Incorporating option values into the economic evaluation of health care technologies

Stephen Palmer, Peter C. Smith
Centre for Health Economics, University of York, York YO10 5DD UK

Received 26 April 1999; received in revised form 10 February 2000; accepted 8 March 2000

Abstract

Despite uncertainty being intrinsic to economic evaluation of health care, existing techniques for handling uncertainty remain underdeveloped compared to the formal techniques commonly applied in the business sector. This paper develops an alternative approach to handling uncertainty in economic evaluation based on 'option-pricing' techniques. The presence of uncertainty and the degree of irreversibility of a decision make it clear that some flexibility in the timing of a decision is often a desirable characteristic with an economic value. We demonstrate how option-pricing techniques can be applied to the decision rules for economic evaluation in health care. The key determinants of an option value are the presence and type of uncertainty; the ability to defer a decision; and the irreversibility of the decision. The relative significance of each of these for a particular economic evaluation will depend on the particular characteristics of the technology under consideration. © 2000 Elsevier Science B.V. All rights reserved.

JEL classification: H51; I18

Keywords: Option-pricing; Economic evaluation; Uncertainty; Methodology
1. Introduction

Uncertainty is intrinsic to all economic evaluation. Indeed it can be argued that, in the absence of uncertainty, economic evaluation is trivial, obviating the need for highly trained and moderately well-paid economists. The importance of uncertainty in the economic evaluation of health care technologies is no less acute than in any other sector of the economy. Yet, while the business sector has developed a number of formal techniques for handling uncertainty in investment appraisal, methodologies for incorporating uncertainty into health technology evaluation are currently at best crude, and are at worst distinctly misleading.

This paper puts forward an approach to evaluating health technologies based on the ideas of real options. These seek to integrate the uncertainty and irreversibility associated with a technology into a unifying theory of economic evaluation, which offers the decision-maker a systematic framework for handling the degree of uncertainty inherent in evidence on the cost-effectiveness of a health technology. The intention is to indicate how it is possible in many circumstances not only to indicate the magnitude of certain types of uncertainty, but also to quantify its impact on any economic evaluation.

The paper is organized as follows. Section 2 discusses existing techniques for handling uncertainty in economic evaluation, and explores the use of an alternative approach, known as ‘real options’. Section 3 puts forward some formal real options models. Section 4 examines the implications of the options approach for health technology assessment. Concluding comments are given in Section 5.

2. Uncertainty and the options approach

Hitherto, two techniques for handling uncertainty due to both sample and non-sample related variation have traditionally been considered relevant in health care evaluation: sensitivity analysis and statistical analysis (Briggs and Gray, 1999). Both of these approaches offer a central point estimate of cost effectiveness (in the form, for example, of cost per QALY) in conjunction with certain bounds within which the estimate can be expected to lie.

Clearly, the reporting of such bounds marks a major step forward in the evaluation of health care technologies. However, it frequently presents the decision-maker with a major problem in coming to a judgement as to whether the reported uncertainty is enough to reject, or at least to defer implementation of, a new technology deemed cost-effective according to the point estimate. While a decision may be straightforward when the estimated bounds (e.g., 95% confidence limits) of the cost per QALY lie entirely to one side of the ratio considered acceptable, a problem arises when the estimated range straddles this critical threshold value. The decision as to whether to implement then becomes a matter of
judgement, and existing techniques offer little quantitative guidance on which to base this judgement (Briggs and Gray, 1999).

The central insight of the options approach to investment appraisal is that most investment decisions have three important characteristics.

1. There exists a degree of uncertainty about the future state of the world.
2. The investment entails an essentially irreversible commitment of resources.
3. There is usually some discretion as to the timing of the investment.

Conventional cash flow techniques consider the investment as being ‘now or never’, and little attention is paid to the possibility of deferring a decision until some later time, when better information regarding costs and benefits may become available. Yet in practice deferral is one of the most important (and frequent) decisions taken, in order that further information can be assembled (Claxton, 1999). It is furthermore important to recognize that an interest in uncertainty in economic evaluation implies a certain sunk cost associated with implementing a new technology. If there was no such ‘implementation cost’, then policy-makers could, if necessary, costlessly switch between technologies as new evidence becomes available. The concern with uncertainty and reversibility suggests that decision-makers are seeking some reassurance that the policy they are implementing is sustainable, because it cannot be reversed without costs. This being the case, it is clear that some flexibility in the timing of an investment decision is often a desirable characteristic with an economic value. Option-pricing theory seeks to value such flexibility.

Options are ubiquitous in economic life. Hitherto, most academic and practitioner emphasis has been on financial options, in the form of various sorts of derivative securities. However, as Trigeorgis (1996) notes, there is no reason to exclude more concrete situations from the analytic framework. These have become known as real options. This approach has already been used in environmental economics to assess the potential irreversible impact of proposed developments on the environment (Arrow and Fisher, 1974). Further examples include the valuation of mineral rights or film rights, or decisions to invest in research and development, which may confer an option to enter a market (Amram and Kulatilaka, 1999). Indeed, it is probably the case that option considerations are dominant when real investment decisions appear to fly in the face of conventional net present value (NPV) calculations. For example, it is well documented that firms are much more cautious about big investments (market entry) than NPV calculations suggest they should be, and are also reluctant to exit markets even when NPV rules would suggest abandonment. In this respect, apparent conservatism in the investment market is readily explained in terms of reluctance to make irreversible decisions in the presence of uncertainty, and the associated retention of options. A decision to proceed with an irreversible decision is equivalent to a loss of a hitherto available option. In the same way, certain investments (for example, in R&D) which cannot
be justified on conventional NPV grounds may be made because they create options with an economic value.

3. Options models

In order to explore some of the important issues underlying option-pricing theory, we present a very simple stylized model within a cost-benefit framework, with the concept of ‘net social benefit’ (NSB) replacing the NPV criterion. Consider an investment of instantaneous cost \( C = 4500 \) which yields expected benefits with financial value \( B \) in perpetuity. Depending on the future state of the world, those benefits might be large \( (L = 400) \) with probability \( q \) or small \( (S = 200) \) with probability \( (1 - q) \). We assume \( q = 0.4 \) and a discount rate \( r \) of 5% per annum.

If the investment is to be made now, then the expected future stream of benefits is \( [qL + (1 - q)S] = [0.4 \times 400 + 0.6 \times 200] = £280 \) per annum. Hence, the NSB of the investment is as follows:

\[
NSB_1 = -C + \sum_{t=1}^{\infty} \frac{qL + (1 - q)S}{(1 + r)^t} = -4500 + 280/0.05 = 1100.
\]

If, on the other hand, investment can be deferred for 1 year, then the future stream of benefits would be known with certainty. The current NSB of the project if benefits are favourable will be:

\[
NSB_2 = \frac{1}{1 + r} \left[ -C + \sum_{t=1}^{\infty} \frac{L}{(1 + r)^t} \right] = \frac{1}{1.05} \left[ -4500 + 400/0.05 \right] = 3333
\]

while the current NSB of the project if benefits turn out to be unfavourable will be

\[
NSB_3 = \frac{1}{1 + r} \left[ -C + \sum_{t=1}^{\infty} \frac{S}{(1 + r)^t} \right] = \frac{1}{1.05} \left[ -4500 + 200/0.05 \right] = -476.
\]

Thus, the decision-maker would only implement in 1 year’s time if the benefits turned out to be favourable. Otherwise, the project would be abandoned (with NSB therefore zero rather than \(-476\)). The NSB of the project with deferral is therefore \( 0.4 \times NSB_2 + 0.6 \times 0 = 1333 \). Note that this exceeds the NSB of the projected implemented immediately by 233.

That is, although the benefits are deferred for 1 year, the loss arising from the delay in implementation is more than offset by the improved information which
permits the decision-maker to abandon the project in unfavourable circumstances. Note that in this situation the deferral value of the project can be quantified at 233, the difference between the NSB with and without deferral. In effect, the deferral value is the difference between the benefits arising from abandonment in unfavourable circumstances and the costs of deferring immediate implementation. Clearly, this deferral value is closely linked to the value of information (in this example, the value of perfect information, deferred for 1 year).

This example is very artificial. In practice, future information arrives slowly and imperfectly. However, at least conceptually, the example can be readily extended to many time periods and many more states of the world. It highlights a number of issues that are features of most option valuations. Notably, other things being equal the greater the level of uncertainty implicit in the decision (either in cash flows or discount rates), the greater the value of the option, as it becomes more worthwhile to await new information. An associated issue is that, as the time horizon over which a decision may be deferred increases, so the value of the option increases.

Uncertainty in option-pricing theory is modelled as a stochastic process, in which the variable of interest evolves over time in a partially random fashion. Clearly, the precise form of stochastic process to be used in any modelling work must depend on the nature of the problem under investigation. In this respect, two particularly important useful tools are the Wiener process and the Poisson process. Under the Wiener process (also known as Brownian motion), an underlying random variable varies incrementally with known variance in each time period. The simplest form of this form of uncertainty is the random walk, in which the best predictor of tomorrow’s value is today’s value, with the change following a normal statistical distribution. In health care, it might be used to model a continuously varying variable such as (say) the prevalence of a disease. Quite frequently a ‘drift’ is introduced into the Wiener process, which allows a systematic trend to be modelled independently of the random element. This might be incorporated (say) to model an expected downward drift in the price of a drug. The mathematical form of this process can be written as:

\[ \text{d} x = \alpha \text{d} t + \sigma \text{d} z \]

where \( x \) is the variable of interest, \( \text{d} x \) is its movement in small time \( \text{d} t \), \( \alpha \) is the drift parameter, \( \text{d} z \) is the stochastic change in time \( \text{d} t \), and \( \sigma \) is the standard error of the random change per unit time period. A number of further generalizations of the basic Wiener process can be introduced where necessary.

In the simplest form of a Poisson process, a random variable can take only two values and has a fixed probability in each time period of changing from one to the other. This process is used to model situations in which the variable of interest is subject to rare but critical ‘jumps’. In health care, it might for example be used to model the emergence of a new drug (which constitutes a discrete shock to the associated market). The magnitude of the jump may also be allowed to vary (for
example, the improved efficacy offered by the new drug might be allowed to vary stochastically).

The structure of the simple example given above is similar to the familiar decision tree of conventional decision analysis, and one solution method could be the familiar method of ‘folding back’ the decision tree using backward induction. Dynamic programming is a more general and efficient method of solving such problems. With modern computing capability, calculation of multiperiod, multiple state examples may often be feasible, the most problematic issue often being the availability of relevant data rather than the solution method. Dynamic programming can also be extended to continuous rather than discrete time. Although other methods exist for solving option-pricing problems (such as contingent claims analysis) it is probably most appropriate in health care settings to use the dynamic programming formulation, in which the discount rate is fixed exogenously.

Within this framework, we now consider a more general treatment of an investment decision, which is analogous to a typical decision as to whether or not to implement a new health technology. We consider the benefits of the technology to be valued as \( V \), and there is an investment cost of \( I \). Let the discount rate be \( r \). Then suppose the value of the project evolves over time according to the geometric Brownian motion

\[
dV = \alpha V dt + \sigma V dz
\]

which unfolds over an infinite time horizon. Here, \( V \) is the variable of interest (the value of the technology), \( dV \) is its movement in small time \( dt \), \( \alpha \) is the drift parameter, \( dz \) is the stochastic change in time \( dt \), and \( \sigma \) is the standard error of the random change per unit time period. Note that, compared with the pure Wiener process, the formulation of this process introduces a term in \( V \) on the right hand side, which effectively allows us to model percentage changes rather than absolute changes in the variable \( V \).

Using dynamic programming, Dixit and Pindyck show how this formulation gives rise to a differential equation

\[
\frac{1}{2} \sigma^2 V^2 F''(V) + \alpha VF'(V) - \rho F(V) = 0
\]

with a set of associated boundary conditions. The function \( F(V) \) is the current value of the investment if its current estimated benefits are \( V \) (that is, before implementation, \( F(V) \) comprises the sum of the current ‘intrinsic’ value \( V \) and the deferral value of having the potential to implement in the future).

The set of equations given above has a solution of the form

\[
F(V) = AV^B
\]

where

\[
A = \frac{(\beta - 1)(\beta - 1)}{\beta^B \Gamma(\beta - 1)}
\]
and

\[ \beta = \frac{1}{2} - \frac{\alpha}{\sigma^2} + \sqrt{\left(\frac{\alpha}{\sigma^2} - \frac{1}{2}\right)^2 + 2 \rho/\sigma^2}. \]

More importantly, the solution yields an optimal value of \( V \), denoted \( V^* \), which is such that once \( V \) exceeds \( V^* \), the technology should be implemented. \( V^* \) is given by:

\[ V^* = \frac{\beta}{\beta - 1} I. \]

From the point of view of this paper the key observation is that the critical value \( V^* \) is not a constant. With no uncertainty (\( \sigma = 0 \)), \( V^* = I \). This reflects the traditional cash flow rule that in order to invest, benefits must simply exceed costs. However, if uncertainty exists (\( \sigma > 0 \)) then it follows that \( \beta/(\beta - 1) > 1 \) and hence \( V^* > I \). Consequently, the application of the simple NSB rule is incorrect; the combined impact of uncertainty and irreversibility results in a divergence between \( V^* \) and \( I \).

The critical value, \( V^* \), and the size of the divergence between \( V^* \) and \( I \), depends on the value of \( \beta \), which is itself determined by the discount rate \( \rho \), the drift parameter \( \alpha \) and the stochastic variance parameter \( \sigma \). In order to illustrate the importance of this result, consider the perfectly reasonable situation in which the annual discount rate is 5% (\( \rho = 0.05 \)), there is no drift (\( \alpha = 0 \)) and the estimated value of \( V \) has an annual standard deviation of 10%. This implies that \( \sigma = 0.1 \). Then it can be readily shown that \( \beta = 3.7 \), and the critical value \( V^* \) becomes 1.37. That is, estimated benefits must 37% higher than costs before immediate implementation is optimal. More generally, Fig. 1 illustrates the relationship between the amount of uncertainty \( \sigma \) and \( V^* \), given \( \rho = 0.05 \) and \( \alpha = 0 \). Note the rapid increase in the critical value \( V^* \) associated with quite modest increases in uncertainty.

So far, we have presumed that the decision to proceed with a new technology is completely irreversible. That is, once the investment \( I \) has been made, the technology cannot be abandoned. In practice, decisions are to differing extents reversible. This possibility is readily modelled by introducing the notion of a ‘scrap’ value \( S \) of the technology. This is the amount of the investment that can be recouped if a decision to implement a new technology is subsequently reversed. Clearly, \( I > S \) if there is an element of irreversibility. However, \( S \) may be positive (for example, if redundant equipment has a scrap value) or negative (perhaps if the political costs of abandonment are considered to be high).

It is then necessary to introduce the notion of two states of the world (the technology not implemented, and the technology implemented). The preceding analysis presumed that the technology was not yet implemented, and yielded a range for \( V \) of \((0, V_{37})\) within which \( V \) could vary without that state being changed.
Fig. 1. Required value of benefits as a function of uncertainty.

The upper limit of the range is $V_H$, the critical value $V^*$. If the technology has been implemented, it is presumed to continue in use within the range $(V_L, \infty)$, where $V_L < V_H$, reflecting the inertia caused by the option value.

Dixit and Pindyck (1994) show how this more complex situation can be accommodated by separately modelling the value of the technology in the two states of the world: not implemented and implemented. The first state of the world is as above, giving an option value of the technology:

$$F_1(V) = A_1 V^{\beta_1}$$
valid in the range $(0, V_H)$, where the subscript denotes values identified in state one (not implemented). In state two (implemented), the structure of the process remains unchanged, although different parameter estimates are obtained, and the implemented value $V$ of the project must be added to the solution, yielding:

$$F_2(V) = A_2 V^{\beta_2} + V$$
valid in the range $(V_L, \infty)$.

In conjunction with appropriate boundary conditions, these equations can be solved (using numerical rather than analytic methods) to yield estimates of unknowns $A_1$, $A_2$, $\beta_1$, $\beta_2$, $V_L$ and $V_H$. In general, if $0 < S < I$, the result will be a curve similar to but less steep than that shown in Fig. 1. In the extreme case of $S = I$ (complete reversibility), there is no benefit from delaying implementation if $V > I$, so that the curve in Fig. 1 becomes a horizontal line, and with $V_L = V_H = I$. This situation applies in the unusual case when the decision-maker can costlessly switch from one technology to another as new evidence becomes available. This case illustrates that it is essential for there to be some degree of irreversibility for
the options approach to be useful. It also illustrates the method of computing $V_t$, the value of $V$ at which an implemented health technology should be abandoned.

The exposition here has made a number of limiting assumptions. In particular, we have presumed that the value of the technology follows a Wiener rather than Poisson process. We have also assumed that new information results in a change in the estimate of the value of the technology, but no change in its standard error. Alternatives to these assumptions are readily modelled using numerical methods, and offer no conceptual challenge to the general principles we set out.

4. Implications for health technology assessment

Crucial to the options approach is the notion that the passage of time will tend to reveal new estimates for key sources of uncertainty. In health care this will often be the case. For example:

- the equilibrium price of new drugs or capital equipment may change once the initial stage of the product life cycle has ended;
- estimates of the long term benefits of a therapy and the generalisability of the results may change as more trials become available, and longer term outcomes are reported;
- the external validity and generalisability of the results of pharmacoeconomic evaluations will become evident when the results of late phase III, post-marketing and phase IV studies are reported;
- estimates of population costs and benefits may change as more epidemiological evidence is assembled.

The results outlined above may therefore have crucial implications for the economic evaluation of health care technology. They imply that the presence of even modest degrees of uncertainty may give rise to substantial increases in the traditional cost-effectiveness ratio. In this section we summarize the principal issues they give rise to from an options perspective. In doing so we recognize that there is a continuing debate about the most appropriate evaluative framework to adopt (Stinnett and Mullahy, 1998; Laska et al., 1999; Claxton, 1999). We do not seek to enter that debate here, and wish to demonstrate only that option considerations are crucial whichever framework is adopted. We focus our discussion on the familiar cost-effectiveness ratio (ICER), although the principles we set out hold for all economic evaluation methodologies.

Under these methods, the relative efficiency of an intervention is traditionally assessed with reference to a critical ratio (or threshold value of the ICER) which is used to determine whether an intervention should be adopted (Birch and Gafni, 1992; Johannesson and Weinstein, 1993; Johannesson, 1995). The insight offered by the options perspective is that a technology’s ICER should first be adjusted
according to the degree of uncertainty and reversibility implicit in the evaluation before being compared with this critical ratio. Only if the magnitude of uncertainty and reversibility were equal across all technologies would such adjustment be unnecessary.

The estimation of the ICER adjusted in line with option considerations is achieved with minor modifications to the NSB calculations reported previously. Suppose that the critical cut-off value for the ICER has been set at £8000 per QALY, and that this benchmark criterion applies only in a situation of perfect information and no uncertainty \((\sigma = 0)\) or costless reversibility \((I = S)\). Suppose then that a new intervention is being appraised for which uncertainty exists \((\sigma > 0)\) and there is a degree of irreversibility \((I > S)\). Then the intervention’s ICER should be altered as follows using option price techniques. We have illustrated the situation in which the annual discount rate is 5\% \((r = 0.05)\), there is no drift \((\alpha = 0)\) and the estimated value of \(V\) has an annual standard deviation of 10\%. Under these condition it was shown that the ratio \(V^*/I\) for the NPV calculations became 1.37. Hence, according to the revised net-benefit criterion, the benefits of a project would have to exceed the costs by 37\% before immediate investment was considered the optimal strategy. The adjustment to the ICER is directly analogous. Suppose that the unadjusted point estimate of the ICER is £7000 per QALY which lies within the acceptability range. The revised decision rule implies multiplying the calculated ICER by 1.37 (i.e. £9600 per QALY). That is, formal consideration of uncertainty and reversibility leads to a quantitative change in the computed ICER and (in this case) a revised decision regarding implementation.

Although this example emphasizes the importance of uncertainty in evaluations, the criterion of irreversibility is equally important in determining option values in health technology assessment. The most obvious example of an irreversible decision is one in which a major piece of capital equipment is purchased. Yet if this equipment is readily resold on the open market, the implementation may not be as irreversible as it appears. On the other hand, programmes involving little setup costs may not be readily reversible if perceptions and expectations have changed such that reversal of the policy is no longer feasible. We believe that the great interest shown in the uncertainty surrounding economic evaluations suggests that irreversibility is a major issue in the appraisal of many health care technologies. However, the extent of the irreversibility will clearly differ between applications, leading to associated variations in the required adjustment to the ICER.

5. Conclusion

We have sought to demonstrate that option considerations are likely to be important in health care technology assessment, and that they may lead to major adjustments to the estimation of a technology’s ICER. The key determinants of
any option value, and hence of the adjustment to the ICER in health care are the magnitude of uncertainty in parameter estimates, the extent to which deferral is possible, and the extent to which the decision to implement is irreversible. The approach is consistent with all accepted methodologies for the economic evaluation of health care interventions, and renders unnecessary the estimation of sampling distributions, which has been the focus of much recent attention in evaluation methodology. In this sense, it is consistent with Claxton’s (1999) “irrelevance of inference” argument, which is derived from an alternative decision-theoretic perspective.

The paper has concentrated on the macro implications of option-pricing for particular technologies. There is however no reason why it should not also be applicable at the microlevel of the individual patient. For example, treatments exhibit different degrees of uncertainty and different degrees of reversibility. Furthermore, there are conditions where there may be the ability to defer a decision until more definitive information is available could be considered a relevant strategy at this level (e.g. watchful waiting in the management of small abdominal aortic aneurysms and benign prostatic hyperplasia). An option-pricing approach may offer the possibility of more systematic advice on the preference ordering for particular treatments for the individual.

One final consideration is that all of the analysis described here presumes an essentially passive approach towards the emergence of new information. It may well be fruitful, however, to seek to integrate the options approach with decision analytic approaches toward acquisition of effectiveness information (Claxton, 1999). The intention would be that, rather that acting as a passive recipient of new information, the regulator should be able to make judgements about where research effort should be directed towards accumulating more information. This line of enquiry offers a potentially fertile agenda for future research.

Acknowledgements

The authors would like to acknowledge the help of colleagues Karim Abadir, Karl Claxton, Christine Godfrey, Hugh Gravelle, Mark Sculpher, participants in seminars held at the Universities of York and Durham, and the journal’s referees.

References


