td>
<td>310.446</td>
</tr>
</tbody>
</table>
Exhibit 1. The Matlab commands to calculate duration for a two-bond portfolio.

```matlab
clear all
format long
p=[-1.9090409307 0.03 0 0.04 0.03 0 0.04 0.03 0 0.04 1.03 0 0.04 0 0 0.04 0 0 1.04]
r=roots(p)
r1=r([17 1:16 18])
a=1.055^(1/3)-1.05^(1/3)
newdur=(prod(abs((r1(1)+a)-r1(2:18)))/(r1(1)+a)^18)*a*100
```
Work 8
Osborne, M.
A resolution to the NPV-IRR debate?
A resolution to the NPV – IRR debate?

by

Michael J. Osborne

Middlesex University Business School

Key words
Capital budgeting, complex plane, internal rate of return, net present value.

JEL classifications
C00, C60, C63, E22, E40, G00, G1, G24, G30, G31, O16, O22

Abstract
Two criteria for choosing between capital investment projects are net present value (NPV) and internal rate of return (IRR). Sometimes they provide inconsistent rankings. This inconsistency sparked a debate about which criterion is better. The debate has lasted more than 100 years.

This paper describes a new approach to the debate. The time value of money equation is a polynomial, and a polynomial of order \( n \) does not have a single root. It has \( n \) roots. The result of taking into account the \( n \) solutions for IRR is a new equation for NPV that suggests a resolution to the debate.

*1 email: mo613@live.mdx.ac.uk
Middlesex University Business School, The Burroughs, Hendon, London, NW4 4BT

*2 Acknowledgements
Thanks for advice in the early stages of the research go to Hani Redha, Jaffar Hussain, Duncan Larraz and Jay Marchand. Thanks also go to Peter Carr, Head of Quantitative Financial Research at Bloomberg LP, for his thoughts and permission to quote from some correspondence, Chris Deeley of Charles Sturt University for advice on both style and substance, and Ephraim Clark and Yacine Belghitar of Middlesex University for their comments and encouragement. Finally, Carter Daniel of Rutgers University is owed a special thank you for his detailed advice on style and presentation.
A resolution to the NPV – IRR debate?

1 Introduction

Two of the most important criteria for choosing among investment projects are net present value (NPV) and internal rate of return (IRR). In many circumstances investment projects are ranked in the same order by both criteria. In some situations, however, the two criteria provide different rankings. A difference between rankings implies inconsistent recommendations about ‘best project’. This inconsistency gives rise to a debate in the literature about which criterion is superior. The debate is an old one. Dorfman (1981) traces it back to Irving Fisher (1907) and Boehm-Bawerk (1889).

In the 1950s Lorie and Savage (1955) and Hirshleifer (1958) exposed a number of deficiencies in the internal rate of return. This work shifted academic opinion in favor of NPV. The arguments were powerful because, in the 1980s, Jensen and Smith (1984) summarize the capital budgeting literature of the time and conclude that ‘analysis of the firm’s investment decisions has been well understood for so long that now the best discussions are in textbooks’. They then refer the reader to Brealey and Myer’s 1981 edition of Principles of Corporate Finance, a notable textbook of the time.

A half-century has passed since the work of Lorie and Savage and Hirshleifer. The understanding described by Jensen and Smith in 1984 remains unchanged. Brealey et al. (2009) remains one of the best graduate textbooks on corporate finance and the authors continue to argue for the superiority of NPV partly on the basis of deficiencies in IRR, now called pitfalls.
One reason why debate persists is that financial practice has not caught up with theory. Large corporations and banks use both criteria, but often prefer IRR, perhaps because of the ease of comparison with the cost of capital. Surveys from the last decade include Payne et al. (1999) and Kester et al. (1999), who report on practices that vary from country to country; sometimes IRR is preferred, sometimes NPV. Ryan and Ryan (2002) report on studies between 1960 and 1996 that show IRR dominated company practice during that period while their own study of Fortune 1000 companies shows NPV is narrowly preferred to IRR. Surveys by Graham and Harvey (2001) and Liljeblom et al. (2004) show NPV used more than IRR. Brounen et al. (2004) apply Graham and Harvey’s methodology to European companies and show NPV is used more in Germany and the Netherlands but IRR is used more in the UK and France.¹

Brealey et al. (2009) comment that ‘for many companies DCF means IRR, not NPV’. They may echo the opinion of many when they write ‘we find this puzzling’.

An answer to the puzzle may lie in our understanding of the two concepts. The argument in this article is that the ‘50-year understanding’ is not entirely correct. NPV is superior, but not for the reasons traditionally offered. Two of the IRR pitfalls are analysed. The first pitfall, that IRR yields multiple solutions, is shown not to be a pitfall. The multiple solutions are useful because they can be used to solve the second pitfall, the problem of inconsistent rankings of mutually exclusive projects. A new equation for NPV is derived containing all possible solutions for IRR as components. This multiple-

¹ In Europe, the payback criterion dominated NPV and IRR in all countries. That puzzle is not discussed here.
IRR criterion is identical to NPV. The traditional approach to IRR focuses on only one rate, thereby ignoring information from the cash flows conveyed by the other, unorthodox rates of interest.

A positive feature of the new approach is that the proposed resolution to the debate is amicable. While the analysis supports the academic preference for NPV, it also shows that the IRR remains a fundamental concept. This is because the margins of all possible IRRs over the cost of capital are the only components of NPV.

2 Not one interest rate, but many

The concepts of NPV and IRR are derived from the time value of money equation (TVM). Equation (1) shows one manifestation of a TVM equation.

\[
NPV = -I_0 + \sum_{i=1}^{n} \frac{c_i}{(1 + r)^i}
\]  

(1)

In equation (1), \(NPV\) is net present value. \(I_0\) is the initial investment in period zero (a positive number shown with a minus sign to indicate that money is paid away). The \(c_i\) represent the cash flows in the periods \(i = 1\) to \(n\) (usually positive to reflect money received, but some values could be negative to indicate additional investment). Finally, \(r\) is the cost of capital and \(n\) is the number of periods.

The internal rate of return is the rate that brings the present value of the returns into equality with the initial investment, i.e., NPV is zero. The rate is shown as \(R\) in the variant of equation (1) that is equation (1a).
\[ 0 = -I_0 + \sum_{i=1}^{n} \frac{c_i}{(1 + R)^i} \]  \hspace{1cm} (1a)

An analysis of the relationship between NPV and IRR would, ideally, be accomplished by combining (1) and (1a) into a single equation containing both concepts, i.e. NPV from equation (1) is expressed in terms of R from equation (1a), or vice versa. Unfortunately, this combination is difficult because the equations are polynomials, and polynomials possess two features that prevent the necessary rearrangements.

First, when the number of periods, \(n\), is greater than four, it is impossible to solve algebraically for \((1+R)\). In this circumstance, NPV cannot be expressed algebraically in terms of IRR, or vice versa.

Secondly, there are many values of \((1+R)\) that satisfy equation (1a). In general, where there are \(n\) time periods, there are \(n\) IRRs. For this reason, IRR is better expressed as \(R_i\) from \(i = 1\) to \(n\) than just \(R\). Real, internal rates of return other than the orthodox rate are either ignored in the literature, or, if addressed, are seen as a problem (see Teichroew et al. (1965) for an early example, or Brealey et al. (2009) for a more recent one). The complex solutions are usually ignored, an exception being Dorfman (1981). Like Dorfman (1981), this study uses all rates, real and complex, positive and negative.  

\[ \text{Dorfman examines the case where the proceeds of an investment are serially reinvested in projects of the same type, and the process is continued indefinitely. He shows that the growth path of a dollar placed in such an investment depends on all the roots of the internal rate of return equation, complex as well as real. Dorfman’s mode of analysis is different from that adopted here therefore the relationship between the two} \]
In financial literature, any mention of multiple solutions for the interest rate usually refers to the real positive solutions. Such solutions occur when the cash flow changes sign more than once. For example, the adjective ‘non-normal’ is used by Brigham et al. (2004) to describe a project that ‘calls for a large cash outflow either sometime during or at the end of its life’. In the context of ‘non-normal’ cash flows, Descartes' sign rule is often quoted.\footnote{For a statement of Descartes' sign rule see Weisstein (1999).} That is, the number of changes of sign in the coefficients (cash flows) corresponds to the maximum number of real, positive roots.\footnote{Note the distinction between the roots -- values of $\left(1+R\right)$ -- and the values of $R$ that they imply.}

Descartes' rule also states that if the signs are reversed on all the coefficients attached to odd powers, then the number of changes of sign in the coefficients (cash flows) corresponds to the maximum number of real, negative roots. A negative root implies the existence of an interest rate that is less than minus 100%. Most financial calculators and spreadsheets will not calculate such interest rates.

analyses is left for future research. Suffice it to say that Dorfman’s analysis uses all rates as ingredients in a formula for a financial concept (a growth rate) as does the analysis in this paper (NPV). In contrast, while Hazen (2003) and Hartman and Schafrick (2004) make explicit mention of the complex rates, they do so as rates \textit{per se} in the context of the multiple-IRR literature rather than as ingredients in a financial formula. The recent multiple-IRR literature is summarized in Zhang (2005); the relationship between the analysis in this paper and the analyses in that literature is also left for future research.
In this article, the term ‘multiple solutions’ is used in its widest possible sense, because it refers to all \( n \) solutions for the rate of interest possessed by any \( n \) period TVM equation. They can be real (positive and negative), and they can be complex. The complex solutions for \((1+R)\) are of the form \( x \pm yi \) where \( i = \sqrt{-1} \). Whether the cash flow is 'well-behaved' or ‘non-normal’, there are always \( n \) solutions, most of them are complex, and they are all employed in the current analysis.\(^5\)

3 **A numerical example**

This section sets up simple, numerical, examples to illustrate inconsistent ranking. The following section uses these examples to show the location of the multiple IRRs. Table 1 shows cash flows for two projects, A and B. The NPV and IRR for each project are shown in Table 2.

[ Table 1 about here ]

[ Table 2 about here ]

The relationship between NPV and a range of interest rates is shown in Fig. 1. The two IRRs, when NPV is zero, are shown on the x-axis. NPV, measured on the y-axis, varies with the cost of capital.

---

\(^5\) If there are \( m \) real solutions (positive and negative) then the remaining \((n-m)\) solutions will be complex. There will always be an even number of complex solutions, because they come in complex, conjugate pairs whenever the coefficients are real (see Hughes et al. (2004)), and they are always real in a financial equation.
Fig. 1 illustrates the inconsistency of the two criteria when appraising projects A and B. Project B has the higher IRR. If the cost of capital is above \( r_e \) (the rate of interest where the NPVs for the two projects are equal), then project B also has the higher NPV. In this situation, the two criteria give the same rank order. If the cost of capital is below \( r_e \) then the NPV for project A exceeds that for B. In this case, the criteria give rank orders that disagree. A cost of capital of 5% is used to calculate the NPVs in Table 2 and the vertical line identifies this situation in Fig. 1.

It is possible to look at the problem in a different way by exposing all the IRRs for the two projects. This is done in the next section.

4 Many internal rates of return

As mentioned earlier, in most circumstances the TVM equation (1a) cannot be solved algebraically for the internal rate of return, \( R \). But it can be done numerically. Given suitable software, numerical solutions can be found for all rates, including the unorthodox rates often dismissed in the literature.\(^6\) Once calculated, they can be labeled

\(^6\) Some scientific hand calculators are capable of solving TVM equations for all the IRRs within a reasonable period of time when the number of time periods is small, e.g., when \( n \) is less than 20. For higher values of \( n \), more advanced mathematical software (Matlab, Mathematica, Maple, etc) and the faster processor of a PC are needed. Several hundred solutions can be calculated within seconds, and several thousands within minutes.
and incorporated into a meaningful equation that displays the sought-after link between NPV and IRR, and facilitates discrimination between the two criteria.

In this section, all the numerical solutions are calculated for projects A and B. The four possible rates of interest for each of the two projects are in column two of Table 3. They are the $Z_i/ = (1 + R_i/)$, for $i = 1$ to 4 and $j = Projects A or B$. These values are depicted in Fig. 2. The figure is an Argand diagram that shows the complex plane with the real number line on the x-axis and units of $\sqrt{-1}$ on the y-axis. The unit circle is shown to provide scale.

[ Table 3 about here ]

The third column of Table 3 shows absolute values for the rates of interest, $|R_i/|$, that are implied by the values for $Z_i/ = (1 + R_i/)$. When the analysis is confined to the real number line the difference between $(1+R)$ and 1 is simply $R$. When the analysis is extended to the complex plane the absolute values of the differences between the points $Z_i/ = (1 + R_i/)$ and the point $(1,0)$ are the values of $|R_i/|$.

[ Fig. 2 about here ]

The first number in column three for each project is the orthodox value of the IRR for that project. For example, $|R_1/| = 0.219861 = 21.99\%$ is the orthodox IRR of project A. The orthodox value is a distance that lies on the real number line to the right of the
point \((1,0)\). The remaining values in the final column (and distances in Fig. 2), being the absolute values of the many IRRs other than the orthodox rate, are not so familiar. These are real numbers because they are absolute values. In Fig. 2, they are distances that extend away from the point \((1,0)\), either into complex space or leftward along the number line into negative territory.

At this point in the analysis, the meaning and utility of these unfamiliar interest rates remain open questions, as does their part in exposing the link between NPV and IRR. Before trying to answer these questions, an interim result is described.

5 An interim result

Any polynomial can be rearranged into the form of equation (2) below, called here the ‘special form’. The special form is characterized by the value -1 at the beginning, the value 1 discounted over \(n\) periods at the rate \(z\) at the far end, and a range of discounted values, \(b_i\), between the two. The whole is set equal to zero.

\[
-1 + \sum_{i=1}^{n} \frac{b_i}{(1 + z)^i} + \frac{1}{(1 + z)^n} = 0
\]  

(2)

Many different arrays of values can exist for the parameters \(b_i\). In a financial equation these parameters are composed of cash flows and, as such, are always real (positive or negative). For any particular array of parameters, there are \(n\) possible solutions for \(z\). A special relationship (3) exists between a given array of parameters and all possible solutions for \(z\) associated with that array.
Equation (3) shows that the absolute value of the sum of the parameters in the special form (2) is equal to the product of the absolute interest rates. This result is proved in Appendix A and it is a key element in what follows.

6 A new equation for NPV

The new equation that links NPV and IRR is developed using the interim result from the previous section.

First, I derive the simplest possible case and then generalize it. Earlier in this paper it was stated that equations (1) and (1a) are not easily combined and manipulated into an algebraic expression linking NPV and the IRR. When the order of the polynomial is greater than four it is impossible. When the order is two, three or four the level of difficulty is high. In the simplest case, however, when \( n = 1 \), the combination can be done. From \( NPV = -I_0 + \frac{c}{1+r} \) and \( 0 = -I_0 + \frac{c}{1+R} \) it is a simple deduction that

\[
\frac{NPV}{I_0} = \frac{(R - r)}{(1 + r)}.
\]  

The relationship between \( r \), the cost of capital, and \( R \), the internal rate of return, can be shown in two different ways – an additive relationship or a multiplicative relationship.
The additive relationship is \((1+R) = (1+r+a)\) in which \(a\) is the difference between the cost of capital and the internal rate of return. It follows that \((R-r) = a\) and therefore

\[
\frac{NPV}{I_0} = \frac{a}{(1+r)}.
\] (4a)

The multiplicative relationship is \((1+R) = (1+r)(1+m)\) in which \(m\) is the interest rate that marks up the cost of capital to the internal rate of return. From this we deduce that \(\frac{a}{(1+r)} = m\). Equation (5) follows.

\[
\frac{NPV}{I_0} = m \quad \text{or} \quad NPV = m \cdot I_0.
\] (5)

Equation (5) is the simplest version of the new equation for NPV. It states that in the case of a one-period cash flow, the ratio of NPV to the initial investment is simply the multiplicative mark-up of the single internal rate of return on the cost of capital.

The most important step in the argument is that the last result generalizes. Given \(n\) periods in the TVM polynomial there are \(n\) solutions for IRR in equation (1a). Assuming a single cost of capital, there must be \(n\) multiplicative mark-ups of these IRRs on the cost of capital. The new generalized equation for NPV is composed of all the multiplicative mark-ups.

\[
\left|\frac{NPV}{I_0}\right| = \prod_{1}^{n}|m_i| \quad \text{or} \quad |NPV| = \prod_{1}^{n}|m_i| \cdot I_0.
\] (6)
Equation (6) shows that NPV is a multiple of the initial investment. The multiple is no longer the lone multiplicative mark-up of the single IRR on the cost of capital, as it was in the simplest case. In the many-period case the multiple is the product of all multiplicative mark-ups of every possible IRR over the cost of capital. This result is not obvious. It shows that, far from multiple IRRs being a problem, every IRR is equally important, because net present value is composed of them all.

The proof of equation (6) is in Appendix B. The proof uses the interim result from the special form of the TVM polynomial.

The new equation (6) above is written in the multiplicative form. It can be restated using additive mark-ups. It is useful to do this because the additive mark-ups connect more easily than the multiplicative with the visual presentation in Fig. 3.

\[
\frac{NPV}{I_0} = \left(\prod_{i=1}^{n} |a_i| \right) \frac{(1 + r)^n}{(1 + r)^n} \tag{6a}
\]

In the four period examples for Projects A and B the equation that applies is:

\[
\frac{NPV}{I_0} = \frac{a_1 a_2 a_3 a_4}{(1 + r)^4} = \frac{\prod_{i=1}^{4} |a_i|}{(1 + r)^4}
\]
The differences, $a_i$, in the numerator of this equation are depicted in Fig. 3. One set is shown for Project A and another for Project B.

[Fig. 3 about here]

The critical difference between Fig. 2 and Fig. 3 is that the locus of the lines is no longer $(1, 0)$ but instead $(1+r, 0)$. The lines radiating from their new locus have been re-labelled the $a_i^j$ instead of $R_i^j$ to indicate their change of meaning, from interest rates to the differences between interest rates. Each $a_i$ is a difference between the $i^{th}$ IRR and the cost of capital.

An alternative view of the additive approach is as follows. Equation (1a) is restated to show that NPV can be reduced to zero by means of the additive shift, $a$, to the interest rate, i.e., $R = r + a$.

\[ 0 = -I_0 + \sum_{i=1}^{n} \frac{C_i}{(1 + r + a)^i} \quad (1b) \]

Given a single value for the cost of capital, $r$, there will be $n$ possible values of the spread, $a$, that satisfy equation (1b). These $n$ values appear in the numerator of the additive version of the new equation for NPV.

The continuous lines in Fig. 3 represent the distances between all the IRRs for project A (the solid circles) and the cost of capital, $r$ (the grey circle). The product of
these distances, suitably discounted at the cost of capital, is the NPV per dollar for project A. The broken lines and hollow circles act similarly for project B.

The measured distances (differences between all IRRs and the cost of capital) for the two projects are shown in Table 4. When fed into the new equation (6a) they produce the NPVs per dollar shown in Table 2. The accuracy of the results is determined by the precision with which the roots are located. Modern mathematical software can, within seconds, locate thousands of roots to an accuracy of $1 in a trillion.

[ Table 4 about here ]

Using Fig. 3, it is possible to visualize what happens to the product of the set of rays for each of the two projects while the cost of capital, \( r \), moves forwards and backwards along the real number line, above or below \( r_e \). That NPV is the preferred measure of return of an investment becomes apparent. At a cost of capital above \( r_e \), project B is preferred because it has the largest product of the differences between its internal rates of return and the cost of capital. Alternatively, at a cost of capital below \( r_e \), project A is preferred because it now has the largest product of the differences between its internal rates of return and the cost of capital. The new 'multiple IRR' criterion and the ‘NPV per dollar invested’ criterion are identical, and always provide consistent rankings.

The information set and procedure necessary to calculate NPV using the new approach can be compared with those required by the orthodox approach.
The orthodox approach requires the set \( \{I_0, c_i, r\} \) as input to equation (1). The calculation is elementary and can be done on any spreadsheet or hand-calculator.

The new approach requires the information set \( \{I_0, c_i\} \) as input to equation (1a) in order to calculate all values of the internal rate of return. The cost of capital \( r \) is additional input to calculate the distances between it and all values of the IRR. Therefore the complete set of information required for the new approach is \( \{I_0, c_i, r\} \), the same as for the orthodox approach. A difference between the two approaches is that the new calculation is more complicated than the orthodox. Sophisticated software is required to calculate the complete set of roots for equation (1a).

It follows that the value of the new approach is not in the practical calculation of NPV, but rather in the light it sheds on the relationship between NPV and IRR, its elucidation of why NPV is a more useful tool for investment appraisal, and its demonstration of the underlying importance of IRR as a component.

7 A question for future work

The analysis described here makes use of all possible internal rates of return, negative as well as positive, complex as well as real. What financial meaning can be attributed to interest rates other than the orthodox? The question can be approached in three ways.

The first is to focus on the solutions for \( (1+R_i) \) from equation (1a). They are employed to calculate the differences between interest rates, i.e.
\[|1 + R_i| = |R_i - r| = |a_i|.\] The elements \((1 + R_i)\) are of the form \(x \pm yi\) where \(i = \sqrt{-1}.\) They are points (not distances) in the complex plane. What can be said about these complex solutions, especially their imaginary components? The following thought experiment is due to Peter Carr.

‘Suppose an investor invests a dollar at year zero and it turns into four dollars at the end of two years. The gross return over the two years is four. The annualized gross return is two since \(1 \times 2 \times 2 = 4.\) Now suppose the investor invests a dollar at year zero and it turns into negative one dollar at the end of two years. In other words, the investor owes another dollar. This can’t happen with a limited liability asset but it can happen with a forward contract or a leveraged stock position. What is the annualized gross return for our unfortunate investor? It has to be the imaginary numbers \(i\) and \(-i\) since \(1 \times i \times i = -1\) and \(1 \times (-i) \times (-i) = -1.\) Complex interest rates allow us to annualize gross returns even if they are negative. ... Thus, I can be interested directly in complex interest rates.’

Secondly, what can be said about the absolute values of the differences between interest rates, the \(|a_i|\)? They are distances (not points) in the complex plane, therefore they are real numbers and do not contain imaginary components. One of the elements in the product on the right hand side of equation (6a), \(|a_i|\), is the difference between the orthodox IRR and the cost of capital; as such, it has a clear financial meaning. On the left hand side of equation (6a) is the net present value of the investment per dollar invested; the meaning of this is also clear. Between these two meaningful elements lie the \((n-l)\)

---

7 Note the distinction between the use of \(i\) as the imaginary component of a complex number and the customary use of \(i\) as a counter, as in \(R_i\) from \(i = 1\) to \(n.\)
unorthodox rates. The cluster of unorthodox rates connects these two meaningful objects therefore it is hard to dismiss the rates as meaningless. What is their meaning as a cluster or when studied individually? The question demands further work because there is no clear answer.

Finally, we need not question the meaning of the unorthodox rates. The rates can be seen as catalysts. This interpretation is in the spirit of the following quote by the mathematician, Hadamard: ‘The shortest path between two truths in the real domain passes through the complex domain’. The diversion into unfamiliar territory is justified by the novel link it exposes between NPV and IRR; that the one is a composite of the many values of the other. The end justifies the means.

8 Concluding remarks

The single, orthodox IRR criterion uses only part of the information to be gleaned from the TVM equation for an investment project. It employs one difference between interest rates, that between the orthodox IRR and the cost of capital. In contrast, NPV uses all the differences between every possible IRR for a project and its cost of capital; therefore NPV is a richer concept than the orthodox IRR alone. The objective of the paper has been to expose this hitherto hidden richness. The argument in favor of NPV is buttressed, and IRR loses two pitfalls and is raised in status to fundamental component.

Finally, the more detailed picture of the TVM equation provided by the extra dimension of the complex plane is illuminating in the context of capital budgeting. The

---

8 Quoted, with permission, from correspondence with the author.
TVM equation is ubiquitous in finance. It follows that much territory remains to be explored from the new perspective. One such study exists already in the context of risk management in fixed income markets. Osborne (2005) uses similar analysis to produce an accurate version of duration without the need for convexity. Other applications must exist.

\(^9\) Quoted in Nahin (1998)
Appendix A

Proof of a relationship between the coefficients and roots of the special form of the TVM equation

In the text it is stated that any TVM equation can be rearranged into a special form, which is repeated below for convenience.

\[-1 + \sum_{i=1}^{n} \frac{b_i}{(1+z)^i} + \frac{1}{(1+z)^n} = 0\]  

(2)

It is asserted that the parameters and the roots of the special form are linked in a particular way: the absolute value of the sum of the parameters is equal to the product of the absolute values of the interest rates.

\[\left|\sum b_i\right| = \prod|z_i|\]  

(3)

A proof of the statement is as follows. First, in the special form, set \(Z = (1+z)\) and multiply throughout by \(Z^n\), then factorize the equation and take absolute values. The result is:

\[|Z^n - b_1 Z^{n-1} - \ldots - b_n Z - b_n - 1| = |Z - Z_1||Z - Z_2|\ldots|Z - Z_{n-1}|Z - Z_n|\]

Now set \(Z = 1\). The left-hand side of becomes the absolute sum of the parameters, while the right hand side becomes the product of the absolute values of \(z_i\), since \(Z_i = (1+z_i)\) therefore \(|Z_i| = |1 - Z_i|\). The proof is now complete.
Appendix B

A derivation of the new equation for NPV

Equation (1a) and the equation showing the multiplicative method of incrementing from
the cost of capital to the internal rate of return are repeated below.

\[ 0 = -I_0 + \frac{c_1}{(1 + R)} + \frac{c_2}{(1 + R)^2} + ... + \frac{c_n}{(1 + R)^n} \]  \hspace{1cm} (1a)

\[ (1 + R) = (1 + r)(1 + m) \]

These equations are combined and manipulated into the special form.

\[ 0 = -1 + \frac{\left( \frac{c_1}{I_0} \right)}{1 + r} + \frac{\left( \frac{c_2}{I_0} \right)}{(1 + r)^2} + ... + \frac{\left( \frac{c_n}{I_0} \right)}{(1 + r)^n} - 1 + \frac{1}{(1 + m)^n} \]

The result (3) about the special form (2) is applied to the last equation.

\[ \left| \frac{\left( \frac{c_1}{I_0} \right)}{1 + r} + \frac{\left( \frac{c_2}{I_0} \right)}{(1 + r)^2} + ... + \frac{\left( \frac{c_n}{I_0} \right)}{(1 + r)^n} - 1 \right| = \prod_{i}^{n} |m_i| \]

The expression for NPV, equation (1) from the text, is rearranged as follows:

\[ \frac{NPV}{I_0} = \frac{\left( \frac{c_1}{I_0} \right)}{1 + r} + \frac{\left( \frac{c_2}{I_0} \right)}{(1 + r)^2} + ... + \frac{\left( \frac{c_n}{I_0} \right)}{(1 + r)^n} - 1 \]

The new expression for NPV, equation (6), is a combination of the last two equations.

\[ \left| \frac{NPV}{I_0} \right| = \prod_{i}^{n} |m_i| \]  \hspace{1cm} (6)
References


Fisher, I. (1907). The Rate of Interest, New York: Macmillan


Table 1. The cash flows for two investment projects, A and B.

<table>
<thead>
<tr>
<th></th>
<th>I₀</th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>C₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project A</td>
<td>-100</td>
<td>30</td>
<td>35</td>
<td>45</td>
<td>60</td>
</tr>
<tr>
<td>Project B</td>
<td>-100</td>
<td>50</td>
<td>50</td>
<td>40</td>
<td>20</td>
</tr>
</tbody>
</table>
Table 2. The NPV and IRR for projects A and B.

<table>
<thead>
<tr>
<th></th>
<th>NPV</th>
<th>IRR %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project A</td>
<td>48.5523</td>
<td>21.9861</td>
</tr>
<tr>
<td>Project B</td>
<td>43.9781</td>
<td>25.4263</td>
</tr>
</tbody>
</table>

* NPV calculated on the assumption of a cost of capital of 5%.
Table 3. The multiple roots and interest rates for projects A and B

<table>
<thead>
<tr>
<th>Project A</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>$Z_i^A = (1 + R_i^A)$</td>
<td>Implied values of $</td>
</tr>
<tr>
<td>1</td>
<td>1.219861</td>
<td>0.219861</td>
</tr>
<tr>
<td>2</td>
<td>-0.081110 + 0.801636 $\sqrt{-1}$</td>
<td>1.345890</td>
</tr>
<tr>
<td>3</td>
<td>-0.757641</td>
<td>1.757641</td>
</tr>
<tr>
<td>4</td>
<td>-0.081110 - 0.801636 $\sqrt{-1}$</td>
<td>1.345890</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Project B</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$I$</td>
<td>$Z_i^B = (1 + R_i^B)$</td>
<td>Implied values of $</td>
</tr>
<tr>
<td>1</td>
<td>1.254263</td>
<td>0.254263</td>
</tr>
<tr>
<td>2</td>
<td>-0.127131 + 0.55022 $\sqrt{-1}$</td>
<td>1.254263</td>
</tr>
<tr>
<td>3</td>
<td>-0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>-0.127131 - 0.55022 $\sqrt{-1}$</td>
<td>1.254263</td>
</tr>
</tbody>
</table>
Table 4. The differences between all the IRRs and the 5% cost of capital for projects A and B.

<table>
<thead>
<tr>
<th></th>
<th>Project A</th>
<th>Project B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1 =</td>
<td>R_1 - r</td>
<td>$</td>
</tr>
<tr>
<td>$a_2 =</td>
<td>R_2 - r</td>
<td>$</td>
</tr>
<tr>
<td>$a_3 =</td>
<td>R_3 - r</td>
<td>$</td>
</tr>
<tr>
<td>$a_4 =</td>
<td>R_4 - r</td>
<td>$</td>
</tr>
</tbody>
</table>
Fig. 1 NPV and the orthodox IRR agree at high rates of interest but disagree at low rates.
Fig. 2. All the IRRs for the two projects, A and B, are rays from the point $(1,0)$. They are the many solutions for $R$ in equation (1a).

All the $Z_i^j$ are points (roots) for project $j$, $j = A$ or $B$
All the $R_i^j$ are distances between the point $(1,0)$ and the $Z_i^j$. 
Fig. 3. The differences between all possible IRRs and the cost of capital are rays from the point \((1+r, 0)\). They are the many solutions for \(a\) in equation (1b).

All the \(Z_i\) are points (roots) for project \(j\), \(j = A\) or \(B\).
All the \(a_i\) are distances between the point \((1+r, 0)\) and the \(Z_i\).
Work 9

Osborne, M.

The Cambridge controversies in the theory of capital: revisiting the reswitching puzzle

The Cambridge Controversies in the Theory of Capital:  
Revisiting the Reswitching Puzzle  
Michael J. Osborne *


Key words  
Capital, complex plane, interest rate, reswitching

JEL classifications  
B1, B2, B5, C6, E1, E2, E4

Abstract  
A solution is proposed to the reswitching puzzle.

When two techniques of production are compared reswitching can occur.  
Reswitching is when a technique begins by being cheapest at a low interest rate,  
switches to being more expensive at a higher rate, and then reswitches to being  
cheapest at yet higher rates. Some believe the inconsistency undermines the  
foundations of neoclassical economics.

The time value of money (TVM) equation is at the core of the reswitching puzzle.  
The equation takes the form of an n$^{th}$ order polynomial having n roots (interest rates).  
In most economic and financial analyses only one root is used. The remaining (n-1)  
roots, mostly complex or negative, are often ignored.

The approach in this article employs all n solutions for the interest rate in a new  
expression relating differences in the dependent variable to differences in interest  
rates. The analysis is applied to the Sraffa-Pasinetti example of reswitching. The new  
expression provides a different perspective on the TVM equation. From the new  
perspective, reswitching does not occur.

The ‘multiple interest rate’ approach to the TVM equation provides insights into  
issues other than reswitching. The issues include the NPV versus IRR debate in  
capital budgeting and the quest for an accurate equation for duration in bond  
mathematics. Reswitching is one problem in a class of similar problems having the  
TVM equation at their core. The TVM equation is ubiquitous in economics and  
finance therefore it is likely that the ‘multiple interest rate’ approach will find other  
applications.

* email: m.osborne@shu.ac.uk  
Sheffield Business School, Sheffield Hallam University,  
Howard Street, Sheffield, S1 1WB, UK
The Cambridge Controversies in the Theory of Capital: Revisiting the Reswitching Puzzle

1 Introduction

This article proposes a solution to a puzzle in economic theory: reswitching. The solution is provided by ‘multiple interest rate’ analysis. The analysis has implications for other areas of economics and finance.

The reswitching puzzle is part of the Cambridge controversies in capital theory. The controversies surfaced at the beginning of the twentieth century, intensified into the ‘Cambridge Controversies’ during the 1960s, and have simmered since. A high point of the debate is the symposium on reswitching in the Quarterly Journal of Economics (QJE) in 1966 containing six articles on the topic: Bruno et al. (1966); Garegnani (1966); Levhari and Samuelson (1966); Morishima (1966); Pasinetti (1966); and Samuelson (1966). A comprehensive survey of the controversies is in Harcourt (1972). Cohen and Harcourt (2003) is a recent review.

When two techniques of production are compared, reswitching is the possibility that one technique can be cheapest at a low interest rate, switch to being more expensive at a higher rate, and reswitch to being cheapest at even higher rates. For some, this inconsistency undermines the foundations of neoclassical economic theory.

Samuelson (1966) expressed his concern about reswitching thus:

‘The phenomenon of switching back at a very low interest rate to a set of techniques that had seemed viable only at a very high interest rate involves more than esoteric technicalities. It shows the simple tale told by Jevons, Bohm-Bawerk, Wicksell, and other neoclassical writers ... cannot be universally valid.’

Nearly forty years later, Cohen and Harcourt (2003) agree that reswitching causes problems for neoclassical economics.

---

1 Paradoxes in Capital Theory: A Symposium, Quarterly Journal of Economics, 80(4), Nov. 1966
'Looking back over this intellectual history, Solow (1963, p.10) suggested that “when a theoretical question remains debatable after 80 years there is a presumption that the question is badly posed – or very deep indeed.” Solow defended the “badly posed” answer, but we believe that the questions at issue in the recurring capital controversies are ‘very deep indeed.’

This article contains a new approach to the puzzle. The analysis of reswitching employs the time value of money (TVM) equation. The TVM equation is a key equation in economics and finance. It takes the form of an \( n \)th order polynomial having \( n \) roots (interest rates). In most economic and financial analyses, including the reswitching debate, it is usual to employ only one root, namely the root yielding a positive, real interest rate. The remaining \( (n-1) \) roots are mostly complex or negative, and are often ignored. When not ignored, the unorthodox roots are seen as a problem. One of the earliest examples of the latter is Lorie & Savage (1955); one of the most recent is Brealey et al. (2009).

In this article it is argued that all roots (interest rates), including the unorthodox, can be employed to shed light on the reswitching puzzle. Far from being a problem, multiple interest rates are part of the solution.\(^2\)

Section 2 describes a numerical example of reswitching taken from one of the contributions to the QJE symposium: Pasinetti (1966). The example is based on material in Sraffa (1960).\(^3\) Section 3 recalls a well-known result about the factorization of polynomials. In section 4 the result is applied to the Sraffa-Pasinetti

\(^2\) Harcourt (1972) notes the contribution of Bruno et al. (1966). They mention multiple roots in the context of the reswitching debate, as do Hagemann and Kurz (1976). However, their contributions refer only to the possibility of multiple real roots; they do not analyze all roots including complex roots.

\(^3\) Velupillai (1975) reports a ‘general consensus that the phenomenon of reswitching of techniques was first brought to the attention of Academic Economists by Joan Robinson, David Champernowne, and Piero Sraffa’ but goes on to note that Irving Fisher’s 1907 book, The Rate of Interest, pp. 352-353, contains an example of reswitching. The example is brief and contained in an appendix but it is unmistakably reswitching and, moreover, part of a critique of Bohm-Bawerk’s methodology. Velupillai acknowledges that Fisher did not draw out the implications of the phenomenon as Sraffa and his colleagues did; therefore the consensus remains.
example. A new expression is derived for differences in wage ratios resulting from differences in rates of interest (profit). The expression employs explicitly all possible interest rates, not just the orthodox. The new perspective provided by the expression shows reswitching is no longer a concern. Section 5 contains some general discussion about the analysis. The final section is the conclusion.

2 An example of reswitching: Sraffa-Pasinetti

The numerical example in Pasinetti (1966) is adapted from Sraffa (1960). Pasinetti’s objective is to refute Levhari and Samuelson (1966), who set out to demonstrate reswitching cannot happen. Samuelson (1966) subsequently accepts Pasinetti’s argument and admits the possibility of reswitching. Pasinetti’s refutation, however, is not the end of the story.

The full details of the Sraffa-Pasinetti model are not presented here; instead the focus is on the particular analysis that Pasinetti uses to demonstrate the existence of reswitching. He creates two economic systems, a and b, each of which possesses a relationship between the wage rate and the rate of interest:

\[
\begin{align*}
  w_a &= \frac{1 - 0.8(1 + r)}{20(1 + r)^8} \quad \text{(1)} \\
  w_b &= \frac{1 - 0.8(1 + r)}{(1 + r)^{25} + 24} \quad \text{(2)}
\end{align*}
\]

'Since the wage rate in 'a' and the wage rate in 'b' are expressed in terms of the same physical commodity ... the two technologies can now be compared. Clearly, on grounds of profitability, that technology will be chosen which – for any given wage rate – yields the higher rate of [interest]. Or alternatively (which comes to the same thing) that technology will be chosen which – for any given rate of [interest] – yields the higher wage rate.

\[\text{4 Both words 'profit' and 'interest' are used in this context in the reswitching literature. To avoid confusion and repetition, from this point onwards the word 'interest’ is employed.}\]
In order to find this out, it is sufficient to compute the values of \( w_a \) and \( w_b \), in expressions ... (1) and (2) ... for any given level of \( r \).’ (Pasinetti, 1966)

Pasinetti’s Fig. 1 (p. 507) has \( w_a \) and \( w_b \) on the vertical axis and \( r \) on the horizontal axis. It shows that ‘... the curves representing \( w_a \) and \( w_b \) intersect each other three times. There are three distinct levels of the rate of [interest], namely \( \approx 3.6 \) per cent, \( \approx 16.2 \) per cent, and 25 per cent, at which \( w_a = w_b \), i.e., at which the two technologies are equally profitable. These three points of intersection correspond to the switching from one technology to the other as the rate of [interest] is increased from zero to its maximum.’

In this article Pasinetti’s result is displayed slightly differently. Let \( w = w_a/w_b \). Combine Eqs. (1) and (2) to produce Eq. (3).

\[
W = \frac{(1 + r)^{25} + 24}{20(1 + r)^8}
\]  

(3)

Fig. 1 employs Eq. (3) to display the result; it is a variant of Pasinetti’s Fig. 1. The ratio \( w = w_a/w_b \) is on the vertical axis (Pasinetti displays \( w_a \) and \( w_b \) separately) and \( r \) is on the horizontal axis. The range of \( r \) is 0% to 25%. The curve crosses the horizontal line \( w = 1 \) at values of \( r \approx 3.6\% \) and \( r \approx 16.2\% \). Thus, switching and then reswitching between techniques takes place as the interest rate increases.

[Fig. 1 about here.]

Reswitching is a puzzling feature of the relationship between the wage ratio and interest rate embodied in Eq. (3). It is argued here that the relationship between \( w \) and \( r \) is subtler than it appears, the subtlety arising from the form of the function. Eq. (3) is a polynomial, therefore for each value of \( w \) there are twenty-five values of... 

---

5 When the rate of interest is 25 percent the values for \( w_a \) and \( w_b \) obtained from Eqs. (1) and (2) are both zero which implies their ratio is undefined. However, Eq. (3) does define \( w \) when \( r = 25\% \). The ratio at this point is 2.42; therefore, within the relevant range of interest rates, there are only two values of \( r \) when \( w \) is unity.
that solve the equation. In general, an \( n \)th order TVM polynomial has \( n \) solutions for \( r \). Mathematically each value is as valid as any other. In order to explore the role played by every possible solution for the interest rate, an interim result is needed. This result enables a transformation of the wage equation to one in which all interest rates are not only visible, but functional.

3 The factorization of polynomials and Viete’s formulas

Aleksandrov et al. (1969) summarize a well-known result about factorization of polynomials.

‘If we accept without proof the so-called fundamental theorem of algebra that every equation \( f(x) = 0 \), where \( f(x) = x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_n \) is a polynomial in \( x \) of given degree \( n \) and the coefficients \( a_1, a_2, \ldots, a_n \) are given real or complex numbers, has at least one real or complex root, and take into consideration that all computations with complex numbers are carried out with the same rules as with rational numbers, then it is easy to show that the polynomial \( f(x) \) can be represented (and in only one way) as a product of first-degree factors

\[
f(x) = (x - a)(x - b)(x - c)\ldots(x - l)
\]

where \( a, b, \ldots, l \) are real or complex numbers.’

Furthermore: ‘Multiplying out the expression \((x - a)(x - b)(x - c)\ldots(x - l)\) and comparing the coefficients of the same powers of \( x \), we see immediately that

\[
\begin{align*}
-a_1 &= a + b + c + \ldots + l, \\
+a_2 &= ab + ac + \ldots + kl, \\
-a_3 &= abc + abd + \ldots, \\
&\vdots \\
\pm a_n &= abc\ldots l
\end{align*}
\]

which are Viete’s formulas.’ Aleksandrov et al. (1969)

These results, the factorization of a polynomial and the relationships between its parameters and its zeros, provide the route to a new wage function and its interpretation. It is shown that the wage ratio is expressible not only as a function of the interest rate and the parameters, as in Eq. (3), but also as a function of the interest rate and the zeros of Eq. (3). The latter function is new and provides the reinterpretation of reswitching.
4 Multiple interest rates in the Sraffa-Pasinetti example

Consider Eq. (3), which is in levels. Assume that the rate of interest, $r$, shifts to $R$. The wage ratio, $w$, becomes $W$. This is shown in Eq. (4), also in levels.

$$W = \frac{(1 + R)^{25} + 24}{20(1 + R)^8}$$

(4)

The next step introduces a third equation to bridge Eqs. (3) and (4), an equation in which $(w - W)$ is a function of $(r - R)$, i.e., the new function is in differences instead of levels. It is this third equation that provides the new view of reswitching.

The result about the factorization of a polynomial is applied to Eq. (3) to produce Eq. (5). The values $(1 + r_i)$ in (5) are the zeros of Eq. (3).

$$(1 + r)^{25} - 20w(1 + r)^8 + 24 = [(1 + r) - (1 + r_1)][(1 + r) - (1 + r_2)]...[(1 + r) - (1 + r_{25})]$$

or, more succinctly,

$$(1 + r)^{25} - 20w(1 + r)^8 + 24 = (r - r_1)(r - r_2)...(r - r_{25})$$

(5)

Substitute $R$ for $r$ in Eq. (5) to produce (6).

$$(1 + R)^{25} - 20w(1 + R)^8 + 24 = (R - r_1)(R - r_2)...(R - r_{25})$$

(6)

Eq. (4) is rearranged and substituted into (6) to produce (7).

$$20W(1 + R)^8 - 20w(1 + R)^8 = \prod_{i=1}^{25} (R - r_i)$$

(7)

The last equation is rearranged into Eq. (8).
\[
(W - w) = \frac{1}{20(1 + R)^8} \prod_{i=1}^{25} (R - r_i)
\] (8)

Eq. (8) is the required equation in differences that bridges Eqs. (3) and (4). There are many observations to be made about Eq. (8).

First, the left-hand side variable, the difference in wage ratios, depends on the differences between the new interest rate, \(R\), and all twenty-five initial values for \(r\) implied by the zeros of Eq. (3). One of the initial values of \(r\), call it \(r_1\), is designated the orthodox value that comes to mind when examining Eq. (3). Therefore the orthodox shift in the interest rate is \((R - r_1)\). By inspection, the relationship in Eq. (8), between the shift in the wage ratio \((W - w)\) and the shift in the orthodox interest rate \((R - r_1)\), is more complicated than that suggested by comparing equations (3) and (4). The complication requires some unpacking.

Secondly, it is possible to visualize the workings of Eq. (8). The \(i^{th}\) element \((R - r_i)\) in Eq. (8) is the difference \([(1 + R) - (1 + r_i)]\). The set \((1 + r_i)\), from \(i=1\) to \(n\), consists of the zeros of Eq. (3). As will be shown, most of these zeros reside in the complex plane, off the real number line. If absolute values are taken on both sides of Eq. (8) it becomes (9).

\[
|W - w| = \frac{1}{20(1 + R)^8} \prod_{i=1}^{25} |R - r_i|
\] (9)

The elements \(|R - r_i|\) are distances in the complex plane. The distances are rays between the zeros of Eq. (3) and the new interest rate \((1 + R)\). The new rate is the locus of the set of rays. This situation is easily demonstrated using the numbers in the Sraffa-Pasinetti example.

In Pasinetti (1966) the prescribed range for the interest rate is 0% to 25% therefore a convenient value for the initial interest rate is zero. When \(r = r_1 = 0\), Eq. (3)
implies \( w = 1.25 \). Knowing this value, Eq. (3) is solved for all twenty-five values of \( r \) that satisfy it. If these initial values for \( w \) and \( r_i \) are inserted into Eq. (9) then the relationship that remains is that between \( R \) and \( W \).

Fig. 2 is an Argand diagram showing all the roots of Eq. (3) when \( w = 1.25 \). Each ray joins a root to the locus. The locus is positioned arbitrarily at \((1 + R) = 1.1\) \( (R = 10\%)\). As the locus moves backwards or forwards along the real number line between 0% and 25%, the twenty-five rays change length. The product of these lengths affects the size of the overall product in the numerator of Eq. (9). In this way, Fig. 2 illustrates what is happening in Eq. (9) as \( R \) changes value and affects \( W \).

The third observation about the difference equation concerns the absolute values in Eq. (9). To understand how \( W \) behaves as \( R \) shifts, the signs (+/-) of some elements must be determined. Absolute values are released from some (but not all) elements on both sides of Eq. (9) and correct signs are identified. For elements of the set \( r_i \) that lie off the real number line the absolute values of \((R - r_i)\) are retained; for elements of the set that lie on the real number line the absolute values are removed and the signs of the ‘wholly real’ differences determined. The sign of the overall product then becomes apparent.

There are three real roots (interest rates) that satisfy the relevant equation:

\[
1.25 = \frac{(1 + r)^{25} + 24}{20(1 + r)^8}.
\]

They are -0.9907 (-199.07%), 1.0000 (0%) and 1.1891 (18.91%). As \( R \) passes through the range 0% to 25%, it lies consistently to the right of the first two values in the list, therefore their differences \((R - r_i)\) are positive. The value of 18.91%, however, lies inside the range 0% to 25%. As \( R \) passes through the relevant range the difference is negative when \( R \) is to the left of 18.91% and positive when it is to the right. The sign of the overall product of differences varies accordingly. Eq. (9) is modified to Eq. (10) to reflect this situation.
\[ W - 1.25 = \left( \frac{1}{20} \right) \left( \prod_{1}^{3} (R - r) \prod_{4}^{25} |R - r| \right) \frac{1}{(1 + R)^8} \] (10)

Table 1, Col. 1, contains values of R in the range 0% and 25%. Col. 2 contains the wage ratios over the range calculated in the orthodox manner from Eq. (4). Col. 3 contains the sign-adjusted product \( \prod_{1}^{3} (R - r) \prod_{4}^{25} |R - r| \) as outlined in the previous paragraphs. Col. 4 contains the composite variable comprised of differences between rates, suitably discounted. Col. 5 contains the wage ratios determined by the new Eq. (10) as R varies from 0% to 25%. The route to the wage ratios in Col. 5 is different from the route to the wage ratios in Col. 2, yet the numbers are identical.

[Table 1 about here.]

The fourth observation about the new equation for differences in wage ratios is the most important. Given a single assumption about the initial value for \( r \) in Eq. (3), all other initial values for \( w \) and \( r_i \) are determined. When these initial values are inserted into (10) the resulting equation is the same as the levels equation (4) in one significant respect: inputting a given value for \( R \) into both equations yields the same value for \( W \). The two equations, however, have entirely different structures. This fact has implications.

'A mathematical variable \( x \) is “something” or, more accurately, “anything” that may take on various numerical values.' - Aleksandrov (1969)

In economics and finance, numerical values usually reside on the real number line; they are always relative to some fixed point on the line, usually zero. For example, in Eq. (4), \( R \) departs from 0% and moves along the real number line to 25%. There is one fixed point. It is zero. The value that varies is \( (R-0) = R \); therefore \( R \) is the variable.
Eq. (10) is different. Fig. 2 shows that $R$ moves along the number line relative to twenty-five fixed points. Only three of the points are on the real number line and only one of them is zero. The remaining fixed points are distributed close to the unit circle in the complex plane. Twenty-five rays change length simultaneously, most at angles to the real number line. Faced with the structure of Eq. (10) it is difficult to maintain that the independent variable is $(R-r_1) = (R-0) = R$ alone. The independent variable is better described by the composite variable comprising every element that contains $R$:

$$
\prod_{i=1}^{3} (R-r_i) \prod_{j=4}^{25} |R-r_j| \frac{1}{(1+R)^8}.
$$

The relationship between the wage ratio and the composite independent variable in Eq. (10) is monotonic because the structure of the equation is linear. From this perspective, reswitching does not occur in the Sraffa-Pasinetti example. The situation is illustrated in Fig. 3.

[Fig. 3 about here.]

In order to emphasize the last result, Eq. (10) is restated in another form. The orthodox increment in the interest rate is expressed as $(R-r_1) = \Delta r$ which, because $r_1=0$, is equal to $R$.

$$
W = 1.25 + \left( \frac{1}{20} \right) \left( \prod_{i=1}^{3} (R-r_i) \prod_{j=4}^{25} |R-r_j| \frac{1}{(1+R)^8} \right) \Delta r.
$$

The wage ratio on the left side of this equation is expressed as a function of the difference in the orthodox interest rate, $\Delta r = R$, on the far right. It is this relationship graphed in Fig. 1.
There is a problem. If Fig. 1 is to represent the true relationship between the wage ratio and the interest rate, $R$, then all the elements that stand between the two variables in Eq. (10) ought to be fixed parameters. In fact, most of the elements vary with $R$; therefore they are components of the independent variable. It follows that the horizontal axis in Fig. 1 represents only one element of the independent variable; it does not represent the entire independent variable. The result is reswitching.

5 Generalizing from the solution to the reswitching puzzle

Why resurrect the reswitching puzzle and offer a solution to it? The objective is a better understanding of the comparative statics of the TVM equation. Different inputs (interest rates) produce different outputs (present or future values). What is the relationship between input and output?

A more general example of the TVM equation than the Sraffa-Pasinetti model is as follows. Eqs. (11) and (12) are equations for the present values of an arbitrary cash flow at two different interest rates. The interest rate shifts from $r$ to $R$ to produce the change in present value from $p$ to $P$.

\[
p = \sum_{i=1}^{n} \frac{c_i}{(1 + r)^i}
\]

\[
P = \sum_{i=1}^{n} \frac{c_i}{(1 + R)^i}
\]

Eq. (13) is derived from Eqs. (11) and (12). It states that the relative shift in value is the discounted product of all shifts in the interest rate.

\[
\left| \frac{\Delta p}{p} \right| = \prod \left| \frac{R - r_i}{(1 + R)^n} \right|
\]
The proof of Eq. (13) is not given here. As with the proof for the Sraffa-Pasinetti example above, it involves the factorization theorem.

The number of ways $\Delta r = (R - r_1)$ can impact $\Delta p$ is infinite if no constraints are placed on the other elements of the cash flow, i.e., the $c_i$ are also allowed to shift. However, there is only one way $\Delta r$ can impact $\Delta p$ if the other elements of the cash flow are held constant. The TVM polynomial, Eq. (12), contains the other elements of the cash flow as constant parameters while different values of $R$ are passed through the equation to give different values of $P$.

There is an alternative approach. When employed simultaneously, the entire cluster of interest rates (implied by the zeros of Eq. (11)) contains implicitly all information about the cash flows (parameters of Eq. (11)). That the two sets of information are equivalent is demonstrated by Viete’s formulas. In the light of this knowledge the factorization theorem is applied to the construction of an equation in differences. The construction requires the inclusion of differences between the new interest rate and all initial interest rates. An outcome of the exercise is the realization that the independent variable is more complicated than it appears. The orthodox change in the interest rate, $\Delta r = R - r_1$, by itself, is not sufficient to explain $\Delta p$.

The structure of the initial equation can be adapted to the problem under investigation. Eq. (3) is a specific structure associated with the Sraffa-Pasinetti model. Eq. (11) is a more general structure. Whatever the structure, the general principle remains the same: the extent to which the dependent variable changes in response to a shift in the interest rate can be shown to depend on all initial values of the interest rate before the shift, not just the orthodox value. The independent variable is a composite variable comprising all possible adjustments to the interest rate. The difference in the orthodox interest rate by itself is not adequate to explain the change in the dependent variable. Ignoring this fact gives rise to puzzling relationships between variables. Reswitching is one example.
The analysis developed here, with its use of differences rather than levels, and the employment of an entire cluster of initial interest rates, holds implications for other topics in economics and finance. In the context of bond mathematics, Osborne (2005) produces a formula sought since Macaulay (1938): an algebraic formula for the interest elasticity of bond price that provides accurate results. The new formula has no need for convexity or the other terms of a Taylor series expansion. The formula employs the differences between the new interest rate and all initial interest rates. In the context of capital budgeting, Osborne (2010) shows that net present value (NPV) per dollar invested is composed of the mark-downs of the cost of capital relative to all internal rates of return (IRR), thereby contributing to the debate about the relative merits of NPV and IRR as investment criteria. This list of topics open to the ‘multiple interest rate’ approach cannot be complete; others must exist.

Reswitching, therefore, is a puzzle in a list of similar puzzles in economics and finance, each causing debate in its own field. The debates have a common factor: the TVM equation. One of the most useful questions that can be asked of the equation is how value varies under different assumptions about the interest rate. This deceptively simple question has proved difficult to answer satisfactorily. The long histories of reswitching and other debates mentioned above are testament to the difficulty. Reswitching is an anomaly in the Kuhnian sense (Kuhn, 1962). It is one indicator among several of an issue that permeates economics and finance.

6 Conclusion

In this article the reswitching phenomenon is re-examined in the context of the Sraffa-Pasinetti model. The phenomenon does not occur when it is analyzed

6 Dorfman (1981) is possibly the earliest example; Dorfman’s mode of analysis, however, is different from that adopted here. Dorfman uses all interest rates in their ‘raw’, complex form, whereas the approach described here uses absolute differences between interest rates, which are real numbers. The relationship between the two modes of analysis is an open question and is left for future research.

The ‘multiple interest rate’ literature in the context of capital budgeting is summarized in Magni (2010). Most authors discuss only the multiple real rates. Hazen (2003) and Pierru (2010) are recent exceptions in which there is explicit use of complex rates. Their discussion is of rates per se, used individually as investment criteria (internal rates of return); they do not use the rates simultaneously as ingredients in a formula, as in Dorfman’s article or this one.
using a new TVM equation expressed in differences rather than levels, an equation containing all possible shifts in interest rates, rather than the single, orthodox shift alone. The methodology can be generalized: it applies to a TVM polynomial of any order; and it applies to topics other than reswitching.

Questions remain. First, what are the implications of the analysis described in this work for the capital controversies overall? There was more to the Cambridge capital controversies than the reswitching puzzle. For example, can the new approach be adapted from comparative static analysis at a moment in time to the analysis of a process through time?7 Secondly, as noted above, what are the implications for other topics in economics and finance that employ the TVM equation? The TVM equation is ubiquitous in economics and finance therefore other applications are likely. Finally, what economic or financial meaning can be attributed to all possible solutions for the interest rate, especially the complex? Answers to these questions are left for future research. In the meantime, this article provides a different perspective on a famous debate.

Acknowledgements
Ian Davidson (Sussex) suggested the possibility that the ‘multiple interest rate’ technique may be relevant to the reswitching debate. Ephraim Clark (Middlesex) supplied helpful comments and Carter Daniel (Rutgers) improved style and presentation. Lucian Tipi (Sheffield Hallam) gave valuable advice, as did Graham Partington and colleague (Sydney). Geoff Harcourt (Cambridge) suggested edits, gave guidance to the literature, provided encouragement and gave comments on the substance. Heinz Kurz (Graz) and Pierangelo Garegnani (Rome) also made valuable comments. I should like to express my thanks to all without implicating anyone in the methodology or the conclusions.

7 See Harcourt (1972, p.122): ‘Following Joan Robinson’s strictures that it is most important not to apply theorems obtained from the analysis of differences to situations of change …, modern writers usually have been most careful to stress that their analysis is essentially the comparisons of different equilibrium situations one with another and that they are not analyzing actual processes.’ No attempt is made to incorporate the passage of time into the analysis outlined here. It remains a challenge.
References


Bruno, M., Burmeister, E. & Sheshinski, E. 1966. The nature and implications of the reswitching of techniques, Quarterly Journal of Economics, 80(4)


Fisher, I. 1907. The Rate of Interest, New York: Macmillan

Garegnani, P. 1966. Switching of techniques, Quarterly Journal of Economics, 80(4)

Hagemann, H. & Kurz, H. 1976. The return of the same truncation period and reswitching of techniques in neo-Austrian and more general models, Kyklos, 29(4)


Kuhn, T. 1962. The Structure of Scientific Revolutions, University of Chicago Press

Levhari, D. & Samuelson, P. 1966. The nonswitching theorem is false, Quarterly Journal of Economics, 80(4)


Table 1. The numbers of the Sraffa-Pasinetti example. The wage ratios in Col. 2 and the numbers in Col. 5 are identical yet they are calculated in two different ways. Col. 2 contains values for \( w_a/w_b \) produced by Eq. (4). Col. 5 is calculated from Eq. (10) using the sign-adjusted products of the differences between interest rates. Note the change of sign in Col. 3 that occurs as the interest rate, \( R \), passes by one of the three real roots at 18.91 percent. The causal variable (the composite variable containing \( R \)) is in Col. 4; it is the variable on the x-axis of Fig. 3.

<table>
<thead>
<tr>
<th>Col. 1</th>
<th>Col. 2</th>
<th>Col. 3</th>
<th>Col. 4</th>
<th>Col. 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td>( w_a/w_b )</td>
<td>( \prod_{i=1}^{3}(R-r_i)\prod_{i=4}^{25}</td>
<td>R-r_i</td>
<td>)</td>
</tr>
<tr>
<td>0</td>
<td>1.2500</td>
<td>0.0000</td>
<td>0.0000</td>
<td>1.2500</td>
</tr>
<tr>
<td>0.01</td>
<td>1.1674</td>
<td>-1.7890</td>
<td>-1.6521</td>
<td>1.1674</td>
</tr>
<tr>
<td>0.02</td>
<td>1.0942</td>
<td>-3.6509</td>
<td>-3.1160</td>
<td>1.0942</td>
</tr>
<tr>
<td>0.03</td>
<td>1.0299</td>
<td>-5.5755</td>
<td>-4.4013</td>
<td>1.0299</td>
</tr>
<tr>
<td>0.04</td>
<td>0.9742</td>
<td>-7.5484</td>
<td>-5.5155</td>
<td>0.9742</td>
</tr>
<tr>
<td>0.05</td>
<td>0.9268</td>
<td>-9.5500</td>
<td>-6.4638</td>
<td>0.9268</td>
</tr>
<tr>
<td>0.06</td>
<td>0.8875</td>
<td>-11.5543</td>
<td>-7.2493</td>
<td>0.8875</td>
</tr>
<tr>
<td>0.07</td>
<td>0.8564</td>
<td>-13.5272</td>
<td>-7.8730</td>
<td>0.8564</td>
</tr>
<tr>
<td>0.08</td>
<td>0.8333</td>
<td>-15.4248</td>
<td>-8.3335</td>
<td>0.8333</td>
</tr>
<tr>
<td>0.09</td>
<td>0.8186</td>
<td>-17.1910</td>
<td>-8.8267</td>
<td>0.8186</td>
</tr>
<tr>
<td>0.10</td>
<td>0.8125</td>
<td>-18.7550</td>
<td>-8.7494</td>
<td>0.8125</td>
</tr>
<tr>
<td>0.11</td>
<td>0.8155</td>
<td>-20.0280</td>
<td>-8.6907</td>
<td>0.8155</td>
</tr>
<tr>
<td>0.12</td>
<td>0.8280</td>
<td>-20.8990</td>
<td>-8.4408</td>
<td>0.8280</td>
</tr>
<tr>
<td>0.13</td>
<td>0.8507</td>
<td>-21.2306</td>
<td>-7.9861</td>
<td>0.8507</td>
</tr>
<tr>
<td>0.14</td>
<td>0.8845</td>
<td>-20.8527</td>
<td>-7.3101</td>
<td>0.8845</td>
</tr>
<tr>
<td>0.15</td>
<td>0.9303</td>
<td>-19.5566</td>
<td>-6.3931</td>
<td>0.9303</td>
</tr>
<tr>
<td>0.16</td>
<td>0.9894</td>
<td>-17.0861</td>
<td>-5.2117</td>
<td>0.9894</td>
</tr>
<tr>
<td>0.17</td>
<td>1.0631</td>
<td>-13.1285</td>
<td>-3.7388</td>
<td>1.0631</td>
</tr>
<tr>
<td>0.18</td>
<td>1.1529</td>
<td>-7.3029</td>
<td>-1.9428</td>
<td>1.1529</td>
</tr>
<tr>
<td>0.19</td>
<td>1.2606</td>
<td>0.8534</td>
<td>0.2122</td>
<td>1.2606</td>
</tr>
<tr>
<td>0.20</td>
<td>1.3884</td>
<td>11.9008</td>
<td>2.7677</td>
<td>1.3884</td>
</tr>
<tr>
<td>0.21</td>
<td>1.5385</td>
<td>26.5165</td>
<td>5.7708</td>
<td>1.5385</td>
</tr>
<tr>
<td>0.22</td>
<td>1.7137</td>
<td>45.5174</td>
<td>9.2747</td>
<td>1.7137</td>
</tr>
<tr>
<td>0.23</td>
<td>1.9170</td>
<td>69.8865</td>
<td>13.3399</td>
<td>1.9170</td>
</tr>
<tr>
<td>0.24</td>
<td>2.1517</td>
<td>100.8043</td>
<td>18.0346</td>
<td>2.1517</td>
</tr>
<tr>
<td>0.25</td>
<td>2.4218</td>
<td>139.6862</td>
<td>23.4355</td>
<td>2.4218</td>
</tr>
</tbody>
</table>
Fig. 1. The Sraffa-Pasinetti example: the wage ratio $w = w_a/w_b$ on the y-axis varies with the rate of interest on the x-axis. The wage ratio is equal to unity at two values of the rate of interest: 3.6 percent and 16.2 percent; therefore reswitching is apparent.
Fig. 2. All twenty-five roots of Eq. 3, i.e., \(1.25 = \frac{(1+r)^{25} + 24}{20(1+r)^8}\), are plotted in the complex plane. Twenty-two are in eleven, conjugate complex pairs and three are real at -0.9907, 1.000 and 1.1891. The rays stretch between each of the roots, \((1+r)\), and the locus \((1+R)\). The locus in the figure is arbitrarily set at 1.1. As \((1+R)\) moves back and forth along the real number line between 1.00 and 1.25, the rays change length and their product changes value, thereby affecting the value of the numerator on the right-hand side of Eq. (9). The unit circle is shown to provide scale.
Fig. 3. The Sraffa-Pasinetti example: the wage ratio \( w_a/w_h \) on the y-axis varies with the composite variable containing differences between rates of interest on the x-axis. The graph is based on Eq. (10) in the text. The causal variable is in Col. 4 of Table 1. The wage ratio is equal to unity at only one value of the causal variable because the relationship is linear; there is no reswitching.
Work 10
Osborne, M.
The meaning of internal rates of return:
An addendum to ‘A resolution to the NPV-IRR debate’
SSRN http://ssrn.com/abstract=163481, 2010c
The meaning of internal rates of return:
An addendum to ‘A resolution to the NPV-IRR debate?’

Michael J. Osborne *

1st draft, Jun. 2010; 2nd draft, Jul.; 3rd draft, Sep. 2010

Key words:
Capital budgeting; complex plane; internal rate of return; net present value

JEL Classifications:
C00, C60, E22, E40, G00, G1, G24, G30, G31, O16, O22

Abstract
All internal rates of return (IRRs) are shown to have meaning as well as use. IRRs are interpreted as the units in which value is measured and the quantities of such units. On this basis it is argued that the orthodox interpretation of internal rate of return as an investment criterion may not be correct. The argument runs counter to a long tradition in the capital budgeting literature.

* email: m.osborne@shu.ac.uk
Sheffield Business School, Sheffield Hallam University, Howard Street, Sheffield, S1 1WB
The meaning of internal rates of return:
An addendum to ‘A resolution to the NPV-IRR debate?’

1 Introduction

The core equation used in capital budgeting is the time value of money (TVM) equation; it is a polynomial. Gauss (1777-1855) proved the Fundamental Theorem of Algebra about polynomials. The theorem states that every polynomial has \( n \) roots where \( n \) is the order of the polynomial (see Stillwell, 1989). The result means that the equation for the internal rate of return (IRR) must have \( n \) solutions for IRR. In a previous article, Osborne (2010) built on this foundation to demonstrate that NPV per dollar invested can be expressed as the product of all mark-ups of the \( n \) IRRs over the cost of capital. In this way, all possible IRRs (unorthodox as well as orthodox) find use as building blocks of NPV. A problem with the previous article is that no financial meaning is attributed to the unorthodox IRRs. The mathematics lacks an interpretation. This addendum is an attempt to remedy the deficiency. It contains a suggestion about the meaning of all IRRs and, in consequence, a new interpretation of the orthodox IRR.

2 The multiple interest rate literature

The bulk of the multiple interest rate literature does not address all possible interest rates. It addresses the restricted set of all real rates. Magni (2010) provides a recent, comprehensive survey of the literature in the context of capital budgeting. The works that address all \( n \) solutions to the TVM equation are few in number.

Perhaps the earliest reference to all possible IRRs is found in an exchange of views in the 1930s. Boulding (1936a) contains a discussion of investment appraisal. In a lengthy appendix, Boulding describes the difficulty of calculating the internal rate of return and presents a procedure for finding a solution.

Wright (1936) replied to Boulding’s article. His reply is interesting because, possibly for the first time in the academic literature, a researcher points out the existence
of all \( n \) solutions in unequivocal terms. He makes explicit reference to the fundamental theorem of algebra and writes:

\[ \text{The theorem asserts that any algebraic equation of degree } n \text{ has } n \text{ solutions.} \]
\[ \text{Applied to Mr. Boulding's definition of the interest rate, it means that several different and reasonable values may be obtained for } i \ldots \]
(Wright, 1936)

In his reply to Wright, Boulding (1936b), takes a stance that will - with rare exceptions - characterize most research on the subject of multiple IRRs for the next 70 years.

\[ \text{Now it is true that an equation of the } n\text{th degree has } n \text{ roots of one sort or another, and that therefore the general equation for the definition of a rate of interest can also have } n \text{ solutions, where } n \text{ is the number of "years " concerned. Indeed, if we adopt continuous compounding, as in strict theory we should, the theoretical number of solutions is infinite! Nevertheless, in the type of payments series with which we are most likely to be concerned, it is extremely probable that all but one of these roots will be either negative or imaginary, in which case they will have no economic significance.}^{1} \]

Samuelson (1937) discusses the papers by Boulding. He also refers to the possibility of ‘a multiplicity of solutions’ but does not elaborate on the nature of the multiplicity.

Thus, economists became aware of the issue of multiple solutions to the time value of money equation sometime before, or during, the 1930s. From the beginning, most of the multiple solutions, especially the complex, were dismissed. Boulding’s comment, that the negative and ‘imaginary’ solutions have no economic significance, is

\[^{1}\text{The reference to imaginary solutions is not strictly correct. Few solutions are purely imaginary. Most solutions are complex and therefore contain both real and imaginary components.} \]
probably indicative of the opinion of many researchers in the field and the reason why most of the works in the multiple interest rate literature surveyed by Magni (2010) focus on the restricted set of real solutions.

In the intervening period, authors who explicitly mention the possibility of complex solutions include Hirschleifer (1958) and Feldstein and Flemming (1964). They point out that wholly positive or wholly negative cash flows, with no change of sign at all, give rise to imaginary solutions. Hirschleifer concludes, presciently in the light of the ideas in this article, that the idea that IRR ‘represents a growth rate in any simple sense cannot be true’. Feldstein and Flemming comment only that ‘the examples given are rather peculiar’.

Dorfman (1981) is a significant exception to the multiple interest rate literature during the twentieth century, not only because he addresses all $n$ solutions, but also because of his mode of analysis. He examines the case where the proceeds of an investment are serially reinvested in projects of the same type, and the process is continued indefinitely. He shows that the growth path of a dollar placed in such an investment depends on all the roots of the internal rate of return equation. Thus, there is an essential similarity between Dorfman’s analysis and the analysis in Osborne (2010) and this addendum, namely that both employ all interest rates as components of another financial concept, rather than as interest rates per se. There is also an essential difference between the two modes of analysis. Dorfman employs all complex solutions in their raw form, i.e., in the form $a + b.i$. In contrast, the analysis in Osborne (2010) uses the absolute values of differences between complex interest rates. Such differences are real numbers. The mathematical relationship between the two modes of analysis remains an open question for future research.²

² There is another reason to praise Dorfman’s contribution. Researchers today benefit from the existence of computers and sophisticated mathematical software, such as Mathcad, Mathematica, Matlab and Maple, with which to explore the complex plane. The software was mostly developed during the 1980s. Dorfman’s analysis was done nearly a decade before the software became widely available therefore his work is truly pioneering.
The literature in the twenty-first century exhibits increasing interest in the unorthodox solutions. Hazen (2003), Hartman and Schafrick (2004), Pierru (2010) and Magni (2010) all refer to the complex rates. The authors take different approaches, however.

Hazen sees every possible IRR as an investment criterion. On his own admission, the procedure he describes is ‘roundabout’, although it appears to work for any IRR. He writes ‘there is no need to discard “unreasonable” or “extreme” internal rates – all are equally valid’. Hazen uses all IRRs to convert the original cash flow for a project into \( n \) alternative cash flows called ‘investment streams’, each stream using one of the \( n \) IRRs. The net present value of each investment stream is determined by valuing it at the cost of capital. The original project is judged to be profitable, or not, according to whether an investment stream is profitable, or not.

Importantly, Hazen makes explicit use of complex valued IRRs. He considers only the real part of complex valued IRRs and only the real part of complex valued investment streams. In doing so he reaches the same conclusions as he does using the real valued IRRs. It does not appear to matter which of the many IRRs (and its associated investment stream) is used to produce a decision. This ‘roundabout’ method need only be applied once, and a real-valued IRR will serve. Thus, in Hazen’s analysis, complex valued results appear to be superfluous. Moreover, since the NPV criterion is simpler, Hazen urges use of NPV. He shows that the IRR criterion can give invest/not-invest decisions for single projects that are consistent with the NPV criterion. Thus, for Hazen, …

‘... the problem of multiple or non-existent [read complex] internal rates of return – universally regarded as a fatal flaw for the IRR method – is not really a flaw at all, and can easily be dealt with conceptually and procedurally’.

The IRR methodology suggested by Hazen does suffer from problems. First, when comparing mutually exclusive projects, a decision using his method can conflict
with the decision from the NPV criterion, the well-known pitfall introduced by Lorie and Savage (1955). Secondly, the method only considers the real parts of the complex results, ignoring any information contained in the imaginary components. Thirdly, the method does not provide an interpretation of complex rates. In Hazen’s own words:

‘We are currently unaware of an economic interpretation of complex-valued rates of return or complex-valued investment streams, and without such an interpretation, it would be hard to justify any economic recommendation without resort to performance measures such as present value.’

Hartmann and Schafrick (2004) adapt an approach first suggested by Cannaday et al. (1986). They partition the cash flow into lending and borrowing periods according to whether the first derivative of the present worth function with respect to the interest rate is positive or negative. The divisions between partitions are where the first derivatives of the present worth function are zero. The relevant IRR is found in the partition containing the cost of capital. The usual comparison of the IRR with the cost of capital is made, explicitly acknowledging the existence of complex roots, although not employing them. They avoid the issue. In their words, ‘in our partitioning scheme, they [the complex internal rates of return] are removed from the analysis’. They do this removal by determining which partitions have complex roots assigned to them and collapsing together partitions such that a real root can be assigned to every partition.

Nevertheless, Hartman and Schafrick puzzle about the complex solutions:

‘In our partitioning scheme, they [the complex internal rates of return] are removed from the analysis and we assume that a project’s status (loaning or borrowing) does not change in the new partition. ... Unfortunately, while our method of collapsing partitions allows for correct analysis in the presence of complex roots, it muddles our definition of a project being loaning or borrowing according to the slope of the present worth. This might signal that complex roots do have meaning, although we do not have an interpretation at this time.’

Hartman and Schafrick (2004)
Pierru (2010) examines complex interest rates in the context of a portfolio of two assets. ‘When a project involves the joint production of two outputs whose markets are subject to different risks, our approach allows the project’s cash flows to be discounted at a single (but complex) rate.’ The single complex rate is interpreted to represent several different real rates at the same time. A difficulty is that the interpretation is confined to a narrow range of applications. Pierru acknowledges this when he writes ‘we are aware of the apparently limited practical interest of the interpretations proposed …’.

In addition to offering the literature review mentioned earlier, Magni (2010) discusses complex rates explicitly but adopts yet another approach to them. He circumvents the possibility of awkward, i.e., complex rates, by finding an average of many specially defined one-period rates of return embedded in the investment. In Magni's words, this average internal rate of return (AIRR) approach ‘wipes out complex valued numbers’.

This kind of analysis employs similar reasoning to Fisher (1907, 1930), Hirschleifer (1958) and Bailey (1959). They too circumvent the multiple interest rate issue by breaking up one long investment into separate short investments, each with its own rate of interest.

Magni ‘presents the notion of Average Internal Rate of Return … [as a concept that] … generalizes the usual IRR notion’. He comments that ‘unfortunately, the venerable internal rate of return … is not a reliable profitability index because it may not exist, multiple roots may arise, and, in general, is incompatible with the NPV’.

In the next section we review briefly the main result in Osborne (2010) that shows how all the multiple IRRs may be employed with the cost of capital, and how, taken as a cluster, they are not only compatible with NPV, they are NPV, and therefore do contribute to a reliable profitability index.
3 A brief summary of the result that uses all IRRs

This summary borrows heavily from Osborne (2010). Equation (1) shows a typical equation for net present value (NPV).

\[
NPV = -I_0 + \sum_{i=1}^{n} \frac{c_i}{(1 + r)^i}
\] (1)

In equation (1), \(NPV\) is net present value. \(I_0\) is the initial investment in period zero (a positive number shown with a minus sign to indicate that money is paid away). The \(c_i\) represent the cash flows in the periods \(i = 1\) to \(n\) (usually positive to reflect money received, but some values could be negative to indicate additional investment). Finally, \(r\) is the cost of capital and \(n\) is the number of periods.

The internal rate of return is the rate that brings the present value of the returns into equality with the initial investment, i.e., \(NPV\) is zero. The rate is shown as \(R\) in the variant of equation (1) that is equation (1a).

\[
0 = -I_0 + \sum_{i=1}^{n} \frac{c_i}{(1 + R)^i}
\] (1a)

The relationship between the cost of capital and the IRR can be expressed as 
\((1 + R) = (1 + r)(1 + m)\) in which \(m\) is the interest rate that marks up the cost of capital to the internal rate of return. Assuming a single cost of capital, there must be \(n\) multiplicative mark-ups of the \(n\) IRRs over the cost of capital, i.e., \((1 + R_i) = (1 + r)(1 + m_i)\) for all \(i\) from 1 to \(n\).

Osborne (2010) presents and proves the new equation for NPV per dollar - equation (2).

\[
\left| \frac{NPV}{I_0} \right| = \prod_{i=1}^{n} |m_i|
\] (2)
Equation (2) shows that all IRRs, the unorthodox as well as the orthodox, serve as components of NPV. To be precise, the product of the mark-ups of all \( n \) IRRs over the cost of capital is NPV per dollar invested. While the equation demonstrates a use for all IRRs, it says nothing about their meaning.

4 The meaning of IRRs

An initial assumption is that the price of a product, taken alone, has no meaning in the sense that it conveys nothing about the value of the product. Products only have relative values: values relative to other products at the same moment in time, or relative to the same product at another moment in time.

The relative value of two products at a moment in time is the ratio of their individual prices. For example, if \( P_A / P_B = $14 / $10 = 1.4 \) then good A is 40% more valuable than good B. The additional value is the difference in price relative to one of the prices. It is represented as follows: \( \frac{P_A}{P_B} - 1 = \frac{P_A - P_B}{P_B} = 0.4 \). The additional value is always a pure number because the currency units cancel.

The pure number that represents additional value at a moment in time is not normally divided into specific units, although it could be. For example, the statement that good A is 40% more valuable than good B, is the same as saying good A is more valuable than good B by 40 units of 0.01 or 1% each. The unit could be 5% in which case the additional value is 8 units of 0.05 each. In wholesale finance one percentage point is often divided into 100 parts, known as basis points. One basis point is therefore \( 1/(100 \times 100) = 1/10,000 = .0001 \) of the whole. Thus good A is more valuable than good B by 4,000 units of one basis point each.

\[
\frac{P_A}{P_B} - 1 = \frac{P_A - P_B}{P_B} = 0.4 = 40 \times 0.01 = 8 \times 0.05 = 4,000 \times 0.0001
\]
Taking this view, additional value, or the difference between two prices relative to one of them, is the product of the unit of value and the number of units. This simple structure is called here the ‘standard value structure’.

The price of good A today can be compared with its price yesterday or its likely price tomorrow. When time is included, the habit is to express the increase in value per unit time. The usual unit is the year and the rate of increase in value per year is variously known as the annual percentage rate (APR), compound annual rate (CAR), yield to maturity (YTM) or internal rate of return (IRR), depending on context. If \( P_A^i \) is the price of good A at time \( i \), and the price of good A increases by 40% over four years, then the calculation is as follows:

\[
\frac{P_A^4}{P_A^0} = 1.4 \text{ implies } 1.4 = (1+r)^4.
\]

Therefore \( r = (1.4)^{1/4} - 1 = 0.087757 = 8.7757\% \) per year.

In this calculation the unit of time is standardized to a year. Given the number of years, the interest rate that emerges from the calculation is an annual rate. If the standard value structure described above applies to this situation, it ought to be possible to express the increment in price relative to the original price as a unit of value multiplied by a number of units. An obvious candidate for the unit of value is the rate of interest per period, \( r \). In which case

\[
\frac{P_A^4 - P_A^0}{P_A^0} = r \cdot X \text{ where } X \text{ is the number of units.}
\]

It follows that \( 0.4 = 0.087757 \cdot X \) and \( X = 4.5580 \) units. The standard value structure has been imposed on the calculation. At this stage, the origin of \( X \) is not clear. It has been declared, and not derived from within the problem.

If the assumptions change, for example, if time is measured in half-years, then the details of the calculation change but the format remains the same.
\[
\frac{P_A^8}{P_A^0} = 1.4 \text{ implies } 1.4 = (1+r)^8.
\]

It follows that \( r = 1.4^{1/8} - 1 = 0.042956 = 4.2956\% \text{ per half-year.} \)

If the standard value structure is applied to the calculation, there is a multiplier \( X \) such that \( \frac{P_A^8 - P_A^0}{P_A^0} = 0.4 = 0.042956.X \) where \( X \) is the number of units, calculated to be 9.3119. Again, we can query the imposition of the structure and request a derivation for \( X \).

At this point an interpretation of the meaning of the \((n-1)\) unorthodox interest rates is in sight. The interpretation employs the notions of the special form of the TVM polynomial, and the special relationship between the coefficients and roots of the special form. These notions are introduced in Osborne (2010) and are reproduced for convenience in the appendix to this addendum. We begin with the first example in which time is measured in years and \( P_A^4 = (1 + r)^4 P_A^0 \). This equation is transformed into the special form.

\[
-1 + \frac{\left( \frac{P_A^4}{P_A^0} - 1 \right)}{(1+r)^4} + \frac{1}{(1+r)^4} = 0
\]

The special form implies the special relationship
\[
\left| \frac{P_A^4}{P_A^0} - 1 \right| = \prod_{1}^{4} |r_i|.
\]

The left side of the last equation has value 0.4. On the right, the orthodox interest rate is labeled \( |r_i| \) and is known to be 0.087757.

Therefore
\[
\left| \frac{P_A^4}{P_A^0} - 1 \right| = \prod_{1}^{4} |r_i| = |r_1| \prod_{1}^{4} |r_i| \text{ becomes } 0.4 = 0.087757 \prod_{2}^{4} |r_i|.
\]
The product of the absolute values of the three unorthodox interest rates is
\[ \prod_{j=2}^{4} |r_j| = \frac{0.4}{0.087757} = 4.5580. \] This result is easy to check using the original values for the roots.\(^3\)

Therefore
\[ \left| \frac{P_A^8}{P_A^0} - 1 \right| = \prod_{j=1}^{4} |r_j| = |r_1| \prod_{j=2}^{4} |r_j| \] becomes \(0.4 = 0.087757 \times 4.5580.\)

This last equation displays the standard value structure. On the left is the change in price relative to the original price. The ratio is a pure number representing additional value. On the right is the unit of value multiplied by the number of units. The unit of value is the orthodox interest rate, \(r_1\). The number of units is the product of the \((n-1)\) remaining rates. In this way the entity X is derived from within the problem.

Turning to the second example, in which time is measured in half-years and \(P_A^8 = (1 + r)^8 P_A^0\), we can transform the equation into the special form.

\[ -1 + \frac{\left( \frac{P_A^8}{P_A^0} - 1 \right)}{(1 + r)^8} + \frac{1}{(1 + r)^8} = 0 \]

The special form implies the special relationship
\[ \left| \frac{P_A^8}{P_A^0} - 1 \right| = \prod_{j=1}^{8} |r_j|. \]

\(^3\) It is possible to visualize the result and calculate it with a hand calculator. The example has the same structure as a zero coupon bond. The roots of a zero coupon bond are distributed evenly around a circle of radius \((1+r)\) where \(r\) is the orthodox interest rate, in this case 0.087757. Therefore the four roots are located at 1.087757, -1.087757, 1.087757\(i\) and -1.087757\(i\) where \(i = \sqrt{-1}\). The absolute values of the interest rates are the distances between the four roots and the point \((1,0)\). It is left to the reader to apply Pythagoras’ theorem several times to establish that the product of all four distances is 0.4, and the product of the three unorthodox distances is 4.5580.
The left side of the last equation has value 0.4. On the right, the orthodox interest rate is labeled $|r_1|$ and is known to be 0.042956.

Therefore \[ \left| \frac{P_A^8}{P_A^0} - 1 \right| = \prod_{i=1}^{8} |r_i| = |r_1| \prod_{i=2}^{8} |r_i| \] is rewritten as \[ 0.4 = 0.042956 \prod_{i=2}^{8} |r_i|. \]

The product of the absolute values of the seven unorthodox interest rates is \[ \prod_{i=2}^{8} |r_i| = \frac{0.4}{0.042956} = 9.3118. \] This result is not as easy to check as the last result.\(^4\)

It follows that \[ \left| \frac{P_A^8}{P_A^0} - 1 \right| = \prod_{i=1}^{8} |r_i| = |r_1| \prod_{i=2}^{8} |r_i| \] and \[ 0.4 = 0.042956 \times 9.3118. \]

Once again, the interpretation is that the standard value structure applies: the additional value (measured by the ratio of the change in price to the original price) is equal to the unit of value multiplied by the number of such units. The unit of value is the orthodox interest rate, $r_1$, and the number of such units is the product of the absolute values of the ($n-1$) other interest rates.

The examples given so far in this section have simple structures. First, there are comparisons between the prices of two different goods, $P_A$ and $P_B$, at the same moment in time. Secondly, there are comparisons between the prices of the same good at two different moments in time, e.g., \[ P_A^0 \text{ and } P_A^4. \]

We turn now to the formulas for NPV and IRR. The notions of NPV in the form of equation (1) and IRR in the form of (1a) introduce two complications into the analysis. First, we are no longer comparing just two prices or cash flows, as in the earlier

---

\(^4\) The result is more difficult to calculate because it involves 8 roots around a circle of radius 1.042956. It is not obvious that the product of the 8 distances radiating from the point (1,0) to the roots on the circumference of the circle is precisely 0.4. It is perhaps even less obvious that the product of the 7 ‘unorthodox’ distances across the circle is 9.3118. Access to advanced mathematical software such as Maple, Mathematica or Matlab is necessary to confirm it.
examples; we are considering many cash flows distributed across time. Secondly, we are comparing values that result from inputting different interest rates. Despite these complications similar reasoning about relative value applies. The new equation for NPV per dollar invested is repeated below.

\[
\frac{NPV}{I_0} = \prod_{1}^{n}|m_i| \tag{2}
\]

The new equation displays the standard value structure. On the left-hand side is the increment in value relative to the initial investment. On the right-hand side is the product of the mark-ups of all IRRs over the cost of capital. One of them is the mark-up of the orthodox IRR over the cost of capital, \( m_i \), which is interpreted to be the unit of value. The product of the remaining mark-ups is the quantity of units of value. In this way the mark-ups of all IRRs over the cost of capital are shown to have meaning.

This interpretation of the meaning of the whole cluster of interest rates has implications for the meaning of the orthodox IRR. The size of the unit and the number of units are inextricably related, indeed simultaneously determined, and it is their combination that impacts the value of an investment. The implication is that neither IRR, nor the mark-up of IRR over the cost of capital, can be an investment criterion. The new interpretation runs counter to the orthodox interpretation in financial literature.

The mark-up of the orthodox IRR over the cost of capital is the unit of value and its size is determined by the length of the chosen sub-period of time. The total time period may be broken into any number of desired sub-periods and the resulting interest rate per sub-period calculated. Whatever the length of the sub-period, the fundamental mathematical relationship remains true: the total increment in value is the product of the absolute values of the mark-ups in all rates. The suggested interpretation also remains: one mark-up is the unit of value, and the remaining mark-ups are multiplied together to give the quantity of units.
Thus, there is no longer a need to be concerned about the existence of multiple
IRRs, real or complex; or to be concerned about picking the one ‘true’, real IRR and
discarding the rest. All IRRs are determined simultaneously and every IRR has a part to
play in the NPV cluster; moreover, they all have meaning.

In the following, penultimate section, additional arguments are offered for the
views expressed in Osborne (2010) and this addendum.

5 What really happens when an interest rate shifts?

Equation (2) is a key equation. It is in multiplicative guise because it contains all
the mark-ups \((m_i)\) of the IRR over the cost of capital. It can also be expressed in additive
guise. This alternative, equivalent expression (2a) employs all the additive increments
\((a_i)\) of the IRR to the cost of capital, i.e., \(R_i = r + a_i\) for \(i = 1\) to \(n\).

\[
\left| \frac{NPV}{I_0} \right| = \prod_{i=1}^{n} |a_i| \quad (2a)
\]

On the right-hand side of (2a) there is the element \(|a_1| = |R_1 - r|\) which is the
difference between the first IRR and the cost of capital. It is convenient to think of the
first IRR as the orthodox IRR. Equation (2a) can, therefore, be rewritten as (2b).

\[
\left| \frac{NPV}{I_0} \right| = \prod_{i=2}^{n} |R_i - r| \quad (2b)
\]

On the left-hand side of (2b) is NPV per dollar invested, which is the difference in
the value of a project that results from valuing it at the cost of capital rather than the
internal rate of return. NPV per dollar is expressed as a function of the difference
between the orthodox IRR and the cost of capital on the far right-hand side. The
relationship between these two differences is often graphed in financial textbooks. As the
cost of capital falls away from IRR, the value of NPV rises from zero to some positive number.

There is a problem. An element stands between NPV per dollar on the left-hand side and the difference between orthodox interest rates, $R$ and $r$, on the right-hand side. For the foregoing description to be true in the causal sense, this element ought to be a fixed parameter. In fact the element also contains the cost of capital; therefore the element is not a parameter. In reality, every element on the right hand side of (2a) and (2b) changes as we select new costs of capital at different distances from IRR in order to obtain different values for NPV. The causal variable is not limited to the orthodox mark-up alone.

This situation is represented in Figure 1. Instead of the usual graph with NPV on the vertical axis and the rate of interest on the horizontal axis, the two axes are shown vertically. Causation is read from right to left. In the first instance, the orthodox IRR ($R$ in the figure) is fed into the TVM polynomial. The value of NPV that results is zero. In the second instance, another rate of interest, the cost of capital ($r$), is fed into the TVM polynomial. The value of NPV that results is some positive number.

[ Figure 1 about here ]

Two conclusions are usually drawn from this comparative static calculation. The first is that the difference in NPV is triggered by the shift in interest rate from IRR to the cost of capital, i.e., the shift from $R$ to $r$. This conclusion is true. The second is that the increase in NPV is uniquely associated with that shift. This conclusion is not true. Equation (2) shows that the interest rate pulls away from all values of the IRR, not just the orthodox. The relationship between NPV on the left-hand side and the difference between IRR and the cost of capital on the right-hand side is more complicated than it looks. There is more going on behind the scenes in the complex plane than appears on the real number line alone.
Another view of the situation is as follows. When an interest rate per period is applied to a cash flow, whether discounting or compounding, the rate is applied to each dollar in each period. The total value that results from the entire process does not depend on this rate only; it also depends on how many dollars it is applied to, and when. The product of the \((n-1)\) unorthodox rates captures the joint impact of the quantities and their timing.

The argument can be couched another way. Equation (1a) defines the IRRs. If, instead of calculating the IRRs from the cash flows, we reconstruct the cash flows from the solutions to the IRR equation, all solutions are needed, not just the orthodox solution. This means that all the solutions for \((1+R_i)\), taken as a cluster, embody all the information in the cash flows. It is, therefore, not too surprising that if the cash flows are revalued by a shift in the interest rate, the revaluation requires all possible shifts, not just one. The purpose of this addendum is to suggest financial meaning for all shifts.

6 Conclusion

Since the time of Fisher (1907) the margin of IRR over the cost of capital has been characterized in the financial literature as an investment criterion to rival the criterion of net present value. There has been much debate about the relative merits of the two criteria.

The analysis in Osborne (2010) shows that NPV per dollar is composed of the mark-ups of every possible IRR over the cost of capital, every mark-up being simultaneously determined. This raises a question mark over the continued interpretation of a single mark-up as an investment criterion. The alternative interpretation offered here is that the orthodox mark-up of IRR over the cost of capital is the unit of value in which the value of an investment is measured. The product of all other mark-ups of IRR over the cost of capital is the quantity of such units. The mark-ups work together, as a single cluster. They comprise the standard value structure.
By way of a pointer to further work, the possibility is noted that the analysis described in this addendum may be applied to other topics in finance employing key interest rates such as the annual percentage rate (APR), the compound annual rate (CAR) and the yield to maturity (YTM).
Appendix

Proof of the special relationship between the coefficients and roots of the special form of the TVM equation

Any TVM equation can be rearranged into a special form, which is repeated below for convenience.

\[-1 + \sum_{i=1}^{n} \frac{b_i}{(1+z)^i} + \frac{1}{(1+z)^n} = 0\]

The parameters and the roots of this special form are linked in a particular way: the absolute value of the sum of the parameters is equal to the product of the absolute values of the interest rates.

\[|\sum b_i| = \prod |z_i|\]

A proof of the statement is as follows. First, in the special form, set \(Z = (1+z)\) and multiply throughout by \(Z^n\), then factorize the equation and take absolute values. The result is:

\[|Z^n - b_1Z^{n-1} - ... - b_{n-1}Z - b_n - 1| = |Z-Z_1||Z-Z_2|...|Z-Z_{n-1}|Z-Z_n|\]

Now set \(Z = 1\). The left side becomes the absolute sum of the parameters, while the right side becomes the product of the absolute values of \(z_i\), since \(Z_i = (1+z_i)\) therefore \(|z_i| = |1 - Z_i|\). The proof is now complete.
References


Figure 1
In the TVM polynomial for NPV, NPV=zero maps back to R, the orthodox IRR, and NPV=‘some positive number’ maps back to r, the cost of capital. In the new equation (2b) that is derived from the TVM polynomial, the change in value of NPV per dollar, from zero to some positive value, does not map back to the single difference between the orthodox IRR and the cost of capital; rather, it maps back to the differences between all possible IRRs and the cost of capital.