

**Reviewing Discount Rates in ONS Valuations**

**Report for the Office for National Statistics**

**Version 2**

**28th November 2017**

**Professor Mark C. Freeman**

**Dr Iain Clacher**

**Professor Karl Claxton**

**Dr Ben Groom**

## Executive summary

This report responds to a request from the Office for National Statistics (ONS) for support in conducting an external review of the discount rates that should be used in the production of its outputs. The ONS produces valuations of a wide range of topic areas, such as the environment (natural capital), the workforce (human capital), health output, and pensions. Each of these areas applies a discount factor to convert future returns into present day values.

The ONS commissioned us to produce an external review of the discount rates that might be used in the production of its outputs while remaining within the guiding framework of the Ramsey Rule, as advocated by UK Treasury's Green Book. Specifically, we were asked to identify considerations impacting on the parameters to be used in the formula in specific topic areas, and propose appropriate values for each component of the Ramsey formula for each area as required. Recommendations about values for all discount rates that might be used in the ONS were asked to be fully justified on the same conceptual basis, with each value derived from a common set of principles of when to make changes to the standard 3.5% rate.

In this document, we report our findings in each of the following areas

- |                                   |           |                    |
|-----------------------------------|-----------|--------------------|
| • Environmental and human capital | Section 1 | Lead: Dr Groom     |
| • Health                          | Section 2 | Lead: Prof Claxton |
| • Investments and R&D             | Section 3 | Lead: Prof Freeman |
| • Pensions                        | Section 4 | Lead: Dr Clacher   |

We recommend that the ONS uses the UK Treasury's Green Book real rate of 3.5% in almost all circumstances for discounting purposes. The theoretical case for using this rate is clearly set out in the Green Book and this has stood the test of time.<sup>1</sup> By using this rate for a wide range of asset classes, there is consistency of treatment across the accounts, and comparability with the assessment of social costs and benefits in other areas of discounting, including the appraisal of public project proposals. Even in the case of pensions liabilities, we believe that this rate is the most suitable to use.

For the environment and health, where the relative prices of these goods are expected to change over time, we recommend that this is dealt with in the explicit pricing of the costs and benefits rather than from a "dual discounting" adjustment to the 3.5% real rate. While the two approaches are quantitatively equivalent, it is conceptually correct to separate out issues of scarcity through pricing from the considerations of welfare over time that underlie the discounting debate. Ultimately, there is no fundamental difference between environmental, health and other assets from a discounting perspective, and the appropriate cost of capital should reflect this.

---

<sup>1</sup> See also the companion report, Freeman, M.C. & B. Groom (2016), "Discounting for Environmental Accounts: Report for the Office for National Statistics", Office for National Statistics.

The single case in which we recommend that the discount rate deviates away from 3.5% real is for non-financial investment capital held by the private (market) sector. As the ONS wishes to estimate a user cost of capital, its methods should reflect the corporate finance discount rate literature. In contrast to the risk-free approach taken by HM Treasury's Green Book, it is well documented that private sector investors require compensation for risk. We recommend that the ONS uses the Capital Asset Pricing Model to determine such discount rates as this is by far the most commonly used model in the private sector.

In principle, the 3.5% rate could be used for a range of purposes, including allowing the ONS to establish wealth accounts. The World Bank will publish a report later this year or early next year which does this across countries following the accounting requirements of the System of National Accounts, like the ONS, and using similar valuation techniques: e.g. the Jorgenson-Fraumeni framework is used for human capital. For purposes of comparability and harmonisation across nations the World Bank uses a constant discount rate across all countries of 5%. The ONS could have undertaken a similar exercise before our discounting advice, but the result would have been less accurate. The advice contained within the report does a much more nuanced job in relation to discounting because it recommends, i) harmonising around the Green Book guidelines for the purpose of estimating wealth; and ii) that attention be paid to some specific dynamic issues associated with particular resources when calculating present values including relative prices when dealing with environmental capital, survival rates/hazard rates when dealing with human capital, and the opportunity cost issues of the health sector. These are matters that are not dealt with as explicitly in the World Bank calculations, and might also have been overlooked under previous ONS practices.

The report proceeds as follows:

#### *Discounting environmental and human capital*

The section on discounting the environment explains how the specific features of environmental resources and eco-system services and stocks influence how changes in environmental stocks and flows should be treated from the perspective of Cost-Benefit Analysis (CBA) and natural capital valuation. Discounting environmental goods and services differs from discounting units of consumption due to the way in which environmental goods and services are becoming more or less scarce over time. Economic scarcity of environmental resources will depend on several features of the resource itself: i) its substitutability with the other goods and services that make up consumption; ii) the ecological thresholds associated with environmental resources, beyond which flows of environmental benefits fall to zero (e.g. due to the collapse of a fishery or eco-system); iii) the rate of physical depletion of the environmental resources over time; and, iv) the preferences for environmental goods.

Two effectively identical approaches can be used in CBA to account for the changing scarcity of environmental resources. First, one could have good specific discount rates, one for consumption and one for the environment. This is known as "dual-discounting". Alternatively, one could reflect changing scarcity in appropriate changes in the relative accounting prices for consumption and environmental resources. Ultimately we recommend the latter since it is in line with existing Green Book guidance that makes clear that the issue of relative prices is an important consideration in any CBA study. This recommendation does, though, come with some important caveats. Finally, valuation of a natural capital stock requires a slightly different approach. Relative prices remain important, but in addition the flows of services from the environmental capital stock need to be discounted using a rate of discount that reflects the standard social discount rate, the growth rate of renewable resources and the behavioural response to adjustments to the stock. Together these make up the *rate of return to the natural capital stock*. Calculation of this net rate of discount requires ecological information, as well as empirical evidence of economic behaviour. These principles of environmental discounting are illustrated using examples.

The current ONS approach to the estimation of human capital uses the Green Book guidance on discounting to calculate the life-time earnings of individuals, following the income-output approach of Fraumeni and Jorgenson (1989), before aggregating across the population. Our recommendations are to continue with this practice. Two issues do arise concerning the growth of wages over time, and the use of the survival rates in the calculation of expected lifetime earnings. The first issue is whether or not using GDP growth to account for wage increases in the future has the effect of double counting the return to human capital that is already reflected in wages levels. The other point is whether or not the survival rates that are used to calculate lifetime earnings are not doubly discounted by the Green Book discount rate, which already contains a mortality risk component. These are issues to be pursued by further research.

### *Health discounting*

There is a history of changing and sometimes conflicting recommendations about discounting policies for health projects that arise from alternative normative positions taken and judgements about the empirical questions that follow.

The choice for the ONS is whether it wishes to reflect the normative position that has been adopted in most evaluations of health care projects for decision making bodies in the UK, that the objective of health care expenditure is to improve health, or a broader view of welfare that would be consistent with other accounts and the welfare arguments that underpin the Ramsey Rule.

Given the need for consistency between accounts and the importance of being able to explicitly quantify other impacts beyond measures of health and public health expenditure it would seem appropriate to convert all effects into streams of

consumption gains and losses discounted at a Social Time Preference Rate (STP) that would be the relevant rate to apply in all contexts where benefits and costs have been expressed in terms of consumption, including any decline to reflect the impact of uncertainty in the estimate of STP.

Separating the rate of change of values/prices from the pure consumption/utility arguments for discounting STP avoids embedding multiple arguments in the discount rate for health and health care costs. The separate and explicit accounting for these arguments allows available evidence to be used transparently and consistently, while preserving the possibility of accountable deliberation about evidence, values and unquantified arguments in decision-making processes.

In addition to STP this also requires the following quantities specific to health to be assessed: i) the marginal productivity of health care expenditure in producing health; ii) the consumption value of health; iii) The marginal productivity of health care expenditure in producing net production outside the health care sector; and iv) how these quantities are likely to evolve over time. Reasonable default assumptions for each are possible based on the current evidence, such as it is.

#### *Discounting investments and R&D*

At present, it is standard practice in National Accounting to use ex-post measures when estimating the discount rate for non-financial investment capital, and this is the basis for current ONS methods. Such discount rates are based on realised gross operating surpluses. Not only can this lead to certain empirical problems – for example, the calculation of negative costs of capital following prolonged periods of economic weakness – but, more fundamentally we know that this is not how practitioners actually go about determining required rates of return.

In the private sector, there is extensive survey evidence that shows that corporations rely heavily on the Capital Asset Pricing Model (CAPM). While this model has a number of weaknesses, including the restrictive theoretical assumptions and poor empirical support, this nevertheless remains best practice amongst firms. In particular, there is evidence that this model works better at the project level than for equity returns. We therefore recommend that the ONS considers the use of the CAPM for private sector user costs of capital. In the public sector, the Treasury recommends Green Book rates, and we see no compelling reasons why this should also not be the approach taken by the ONS in this context. We further show that there is theoretical consistency between the CAPM and Green Book approaches.

Two main issues remain. First, R&D is undertaken by firms in the knowledge that this activity can either be discontinued or expanded in future. This is most commonly modelled in finance not through Net Present Value calculations but instead through the use of Real Options. While an analysis of these methods lies outside the scope of this report, the ONS might consider such approaches in the future for R&D capital. Second, PFI deals, which combine a public sector user with a private sector capital

provider, must reconcile the Green Book and CAPM approaches. This is not a simple task because of the absence of risk premia in Treasury discount rate recommendations.

### *Pensions discounting*

Currently, the ONS applies European best practice for the discounting of pension liabilities for National Accounting purposes. However, this is not consistent with the discounting of other governmental liabilities which use a social rate of time preference e.g. a Ramsey discount rate. In looking at the discounting of pension liabilities for National Accounts, the ONS has to consider a number of questions going forward: (i) Why is comparability across National Accounts desirable when it prevents straightforward comparisons with other parts of governmental liabilities that use STP? (ii) Why is the yield on a basket of European government debt the most appropriate measure for discounting in National Accounts? (iii) What alternatives can meet the principles for discounting that resulted in the current approach to discounting pensions in National Accounts?

For all government accounts including National Accounts, discounting using an explicit STP is the most appropriate to achieve these goals.

- The current rate used is based on the average yield on a basket of long-dated European sovereign and has remained fixed at 3% real or 5% nominal, which currently gives the same rate as STP.
- The application of a STP rate allows comparability across government accounts and allows for consistency in the valuation of other governmental liabilities.
- Such an approach should also be used across countries and explicit country specific adjustments can be made so that comparisons are possible.
- If an explicit STP were applied, then where disagreement exists regarding any assumption that goes into its construction, this can be resolved via a transparent process, which is not possible under the current approach.

## Section 1: Discounting for environmental and human capital

The Natural Capital Committee has recommended that natural capital should be accounted for systematically in the National Accounts. As a separate measure, the NCC has a longer-term vision to construct *comprehensive* or *inclusive* wealth accounts. Mayer (2014) stresses that these two objectives, while related, ought not to be mixed up. Following the NCC recommendations, only the latter is influenced by discounting procedures.

For the NCC, building natural capital accounts for the purpose of national accounting requires evaluating the flows of income in each year stemming from different sorts of natural capital, as well as the costs of maintaining and restoring natural capital. Each would be a separate item in the National Accounts and the difference between them would indicate the income net of restoration costs for accounting purposes, in the same way as rental income minus maintenance costs would reflect the flow of income from a housing asset.

The compilation of inclusive wealth accounts has the objective of providing better more well-rounded measures of economic performance than the standard measures of GDP/GNP or their per capita equivalents. The latter are deficient as measures of long-term economic performance since they ignore the fact that GDP could be high today due to the depreciation of natural capital, e.g. running down of non-renewable resources such as oil, or the irreversible depletion of renewable resources such as fisheries. Such measures of economic performance do not provide a signal of whether a growth trajectory is sustainable in the long-term, neither do they have any direct welfare significance in theory (Weitzman 1976; Hamilton and Hartwick 2014). GDP growth ignores the erosion of natural capital, and hence the erosion of an important component of national wealth, along with physical and human capital.

Theoretical contributions in the field of sustainability economics show that only if some measure of *comprehensive/inclusive* wealth is maintained over time will the path of future well-being be sustainable in the sense of leaving future generations no less well-off. Evaluating wealth, and forming inclusive wealth accounts is a primary objective of the NCC. In order to do this it is necessary to provide a valuation of natural wealth. This requires the valuation of the present value of services flowing from natural capital, and to do this requires guidance on the appropriate discount rate with which to calculate present values over long-time horizons.

### 1.1 Discounting and relative prices

Where comprehensive/inclusive wealth accounts are to be evaluated, or environmental costs and benefits are to be assessed as part of a cost-benefit analysis, non-marketed environmental commodities (clean air, clean water, amenity values, but more broadly, eco-system services) may require slightly different treatment compared to commonly marketed consumption goods. One proposal that has been influential in the theoretical literature, and more recently at the policy level, is to use “dual” discount

rates, i.e. different discount rates for different types of good. The idea has been discussed for many years (e.g. Weikard and Zhu 2005, Sterner and Persson 2008; Baumgartner *et al.* 2014).

The general point is that, beyond the typical marketed consumption goods that are evaluated in CBA, non-marketed goods associated with the environment (and possibly health) should enter as separate arguments of social welfare and should therefore be treated differently in CBA to reflect some of their special characteristics, such as substitutability, and future trajectories. Formally, environmental goods (E) should appear as a separate argument in the utility function alongside consumption (C):  $U = U(C, E)$ .

In the case of environmental goods, it is often argued that the shadow price (social value in some numeraire) of non-marketed environmental goods, such as habitat, wildlife or environmental quality in general, will be increasing over time as these services become economically scarcer relative to typical consumption goods, either as demand for these services increases, or their physical quantities or qualities decline (Fisher and Krutilla 1972; Sterner and Persson 2008; Drupp 2016). In general, the precise trajectory of this increase in the shadow price will depend on three factors: 1) the rate of physical growth of the environmental goods in question, and hence how physically scarce they are becoming over time; 2) the substitutability of environmental goods with typical consumption goods, reflecting how difficult it is to maintain well-being; and 3) growth in consumption. Intuitively, an environmental good that is becoming scarcer and which is not easily substituted by other goods will face a sustained increase in its shadow price in the future.

There are two entirely equivalent ways of dealing with this issue in CBA. Weikard and Zhu (2005) has the most helpful exposition of this point. The first is to recognise that the increase in the shadow price reflects a change in relative prices of environmental goods compared to consumption goods. It is important therefore to and ensure that these increasing values are reflected in the shadow or "accounting" prices that are used to evaluate the benefits associated with environmental commodities in CBA. Once the values have been placed in terms of consumption, the present value of these suitably valued benefits (or costs) can then be calculated using the appropriate consumption discount rate. The general point about accounting for relative price changes is already embodied in the current Green Book Guidelines (HMT 2003, pp. 20-25), indicating that relative price concerns are well understood in CBA.

Equivalently, one could simply base future valuations of the environment on today's shadow prices, assume they remain constant in real terms over time, and reflect the change in relative prices in a separate discount rate for environmental goods. This is the essence of "dual" discounting. While the mechanics and emphasis of these two options differ, and one or other approach may be preferred for procedural reasons, done properly, the practical outcome, if correctly calibrated, is the same in each case.



In the case of environmental goods and resources, the question remains: what determines the increase in scarcity and the change in the shadow price over time? Or equivalently, what factors determine the environmental discount rate? The following section elaborates on the simple economics of dual discounting to shed some light on these questions.

### 1.1.1. Dual discounting and relative prices: simple theoretical results

The theory behind dual discounting or relative prices is shown more formally in the Appendix to this report. On the question of how to determine the change in the relative prices with which to evaluate environmental goods, two inputs are required, one theoretical and one empirical.

The simple theoretical approach starts by recognising that social welfare may depend explicitly on environmental goods or stocks. So instead of the standard felicity function dependent purely on consumption,  $U = U(C)$  which leads to the standard Ramsey formulation of the social rate of time preference:  $S RTP = \delta + \eta g$ , instantaneous felicity is given by  $U = U(C, E)$  which explicitly depends on the environment,  $E$ , which is not perfectly substitutable with consumption. Given there are two arguments to the felicity function, each measuring different quantities, each has its own social rate of time preference. Appendix 1 shows that the discount rates for consumption and environment in this case become:

$$\rho_C(t) = \delta + \eta_{CC} g_C + \eta_{CE} g_E \quad (2)$$

$$\rho_E(t) = \delta + \eta_{EE} g_E + \eta_{EC} g_C \quad (3)$$

where the subscripts indicate the order of differentiation in relation to  $E$  and  $C$ .<sup>2</sup> Both (2) and (3) are recognisable as simple extensions to the standard Ramsey Rule for consumption,  $\rho_C$ , and environment,  $\rho_E$ . In principle, practical application of these discount rates simply requires empirical or expert estimates of the parameters of equations (2) and (3).<sup>3</sup>

Whereas there are many estimates for the elasticity of the marginal utility of consumption,  $\eta_{CC}$ , (see e.g. Groom and Maddison 2017) the elasticity of marginal utility of the environment,  $\eta_{EE}$ , and the cross elasticity parameters ( $\eta_{CE}, \eta_{EC}$ ) are less well understood.<sup>4</sup> In the absence of empirical estimates, additional theoretical assumptions are typically used to simplify the analysis.

---

<sup>2</sup>  $\eta_{xy} = - \frac{\partial^2 U / \partial x \partial y}{\partial U / \partial x}$

<sup>3</sup> For an extension of these principles to cases in which benefits and costs are uncertain, see Gollier (2010).

<sup>4</sup> Groom and Venmanns (2017) use experiments to try and estimate these parameters, but the work is highly preliminary. Willingness to pay studies have provided some evidence on these parameters also, see e.g. Drupp (2016).

Appendix 1 shows that if it is assumed that utility takes the constant elasticity of substitution (CES) form, the difference between consumption and environmental discount rates, respectively  $\rho_C$  and  $\rho_E$ , can be reduced to a very simple formula:

$$\rho_C - \rho_E = \frac{1}{\sigma}(g_C - g_E) \quad (4)$$

Where  $g_C$  is the growth rate of consumption goods,  $g_E$  is the growth rate of environmental goods (typically per capita, annualised), and  $\sigma$  is the elasticity of substitution between environmental and consumption goods.

Equation (4) is the difference between the consumption and environmental discount rates. Yet Weikard and Zhu (2005) show that this difference also reflects the rate at which the relative prices between environment and consumption goods should change if the relative pricing approach, rather than the dual discounting approach, to environmental scarcity is taking. It is the ‘inflation’ term for (real) environmental values.

Specifically, in Equation (4) if  $\sigma \rightarrow \infty$ , environment and consumption are perfectly substitutable and there is no difference in the dual discount rates, and relative prices remain constant. Essentially the environment becomes a regular part of the consumption bundle and has no special impact on social welfare. If  $\sigma = 0$ , then no amount of additional consumption goods can compensate for *any* loss in the environmental goods. While these two cases represent the extremes, they are probably not unrealistic extremes for certain types of natural capital stocks or flows. There are many cases which are likely to lie in between these extremes though.

The case where  $\sigma = 0$  can be thought of as a manifestation of *strong sustainability* in the parlance of sustainable economic development (e.g. Neumayer 2010; Ten Brink 2010; Drupp 2016). Inversely,  $\sigma > 0$  is a manifestation of weak sustainability, in which some substitutability is possible to maintain utility.

### 1.1.2. Advice on dual discounting: part 1

Equation (4) is expressed in terms of differences in the consumption and environmental discount rates, but as discussed this simply reflects the rate of change in the relative prices of consumption and environment. In terms of implementation, our preference would be, wherever possible, to reflect these changes in relative prices in the valuation of environmental goods, rather than taking the dual discounting approach. This has the benefit of not necessitating a large departure from the Green Book discounting guidelines, and follows the Green Book guidance on accounting for relative price changes. Indeed, this is precisely the approach that the Dutch government took in their recent review of CBA practices (Discontovoet 2016).

There are two caveats to this advice. As shown in the Appendix, the STP for consumption when environmental goods are taken into account separately, and utility

is given by  $U = U(C, E)$ , differs from the standard case shown in the Green Book where social welfare is assumed to be:  $U = U(C)$ . Strictly speaking the relative pricing approach that we recommend would require the eventual consumption units to be discounted using  $\rho_C$  rather than  $\rho = 3.5\%$  from the standard Ramsey Rule. Some work would have to be undertaken to see how much this adjustment would change outcomes.

Second, when substitutability becomes very limited, and  $\sigma \rightarrow \infty$ , future valuations are likely to be very high. Here more introspection about the substitutability, irreversibility and the extent to which environmental assets are somehow essential, should be brought to bear on decision-making and valuation. Some implications of substitutability are discussed in the following section, but Drupp (2016) points out that the elasticity of substitutability can be calculated using the inverse of the elasticity of marginal willingness to pay for the environment, estimates of which range between 0.5 and 5, depending on the resource in question and the empirical methods used.

### 1.1.3. Dual discounting, substitution, thresholds and strong sustainability

In a recent paper, Drupp (2016) introduces subsistence requirements for environmental stocks into the Weikard and Zhu (2005)/Hoel and Sterner (2007) framework. This minimum bound subsistence requirement on environmental stocks can be understood as reflecting an environmental threshold below which ecosystem services go to zero, or environmental quality disappears. More generally, the threshold can be thought of a strong sustainability constraint, or a level of *critical natural capital* below which the services that contribute to well-being become miniscule. The framework is adapted simply as follows.

First, the CES utility function takes the form:

$$U(C, E) = \frac{1}{(1-\eta)} \left[ (1-\gamma)C^{1-\frac{1}{\sigma}} + \gamma(E - \bar{E})^{1-\frac{1}{\sigma}} \right]^{\frac{(1-\eta)\sigma}{\sigma-1}} \quad (5)$$

where  $\bar{E}$  reflects threshold/subsistence levels of the environment. With this simple analytical addition, the difference in the dual discount rates becomes in (4) rather becomes:

$$\rho_C - \rho_E = \Delta\rho_t = \frac{1}{\sigma} \left( g_C - \left( \frac{E_t}{E_t - \bar{E}} \right) g_E \right) \quad (6)$$

The implications for the dual discount rates are clear. The inclusion of a threshold/strong sustainability means that the difference between the dual discount rates depends on the change in the environment good over time,  $E_t$ .

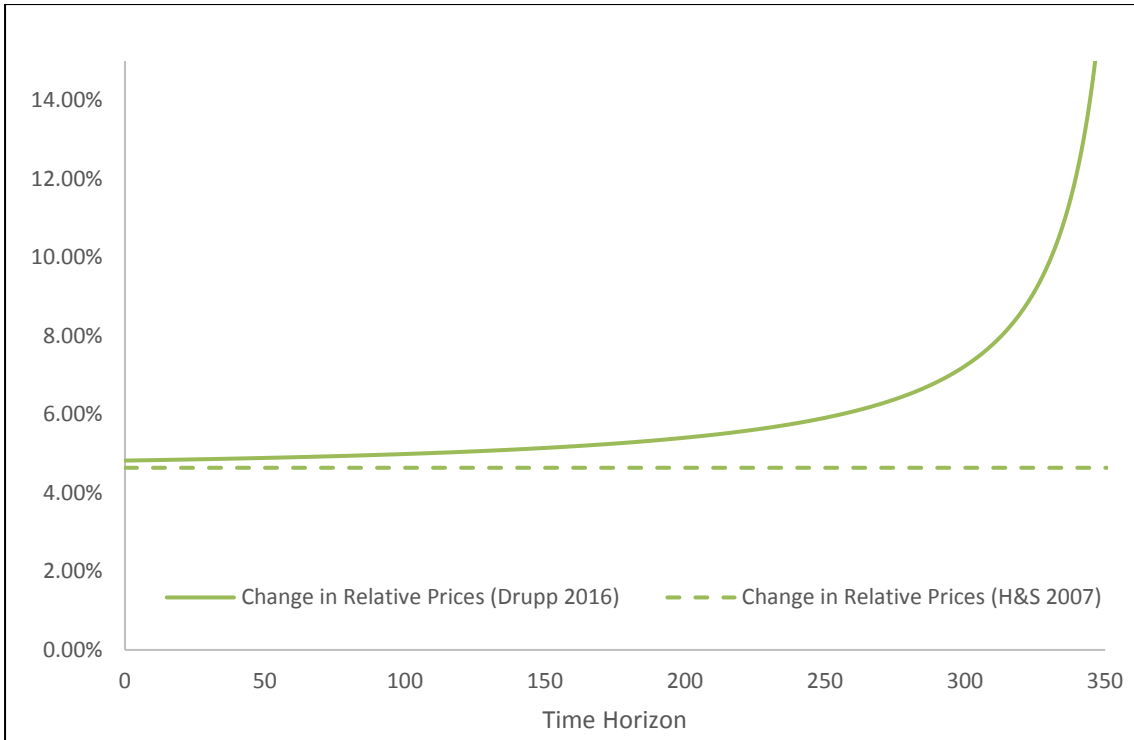


Figure 1. The difference in discount rates for consumption and environment (%):  $\Delta\rho_t = \rho_C - \rho_E$ , for Equation (4) (Hoel and Sterner 2007) and Equation (6) (Drupp 2016).

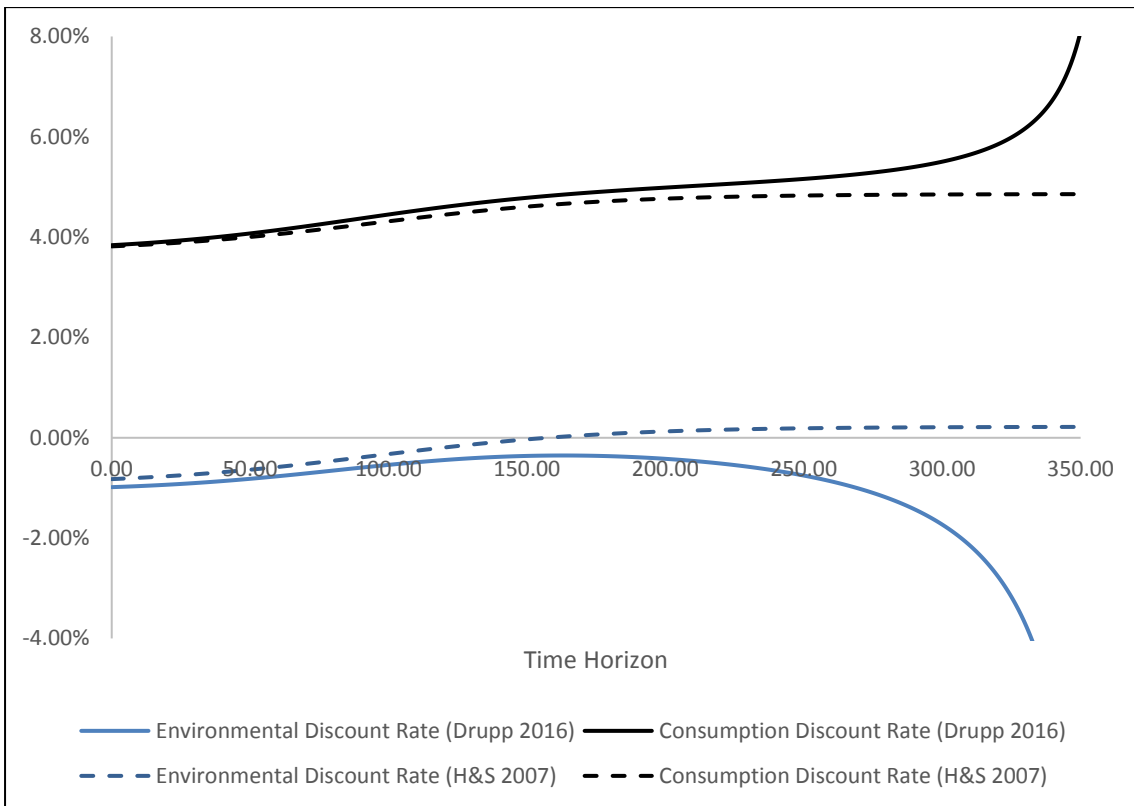


Figure 2. Consumption and Environmental Discount Rates (%) Hoel and Sterner (2007) and Drupp (2016): Parameters of Equation (5):  $\delta = 1\%$ ,  $\eta = 1.5$ ,  $\sigma = 0.5$ ,  $\gamma = 0.1$ ,  $E_0 = C_0 = 1$ , while:  $\bar{E} = 0.15$ ,  $g_C = 1.8\%$ ,  $g_E = -0.52\%$ .

To see this suppose that in both (4) and (6) the expected growth of consumption,  $g_C$ , and environment,  $g_E$ , are both constant, and remember also that the parameter  $\sigma$  is also constant. In equation (4) this means that the difference between the dual discount rates is a constant percentage, as in Example 1. However, in equation (5) the term  $E_t / (E_t - \bar{E})$  will approach unity if  $E_t$  is growing ( $g_E > 0$ ), and will approach infinity as  $E_t$  declines ( $g_E < 0$ ) to  $\bar{E}$ , the threshold value. In the latter case this means that the environmental discount rate would become very low indeed relative to the consumption discount rate, and possibly negative.

Interpreted in terms of relative prices, equation (6) states that in a CES framework, with an environmental threshold the relative price of the environment grows at increasing rates as the environment approaches the threshold  $\bar{E}$ . Figure 1 provides an example of this outcome.

Figure 1 shows the difference between consumption and environmental discount rates for the standard Hoel and Sterner (2007)/Sterner and Persson (2008) approach shown in equation (4), and the Drupp (2016) approach with an environmental threshold shown in equation (5).<sup>5</sup> In the latter case the simulations assume that consumption growth is positive, while growth of the environmental good is negative, and approaching the threshold. This leads to a rapid increase in the growth of the relative price for the environmental good for time horizons of 200 years or more, compared to the constant level of growth of relative prices in the Hoel and Sterner (2007) case. Figure 2 shows how this relative price effect is manifested in the discount rates.

Figure 2 illustrates the impact that approaching the threshold has on the discount rates appropriate for different time horizons, in this deterministic case, compared to Hoel and Sterner (2007) who ignore the threshold effect. There are two important impacts: i) the environmental discount rate declines rapidly for horizons of 300+ years; ii) the consumption discount rate increases rapidly at the same time horizon. This reflects the increasing value of an additional unit of environmental goods in the future and the declining value of consumption goods when we are approaching an environmental threshold.

Both Figures 1 and 2 use the parameterisation of Hoel and Sterner (2007). The horizons over which the threshold has any practical effect are long and way beyond the time horizons that have been the focus of the ONS so far. Nevertheless, Hoel and Sterner (2007) were interested in climate change, and ecosystem services in general. The relevant threshold and time horizon will depend on the resource in question.

#### 1.1.4. Advice on dual discounting: part 2

As before, even when thresholds and subsistence requirements are present, rather than using dual discount rates for consumption and environmental goods, it is

---

<sup>5</sup> For further scenarios see Drupp (2016).

recommended that the shadow prices used to value the environmental goods over time are adjusted to reflect scarcity. The only difference here is that the shadow prices need to reflect the thresholds or subsistence requirements associated with the environmental good in question.

The analysis is extremely stylised in the previous section, and while subsistence levels of consumption might be easy to identify, subsistence levels and thresholds associated with environmental goods and ecosystems are less easy to identify. Each environmental resource will be different and it might be difficult to establish the level of, e.g., groundwater extraction at which the threshold is met, or the exploitation of an ecosystem that eventually removes all eco-system services. Yet, the valuation of streams of environmental services over time ought to account for the ecological facets of the resource in question. This comes in addition to the need to identify the secular growth rate of environmental resources, and the elasticity of substitutability as discussed above. These are stringent empirical requirements, and best guesses will be required in general.

In the “Advice on dual discounting: part 1”, it was noted that the advice to use relative prices is subject to one major caveat. For emphasis this caveat is repeated here. Equations (2) and (3) show that the dual discounting approach requires both the consumption discount rate and the environmental discount rate needs to be augmented. Together these discount rates capture the fact that changes in relative prices come from two sources: 1) changing absolute scarcity of the environmental good; and, 2) changing absolute scarcity of the consumption good. In the example illustrated in Figures 1 and 2, environmental goods become more scarce, and consumption goods less scarce. The reduced scarcity of consumption goods comes from two sources: 1) growth in consumption; **and**, 2) the physical scarcity the environmental good. The augmented discount rate in Equation (2) captures both of these effects.

In sum, it would be a mistake to simply use the formulas in Equations (4) or (6) to adjust the shadow price of environmental resources and then to use the standard Ramsey Rule ( $\delta + \eta g_c$ ) to discount the consumption equivalents. Example 1 shows that doing so would **not** maintain the equivalence of dual discounting and adjustments to relative prices.

**Example 1: The importance of augmenting the Social Rate of Time Preference for Consumption in the Dual Discounting Framework.**

Table 1 shows the appraisal of a simple project which has costs and benefits in terms of consumption, C, and environmental goods, E (Columns 2 and 3 resp.) and the background parameters are taken from the environmental threshold example in Figures 1 and 2. To evaluate this project two approaches are taken. **First**, a dual discounting approach is taken, in which the consumption discount rate (Equation (A4) from Appendix 3.A1) is used to calculate present value of the consumption impact of the project, and an environmental discount rate (Equation (A6) from Appendix 3.A2) is used to calculate the present value of the environmental impact. **Second**, a relative price adjustment approach is taken whereby the shadow values of the environmental goods are inflated using Equation (6) (Column 4) and then discounted using the consumption discount rate (Equation (A5)). The NPV of these two exercises are shown at the bottom of Table 1. The fact that the NPVs at the bottom of columns 2 and 4 are equal shows the equivalence of these two approaches in project appraisal.

Year	Project			Discount Rates	
	E	C	E (Relative Price Adjusted)	Environmental Discount Rate	Consumption Discount Rate
0	0	-1	0.00	-0.98%	3.84%
10	0	-1	0.00	-0.96%	3.87%
20	0	-1	0.00	-0.94%	3.91%
30	-2	-1	-8.60	-0.90%	3.96%
40	-2	-1	-14.06	-0.86%	4.01%
50	0.75	-1	8.65	-0.82%	4.07%
60	0.75	2	14.26	-0.77%	4.14%
70	0.75	2	23.59	-0.71%	4.22%
80	0.75	2	39.22	-0.65%	4.29%
90	0.75	2	65.57	-0.60%	4.37%
100	0.75	2	110.31	-0.54%	4.45%
NPV	<b>1.94<sup>a</sup></b>	<b>-2.37</b>	<b>1.94<sup>b</sup></b>		
			<b>NPV (Dual Discounting)</b>		<b>-0.44</b>
			<b>NPV (Relative Price)</b>		<b>-0.44</b>
			<b>NPV (RP) Standard Ramsey</b>		<b>3.46<sup>c</sup></b>

Table 1. Project Appraisal using Dual Discount Rates or Relative Price Adjustments. <sup>a</sup> Calculated using the environmental discount rate in column 5. <sup>b</sup> Calculated using adjusted relative prices in column 4 and the consumption discount rate in column 6. <sup>c</sup> Calculated using the adjusted prices and the standard Ramsey Rule ( $\delta + \eta g_C$ ) with parameters as in Figure 2.

A **third** approach is also taken, which is a variant of the second. Now we suppose that the adjusted environmental values are discounted using the standard Ramsey Rule ( $\delta + \eta g_C$ ). The NPV calculation in this case is shown at the bottom of Table 1. This approach severely overestimates the NPV of the project because it fails to adjust the consumption discount rate for the increased scarcity of the environmental good, and the term structure of the consumption discount rate shown in Figure 2. This approach would only coincide with approaches 1 and 2 under special circumstances, highlighting the need for care in calibrating the discount rate when relative prices of the environment are changing.

## 1.2 Natural capital valuation: Estimating the present value of the stream of benefits

The approach to discounting in the previous sections can be used in CBA to evaluate marginal projects which have implications for consumption opportunities over time, and implications for the flow or stock of environmental assets. What dual discounting alerts us to is the need to be extremely careful in calculating the shadow or “accounting” prices with which environmental stocks and flows are valued, and to use prices that reflect the underlying determinants of the scarcity of those resources.

A more important reason to be interested in valuing natural capital is the creation of comprehensive/inclusive wealth accounts. Comprehensive wealth accounts value all elements of the nation’s wealth by valuing national assets (physical capital, human capital, natural capital, etc.). This too is an ambition of the Natural Capital Committee (Mayer 2014). By building up a picture of the comprehensive/inclusive wealth of a country it becomes possible to make more precise statements about economic performance and, in particular, whether or not a country is likely to be on a sustainable economic development path. Theoretical results have shown that if Genuine Savings/Investment (aggregate savings/investment net of resource depletion, depreciation and degradation) is consistently positive, and the measure of comprehensive wealth is non-declining over time, then a sustainable path of well-being is possible (see, e.g. Hamilton and Hartwick 2014).

To this end, Fenichel and Abbot (2014) provide a capital theoretic framework for the evaluation natural resource stocks which embodies all of the relevant determining factors of the value of a resource stock. These include the ecological factors discussed above, the production/demand side factors of the resource, the discount rate and, importantly, the institutional/regulatory determinants of resource value. In the end, Fenichel and Abbot (2014) provide a workable (albeit data and computationally intensive) valuation methodology for natural resource stocks, along with worked examples: Groundwater and fisheries.

### 1.2.1. The basic F&A formula for the valuation of natural capital

Many previous studies defined the asset value of natural capital simply as the present value of the marginal net benefits arising from an additional unit of the resource stock (Barbier 2011). The basic idea was that if a natural resource provides an annual flow of net benefits given by some function  $W(H_t(S_t), S_t)$  which depends on stock of the resource,  $S_t$ , and the harvesting rule  $H_t(S_t)$  (e.g. tons of fish caught under open access/common property/regulated fishery, or flow of ecosystem services), and the instantaneous *marginal* value of the stock is given by  $W_s(H_t(S_t), S_t)$ , then the asset value of this marginal unit of the stock is equal to the present value of these marginal net benefits over the relevant time horizon. Typically, the marginal value is often thought of as an annuity (a constant flow in each time period), so if the time horizon is infinite the marginal asset value,  $p$ , is given by:



$$p = \frac{W_s(H(S), S)}{\rho} \quad (7)$$

which is a standard infinite horizon annuity value formula, where  $\rho$  is the appropriate social discount rate. However, Fenichel and Abbot (2014) note that this formula is incomplete for all but a limited class of resources. The more general case stems from methods used in natural resource economics and dynamic optimisation which take into account the growth dynamics of the natural resource, and economic factors beyond those contained in Equation (7). In sum, Fenichel and Abbot (2014) show that the marginal asset value, or shadow price of the resource stock,  $p$ , is given by:

$$p = \frac{W_s(H(S), S) + \dot{p}}{\rho - (G_s(S) - f_s(S, H(S)))} \quad (8)$$

The numerator of this expression for the marginal stock value in (8) contains two components: i)  $W_s(H_t(S_t), S_t)$ : the instantaneous flow of marginal benefits from the natural resource; ii)  $\dot{p}$ : future changes in the shadow price, which can be understood as the expected *capital gain* reflecting changes in the economic scarcity of the resource.

The denominator, essentially a **net discount rate**, contains 3 components: i) the social discount rate appropriate for the units in which  $W_s(H_t(S_t), S_t)$  is measured, e.g. a consumption discount rate; ii)  $G_s(S)$ : the marginal growth rate of the natural resource stock, e.g. the change in the recruitment of fish, or the physical growth of trees as the stock changes; iii)  $f_s(S, H(S))$ : this is the marginal effect of the stock on the impact of harvest decisions on the growth of the stock. So, in addition to the discount rate, there are two other reasons why we might not want to hold the asset any longer, and would prefer to harvest: i) the growth rate of the stock; and, ii) changes in the marginal impact of harvest on the growth of the stock. For instance, if  $G_s > 0$  then the intrinsic rate of return of the resource is increasing with the stock/asset holding, and the marginal shadow value of the stock is increased in present value terms. If  $f_s(S, H(S)) > 0$  then additional stocks increase the negative impact of harvesting decisions on the growth rate of the stock, hence reducing the present value of the in situ stock. Taken together, the denominator reflects what F&A refer to as “**net rate of capital productivity**” (Fenichel and Abbot 2014, p8).

In relation to discounting, one important point to take from this analysis is that in valuing natural capital it is not sufficient to discount the expected instantaneous gains using the social discount rate,  $\rho$ , be it a consumption discount rate or whatever. In addition the intrinsic growth aspects of the resource itself ( $G_s$ ), and the impact of human activity on growth ( $f_s$ ) must be taken into account. It is this **net discount rate (net capital productivity)**, that determines the value of natural capital.

Another important point is that the economic programme, which determines the relationship between the harvest level,  $H$ , and the stock,  $S$ , and is embodied in the function  $H(S)$ , is also an important determinant of natural capital values. This programme determines the valuation since it may reflect open access arrangements, or well regulated scenarios. Finally, the pricing formula on equation (8) also holds in non-optimal scenarios, and so is quite general.

The discussion of (8) has been extremely abstract. In the following section a practical interpretation is given following Fenichel and Abbot (2014).

### 1.2.2 Valuing natural capital: A fisheries example

To overcome the abstraction above, Example 2 provides a well-defined example of how Equation (8) could be interpreted in practice in the context of fisheries. In order to do this, the components of Equation (8) must be defined. Table 2 has the details.

Description	Function in Equation (8)	Fisheries Interpretation	Details
<b>Instantaneous value of resource use</b>	$W(S, H(S))$	$W = m * h - cE$	$h$ = harvest, $E$ = effort, $c$ = costs, $m$ = price. $W$ is measured in cash terms.
<b>Growth Function of Resource</b>	$G(S)$	$G(S) = rS(1 - S / K)$	$r$ = intrinsic growth of stock, $K$ = carrying capacity. Logistic growth function. Estimated empirically conditional on existing institutions using time series data. From this $G_s(S)$ can be obtained.
<b>Change in Resource Stock</b>	$f(S, H(S))$	$f(.) = H(S)$	The stock declines by the amount harvested. $f(.) = H(S)$ . $\Rightarrow f_s(.) = H_s(S)$
<b>Discount Rate</b>	$\delta$	$\delta$	The appropriate rate for the cash value of the fishery
<b>“Economic programme”</b>	$H(S)$	$H(S) = \alpha S^\gamma$	Empirical relationship estimated conditional on behaviour in an un-regulated fishery.

Table 2. The functions underlying the asset value of a fishery (Source: Fenichel and Abbot 2014).

Table 2 shows precisely that the valuation of natural capital assets, while highly dependent on the discount rate,  $\delta$ , it also depends on a whole host of other characteristics of the problem at hand. For fisheries this depends on the ecology of the fishery itself,  $G(S)$ . It also depends on the economic and institutional factors such as the harvesting technology (fishing boats, radar, satellite images, nets), and,

importantly, the institutions under which the fishery is regulated (open access, common property, Individual Tradable Quotas, etc.). Social discounting is an important part of this calculation, and the asset values will be sensitive to the choice of discount rate, but there are other elements of the return to natural capital that also need to be scrutinised.

### **1.3. Conclusion on environmental discounting**

This section has argued that the appropriate discount rate for environmental commodities depends upon the context in which environmental valuation is being undertaken. In CBA of marginal projects or the valuation of streams of benefits from environmental assets, there are arguments for using good specific dual discount rates that embody the changing relative scarcity of the environmental resource in question. Our recommendation is to rather reflect changing scarcity of environmental resources in the shadow prices for those resources, convert all benefit streams into consumption and discount them using an appropriate consumption discount rate or Social Rate of Time Preference.

The benefit of this approach is that each environmental resource is likely to have different characteristics: substitutability, thresholds, subsistence levels, and other ecological characteristics, which would warrant different discount rates being deployed for different environmental values. To fix a menu of discount rates for different environmental goods might be practically helpful, but administratively difficult in our view. In our opinion it would be better to organise specific valuation studies oriented to estimating shadow prices, and argue about the discount separately.

This general point is illustrated in recent work that shows how to value natural capital assets, say as part of a programme comprehensive/inclusive wealth measurement. What this theory shows is that the shadow price of a stock of natural capital should be discounted using a net discount rate which reflects all the ecological and economic determinants of the assets rate of return. Normally this “net rate of natural capital productivity” contains but is not equal to what is typically understood to be the social discount rate. Once again, this is an argument for separating out issues of social discounting from estimating expected changes in scarcity values (shadow prices) and the ecological facets of the environmental resource in question.

### **1.4. Discounting for human capital**

Of all the elements of national wealth (physical, human, environmental), it is human capital that is the most important in terms of magnitude. The trajectory of human capital values is a key determinant of future growth prospects and the growth of well-being in society. Human capital is important from the perspective of obtaining a measure of comprehensive wealth and hence developing measures of economic performance that move beyond the typical flow measures such as GDP, which are only loosely related

to well-being, and have no bearing on the durability/sustainability of growth (Weitzman 1976; Stiglitz et al 2009).

Although human capital can be measured in terms of educational attainment, years of schooling and the likelihood of obtaining employment, monetary measures are required in order for human capital to be comparable to other components of a comprehensive wealth measure. Depending on the approach taken, as with other measures of wealth, the monetisation of human capital stocks usually reflect the present value of a stream of benefits over a specified time horizon. Calculating the present value requires a discount rate to be applied to these benefits, and in this section we discuss the appropriate discount rate for this purpose. First, we briefly summarise the way in which human capital has been evaluated and estimated in the past.

#### *1.4.1 The valuation of human capital*

There are three typical approaches to valuing human capital (ONS 2015):

- 1) Educational attainment: years of schooling, levels of attainment, etc.
- 2) The input or cost based approach: adding up the cost of all educational inputs undertaken by an individual.
- 3) Output or Income based approaches: using the discounting value of earnings over the expected lifetime of the individual

The ONS use method 3 since this is widely seen as the most reliable and accurate approach in that it is forward looking, provides monetary measures and data is available to estimate this measure.

#### *1.4.2 The output/income approach to human capital*

The standard approach to measuring human capital in the national accounts, so that the value can be compared year on year, stems from the Jorgensen-Fraumeni framework (e.g. Jorgensen and Fraumeni 1989, 1992). The J-F approach evaluates human capital based on the streams of earnings that are expected for a person of a given age, gender and educational background. The value of human capital is calculated as the present value of earnings over a time horizon which depends on the expected lifetime. Typically, market wages are used to estimate earnings, coupled with some estimate of expected earnings growth over the relevant time horizon. The latter depends on the level of education and the likelihood of obtaining employment at various stages of transition from school to labour market. Expected lifetime is calculated using an estimate of the likelihood of mortality.

There is a large literature on the estimation of human capital, how it can be calculated by gender, and how individual measures can be aggregated to reflect the demographic characteristics of the country in question and obtain a total asset value. The calculation

of human capital values starts with an evaluation of lifetime earnings for individuals of different ages, and then aggregates across these estimates to reflect the demographic composition of the country. For instance, for each age  $a$ , and education level  $e$ , lifetime incomes at time  $t$  are typically given by (Fraumeni et al 2015):<sup>6</sup>

$$LE_{75} = 0 \quad \text{if } age = 75, e = \text{anything} \quad (1)$$

$$LE_{t,a,e} = y_{t,a,e} + \left[ \frac{(1+g_w)}{(1+\rho)} sr_{t,a,e} \right] LE_{t,a+1,e} \quad \text{if } age = 35 - 74, education =: e \quad (2)$$

$$LE_{t,a,e} = y_{t,a,e} + senr_{t,a,e} \left[ \frac{(1+g_w)}{(1+\rho)} sr_{t,a,e+1} \right] LE_{t,a+1,e+1} + (1-senr_{t,a,e}) \left[ \frac{(1+g_w)}{(1+\rho)} sr_{t,a,e} \right] LE_{t,a+1,e} \quad \text{if } age = 14 - 34, education =: e \quad (3)$$

Where:

$y_{a,e}$  = average yearly income for person of age  $a$  and education  $e$  at time  $t$

$\rho$  = the discount rate;

$g_w$  = the growth rate of wages;

$sr_{t,a,e}$  = the survival rate at time  $t$ , age  $a$ , and education level  $e$ ;

$senr_{t,a,e}$  = the school enrolment rate at at time  $t$ , age  $a$ , and education level  $e$ .

Equation (2) is the sum of today's annual wage and the discounted value of all expected lifetime earnings at age  $a + 1$  under the assumption that no additional education is undertaken. Equation (3) is similar, except the likelihood (frequency in society) of being in education, and having a different future lifetime earnings trajectory as a consequence, is taken into account. The overall value of human capital is then calculated backwards, as it were, starting at the end of the lifetime, and summing up over the history of the agent accounting for educational and implicitly occupational choices.

For illustrative purposes, the summation in Equation (2) is undertaken iteratively for an individual at time  $t$ , with education level  $e$ , and age at time  $t$  of  $a_t$ . With income at

---

<sup>6</sup> The dependence on gender is suppressed here for notational simplicity, but is a feature of the theory, the ONS calculations and the use of the J-F framework in the US national accounts.

time  $t$ , age  $a$ , education  $e$  is given by  $y_{t,a,e}$ , this leads to the following formula for the total value of human capital:<sup>7</sup>

$$LE_{t,a,e} = \underbrace{\sum_{a=a_t}^A \frac{y_{t,a,e} (1+g_w)^{a-a_t}}{(1+\rho)^{a-a_t}}}_{\text{Present Value of Lifetime Earnings}} sr_{t,a-1,e} \quad (4)$$

The value of the aggregate human in the economy is then calculated by summing up the present value of lifetime earnings across the population over the distributions of age, education (and gender). ONS (2015) shows how this is done in practice.

Just as with the estimate of natural capital values discussed above, the income approach to the valuation of human capital is rooted in capital theory. If one calculates the marginal value of human capital, that is the shadow price of the human capital stock, the formula that emerges is entirely analogous to the one described for natural capital as derived by Fenichel and Abbot (2014) and Jorgensen (1963).

As with any approach, there are questions over the accuracy of such capital valuations. For instance, there is concern about whether market prices properly reflect the true marginal productivity of an individual, or whether there are externalities that need to be taken into account. Whether rigidities or institutional structures in the labour market, such as unions or minimum wages, distort the market is another question. Nonetheless, the J-F framework, evaluated with market prices is a typical approach (Fraumeni et al 2015).

#### 1.4.3. Discounting human capital

The approach taken to discounting human capital by the ONS is to use the Green Book discounting guidelines (ONS 2015, ch4). This means that in Equation (4) the values of annual income  $y_{t,a,e}$  are measured in terms of consumption and the discount rate  $\rho$  is the Social Rate of Time Preference which is reflected by the Ramsey rule that was discussed in previous sections of this report:  $SRTP = \delta + \eta g$ , with the declining term structure that begins at time horizons beyond 30 years. This is the approach that we would recommend.

Yet, there are some other issues that can arise when thinking about discounting human capital which relate not to the discount rate itself, but to the growth of wages and the hazards associated with survival rates.

---

<sup>7</sup> The full formula is:  $V^H = \underbrace{\left( \sum_{a=a_t}^A \frac{y_{t,a,e} (1+g_w)^t}{(1+\rho)^t} sr_{t,a-1,e} \right)}_{\text{Present Value of Lifetime Earnings}} + \underbrace{\frac{(1+g_w)^{A+1}}{(1+\rho)^{A+1}} sr_{t,A+1,e} LE_{t,A+1,e}}_{\text{Terminal Value } \approx 0}$  but the terminal

values are typically ignored or considered to be negligible.

In the spirit of dual discounting, one could re-write equation (4) as the following expression, in which wages are discounted using a ‘net discount rate’ which reflects the growth of wages over time, and the hazard rate associated with survival probabilities:

$$LE_{t,a,e} \approx \sum_{a=a_t}^A \frac{y_{t,a,e}}{(1 + \rho^*)^{a-a_t}} \quad (5)$$

Where  $\rho^* = \rho - g_w + \phi$ , can be thought of as a net discount rate which takes into account the growth rate of wages,  $g_w$ , and the hazard associated with the survival probability at age  $a$ .<sup>8</sup> This shows how the survival rates and the growth of wages can be understood in the similar way as the discount rate. Doing this raises two important issues for discounting: i) Double counting; and, ii) Double discounting.

**Potential problem 1:** *Dual discounting and the growth of wages: is this double counting?*

Consider the ONS case where the discount rate follows the Green Book guidance. In this case the net discount rate becomes:

$$\begin{aligned} \rho^* &= \delta + \eta g - g_w + \phi \\ \text{if } g &= g_w \Rightarrow \delta + (\eta - 1)g + \phi \\ \text{if } \eta &= 1 \Rightarrow \delta + \phi \end{aligned} \quad (6)$$

It is typical in applied work, it seems, to assume that wage growth is equal to growth in GDP, hence the first simplification. In the current Green Book guidelines  $\eta = 1$ , hence the second simplification.

However, this raises the question of double-counting in the growth component since wage growth reflects growth in human capital itself. In the general formulation of lifetime earnings, it would be better to value wages at each point in time, and assume that the wage growth component is net of the growth of human capital itself. Essentially, the growth component should only reflect those parts of productivity which affect human capital but which stem from technological change or changes in physical capital stocks. This is an empirical question of how to calibrate equation (4) and its more general counterparts.

---

<sup>8</sup> As a very rough approximation, just to make the point, the survival rate,  $sr_{t,a,e}$  is essentially a probability of living for  $a$  years or longer. One could model this probability as a survivor function of a hazard rate  $\phi$  as follows:

$P(a \geq A) = \frac{1}{(1 + \phi)^a}$  as a rate. The rate  $\phi$  would be the average hazard rate. The inclusion of this hazard rate in the net discount rate is another approximation.

## **Potential problem 2: Double discounting of survival rates**

In the ONS case, the net discount rate reduces to:

$$\rho^* = \delta + \phi$$

Where  $\delta$  is the utility discount rate, and  $\phi$  is the hazard rate associated with the survival probability. In principle this reduction to the net discount rate focusses attention on the pure rate of time preference, and making sure that one gets this right for social discounting. In the case of the Green Book guidance, however, the utility discount rate contains two components: i)  $\bar{\delta}$ : the pure rate of time preference; and, ii)  $L$ : a generalised hazard:

$$\delta = \bar{\delta} + L \quad (7)$$

The generalised hazard has typically been calculated to reflect the risk of societal collapse or some other catastrophe at the societal level. The empirical work that is used to calculate this hazard often uses mortality rates as data (HMT 2003). For this reason it is quite likely that there is a significant overlap of this societal hazard component,  $L$ , and the individual survival hazard. This is an open question requiring future research.

### **1.5. Conclusion to discounting for human capital**

It would be tempting to approach the estimation of human capital using the net discount rate derived in the text. However, as with dual discounting, we do not recommend it since including these elements in the discount rate tends to hide some of the moving parts of the analysis, namely, assumptions concerning wage growth. We rather recommend that the current approach, following the Green Book guidelines, should continue, with assumptions on wage growth both revisited (particularly in relation to the double counting issue), and made explicit.

However, thinking about the valuation of human capital through the lens of the net discount rate does illustrate some potential problems with the current approach to evaluating human capital: 1) double counting; 2) double discounting, which have the opposite effect on the valuation of human capital. So, while the basic guidance is to carry on with the approach outlined in ONS (2015), Further research is required to establish the extent to which 1) and 2) are genuine problems.

### **1.A. Appendix**

#### **1.A.1. Dual discounting in theory<sup>9</sup>**

Suppose that instantaneous utility depends on consumption  $C$  and a stock of

---

<sup>9</sup> This is reproduced from the [Stage 1 report](#) for convenience.



environmental goods, E. Intertemporal Social Welfare is then given by:

$$W = \int_0^{\infty} U(C_t, E_t) \exp(-\delta t) dt$$

where  $\delta$  is the utility discount rate (which here does not differ between environmental and consumption goods). There is no uncertainty. The social discount factor with which to “price” changes in the quantities of each of the arguments, consumption and environment, from the perspective of today ( $t = 0$ ) is given by:<sup>10</sup>

$$P_i(t, 0) = \frac{U_i(C_t, E_t)}{U_i(C_0, E_0)} \exp(-\delta t) \quad \text{for } i = C, E$$

The associated discount rates are given by the rate of change of this price over time. For C and E respectively this leads to two separate social discount rates:

$$\rho_C(t) = \delta - C \frac{U_{CC}(t)}{U_C} g_C - E \frac{U_{EC}(t)}{U_C} g_E \quad (\text{A1})$$

$$\rho_E(t) = \delta - E \frac{U_{EE}(t)}{U_E} g_E - C \frac{U_{CE}(t)}{U_E} g_C \quad (\text{A2})$$

This should be compared to the standard single good framework of Ramsey in which the social discount rate for consumption goods is simply:

$$\rho = \delta - C \frac{U_{CC}}{U_C} g_C$$

Which is usually written as  $\rho = \delta + \eta g$ . This is the typical framework for the analysis of dual (meaning separate) discounting of environmental benefits and costs on the one hand, and consumption goods on the other.

What this means is quantities of consumption (an index of all consumption of apples, oranges, etc. usually measured in money terms) should be discounted using  $\rho_C$ , and quantities of environmental goods (changes in air quality, or changes in benefit from forested areas or ecosystem services in general) should be discounted using  $\rho_E$ .

Weikard and Zhu (2005) show that the pricing and discounting approaches in Example 1 are equivalent from a welfare perspective. They show that the rate of change in the shadow price of the environment,  $p$ , is equivalent to the difference between A1 and A2. They do this by first reminding us that the shadow price for the environment in terms of consumption is given by:  $p = \frac{U_E}{U_C}$ , and that the rate of change of this is given

by:

<sup>10</sup> Adapted from (Traeger 2011, p 216).

$$\frac{\dot{p}}{p} = \frac{\dot{C}}{C}(\eta_{CC} + \eta_{EC}) + \frac{\dot{E}}{E}(\eta_{EE} + \eta_{CE}) \quad (\text{A3})$$

Which is just the difference between (A1) and (A2) with  $\eta_{ij} = -x_i \frac{U_{ij}}{U_i}$ . Following Hoel and Sterner (2007) suppose that preferences are Constant Elasticity of Substitution (CES):

$$U(C, E) = \frac{1}{(1-\eta)} \left[ (1-\gamma)C^{1-\frac{1}{\sigma}} + \gamma E^{1-\frac{1}{\sigma}} \right]^{\frac{(1-\eta)\sigma}{\sigma-1}}$$

then it can be shown that (Hoel and Sterner 2007, p 9-12 and Appendix):

$$\rho_C = \delta + \left[ (1-\gamma^*)\eta + \gamma^* \frac{1}{\sigma} \right] g_C + \left[ \gamma^* \left( \eta - \frac{1}{\sigma} \right) \right] g_H$$

$$\rho_E = \delta + \left[ (1-\gamma^*) \left( \eta - \frac{1}{\sigma} \right) \right] g_C + \left[ \gamma^* \eta + (1-\gamma^*) \frac{1}{\sigma} \right] g_H$$

and hence:

$$\rho_C - \rho_E = \frac{1}{\sigma} (g_C - g_H)$$

where  $\gamma^* = \gamma E^{1-\frac{1}{\sigma}} \left( (1-\gamma)C^{1-\frac{1}{\sigma}} + \gamma E^{1-\frac{1}{\sigma}} \right)^{-1}$ .

We can now say the following in the Hoel and Sterner framework:

- 1)  $\rho = \rho_C$  if either i)  $\gamma^* = 0$ ; ii)  $g_C = g_E$ ; or, iii)  $\eta\sigma = 1$ . This illustrates the importance of relative growth and substitutability in this analysis.
- 2)  $\rho_C = \rho_E$  if either i) ; or, ii)  $E$  and  $C$  are perfect substitutes, i.e.  $\sigma \rightarrow \infty$
- 3) If  $g_C > g_E$  and then will tend to 1 over time. This means that the limits of the two discount rates are:

$$\rho_C = \delta + \frac{1}{\sigma} (g_C - g_E) + \eta g_E$$

$$\rho_E = \delta + \eta g_E$$

This implies a term structure of social discount rates as consumption patterns change.

- 4) Substitutability is a key issue: all the results depend on  $\sigma$ : the elasticity of substitution between  $E$  and  $C$ .

### 1.A.2. Dual discounting with subsistence/thresholds

The starting point for Drupp (2016) is a welfare function which takes the following form as in Equation (5):

$$U(C, E) = \frac{1}{(1-\eta)} \left[ (1-\gamma)C^{1-\frac{1}{\sigma}} + \gamma(E-\bar{E})^{1-\frac{1}{\sigma}} \right]^{\frac{(1-\eta)\sigma}{\sigma-1}} \quad (\text{A4})$$

Using the same procedures as in Appendix 3.A1, this leads to good specific discount rates of the form:

$$\rho_C(t) = \delta + \frac{\gamma(1-\theta)(E_t - \bar{E})^\theta + (1-\gamma)\eta C_t^\theta}{\alpha(E_t - \bar{E})^\theta + (1-\gamma)C_t^\theta} g_C + \frac{\gamma E_t (E_t - \bar{E})^{\theta-1} (\gamma - (1-\theta))}{\alpha(E_t - \bar{E})^\theta + (1-\gamma)C_t^\theta} g_E \quad (\text{A5})$$

and

$$\rho_E(t) = \delta + \frac{E_t}{E_t - \bar{E}} \frac{\gamma\eta(E_t - \bar{E})^\theta + (1-\gamma)(1-\theta)C_t^\theta}{\alpha(E_t - \bar{E})^\theta + (1-\gamma)C_t^\theta} g_E + \frac{(1-\gamma)C_t^\theta (\gamma - (1-\theta))}{\alpha(E_t - \bar{E})^\theta + (1-\gamma)C_t^\theta} g_C \quad (\text{A6})$$

These two equations are used in Figures 1 and 2 and in Example 1 in the text.

## References

Arrow K, M Cropper, C Gollier, B Groom, G Heal, R Newell, W Nordhaus, R Pindyck, W Pizer, P Portney, T Sterner, R Tol, M Weitzman (2013). Determining benefits and costs for future generations. *Science*, 341: 349–50.

Baumgärtner S, AM Klein, D Thiel, K Winkler (2015). Ramsey discounting of ecosystem services. *Environmental and Resource Economics*, 61: 273-296.

Discontovoet (2016). Rapport Werkgroep Discontovoet, Government of the Netherlands:

<https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2015/11/13/rapport-werkgroep-discontovoet-2015-bijlage/rapport-werkgroep-discontovoet-2015-bijlage.pdf>

Drupp MA, MC Freeman, B Groom, F Nesje (2015). Discounting disentangled: An expert survey on the determinants of the long-term social discount rate. Grantham Research Institute Working Paper No.172. London School of Economics.

Drupp, M.A. (2016). Limits to Substitution between Ecosystem Services and Manufactured Goods and Implications for Social Discounting. Available as [SSRN Working Paper](#).

Ekins, P. (2003). Identifying critical natural capital: Conclusions about critical natural capital, *Ecological Economics* 44: 277-292.

Fenichel E P and Joshua K. Abbott (2014). [Natural Capital: From Metaphor to Measurement](#) *Journal of the Association of Environmental and Resource Economists* 1: 1-27

Fenichel E P and Joshua K. Abbott (2014). [Natural Capital: From Metaphor to Measurement](#) Journal of the Association of Environmental and Resource Economists 1: 1-27

Fisher, Anthony C., John V. Krutilla, and Charles J. Cicchetti. (1972). The Economics of Environmental Preservation: A Theoretical and Empirical Analysis. American Economic Review, 62: 605-19.

Fraumeni, B.M. and G. Liu, "Human Capital: Country Estimates Using Alternative Approaches," chapter 4 in UNU-IHDP and UNEP (2014). *Inclusive Wealth Report 2014, Measuring Progress Toward Sustainability*, Cambridge: Cambridge University Press, pp. 109- 122, 2014.

Fraumeni, B.M., Christian, M. S., and Samuels, J. D. (2015). The Accumulation of Human and Nonhuman Capital, Revisited, Paper Prepared for the IARIW-OECD Special Conference: "W(h)ither the SNA?" Paris, France, April 16-17, OECD.

Gollier C (2012). Pricing the planet's future: The economics of discounting in an uncertain world. Princeton University Press, Princeton.

Gollier, C. (2010). Ecological discounting, Journal of Economic Theory, 145: 812-829.

Groom B and Maddison D J (2017). 'Four New Estimates of the Elasticity of Marginal Utility for the UK'. Forthcoming in Environmental and Resource Economics. Working paper version: Centre for Climate Change Economics and Policy Working Paper No. 141.

Groom B and Venmanns F (2017). Estimates of the elasticity of marginal utility for the environment: preliminary experimental results. *Mimeo*, LSE.

Hamilton K and Hartwick J (2014). Wealth and Sustainability. Oxford Review of Economic Policy, 30: 170-187.

Heal, G. (2009). Climate Economics: A Meta-Review and Some Suggestions for Future Research, Review of Environmental Economics and Policy 3(1): 4-21.

HM Treasury (2008). Intergenerational wealth transfers and social discounting: Supplementary Green Book guidance. HM Treasury, London, UK.

HMT (2003). Guidelines on Cost Benefit Analysis. HM Treasury, UK.

Hoel, M and Sterner, T (2007). Discounting and relative prices. Climatic Change, 84: 265-280.

Jorgenson, D (1963). Capital Theory and Investment Behavior. American Economic Review 53, no. 2 (1963): 247-259

Jorgenson, D. W. and B. M. Fraumeni (1989). The Accumulation of Human and Non-Human Capital, 1948-1984, in R. Lipsey and H. Tice eds., *The Measurement of*

*Saving, Investment and Wealth*, Chicago, University of Chicago Press, NBER, pp. 227-282, 1989.

Jorgenson, D. W. and B. M. Fraumeni (1992) 'Investment in Education and U.S. Economic Growth,' *Scandinavian Journal of Economics*, Vol. 94, supplement, pp. S51-70, 1992a.

Mayer C (2014). Unnatural capital accounting. <http://nebula.wsimg.com/5e358684df379a17f7eff52ca411d7a4?AccessKeyId=68F83A8E994328D64D3D&disposition=0&alloworigin=1>

MEA, 2005, Millennium Ecosystem Assessment (2005), Ecosystems and Human Well-Being: Synthesis, Washington, DC: Island Press.

Neumayer, E. (2010). *Weak Versus Strong Sustainability: Exploring the Limits of Two Opposing Paradigms*, Cheltenham, UK, Edward Elgar, 3rd Edition.

ONS (2015). Human capital estimates: 2015 Estimates of national and regional human capital in the UK from 2004 to 2015. ONS. <https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/humancapitalestimates/2015#how-we-measure-human-capital>

Stern NH (2007). *The economics of climate change: the Stern review*. Cambridge University Press, Cambridge, UK.

Stern T and Persson M (2008). An even sterner review: Introducing relative prices into the discounting debate. *Review of Environmental Economic Policy*, 2: 61-76.

Stiglitz J.E., A. Sen and J.-P. Fitoussi (2009), Report by the Commission on the Measurement of Economic Performance and Social Progress, [http://www.stiglitz-sen-fitoussi.fr/documents/rapport\\_anglais.pdf](http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf)

Ten Brink, P. (Ed.) (2011). *The Economics of Ecosystems and Biodiversity in National and International Policy Making*, London, Earthscan.

Traeger CP (2011). Sustainability, limited substitutability and non-constant social discount rates. *Journal of Environmental Economics and Management*, 62: 215-228.

Weikard, H.-P. and X. Zhu (2005). Discounting and environmental quality: when should dual rates be used?, *Economic Modelling*, 22: 868-878.

Weitzman, M. L. (1976). 'On the Welfare Significance of National Product in a Dynamic Economy', *The Quarterly Journal of Economics*, 90: 156–162.

Weitzman, M.L. (1976). On the welfare significance of national product in a dynamic economy. *Quarterly Journal of Economics* 90, 156–162.

## Section 2: Discounting health and healthcare costs

### 2.1. A brief history of discounting health and health care costs in the UK.

The appropriate discounting of health effects and health care costs in the evaluation of health projects, programmes and technologies has been a source of debate and confusion over a number of years (e.g., Brouwer 2005, Claxton *et al.* 2011, Nord 2011 and Paulden *et al.* 2017). This is illustrated by the recent history of discounting policies. For example, the UK's National Institute for Health and Clinical Excellence (NICE) initially required that costs and health effects be discounted at a real rate of 6% and 1.5% respectively (NICE 2001), reflecting guidance from the UK Department of Health (DH 1996) which was in place at that time. This guidance was reissued in 2004 suggesting rates of 3.5% for costs and 1.5% for health effects which was closely based on the reasoning in Gravelle and Smith, 2001. In 2004, NICE also amended its discounting guidance (NICE, 2004), but instead required that costs and health effects should both be discounted at the 3.5% STPR specified in the Treasury Green Book. The NICE policy of common discounting of costs and health effects at 3.5% was maintained in subsequent guidance (NICE 2008). NICE later amended this guidance indicating that a lower common rate of 1.5% could also be considered when there are long term and substantial health benefits, which are 'highly likely' to be achieved, and where introduction of the technology does not commit the NHS to significant irrecoverable costs (section 6.2.19, NICE 2013) (see 1.4). A common discount rate for health and health care costs is in line with recommendations from the US<sup>11</sup> and most assessment bodies in other countries.<sup>12</sup>

The draft 'best practice' report from the Appraisal Alignment Working Group (AAWG)<sup>13</sup> suggest a discount rate of 1.5% for health (e.g., QALYs, and the ASCOT equivalent for social care) and health care costs, but 3.5% applied to other impacts which fall outside public expenditure on health care. The lower discount rate on health and health care costs reflects a view that the consumption value of health is expected to grow in the long run, i.e., it adopts a dual discounting approach. Applying the same lower rate to health care costs recognises that the opportunity costs of NHS costs fall on health outcomes rather than consumption. Estimates of health opportunity costs (the marginal productivity of NHS expenditure) and how it is likely to evolve over time is dealt with separately and explicitly rather than being embedded in the discount rate for health care costs (see 2.2.3).

---

<sup>11</sup> For example, the Washington Panel (Lipscomb *et al.* 1996) recommended common discounting at a rate of 3%. The recent update to this guidance (Neumann *et al.* 2016) continues to recommend common discounting at 3% although the reasoning has been shown to be fatally flawed (Paulden *et al.* 2017).

<sup>12</sup> Exceptions include the Belgian Federaal Kenniscentrum voor de Gezondheidszorg (KCE) guidelines for economic evaluation of pharmaceutical products require a discount rate of 3.5% for costs and 1.5% for health benefits (Cleemput *et al.*, 2008, p. 28) and the Dutch Health Insurance Board require a discount rate of 4% for costs and 1.5% for health benefits (CVZ, 2006, p. 10).

<sup>13</sup> The AAWG includes policy and analytic staff who work in, or give advice to, DH and its ALBs (e.g., DH, NICE, PHE, NHSE, JCVI etc) on the cost-benefit and cost-effectiveness of programmes, technologies and policies. The working group is an advisory rather than a decision making body.

In view of this history of changing and sometimes conflicting recommendations it would not be surprising if practitioners, policy makers and other stakeholders (importantly clinical communities) are confused about why health should be discounted, whether health care costs should be discounted at the same rate as health and what discount rates should be applied.

Many of these conflicts and contradictions are more apparent than real and arise from different judgements about normative and empirical questions:

- i. whether the social objective of the health care expenditure and the decision maker's economic analysis claims to inform is to maximise welfare or health itself;
- ii. the health opportunity costs associated with constraints on the growth in public health care expenditure;
- iii. expected changes in health opportunity costs and the consumption value of health over time;
- iv. the social time preference for health and for consumption.

Lack of clarity has tended to be compounded when discounting is used to implicitly account for expected changes in health opportunity costs and the consumption value of health over time (i.e., multiple dual discounting) rather than representing effects as either a time stream of health gained and health forgone (discounted at an appropriate rate for health) or valuing this time stream of health effects at their equivalent consumption value (discounted at a rate for consumption).

## **2.2. The objective of health care expenditure is to improve health**

This normative position views decision making bodies and institutions in health as the agents of a principal (e.g., a socially legitimate process such as government) which allocates resources and devolves powers to the agent, giving it a responsibility to pursue specific, measurable and therefore narrowly defined objectives that are regarded as socially valuable, e.g., improving health. In these circumstances economic analysis cannot be used to make claims about social welfare or the optimality or otherwise of the resources allocated to health care. Its role is more modest, claiming to inform social decisions in health, revealing the implied values and exposing the implications of social choices made by the principal. It is this role that economic analysis has tended to play in health policy, especially in the UK, and underpins much of the evaluation of health care projects and cost-effectiveness analysis that has been conducted (Drummond *et al.* 2015, Coast *et al.* 2008).

### *2.2.1. Why discount health?*

In this context, the reason to discount future health effects cannot appeal to preferences and the type of welfare arguments that underpin STPR based on the Ramsey Rule, but instead to the opportunity costs of financing health care. The health care costs of a project could have been invested elsewhere in the economy or used

to reduce public borrowing at a real rate of return, which would provide more health care resources in the future and generate greater health benefits. Health care transforms resources into health so from the perspective of a social planner trading health care resources over time is to trade health. Therefore, if health care costs are discounted to reflect the opportunity cost of financing health care, their health effects must be discounted as well.<sup>14</sup> Since the social planner in health care is not able to make investments in the private sector the opportunity cost they face is the rate of return on debt reduction rather than higher estimates of the social opportunity cost of capital based on market rates (Spackman 2017). For example, real yields on UK government bonds reflect the marginal cost of increasing health care expenditure available to government (Paulden and Claxton 2012). In this context the broader question of the social opportunity costs of public expenditure including the macroeconomic choice of levels and mix of taxation and borrowing (Spackman 2017) can be regarded as the responsibility of government rather than spending departments or decision making bodies such as NICE.

### 2.2.2. Representing the effects of health care projects

Estimates of the additional health care costs ( $\Delta c_h$ ) and additional health effects ( $\Delta h$ ) (e.g., measured as QALYs gained) of a health care project or intervention are commonly presented as incremental cost-effectiveness ratios (ICER). These provide a useful summary of how much additional resource is required to achieve a measured improvement in health (the additional cost per QALY gained). Whether the intervention will improve health outcomes overall, because the cost per QALY it offers is judged to be cost-effective, requires a comparison with a 'threshold' ( $k_h$ ) that reflects the likely health opportunity costs, i.e. the improvement in health that would have been possible if the additional resources required had, instead, been made available for other health care activities (the marginal productivity of health care expenditure). A project will improve health overall if the additional cost per QALY it offers is less than the cost-effectiveness 'threshold' ( $\Delta c_h/\Delta h < k_h$ ).

Some assessment of health opportunity cost and its evolution over time is required. For example, if the threshold is expected to grow in real terms ( $gk_h > 0$ ), because the marginal productivity of health care expenditure is expected to decline (e.g., due to real growth in health expenditure), then future costs are less important because they will be expected to displace less health. The relative importance of future health care costs can be reflected in the following ways, which have different implications for discounting policy (see Table 1):

- i. The health benefits and costs of the project can be reported as a stream of expected health gained and forgone each period ( $t$ ) by applying the threshold

---

<sup>14</sup> This is commonly illustrated by a comparison of terminal and present values. The cost per QALY of a project with immediate costs and additional health benefits all occurring at a future point in time is the same whether costs are expressed at their terminal value when the health benefits occur, or discounting the health benefits back to their present value at the same rate (Nord 2011).



relevant to that period,  $k_{ht}$ , to the health care costs that occur that period ( $\Delta h_t - \Delta c_{ht}/k_{ht}$ ). This time stream of health effects ((2) and (3) in Table 1) can then be discounted at a rate which reflects a social time preference for health ( $D_h=r_h$ , see 2.2.5).

- ii. Alternatively, health benefits can also be valued as the health care resources required to deliver similar benefits elsewhere. The effects of the project can be reported as a stream of expected health care resources gained and forgone in each period by applying the relevant 'threshold' to the health benefits that occur in that period ( $\Delta h_t \cdot k_{ht} - \Delta c_{ht}$ , (6) and (7) in Table 1). This time stream of health care resources can then be discounted at a rate, which reflects the marginal opportunity cost, faced by government, of increasing public health care expenditure ( $r_s$ ), e.g., real yields UK government bonds.
- iii. If the effects of a project are reported as an incremental cost effectiveness ratio ( $\Delta c_{ht}/\Delta h_t < k_h$ ) it must be compared to a single 'threshold' relevant to the current period ( $k_{ht}$ ). However, some account must still be taken of expected changes in health opportunity costs. For example, if health opportunity costs are expected to grow in real terms ( $g_k > 0$ ), because the marginal productivity of health care expenditure is expected to decline, then future costs are less important because they will be expected to displace (or any additional resources could deliver) less health. In some circumstances this can be achieved by discounting the additional health care costs at a rate that accounts for any growth in the 'threshold', reflecting the relative importance of future costs ( $D_c = r_h + g_k$ <sup>15</sup>, and  $D_h=r_h$ ), i.e., a form of dual discounting which reflects expected changes in the marginal productivity of health care expenditure rather than changes in the value of health relative to consumption (Claxton *et al.* 2011).

Most analysis of health care projects and interventions in the UK implicitly adopts this type of normative position but generally reports results as cost effectiveness ratios rather than net health benefits (column (4)-(5) in Table 1) or the equivalent net effect on health care resources (column (6)-(7) in Table 1) (Phelps and Mushlin 1991, Stinnett AA, Mullahy 1998). This can be seen as an historic norm which may reflect reluctance on the part of decision-making and advisory bodies to be explicit about how much society can afford to pay to improve health and how this is likely to evolve over time.<sup>16</sup> Until recently there has also been a lack of evidence about the likely health opportunity costs (Culyer *et al.* 2007). Consequently, implicit assessments have been embedded in how costs and health effects are discounted. This has contributed to a

---

<sup>15</sup> This approximation is based on the plausible assumption that  $r_h$  and  $g_k$  are small.

<sup>16</sup> Since 2004 NICE has published an explicit range for the cost-effectiveness thresholds used in its deliberative decision-making process (£20,000 to £30,000 per QALY) (NICE, 2004). Although NICE makes clear that the threshold ought to represent the health consequences of additional NHS costs, this range was, in fact, founded on the values implied by the decisions it made between 1999 and 2003 (Rawlins and Culyer, 2004). It has become an established norm, which is intended to represent how NICE makes its decisions rather than an evidence based assessment of the likely health opportunity costs

lack of clarity about discounting policy, what a cost effectiveness ‘threshold’ ought to represent and how it might be informed with evidence.

**Table 1.**

**Reporting the effects of a project with health benefits and health care costs**

	Effects of the project		Health Effects		Equivalent health care resources		Equivalent consumption effects	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Time	Additional health benefits	Additional health care costs	Benefits	Costs	Benefits	Costs	Benefits	Costs
1	$\Delta h_1$	$\Delta c_{h1}$	$\Delta h_1$	$\Delta c_{h1}/k_{h1}$	$k_{h1} \cdot \Delta h_1$	$\Delta c_{h1}$	$V_{h1} \cdot \Delta h_1$	$V_{h1}(\Delta c_{h1}/k_{h1})$
..	..	..	..	..	..	..	..	..
t	$\Delta h_t$	$\Delta c_{ht}$	$\Delta h_t$	$\Delta c_{ht}/k_{ht}$	$k_{ht} \cdot \Delta h_t$	$\Delta c_{t1}$	$V_{ht} \cdot \Delta h_t$	$V_{ht}(\Delta c_{ht}/k_{ht})$
..	..	..	..	..	..	..	..	..
T	$\Delta h_T$	$\Delta c_{hT}$	$\Delta h_T$	$\Delta c_{hT}/k_{hT}$	$k_{hT} \cdot \Delta h_T$	$\Delta c_{hT}$	$V_{hT} \cdot \Delta h_T$	$V_{hT}(\Delta c_{hT}/k_{hT})$

It is also possible to express the effects of a project which only has health effects and additional health care costs as the equivalent consumption value of the health gained ( $V_{ht} \cdot \Delta h_t$ ) and the health forgone ( $V_{ht}(\Delta c_{ht}/k_{ht})$ ) in each time period (see columns (8) and (9) in Table 1). To do so requires some assessment of the consumption value of health ( $v_{ht}$ ) and how it is likely to evolve over time (see Section 3.2.1 for a more detailed discussion and a brief consideration of available evidence).

For the type of projects illustrated in Table 1, where there are no effects outside health and health care costs (or where the social planner has decided that other effects should be set these aside when considering this type of health care project<sup>17</sup>), the equivalent consumption value of health does not influence the decision as it simply rescales any net health benefit or net health loss (both sides of  $\Delta h > \Delta c_h/k_h$  are multiplied by the same quantity). The key, however, is that health care costs cannot be treated as if they are private consumption costs, because  $k_{ht}$  and  $v_{ht}$  cannot be assumed to be necessarily and always equal.

For example, reported values of  $v_{ht}$  tend to be higher than available estimates of a ‘supply side’ assessment of health opportunity costs (Vallejo-Torres et al, 2016). This

<sup>17</sup> There are reasons to set aside explicit and quantitative consideration of other effects if they are likely to conflict with other important social arguments that are difficult to specify let alone quantify, e.g., equity and the benefits of social solidarity offered by collectively funded health care. This is the explicit decision that has been taken in the UK by NICE and UK DH after considering the benefits and potential costs of quantifying these wider effects in the decision making process (refs Claxton et al 2015b and #Claxton et al 2010#).

suggests a common discrepancy between the demand and supply side of health care systems. If these estimates of  $v_{ht}$  are regarded as an appropriate expression of social value, the difference between  $v_{ht}$  and  $k_{ht}$  would indicate that health care from collectively pooled resources is 'underfunded' compared to individual preferences about health and consumption.<sup>18</sup> It is consistent with the view that the public funding of health care is not matching individual preferences and public expectations of their health care system. However, given the difficulties faced in the public financing of health care and the welfare losses associated with socially acceptable means of public finance, this is what might be expected. The balance of evidence suggests that  $v_{ht}/k_{ht} > 1$  which represents the value of NHS £ relative to a private consumption £ or the shadow price of public health expenditure.

Consideration of  $k_{ht}$  is essential when comparing different health care projects competing for available health care resources and especially when they have effects outside health and health care costs (see Section 2.3.2). However, it is also relevant when considering broader questions of whether public resources available for health care should be increased. For example, it helps to inform two key questions:

- i. Is there a strong case for increasing health expenditure because the current scale of the discrepancy between the supply and demand side ( $k_{ht} < v_{ht}$ ), means projects are being rejected that would have offered net social benefits if total expenditure was increased to the point where  $k_{ht}=v_{ht}$ .
- ii. How much of an increase in health expenditure would be required to ensure  $k_{ht}=v_{ht}$  (how is  $k_{ht}$  likely to evolve as total expenditure is increased over time and how is  $v_{ht}$  likely to evolve over the same period).

The only circumstance in which evidence about  $k_{ht}$  could be reasonably disregarded is if it is assumed that public health expenditure will be immediately increased to the point that  $k_{ht}=v_{ht}$ . Since ONS is not in a position to set public finances it would not be appropriate to evaluate projects (whether or not they have effects outside health care or need to be compared to projects in other sectors) based only on  $v_{ht}$  or adopt discounting policies that implicitly make such assumptions. It would be better to establish discounting policies that are founded on the evaluation of projects which include empirically based assessment of how  $k_{ht}$  and  $v_{ht}$  are expected to evolve.

---

<sup>18</sup> For example, the UK DH has adopted £15,000 per QALY to assess health opportunity costs and until recently £60,000 per QALY as an estimate of the consumption value of health based on deriving QALY effects from VSL estimates. This would suggest that one health care £ is worth £4 of private consumption effects, which is especially important when there are other impacts which fall outside constrained public expenditure.

### 2.2.3. Assessment of health opportunity costs

The problem of estimating a cost-effectiveness ‘threshold’ that represents expected health opportunity costs is the same as estimating the relationship between changes in health care expenditure and health outcome. Recent research used national data on expenditure and outcomes in different disease areas (programme budget categories) reported at a local level in the UK NHS (Martin *et al.* 2008, Martin *et al.* 2012 and Claxton *et al.* 2015a). By exploiting the variation in expenditure and mortality outcomes, the relationship between changes in expenditure and mortality can be estimated while accounting for endogeneity. By using the effect of expenditure on the mortality and life-year burden of disease as a surrogate for the effects on a more complete measure of burden (one that also includes the quality of life burden of disease), a cost per QALY threshold that reflects the likely impact of changes in expenditure on both mortality and morbidity can be estimated (e.g., £13,000 per QALY for 2007/08 expenditure).

Analysis of three waves of expenditure and outcome data did not provide evidence of growth in the threshold at a time when there was real growth in health care expenditure. A lack of evidence of declining marginal productivity may be due to changes in level and type of demand for health care as well as increases in productivity through improvements in allocative and technical efficiency. The DH is supporting the ongoing re-estimation of cost per QALY thresholds for subsequent waves of data, which will provide a longer series of cross sectional estimates and opportunities for panel data estimation.<sup>19</sup>

In response to this accumulating evidence, the DH adopted an estimate of health opportunity costs of £15,000 per QALY, which is used to inform the impact assessments it conducts. The DH expects that in the longer run the marginal productivity health care expenditure will fall as greater real increases in health care expenditure start to reflect the expected growth in the consumption value of health. However, in the short to medium term, increases in demand, modest increases in real funding and increases productivity through innovation in health and medicine means that the threshold is not expected to increase. The DH assumes that a threshold of £15,000 per QALY will remain constant in real terms for the next 10 years and then grow at 2% pa with anticipated periodic re-estimation of this marginal cost per QALY. These assessments are also reflected in the draft ‘best practice’ report from AAWG, which also suggests that health opportunity costs are dealt with explicitly and separately from discounting.

This seems to represent a reasonable and thoughtful assessment based on the balance of evidence such as it is. It provides a useful and consistent default

---

<sup>19</sup> The analysis of 10 waves of data are now complete and suggests that  $k_t \leq £15,000$  per QALY and there is no evidence of growth in real terms. Estimation of panel data is underway but initial results suggest similar estimates to those based on cross sectional analysis. See <https://www.york.ac.uk/che/research/teehta/health-opportunity-costs/re-estimating-health-opportunity-costs/#tab-2>

assessment of health opportunity costs and how they are likely to evolve. They can be used to report time streams of health care expenditure as health effects or time streams of health benefits as the equivalent health care resources. It means the difficulties and lack of clarity associated with reporting cost-effectiveness ratios and the requirement for dual discounting can be avoided.

#### 2.2.4. Other impacts

Health care projects often impose cost or offer benefits beyond measures of health and public expenditure on health care; for example, the net production effects of improved survival and quality of life (e.g., Meltzer 2013). Some implicit assessment of whether other benefits can justify net health losses is required in deliberative decision making processes (e.g., NICE Appraisal). An explicit consumption value of health allows them to be expressed as their health or health care resource equivalent (see Table 2 and Sections 2.3.1 and 2.3.2).

#### 2.2.5. Time preference for health

The normative position that underpins much of the evaluation of health care projects, especially in the UK, takes the values implied by the outcome of legitimate processes (e.g., government setting the budget for health care) as a partial but revealed expression of some unknown latent social welfare function that may include many conflicting arguments, e.g., health equity, social solidarity among many others that are difficult to specify let alone quantify (Drummond *et al.* 2015). Similarly the social choice of how resources are devoted to health care over time and the resulting health in each period reveals something about society's willingness to trade current and future health, i.e., the choices of the principal in setting budgets based on expectations about the marginal productivity of health care in each period implies values for  $k_{ht}$ . Therefore, a revealed social time preference for health<sup>20</sup> can be based on the rate at which government can borrow or save ( $r_s$ ) and whether the threshold is expected to grow ( $gk_h$ ) because this indicates the relative value (in terms of health care resources) of current compared to future health ( $r_h = r_s - gk_h$ ) (Paulden and Claxton 2012).

### 2.3. The objective of health care expenditure is to improve welfare

Traditionally economic analysis (e.g., Boadway and Bruce, 1984) adopts a view of social welfare resting on individual preferences revealed through markets and their surrogates or modified by an explicit welfare function. Analysis based on this normative position (e.g., cost-benefit analysis) is less well represented in the evaluation of health projects, partly due to the difficulty of decision making bodies being willing to identify a welfare function carrying some broad consensus or social legitimacy (Arrow 2012), particularly if health is felt to be unlike other goods (e.g., Broome 1978, Sen 1979, Brouwer *et al.*, 2008). Nevertheless, health must inevitably

---

<sup>20</sup> This is the time preference for health, as distinct from pure time preference (for utility) or STP for consumption (see 2.5 below).

be traded with other welfare arguments, most notably consumption, by social planners whilst taking account of the budget constraints they face.

This normative position regards purpose of health care expenditure as improving a broader notion of welfare rather than health itself. If consumption and health are the only arguments or are separable from others then decisions which maximise the consumption value of health will also maximise social welfare (Gravelle *et al.*, 2007). In this context the reason to discount future health effects can be based on preferences and the type of welfare arguments that underpin STPR based on the Ramsey Rule. This provides a clear link between the social time preference rate for consumption and health (Gravelle and Smith, 2001).

The relative importance of future health care costs and the consumption value of health gained and forgone can be reflected in the following ways which have different implications for discounting policy:

- i. The health benefits and costs of a project can be reported as a stream of expected health gained and forgone each period by applying the ‘threshold’ relevant to that period ( $\Delta h_t - \Delta c_{ht}/k_{ht}$ ). These health effects can be valued by applying a consumption value of health relevant to that period  $v_{ht}(\Delta h_t - \Delta c_{ht}/k_{ht})$  (See Table 1). The stream of consumption gains and losses can be discounted at a rate which reflects a STP that would be the relevant rate to apply in all contexts were benefits and costs have been expressed in terms of consumption.
- ii. This can also be expressed as a comparison of the cost effectiveness ratio of the project to the current period cost-effectiveness ‘threshold’ if the discount rate applied to  $\Delta h_t$  is amended to reflect growth in the consumption value of health ( $D_h = r_c - gv_h$ ) and the discount rate applied to  $\Delta c_h$  is amended to reflect growth in the consumption value of health forgone and changes in the rate at which future health will be forgone ( $D_c = r_c - gv_h + gk_h$ ) (Claxton *et al.*, 2011).<sup>21</sup>

This approach (in ii) poses more difficulties and potential for confusion, with dual discounting being used to account for changes in the value of health and changes in the marginal productivity of health expenditure as well as time preference. The separate and explicit accounting for each of these effects (in i) would appear more transparent, accountable and comparable.

The DH and the draft of the AAWG ‘best practice’ report suggests that health opportunity costs are dealt with explicitly and separately from discounting (i.e., health care costs are treated as health losses). They recommend a discount rate of 1.5% applied to health and health care costs, which embeds the expectation that the consumption value of health will grow at 2% based on the real consumption growth assumed in the Green Book STPR of 3.5% and an (implicit) assumption of an income

---

<sup>21</sup> This approximation is based on the plausible assumption that  $r_h$ ,  $g_v$  and  $g_k$  are small.

elasticity of demand of health of one.<sup>22</sup> They recommend that 3.5% is applied to effects on other non-health benefits and costs. Therefore, this policy uses dual discounting to account of an expectation of  $gv_h=2\%$  but applies it equally to health benefits and health care costs, recognising health opportunity costs and an assumption that  $gk_h=0$  over the next 10 years.<sup>23</sup>

### 2.3.1. Assessing the consumption value of health

There is a large literature which has used stated preferences (contingent valuation and discrete choice experiments) to estimate the consumption value or willingness to pay for a QALY (e.g., Pinto-Prades 2009, Mason *et al.* 2009). The estimates reflect the demand for health and imply what health care expenditure ought to be, rather than a 'supply side' assessment of the marginal productivity of health care expenditure. Most estimate how much consumption an individual is willing to give up to improve their own health. A few try to elicit how much individuals believe society should pay to improve health more generally. A wider literature, that extends beyond health, estimates the value of a statistical life (VSL) based on how much consumption individuals are willing to give up to reduce their mortality risk (Hammit 2000, Robinson *et al.* 2016). Some studies are based on stated preferences (e.g., Lindhjelm 2011) but others identify situations where individuals make choices that imply a value, e.g. revealed preferences in the labour market. A cost per QALY can be derived from these studies by making assumptions about age and gender distribution, conditional life expectancies and quality of life norms.

Recent reviews of this literature reveal wide variation in values (Vallejo-Torres *et al.*, 2016; Ryen and Svensson, 2015;). However, some patterns do emerge: estimates based on VSL studies tend to be higher than those based on willingness to pay for a QALY; values are not proportional to the scale of health gains and differ depending on whether QALY gains are through quality improvement or survival benefits. Reported values also tend to be higher than available estimates of a 'supply side' assessment of health opportunity costs (Vallejo-Torres *et al.* 2016). This suggests a discrepancy between the demand and supply side of health care systems. For example, if these estimates are regarded as an appropriate expression of social value, the difference would indicate that health care from collectively pooled resources is 'underfunded' compared to individual preferences about health and consumption. However, given the difficulties faced in the public financing of health care and the welfare losses associated with socially acceptable means of taxation this is what might be expected.

The balance of evidence suggests that  $v_h/k_t > 1$ , which would indicate that public expenditure available for health care is relatively scarce and more valuable than the same amount of private consumption. Until recently, the DH took £60,000 per QALY as an estimate of the consumption value of health based on deriving QALY effects

---

<sup>22</sup> This happens to nullify the wealth effect in the Ramsey Rule based on the Green Book (see 2.3.1).

<sup>23</sup> They recommend that the discount rate for health care costs that occur after 10 years should increase to 3.5% based on an assessment that in the longer run  $gk_h$  will match  $gv_h$  and grow at 2%

from VSL estimates. This appears somewhat higher than recent reviews of the literature and the DH is currently reviewing this estimate. It would suggest that one NHS£ is worth £4 of private consumption effects, which is especially important when there are other impacts which fall outside budget constrained public expenditure (see Section 2.3.2).

The evident difficulties in eliciting willingness to pay for QALY gains means that there is also limited empirical evidence of how these values change with income. However, reviews of the literature that have investigated the relationship between the VSL and income (e.g., Viscusi and Aldy 2003; and Hammitt and Robinson 2011) suggests that earlier cross sectional studies of wage-risk premiums indicate income elasticities  $<1$ , but longitudinal or cohort studies typically estimate elasticities  $>1$ . (e.g., Costa and Kahn 2004). The reasons for these differences may be that cross-sectional studies are more likely to reflect changes in realised income, whereas longitudinal or across cohort studies are more likely to capture the impact of permanent income (e.g., Getzen 2000; Aldy and Smyth 2014). Despite the empirical difficulties the balance of evidence suggests that the consumption value of health increases with income. Assuming an income elasticity of demand of health  $\geq 1$  would not be unreasonable.

There are also sound theoretical reasons why the value of health would be expected to grow with consumption (e.g., Parsonage and Neuburger 1992, Gravelle and Smith 2001, Hall and Jones 2007). The intuition can be expressed in the same way as the expected increase in value of environmental goods; that the growth in consumption is likely to outstrip the growth in health so health will become scarcer relative to consumption. Since consumption is an imperfect substitute for health the value of health will increase. These arguments can be made using behavioural models of individual choices of health affecting activities over time e.g., purchasing health care. The growth in the value of health will be determined by income growth, the income elasticity of demand for health care and the elasticity of the marginal productivity of health care. Alternatively, health can be included as a separate argument in a social welfare function where it is valued in its own right, in part, because a healthier state increases the marginal utility of income and an indirect effect through income due to uninsured health care costs and/or increased productivity of being in a healthier state. These insights indicate there are compelling reasons to believe the value of health will grow with income and it is likely to grow at a faster rate if there is a direct effect of health on utility and an indirect effect through income.

Gravelle and Smith (2001) identified a number of special cases. For example if health has no effect on income and the utility effect of health is constant over time then  $g_v$  will be equal to rate at which marginal utility of income declines (the wealth effect in the Ramsey Rule). Alternatively, when health affects income but has no direct effect on utility  $g_v$  is equal to real growth in income. These special cases just happen to indicate  $g_v = 2\%$  based on the values used by Treasury in the Ramsey rule. Although theoretical arguments point to number of empirical questions, a simple but reasonable assessment could be based on growth in consumption (embedded in the STPR) and



the income elasticity of demand for health. The DH and the draft 'best practice' report from AAWG accepts a STPR of 3.5% but a suggest a discount rate of 1.5% for health reflecting a view that the consumption value of health is expected to grow at 2%. Given real growth of 2% this implies an income elasticity of demand for health of one. This might be regarded as a reasonable albeit potentially conservative assessment given empirical evidence and theoretical insights.

### 2.3.2. Other impacts

Adopting an explicit consumption value of health allows cost and benefits beyond measures of health and public health expenditure to be included as a stream of consumption gains and losses alongside the stream of the consumption gains and losses associated with health benefits and health opportunity costs. Once the effects on health, health care costs and other impacts are expressed as equivalent streams of consumption, they can be discounted at a STPR. This is illustrated in Table 2 where a project which has health benefits and health care costs also imposes costs on private consumption ( $\Delta c_{ct}$ ) or offers private consumption benefits (i.e., when  $\Delta c_{ct} < 0$ ).

**Table 2.**  
**Reporting the effects of a project on health, health care costs and consumption**

(1)	Effects of the project			Effects on health	Effects on consumption	Net consumption effects
	(2)	(3)	(4)	(5)	(6)	(7)
	Additional health benefits	Additional health care costs	Consumption costs	Net health benefits	Net consumption costs	Equivalent consumption effects
1	$\Delta h_1$	$\Delta c_{h1}$	$\Delta c_{c1}$	$\Delta h_1 - \Delta c_{h1}/k_{h1}$	$\Delta c_{c1} + k_{c1} \cdot \Delta c_{h1}$	$v_{h1}(\Delta h_1 - \Delta c_{h1}/k_{h1}) - (\Delta c_{c1} + k_{c1} \cdot \Delta c_{h1})$
..	..	..	..	..	..	..
t	$\Delta h_t$	$\Delta c_{ht}$	$\Delta c_{ht}$	$\Delta h_t - \Delta c_{ht}/k_{ht}$	$\Delta c_{ct} + k_{ct} \cdot \Delta c_{ht}$	$v_{ht}(\Delta h_t - \Delta c_{ht}/k_{ht}) - (\Delta c_{ct} + k_{ct} \cdot \Delta c_{ht})$
..	..	..	..	..	..	..
T	$\Delta h_T$	$\Delta c_{hT}$	$\Delta c_{hT}$	$\Delta h_T - \Delta c_{hT}/k_{hT}$	$\Delta c_{cT} + k_{cT} \cdot \Delta c_{hT}$	$v_{hT}(\Delta h_T - \Delta c_{hT}/k_{hT}) - (\Delta c_{cT} + k_{cT} \cdot \Delta c_{hT})$

However, the effect on consumption ((6) in Table 2) also requires some assessment of the other (non-health) opportunity costs associated with additional health care costs. Therefore, once other effects beyond health and health care costs are included, some assessment of either the consumption opportunity costs of health care expenditure ( $k_{ct}$ ) or the consumption effects of changes in health is also required (whether they

are gains,  $\Delta h_t$ , or opportunities lost,  $\Delta C_{ht}/K_{ht}$ <sup>24</sup>. The net effects of the project on both health and consumption can then be reported as two time streams of net health and net consumption effects ((5) and (6) in Table 2).<sup>25</sup> Once the time stream of net health effects are transformed into equivalent consumption (using  $v_{ht}$ ) the net consumption effects can be discounted at the STPR ((7) in Table 2).

As part of efforts to inform value based pricing of branded medicines (DH 2010 NICE 2014), the DH undertook work to estimate the 'wider social benefits' associated with changes in health outcome. These were characterised as the consumption value of production effects net of additional consumption due to improvements in survival and quality of life. These estimates included valuation of marketed and non-marketed production and consumption, by age, gender, broad areas of disease area (ICD codes) and whether effects fall on quality or length of life (see Appendix B of Claxton *et al.* 2015b). They can provide some default assessment of the net production effects likely to be associated with the particular type of health benefits offered by a health care project.

Importantly, they can also be linked to evidence of health opportunity costs (reported by age gender ICD code, quality and survival effects) to estimate the net production effects of changes in health care expenditure. The DH estimate that the marginal £ in the NHS budget provides 63p worth of net production gains ( $k_{c1}=1.59$ ). The stream of consumption losses due to these net production opportunity costs of the health care costs of a project can be included in net consumption effects ((7) in Table 2) and discounted at a STPR. Although there is little guidance on how this aspect of opportunity costs is likely to evolve, a default assumption that the real value of net production effects of the health effects of expenditure will grow at the same rate as consumption would seem reasonable.

Health projects may well have impacts on other categories of public expenditure. Estimates of  $v_{ht}/K_{ht}$  in the health sector might be used to shadow price other forms of public expenditure (where the equivalent estimates for that sector are absent) since resource allocation and expenditure decisions by government, Treasury and spending departments would be expected to equalise this ratio across sectors (x) given an overall political choice of total public expenditure, i.e., it may not be unreasonable to assume  $v_{ht}/K_{ht} = v_{xt}/K_{xt}$  when considering impacts on other categories of public sector.

---

<sup>24</sup> These alternatives will be equivalent if the causal consumption effects of health care expenditure run only through the health effects of health expenditure, rather than, in part at least, directly from health expenditure itself. Insofar as health expenditure has a positive impact on economic growth compared to other forms of expenditure then restricting attention to the consumption effects of changes in health is likely to underestimate the consumption opportunity costs of health care costs.

<sup>25</sup> It should be noted that attempts to estimate and explicitly account for the consumption opportunity costs of health care expenditure are particularly limited, even in high income settings, but do exist (Claxton *et al.* 2015b). Although there is currently little evidence in lower income setting to support such assessment some default assumptions based on what is already known about the relationship between changes in health and economic growth should be possible (see Section 2.2.2).

The implications for discounting policy is that it becomes more difficult and opaque to try and embed all relevant arguments in how health and health care costs are discounted, i.e., reflecting changes in the value of health and in the marginal productivity of health care expenditure in terms of health and net production. AAWG 'best practice' draft report suggests that quantification and conversion to a common numeraire be done separately and explicitly, allowing available evidence to be used transparently and consistently, while preserving the possibility of accountable deliberation about evidence, values and unquantified arguments.

#### **2.4. Uncertainty, risk and time horizon**

The horizon for many evaluations for health care interventions are often less than 30 years or generally do not extend much beyond that. For example, insofar as a health care intervention impacts on mortality risk the time horizon for costs and benefits need only extend to the survival of the cohort of current beneficiaries. However, projects which require commitment of irrecoverable costs, also require an assessment over the survival of future incident cohorts that will benefit from this investment. Similarly, the value of information generated by clinical research also includes future patient populations, as do preventative interventions and infectious and communicable diseases. If all effects are expressed as streams of consumption gains and losses then discounting using STPR would be appropriate, including any decline over longer time horizons to reflect the impact of uncertainty its components.

Considerable efforts have been made in the evaluation of health care projects to characterise all sources of uncertainty, value the consequences and establish how these should inform project choice; for example, whether the approval of a cost-effective project ( $NPV > 0$ ) should be delayed or access restricted until further research is conducted or until sources of uncertainty resolve overtime (e.g., the entry and change in price of competing interventions). The impact of irrecoverable costs and the real option value of delay have been examined as well as the impact of approval on the opportunities to acquire evidence that would benefit future patient populations. The impact of uncertainty on resource allocation across projects under alternative budgetary policies and the implications uncertain non-marginal budget impacts have also been examined.

This type of analysis starts to unpick the reasons for the appearance of risk aversion in project choice and undermines the justification for embedding a common risk premium in discount rates. The evaluation of health care projects is increasingly attempting to model explicitly many of the effects that are otherwise embedded in 'catastrophic risk' element of the Treasury discount rate. These considerations might (in part) explain why NICE is willing to consider a lower discount rate when there are substantial health benefits, which are 'highly likely' to be achieved, and where introduction of the technology does not commit the NHS to significant irrecoverable costs (section 6.2.19, NICE 2013).

Although the application of this type of analysis (value of information, Bayesian decision theory and real options) is well developed in the evaluation of health projects,<sup>26</sup> it is far from universal. Therefore, applying a risk free STPR to all health projects may be premature. Nonetheless, some project evaluations, may have already accounted for the consequences of some of these project specific risks in a way that others in health or other accounts may not.

## 2.5. Recommendations and considerations

The two alternative normative positions described above have implications for the valuation of effects and discounting. What distinguishes them is whether the social values ought to reflect those implied by the outcome of legitimate processes (e.g., government setting budgets for health care) or a notion of welfare founded on individual preferences. For example, the former suggests a social time preference for health of  $r_s - gk_h$  and that latter,  $r_c - gv_h$ . The distinction is whether social value is expressed by  $k_{ht}$  or  $v_{ht}$  and whether it is the opportunity cost of financing health care or the welfare arguments that underpin the Ramsey Rule that justify discounting.<sup>27</sup>

The choice for the ONS is whether they wish to reflect the normative position that has been adopted in most evaluations of health care projects for decision-making bodies in the UK, or a broader view of welfare that would be consistent with other accounts and the welfare arguments that underpin the Ramsey Rule.

Given the need for consistency between accounts and the importance of being able to explicitly quantify other impacts beyond measures of health and public health expenditure, it would seem appropriate to convert all effects into streams of consumption gains and losses discounted at a STP that would be the relevant rate to apply in all contexts were benefits and costs have been expressed in terms of consumption, including any decline to reflect the impact of uncertainty in the estimate of STP.<sup>28</sup>

This approach avoids embedding multiple arguments in the discount rate for health and health care costs. The separate and explicit accounting for these arguments illustrated in Tables 1 and 2 allows available evidence to be used transparently and consistently, while preserving the possibility of accountable deliberation about evidence, values and unquantified arguments in decision-making processes.

In addition to STPR, it also requires the following quantities specific to health to be assessed:

---

<sup>26</sup> A characterisation of 'all' sources of uncertainty is required by NICE appraisal and value of information analysis is recommended. NICE is considering how more formal analysis of the value of additional evidence and irrecoverable costs can inform when it should make only in research recommendations.

<sup>27</sup> The actual differences may be modest if  $g_k$  and  $g_v$  are similar and the real rate at which government can borrow is regarded as a reasonable proxy for STPR as some argue it is (Council of Economic Advisers 2017).

<sup>28</sup> There will be a decline in the risk free STPR due to uncertainty in consumption growth. However, any risk premium included in STP is likely to increase with term structure.

i. *The marginal productivity of health care expenditure in producing health ( $k_{h1}$ )*

The DH and the draft 'best practice' report from AAWG has adopted an estimate of health opportunity costs of £15,000 per QALY which reflects a reasonable assessment of the balance of current evidence.

ii. *Future changes in the marginal (health) productivity of health care expenditure ( $gk_h$ )*

The DH assumes that a threshold of £15,000 per QALY will remain static in real terms for the next 10 years and then grow at 2% pa. These assessments are also reflected in the draft 'best practice' report from AAWG and appear to be a reasonable assessment given currently available evidence. Periodic re-estimation and re-assessment of this marginal cost per QALY and its likely evolution is anticipated.

iii. *The consumption value of health ( $v_{ht}$ )*

Until recently the DH took £60,000 per QALY as an estimate of the consumption value of health based on deriving QALY effects from VSL estimates. This appears somewhat higher than recent reviews of the literature and the DH and others are currently reviewing this estimate. ONS may wish to adopt a value based on the outcome of this work.

iv. *Future changes in the consumption value of health ( $gv_h$ )*

There are compelling empirical and theoretical reasons why the value of health would be expected to grow with growth in consumption. Although theoretical arguments point to number of empirical questions, a simple but reasonable assessment could be based on growth in consumption (embedded in the STPR) and the income elasticity of demand for health. The DH and the draft 'best practice' report from AAWG reflects a view that the consumption value of health is expected to grow at 2%, which, given real growth of 2% embedded in the STPR, implies an income elasticity of demand for health of one. This might be regarded as a reasonable, albeit potentially conservative, assessment given empirical evidence and other theoretical arguments.

v. *The marginal productivity of health care expenditure in producing net production outside the health care sector*

The DH estimate that a marginal £ of NHS expenditure provides 63p worth of net production gains ( $k_{c1}=1.59$ ). This is based on evidence of the health effects of marginal health expenditure (by age, gender, disease and survival vs quality effects) and the link between these changes in health and net production. As such it provides some assessment of the net production opportunity costs of health care expenditure which can be revised as evidence evolves.

vi. *Future changes in the marginal (net production) productivity of health care expenditure.*

There is currently little direct evidence on how the net production opportunity cost of health care expenditure is likely to evolve. However, a default assumption, which can be revised as evidence evolves, might be that the real value of net production effects of the health effects of expenditure will grow at the same rate as consumption.

Although avoiding dual discounting has much to recommend it, it would require ONS to transform all health effects (benefits and health opportunity costs) into time streams of consumption using an explicit value for  $v_{h1}$  as well as  $gv_h$ , which may conflict with other unquantified social objectives. Given there are also wide and disputed variations in values for  $v_{h1}$  ONS could, for pragmatic reasons, adopt a similar approach to DH by embedding  $gv_h$  in the discount rate for health and health care costs. A reasonable assumption (made implicitly by DH) would be an income elasticity of demand of health of one, in which case  $gv_h$  would be equal to the growth rate assumed in the STPR, so the discount rate for health gains and losses would be 1.5% (i.e., the Green Book rate of 3.5% minus 2% for  $gv_h$ ). Since current evidence also suggests that  $gk_h$  has not grown over the past decade, discounting health care costs over the medium term also at 1.5% (in common with DH and AAWG) would not be unreasonable. This would achieve consistency with DH and AAWG while being founded on reasonable assumptions supported by an assessment of current evidence.<sup>29</sup>

## References

- Aldy, J.E., Smyth, S.J. (2014) Heterogeneity in the value of life. Resources for the Future, Discussion Paper No. 14-13.
- Arrow, K.J. (2012) Social Choice and Individual Values, with an introduction by Eric S Maskin, Newhaven: Cowles Foundation.
- Boadway RW, Bruce N. 1984. Welfare Economics. Blackwell: Oxford.
- Broome, J. (1978). Trying to value a life. Journal of Public Economics, 9: 91-100.
- Brouwer WBF, Niessen LW, Postma MJ, Rutten FFH. (2005). Need for differential discounting of costs and health effects in cost effectiveness analyses. British Medical Journal, 331: 446-448.
- Brouwer W, Culyer AJ, Job N, van Exel A, Rutten FH. (2008). Welfarism vs. extra-welfarism. Journal of Health Economics 27: 325–338.

---

<sup>29</sup> It would also be the same as a discounting policy which was founded on the alternative non welfarist normative position if  $gk_h$  was also assumed to be zero and  $r_s$  was 1.5%, which would not be unreasonable.

Claxton, K, Paulden, M, Gravelle, H, Brouwer, W & Culyer, AJ. (2011). Discounting and decision making in the economic evaluation of health-care technologies, *Health Economics*, 20: 2-15.

Claxton K, Martin S, Soares MO, Rice N, Spackman E, Hinde S, Devlin N, Smith PC, Sculpher M. (2015a). Methods for the estimation of the NICE cost effectiveness threshold. *Health Technology Assessment*, 19:

Claxton K, Sculpher M, Palmer S, Culyer AJ. (2015b). Causes for concern: is NICE failing to uphold its responsibilities to all NHS patients? *Health Economics*, 24:1-7.

Coast, J., Smith, R., and Lorgelly, P. (2008), Welfarism, extra-welfarism and capability: the spread of ideas in health economics. *Social Science and Medicine*, 67: 1190-98.

Costa DL, Kahn ME. Changes in the value of life, 1940–1980. (2004). *Journal of Risk and Uncertainty*, 24: 159–180.

Council of Economic Advisers Discounting for public policy: theory and recent evidence on the merits of updating the discount rate. Council of Economic Advisers Issue Brief, January 2017.

Culyer AJ, McCabe C, Briggs A, Claxton K, Buxton M, Akehurst R, Sculpher M, Brazier J. (2007). Searching for a threshold, not setting one: the role of the National Institute of Health and Clinical Excellence. *Journal of Health Services Research and Policy*, 12: 56-58.

Department of Health. (1996). *Policy Appraisal and Health: A Guide from the Department of Health*. Department of Health: London.

Department of Health. (2004). *Policy Appraisal and Health: A Guide from the Department of Health (Reissued)*. Department of Health: London.

Department of Health. (2010). *A New Value-Based Approach to the Pricing of Branded Medicines - a Consultation*. Department of Health: London.

Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. 4th ed. Oxford: Oxford: Oxford University Press, 2015.

Getzen, T.E., (2000) Healthcare is an individual necessity and a national luxury: applying multilevel decision models to the analysis of health care expenditures. *Journal of Health Economics*, 19: 259-70.

Gravelle H, Smith D. (2001). Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Economics*, 10: 587-599.

- Gravelle H, Brouwer WBF, Niessen LW, Postma MJ, Rutten FFH. (2007). Discounting in economic evaluations: stepping forward towards optimal decision rules. *Health Economics*, 16: 307-317.
- Hall RE, Jones CI. (2007). The value of life and the rise in health spending. *Quarterly Journal of Economics*, 122: 39-72.
- Hammit JK. (2000). Valuing mortality risk: theory and practice. *Environmental Science and Technology* 34: 1396-1400.
- Hammit, J.K., Robinson, L.A., (2011). The income elasticity of the value per statistical life: transferring estimates between high and low income populations. *Journal of Benefit-Cost Analysis*, 2: 2152-2812.
- Lipscomb J, Weinstein MC, Torrance GW. Time preference. In *Cost-Effectiveness in Health and Medicine*, Gold MR, Siegel JE, Russell LB, Weinstein MC (eds). Oxford University Press: Oxford, 1996; 214–246.
- Lindhjelm H, Navrud S, Braathen NA, Biaisque V. 2011. Valuing mortality risk reductions from environmental, transport, and health policies: a global meta-analysis of stated preference studies. *Risk Analysis*, 31: 1381-1407.
- Mason H, Jones-Lee M, Donaldsson C. (2009). Modelling the monetary value of a QALY: a new approach based on UK data. *Health Economics*, 18: 933-950.
- Meltzer, D. O. (2013), 'Future costs in medical cost-effectiveness analysis ', in A. Jones (ed.), *The Elgar Companion to Health Economics* (Cheltenham: Edward Elgar Publishing Ltd), 481-89.
- Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG, editors. *Cost-effectiveness in health and medicine. Second edition. 2nd edition.* New York: Oxford University Press; 2016.
- NICE. (2001). *Technical Guidance for Manufacturers and Sponsors on Making Submissions for a Technology Appraisal.* NICE: London.
- NICE. (2004). *Guide to the Methods of Technology Appraisal.* NICE: London.
- NICE. (2008). *Guide to the Methods of Technology Appraisal.* NICE: London.
- NICE. (2013). *Guide to the Methods of Technology Appraisal.* NICE: London.
- NICE. (2014). *Consultation Paper: Value-Based Assessment of Health Technologies.* NICE: London.
- Nord, E. (2011). Discounting future health benefits: the poverty of consistency arguments. *Health Economics*, 20: 16-26.



- Parsonage M, Neuburger H. (1992). Discounting and health benefits. *Health Economics*, 1: 71-76.
- Paulden, M. and Claxton, K. (2012). Budget allocation and the revealed social rate of time preference for health', *Health Economics*, 21: 612-18.
- Paulden M, O'Mahony JF, and McCabe C. (2017). Discounting the Recommendations of the Second Panel on Cost-Effectiveness in Health and Medicine. *PharmacoEconomics*, 35: 5-13
- Pinto-Prades JL, Loomes G, Brey R. (2009). Trying to estimate a monetary value for the QALY. *Journal of Health Economics*, 28: 553–562.
- Phelps, C.E. and Mushlin, A.I. (1991), On the (near) equivalence of cost-effectiveness and cost-benefit analyses. *International Journal of Technology Assessment in Health Care*, 7: 12-21.
- Rawlins, M. D. and Culyer, A. J. (2004), National Institute for Clinical Excellence and its value judgments. *British Medical Journal*, 329: 224-27.
- Robinson LA, Hammitt JK, Chang AJ, and Resch S. (2016). Understanding and improving the one and three times GDP per capita cost-effectiveness thresholds. *Health Policy Planning*,
- Ryen, L. and Svensson, M. (2014). The willingness to pay for a quality-adjusted life year: a review of the empirical literature. *Health Economics*,
- Sen, A. (1979), Personal utilities and public judgments. Or what's wrong with welfare economics. *The Economic Journal*, 89: 537-58.
- Spackman M. (2017). Social discounting: the SOC/STP divide. Grantham Research Institute on Climate Change and the Environment, Working Paper No. 182.
- Stinnett AA, Mullahy J. (1998). Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 18 (suppl): S68-S80.
- Martin S, Rice N, Smith PC. (2012). Comparing costs and outcomes across programmes of health care. *Health Economics*. 3: 316-337.
- Martin S, Rice N, and Smith PC. (2008). Does health care spending improve health outcomes? Evidence from English programme budgeting data. *Journal of Health Economics*, 27: 826-42.
- Vallejo-Torres L, García-Lorenzo B, Castilla I, Valcárcel-Nazco C, Lidia García-Pérez L, Linertová R, Polentinos-Castro E, Pedro Serrano-Aguilar P. (2016). On the Estimation of the Cost-Effectiveness Threshold: Why, What, How? *Value in Health*.

Viscusi, W.K. and J.E. Aldy, (2003). The value of a statistical life: A critical review of market estimates throughout the world. *Journal of Risk and Uncertainty*, 27: 5-76.

### Section 3: Discount rates for non-financial productive assets

In national accounting, discount rates play a central role in linking a flow element, the cost of capital services or “user cost of capital”, with the balance sheet item of net capital stock. “Whereas the introduction of costs of capital services into the accounts has been of interest in itself, they should also be internally consistent with measures of the net capital stock so that the volume and price measures of capital services, depreciation and net income aggregates in the national accounts as well as balance sheets are fully integrated” (OECD, 2009, p.25). This section considers the use of discount rates in this specific context.

Consider a non-financial asset,  $i$ , that, after time  $t$ , will pay a gross operating surplus  $c_{i,t+1}, c_{i,t+2}, \dots$  to its owner-user in each subsequent time period. Then under the standard discounting approach that is used by the Office for National Statistics, the value of the asset to the owner-user at this time,  $p_{i,t}$ , is given by

$$p_{i,t} = \frac{c_{i,t+1}}{(1+r_i)} + \frac{c_{i,t+2}}{(1+r_i)^2} + \frac{c_{i,t+3}}{(1+r_i)^3} + \dots$$

for some discount rate  $r_i$  that is asset specific. Under the assumption that the economic environment is fully competitive so that monopolistic rents are not available, the gross operating surplus, at least in expectation, will be exactly equal to the cost of capital services.<sup>30</sup> Using discount rates in this way therefore links user costs of capital with the stock of capital and also allows for the calculation of depreciation schedules for capital stock. The OECD (2009) provides a detailed discussion of this background. The specific issue to be addressed here is on how the discount rate should be estimated.

#### 3.1. Ex-ante vs ex-post approaches

At present, the most common preference in national accounting, as recommended by OECD guidance, for example, is to derive the discount rate using an ex-post approach.<sup>31</sup> By multiplying both sides of the NPV equation by  $(1+r_i)$ :

$$p_{i,t}(1+r_i) = c_{i,t+1} + \frac{c_{i,t+2}}{(1+r_i)} + \frac{c_{i,t+3}}{(1+r_i)^2} + \dots = c_{i,t+1} + p_{i,t+1}$$

And rearranging gives the user cost of capital as:

$$c_{i,t} = p_{i,t}(1+r_i) - p_{i,t+1} = p_{i,t}(1+r_i) - p_{i,t}(1+\pi_{i,t})(1-\delta_{i,t}) \approx p_{i,t}(r_i + \delta_{i,t} - \pi_{i,t})$$

<sup>30</sup> In a limited number of situations, the cost of capital services might be observable from the rental paid by the asset's user to the owner. This, though, is not a common circumstance, and rentals also include the cost elements of managing the rental business.

<sup>31</sup> As a counterexample, UN guidance prefers an exogenous approach to discount rates.

where  $\pi_{i,t}$ ,  $\delta_{i,t}$  are respectively the revaluation in price of the production asset over the year had it remained unused, and its annual rate of depreciation through ageing. It then follows that the discount rate is given by:

$$r_i = \frac{c_{i,t}}{p_{i,t}} + \pi_{i,t} - \delta_{i,t}$$

Since the net stock of the capital asset, the gross operating profit, the depreciation schedule and the inflation rate are all assumed known, the discount rate can be derived.

There are a number of problems with this approach; we briefly discuss three here. First, there is the issue of endogeneity. The discount rate is an input into the depreciation schedule of the asset over time. Despite this, the expression for the discount rate has  $\delta_{i,t}$  as an exogenous variable. This problem of implicit endogeneity can be overcome under the assumption of a geometric age-efficiency profile, when  $\delta_{i,t}$  exactly matches the age-efficiency profile (OECD, 2009), but it is not always appropriate to make this assumption. As examples, intellectual property does not decline over time, and the observed market prices of aging transport equipment are inconsistent with a geometric age-efficiency profile. In these cases, it is necessary to solve systems of non-linear equations to co-solve for the discount rate and the depreciation schedule, which is a complex process. Second, there is a core difference between expected and realised benefits from projects, even when aggregated across assets and time. In sustained periods of growth (decline), realised profits are likely to exceed (underperform against) expected profits, leading to an over (under) estimate of the true cost of capital. For some, this is seen as being an advantage of using an ex-post approach: it does not require the ONS to account for unanticipated profits and losses to capital. Yet there remains a clear distinction between the costs that a user requires in expectation for investing capital in an asset and the realised return that will, in practice, exceed or fall short of this expectation. This, after all, is the risk inherent in project ownership. Ahmad (2004) discusses this point more widely in the context of placing capital services in the production account. Third, there is the difficulty of assigning gross operating profits to a single set of assets in production processes that may involve many different asset classes in the supply chain. This is particularly true in situations that involve non-produced assets (e.g., land, other natural resources, radio spectrum), inventories (e.g., works in progress, materials and supplies), and other 'intangible' assets which are not included in the national accounts asset boundary.

The greatest difficulty with this approach, though, is that it does not reflect how, in practice, either private or public sector users do actually allocate costs to capital in most contexts. This point is particularly important in the context of this wider report, where we have been asked to assess the consistency of discounting practices by the ONS across its wide remit of activities.

As is evidenced by all the main textbooks in the field of Corporate Finance, and in surveys of practice (see, for example, Bruner *et al.* 1998; Arnold and Hatzopoulos, 2000; Gitman and Vandenberg, 2000; Burns and Walker, 2009; Bancel and Mitto 2014), the Capital Asset Pricing Model dominates in the corporate world. In the public sector, consumption based asset pricing models that underlie the Ramsey Model are most widely used. While these are different models they have a shared theoretical basis providing consistency between public and private sector discounting.

Consider a user in either the public or private sector who is considering investing in an asset today for a price of  $p_{i,t}$  in exchange for a future stream of gross operating surpluses;  $c_{i,t+1}, c_{i,t+2}, \dots$ .<sup>32</sup> Then the price of this asset can be decomposed into  $p_{i,t} = \rho_{i,t,t+1} + \rho_{i,t,t+2} + \dots$  where  $\rho_{i,t,t+\tau}$  is the price that the user would pay at time  $t$  for receipt of the gross operating surplus  $c_{i,t+\tau}$  at time  $t + \tau$  and zero at all other times. To determine  $\rho_{i,t,t+\tau}$ , the most basic model is known as the “fundamental theorem of asset pricing” (see, for example, Cochrane 2005). This states that, for some  $\varphi_{t,t+\tau}$ , which is known as the “pricing kernel” or “stochastic discount factor”, which is independent of the asset/project,  $i$ , being valued:

$$\rho_{i,t,t+\tau} = E_t[c_{i,t+\tau}\varphi_{t,t+\tau}] = \frac{E_t[c_{i,t+\tau}]}{(1 + r_{i,t,t+\tau})^\tau}$$

Re-arranging this gives a highly generic expression for the discount rate that can be used in (almost) all situations:<sup>33</sup>

$$r_{i,t,t+\tau} = \left( E_t[\varphi_{t,t+\tau}] + Cov_t \left[ \frac{c_{i,t+\tau}}{E_t[c_{i,t+\tau}]}, \varphi_{t,t+\tau} \right] \right)^{-1/\tau} - 1$$

The user cost of capital is then calculated as  $\sum_\tau \rho_{i,t,t+\tau} r_{i,t,t+\tau}$ . Taking this formal approach to valuation reveals several things about the Net Present Value equation used in estimating capital stocks:

1. It is the expected benefit to the user,  $E_t[c_{i,t+\tau}]$ , that should be used in the equation and not the realised benefit. Therefore, from a purely theoretical perspective, the ex-ante approach is to be preferred.
2. The discount rate varies both over time ( $t$ ) and, at any given time, by the maturity of the cash flow ( $\tau$ ). While it is common in many applications of the NPV equation to use a single discount rate for all maturities, this is not theoretically robust.

<sup>32</sup> It should be noted that the public/private distinction that is prevalent in the discounting literature does not exactly mirror the way in which national accounts reflect a market/non-market sector split. All corporations, public, private or foreign owned, are considered market producers, general (central and local) government units are considered non-market producers, while households may be considered market or non-market depending on their activity. It is the market/non-market criteria that currently determines whether the ONS include a rate of return or not, rather than the public/private one.

<sup>33</sup> The key assumption is that valuation is linear: the price of buying two of project A and three of project B must equal twice the price of one project A plus three times the price of one project B. Gross operating surpluses must also have finite variance.

3. The discount rate will incorporate both a risk-free element (determined by  $E_t[\varphi_{t,t+\tau}]$ , which is the same for all assets) and a risk premium (determined by the covariance term, which varies from asset-to-asset through  $c_{i,t+\tau}$ ).

While the fundamental theorem is useful for providing a general conceptual framework for understanding discounting in this context by illustrating the theoretical strength of the ex-ante approach, in practice it is necessary to make additional assumptions to identify the pricing kernel. It is in this regard that the private and public sector differ in determining the user cost of capital in practice. Armitage (2017) provides a detailed discussion, particularly in the context of long-term discounting, of the differences in approach taken by corporations and governments when calculating discount rates.

### 3.2. User costs in the private sector

As evidenced by a wide range of surveys of corporate finance, cited in the previous section, the cost of private capital is most commonly determined through the Capital Asset Pricing Model (CAPM). This is based on the assumption that investors wish to maximise the expected return to their capital while reducing the risk (as measured by the standard deviation) of their realised returns. In this case, the pricing kernel is a simple linear function,  $\varphi_{t,t+\tau} = a + bc_{m,t+\tau}$ , of the payoffs,  $c_{m,t+\tau}$ , from a value-weighted market portfolio of all available projects. This leads to the well-known formula

$$r_{i,t} = r_{ft} + \beta_i E_t[r_{m,t} - r_{ft}], \text{ where } \beta_i = \frac{\text{Cov}[r_{i,t}, r_{m,t}]}{\text{Var}[r_{m,t}]}$$

where  $r_{ft}$  is a risk-free rate and  $r_{m,t}$  is the return to the market portfolio.

There are a number of theoretical difficulties with using the CAPM. First, it is a single-period, rather than a multi-period, model; the formula for the rate of return does not depend on the horizon of the cash flow,  $\tau$ . While formal intertemporal variations of the CAPM do exist (Merton, 1973; Campbell, 1993), they are not commonly applied in practice; instead, it is usual to apply the single-period model at all cash flow horizons as if it had been derived in an inter-temporal context. An important consequence of this is that it is then not theoretically clear what maturity of Treasury security should be used to identify the risk-free rate. Two predominant schools of thought exist. The first equates  $r_{ft}$  with the yield on the 3-month maturity Treasury bill, as this has the lowest risk of any marketable security. The second prefers to use the yield to maturity on the Treasury strip (zero-coupon bond) that matches the horizon of the cash flow. While these longer-life bonds suffer from inflation risk, and so are not “risk-free”, they better capture the yield curve at the time when the analysis is being undertaken. Within an intertemporal CAPM, resolving these issues is theoretically straightforward, but it becomes ambiguous when a single-period model is used in a multi-period

context. Freeman (2009) discusses this point in detail and finds the theoretical case for using bond yields stronger than for using bill yields, particularly for real (inflation-adjusted) valuation. In the corporate sector, the strong preference is also to use bond yields rather than bill yields (e.g., Bancel and Mittoo 2014). This allows for the user cost of capital to depend on the life of the asset, consistent with the theoretical framework outlined above.

Second, there is considerable difficulty identifying the “market portfolio” for use in this model. In theory, this contains all available assets. But since we can invest in fine art, fine wine, human capital, and a range of other assets around the world, such a portfolio cannot be observed. It is therefore necessary instead to choose a market proxy using a limited number of assets, and then use this as a surrogate for the true market portfolio. The most common choice is a broad equity market index such as the FTSE100 in the UK and the S&P500 in the US. But apparently small differences between the true market portfolio and its surrogate can have significant effects on the accuracy of the model. As famously demonstrated by Roll (1977), if this surrogate portfolio is mean-variance efficient in that it has the highest expected return for a given variance, then the CAPM is tautologically accurate: the CAPM equation is just a mathematical restatement of the mean-variance efficiency of the chosen surrogate. If, instead, a surrogate portfolio is chosen that is not mean-variance efficient, then there is no reason to believe that the CAPM will hold even as an approximation. Kandel and Stambaugh (1995), for example, demonstrate that even if the chosen surrogate is only slightly mean-variance inefficient, then the relationship between the estimated beta and the cost of capital might not even approximate the linear CAPM model.

Third, there are a number of other core assumptions that underlie the CAPM, which include all investors agreeing on the expected return and (co-)variance to (between) each asset. Even in a domestic context this is unlikely to be the case; even less so in an international context with investors taking gains in different currencies. It is also necessary to accept that investors care only about the mean and variance of returns (and not the higher moments of the distribution), and that all investors can lend and borrow at the same risk-free rate. Given the stringency of these assumptions, many would argue that, while the model is an elegant description of how risk influences valuation, it does not capture all the salient features of real-world investment.

These theoretical issues are particularly important because of the lack of empirical support for the CAPM: see, for example, Fama and French (2004) for a review. Nearly all tests of the validity of the model use returns on equities as their core data as these are easily observable. While early studies appeared to show a clear positive, and near-linear, relationship between the betas of different portfolios of shares and their average returns, more recent work shows that this effect disappears when other factors which may proxy for risk are taken into account. One of the most important academic papers in finance of recent times is Fama and French (1992), which has over fifteen thousand citations according to Google Scholar. They show that once both the size of a firm and whether or not it sits in a growth industry are taken into

account, beta has little if any explanatory power for the difference in average observed returns for assets with different systematic risk (beta). Of course, such empirical tests have been subject to extremely widespread debate and critique; for example, that such tests are based on realised returns, not unobservable expected returns, and do not always allow for changes in parameter values over time. Nevertheless, it is accurate to say that the CAPM, despite being over half a century old, remains unproven in its ability to explain the cross-section of behaviour of different equity prices. As a consequence, some have now argued that adjusting the cost of capital not only for beta but also for the firm's market capitalisation and whether it sits in a growth or value industry, through a "three-factor" model, is most appropriate (e.g., Estrada 2011). While there is a strong academic case for supporting this approach, as yet it has not been widely implemented in practice.

This leads to a significant paradox. Why is the CAPM model still in such widespread practical use if the empirical evidence to support it is so hotly contested and if the theoretical assumptions are so restrictive? Is it that, as Joshi (2013) suggests, "practitioners continue to use the model in their investment processes (to a varying degrees) because they lack a better alternative"? Or is it that the theoretical appeal and elegance of the model, together with its continued widespread dissemination in world-leading Business Schools, lead practitioners to overlook its empirical and theoretical shortcomings? Da, Guo and Jagannathan (2012) argue instead that the poor empirical performance of the model arises instead because tests have been run on equity returns, and corporations are considerably more complex operations to value than projects themselves. In particular, firms have embedded "real options" to grow or modify existing projects and the valuation of such options falls outside the standard NPV framework. Once these options effects are removed, they claim that the use of the CAPM in a project context is still justified.

Therefore, despite the many difficulties with the model that have been extensively discussed in the academic literature, the Office for National Statistics could reasonably justify the use of the CAPM to estimate the user cost of capital for the private sector. The two main motivations for this are that it is a genuinely ex-ante model, as the theory of valuation dictates, and is very widely used in practice. In addition, while the ability of the CAPM to explain equity returns has been widely disputed, the case is stronger for using the model on the primary assets of the firm. However, should it decide to do so, there are a number of empirical issues for the ONS to deal with. How these are addressed in practice is discussed in Bancel and Mittoo (2014).

1. The risk-free rate. As discussed above, because the CAPM is a single-period model applied in a multi-period context, the "risk-free asset" is ambiguously defined. Treasury bills and longer-term Treasury bonds do not generally have the same yields because of the shape of the term structure of interest rates. We would recommend that the ONS follows the common corporate practice of preferring bond yields over bill yields, as academically justified by Freeman (2009).



2. Beta estimates for portfolios of projects. The standard technique is to start initially with equity betas of companies that specialise in similar projects. These can be easily calculated from first principles using stock market returns, or obtained from financial websites. Next, these are “de-gearred” to account for the difference between equity risk and asset risk that arises from financial leverage. This process is described in all mainstream textbooks on corporate finance. Individual betas can then be averaged within any given sector. This helps remove the estimation errors that arise within single calculations. This average sector asset beta can then be used in the CAPM to calculate the user cost of capital. However, given the presence of real options in corporations, it can also be argued that the asset beta needs to be further adjusted to remove this effect. Bernardo *et al.* (2012) discuss this point and provide average project betas for a range of industries in the US. This, we believe, provides a useful starting point for estimating the systematic risks of different sets of similar projects. However, as CAPM betas have been widely reported to vary significantly over time, this introduces an additional complexity in this area. While there are a range of empirical approaches that can be used to address this issue, (see, for example, Choudhary and Wu, 2008), again it is not common to use these in practice. As a consequence, beta estimates by industry are likely to be imprecise, resulting in estimation error in the ex-ante cost of capital.
3. The equity premium,  $E_t[r_{m,t} - r_{ft}]$ . There is a vast academic and practitioner literature on how this parameter can be estimated. Traditionally, the preference has been to use realised historic averages of the difference between equity and Treasury returns. This approach, though, has now largely fallen out of favour. This is because, to get any precision in the estimate using this technique, it is necessary to use very long time horizons of returns. This is problematic because there is increasing evidence that the equity premium value changes over time (see, for example, Lettau *et al.* 2008). Theoretical approaches are also not commonly used as they generally produce estimates of the equity premium that are commonly believed to be too low; this is the famous “equity premium puzzle” of Mehra and Prescott (1985). For this reason, there is now increased interest in genuine ex-ante measures of this risk premium as obtained by surveying experts. Pablo Fernandez at IESE Business School, Universidad de Navarra, has run and reported on a number of such surveys recently. Overall, there is something of a consensus that the value of this variable is around 4% at present when using Treasury bond yields as the proxy for the risk-free rate (e.g., FSA 2012).
4. The beta estimate for the vast majority of industries will combine the systematic risks of a range of complex projects within that industry. All of these are likely to differ in their risk profile. When using industry average values, disentangling these betas into their constituent parts would be extremely challenging.

Take an example. Suppose the ONS wishes to estimate the user cost of capital on shipping containers owned in the private sector using the ex-ante CAPM approach. According to Bernardo *et al.* (2012), this sector has a project beta of 0.73. The yield

on 10-year UK bonds at the time of writing is close to 1% nominal and using an equity premium of 4% over bond yields, this gives an ex-ante discount rate of 3.9% nominal. Therefore, for each £1m of capital employed in this sector, the user cost is £39,000 per annum. Note that this will be realised through a combination of the anticipated gross operating profits and expected capital gains/losses associated with owning these assets.

We briefly note that this approach is not appropriate for all assets; specifically we refer here to those employed in research and development. The core purpose of such activity is to generate initiatives that can either be later expanded or discarded. Such projects violate a core assumption of discounting models that, once a project is set up and running, then it is then left broadly unchanged from its original plans. By contrast, R&D activity is more accurately modelled through the use of real options. While a discussion of this framework lies outside the scope of the current report, in general the user cost of capital will be much lower than is implied by the NPV model because of the value of flexibility that is implicit in such activity. We refer the ONS to one of the large range of textbooks on this topic for more detail (e.g., Guthrie, 2009).

### 3.3. User costs in the public sector<sup>34</sup>

As widely discussed elsewhere in this report, the public sector does not use the CAPM in order to determine the appropriate discount rate, preferring instead to use a consumption based approach to determine value for money. This can also be interpreted within the context of the fundamental theorem. In this case, the pricing kernel is given by  $\varphi_{t,t+\tau} = u'(C_{\tau}, \tau) / u'(C_t, t)$  where  $C_t$  is average real per-capita consumption at time  $t$  and  $u'(C_t, t)$  is the marginal utility derived from that consumption. In the absence of a risk premium, and with an isoelastic utility function, this results in the standard Ramsey equation. This is the framework of the Treasury Green Book and results in the standard 3.5% real discount rate when parameters are chosen on a normative/prescriptive basis. This has the major advantage of providing consistency across different areas of discounting within the public sector and reconciles this aspect of national accounting with Treasury valuation recommendations.

Alternatively, the ONS could more closely mimic private sector practices by interpreting the Ramsey equation within a descriptive framework, which would link the user cost of capital with the observed yields on bonds. The average cost of existing borrowing for the UK government is estimated to be around 3.4% (NAO, 2015), which is a little below the Treasury rate because it is not inflation-adjusted. By contrast, this is significantly above the 1% nominal rate that the government can borrow at present for ten years, and is close to the risk-adjusted cost of capital for shipping containers

---

<sup>34</sup> Current international guidance (and European regulation) means that the ONS cannot estimate a rate of return on assets for non-market units beyond their consumption of fixed capital (depreciation). This is mainly for reasons of international comparability and the definition of non-market output in national accounts terms. Nevertheless, for satellite accounts and for related analytical work, such as estimating public sector productivity, the effect of such a return is of great interest.

calculated in the previous section. This emphasises that user costs of capital for new funding are significantly lower than the rates that the government has historically tied itself into.

A specific issue arises in the case of assets that are held under PFI deals. According to the NAO (2015), the implied cost of capital on such projects is close to 7.3%. This is because, when using private finance in the public sector, the funder requires compensation for the risk involved. This potentially creates distortions in the market: “We also believe that many public-private partnerships exist just because of the discrepancies in the way the two sectors evaluate the cost of risk. This potentially generated a massive transfer of risk from the private sector to the public one” (Baumstark and Gollier, 2014). Nevertheless, when calculating a user cost of capital for PFIs, the ONS will need to decide whether to use the government’s own risk-free approach or incorporate risk into the discount rate in the way that the private sector funder requires. This discrepancy can be partially overcome if, within the consumption approach preferred by the Treasury, a risk premium is incorporated into the Governmental discount rate. As we have discussed elsewhere in this report and its companion, we believe that there are strong reasons for both the Treasury and the ONS to adapt their approaches to more explicitly incorporate risk.<sup>35</sup>

### **3.4. Conclusion**

The preference for calculating the user cost of capital in productivity analysis is to take an ex-post approach. We find this somewhat surprising because this is not how users do actually determine their cost of funds. In the private sector, it is well-evidenced that corporations use the Capital Asset Pricing Model. In the governmental sector, the Treasury Green Book rate is widely used in other contexts. We see few compelling reasons to take a different approach here from what occurs in other areas of practice; the theory of discounting dictates that an ex-ante approach should be preferred. While this will require the ONS to account for the difference between the cost of user capital and the realised gross operating surplus, a simple “unanticipated returns” line would appear to fully reconcile this difference (see also Ahmad 2004).

There are, though, a number of difficult choices for the ONS to take if it decides to invoke an ex-ante approach. Within a CAPM context, both the beta of the asset class and the equity premium itself are not easily observable, change over time, and are subject to considerable controversy over how best to estimate. In a governmental context, the ONS must decide whether to use the Green Book rate based on a normative approach to discounting, or take a more descriptive approach by using the yields on Government bonds. In the latter case, the ONS would also need to decide whether to use interest rates that the government has previously locked into, or the cost of new capital as determined by today’s Treasury bond yields. Finally, the issue of whether to incorporate a risk premium into the governmental cost of capital should

---

<sup>35</sup> Freeman, M.C. & B. Groom (2016), “Discounting for Environmental Accounts: Report for the Office for National Statistics”, Office for National Statistics

be determined. This issue becomes particularly apparent because the observed cost of funds on PFI initiatives are so much higher than both prevailing bond yields and the Green Book rate. As we have explained elsewhere in these reports, we believe there are strong reasons for both the Treasury and the ONS to adjust their practices to more explicitly account for risk in the public sector discount rate.

Finally, we observe that the Net Present Value approach itself is not appropriate for use in all sectors when determining the user cost of capital. This is particularly true for situations where managers have options to expand, contract or abandon projects; for example, when considering assets involved in research and development activities. Here, the ONS should consider further consulting the literature on real options.

## References

Ahmad, N. (2004). Introducing capital services into the production account. Paper for the Canberra II Group (Second) Meeting on the Measurement of Non-Financial Assets

Arnold, G.C., & Hatzopoulos, P.D. (2000). The theory-practice gap in capital budgeting: Evidence from the United Kingdom. *Journal of Business Finance and Accounting*, 27: 603-626.

Armitage, S. (2017). Discount rates for long-term projects: The cost of capital and social discount rate compared. *The European Journal of Finance*, 23: 60-79.

Bancel, F., & Mittoo, U.R. (2014). The gap between the theory and practice of corporate valuation: Survey of European experts. *Journal of Applied Corporate Finance*, 26: 106-117.

Baumstark L., & Gollier, C. (2014). The relevance and the limits of the Arrow-Lind Theorem. *Journal of Natural Resources Policy Research*, 6: 45-49.

Bernardo, A.E., Chowdhry, B., & Goyal, A. (2012). Assessing project risk. *Journal of Applied Corporate Finance*, 24: 94-100.

Bruner, R.F., Eades, K.M., Harris, R.S., & Higgins, R.C. (1998). Best practices in estimating the cost of capital: Survey and synthesis. *Financial Practice and Education*, 13-28.

Burns, R., & Walker, J. (2009). Capital budgeting surveys: The future is now. *Journal of Applied Finance*, 19: 78-90.

Campbell, J.H. (1993). Intertemporal asset pricing without consumption data. *American Economic Review*, 83: 487-512.

Cochrane, J.H. (2005). *Asset Pricing* (revised edition). Princeton University Press.

Choudhary, T. & Wu, H. (2008). Forecasting ability of GARCH vs Kalman filter method: Evidence from daily UK time-varying beta. *Journal of Forecasting*, 27: 670–689.

Da, Z., Guo, R.-J. & Jagannathan, R. (2012). CAPM for estimating the cost of equity capital: interpreting the empirical evidence. *Journal of Financial Economics*, 103: 204–220.

Estrada, J. (2011). The three-factor model: A practitioner's guide. *Journal of Applied Corporate Finance*, 23: 77–84.

Fama, E.F. & French, K. (2002). The cross-section of expected stock returns. *Journal of Finance*, 47: 427–465.

Fama, E.F. & French, K. (2004). The Capital Asset Pricing Model: Theory and evidence. *Journal of Economic Perspectives*, 18: 25–46.

FSA (2012), Rates of return for FSA prescribed projections, Financial Services Authority, London, England.

Freeman M.C. (2009). The practice of estimating the term structure of discount rates. *Global Finance Journal*, 19: 219–234.

Gitman, L. J., & Vandenberg, P.A. (2000). Cost of capital techniques used by major US firms: 1997 vs. 1980. *Financial Practice and Education*, 53–68.

Guthrie, G. (2009). *Real Options in Theory and Practice*. Oxford University Press. Oxford, England.

Joshi, N. (2013). A new perspective on the validity of the CAPM: Still alive and well (digest summary), *CFA Digest*, 42:

<http://www.cfapubs.org/doi/full/10.2469/dig.v43.n2.30>

Kandel, S. and Stambaugh, R.F. (1995). Portfolio inefficiency and the cross-section of expected returns. *The Journal of Finance*, 50: 157–184.

Lettau, M., Ludvigson S.C., & Wachter, J.A. (2008). The declining equity premium: what role does macroeconomic risk play? *Review of Financial Studies*, 21: 1653–1687.

Mehra, R., & Prescott, E.C. (1985). The equity premium: a puzzle. *Journal of Monetary Economics*, 15: 145–161.

Merton, R.C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41: 867–88.

NAO (2015). *The choice of finance for capital investment*, National Audit Office, HM Treasury, London, UK.

OECD (2009), Measuring capital: OECD manual, 2<sup>nd</sup> edition.

Roll, R. (1977). A critique of the asset pricing theory's tests. Part I: On past and potential testability of the theory. *Journal of Financial Economics*, 4: 129-176.

## Section 4. Pensions discounting.

This section looks at the different discounting approaches that the Office for National Statistics (ONS) could apply to discount pension liabilities in National Accounts. As with all areas of governmental finance, there are several perspectives as to what the 'correct' discount rate for governmental pension liabilities should be and, as with all debates on discounting and the underlying economic rationale for preferring one approach to another, it is a hotly contested issue.

The current approach of the ONS for discounting pension liabilities in the National Accounts complies with the methods set out in the 2008 System of National Accounts (SNA2008) and the 2010 European System of Accounts (ESA2010).<sup>36</sup> The aim of valuing and reporting pensions in National Accounts is to show the pension liabilities of a country, including funded private sector pensions, funded government pensions, unfunded public service pensions, and state pensions. These obligations make up a significant part of the liabilities that sit on a country's balance sheet and so arriving at a best estimate of this is fundamentally important.

The current EU-wide requirement for discounting pension liabilities is to use a stable discount rate of 3% real or 5% nominal when discounting unfunded government pension liabilities and the unfunded part of the local government pension schemes (LGPS). In arriving at these rates, four principles were applied. First, the rate should be based upon a basket of European central government debt securities. Second, these securities should have a maturity of 10 years or longer given the duration of pension liabilities. Third, that the same discount rate should be used across the EU to allow for comparability of National Accounts, and fourth, that the discount rate should be stable to avoid incorporating noise into these estimates.<sup>37</sup> At the time that these rates (3% real or 5% nominal) were agreed they were considered comparable to a social time preference rate (STP) but with the option to adjust the rate in the future. Since setting the rate it has not been changed in the intervening years, despite lobbying to do so to reflect shifts in the yield on government debt post 2008.<sup>38</sup> In looking at the four criteria below, there is an underlying rationale for each one.<sup>39</sup>

The application of a basket of government securities, rather than the use of a single country's long-dated government debt, is motivated by two factors. First, there are differing levels of default risk across countries, as no country is wholly risk-free, and consequently, the securities of different governments have different risk premiums.<sup>40</sup>

---

<sup>36</sup> Levy (2012), Pensions in the National Accounts – A Fuller Picture of the UK's Funded and Unfunded Pension Obligations.

<sup>37</sup> Technical Compilation Guide for Pension Data in National Accounts, The European Commission (Eurostat) <http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-11-027>

<sup>38</sup> This information comes from parties involved in the setting of the original rates and discussions about the appropriate rate post 2008. These rates are set by the Working Group on Ageing Populations and Sustainability, which is now part of the European Commission's Directorate-General for Economic and Financial Affairs.

<sup>39</sup> *ibid*

<sup>40</sup> As a rather stark example, the current yield on 25-year Greek government debt at the time of writing (November 2017) is approximately 5.72% while the yield on German 30-year government debt is approximately 1.17%.

Second, the discounted value of a pension liability is directly related to the yield on government debt of a country. As a result, if there were to be large increases in the award of unfunded state pension benefits in a country, this would naturally increase the risk of default of that country. The consequence of this would be an increase in the yield on government debt to reflect the increased risk. If a country specific yield were to be used as the appropriate discount rate in such circumstances, this increased yield could result in a lower pension liability, given the sensitivities of discounting pension liabilities, thereby masking the true extent of the awarded increase and understate the overall liability of the country.<sup>41</sup>

The use of long-dated government securities is practical given the duration of pension liabilities. In addition, the preferred maturity is one of 30 years as this is a good approximation of the duration of pensions in payment given increasing longevity across Europe.

In setting the same rate for all EU countries, this allows for comparability in the National Accounts across countries. If different country specific rates were used i.e. individual country yields, then this would prevent comparability, and even small changes in the rate that is applied, would have significant impacts on the estimation of the liability.

Consistent with the need for comparability across countries, the use of a stable discount rate allows for comparability through time. If the discount rate were to be updated annually based upon short-run deviations in the market for government securities for example, then it would not be possible to understand how the liabilities in country are evolving. Moreover, and as has been demonstrated in private sector schemes, volatility in the pension liability becomes difficult to explain to stakeholders<sup>42</sup> and in the case of public sector pension liabilities, this would not be easily explained to the taxpayer.

#### **4.1. Discounting and the cost of unfunded public service pension schemes**

While the aim of this report is to examine the discount rate used for the estimation of pension liabilities for National Accounting purposes, it is worth examining how the government uses discount rates to set the contributions to unfunded public service pension schemes and the debates that exist here. The current approach within National Accounting is to apply a uniform discount rate where the government is the pension manager.<sup>43</sup> However, the debates as to what approach is best for setting the discount rate for the estimation of contribution rates also hold for the valuation of liabilities more broadly.<sup>44</sup>

---

<sup>41</sup> See Levy (2012), Pensions in the National Accounts – A Fuller Picture of the UK's Funded and Unfunded Pension Obligations, for an illustration of the magnitude of these sensitivities.

<sup>42</sup> Accounting for Pensions: A Report for the National Association of Pension Funds, Clacher and Moizer (2012).

<sup>43</sup> Levy (2012)

<sup>44</sup> A key issue in all such debates is the question as to what the discount rate is being used for? Under the current approach for National Accounting, the goal appears to be one of comparability of the liabilities. However, this rate does not reflect the cost of the unfunded public service pensions or the state pension.



#### *4.2.1. The SCAPE Discount Rate*

The Superannuation Contributions Adjusted for Past Experience (SCAPE) discount rate is used to set contribution rates for the unfunded public sector pension schemes, which cover teachers, the NHS, police, firefighters, armed forces, and civil servants. Historically, this rate was set at 3.5% plus RPI inflation, which was in line with STPR, although it was never used to discount pension liabilities as prior to this it was assumed that assets were equal to liabilities. However, this rate was reviewed in 2010<sup>45</sup> and was reduced by 0.5 percentage points to 3%, the consequence of which was an estimated increase in the cost of contributions by 3%.<sup>46</sup> In 2011, the ONS set out the method for the first Supplementary Table on Pensions, which is the method now used in the UK National Accounts. At this time, both the SCAPE rate and the EU-Wide rate were exactly same.<sup>47</sup> Although, more recently, the SCAPE rate has been reduced down to 2.8% for 2019-20, which is estimated to increase contributions by £2bn per annum.<sup>48</sup>

## **4.2. Alternatives to the current approach**

In looking at the different approaches to discounting, the way in which pension liabilities are valued for National Accounts is currently the accepted best practice within Europe and achieves comparability across both countries and through time. Moreover, the ONS went further than many European countries in being the first to present some sensitivity around the estimated discount rate and released a supplementary table for the 2010 National Accounts. However, there are alternatives to the current regime, some of which have been used previously in government finances in relation to pensions i.e. the Social Time Preference Rate.

#### *4.2.1. Discounting Consistently with Funded Private Sector Schemes*

Funded private sector defined benefit schemes discount their pension liabilities with reference to market rates. The current approach within the UK is to apply a gilts+ methodology for triennial valuation purposes. Here, the current yield on gilts is taken and an adjustment based on risk is added to reflect sponsor covenant and the investments held by the pension fund with some margin for prudence with respect to investment returns.<sup>49</sup> However, this approach differs from the current accounting standard, IAS 19, which relies on a AA bond yield. Underpinning the approaches that are used within private sector are a number of different drivers, none of which are

---

<sup>45</sup> HM Treasury, Consultation on the Discount Rate Used to set Unfunded Pension Contributions, December 2010.

<sup>46</sup> Independent Public Service Pensions Commission: Interim Report, October 2010.

<sup>47</sup> Levy (2011) Pensions in the National Accounts: Compiling a Complete Picture of UK Pensions Including Unfunded Pensions for Public Sector Employees

<sup>48</sup> Thurley (2016) Public Service Pensions – Employer Contributions, Briefing Paper, House of Commons Library.

<sup>49</sup> There are a number of different estimates provided for regulatory purposes e.g. technical provisions, best estimate, and S.179 valuations.

widely accepted as being correct and there remain fierce debates about how best to discount defined benefit pension liabilities within private sector schemes.<sup>50</sup>

One motivation for using market based corporate bond yields for all government pension liabilities is that such an approach is consistent with how private sector companies account for their pension scheme liabilities. As such, the liabilities disclosed by Government in accounting for their pension obligations would be directly comparable to those of private sector companies.<sup>51</sup>

The underlying intellectual basis for discounting of pension liabilities using a market rate or a market rate with an explicit adjustment comes from financial economics whereby the market rate is the 'true' cost of the pension to the sponsor at a point in time. As such, the discounting of a stream of financial payments should be at a rate that reflects their risk (Modigliani and Miller, 1958), and their covariance with priced risks (Treynor, 1961; Sharpe, 1964; Lintner, 1965). Other methods of discounting, such as expected return on plan assets, which is used in funded public pension promises in the US, does not adhere to the approach of financial economics. Consequently, the current approach skews investment towards risk assets, such as equities, as this inflates the discount rate and understates the 'true' cost liability (Novy-Marx and Rauh, 2011).

However, the approach of financial economics may not be appropriate in a governmental context. For a government, the costs of government debt do not reflect the future cost of meeting pension payments, as the costs and frictions that occur in private sector pension fund management do not affect the government.

In looking at the costs of unfunded pensions, three key issues emerge. First, the government does not have to bear the significant costs associated with fund management, as pensions are often unfunded.<sup>52</sup> These costs are significant for both private sector and funded defined benefit pension schemes such as the LGPS.<sup>53</sup> Moreover, the government stands behind public sector pension schemes as the sponsor and so there is a strong employer covenant. Third, private sector schemes attempt to pay benefits through a mixture of employer and employee contributions and investment returns, if investment returns to a fund are high, then this will lower the cost of provision in the future, which is not something the government can achieve, as the schemes are unfunded.

#### *4.2.2. Discounting using the yield on indexed-linked gilts*

---

<sup>50</sup> See, Our Mad Approach to Pension Fund Deficits, Anthony Hilton, London Evening Standard and Why Anthony Hilton is Wrong about Defined Benefit Pensions, Dan Mikulskis, Professional Pensions.

<sup>51</sup> Public Sector Pension Schemes: Policy Objectives and Options for the Future, The Pensions Policy Institute 2010.

<sup>52</sup> Consultation on the Discount Rate used to set Unfunded Public Service Pension Contributions, Summary of Responses, HM Treasury 2011.

<sup>53</sup> See the Asset Management Market Study Interim Report, Financial Conduct Authority, 2016.

There are many advocates for the discounting of government pension liabilities using the yield on index-linked gilts. The rationale for such an approach is that the liability is best measured with reference to a market rate that reflects the cost, maturity, and risk of a scheme if it were to be funded and backed by assets, and for the government the yield on indexed-linked gilts is the rate that reflects this.<sup>54</sup> Moreover, the yield on indexed-linked gilts prices the value that markets place on future government revenues and as such reflects the cost of provision.<sup>55</sup>

Implicit in the use of index-linked gilts for the valuation of public pension liabilities is the view that these promises are a binding debt of the government. This does not logically follow. While debt is issued under specific terms and, assuming no default and the holding of the debt to maturity, these terms are immutable, this is not the case with public service pension liabilities. Pensions are subject to negotiation and collective bargaining arrangements, which often take place based on cost.<sup>56</sup> As a result, pensions are classified as contingent pension obligations in the National Accounts.<sup>57</sup>

Further, the use of indexed-linked gilts for the valuation of aggregate governmental pension liabilities may not be wholly appropriate as it overstates the cost to government. If an individual were to purchase an unfunded public sector pension in the market, then this would be valued in relation to the yield on indexed-linked gilts, as this would reflect the cost, maturity, and risk of such a scheme to an individual. However, this is not the cost to the government as there are frictional costs to the purchaser of such securities, which the government does not incur.

Moreover, and most importantly from a National Accounting perspective, the yield on indexed-linked gilts can be subject to considerable variation due to external forces. The most recent example of this being quantitative easing and the policy response to the global financial crisis. Consequently, it does not provide a stable discount rate to allow for the long-term valuation and funding of unfunded public sector pension liabilities. Such volatility would prevent any meaningful comparison through time with how the 'true' costs of public service pension schemes evolve.

#### *4.2.3. Discounting using the social time preference rate (STP)*

Historically, the government has used an STP for the discounting of public service pension liabilities. This was set at a real rate above RPI of 3.5% per annum, which was based on the Ramsey Rule following the Green Book.<sup>58</sup> There are a number of advantages to the use of the STP for discounting the value and contributions of

---

<sup>54</sup> This is analogous to the approach of financial economics but applied to government finances.

<sup>55</sup> It is also worth considering the decision to change the Ogden discount rates for the insurance industry and the consequences of applying a negative real discount rate. The Ogden Discount Rate is now -0.75% real as a result of the current yield on gilts.

<sup>56</sup> See for example, the Independent Public Service Pensions Commission: Final Report, 2011.

<sup>57</sup> Levy (2012), Pensions in the National Accounts – A Fuller Picture of the UK's Funded and Unfunded Pension Obligations.

<sup>58</sup> HM Treasury (2003) The Green Book: Appraisal and Evaluation in Central Government, London.

unfunded public service pensions. First, it allows for a consistent valuation approach with respect to other governmental liabilities. Second, the provision of a public service pension is a choice about the use of public funds, and is therefore no different to the decision to allocate public funds to any other project, and it is the costs and benefits of doing so that should be analysed. Third, underpinning the use of a social rate of time preference includes notions of intergenerational fairness and can lead to a more equitable intertemporal allocation of resources (Marini and Scaramozzino, 2000).

However, the application of a STP is not without its issues. First, historically there was no adjustment made for pension liabilities extending beyond 30 years, which the Green Book explicitly allowed for, whereby the discount rate is reduced to 3% for liabilities between 30 and 70 years, which may be appropriate given increasing longevity. Second, the inclusion of catastrophe risk at 1% has been questioned as the 1% may understate liabilities. Catastrophe could therefore increase liabilities and the cost of providing public service pensions, although this is ultimately dependant on the nature of the catastrophe.<sup>59</sup>

#### *4.2.4. Discounting using the GDP growth rate*

The discounting of pension liabilities can also be done based on the GDP growth rate, and this is the preferred approach of the Treasury. The use of the GDP growth rate arguably has its roots in the work of Samuleson (1958) and Aaron (1966) that argues for unfunded pension liabilities to be discounted at the expected rate of growth in the overall wage bill of the government, which over the long-run should closely follow GDP growth. Underpinning the use of GDP growth is the source of income that will be used to meet future pension obligations i.e. tax receipts. If GDP growth is higher, the tax base will be larger, and so this has an in-built affordability measure. In circumstances where GDP growth rate is expected to be lower, then contributions can be increased today to take account of this. Moreover, the use of a GDP growth measure in this way implies some element of intergenerational fairness as this measure equalizes, albeit with uncertainty, the contributions from GDP today that are required to pay for future pensions as a proportion of GDP tomorrow.

Discounting using the GDP growth rate is however, imperfect. For example, the actual rate of GDP growth is not known and to forecast this into the far future will be subject to considerable measurement error. Moreover, the adjustment for intergenerational fairness is implied in this approach rather than being explicit.

### **4.3. Conclusion**

The current approach of the ONS for discounting pension liabilities in the National Accounts complies with the methods set out in the 2008 System of National Accounts

---

<sup>59</sup>Consultation on the Discount Rate used to set Unfunded Public Service Pension Contributions, Summary of Responses, HM Treasury 2011.

(SNA2008) and the 2010 European System of Accounts (ESA2010).<sup>60</sup> The aim of valuing and reporting pensions in National Accounts is to show the pension liabilities of a country, including funded private sector pensions, funded government pensions, unfunded public service pensions, and state pensions. These obligations make up a significant part of the liabilities that sit on a country's balance sheet and so arriving at a best estimate of this is fundamentally important.

Currently, the ONS applies European best practice for the discounting of governmental pension liabilities for National Accounting. However, the rate that is used, like all discount rates, is debatable and may not be the best estimate for understanding the magnitude of the liabilities or their associated costs at a point in time or through time.

In looking at the discounting of pension liabilities for National Accounts, the ONS has to consider:

Why is comparability across National Accounts desirable when it may prevent straightforward comparisons with other governmental liabilities that use STP?

Why is the yield on a basket of European government debt the most appropriate measure for discounting in National Accounts?

What alternatives can meet the principles for discounting that resulted in the current approach to discounting pensions in National Accounts?

Ultimately, discounting using an explicit Social Time Preference Rate is the most appropriate way to meet the principles set out in arriving at the current approach. Moreover, the application of an STP allows comparability across other parts of the government's accounts as it is consistent with the valuation of other governmental liabilities.

As it stands today, the application of a rate based on a basket of long-dated European sovereign debt that is considered equivalent to the social time preference rate i.e. it gives the right number, is not the same as the rate being a social time preference rate.

In the short-run, the rate applied under the current method has not been changed. Currently, there is an understanding from those involved in the setting of the rate that it is for all intents and purposes a social time preference rate. However, this runs the risk that the rate may become more market-based in the future. As the setting and use of discount rates is often a political process, lobbying to shift from the current approach to a "true" discount rate based on the current yield on a basket of long-dated European sovereign debt, would bring in many of the undesirable qualities of market based rates discussed above.

---

<sup>60</sup> Levy (2012), Pensions in the National Accounts – A Fuller Picture of the UK's Funded and Unfunded Pension Obligations.

In addition, while some would argue for the use of the GDP growth rate, which is applied elsewhere, this is not transparent as intergenerational fairness is implied rather than explicitly stated.

In advocating the use of a Social Time Preference Rate, this should be the way that discounting is conducted in all National Accounts. In doing so, explicit adjustments can be made to reflect expectations on economic growth, inflation, catastrophe, and intergenerational fairness both within and across countries.

Finally, it is not clear in the current approach what is being assumed about intergenerational fairness for example. If there was to be an explicit STP applied, the where disagreement exists regarding any assumption that goes into its construction, this can be resolved via a transparent process, which is not possible under the current approach.

## **References**

Aaron, H., (1966). The social insurance paradox. *The Canadian Journal of Economics and Political Science / Revue canadienne d'Economie et de Science politique*, 32: 371-374.

Lintner, L., (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital Budgets. *Review of Economic Statistics*, 47: 13-37.

Marini, G. and P. Scaramozzino, (2000). Social time preference. *Journal of Population Economics*, 13: 639-645.

Modigliani, F, and M. H. Miller, (1958). The cost of capital, corporation finance, and the theory of investment. *American Economic Review* 4: 261-297.

Novy-Marx, R. and J. Rauh, (2011). Public pension promises: how big are they and what are they worth? *Journal of Finance*, 66: 1211-1249.

Samuelson, P. A. (1958) An exact consumption-loan model of interest with or without the social contrivance of money. *Journal of Political Economy*, 66: 467-82.

Sharpe, W.F., (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19: 425-442.

Treynor, Jack L., (1961). Toward a theory of the market value of risky assets, Unpublished Manuscript.