Choosing the future: Markets, ethics and rapprochement in social discounting

Antony Millner (University of California, Santa Barbara)
and Geoffrey Heal (Columbia University)

Global Priorities Institute | September 2021

GPI Working Paper No . 13-2021

Choosing the Future:

Markets, Ethics, and Rapprochement in Social Discounting

Antony Millner*[1,3] and Geoffrey Heal[2,3]

[1]University of California, Santa Barbara
[2]Columbia University
[3]National Bureau of Economic Research

September 8, 2021

Abstract

This paper provides a critical review of the literature on choosing social discount rates (SDRs) for public cost-benefit analysis. We discuss two dominant approaches, the first based on market prices, and the second based on intertemporal ethics. While both methods have attractive features, neither is immune to criticism. The market-based approach is not entirely persuasive even if markets are perfect, and faces further headwinds once the implications of market imperfections are recognised. By contrast, the 'ethical' approach – which relates SDRs to marginal rates of substitution implicit in a single planner's intertemporal welfare function – does not rely exclusively on markets, but raises difficult questions about what that welfare function should be. There is considerable disagreement on this matter, which translates into enormous variation in the evaluation of long-run payoffs. We discuss the origins of these disagreements, and suggest that they are difficult to resolve unequivocally. This leads us to propose a third approach that recognises the immutable nature of some normative disagreements, and proposes methods for aggregating diverse theories of intertemporal social welfare. We illustrate the application of these methods to social discounting, and suggest that they may help us to move beyond long-standing debates that have bedevilled this field.

*Millner: `millner@ucsb.edu`. We are grateful to Maya Eden, Marc Fleurbaey, and Christian Gollier for helpful comments.

1 Introduction

Towards the end of the first century AD the Roman emporer Domitian ordered the construction of an aqueduct to serve the settlement of Segovia, a trading centre serendipitously located in the middle of the Iberian peninsula. The aqueduct was an architectural marvel; it transported water from the mountains over 15 km away, and required the construction of 167 arches. The Segovia aqueduct operated almost continuously for over 1800 years, supplying the town with water until the early 20th century. It remains a tourist attraction to this day.

Nestled in a nature reserve on the Atlantic ocean, about half an hour's drive from central Cape Town, sits Koeberg nuclear power station – the only such facility in Africa. Koeberg is capable of producing 1860MW of zero-carbon electricity, and currently accounts for about 5% of South Africa's electricity production. Carbon-intensive coal power accounts for almost 90% of national production. Like all nuclear power stations, Koeberg produces radioactive waste. Low grade waste is sealed in concrete drums and trucked up the coastal highway to be buried in the Namaqualand desert, a region known for vivid displays of spring wild flowers. High grade waste – spent fuel rods and the like – is too hazardous to be transported. It is stored on site in cooling ponds, across the bay from a rapidly expanding city of 4 million inhabitants. The half life of spent nuclear fuel is roughly 160,000 years.

According to recent climate science (Ricke & Caldeira, 2014), the increase in global average temperature due to a pulse of carbon dioxide emissions is likely greatest after only 10 years. But even after 100 years the temperature response is still at 80% of its peak. Some studies have estimated that even if we stopped emitting carbon dioxide immediately today, sea levels would continue to rise for a further 1000 years (Solomon et al., 2009).

To what extent are large infrastructure projects, nuclear power, and reductions in carbon dioxide emissions 'good things'? The three examples above serve to show that these questions cannot be answered without consideration for the distant future – time horizons extending to multiple decades, centuries, perhaps even millennia. How should such projects be evaluated? If we take the conventional tools of cost-benefit analysis as read, this question largely boils down to how to aggregate project consequences that occur at different points in time. The tool that enables us to convert temporal streams of payoffs into a net present value that captures the social value of a project is the social discount rate (SDR). The SDR is the rate at which the social value of a marginal consumption change declines as it moves further into the future. The values that governments pick for SDRs largely determine the degree of consideration that they give to the future when evaluating investments and policy choices. It is not an overstatement to say that choosing SDRs amounts to choosing a substantial part of the future itself.

Despite its central role in determining the future orientation of government policy, the choice of SDRs – especially for long-term project evaluation – has been a source of controversy for decades. At the heart of this controversy is a fundamental methodological schism in the community of scholars and practitioners who work on this issue. The central line of cleavage concerns the relative importance of *markets* and *ethics* as determinants of appropriate values for SDRs. Paraphrasing the arguments, those who view markets as paramount argue that prices determine the opportunity costs of public investment, and hence efficiency requires that social discount rates should coincide with observed market rates of return. By contrast, proponents of the 'ethical' approach are skeptical of the welfare significance of market prices, especially in the presence of market imperfections and at intergenerational time horizons. Instead they propose to compute SDRs directly from a normative model of society's intertemporal welfare function. This function represents intertemporal distributive objectives in much the same way that the social welfare functions used in optimal income tax theory capture intratemporal distributive objectives. While this approach is less prone to the criticism of naïve market perfectionism, it has its own difficulties. In particular, the SDRs that emerge from this approach depend strongly on normative parameters that capture e.g., societal impatience, aversion to inequality, and attitudes to uncertainty. Small changes in the prescribed values for these parameters can have an enormous effect on the evaluation of projects with long-run consequences. Since these parameters capture primitive normative judgements about social objectives, it will come as no surprise that there is a diverse set of views on how they should be chosen. Critics of the 'ethical' approach argue that it is paternalistic, or point to the inherent indeterminacy of an analytic framework that rests on normative parameters about which there can never be objective scientific consensus.

The first purpose of this article is to lay out rigorous 'steel man' presentations of the approaches to social discounting based on markets and ethics respectively. Due to the at times factional nature of the debate, each of these approaches is rarely presented in its best light, and there are many attractions and difficulties with each of them that have not been widely appreciated. The second purpose is to offer a rapprochement, a 'third way' that we believe has the potential to combine the best of both worlds. This new approach draws on a recent literature on intertemporal social choice; a set of theoretical ideas that allows for the preservation of disagreements about intertemporal societal objectives, but nevertheless seeks to achieve some kind of resolution or consensus that allows us to move beyond them. Although nascent, we believe that this line of work avoids some of the pitfalls of approaches to social discounting that rest on a naïve faith in markets, or on a single normative paradigm that is invariably open to attacks of paternalism or ethical arbitrariness. Although there are many previous fine reviews of social discounting and its role in public cost benefit analysis, much of what we will discuss has been published in the last 15 years, and is thus absent from these treatments.[1] We also believe that our presentation offers something of a fresh perspective, in that we are mostly concerned with the conceptual underpinnings of the dominant approaches to social discounting, rather than calculations of discount rates under various detailed modelling assumptions.

We begin with a discussion of what we can, and cannot, learn from market prices. We suggest that while there are some situations where an approach to social discounting based on observed prices may make sense, on the whole this requires a strong faith in the welfare significance of prices, and is at best incomplete. Indeed, even if markets are perfect, there are a number of reasons to be skeptical about whether prices are telling us what we need to know for the purposes of setting SDRs. We discuss three of these: limitations of revealed preference in dynamic choice contexts, intergenerational issues, and the presence of heterogeneous or erroneous beliefs. We then discuss deficiencies in the implementation of the market-based approach, focussing on the way governments currently handle the maturity- and risk-dependence of discount rates. We close our discussion of this approach with an elaboration of the consequences of market imperfections. We discuss the implications of market incompleteness, and also point out that in the presence of market failures competitive prices do not contain the information needed to determine social discount rates at the social optimum, i.e., after inefficiencies have been corrected by non-marginal policy instruments.

We then provide a detailed treatment of the dominant normative paradigm for setting discount rates, an approach based on computing intertemporal marginal rates of substitution for an idealized planner with discounted utilitarian time preferences. We describe the mechanics of this approach, the strong constraints on social preferences implicit in the choice of the discounted utilitarian form for the planner's objective function, and the

[1] Classic texts on public cost benefit analysis and other reviews of social discounting include Dasgupta et al. (1972); Little & Mirrlees (1974); Lind (1982); Dreze & Stern (1987); Portney & Weyant (1999); Groom et al. (2005); Heal (2005); Dasgupta (2008); Gollier (2012); Arrow et al. (2013, 2014); Gollier & Hammitt (2014); Cropper et al. (2014); Groom & Hepburn (2017); Greaves (2017); van der Ploeg (2020).

debates surrounding how the normative parameters of that welfare function should be chosen.

Our final substantive section presents recent developments in intertemporal social choice, and their potential applications to resolving disagreements about normative parameters that strongly influence SDRs. We show that although there are several promising approaches in the literature, many of them have an important recommendation in common: long-run risk-free discount rates should be very low.

2 What can we learn from the markets?

We begin with a brief treatment of the problem of public investment appraisal in complete, competitive markets, a useful point of departure for three reasons. First, much of the formalism developed in the case of perfect markets can be applied to the more realistic case of imperfect markets with appropriate modifications. Second, a careful treatment of perfect markets will make it clear what conditions must be satisfied in order for this framework to deliver us appropriate values for social discount rates. Finally, even if we take the assumptions underlying the perfect markets model as read, there are a number of important deficiencies in the way it is often implemented in practical applications. Our treatment of this case thus highlights inconsistencies in the practice of social discounting by governments that adhere to a market-based approach.

2.1 Project appraisal in a perfect economy

Consider an exchange economy with I consumers, indexed by $i = 1, 2, \ldots, I$, who have preferences over N goods, indexed by $n = 1, 2, \ldots, N$. These goods can be consumed in any of $T + 1$ time periods indexed by $t = 0, 1, \ldots, T$, and S states of the world indexed by $s = 1, 2, \ldots, S$.[2] We assume that there is no uncertainty about consumption in the initial period. Consumer i's consumption bundle may thus be represented by the $N(TS + 1)$ dimensional bundle $\mathbf{c}^i = (c^i_{n,0}, c^i_{n,t,s})_{n=1,\ldots,N,t=1,\ldots,T,s=1\ldots,S}$. We will assume that consumer

[2] For simplicity's sake our main discussion favours a static uncertainty interpretation of the state space, i.e., the state s represents a single event about which consumers are uncertain at $t = 0$. Dealing with dynamic uncertainty simply requires a redefinition of the state space. In a dynamic framework we can think of a state of the world as a sequence of future events that occur at times $1, \ldots, T$, and we are uncertain about which sequence will be realised right up until the end of the last period. Instead of indexing consumption at time t by the state of the world s, we can index it by the history of events up to time t, so that consumption in states of the world that share a common history up to time t is the same. The relevant state space at time t is then just the set of all possible histories of length t.

i's preferences over consumption bundles are represented by a differentiable and strictly quasi-concave utility function $U^i(\mathbf{c}^i)$ that satisfies Inada-like conditions, and her initial endowment of claims to time and state contingent consumption is $\boldsymbol{\omega}^i$. The function $U^i(\mathbf{c}^i)$ captures consumer i's tastes and beliefs.

Let $p_{n,t,s}$ be the price of good n in time period t and state s, and let $\mathbf{p}$ be the $N(TS+1)$ dimensional vector of these prices.[3] At the beginning of period 0 consumers trade on a complete set of state-contingent futures markets, so that they can buy and sell any good at any date or any state of the world at the prices $p_{n,t,s}$. We also make enough additional assumptions (on preferences and endowments) to ensure that a competitive equilibrium of the economy exists, and may be characterised by first order conditions.

We assume that consumers maximize their utility given their budget constraint:

$$\max_{\mathbf{c}^i} U^i(\mathbf{c}^i) \text{ s.t. } \mathbf{p}\cdot\mathbf{c}^i = \mathbf{p}\cdot\boldsymbol{\omega}^i.$$

The first order conditions for this optimization problem imply that

$$\frac{\frac{\partial U^i(\mathbf{c}^{*i})}{\partial c_{n,t,s}}}{\frac{\partial U^i(\mathbf{c}^{*i})}{\partial c_{n',t',s'}}} = \frac{p_{n,t,s}}{p_{n',t',s'}} \tag{1}$$

where $\mathbf{c}^{*i}$ is the equilibrium consumption bundle of consumer i, determined by (1) and the market clearing conditions.[4] In a complete, competitive market, price ratios capture all consumers' marginal rates of substitution between time and state contingent goods in equilibrium.

To see the implications of this result, let us simplify to the case where there is only one consumption good in the economy (i.e., $N = 1$). We can thus suppress the index n, and we can also choose (certain) consumption at $t = 0$ as the numeraire in this economy. Then from (1) we have

$$\frac{\frac{\partial U^i(\mathbf{c}^{*i})}{\partial c_{t,s}}}{\frac{\partial U^i(\mathbf{c}^{*i})}{\partial c_0}} = \frac{p_{t,s}}{1}. \tag{2}$$

It will be helpful for what follows to provide an interpretation of this result. Suppose that consumer i with initial consumption vector $\mathbf{c}$ has the opportunity to sacrifice a small amount π_0 of current consumption in exchange for a bundle of marginal payoffs $(\pi_{t,s})_{t=1,\ldots,T,s=1,\ldots,S}$. Should she do so? Denoting the vector of net payoffs by $\boldsymbol{\pi}$, this trade

[3] We again adopt the convention that when $t = 0$ the s index is redundant.

[4] These are given by $\sum_i c^{*i}_{n,t,s} = \sum_i \omega^i_{n,t,s}$.

is advantageous iff

$$U^i(\mathbf{c}+\boldsymbol{\pi}) - U^i(\mathbf{c}) > 0 \iff -\pi_0 + \sum_{t=1}^{T}\sum_{s=1}^{S} \pi_{t,s}\left(\frac{\frac{\partial U^i(\mathbf{c})}{\partial c_{t,s}}}{\frac{\partial U^i(\mathbf{c})}{\partial c_0}}\right) > 0, \tag{3}$$

where we have Taylor expanded to first order, neglecting higher order terms due to the marginality of payoffs. Thus we see that the marginal rate of substitution $\left(\frac{\frac{\partial U^i(\mathbf{c})}{\partial c_{t,s}}}{\frac{\partial U^i(\mathbf{c})}{\partial c_0}}\right)$ tells us the *present value* of the payoff $\pi_{t,s}$ for consumer i at the consumption vector $\mathbf{c}$. This fact motivates us to define a set of *consumption discount rates* $\rho^i_{t,s}(\mathbf{c})$ for consumer i as follows:

$$(1+\rho^i_{t,s}(\mathbf{c}))^{-t} := \left(\frac{\frac{\partial U^i(\mathbf{c})}{\partial c_{t,s}}}{\frac{\partial U^i(\mathbf{c})}{\partial c_0}}\right) \tag{4}$$

In what follows we will also often talk about the *risk-free* consumption discount rate $\rho^i_t(\mathbf{c})$, i.e., consumer i's discount rate on a sure transfer of current consumption to time t (i.e., $\pi_{t,s} = \pi_t$ for all states s):

$$(1+\rho^i_t(\mathbf{c}))^{-t} := \sum_{s=1}^{S}\left(\frac{\frac{\partial U^i(\mathbf{c})}{\partial c_{t,s}}}{\frac{\partial U^i(\mathbf{c})}{\partial c_0}}\right). \tag{5}$$

Risk-free discount rates are related to state-contingent discount rates through:

$$(1+\rho^i_t(\mathbf{c}))^{-t} = \sum_{s=1}^{S}(1+\rho^i_{t,s}(\mathbf{c}))^{-t}.$$

Using the definition (4), consumer i's cost-benefit rule becomes:

$$-\pi_0 + \sum_{t=1}^{T}\sum_{s=1}^{S} \pi_{t,s}(1+\rho^i_{t,s}(\mathbf{c}))^{-t} > 0. \tag{6}$$

The finding that all consumers' marginal rates of substitution at their equilibrium allocations are equal to competitive state prices can thus be restated in terms of their consumption discount rates:

$$\forall i,\ \rho^i_{t,s}(\mathbf{c}^{*i}) = \left(\frac{1}{p_{t,s}}\right)^{1/t} - 1 := \rho^*_{t,s}. \tag{7}$$

Consumption discount rates capture consumers' private relative valuations of marginal consumption changes that occur at different times and in different states of the world. Nevertheless, in competitive markets all individuals' marginal valuations are equal. This identification between statistics of consumers' preferences (consumption discount rates) and observable features of the world (competitive state prices) is at the heart of the approach to social discounting based on markets. Since all individuals' consumption discount rates are equal at the competitive equilibrium we can drop the i index, and simply talk about *the* consumption discount rate, which we denote by $\rho^*_{t,s}$. Similarly, we may talk about *the* risk-free consumption discount rate ρ^*_t. The asterisk on these quantities indicates that they are equal to individuals' consumption discount rates evaluated at their equilibrium allocations.

The results developed above in our model of an exchange economy apply equally well in a productive economy. Consumers face similar optimization problems in a model with production,[5] and state prices will again reflect all consumers' marginal rates of substitution. Modelling production explicitly does however provide an additional relationship between competitive prices and economic fundamentals that has proven influential in the practice of social discounting: in a competitive productive economy state prices also reflect aggregate marginal rates of transformation, i.e., private returns on investment. In our model with a single consumption good this relationship can be stated as follows. If $\mathbf{Y}$ is a vector of aggregate net outputs in the economy, and $T(\mathbf{Y})$ the aggregate transformation function,[6] profit maximization implies that

$$p_{t,s} = \left. \frac{\frac{\partial T}{\partial C_{t,s}}}{\frac{\partial T}{\partial C_0}} \right|_{\mathbf{Y}^*}, \tag{8}$$

where $C_{t,s}$ is aggregate consumption is state t, s. As the marginal rate of transformation captures the technological rate of exchange between initial investments (i.e., reductions in C_0) and future consumption, we can think of it as defining a private rate of return on

[5] Consumers' budget constraints must be modified to account for rents from inputs and shares in firms' profits, but are otherwise identical.

[6] The aggregate production set is given by $T(\mathbf{Y}) < 0$. If there are no externalities and individual firms have convex production sets a convex aggregate production set exists, and we can also work with an aggregate transformation function for the economy as a whole, rather than modelling each firm's production decisions individually (see e.g. Varian, 1992, p. 339).

investment $\hat{r}_{t,s}$ in state t, s:

$$(1+\hat{r}_{t,s})^t = \left.\frac{\frac{\partial T}{\partial C_0}}{\frac{\partial T}{\partial C_{t,s}}}\right|_{\mathbf{Y}^*} = \frac{1}{p_{t,s}}.$$

Of course, in a competitive equilibrium private rates of return are equal to consumption discount rates:

$$\hat{r}_{t,s} = \rho^*_{t,s} = \left(\frac{1}{p_{t,s}}\right)^{\frac{1}{t}} - 1.$$

Suppose now that the economy is at a competitive equilibrium, and a government wishes to evaluate a marginal public project that yields $\pi^i_{t,s}$ units of net consumption to individual i at time t in state s. Since initial consumption c_0 is the numeraire, consumer i's marginal utility of income is given by $\lambda^i = \frac{\partial U^i(\mathbf{c}^{*i})}{\partial c_0} > 0$. Since the project is marginal, consumer i's compensating and equivalent variations for the project are equal,[7] and given by

$$\Delta^i = \frac{dU^i}{\lambda^i} = \sum_{t=1}^{T}\sum_{s=1}^{S}\frac{\frac{\partial U^i(\mathbf{c}^{*i})}{\partial c_{t,s}}}{\frac{\partial U^i(\mathbf{c}^{*i})}{\partial c_0}}\pi^i_{t,s} = \sum_{t=1}^{T}\sum_{s=1}^{S}(1+\rho^*_{t,s})^{-t}\pi^i_{t,s} = \sum_{t=1}^{T}\sum_{s=1}^{S}p_{t,s}\pi^i_{t,s}. \tag{9}$$

The project constitutes a Pareto improvement relative to the initial market equilibrium iff

$$\forall i,\ \Delta^i = \sum_{t=1}^{T}\sum_{s=1}^{S}p_{t,s}\pi^i_{t,s} \geq 0,$$

with the inequality being strict for at least one i. This is not a very practicable criterion for project evaluation, as almost all projects will fail to be Pareto improvements. There are two standard ways to proceed beyond this criterion. The first is to introduce a differentiable social welfare function

$$W = W(U^1(\mathbf{c}^1), U^2(\mathbf{c}^2), \ldots, U^I(\mathbf{c}^I))$$

that captures society's normative preferences over distributions of utility across individuals. This construction invariably requires us to make interpersonal utility comparisons across individuals. Let

$$w_i = \left.\frac{\partial W}{\partial U^i}\right|_{U(\mathbf{c}^{*1}),\ldots,U(\mathbf{c}^{*I})}$$

be the *social marginal welfare weight* on individual i at the equilibrium allocation. The

[7]Marginality here should be taken to mean that the project does not affect prices. If prices and incomes change as a result of the project, there is no guarantee that a positive sum of compensating variations across individuals implies the existence of a potential Pareto improvement (Blackorby & Donaldson, 1990).

project improves social welfare iff

$$dW = \sum_{i=1}^{I} w_i dU^i = \sum_{i=1}^{I} w_i \lambda^i \Delta^i = \sum_{t=1}^{T}\sum_{s=1}^{S} p_{t,s} \left(\sum_{i=1}^{I} w_i \lambda^i \pi^i_{t,s} \right) > 0. \tag{10}$$

As this expression shows, project payoffs across heterogeneous individuals are weighted by their contribution to social welfare, computed as the product of the marginal utility of income λ_i and the social marginal welfare weight w_i. λ^i captures the impact of a small change in wealth on consumer i's utility, and w_i captures the impact of a small change in consumer i's utility on social welfare. These quantities capture important information about wealth effects, and society's attitudes to distributive justice, respectively.

The second approach to aggregating project consequences across individuals is to appeal to the Kaldor-Hicks potential compensation criterion. This criterion aims to separate efficiency issues from distributional concerns, and obviates the need for interpersonal comparisons of utility. The Kaldor-Hicks criterion simply requires the winners to be able to compensate the losers, in theory, i.e.,

$$\sum_{i=1}^{I} \Delta^i = \sum_{t=1}^{T}\sum_{s=1}^{S} p_{t,s} \left(\sum_{i=1}^{I} \pi^i_{t,s} \right) > 0. \tag{11}$$

If this criterion is satisfied then there exist hypothetical lump-sum transfers that could make the project a Pareto improvement (and conversely, no such transfers exist otherwise). These transfers need not actually occur, and are almost always infeasible in practice.[8]

For the remainder of the paper we largely ignore the issues that are involved in aggregating project consequences across individuals, and instead focus on intertemporal issues. Regardless of whether we follow an approach based on a social welfare function or the Kaldor-Hicks criterion, the intertemporal issues are formally similar, i.e. aggregation of project payoffs across time is performed using state-contingent discount rates. Nevertheless, it is important to emphasise that the assumption of perfect markets does not remove the issues associated with aggregation of payoffs across individuals from consideration; they lurk in the background, as they do in any welfare analysis.

Finally, although our presentation of these results has focussed on the simple case of a single consumption good, notice that it is straightforward to generalise to an arbitrary

[8] A recent literature defines new compensation tests that rely on feasible transfers given the limited information and instruments of a redistributive planner, and account for the fact that redistribution is itself distortionary (Antras et al., 2017; Tsyvinski & Werquin, 2018; Hendren, 2019).

number of goods. Equilibrium Arrow-Debreu prices for each good can be converted into good-specific time and state contingent discount rates as in the single good case. Public projects that affect individuals' consumption of multiple goods are evaluated using an aggregate cost-benefit rule, in which the project's marginal effect on the consumption of each time and state contingent good is discounted using the appropriate good-specific discount rate, and the project is implemented if and only if the (welfare-weighted) sum of these discounted marginal effects across goods, time, and states, is positive.

***

Taking the perfect markets scenario as read, the problem of evaluating public projects seems to be straightforward – simply calculate the prices of Arrow-Debreu securities from the market prices of assets,[9] and use them to discount project payoffs. This recipe works regardless of how projects are financed. We now draw an expanding circle of criticism around this approach and the way it has been implemented in practice. First we discuss deficiencies of implementation, taking the assumptions of our analysis thus far as read. Second we discuss reasons to be skeptical of the welfare significance of market prices, even if markets are perfect. Finally, we interrogate the perfect markets assumption itself.

2.2 Which market rates?

In the case of perfect markets we can use the market prices of assets to back out the prices of Arrow-Debreu securities, and hence the appropriate discount rates for public cost-benefit analysis. There are two immediate practical questions that arise when implementing this approach: First, what is the maturity dependence of market interest rates? Second, how should the government account for risk when evaluating public projects?

With respect to maturity dependence, the key issue is that the prices markets place on payoffs that occur at different future times are almost never consistent with risk-free discount rates that are independent of maturity. Despite this, several governments that use a market-based approach to set social discount rates, most notably that of the United States (OMB, 2003), recommend that public projects be evaluated using constant discount rates.

[9] If markets are complete, satisfy a no-arbitrage condition, and the law of one price holds, there is a unique positive matrix of state prices that rationalises observed asset prices with their state-contingent payoffs.

The formula that relates risk-free interest rates to the prices of Arrow-Debreu securities is:

$$r_t^f = \left(\frac{1}{\sum_s p_{t,s}}\right)^{\frac{1}{t}} - 1.$$

For this to be independent of maturity we require $\frac{1}{t}\ln\left(\sum_s p_{t,s}\right)$ to be independent of t. This is clearly a very strong constraint on observed prices, and is seldom, if ever, satisfied in practice. Observed yield curves for inflation-indexed government debt (the closest thing we have to a risk-free asset in countries with low default risk) are almost never flat. There is thus an internal inconsistency in the constant rate approach – governments defer to market interest rates when setting social discount rates, but the discount rates they end up prescribing do not reflect the prices that actually prevail on the market.

The second question – regarding project risks – has been the subject of heated discussion amongst economists since the 1970s. A seminal result due to Arrow & Lind (1970) suggested that since the payoffs from risky government projects are distributed across a large number of individuals, the aggregate cost to society of bearing this risk is zero.[10] This seems to suggest that public cost-benefit analysis should not concern itself with project specific risks, i.e., all payoffs should be discounted using risk-free discount rates. However, the Arrow-Lind result depends critically on the assumption that project payoffs are uncorrelated with aggregate macroeconomic risks. If, as is very likely, such a correlation does exist, project risks must be accounted for. This can be achieved either by using time- and state-dependent discount rates, or by working with expected project payoffs, but adjusting discount rates to reflect the fact that the project either amplifies (positive correlation) or attenuates (negative correlation) the risk society faces. In practice the latter approach is often implemented by using a reduced form asset pricing model – the Consumption-based Capital Asset Pricing Model (CCAPM) being by far the most popular choice.[11]

To understand this approach, assume for the moment that a representative agent for the economy exists – we will return to this strong assumption below, but for the moment ask the reader to suspend disbelief. The CCAPM models project payoffs and the representative agent's consumption at maturity t using the random variables B_t and C_t

[10] This result is a consequence of the fact that risk aversion is a second order phenomenon. The Arrow-Pratt approximation tells us that the cost to an individual of bearing a $1/n$th share of a zero mean risk with standard deviation σ is proportional to σ^2/n^2. Hence the aggregate cost of bearing this risk is proportional to $n \times \frac{\sigma^2}{n^2} \to 0$ as $n \to \infty$.

[11] The CCAPM is widely used in the academic literature on social discounting, despite its well known empirical shortcomings (Fama & French, 2004).

respectively. Assume moreoever that $\ln B_t$ and $\ln C_t$ are jointly normally distributed. If the representative agent has a constant elasticity of marginal utility η, it can be shown that the appropriate certainty equivalent discount rate for expected payoffs $\mathsf{E}B_t$ at maturity t is:

$$r_t = r_t^f + \beta_t \eta \sigma_t^2 \tag{12}$$

where r_t^f is the risk-free rate at maturity t, σ_t^2 is the variance of consumption growth at maturity t, and

$$\beta_t = \frac{1}{\sigma_t^2}\mathrm{Cov}(\ln(B_t/B_{t-1}), \ln(C_t/C_{t-1})).$$

The coefficient β_t is project and maturity specific, and captures the correlation between growth in the project's payoffs and aggregate consumption growth. If the project is growth enhancing in low growth states of the world it provides desirable insurance against aggregate risk, and β_t is negative. Conversely, β_t is positive if the project's returns are highest in high growth states of the world. To our knowledge, only one country has trialed the use of project-specific social discount rates based on the CCAPM – France (see Gollier, 2011; Quinet, 2013).[12]

The importance of accounting for both maturity dependence and project-specific risk adjustments in social discounting is illustrated in a recent set of empirical papers (Giglio et al., 2015a,b). The authors exploit an idiosyncratic feature of the real estate markets in the United Kingdom and Singapore – property can be bought as either 'freehold' (an indefinite ownership right) or 'leasehold' (a lease agreement, at the end of which ownership reverts to the freeholder) in these markets. The price differential between a freehold property and a 100 year lease on an otherwise identical property contains information about how the market values cash flows at 100 years and beyond. Using a database on the universe of property transactions in these markets, the authors are able to infer bounds on implied discount rates for this asset class. Remarkably, because leasehold contracts are often very long (extending to 999 years in some cases), these data can be used to constrain implicit discount rates at time horizons far exceeding those for other asset classes. While Giglio et al. (2015a) focusses on obtaining an upper bound on the long-run risk-free rate (their estimate is 2.6%/yr), Giglio et al. (2015b) investigate the term structure of discount

[12]The US government uses constant discount rates of 3%/yr and 7%yr for cost-benefit analysis. 3%/yr is supposed to reflect the average rate of return on risk-free government debt, while 7%/yr is supposed to represent the average return to risky capital (i.e., equities). Although this approach pays lip service to the risk adjustments we discuss in this section, it is a long way from the project specific risk premia that basic asset pricing theory requires. Gollier (2019) presents a model that uses French estimates of betas in different sectors to suggest that the welfare consequence of neglecting project-specific betas could be large.

rates, and how the market adjusts for the riskiness of real estate. They observe that since the average rate of return for real estate is 6%/yr, but the long-run (i.e., >100 year) rate is only 2.6%/yr, the term structure of discount rates for this asset class cannot be flat, and must be downward sloping for at least some maturities. This suggests that using a constant discount rate calibrated to the average rate of return of 6%/yr could substantially undervalue investments that pay off most in the long run, e.g. climate change mitigation measures, which should be discounted at much lower rates. Since climate abatement investments do not have the same risk profile as real estate, Giglio et al. (2015b) also construct a model of climate change as a rare disaster. In their baseline estimates discount rates on climate mitigation investments are low (i.e., always below the risk-free rate) but, unlike real estate, have an increasing term structure. These results highlight the importance of having a procedure for setting discount rates that is consistent with basic principles from asset pricing. Even if we take the perfect markets assumption as read, and attempt to read off social discount rates from market prices, the implied discount rates should reflect the risk profile of the investment in question, as well as the market's valuation of payoffs that occur at different points in time.

2.3 Critiques of the welfare significance of prices in perfect markets

The previous discussion highlights areas where the practice of social discounting diverges from what we know about the implications of market prices for discount rates from economic theory, assuming market perfection. These issues are operationally important, but ultimately do not require a major shift in our thinking about the relevance of market prices for public project evaluation. They are technical complaints about how prices are currently used by those governments that follow a market-based approach to setting social discount rates, but the remedies are conceptually straightforward.

In this subsection we take a more critical look at this procedure. We suggest that, even in the optimistic case of perfect markets, there are still reasons to be skeptical about the welfare significance of market prices for public project evaluation. We focus on three arguments: the limitations of revealed preference as a welfare indicator in dynamic contexts, intergenerational issues, and the implications of belief heterogeneity for welfare measurement.

2.3.1 Limitations of revealed preference in dynamic contexts

The central argument for using market prices to set social discount rates in a perfect market setting is that equilibrium price ratios capture individuals' marginal rates of substitution across times and states of the world, as demonstrated in (1). An implicit assumption in this approach is that these marginal rates of substitution capture all the effects we care about for the purposes of social decision-making. There is however a profound critique of this reasoning due to Caplin & Leahy (2004) (CL). CL make the deep observation that in the context of dynamic choice, observed preferences (as encoded in market prices) are only partial indicators of welfare: 'preferences revealed in the market do not adequately represent tastes'. At its core, their insight is that preferences over future consumption streams do not fully capture information about agents' attitudes to past consumption.

CL's argument is easiest to illustrate in a two period model, with time indexed by $t \in \{1, 2\}$. Following Strotz (1955), who highlighted 'the possibility that a person is not indifferent to his consumption history but enjoys his memory of it', CL consider an agent who may derive utility from past as well as future consumption. Let U^t denote the agent's utility from the consumption stream (c_1, c_2) at time t, and assume that:

$$\begin{aligned} U^1(c_1, c_2) &= u_1(c_1) + \lambda(1) u_2(c_2) \\ U^2(c_1, c_2) &= \lambda(-1) u_1(c_1) + u_2(c_2). \end{aligned} \tag{13}$$

The preferences in (13) are trivially time consistent,[13] but depend on the history of consumption. From the perspective of describing choice over future consumption streams, the term in red is irrelevant, and can be neglected. This term represents the agent's preferences over past consumption in period 2, but since the past is fixed, it has no influence on choices in period 2. However, this does not imply that this term is irrelevant for welfare computations. Indeed, the presence of this term destroys the correspondence between choice and welfare that exists in static applications of revealed preference. To see this consider the Fisher diagram in Figure 1. The agent's equilibrium consumption allocation occurs where the indifference curve associated with $U^1(c_1, c_2)$, denoted IC_1, is tangent to the production possibilities frontier. However, this equilibrium is not welfare maximising for the agent at $t = 2$, who would like to implement the allocation at point Q. The $t = 2$ agent's choices are however constrained by the fact that c_1 has already been chosen – it is in the past, and therefore immutable. The constrained optimal choice of c_2 for the agent at $t = 2$

[13] See Section 3.3 for a detailed definition of time consistency.

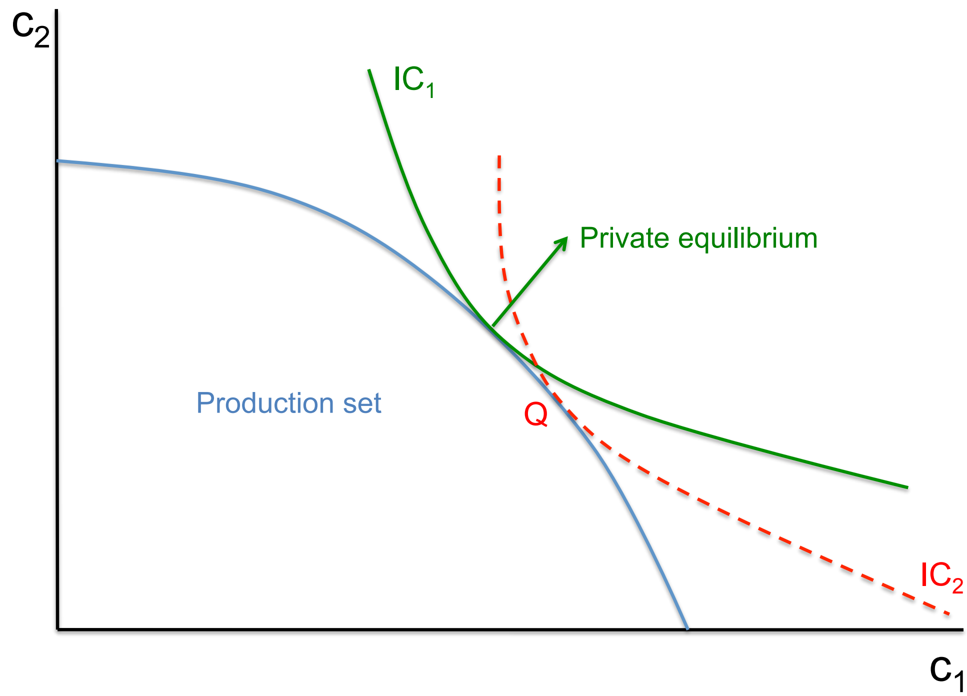

Figure 1: Illustration of the insights in Caplin & Leahy (2004). The solid blue curve denotes the production possibilities frontier. IC_1 denotes the agent's maximal indifference curve at $t = 1$, while IC_2 denotes her maximal indifference curve at $t = 2$. Consumption allocations on IC_2 are not attainable, as period 1 consumption is fixed in period 2.

is the same as the optimal choice for the agent at $t = 1$, but the resulting consumption stream is sub-optimal from the perspective of $t = 2$. Note that time inconsistency plays no role whatsoever in this argument – the agent is perfectly time consistent. The conflict between welfare measures at different times is an inescapable consequence of concern for past consumption, and the fact that the arrow of time flows in only one direction.

What can be concluded from this observation? CL explain that 'while neglecting period 2 preferences makes sense from the viewpoint of private decision making, social policy is quite a different matter. The fact that only period 1 and not period 2 preferences over consumption streams can be put into effect should not be confused with a policy doctrine. The asymmetry in intertemporal control rights should no more determine policy in this case than it would were we dealing with two distinct individuals.' How then should the agent's welfare be defined in this case? One natural procedure is to explore the set of Pareto optimal allocations across the agent's preferences at all times. In the example above this corresponds to finding feasible allocations that maximize $V = w_1 U^1(c_1, c_2) +$

$w_2 U^2(c_1, c_2)$, where $w_1, w_2 \geq 0$. The observed private equilibrium is clearly Pareto efficient, but corresponds to the case $w_2 = 0$. Using a more general model, CL find plausible conditions under which observed intertemporal equilibria are the most impatient Pareto optima. Their argument thus provides an entirely non-paternalistic reason for a social planner to adopt an intertemporal welfare function that does not replicate observed market prices, even if markets are perfect. Moreover, such a planner will often be more patient than the market.

CL's argument is powerful because it applies even to the bread and butter variety consumer preferences that economists find congenial in many applications. For example, (two period) discounted utilitarian preferences are indistinguishable from the preferences in (13) from a revealed preference perspective, and thus the concerns they raise are relevant whenever this model is used. While CL's critique relies on a backward-looking aspect of preferences, this is not required for the general set of concerns they identify to be relevant. Indeed, their argument parallels issues that arise in thinking about the appropriate welfare measure for exclusively forward-looking agents who suffer from time inconsistency problems. There too we have an asymmetry in 'intertemporal control rights'; it is not at all clear that welfare should be identified with the preferences of the 'current self', when those preferences conflict with those of 'future selves'. We will have a lot more to say about time (in)consistency issues below, but simply note here that there is abundant evidence that consumers exhibit this kind of behaviour.[14] This is another, independent, reason to be skeptical of a straightforward identification between observed prices and the welfare measures that are relevant for setting social discount rates.

2.3.2 Intergenerational issues

An arguably even more serious concern with the identification of social discount rates with consumer prices is that these prices will only reflect the preferences of currently living consumers. In our discussion of the CL critique we saw that even if agents are time consistent, current preferences may not reflect the preferences of past or future selves, and thus social welfare may be underdetermined by preferences. This 'missing preferences' critique becomes even more salient in the context of intergenerational decision-making, since those who are most affected by payoffs that occur in say 100 years are not yet born.

[14] See e.g. Benartzi & Thaler (2007); Skinner (2007) for a discussion of evidence of time inconsistency in the context of retirement savings. Cohen et al. (2020) provide an up to date review of recent empirical approaches to estimating time preferences.

Clearly, these considerations are of primary importance when choosing long-run social discount rates.

To illustrate this point it is helpful to consider a simple model of intergenerational altruism, versions of which have been studied by Bernheim (1989); Farhi & Werning (2007). Suppose that a generation of identical individuals is born at time t. Each individual lives for a single period, and each has a single child with identical preferences. All individuals at time t care about their own consumption c_t and the wellbeing of their children V_t, in such a way that the preferences of generation t can be represented by:

$$V_t = U(c_t) + \beta V_{t+1} = \sum_{\tau=0}^{\infty} \beta^{\tau} U(c_{t+\tau}), \tag{14}$$

for some $\beta \in (0, 1)$. The agents in this model are 'purely altruistic' with respect to their children: they care about what their children care about (i.e., V_t), and not just about their children's utility (i.e., $U(c_{t+1})$). It is clear that the preferences of successive generations are time consistent in this case,[15] suggesting that any consumer prices that capture current individuals' marginal rates of substitution between consumption in different time periods will also capture future generations' relative valuations.

So much for each generations' preferences, but how should a social planner evaluate welfare? One possibility is that the planner at time t should only account for the preferences of the current generation, since that generation's preferences already reflect direct concern for the next generation, and indirect concern for all generations after that via the preferences of their children. But that is a rather extreme possibility – it implies that only the current generation's preferences have welfare significance. A more general approach would be for the planner to give the preferences of each generation direct weight in her welfare measure. As in the CL model, the Pareto frontier in such a framework can be captured by maximising welfare functions of the form

$$W_t = \sum_{\tau=0}^{\infty} \alpha_{\tau} V_{t+\tau} = \sum_{\tau=0}^{\infty} \gamma_{\tau} U(c_{t+\tau}), \tag{15}$$

for some weights $\alpha_{\tau} \geq 0$, where

$$\gamma_{\tau} = \left(\sum_{k=0}^{\tau} \alpha_k \beta^{\tau-k} \right). \tag{16}$$

[15] See Section 3.3 for a detailed definition of time consistency.

If $\alpha_0 = 1, \alpha_\tau = 0$ for all $\tau \geq 1$, then only currently living consumers' preferences are relevant for welfare analysis. But the expression (15) makes it plain that this choice corresponds to a highly inequitable point on the Pareto frontier, a kind of 'tyranny of the present'. It seems questionable to identify welfare with the preferences of a single group of agents who happen to have had the good fortune to be born in the present. At the least, we should give each generation's preferences some non-negative weight in social welfare. When we do this, social preferences and current consumers' preferences will differ. Indeed we can say more: social preferences will be more patient than consumers' preferences, since

$$\forall \tau, \quad \frac{\gamma_{\tau+1}}{\gamma_\tau} = \beta + \frac{\alpha_{\tau+1}}{\gamma_\tau} > \beta$$

whenever $\alpha_\tau > 0$ for all τ.

Notice that this argument proceeded in an environment that was very advantageous to the welfare significance of consumer preferences. As the preferences of successive generations are time consistent in this model, it seems natural *a priori* to identify discount rates with prices. Yet, as in our discussion of the CL critique, a deeper look at the implications of this intuitive procedure calls it into question. The CL critique and the intergenerational issues discussed here are temporal mirror images of one another – CL operates backwards in time from the perspective of individual consumers, while this model operates forwards in time from the perspective of altruistic generations. Both apply even if markets are perfect. Thus, even in these optimistic cases, current consumers' preferences are arguably a poor guide to social welfare computations in general. We will return to a discussion of alternative notions of intergenerational welfare in Section 4.

2.3.3 Belief heterogeneity

A third issue concerns the welfare significance of market prices when agents have heterogeneous, and possibly erroneous, beliefs. Traditionally economists tend to distance themselves from strong normative pronouncements on the legitimacy of individuals' preferences. Where possible we defer to the Pareto criterion, which is often seen as a value-free rationality criterion for social comparisons. The Pareto criterion is fundamentally non-paternalistic – it is an incomplete ordering on sets of individual preferences, and all individuals are free to determine their preferences in any manner they please. In a world with uncertainty, individuals' preferences depend on their tastes and their beliefs. The Pareto criterion allows individuals complete sovereignty over both these components. Even if someone believes that the world is flat, this belief is a perfectly legitimate input to preferences, and hence also relevant for social comparisons.

But the world is not flat. There is a material difference between a planner who imposes her tastes on individuals, and a planner who respects individuals' tastes, but makes social comparisons using the best available information about the world, even if that information conflicts with some citizens' beliefs. In the former case there is a clear violation of individual sovereignty, but can we say the same of the latter? Perhaps the planner is simply better informed than some citizens, or less prone to the many biases in judgment under uncertainty that have been documented by psychologists?[16] If the notion of objective truth has any currency, it seems perverse for welfare criteria to account for beliefs that are known to be incorrect or irrational, or to admit plural beliefs that are in direct conflict with one another.[17]

A recent literature has begun to grapple with the difficulties of making welfare comparisons when beliefs are heterogeneous, or erroneous. Gilboa et al. (2014) define a refinement of Pareto dominance called 'No-betting-Pareto'. This criterion recognises that trade between individuals with heterogeneous beliefs can still be desirable (their tastes may differ too, after all), but excludes trades that are Pareto improving solely because of belief heterogeneity when making social comparisons. Brunnermeier et al. (2014) introduce the alternative notion of 'belief-neutral' Pareto efficiency. The motivating thought behind their definition is that welfare should be computed with respect to the 'true' objective probabilities of events, but it is often difficult for planners to identify those probabilities. The authors thus look for a definition of efficiency that is robust for all 'reasonable' beliefs, taken to be the set of convex combinations of all agents' beliefs. These two definitions occupy somewhat different philosophical territory, and lead to different characterisations of efficiency. The important point for our purposes is that both criteria do not take agents' beliefs as given when making social comparisons. Each of them is more demanding than the standard notion of efficiency, and in each case social states are compared from the perspective of a coherent, non-contradictory, set of beliefs. This literature has natural implications for public cost benefit analysis, even though we generally cannot rely on the

[16] Our argument here is not in conflict with the well known arguments for the informational virtues of markets (Hayek, 1945). It is possible to recognise the market's ability to aggregate dispersed local information about tastes, while simultaneously acknowledging that market participants' beliefs are not always rational.

[17] Mongin (2016) observes that there are many ways for a group of people to reach agreement on the ranking of social states, despite disagreements about facts and values. Even if everyone agrees that one policy is superior to another, their reasons for believing this may differ. Ranking social states with the Pareto principle may thus give rise to 'spurious unanimity'.

Pareto principle or its modifications in this domain. The essential message is that since market prices reflect the distribution of consumers' tastes *and* beliefs, there is no reason to expect them to encode the information that is needed for welfare analysis if there is substantial belief heterogeneity in the population.[18]

As an empirical matter, the prevalence of heterogenous, or possibly erroneous, beliefs can hardly be contested (see Brunnermeier et al., 2014, for a summary of the literature). In the case of climate change, for example, Severen et al. (2018) have used hedonic methods to demonstrate that the value of agricultural land depends on forecasts of future climate change, and that this dependence is stronger in counties where larger proportions of the population believe that 'global warming is happening'. In those counties were belief in climate change is very weak, the value of land is likely insufficiently sensitive to projections of the effects of climate change. Bakkensen & Barrage (2018) document similar heterogeneity in beliefs about coastal flooding risks, and show that this can explain the variation in the capitalisation of climate change risks into coastal housing markets. While these are limited examples, the lesson applies to asset prices more generally.

2.3.4 Implications for cost benefit analysis

If we accept the arguments in the previous three subsections, it follows that the marginal rates of substitution that are captured by competitive market prices do not necessarily contain all the information required to assess the welfare consequences of public investments, even if markets are perfect. Nevertheless, when markets are perfect planners can still use them to frictionlessly redistribute consumption across time and states of the world, and prices still reflect opportunity costs in the planner's budget constraint. Does this imply that it is prices, not welfare-relevant marginal rates of substitution, that are important, regardless of the concerns raised in the discussion above?

To address this question, consider a planner with arbitrary preferences over aggregate consumption bundles represented by $U(\mathbf{c})$.[19] The planner's preferences may be assumed to account for issues relating to the CL critique, intergenerational concerns, and to represent

[18] Belief heterogeneity that stems purely from differing priors is less problematic than heterogeneity that stems from differing information sets, or idiosyncratic irrational belief updating. In reality however, it can hardly be contested that heterogeneity stems from all these channels. Indeed agents' priors can strongly reinforce biases in belief updating, e.g. via confirmation bias. See e.g. Millner & Ollivier (2016) for further discussion.

[19] We focus here on preferences over aggregate (or per capita) consumption, and thus again abstract from intratemporal distributive issues. It is straightforward to extend our discussion to include these concerns in the planners' objective function.

beliefs about future states of the world based on the best available evidence. Suppose that the current state of the economy is $\mathbf{c}_0$, and the planner seeks to evaluate a marginal project that yields payoffs $\boldsymbol{\pi} = (\pi_{t,s})_{t=0\ldots T,s=1,\ldots,S}$. Since markets are complete, the planner can in principle achieve any desired redistribution of the project payoffs across time and states by borrowing and lending on the market. Under this view, the value she would obtain from the project is

$$V(\boldsymbol{\pi}) = \max_{\tilde{\boldsymbol{\pi}}} U(\mathbf{c}_0 + \tilde{\boldsymbol{\pi}}) \text{ s.t. } \sum_{t=0}^{T} \sum_{s=1}^{S} p_{t,s}\tilde{\pi}_{t,s} \leq \sum_{t=0}^{T} \sum_{s=1}^{S} p_{t,s}\pi_{t,s}. \tag{17}$$

The envelope theorem now immediately yields

$$\frac{\partial V}{\partial \pi_{t,s}} = \lambda p_{t,s},$$

where $\lambda > 0$ is the Lagrange multiplier on the budget constraint (which is always binding), and hence for marginal projects $\boldsymbol{\pi}$

$$V(\boldsymbol{\pi}) > V(\mathbf{0}) \iff \sum_{t,s} p_{t,s}\pi_{t,s} > 0.$$

When markets are perfect a project seems to improve the planner's welfare measure if and only if it improves her market budget position. So does this mean that the concerns about the welfare significance of prices we laid out above are irrelevant?

In general the answer is no; the redistribution of consumption the planner undertakes in (17) will generally be highly non-marginal, thus violating the implicit assumption that prices are independent of the planner's actions. Indeed, the problem in (17) is equivalent to determining the optimal consumption bundle $\mathbf{c}^*$, subject to the constraint that the market value of $\mathbf{c}^*$ is equal to the market value of $\mathbf{c}_0 + \boldsymbol{\pi}$; the planner is assumed to be able to redistribute the economy's entire aggregate consumption bundle $\mathbf{c}_0$ at constant prices!

A correct account of the effects of non-marginal redistributive actions would require a full general equilibrium model of the economy, but that is beyond the scope of our discussion here. Nevertheless, it is of interest to ask how the planner's ability to use the markets to pursue social objectives might affect project evaluation at the margin, given the current (non-optimal) market equilibrium. One approach to this question is to study a problem where the planner may use the markets to redistribute a marginal project's payoffs, but the redistributed payoffs are themselves constrained to be marginal. In this setting it is reasonable to assume that prices are unaffected by the planner's actions as before. A useful formulation of this problem is as follows:

$$
\begin{aligned}
\tilde{V}(\boldsymbol{\pi}) = &\max_{\tilde{\boldsymbol{\pi}}} \sum_{t,s} \left.\frac{\partial U}{\partial c_{t,s}}\right|_{\mathbf{c}_0} \tilde{\pi}_{t,s} \text{ subject to} \\
&\text{i) } \sum_{t,s} p_{t,s}\tilde{\pi}_{t,s} \leq \sum_{t,s} p_{t,s}\pi_{t,s}, \\
&\text{ii) } -\frac{1}{2}\sum_{t,s}\sum_{t',s'} \left.\frac{\partial^2 U}{\partial c_{t,s}\partial c_{t',s'}}\right|_{\mathbf{c}_0} \tilde{\pi}_{t,s}\tilde{\pi}_{t',s'} \leq -\frac{1}{2}\sum_{t,s}\sum_{t',s'} \left.\frac{\partial^2 U}{\partial c_{t,s}\partial c_{t',s'}}\right|_{\mathbf{c}_0} \pi_{t,s}\pi_{t',s'}. \qquad (18)
\end{aligned}
$$

The objective function in this problem is the marginal effect of the redistributed project payoffs $\tilde{\boldsymbol{\pi}}$. The inequality in i) is the budget constraint, and ii) is a marginality constraint on the redistributed payoffs. This latter constraint says that the second order effects of the redistributed project payoffs on the planner's welfare measure cannot exceed the second order effects of the original payoffs (which we were happy to neglect).[20] Applying the envelope theorem to this problem yields

$$
\frac{\partial \tilde{V}}{\partial \pi_{t,s}} = \lambda p_{t,s} - \mu \sum_{t',s'} \left.\frac{\partial^2 U}{\partial c_{t,s}\partial c_{t',s'}}\right|_{\mathbf{c}_0} \pi_{t',s'},
$$

where $\lambda \geq 0$ is the Kuhn-Tucker multiplier on the budget constraint, and $\mu \geq 0$ is the multiplier on the marginality constraint. A little reflection shows that provided the market value of the project is small, both constraints are binding (i.e., $\lambda > 0, \mu > 0$), and the associated cost benefit rule becomes

$$
\tilde{V}(\boldsymbol{\pi}) - \tilde{V}(\mathbf{0}) > 0 \iff \sum_{t,s} p_{t,s}\pi_{t,s} - \frac{\mu}{\lambda}\sum_{t,s}\sum_{t',s'} \left.\frac{\partial^2 U}{\partial c_{t,s}\partial c_{t',s'}}\right|_{\mathbf{c}_0} \pi_{t,s}\pi_{t',s'} > 0. \qquad (19)
$$

The first term in this expression is again the market value of the project, and the new second term, which is non-negative, reflects the value of a small relaxation of the marginality constraint. The expression in (19) makes is clear that positive market value is sufficient, but not necessary, for a project to pass the cost-benefit test. In general both prices and the planner's preferences may be important, even if markets are perfect. We will return to

[20]By assumption the matrix of second derivatives with elements $H_{(t,s),(t',s')} = \left.\frac{\partial^2 U}{\partial c_{t,s}\partial c_{t',s'}}\right|_{\mathbf{c}_0}$ is negative semi-definite, so the sums in (18) are non-negative. The marginality constraint is required for the problem to be well-posed. If it were not present the optimal solution would require unbounded values of $\tilde{\boldsymbol{\pi}}$, thus again violating marginality.

a detailed discussion of the issues involved in specifying planner preferences in Section 3.

2.4 Market imperfections and their implications

Until now our discussion of the relationship between prices and SDRs has rested on generous market perfection assumptions. Yet in the real world markets are not perfect. The list of standard critiques of the perfect market paradigm is long and well known; it includes information asymmetries, market power, incompleteness, externalities, non-convexities, and issues of distributive justice. What are the consequences of these market failures for social discounting?

At first sight it may seem that market imperfections pose no special difficulties for the relationship between prices and consumption discount rates that we laid out in Section 2.1. Setting concerns about the welfare significance of current consumers' preferences aside, all that was required to identify marginal rates of substitution with (ratios of) state prices in that analysis was that consumers choose optimal consumption bundles taking prices as given. This assumption may hold even in imperfect economies, for example, in the presence of externalities, market incompleteness, or market power. Market equilibria in these cases will generically not be efficient, but at the inefficient equilibrium marginal values still coincide with consumer prices. Can we thus conclude that observed prices still contain all the information that is necessary for setting social discount rates, even in imperfect economies? We discuss two issues that suggest that things are unlikely to be so simple: missing prices, and the informational requirements of non-marginal policy instruments.

2.4.1 Missing prices

A critical assumption of our analysis in Section 2.1 was that markets are complete, i.e., a market for each time and state contingent Arrow-Debreu security exists. This is an extremely demanding assumption. With N goods, S states of the world, and $T+1$ time periods (with no uncertainty at $t=0$), completeness requires $N(TS+1)$ markets to operate at time $t=0$, when trading occurs. However, the Arrow-Debreu model we discussed can be reformulated as a model of sequential trading of goods on spot markets, with a single Arrow security that enables wealth transfers across times and states traded at $t=0$ (assuming that agents have rational expectations about future prices). In this reformulation the number of markets that need to operate is only $TS+1+N$. Even in this parsimonious reformulation however, moderately large values of T and S lead to a very large number of markets.[21]

As Geanakoplos (1990) has observed, "For a quarter of a century, scores of economists have complained about the absurdity of allowing all agents...to meet together at one moment in time, and to trade assets that allow for every conceivable contingency, for all future time." His survey of the literature on incomplete markets points out that there are many reasons why market incompleteness is inevitable. Asymmetric information may mean that the occurrence of a state is not visible to both parties to a potential transaction, so that the transaction cannot occur. Many interested traders may not have access to the markets, either because they are not yet born, or because they are not well-informed about market opportunities or because they do not have access to liquidity. Finally, if the market is thin, the transaction costs of establishing and maintaining a market may be too great for the operation to be profitable (recall that most major markets are run by for-profit corporations).

These observations imply that when markets are incomplete there are no prices for states that are not spanned by the available assets in the economy. By definition, the market has nothing to say about how payoffs that occur in those states should be discounted. For example, as the longest maturity government bonds are typically 30 years, there is arguably no way of pricing risk-free payoffs at greater maturities.[22] Our discussion of the empirical estimation of discount rates from certain real estate markets shows that it may be possible to get some bounds on long-run rates, but these estimates are necessarily limited by the fact that real estate is a risky asset with a particular risk profile. What is needed is a set of assets that spans the payoff space, and that is almost certainly beyond practical reach.

2.4.2 Prices, preferences, and non-marginal policy instruments

As we noted at the beginning of this section, the presence of market failures does not necessarily imply that prices are devoid of welfare content. Prices capture current consumers' marginal rates of substitution at the existing, inefficient, market equilibrium, and are thus relevant but not necessarily decisive inputs for the cost-benefit analysis of marginal projects. And yet this argument may leave some readers uneasy: what is the welfare sig-

[21] This is especially true if we interpret the state space as representing dynamic uncertainty, i.e., uncertainty about which sequence of events will unfold in the future. In that interpretation (discussed in Footnote 2), the relevant state space grows exponentially with the number of time periods in the model, and quickly becomes astronomically large.

[22] Indeed, even if there were a liquid market in much longer maturity bonds, their prices would reflect non-negligible default risks in even the most stable countries.

nificance of observed prices if they do not decentralise efficient allocations? Put another way, perhaps observed prices are simply 'wrong'?

If the market equilibrium is inefficient, a social planner will clearly want to intervene to correct the inefficiency using non-marginal policy instruments. These instruments will alter prices and consumption discount rates. The key question for our purposes is whether all the information that is required to choose these corrective instruments is contained in the competitive prices that we observe *before* the planner intervenes. If it is then the planner can in principle correct inefficiencies without asking any more detailed questions about the appropriate social objective. Social discount rates could then be chosen to reflect the prices that are observed after market failures have been corrected. However, if ex-ante prices do not contain all the information needed to correct market failures, the tight connection between market observables and social discount rates *at the social optimum* breaks down.

Standard microeconomic theory should immediately make us suspicious about whether the state prices that we observe at a point in time can tell us what we need to know to choose instruments to correct market failures. In general, determining the optimal level of these instruments will require knowledge of consumers' preferences. Under certain conditions[23] preferences can in principle be inferred from Walrasian demand functions, i.e. observations of demand as a function of prices and income. But at any fixed moment in time we only observe a single point on the demand function: demand at the equilibrium price. This is not enough information to reconstruct preferences, and hence also not enough information to determine optimal corrective instruments, or social discount rates at the optimum.

To make this point more concrete, let's consider a simple model of an intertemporal externality due to climate change. Externalities, of course, mean that the private and social costs of an activity differ; in the absence of government intervention competitive equilibria are inefficient in this case. For the sake of analytical convenience we'll present a continuous time model; our previous expressions for e.g. consumption discount rates can easily be extended to this case.

Suppose that there is a continuum of identical consumers with unit mass, and that they have standard discounted utilitarian time preferences. Production generates CO_2 emissions that accumulate in the atmosphere, and alter the climate. Temperature change depends on aggregate emissions in the economy, and affects utility directly through a damage function.

[23] See Mas-Colell et al. (1995, pp 75-78) for an elaboration of the so-called 'integrability' conditions for reconstructing preferences from demand functions.

We write the planner's problem in this model as:

$$\max_{c_t} \int_0^\infty u(c_t)(1 - D(T_t)) e^{-\delta t} dt \text{ s.t. } \dot{k}_t = f(k_t) - c_t, \quad \dot{T}_t = \alpha f(k_t) \tag{20}$$

where c_t and k_t are per-capita consumption and capital at time t, $f(k)$ is a production function, T_t is the change in global mean temperature at time t from its reference value at $t = 0$, $D(T_t)$ is an increasing and convex damage function from temperature change, and α is the product of the emissions intensity of output and the temperature change per ton of CO_2 emissions.[24] A necessary condition for a social optimum in this model is

$$f'(k_t)\left[1 - \left(-\alpha \frac{\mu_t}{\lambda_t}\right)\right] = \delta + \eta(c_t)\frac{\dot{c}_t}{c_t} + \alpha f(k_t)\frac{D'(T_t)}{1 - D(T)}. \tag{21}$$

where $\eta(c) = -cU''(c)/U'(c)$, and λ_t and μ_t are the co-state variables associated with the state variables k and T respectively. The left hand side of this equation is the instantaneous rate of return on investment at time t, adjusted for the social cost of the emissions associated with a marginal unit of production at t, i.e., $-\alpha f'(k_t)\frac{\mu_t}{\lambda_t} > 0$. The right hand side is the rate of change of marginal utility at time t, i.e., $-\frac{d}{d\tau}\log\left(u'(c_\tau)(1 - D(T_\tau)) e^{-\delta\tau}\right)\big|_{\tau=t}$. At the social optimum these two quantities must be equal for all times t, otherwise it would be possible to increase social welfare by changing the quantity of investment.

Now compare this with what occurs in a competitive equilibrium. As consumers are infinitesimal, they treat the trajectory of temperature change as exogenous – they neglect the effect of their actions on the climate, and on the welfare of others. Denoting the rental price of capital at time t by r_t, consumers solve:

$$\max_{c_t} \int_0^\infty u(c_t)(1 - D(T_t)) e^{-\delta t} dt \quad \frac{dk_t}{dt} = r_t k_t - c_t \tag{22}$$

Firms' profit maximization conditions yield $r_t = f'(k_t)$, and we find that a necessary condition for a competitive equilibrium in this problem is:

$$f'(k_t) = \delta + \eta(c_t)\frac{\dot{c}_t}{c_t} + \alpha f(k_t)\frac{D'(T_t)}{1 - D(T_t)}. \tag{23}$$

[24] In more complex models emissions contribute to CO_2 concentrations, and temperature responds to increases in CO_2 concentration with some inertia. However, this simple model turns out to be a surprisingly good approximation to the latest models from climate science, which show that temperature change is an approximately linear function of cumulative emissions.

The return on investment in the competitive equilibrium is the left hand side of (23). This differs from the socially optimal return on investment in (21) by the external cost of a marginal unit of investment, i.e., $-\alpha f'(k_t)\frac{\mu_t}{\lambda_t}$. In both cases the return on investment coincides with consumption discount rates on the relevant path for the economy, but only in the latter case are these discount rates immediately identifiable from market observables.

Which of the expressions (21) or (23) should be used to set social discount rates? If we are at the inefficient competitive equilibrium, and no instruments are available to correct the externality, then (23) is the correct expression. Even though this equilibrium is inefficient, consumption discount rates are given by observed private rates of return on capital at this equilibrium. All we need to do is read off the rental price of capital $r_t = f'(k_t)$ on the equilibrium path, and set the discount rate at maturity t equal to $\frac{1}{t}\int_0^t f'(k_\tau)d\tau$.

Now consider the case of a government that takes action to correct the externality by taxing emissions. The government can internalise the externality by making producers pay a tax $\sigma_t = -\alpha\frac{\mu_t}{\lambda_t}$ per unit of output at time t. In this case the private return on investment at time t is $(1-\sigma_t)f'(k_t)$, and we can set the social discount rate at maturity t equal to $\frac{1}{t}\int_0^t(1-\sigma_\tau)f'(k_\tau)d\tau$. Now here is the punchline: the correct value of the Pigouvian tax σ_t, and hence the correct value of the consumption discount rate in the presence of this tax, *is not revealed by market prices in the competitive equilibrium.* The Pigouvian tax σ_t is the ratio of the shadow price of temperature (μ_t) to the shadow price of capital (λ_t). Crucially, shadow prices are not equal to market prices in the presence of externalities. The market price of temperature changes is zero in the competitive equilibrium, but the shadow price of temperature changes is negative.

In order to calculate the Pigouvian tax σ_t, the government must know consumers' preferences, firms' technologies, and the nature of the externality. None of these quantities is generically revealed by observed prices at a point in time. Rental prices of capital in the competitive equilibrium only reveal combinations of preference parameters, endogenous consumption growth rates, and climate damages via (23). These equations cannot be used to uniquely identify preferences, since for each value of t there is one equation, but several unknowns, including two preference parameters (δ and $\eta(c_t)$) and the unobserved future value of consumption growth. Some parameters may of course be calibrated by restricting the form of the utility function and damage function and calibrating on past data, and this is often what is done in practice. But this procedure will invariably introduce errors into the calculation of the Pigouvian tax, as in general no simple parametric forms for these functions will be able to match observed rates of return for all t. The tight link between currently observed prices and consumption discount rates that we had at the competitive equilibrium is broken when the government seeks to intervene to correct the externality.

It's important to be clear about exactly what is being said in this discussion. It may seem that any mention of Pigouvian taxes is out of place in a discussion of social discount rates, as these are non-marginal instruments, whereas the social discount rate captures consumers' relative valuations of marginal consumption changes. Our point, however, is that when the government seeks to correct market failures with non-marginal instruments consumption discount rates will change. Since optimal non-marginal instruments depend on information that is not generically captured in ex-ante prices at a point in time, so too do consumption discount rates *after* the government acts. Ultimately, the government must ask questions about the appropriate notion of welfare that cannot be answered by reading off market observables at a fixed point in time.[25]

This point is not specific to externalities; it applies to other market failures as well. Consider the case of incomplete markets, discussed in the previous subsection. The theory of general equilibrium with incomplete markets shows that competitive equilibria are generically not constrained Pareto efficient in this context: even the assets that do exist in the economy are not used efficiently (Geanakoplos & Polemarchakis, 1986). This implies that government intervention, e.g. in the form of long-run public investments that are not available to private investors, could lead to Pareto improvements. But to know how to intervene in the market, we need information that is not revealed by prices. That information will also be an indirect ingredient of social discount rates after the intervention has been implemented.

3 Normative discounting

In the previous section we presented several reasons for skepticism about whether market prices should determine social discount rates. An argument that is sometimes made, either implicitly or explicitly, is that markets provide the 'least bad' solution to the problem – if not via markets, how else are we to choose discount rates? This section deals with precisely that question – it discusses an alternative approach to social discounting that

[25]An objection to this point could be that we have restricted attention to a single observation of prices and demand, whereas a planner may have access to many observations of demand at different prices. However, Afriat's theorem shows that if finite demand data are rationalizable, there are in general many well-behaved utility functions that can explain the data. See Varian (1982) for a discussion of bounds on specific classes of utility functions from demand data.

does not rely solely on market data. This approach is often referred to as the 'normative approach'. By contrast, the approach to discounting that exclusively uses market observables (prices) is referred to as the 'positive approach'. Both terms are highly misleading. The 'positive' approach is in truth highly normative, as we saw in our discussion in Section 1. It requires us to make judgements about how to aggregate project consequences across different individuals (either via Kaldor-Hicks or an explicit social welfare function), and relies on a specific interpretation of the welfare significance of market prices. In addition, the 'normative approach' is not independent of 'positive' data, as we shall see. Rather than being 'positive' or 'normative', these approaches are distinguished by how seriously they take the issues raised in the previous section.

3.1 Welfarism & Planner Preferences

What is to be done if we take the possibility that prices may not capture the information needed for welfare analysis seriously? There are essentially two approaches in the literature. One is to maintain an evaluation framework that aims to reflect the welfare consequences of policy changes for diverse agents. The problem then becomes one of estimation and aggregation: we need to estimate how policies affect individual wellbeing using non-market methods, and aggregate heterogeneous policy effects, being careful to acknowledge the normative choices that are involved in that exercise. We refer to this as a 'welfarist' approach, as the data that are relevant for policy evaluation in this case are individual wellbeings. The second approach does not hinge on measuring individuals' welfare and aggregating *per se*. In this approach we imagine a social planner who has some *a priori* ethical preferences over alternative distributions of goods across time and states that may or may not be explicitly tied to individuals' wellbeings, and place 'reasonable' constraints on these preferences by imposing certain properties on them. These constraints come in the form of axioms, in much the same way that axioms are used to motivate decision-theoretic models of consumer behaviour. In practice the line between these two approaches is often blurred by the invocation of a representative agent with preferences of a specified form. Such an agent's preferences are intended to be a welfarist measure of diverse individuals' wellbeing. However, as the form of the representative agent's preferences is often highly specialised and implicitly motivated by axiomatic criteria (e.g. discounted utilitarian preferences), and the conditions for representative agents to exist are highly stringent (Gorman, 1953), this interpretation stretches credulity. Far better to acknowledge that in this arena economics comes into close contact with philosophy.

A common objection to the 'planner preferences' approach is that it is paternalistic – the planner imposes his or her own ethical stance on society, irrespective of its consequences for individuals' abilities to pursue the lives that they value.[26] That view is accurate, if somewhat lacking in nuance. While it is true that reduced form models of planner preferences need not be tied to individuals' wellbeing, the models themselves are silent on the data that planners may draw on to justify values for the parameters of a particular preference representation. In practice, calibrating normative parameters may, at least partially, draw on facts relating to individuals' wellbeing. In addition, this critique downplays the normative choices that are required in the welfarist approach – even if we do define an evaluation framework that reflects individuals' wellbeing, aggregating those wellbeings requires normative judgements that are independent of the wellbeing measures themselves. These are not dissimilar to the kinds of judgements that need to be made to specify *a priori* planner preferences. Finally, while it seems clear that the welfarist approach is more comprehensive and ethically defensible than the reduced form planner preferences approach, it places a very high informational burden on the analyst; implementing this approach in practice requires detailed models of the lives and deaths of myriad individual agents. The planner preferences approach, by contrast, usually deals with a single preference relation that depends on only a small number of sufficient statistics that summarise the state of the economy (e.g. aggregate consumption growth), and a small number of parameters that aim to capture key normative tradeoffs.

Because of its wide deployment in the academic and policy literature on social discounting, the remainder of this section will largely focus on 'planner preferences'. A casualty of this narrow focus is that we once again largely neglect distributional issues across individuals. We also abstract from controversial, but important, issues relating to population change, and its implications for normative welfare criteria.[27] For an excellent detailed treatment of both these issues we refer the reader to Fleurbaey & Zuber (2015).

[26] Of course, hybrid approaches, in which an ethical planner interacts with private agents with different preferences, are also possible. See e.g. Barrage (2018).

[27] For the most part our consumption values can be interpreted as consumption per capita, so that we implicitly adopt some form of average utilitarianism. We are well aware of the difficulties with this criterion in a variable population context (see Millner, 2013), but have chosen to focus on other issues in this paper.

3.2 Normative discount rates for expected discounted utilitarians

By far the most commonly deployed model of planner preferences in the literature on social discounting is the standard expected discounted utilitarian model. In this model the value of a consumption bundle $\mathbf{c}$ at any time τ is

$$V_\tau(\mathbf{c}) = \sum_{t=0}^{\infty} \left(\frac{1}{1+\delta}\right)^t \sum_s q_{t,s}^{(\tau)} U(c_{t,s}), \tag{24}$$

where $q_{t,s}^{(\tau)}$ is the subjective probability of state s at time $\tau + t$, conditional on information available at time τ. It is crucial to understand from the beginning that the function defined in (24) should be interpreted as a normative evaluation principle, and not as a positive description of choice. We should think of (24) as a reduced form intertemporal social welfare measure, which aims to capture some of the key normative tradeoffs that are relevant to making intertemporal choices at a high level of aggregation. We begin this section with a derivation and discussion of the consumption discount rates that follow from (24). After that we discuss the assumptions that are implicit in this representation of planner preferences, and alternatives to them.

Applying the formula (4) for the consumption discount rate in state (t, s) to the preferences in (24) immediately yields:

$$(1 + \rho_{t,s})^{-t} = \left(\frac{1}{1+\delta}\right)^t q_{t,s} \frac{U'((1 + g_{t,s})^t c_0)}{U'(c_0)}.$$

where we have defined the consumption growth rate $g_{t,s}$ in state (t, s) via $c_{t,s} = c_0(1 + g_{t,s})^t$, and dropped the τ superscript on $q_{t,s}$ for convenience. This expression is easier to manipulate in the continuous time limit.[28] In this limit we find

$$e^{-\rho_{t,s} t} = e^{-\delta t} q_{t,s} \frac{U'(c_0 e^{g_{t,s} t})}{U'(c_0)}.$$

Assuming that $U(c)$ is concave and twice continuously differentiable, define the elasticity of marginal utility:

$$\eta(c) = -c \frac{U''(c)}{U'(c)}.$$

[28]In this limit we send $\rho_{t,s} \to \rho_{t,s}\Delta t, \delta \to \delta\Delta t, g_{t,s} \to g_{t,s}\Delta t, t \to t/\Delta t$, and take the limit as $\Delta t \to 0$.

Treating this definition as a differential equation for $U'(c)$, we find that

$$(U)'(c) = K \exp\left(-\int_0^c \frac{\eta(x)}{x} dx\right).$$

for some constant K. Make the change of variables $x = c_0 e^{g_{t',s}t'}$ in the integral in the exponent, evaluate $(U)'(c)$ at $c = c_0 e^{g_{t,s}t}$, and impose the initial condition $\lim_{t\to 0} U'(c_0 e^{g_{t,s}t}) = U'(c_0)$ to find

$$(U)'(c_0 e^{g_{t,s}t}) = U'(c_0) \exp\left(-g_{t,s} \int_0^t \eta(c_0 e^{g_{t,s}t'}) dt'\right).$$

Hence,

$$\rho_{t,s} = \delta + g_{t,s} \frac{\int_0^t \eta(c_0 e^{g_{t,s}t'}) dt'}{t} - \frac{1}{t} \ln q_{t,s}.$$

These state-dependent discount rates are rarely used in practice – most applications focus on the risk-free rate, or risk-adjusted rates that reflect correlations between project payoffs and discount factors in close analogy with our discussion of the CCAPM in (12).

The risk-free social discount rate ρ_t is calculated as

$$\begin{aligned} e^{-\rho_t t} &= e^{-\delta t} \sum_s q_{t,s} \exp\left(-g_{t,s} \int_0^t \eta(c_0 e^{g_{t,s}t'}) dt'\right) \\ \Rightarrow \rho_t &= \delta - \frac{1}{t} \ln\left[\sum_s q_{t,s} \exp\left(-g_{t,s} \int_0^t \eta(c_0 e^{g_{t,s}t'}) dt'\right)\right]. \end{aligned} \tag{25}$$

This is a general formula for the risk-free consumption discount rate when preferences are given by (24). It shows that discount rates are related to properties of planner preferences – the pure rate of social time preference δ and elasticity of marginal utility $\eta(c)$ – and to empirical forecasts of consumption growth ($q_{t,s}$).

Most applications of this formula usually impose a lot more structure on the model. It is commonly assumed that $\eta(c) = \eta$, a constant. In this case we find

$$\rho_t = \delta - \frac{1}{t} \ln\left[\sum_s q_{t,s} \exp\left(-\eta g_{t,s} t\right)\right].$$

Defining the cumulant generating function $\kappa_X(t)$ for the random variable X as

$$\kappa_X(y) = \ln\left[\mathsf{E} \exp(yX)\right],$$

where E is the expectation operator, we can write the risk-free rate in this case as

$$\rho_t = \delta - \frac{1}{t}\kappa_{g_t}(-\eta t).$$

Assuming that the cumulant generating function for growth rates has a Taylor expansion at zero, we can write it as:

$$\kappa_{g_t}(y) = \sum_{n=0}^{\infty} \frac{k_n^{(t)}}{n!} y^n$$

The Talyor coefficients $k_n^{(t)}$ are known as the cumulants of the distribution for g_t. They are related to the moments $\mu_n^{(t)}$ of this distribution. The first few cumulants are given by:

$$\begin{aligned}
k_0^{(t)} &= 0 \\
k_1^{(t)} &= \mu_1^{(t)} \\
k_2^{(t)} &= \mu_2^{(t)} - \mu_1^{(t)} \\
k_3^{(t)} &= \mu_3^{(t)} - 3\mu_2^{(t)}\mu_1^{(t)} + 2(\mu_1^{(t)})^3
\end{aligned}$$

If we truncate the Taylor expansion for $\kappa_{g_t}(y)$ at second order, we see that

$$\rho_t \approx \delta + \eta \mathsf{E} g_t - \frac{1}{2}\eta^2 t \mathrm{Var} g_t.$$

Noting that $g_t \sim \frac{1}{t}\ln(c_t/c_0)$, we can write this expression as

$$\rho_t \approx \delta + \frac{1}{t}\eta \mathsf{E}(\ln(c_t/c_0)) - \frac{1}{2t}\eta^2 \mathrm{Var}(\ln(c_t/c_0)).$$

This formula is the *Ramsey Rule* for the risk-free consumption discount rate when the utility function is iso-elastic. The first term is the utility discount rate, or pure rate of social time preference – this captures the planner's impatience with respect to utility flows. The second term is a wealth effect; it captures the planner's aversion to intertemporal consumption inequalities. Since the utility function is concave by assumption, $\eta > 0$, and hence marginal consumption changes in high consumption states are worth less than equivalent changes in low consumption states. This wealth effect is usually the dominant term in empirical calibrations of the Ramsey rule. The third term is a precautionary effect. Since Planners with iso-elastic utility have positive prudence[29] their appetite for

[29]The coefficient of relative prudence is $-cU'''/U'' = \eta + 1 > 0$.

precautionary savings is increased when the future is risky. This effect reduces the risk-free discount rate in proportion to the riskiness of future consumption.

In the classical case where c_t follows a geometric Brownian motion, this expression can be simplified further. In this case we have

$$\ln \frac{c_t}{c_0} \sim \mathcal{N}(\mu t, \sigma^2 t)$$

and hence

$$\rho_t = \delta + \eta\mu - \frac{1}{2}\eta^2\sigma^2. \tag{26}$$

As third order and higher cumulants of g_t are zero when consumption is log-normally distributed, this expression is in fact exact. It is worth repeating the assumptions that got us here: expected discounted utilitarian preferences, iso-elastic utility, and a geometric Brownian motion for consumption.

Equation (26) has played an influential role in discussions of social discounting. Gollier (2012) uses a calibration of this equation to motivate his baseline recommendation for social discount rates. His preferred parameter values for the preference parameters are $\delta = 0, \eta = 2$, while for the empirical parameters he uses summary statistics from empirical studies of consumption growth in the Western world to suggest $\mu = 2\%/\text{yr}$, $\sigma^2 = (0.036)^2$. This leads to a baseline recommendation that the risk-free rate be set at 3.6%/yr. The UK 'Greenbook' uses a deterministic version of this equation (i.e., with $\sigma^2 = 0$) to motivate the discount rates the UK government uses for project evaluation. Its preferred values are $\delta = 1.5\%/\text{yr}$, $\eta = 1, \mu = 2\%/\text{yr}$, leading to a discount rate of 3.5%/yr for maturities of less than 30 years. Although these quantitative recommendations are very similar, notice that they arise from markedly different values for the preference parameters.

While (26) is the simplest and most widely known version of the Ramsey rule, it clearly relies on some very strong normative and empirical assumptions. The core empirical assumptions relate to the model of consumption growth. In the geometric Brownian motion model consumption growth rates are independent and identically normally distributed. A large literature in both asset pricing and social discounting has studied variations of this model that allow for serial correlations in growth rates, uncertainty about the underlying growth process that fattens the tails of the distribution of growth rates, rare disasters, and several other effects. These effects can have qualitatively and quantitatively important consequences for discount rates – term structures are no longer flat when growth rates are serially correlated, and fat-tailed distributions for growth rates can dramatically inflate the magnitude of the precautionary term in (26). In addition, the role of the iso-elastic utility assumption has been interrogated. It has been shown that if $d\eta(c)/dc < 0$, and consumption growth rates are non-negative and independently and identically distributed, then the term structure of risk-free rates is declining. This can be seen directly by inspection of the general formula in (25). Thus the finding of a flat term structure in (26) is not robust either to the model of consumption growth, or to the choice of utility function. Gollier (2012) presents a comprehensive overview of these issues; we refer the reader there for further details.

It is worth pausing briefly to emphasise the difference between the approach we have adopted in this section and that in Section 2. There is not a single price in any of the formulae in this section – all the results follow from a direct assessment of the impact of a public project on a welfare measure that represents the planner's normative intertemporal preferences. Prices are only relevant to the extent that they play an implict role in any underlying model of consumption growth. In practice however, the applied literature tends to focus on exogenous time series models for consumption. While it is possible to interpret these models as equilibrium behaviour that emerges from some underlying model of the economy, that connection is seldom made explicitly in practical applications of discounting formulae. If it were, the implications for choosing discount rates might be quite different. For example, as in our discussion in Section 2.3.4, normative planners may still find it beneficial to use markets to redistribute project payoffs across time and states. If that is feasible (i.e., if the payoff streams from public projects are at least partially tradable), prices still determine opportunity costs, and may still be important inputs to cost benefit analysis.

3.3 Properties of discounted utilitarian time preferences

The discounted utilitarian time preferences that underpin the Ramsey rule are so familiar that it is easy to forget what the justification for using them was in the first place. In this section we drill down into the axiomatic properties of discounted utilitarian time preferences. Our purpose in doing this is to understand the precise assumptions that are implicit in this commonly used tool, and in so doing evaluate their normative credibility. Our discussion will show that, while discounted utilitarian preferences have a number of convenient and attractive features, they are arguably far from being the only plausible model of normative intertemporal preferences.

We begin with a discussion of the famous axiomatic foundation for discounted utilitarianism due to Koopmans (1960). Let $\mathbf{c} = (c_0, c_1, \ldots)$ be an infinite stream of future consumption, and let $(\mathbf{c}|_t\mathbf{c}')$ denote a future consumption stream that consists of the elements of stream $\mathbf{c}$ for the first t time periods and the elements of stream $\mathbf{c}'$ thereafter. Let h be an arbitrary history of consumption. We denote preferences over future consumption streams $\mathbf{c}$ at a history h by $V_h(\mathbf{c})$. Since preferences over consumption streams are conditioned on histories h we admit the possibility that they may differ at different histories.

The two most important of Koopmans' axioms are as follows:

- **Independence** – For all consumption streams $\mathbf{c}, \mathbf{c}', \hat{\mathbf{c}}, \tilde{\mathbf{c}}$, all $t' > t$, and all histories h:
$$V_h(\mathbf{c}|_t\hat{\mathbf{c}}|_{t'}\mathbf{c}) \succeq V_h(\mathbf{c}'|_t\hat{\mathbf{c}}|_{t'}\mathbf{c}') \iff V_h(\mathbf{c}|_t\tilde{\mathbf{c}}|_{t'}\mathbf{c}) \succeq V_h(\mathbf{c}'|_t\tilde{\mathbf{c}}|_{t'}\mathbf{c}'). \tag{27}$$

- **Stationarity** – For all consumption values x, consumption streams $\mathbf{c}, \mathbf{c}'$, and all histories h:
$$V_h(x, \mathbf{c}) \succeq V_h(x, \mathbf{c}') \iff V_h(\mathbf{c}) \succeq V_h(\mathbf{c}'). \tag{28}$$

Koopmans also assumes some other technical axioms (including a continuity axiom). Note that both these axioms concern preferences at a fixed history h. They say nothing about the relationship between preference at different histories.

The Independence Axiom is responsible for the additive separability of preferences. It is easy to show that if $V_h(\mathbf{c})$ takes the form

$$V_h(\mathbf{c}) = U_0^h(c_0) + U_1^h(c_1) + U_2^h(c_2) + \ldots$$

then Independence is satisfied. With some further technical axioms, this representation of $V_h(\mathbf{c})$ can be shown to be necessary for Independence. Additive separability is a strong property – it says that the relative value of a marginal unit of consumption in any two periods only depends on those periods. Formally, separability implies that for any i, j, k, where $i \neq j \neq k$:

$$\frac{\partial}{\partial c_k}\left(\frac{\partial V_h/\partial c_i}{\partial V_h/\partial c_j}\right) = 0.$$

In a nutshell, the value of an additional mouthful of Spam for dinner, relative to its value at breakfast time, does not depend on how much Spam you ate for lunch.

Stationarity is actually also a kind of independence property of preferences. It says that if two consumption streams share a common 'beginning', then our preferences between them should be the same as our preferences between two modified consumption streams that are the same as the original pair, but with their common beginning deleted. Stationarity is also a strong property. For example, it implies that for an arbitrary infinite stream $\mathbf{c}$,

$$\begin{aligned} &(\text{Star Wars I}, \mathbf{c}) \succ (\text{Star Wars II}, \mathbf{c}) \\ \iff &(\text{Star Wars I, Star Wars I}, \mathbf{c}) \succ (\text{Star Wars I, Star Wars II}, \mathbf{c}) \end{aligned}$$

That is, if I prefer watching Star Wars I to Star Wars II, and then continuing with my life ($\mathbf{c}$), I should also prefer watching Star Wars I twice in a row to watching Star Wars I followed by Star Wars II, and then continuing with my life.[30] Clearly there are some situations in which stationarity may not capture important aspects of the interactions between successive consumption values, although it may arguably be a less objectionable property to assume of social preferences over a composite consumption variable.

To see (roughly) that Stationarity and Independence imply discounted utilitarianism, consider two consumption streams $\mathbf{c}, \mathbf{c}'$, where $V_h(\mathbf{c}) \succ V_h(\mathbf{c}')$. By independence this means

$$U_0^h(c_0) + U_1^h(c_1) + U_2^h(c_2) + \ldots > U_0^h(c_0') + U_1^h(c_1') + U_2^h(c_2') + \ldots$$

The Stationarity axiom now says that $V_h(c_0, \mathbf{c}) \succ V_h(c_0, \mathbf{c}') \iff V_h(\mathbf{c}) \succ V_h(\mathbf{c}')$, so setting $c_0 = c_0'$ we have

$$U_1^h(c_1)+U_2^h(c_2)+\ldots > U_1^h(c_1')+U_2^h(c_2')+\ldots \iff U_0^h(c_1)+U_1^h(c_2)+\ldots > U_0^h(c_1')+U_1^h(c_2')+\ldots \tag{29}$$

It is clear that a sufficient condition for this to hold is

$$U_{t+1}^h(c) = (1+\delta_h)^{-1} U_t^h(c). \tag{30}$$

for some $\delta_h > 0$. This can be verified be substituting (30) into the left inequality in (29). With the use of some other technical axioms, Koopmans shows that (30) is necessary, as well as sufficient. Solving (30) explicitly as a function of time, we arrive at our old friend discounted utilitarianism:

$$V_h(\mathbf{c}) = \sum_{t=0}^{\infty} (1+\delta_h)^{-t} U^h(c_t). \tag{31}$$

[30]This example is taken from Machina (1989).

Although the Koopmans axiomatics lead to a discounted utilitarian representation of preferences, one further assumption is needed to get to the standard (deterministic) formula in (24). The key observation is that nothing so far in our discussion tells us how preferences at different histories are related. Thus, if a decision maker obeys the Koopmans axioms, she could have a different discounted utilitarian preference relation at each point in time. To get to the standard result we must impose one of two additional conditions:

- **Time consistency**: For all histories h, consumption values x, and future consumption streams $\mathbf{c}, \mathbf{c}'$:

$$V_h((x, \mathbf{c})) \succeq V_h((x, \mathbf{c}')) \iff V_{(h,x)}(\mathbf{c}) \succeq V_{(h,x)}(\mathbf{c}') \tag{32}$$

Time consistency rules out preference reversals, and implies that if an optimal plan is implemented today, it will remain optimal to follow it tomorrow. In order to understand the constraints this requirement places on preferences, it is helpful to see how time consistency interacts with Stationarity. Notice that the left hand sides of (28) and (32) are the same, so if we impose time consistency on stationary preferences, it must also be the case that the following property holds:

- **Time Invariance**: For all histories h, consumption values x, and future consumption streams $\mathbf{c}, \mathbf{c}'$,

$$V_h(\mathbf{c}) \succeq V_h(\mathbf{c}') \iff V_{(h,x)}(\mathbf{c}) \succeq V_{(h,x)}(\mathbf{c}') \tag{33}$$

Time invariance is a property that is logically distinct from time consistency and stationarity, but is implied by their conjunction. In fact, it is easy to see that any two of these properties – time consistency, stationarity, and time invariance – implies the third. Thus we could equally have imposed time invariance in addition to the Koopmans axioms, and derived time consistency as a consequence.

Time invariance requires that preferences over future consumption streams be independent of translations of the time axis – shifting preferences forwards or backwards in time has no effect on rankings of future consumption streams. An immediate consequence of time invariance is that preferences are *history independent* – historical consumption cannot have any effect on the ranking of future consumption streams. We have shown that stationary preferences that obey the independence axiom are discounted utilitarian. In addition, if these preferences are to be time consistent they must be time invariant. This implies that discounted utilitarian preferences are time consistent if and only if they are the same in each history – i.e. the utility function $U(c)$ and the utility discount rate δ cannot vary with the passage of time.

This discussion has shown that although discounted utilitarian preferences are by far the most commonly used in the literature, they are actually highly restrictive. Not only do they severely constrain how interactions between consumption at different points in time can affect our preferences, they also require us to be entirely ahistorical in our evaluations of future consumption streams.

Our discussion thus far has emphasised time, and not risk/uncertainty. A detailed discussion of axiomatic approaches to intertemporal choice in the presence of risk and/or uncertainty would take us too far afield for our present purposes. We refer the reader to e.g. Hammond & Zank (2014); Ghirardato (2002) for detailed treatments. We note however that two of the standard constraints on rational dynamic choice do not imply expected discounted utilitarian preferences. *Dynamic consistency* is an extension of the time consistency property above to a stochastic context. It requires plans that are made today about how to act at future nodes of a decision tree that depend on the realisation of chance events remain optimal if those events are realised. *Consequentialism* says that preferences at each node of a decision tree should depend only on those nodes that are reachable from the current node. While each of these properties is appealing, combining them does not require that choices be represented by the discounted utilitarian preferences in (24). Johnsen & Donaldson (1985) show that these two criteria together imply that preferences must have a recursive representation, but significantly more structure is required for them to be additively separable across both time and states of the world, and a stationarity axiom is still required to generate exponential utility discount factors. The occasionally heard claim that (24) is required for consistent dynamic choice is manifestly incorrect.

Given that the constraints on planner preferences in (24) are highly restrictive, it is natural to explore the implications of alternative preferences for normative consumption discount rates. It is straightforward to extend the analysis in Section 3.2 to other preferences. See e.g. Bansal & Yaron (2005); Traeger (2014) for a discussion of discount rates derived from Epstein-Zin preferences, Traeger (2014); Gierlinger & Gollier (2017); Collard et al. (2018) for discount rates that account for ambiguity aversion, and Campbell & Cochrane (1999) for discount rates in the presence of habit formation.[31] Backus

[31] Gollier (2010); Traeger (2011); Gueant et al. (2012) provide discussions of discount rates for multi-attribute utility functions, but these do not require us to deviate from the standard time- and state-separable framework.

et al. (2004) is a handy reference for so-called 'exotic' preferences and their applications. While there are important discussions to be had about which (if any) of these alternative preference formulations has normative appeal for social discounting, we hope that this section illustrates that those discussions should also be had about the standard expected discounted utilitarian paradigm.

3.4 Calibrating δ and η: ethical arguments

Representation theorems such as that of Koopmans pin down the functional form of planner preferences, subject to us accepting the axioms they rely on. But to operationalize these representations for social discounting applications we need to specify the free normative parameters of these functional forms. In the case of discounted utilitarianism, this means choosing a utility function $U(c)$, and a utility discount rate δ.

In applications the utility function is often chosen to be iso-elastic. Iso-elastic utility makes income effects very simple – the elasticity of intertemporal substitution is a constant, and equal to $1/\eta$. Iso-elastic utility is also required for balanced growth paths to exist in standard economic growth models – see e.g. Acemoglu (2008). In the time and state-separable model in (24), η also captures risk attitudes; it is the coefficient of relative risk aversion, which is again constant for iso-elastic utility functions. The preferences in (24) do not permit a separation between attitudes to risk and attitudes to intertemporal consumption smoothing. Although iso-elastic utility functions have a number of very convenient properties, it is important to be aware that it is an essentially arbitrary choice from a normative perspective. Ultimately the main reason for the focus on this utility function is analytic simplicity.

How should we interpret and calibrate η for social discounting applications? Gollier (2018) reviews the various methods that have been used in this arena. Positive approaches to calibrating η include empirical estimates of risk aversion from laboratory and field experiments and studies of societal inequality aversion as reflected in observed income tax schedules. A more purely normative approach uses simple examples to inform intuition about ethically appropriate values for η. For example, consider a \$1 transfer from an individual with consumption $2c$ to someone with consumption c. If we weight these individuals' utilities equally, the maximum fraction x of this transfer that can be 'lost in transit', while still making the transfer socially desirable, satisfies

$$U'(2c) \times 1 = U'(c) \times (1 - x) \Rightarrow x = 1 - 2^{-\eta}.$$

For example, if we feel that it is acceptable to lose at most 25% of the transfer, we should set $\eta = -\log_2(3/4) \approx 0.4$. If 75% is an acceptable maximum loss, we should set $\eta = 2$. This example shows how η captures inequality aversion, and can help to form intuition for 'reasonable' values. Clearly however, there is room for a wide range of opinions on the ethically acceptable value of x, and hence inequality aversion η. Commonly deployed values lie in the range 1-4.

Similarly, the value of δ is often chosen either to fit empirical data on the rate of return on capital, or from normative equity considerations. We have already discussed the reasons to be skeptical about attempts to calibrate δ based on market observables, and thus focus on the normative arguments here.

If the time periods in the discounted utilitarian model represent different generations, e.g. we are making choices about something like climate change, it is natural to allow normative fairness considerations to influence our choice of δ. A commonly advocated choice is $\delta = 0$, i.e. intergenerational equity. There are many good arguments for this value, going all the way back to Sidgwick (1874) and Ramsey (1928). Broome (2014) and Greaves (2017) provide recent surveys. In a nutshell, arguments for $\delta = 0$ turn on the undeniably compelling claim that ethical principles should be impartial, i.e., accidents of birth or circumstance should play no role in a sound theory of distributive justice. Since positive δ implies that a util experienced today is worth substantially more than the same util experienced by agents who happen to live their lives in the distant future, it must surely violate any meaningful notion of impartiality.[32]

Since the impartiality arguments for $\delta = 0$ are quite straightforward, and are by now very well-covered ground, we will focus on arguments against $\delta = 0$. A seminal result in this arena is the impossibility theorem of Diamond (1965):

Theorem 1. *Social welfare orders (i.e. complete, transitive, reflexive relations) defined over infinite bounded utility streams, and that satisfy:*

- *Strong Pareto: If $c_t \geq c'_t$ for all t, and there exists τ such that $c_\tau > c'_\tau$, then $\mathbf{c} \succ \mathbf{c}'$.*
- *Continuity: The order is continuous in the sup norm topology.*
- *Equity: The order is indifferent between utility streams that are finite permutations of one another.*

[32] Advocates of this impartiality argument (e.g. Stern, 2007) still sometimes adopt a slightly positive value of δ intended to reflect a constant 'extinction probability' per unit time. From an empirical perspective both the assumed value of this probability (e.g. 0.1%/year), and its constancy over time, are somewhat arbitrary.

do not exist.

This result has been strengthened over the years – see Asheim (2010) for a review of recent developments. So, if we want to define complete, continuous, social preferences over infinite consumption streams that are also equitable, we have to give up Strong Pareto, a property that is usually considered as an uncontroversial efficiency requirement. Alternatively, we can admit Strong Pareto and Equity, but then we'd have to give up completeness. There is thus a very deep conflict between the desire for impartial, sensitive evaluation criteria, and the notion of a numerically representable social relation of the kind we are used to.

A second argument against $\delta = 0$ is that it can give rise to the so-called 'paradox of the indefinitely postponed splurge'. To see an example of this, consider the following optimal intergenerational savings problem:

$$\max_{\{c_t\}} \sum_{t=0}^{\infty} (1+\delta)^{-t} U(c_t) \ \ s.t. \ \ k_{t+1} = (1+r)(k_t - c_t),$$

where

$$U(c) = \begin{cases} \ln c & \eta = 1 \\ \frac{c^{1-\eta}}{1-\eta} & \eta \neq 1, \end{cases}$$

and we assume that $\delta \geq 0, r > 0, \eta \geq 0$. Standard calculations (see e.g. Dasgupta, 2008) show that an optimal program exists in this problem if and only if $(1+r)^{1-\eta} < 1+\delta$. This condition is always satisfied when $\eta > 1$, but may not be when $\eta \leq 1$. When an optimum exists, the optimal savings rate $s_t = (k_t - c_t)/c_t$ is constant, and given by:

$$s^* = (1+r)^{-\frac{\eta-1}{\eta}} (1+\delta)^{-\frac{1}{\eta}}. \tag{34}$$

Consider the case $\eta = 1$. In this case, as $\delta \to 0$, the optimal savings rate $s^* \to 1$ in every generation. Thus each generation values the future so highly that it starves itself today for the sake of future generations. This is true of every generation, so no one ever benefits from this selfless altruism. Hence the paradox. While this result is clearly unpalatable, it is not a generic feature of optimal programs as $\delta \to 0$; it relies on the fact that utility is unbounded above for $\eta = 1$. In fact, we can see from (34) that when $\eta > 1$ (i.e., utilities are bounded above), $s^* \to (1+r)^{-(\eta-1)/\eta} < 1$ as $\delta \to 0$. Nevertheless, efficient savings rates in this limit may still be unpalatably high. For example, if $r = 5\%/\text{yr}$, $\eta = 2$, we have $s^* \to 97.6\%$ when $\delta \to 0$ – this still seems an excessively burdensome prescription. The fact that very low values of δ tend to favour very high savings rates is symptomatic of a broader set of concerns about 'excessive sacrifice'.[33]

Consider a project that yields a small utility benefit ϵ to the next T generations. If $\delta = 0$, then no matter how small we make ϵ, there is always a T large enough so that the project is welfare improving, even if it costs the current generation all of its utility. This is the real heart of the 'excessive sacrifice' line of complaint against $\delta = 0$. The deep tension between impartiality and the rights of the individual is eloquently summarised by Arrow (1999): 'I do not think that this dilemma arises merely out of the specifically utilitarian formulation that welfare economists find so congenial. It is a conflict between a basic property of morality...that of universalizability, and a principle of self-regard, of the individual as an end and not merely as a means to the welfare of others. In a favourite quotation of mine, Hillel, the first-century rabbi, asked,"If I am not for myself, then who will be for me? If I am not for others, then who am I? If not now, when?" One can only say that *both* the universal other and the self impose obligations on an agent.'

It is tempting to object that both the 'non-existence' and 'excessive sacrifice' arguments against $\delta = 0$ are an artefact of the assumption of an infinite time horizon. Certainly it is a mathematical fact that both problems, conceived in their narrowest sense, evaporate in a model with a finite time horizon. It is also a physical fact that humanity will almost certainly cease to exist in finite time. However, these observations merely deflect attention from the conceptual concerns raised by these arguments. The excessive sacrifice concern emerges approximately for very large time horizon models, or for models where $\delta > 0$ but arbitrarily small – optimality may still require sacrifices that are unpalatable in such models. Similarly, while an undiscounted intertemporal welfare function avoids all the difficulties of Diamond's theorem with a finite time horizon, the ranking of consumption streams can depend very strongly on where the temporal cutoff is set when $\delta = 0$. Since any choice for the time horizon is essentially arbitrary, or at best highly uncertain, this seems an undesirable feature for a normative criterion. One might then invoke the principle that the time horizon should be large enough that the terminal conditions (i.e., the assumed final values of state variables) of any intertemporal optimization problem we are interested in should not have much influence on the optimal solution, since any choice of terminal

[33] In general, Asheim & Buchholz (2003) observe that any efficient and increasing consumption stream is the optimal path for some undiscounted welfare function, so it is perfectly possible to find undiscounted objectives that favour arbitrary savings patterns. The question then becomes whether those objectives satisfy other desirable normative criteria.

values will be arbitrary. But this requirement is, to all intents and purposes, identical to requiring a very large, if not infinite, time horizon when $\delta = 0$. It seems that when it comes to equitable evaluation of very long utility streams there is no free lunch.

One final argument against $\delta = 0$ is, of course, that it may not reflect the intertemporal objectives of currently living consumers, both with respect to their own consumption, and with respect to their concern for their distant descendants. Supporters of this argument see an imposition of $\delta = 0$ as a deeply paternalistic move. From an ethical perspective this view can be critiqued for confusing an 'is' with an 'ought': the fact that agents behave a certain way does not imply that this behaviour should be lifted to the status of a normative principle for social decisions. Nevertheless, for commentators who adopt a narrow normative framework that identifies the social good with the satisfaction of the preferences of currently living consumers, $\delta = 0$ does not seem a plausible value.

We take up the theme of disagreements about normative parameters once more in Section 4, but now turn to practical issues that arise when evaluating public projects in imperfect economies.

3.5 Project evaluation in imperfect economies

Thus far our discussion has focussed entirely on the consumption impacts of public projects. Of course, projects may be financed by a variety of means that do not involve direct consumption changes, and they may also have effects on other economic variables, e.g. private investment. When markets are perfect all these effects can be handled using a unified approach, as consumption discount rates coincide with market rates of return on all margins. This is no longer the case when there are distortions in the economy. An extensive literature discusses project evaluation in such second-best situations.[34] The central issue is whether there are general rules governing project evaluation in these cases, or whether the outcomes are specific to the distortions that operate in the economy. We close this section with a small digression from our discussion of the 'pure' normative approach to discuss this issue. We show that there are some general principles, though their implementation is not always straightforward.

To see the basic idea in a simple example consider Figure 2. This represents a two-period economy, with present consumption on the horizontal axis and future consumption on the vertical axis. Think of the vertical axis as representing a composite of all future

[34]Classic contributions on this topic in a general equilibrium setting include Arrow & Kurz (1970); Sandmo & Dreze (1971). We discuss some of the partial equilibrium literature below.

consumption. The intertemporal production frontier is shown as a red line and the blue lines are indifference curves for a representative consumer, or social planner. Clearly the first best is the point A where an indifference curve is tangent to the frontier. Here the consumer's marginal rate of substitution and the producer's marginal rate of transformation are equal – the consumption discount rate and the rate of return on investment are the same. Now suppose however there are distortionary taxes in the economy which drive these two rates apart: then the system will be at a point like B, which is inefficient. At such a point, what rate should be used to evaluate a small change in the economy's configuration? Clearly a change is beneficial if and only if it is feasible and it moves the economy above the indifference curve through B, and for small changes this is true if and only if the proposed (feasible) change has positive value at the prices given by the slope of the tangent to the indifference curve at B, which means positive value at the consumption discount rate. So a move from B to C is beneficial, even though neither B nor C is efficient. Note that the slope of the production frontier at B has no role to play in evaluating a project in this case. So here at least using the consumption discount rate is the right approach. How robust is this conclusion?

The answer is that it is robust, provided that we can accurately identify all of the consequences of a project and value them appropriately. In case this sounds trivial let

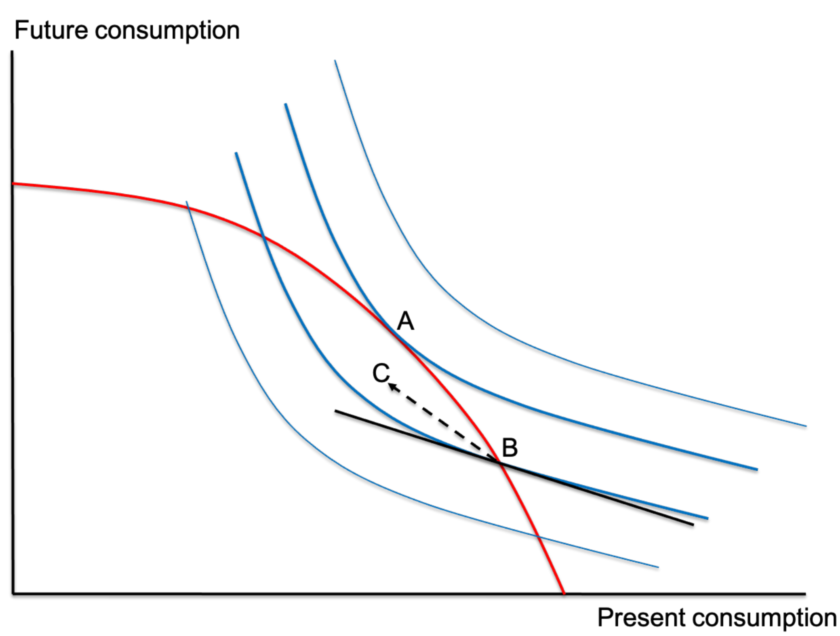

Figure 2: Appraising marginal projects in distorted economies.

us illustrate the possible complexities. Consider an investment in wind power, which will provide carbon-free electricity for thirty years. This investment may be paid for by a government or by a private company. In the former case it might be financed by issuing bonds, or by a tax on profits, labor income or sales. In the latter case it could be financed by issuing corporate bonds or equity, or from retained earnings. Clearly the implications of these seven alternatives could be quite different, both now and in the future. An income tax would reduce savings and consumption, and hence reduce output and profits. Financing via government bonds might reduce private investment and hence future output and profits, and so on. In terms of Figure 2, each of these would be represented by a different movement from B, even though they all lead to the same wind farm. However, provided that we identify precisely what these effects are, and value them with the appropriate shadow prices, we can keep discounting using the consumption discount rate.

The literature sometimes distinguishes between the direct and indirect effects of a project. In the case of a wind farm financed by an income tax, the direct effects would be the generation of power by the farm and the employment in constructing and operating the farm, together with the drops in consumption and savings resulting from the tax. The indirect effects would include the change in the consumers' work-leisure tradeoff because of the lower effective wage rate, the change in producers' outputs because of the drop in demand, and any consequences of the reduction in consumer savings. The direct effects are obvious to the managers of the project, whereas the indirect effects may be hard to evaluate. It is these that make it hard to be sure where a project starting at a point such as B in Figure 2 will actually take us. It is certainly possible that a project could have positive value if we consider the direct effects only, but negative value when all effects are taken into account - or vice versa.

To illustrate more formally how project evaluation proceeds in a non-optimal economy, consider a deterministic model in which the planner's intertemporal preferences are given by

$$V(\mathbf{c}) = V(c_0, c_1, c_2, \ldots),$$

and suppose that some non-optimal resource allocation mechanism operates in the economy. We represent this mechanism as a general (non-optimal) mapping between consumption and investment at time t, and consumption and investment at all prior times $\tau < t$:

$$c_t = C_t(c_{t-1}, i_{t-1}, c_{t-2}, i_{t-2}, \ldots, c_0, i_0) \tag{35}$$

$$i_t = I_t(c_{t-1}, i_{t-1}, c_{t-2}, i_{t-2}, \ldots, c_0, i_0) \tag{36}$$

for some set of functions $C_t(\cdot), I_t(\cdot)$. Now consider valuing a public project that leads to direct marginal changes in consumption and investment at time t, denoted by dc_t and di_t respectively. We can compute the total effect of this project by totally differentiating (35–36), holding consumption and investment values at times $0, \ldots, t-1$ fixed as they are in the past when the project's payoffs are realized:

$$dc_{t+\tau} = \sum_{n=1}^{\tau-1} \frac{\partial C_{t+\tau}}{\partial c_{t+\tau-n}} dc_{t+\tau-n} + \sum_{n=1}^{\tau-1} \frac{\partial C_{t+\tau}}{\partial i_{t+\tau-n}} di_{t+\tau-n} \tag{37}$$

$$di_{t+\tau} = \sum_{n=1}^{\tau-1} \frac{\partial I_{t+\tau}}{\partial c_{t+\tau-n}} dc_{t+\tau-n} + \sum_{n=1}^{\tau-1} \frac{\partial I_{t+\tau}}{\partial i_{t+\tau-n}} di_{t+\tau-n}. \tag{38}$$

To understand all the indirect consequences of this project on consumption and investment at future times $t+\tau$ for $\tau \geq 1$, we need to solve this linear system of difference equations forwards in time, given the stated initial conditions at $\tau = 0$. Let $\Delta_{t,\tau}(dc_t, di_t)$ be the time series of consumption changes that emerges from this procedure, given initial conditions (dc_t, di_t). By linearity, we have $\Delta_{t,\tau}(dc_t, di_t) = \Delta_{t,\tau}(1,0)dc_t + \Delta_{t,\tau}(0,1)di_t$. Defining the *shadow price* of investment at time t as

$$\nu_t = \sum_{\tau=1}^{\infty} \frac{\frac{\partial V}{\partial c_{t+\tau}}}{\frac{\partial V}{\partial c_t}} \Delta_{t,\tau}(0,1),$$

the appropriate cost benefit rule for a project that gives rise to a sequence of marginal consumption and investment changes given by $(dc_t, di_t)_{t=0,\ldots,\infty}$ is:

$$\sum_{t=0}^{\infty} dc_t(1+\rho_t)^{-t} + \sum_{t=0}^{\infty}\sum_{\tau=1}^{\infty}(1+\rho_{t+\tau})^{-(t+\tau)}\Delta_{t,\tau}(1,0)dc_t + \sum_{t=0}^{\infty}(1+\rho_t)^{-t}\nu_t di_t > 0. \tag{39}$$

The first term in this expression accounts for the direct effect of the project on consumption, the second term for indirect consumption effects, and the third term for changes in investment, valued at appropriate shadow prices.

The shadow prices and indirect consumption effects in (39) are of course dependent on the resource allocation mechanisms that operate in the economy. If the economy is at an intertemporal welfare optimum, the envelope theorem implies that only the first term of (39) is non-zero. However, in an economy with distortions, all the terms in (39) are relevant. This analysis illustrates the complexity of this exercise in general. A series of early papers examined the implications of these complexities for project appraisal in various simplified cases, see e.g. Marglin (1963); Feldstein (1964); Baumol (1968); Bradford (1975). See also Dreze & Stern (1987); Arrow et al. (2003) for discussions of shadow prices and their role in policy appraisal in imperfect economies. Notice that although the empirical consequences of market imperfections are somewhat formidable, the approach here is conceptually straightforward: keep track of all the indirect effects of consumption changes, and convert all project consequences for investment into consumption equivalents using appropriate shadow prices, before discounting using the consumption discount rate as normal.

4 Rapprochement: Social choice and social discounting

Our discussion of the market based approach to social discounting showed that there are a number of serious concerns about whether observed market prices capture the information that is needed to perform cost benefit analysis of intertemporal projects, especially if those projects affect outcomes in the distant future where markets are likely incomplete, intergenerational issues are salient, and externalities due to e.g. climate change are relevant. The alternative to the market approach is to try to estimate the welfare consequences of public projects directly, via a normative social discount rate. However, our discussion of that approach showed that it too suffers from some serious difficulties and indeterminacies. These arise when one asks which normative axioms are most appealing as foundations of 'planner preferences'. In addition, even supposing that we agree on a given representation for planner preferences (e.g., discounted utilitarianism) there are still free parameters that must be specified based on a combination of normative reasoning and empirical estimation. For example, given the tradeoffs involved in choosing appropriate values for δ and η in the discounted utilitarian paradigm, it is no surprise that informed opinions on their values vary widely. This is amply demonstrated in Figure 3, which plots the results of a recent survey (Drupp et al., 2018) that asked economists who have published papers on social discounting which values of δ and η are most appropriate for public project appraisal.

The variation in this figure is truly astonishing when it is translated into estimates of the social value of payoffs in the distant future. For example, assuming that consumption growth is a deterministic 2%/yr, the value of a \$1,000,000 payoff in 100 years is \$1,000,000 if $(\delta, \eta) = (0, 0)$ (bottom left of Figure 3), or \$2 if $(\delta, \eta) = (6\%/\text{yr}, 4)$ (top right of Figure 3). These two economists disagree about the value of 100 year payoffs by a factor of 500,000!

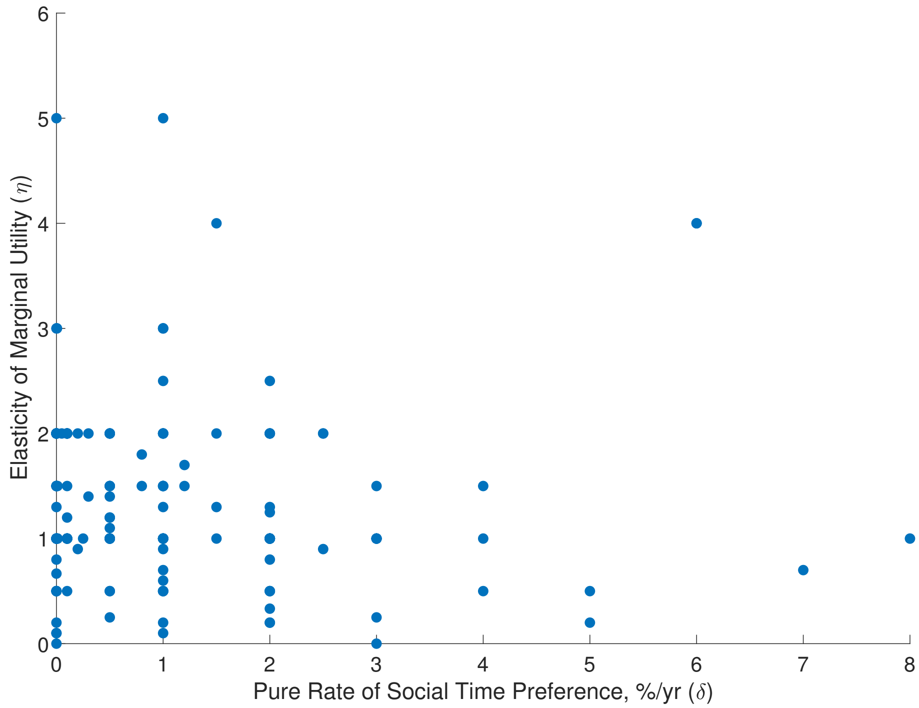

Figure 3: Values of δ and η recommended by a sample of economists who have published papers on social discounting. Data from Drupp et al. (2018).

This is an extreme example, but even much more modest differences of opinion about these parameters become highly significant at long maturities.

There are at least three possible responses to this diversity of opinion. One is to conclude that economics has little of value to say about how long-run projects should be evaluated. Since essentially any policy can be justified by appealing to an appropriate (δ, η) pair, and deciding between these values is not a question that can be settled definitively by objective means, there is, in this view, little more to be said. A second response would be to insist that there is only one (δ, η) pair that is 'correct', and those who deviate from this value are making an ethical or methodological mistake. A third response is to accept that judgements of the kind that are required to settle on values for (δ, η) are invariably subject to good-faith disagreements, and to attempt to find tools that can aid policy analysis despite these disagreements.

The first of these responses is excessively pessimistic about the role of analytical and moral reasoning in debates about intertemporal social decision-making. The second response captures the views of many commentators on social discounting. For example, advocates of the 'impartiality' argument for $\delta = 0$ (including e.g. Pigou, Solow, Stern, and Gollier) seem to believe that any positive value of δ constitutes, in Harrod's words, 'a polite expression for rapacity and the conquest of reason by passion', and give short shrift to alternative views. The difficulty is that there is also a group of equally thoughtful and well-informed commentators (including e.g. Arrow, Samuelson and Nordhaus), for whom the 'excessive sacrifice', 'non-existence', or 'paternalism' critiques of $\delta = 0$ are persuasive arguments against this value. It is hard to believe that any of these commentators is guilty of any rudimentary conceptual errors. Rather, it seems more likely that each of them is grappling with the conflicting obligations to the self and to others, and finding some compromise between them that they find attractive based on their methodological proclivities and moral intuitions.

This brings us to the third response to the normative disagreements illustrated in Figure 3. Instead of asking what the 'correct' values of normative parameters are, we could ask how to proceed given persistent good-faith ethical disagreements. Is there a set of principles for aggregating across normative opinions that gives each its due, but leads to a fair compromise solution? That is the topic of this section.

Pivoting from seeking a single 'correct' normative theory to asking how to deal with persistent normative disagreements is not without precedent in the literature. In his magisterial work 'The Idea of Justice', Amartya Sen argues that there is often a plurality of (mutually inconsistent) ethical theories that may be brought to bear on a given issue, each of which may have something to recommend it. Importantly, this pluralist viewpoint does not mean that 'anything goes' ethically. Sen (2010, p. x) explains that "There is a need for reasoned argument, with oneself and with others, in dealing with conflicting claims, rather than for what can be called disengaged toleration, with the comfort of such a lazy resolution as: you are right in your community and I am right in mine. Reasoning and impartial scrutiny are essential. However, even the most vigorous of critical examination can still leave conflicting and competing arguments that are not eliminated by impartial scrutiny." The methods of social choice may be viewed as a rationality technology for adjudicating between those ethical claims that survive a process of public reasoned scrutiny. These methods may themselves be inherently plural, but this is not a reason for pessimism about their potential for dealing productively with normative disagreements. There is a difference in kind between debates about the rationality principles that social choice procedures should obey, and debates about basic ethical principles or normative parameters of a single planner's preferences. Moreover, as we shall see, some important conclusions may be reasonably robust across several different social choice approaches. With this background in mind, we now turn to some specific proposals that have been discussed in the recent literature.

4.1 Utilitarian aggregation and voting

In this section we discuss issues that arise when we apply standard tools from social choice theory and welfare economics to aggregate diverse normative theories of intertemporal social welfare. The primitives of the analysis are a set of N theories of intertemporal social welfare that boil down to numerically representable preference relations on (aggregate) consumption streams: $\{V_h^{(i)}(\mathbf{c})\}_{i=1...N}$. Here $V_h^{(i)}(\mathbf{c})$ denotes a cardinal representation of theory i's normative ranking of future consumption streams $\mathbf{c}$ at history h. We assume in addition that the $V_h^{(i)}(\cdot)$ are 'interpersonally comparable'.[35] We are interested in aggregation rules Ω_h which also deliver a cardinal preference relation W_h over future consumption streams at each history h: $W_h(\mathbf{c}) = \Omega_h(V_h^{(1)}(\mathbf{c}), V_h^{(2)}(\mathbf{c}), \ldots, V_h^{(N)}(\mathbf{c}))$. We will assume that the individual theories V_h^i are of the discounted utilitarian form (we again specialise to the deterministic case for simplicity):

$$V_h^{(i)}(\mathbf{c}) = \sum_{t=0}^{\infty} U_h^{(i)}(c_t)(1+\delta_h^{(i)})^{-t}.$$

In this section we specialize to 'utilitarian' aggregation rules in which Ω_h is a linear function. These rules are only utilitarian in form, not interpretation; recall that $V^{(i)}$ represents intertemporal welfare according to theory i, and not a measure of individual wellbeing. In this case, we have

$$W_h(\mathbf{c}) = \sum_{i=1}^{N} w_h^{(i)} \sum_{t=0}^{\infty} U_h^{(i)}(c_t)(1+\delta_h^{(i)})^{-t}. \tag{40}$$

To avoid non-generic cases, we assume there exist $i \neq j$ such that $w_h^{(i)} > 0, w_h^{(j)} > 0$ and $\delta_h^{(i)} \neq \delta_h^{(j)}$.

[35] Technically, we require social preferences to be unaffected by transformations of the form $V_h^{(i)}(\mathbf{c}) \rightarrow \alpha V_h^{(i)}(\mathbf{c}) + \beta_i$ for any $\alpha > 0, \beta_i \in \mathbb{R}$. The assumption of interpersonal comparability across theories is not innocuous – see Harsanyi (1955) for the definitive discussion of how this requirement may be interpreted.

4.1.1 Stationarity, time consistency, and time invariance in a collective context

An important feature of the preferences in (40) is that they no longer satisfy the stationarity axiom (28). This is obvious in the context of our discussion of the Koopmans axioms – there we showed that continuous preferences that satisfy the independence axiom are stationary if and only if they are discounted utilitarian. This never occurs in (40) as long as there are two $V_h^{(i)}$ with non-identical values of $\delta_h^{(i)}$ that have positive weight. Suppose that we now impose an additional assumption on (40) – time invariance. In this case we have $U_h^{(i)}(\cdot) = U^{(i)}(\cdot), \delta_h^{(i)} = \delta^{(i)}, w_h^{(i)} = w^{(i)}$ for all histories h. Since Stationarity and Time Invariance together imply Time Consistency, we have

$$\text{NOT Stationary} \Rightarrow \text{NOT Time Invariant OR NOT Time Consistent.}$$

As we have assumed time invariance, the fact that (40) is not stationary leads to a violation of time consistency. This is essentially the content of the results in Jackson & Yariv (2015), who claim that continuous aggregation rules that respect unanimity (i.e., the Strong Pareto principle applied across the $V^{(i)}$), the independence axiom, and time consistency do not exist. One can find similar claims in a wide range of classic and contemporary literature (see e.g. Marglin, 1963; Feldstein, 1964; Adams et al., 2014). And yet, time consistent utilitarian aggregation *is* possible. Millner & Heal (2018b) show that the 'impossibility' of time consistent utilitarian aggregation relies critically on the assumption of time invariance. If we drop time invariance it is a trivial matter to find time consistent versions of (40). If we assume that the weights $w_h^{(i)}$ only depend on calendar time τ and not the full history of consumption, a necessary and sufficient condition for time consistency is:

$$w_\tau^{(i)} = \frac{w_0^{(i)}(1+\delta^{(i)})^{-\tau}}{\sum_j w_0^{(j)}(1+\delta^{(j)})^{-\tau}}. \tag{41}$$

where $w_0^{(i)} \geq 0$, $\sum_i w_0^{(i)} = 1$. Millner & Heal (2018b) provide a critical discussion of the role of time invariance in collective intertemporal choice, arguing that it is a normatively and descriptively problematic assumption when a fixed group of agents decide on resource allocation over time.[36]

Time invariance is however a more compelling modeling assumption if we are attempt-

[36] If we view the model as one of consumers' wellbeing the weights in 41 can be interpreted as requiring us to account for agents' total lifetime wellbeing when computing social welfare (Millner & Heal, 2018b). Calvo & Obstfeld (1988) come to a related conclusion in a model in which cohorts of agents are born in each instant and face some hazard rate of death that depends on their age.

ing to model conflicts between different groups of decision-makers, as might occur in intergenerational decision-making. In that case we can think of the time invariant version of (40) as a principle that each generation uses to aggregate the normative intertemporal preferences of its constituents. The core assumption in this approach is that the only thing that is relevant for decision making within a given generation is the degree of concern that *current* agents exhibit towards the future. The concerns of past or future generations are not accounted for explicitly; they are only relevant to the extent that they implicitly influence the normative judgements of current agents. If we assume that the distribution of normative views does not vary with time we end up with a model in which aggregate preferences over future consumption streams at all histories h are given by:

$$W_h(\mathbf{c}) = \sum_{t=0}^{\infty} \tilde{U}(c_t)(1+\tilde{\delta}_t)^{-t}, \tag{42}$$

where

$$\tilde{U}(c) = \sum_{i=1}^{N} w^{(i)} U^{(i)}(c) \tag{43}$$

$$\tilde{\delta}_t = \left(\sum_{i=1}^{N} w^{(i)} (1+\delta^{(i)})^{-t} \right)^{-\frac{1}{t}} - 1. \tag{44}$$

We begin our discussion of these preferences with an examination of their implications for the risk-free consumption discount rate. Our analysis has some similarities to Gollier & Zeckhauser (2005); Jouini et al. (2010); Heal & Millner (2014). After that we discuss some of the broader issues that arise when social preferences are time inconsistent.

4.1.2 Implications of utilitarian aggregation for discounting

From the aggregated utility function in (43), we can define the aggregate elasticity of marginal utility

$$\tilde{\eta}(c) = -c \frac{\sum_{i=1}^{N} w^{(i)} U^{(i)''}(c)}{\sum_{i=1}^{N} w^{(i)} U^{(i)'}(c)} = \frac{\sum_{i=1}^{N} w^{(i)} \eta^{(i)}(c) U^{(i)'}(c)}{\sum_{i=1}^{N} w^{(i)} U^{(i)'}(c)}. \tag{45}$$

Assuming that $U^{(i)\prime}(c) = c^{-\eta^{(i)}}$, and writing $c_t = c_0 e^{g_t t}$ we simplify further to find that

$$\tilde{\eta}(c_0 e^{g_t t}) = \frac{\sum_{i=1}^{N} w^{(i)} \eta^{(i)} e^{-\eta^{(i)} g_t t}}{\sum_{i=1}^{N} w^{(i)} e^{-\eta^{(i)} g_t t}}$$

Again taking the continuous time limit for analytical convenience, and focussing on a deterministic model for simplicity, we see from (25) that the risk-free consumption discount rate at maturity t in this model is given by,

$$\rho_t = \tilde{\delta}_t + \frac{g_t}{t} \int_0^t \tilde{\eta}(c_0 e^{\tau g_\tau}) d\tau, \tag{46}$$

where the continuous time version of the expression in (44) is

$$\tilde{\delta}_t = -\frac{1}{t} \ln \left(\sum_{i=1}^{N} w^{(i)} e^{-\delta^{(i)} t} \right).$$

Notice that

$$\frac{d}{dt} \tilde{\delta}_t < 0, \quad \lim_{t \to \infty} \tilde{\delta}_t = \min_i \delta^{(i)}.$$

Moreover, if the growth rate is constant, i.e. $g_t = g$ for all t, and defining $\tilde{\eta}_t(g) = \tilde{\eta}(c_0 e^{gt})$, it is straightforward to show that

$$\operatorname{sgn}\left(\frac{d}{dt} \tilde{\eta}_t(g) \right) = -\operatorname{sgn}(g), \quad \lim_{t \to \infty} \tilde{\eta}_t(g) = \begin{cases} \min_i \eta^{(i)} & g > 0 \\ \max_i \eta^{(i)} & g < 0. \end{cases}$$

The aggregate pure rate of social time preference $\tilde{\delta}_t$ is a generalized mean of the $\{\delta^{(i)}\}$, and declines with maturity to the lowest value in the infinite future. Similarly, the aggregate elasticity of marginal utility $\tilde{\eta}_t(g)$ is a weighted average of the $\{\eta^{(i)}\}$, with weights that vary with maturity in such a way that $\tilde{\eta}_t(g)$ declines with maturity if $g > 0$, or increases with maturity if $g < 0$. Putting these pieces together, it is easy to see that when consumption growth is constant, the risk-free rate ρ_t declines with maturity to a limiting value of $\min_i\{\delta^{(i)} + \eta^{(i)} g\}$.

We illustrate the maturity dependence of the utilitarian risk-free discount rate in Figure 4. In this figure we have calibrated the values of $\delta^{(i)}$ and $\eta^{(i)}$ to the survey data in Figure

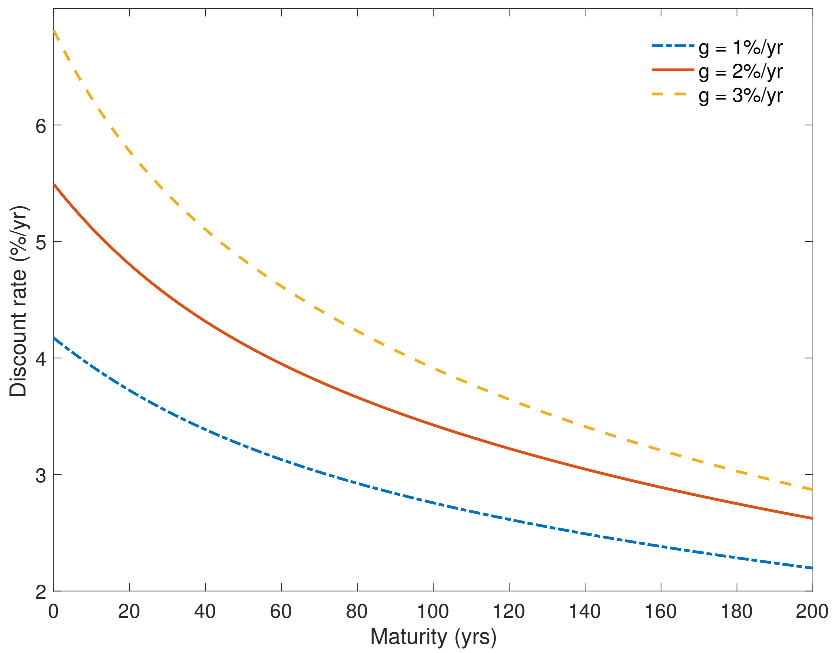

Figure 4: Risk-free consumption discount rate under utilitarian aggregation of the survey data in Figure 3, assuming constant deterministic consumption growth rate g.

3, chosen weights $w^{(i)} = \frac{\delta^{(i)}}{\sum_i \delta^{(i)}}$,[37] and used equations (44 –45) to compute the aggregate discount rate for different constant consumption growth rates g. The most striking difference between the discount rates in Figure 4 and those that follow from the 'single theory' Ramsey rule (26) is that now the risk-free rate has a declining term structure. This is a direct consequence of utilitarian aggregation across theories.

4.1.3 Time inconsistency, non-marginal policies, and voting

The fact that time invariant utilitarian aggregation leads to a time inconsistency problem might seem like a knockdown argument against this approach. Yet in the context of intergenerational choice there is no reason to insist on time consistency in general – we are modeling the collective normative preferences of successive groups of people, and it seems excessively demanding to require that these be aligned across groups who may be separated by decades or centuries. Time consistency is perhaps a reasonable constraint to impose on an idealized normative planner who accounts for the interests of all generations, into the

[37]This choice of weights ensures that agents with different values of $\delta^{(i)}$ contribute equally to the aggregate welfare function on constant utility paths, since $\int_0^\infty \bar{U} e^{-\delta^{(i)} t} dt = \frac{\bar{U}}{\delta^{(i)}}$.

indefinite past and future, in every time period. But the problem we are concerned with here focusses on a more limited, pragmatic, notion of normativity, i.e., one that emphasizes the normative views of those agents who have agency over *current* choices.

As a practical matter, time inconsistency poses no real difficulties for marginal policy analysis. Consumption paths are exogenous in this case, so the discount rates derived in the previous section can be applied at each point in time given an estimate of consumption growth. Things do however become more complex if we aim to use the utilitarian aggregation approach for non-marginal analysis. Thinking about intertemporal equilibrium in models of time inconsistent preferences leads us to treat the interaction between successive groups as a dynamic game. This follows since rational groups will choose the best available action given their beliefs about the actions of future groups, whose preferences are different from their own. We are in equilibrium if all groups best respond given their conjectures about the actions of future groups, and their conjectures are correct. This approach to handling time inconsistency goes back to Phelps & Pollak (1968), and is now a conventional part of the behavioral economist's toolkit (Laibson, 1997). This is one area where these tools may also have something to contribute to normative analysis.

Millner & Heal (2018a) present an analysis of a linear production economy in which consumption is chosen by a sequence of utilitarian 'committees' that are constituted afresh in each period. Each committee aggregates its members' views on the appropriate value of δ in each period, and chooses the value of consumption in that period only.[38] The equilibrium of this model can only be characterized analytically if committee members share a common logarithmic utility function. In this case the equilibrium consumption path is equivalent to the optimal path according to a single discounted utilitarian agent with pure rate of time preference given by the weighted harmonic mean of the individual's δs:

$$\hat{\delta} = \left(\sum_i w^{(i)} [\delta^{(i)}]^{-1} \right)^{-1}. \tag{47}$$

From the perspective of a committee in any period τ, the equilibrium summarized in (47) is inefficient. There exist feasible consumption paths that would increase its aggregate welfare measure W_τ. However, owing to the time inconsistency of committees' preferences, these paths are not implementable. Any future committee at time $\tau' > \tau$ can increase its welfare measure by deviating from the time τ committee's optimal plan. The time τ

[38] The focus on heterogeneity in δ is largely for technical reasons; a linear equilibrium of the resulting dynamic game exists in this case, but not in the case where there is heterogeneity in both δ and η.

committee knows this, anticipates the behavior of all future committees, and reacts optimally to this knowledge. Since all committees behave this way, the resulting equilibrium is inefficient, but fully rational. The fact that utilitarian committees choose inefficient consumption plans in equilibrium suggests that alternative methods for aggregating member's opinions could improve on utilitarian aggregation. The most natural alternative to consider is majoritarian voting.

Since consumption paths are high dimensional objects, standard voting theory suggests that it will be difficult to come up with workable voting rules. Indeed, this is the content of a second result in Jackson & Yariv (2015) – they show that under a 'local non-dictatorship' condition (i.e., whenever all but one agent prefers $\mathbf{c}$ to $\mathbf{c}'$, the voting rule does too), voting over consumption streams is generically intransitive. There is however a way around this negative result. Instead of considering a once-off vote over a full consumption stream, we can consider a sequence of votes over current consumption. Millner & Heal (2018a) observe that if each committee votes over their consumption choice only, but anticipates the outcome of the votes of future committees, then there is a well-defined equilibrium. The equilibrium consumption path in this case corresponds to the optimal path of the median committee member (see also Boylan & McKelvey, 1995).

One can also show that all committee members have single-peaked preferences over optimal discounted utilitarian consumption paths. Combining this fact with the fact that the utilitarian equilibrium is observationally equivalent to a discounted utilitarian optimal path, and employing the properties of single-peaked preferences (Downs, 1957), we see that a majority of agents will always prefer the voting equilibrium to the equilibrium implemented by the utilitarian committee, regardless of how preferences are aggregated in the utilitarian welfare function. Moreover, the voting method can dominate the utilitarian aggregation method according to the utilitarian committee's own welfare function. If utility functions are logarithmic, and the aggregation weights w_i are again chosen equitably (i.e., $w_i = \frac{\delta_i}{\sum_i \delta_i}$), a sufficient condition for this to occur is:

$$\left(\frac{1}{N}\sum_i [\delta^{(i)}]^{-1}\right)^{-1} < \delta_{\text{median}} < \frac{1}{N}\sum_i \delta^{(i)}.$$

That is, the harmonic mean must be less than the median, which must be less than the arithmetic mean. Since the harmonic mean is always less than the arithmetic mean, this is really only a constraint on the median – it is satisfied for most positively skewed distributions, and in particular for the data sample in Figure 3. This finding can also be shown to hold for other iso-elastic utility functions (Millner & Heal, 2018a).

4.1.4 Intergenerational Pareto

In Section 4.1.1 we observed that the time consistency of utilitarian aggregation can be restored if we abandon time invariance. In this section we consider a method for restoring time consistency with time invariance due to Feng & Ke (2018). These authors adopt the normative position of the 'idealised intergenerational planner' we mentioned above, i.e., a planner who accounts for the views of all agents in current and future generations at once. Their key insight is that applying a version of the Pareto property *across* generations, rather than just within the current generation, gives one much more freedom in the form of the aggregate preference relation. Indeed, there is so much freedom that one can construct examples where the aggregate intergenerational planner has exponential discounted utilitarian time preferences, and is hence time consistent.

We consider the simplest version of their model. Feng & Ke (2018) assume history independent preferences, so we can index all preferences by time τ. Within each generation individual i is assumed to have preferences over (lotteries over) future aggregate consumption given by

$$V_\tau^{(i)}(\mathbf{p}) = \sum_{s=\tau}^{T}(1+\delta_i)^{-(s-\tau)}u_i(p_s). \tag{48}$$

Here $\mathbf{p} = (p_0, p_{\tau+1}, \ldots, p_T)$ is a sequence of lotteries over aggregate consumption, δ_i is individual i's utility discount rate, $u_i(p_\tau)$ is individual i's assessment of the expected social utility of the (static) lottery p_τ, and T is the time horizon.[39] In addition, they assume that the planner's preferences $W_\tau(\mathbf{p})$ are exponential discounted utilitarian:

$$W_\tau(\mathbf{p}) = \sum_{s=\tau}^{T}(1+\delta_W)^{-(s-\tau)}u(p_s). \tag{49}$$

The main axiom their results rely on is:

- **Intergenerational Pareto**: Social preferences $W_\tau(\mathbf{p})$ satisfy intergenerational Pareto if in each period τ, and for any sequences of lotteries $\mathbf{p}, \mathbf{p}'$, $V_{\tau+s}^{(i)}(\mathbf{p}) \succeq V_{\tau+s}^{(i)}(i)(\mathbf{p}')$ for all $i, s \geq 0$ implies $W_\tau(\mathbf{p}) \succeq W_\tau(\mathbf{p}')$, and $V_{\tau+s}^{(i)}(\mathbf{p}) \succ V_{\tau+s}^{(i)}(\mathbf{p}')$ for all $i, s \geq 0$ implies $W_\tau(\mathbf{p}) \succ W_\tau(\mathbf{p}')$.

[39] The analysis in Feng & Ke (2018) assumes that T is finite, but their results have been extended to infinite T in Feng et al. (2020).

Unlike the unanimity (i.e., Strong Pareto) property that motivated (40), and acted across types within the current period, Intergenerational Pareto acts across all types in all future periods at once. Applying some well known results of Harsanyi (1955), it can be shown that $W_\tau(\mathbf{p})$ satisfies Intergenerational Pareto iff

$$W_\tau(\mathbf{p}) = \sum_{s=\tau}^{T} \sum_{i=1}^{N} \omega_{\tau,s}^{(i)} V_s^{(i)}(\mathbf{p}) \tag{50}$$

for some weights $\omega_{\tau,s}^{(i)} \geq 0$. Note that instead of the N weights $w_\tau^{(i)}$ that we had in (40), planner preferences in period τ now depend on $N \times (T + 1 - \tau)$ weights. This additional freedom in the representation is at the heart of the results in Feng & Ke (2018). The non-negativity of the weights $\omega_{\tau,s}^{(i)}$ in the representation of planner preferences places constraints on the possible values of the planners' discount factor δ_W. Suppose for example that all individuals have the same utility function, but their discount rates differ. In this case the planner's preferences satisfy intergenerational Pareto if and only if there are weights $\omega_s^{(i)} \geq 0$ and a constant utility discount factor $\delta_W > 0$ such that for t,

$$(1 + \delta_W)^{-(t-1)} = \sum_{s=1}^{t} \sum_{i=1}^{N} \omega_s^{(i)} (1 + \delta^{(i)})^{-(t-s)}.$$

It is relatively straightforward to show that this implies $\delta_W < \max_i \delta^{(i)}$. This is not much of a constraint on the planner's discount rate, but it turns out that the upper bound can be very considerably strengthen if individuals have different utility functions. The main result in Feng & Ke (2018) shows that if the individual utility functions $u^{(i)}(\mathbf{p})$ are linearly independent, then $W_\tau(\mathbf{p})$ respects intergenerational Pareto if and only if the aggregate planner's instantaneous utility function is a (strict) convex combination of individuals' utility functions, and

$$\delta_W < \min_i \delta^{(i)}.$$

Thus, when individuals' utility functions are linearly independent (surely the generic case), the upper bound on the planner's discount rate jumps from the maximum to the minimum of individuals' discount rates. These results show that if we impose discounted utilitarian preferences on the aggregate planner, she must exhibit more patience than any individual within a generation. From our discussion of the Koopmans axioms, we know that we could have arrived at the same conclusion if instead of imposing discounted utilitarian preferences directly, we imposed time consistency and time invariance on (50). Thus, extending our concept of which agents have standing for current decisions beyond the current generation, and imposing time consistency and time invariance, seems to require us to be extremely patient when making intergenerational decisions.

4.1.5 'Robust' aggregation weights

One potential operational difficulty with the utilitarian approach to aggregating intertemporal preferences is that it introduces an additional set of normative parameters, i.e., the aggregation weights $w_h^{(i)}$. In the analysis in Figure 3 we specified the aggregation weights by appealing to an intuitive fairness criterion – welfare functions with different values of δ should contribute equally to aggregate welfare when utility streams are constant and common to all theories. In Section 4.1.4 the weights were partially, but not totally, constrained by the requirement that the aggregate welfare function be discounted utilitarian. Is there some other, more systematic, way of specifying aggregation weights?

A partial affirmative answer is provided in a paper by Chambers & Echenique (2018). They consider the case where individuals share a common utility function but favour different utility discount rates. They then write down a set of plausible axioms – versions of Strong Pareto, continuity, interpersonal comparability, and an intergenerational inequality aversion axiom – under which

$$W(\mathbf{c}) = \min_{\vec{w}\in\Gamma} \sum_{i=1}^{N} w_i \frac{\delta^{(i)}}{1+\delta^{(i)}} \sum_{t=0}^{\infty} U(c_t)(1+\delta^{(i)})^{-t},$$

where $\vec{w} = (w_1, w_2, \ldots, w_N)$, $\sum_{i=1}^{N} w_i = 1$, and Γ is a convex set of possible weight vectors $\vec{w}$. Thus according to this rule we should evaluate utilitarian welfare using the most pessimistic aggregation weights in the set Γ.[40] Crucially, the most pessimistic weights will vary with the consumption sequence $\mathbf{c}$, and will depend on the set Γ of plausible weight vectors. Nevertheless, this approach provides a principled way of choosing the w_i that is robust to some 'uncertainty' about how different theories should be weighted.

4.2 Non-dogmatic social discounting

The previous subsection took an approach to the problem of aggregating diverse theories of intertemporal social welfare based on classical social choice theory. We hypothesise

[40]This result is analogous to the results in Gilboa & Schmeidler (1989) in the context of decision under ambiguity.

some aggregate preference relation that somehow combines each of the theories, place reasonable constraints on it, and see what follows from these constraints. Implicit in all these approaches is some 'meta-planner' whose authority is acknowledged by devotees of all theories. All actors cede authority to this meta-planner, who makes a collective choice on their behalf. In this section we discuss an alternative approach to dealing with normative disagreements about intertemporal welfare functions. The model we discuss in this section dispenses with any meta-planner; agents retain complete sovereignty over their own normative judgements, and are free to choose all the normative parameters of their welfare functions idiosyncratically. We only require them to exhibit a certain openness of mind – advocates of each normative theory are required to admit the possibility of a future change of heart, and to form their current normative judgements with one eye on their possible future selves. Following Millner (2020), planners who do this will be called 'non-dogmatic'.

More formally, we once again assume a set of N history-independent theories of intertemporal social welfare, represented by the welfare functions $V_\tau^{(i)}(\mathbf{c})$. Non-dogmatic social time preferences can be written in the following recursive form:

$$V_\tau^i = F^i\left(c_\tau, V_{\tau+1}^1, \ldots, V_{\tau+1}^N, V_{\tau+2}^1, \ldots, V_{\tau+2}^N, \ldots\right),$$

where there exists a $t > 0$ such that the functions F^i are strictly increasing in $V_{\tau+t}^j$ for all $j = 1 \ldots N$. The interpretation of this restriction on preferences is as follows: each planner at time τ favors her own idiosyncratic theory of intertemporal social welfare, but admits the possibility that her normative views may change in the future, i.e., a future self may advocate one of the other plausible theories. Each planner is non-dogmatic – instead of imposing their current preferred theory on their future selves, they internalize the preferences of future selves. Current welfare depends directly on the welfare measures that future selves may advocate, and not just on future consumption values according to the current self's preferred theory. Finally, non-dogmatism is persistent: planners are always non-dogmatic. Internalization and persistence together yield a recursive preference system in which current preferences depend on future preferences, each of which is in turn recursively defined. Note that although non-dogmatic planners are required to admit the possibility of a future change of heart and internalize future preferences, the functions $F^i(\cdot)$ are idiosyncratic. Non-dogmatic planners are thus free to advocate their preferred theory of intertemporal social welfare unequivocally.

If we assume further that preferences are additively time separable the functions $F^i(\cdot)$

it can be shown that this implies that

$$V_\tau^i = U^i(c_\tau) + \sum_{s=1}^{\infty} \beta_s^i \sum_{j=1}^{N} w_s^{ij} V_{\tau+s}^j \tag{51}$$

where $\beta_s^i > 0, w_s^{ij} > 0$ for all $s = 1 \ldots \infty, i, j = 1 \ldots N$, and the intratemporal weights w_s^{ij} satisfy $\sum_{j=1}^{N} w_s^{ij} = 1$ for all i. This interdependent preference system defines time preferences that are complete on the set of bounded utility streams, and are increasing in all utilities, if the coefficients β_s^i satisfy $\max_i \sum_{s=1}^{\infty} \beta_s^i < 1$. We assume this condition from now on. The central result of Millner (2020) is as follows:

Theorem 2. *Assume that planners' preferences satisfy (51), and let ρ_s^i be the social discount rate at maturity s according to planner i. Then*

$$\lim_{s\to\infty} \rho_s^i = \lim_{s\to\infty} \rho_s^j$$

for all $i, j \in \{1, \ldots, N\}$.

In words, all non-dogmatic planners agree on the long-run social discount rate, despite arbitrary disagreements about the coefficients β_s^i, w_s^{ij} and utility functions $U^i(\cdot)$.[41]

Although this result may be surprising at first, it has an intuitive origin. Suppose that the planners only place positive weight on selves one time step ahead, i.e. $\beta_s^i = 0$ for $s > 1$. To understand these planners' attitudes to consumption at some future time $\tau + s$, notice that only planners at $\tau + s$ care directly about consumption in that period. Planners at time $\tau+s-1$ care directly about consumption in period $\tau+s-1$, but indirectly about consumption in period $\tau + s$ via a mixture of the preferences of the selves at time $\tau + s$. After s steps back to the present, planners' concerns about $\tau + s$ are mediated through s iterates of a mixing operator that blends their intertemporal weights, and acts on utilities at time $\tau + s$. As s becomes large, this mixing process converges, and all planners' utility weights decline at the same geometric rate. In addition, marginal utilities at large maturities are dominated by the planner whose marginal utility function decreases slowest for large (small) consumption values if asymptotic consumption growth is positive (negative). Thus, in the limit as $s \to \infty$, all planners agree on both the rate of decline of utility weights, and on the rate of decline of marginal utility, i.e., they agree on the social discount rate.

[41] Millner (2020) presents an explicit formula for the consensus long-run social discount rate, but this is not needed for our current purposes.

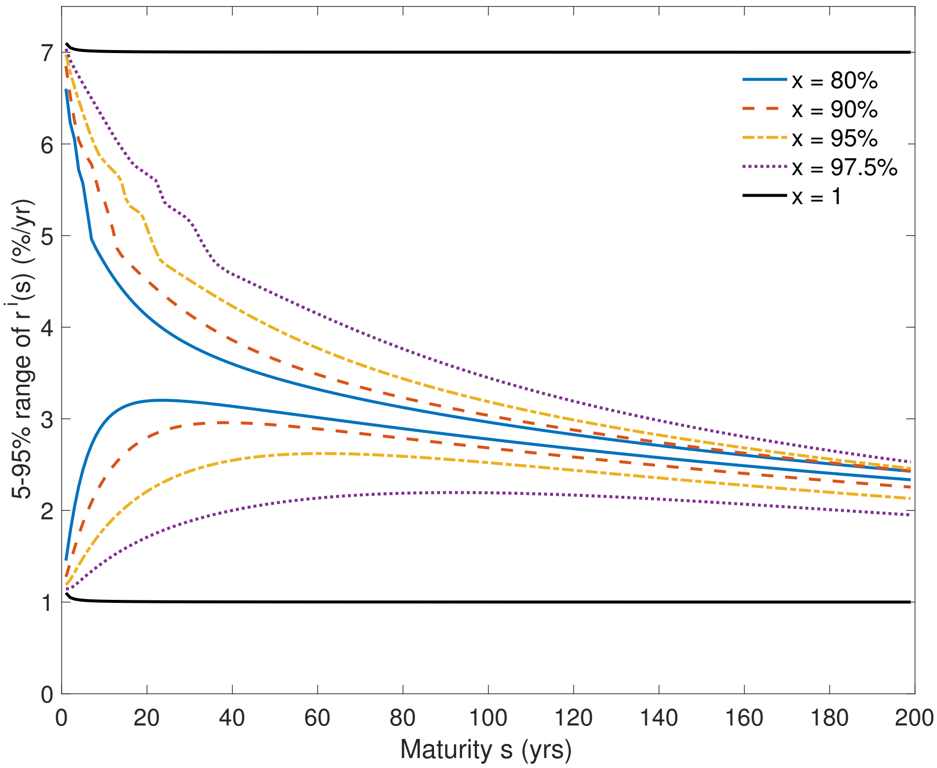

Figure 5: Effects of non-dogmatism on disagreements about discount rates. The figure depicts the 5-95% range of recommended values for the (risk-free) social discount rate, as a function of maturity (s), and the 'degree' of non-dogmatism (x). Reproduced from Millner (2020).

Figure 5 demonstrates the effects of non-dogmatism on disagreements about discount rates at different maturities. The analysis in this figure calibrates the model in (51) to the survey data in Figure 3, and calculates the distribution of recommended discount rates at each maturity as a function of a parameter x that measures the 'degree' of non-dogmatism. The value of x is the probability that each planner sticks with their preferred theory in the next year, so e.g. $x = 97.5\%$ corresponds to a change in views roughly once every 40 years on average. The figure shows that even with this mild degree of humility built into the preferences of diverse planners, non-dogmatism may still yield very substantial reductions in disagreement about how to value payoffs at medium to long maturities. As it is precisely at these long maturities where normative disagreements have the biggest effect on project evaluation, non-dogmatism achieves consensus where it is needed most.

5 Conclusion

There are not many easy take aways from our presentation of the issues involved in setting social discount rates – if anything our discussion shows that this task is in some ways more freighted with technical, ethical, and practical challenges than is commonly appreciated. Nevertheless, the needs of policy analysis trump any intellectual paralysis that these challenges may induce – we must grapple with them, for the sake of our descendants. We close with some brief recommendations that summarise our findings.

First, the market based approach is unlikely to be fit for purpose, especially when it comes to setting discount rates for payoffs that occur in the distant future. We showed that this approach entails substantive normative assumptions – about how to aggregate payoffs across individuals, the welfare significance of prices, and the consequences of heterogeneous beliefs – even in the highly optimistic case of perfect markets. Moreover, the perfect market assumption itself is almost surely untenable; market incompleteness and the presence of externalities are, in our view, serious challenges to this assumption in the context of long-term discounting. Nevertheless, if governments insist on using this approach despite its shortcomings, they should do so in a manner that is consistent with the basics of asset pricing. Discount rates should reflect *current* market prices (e.g. the current yield curve for index-linked government bonds), should very likely vary with maturity, and to the extent possible, should adjust for the correlations between project-specific risks and aggregate consumption risks.

Second, while the normative approach avoids the pitfalls of market perfectionism, it too suffers from serious implementation difficulties. While arguments for appropriate values of normative parameters (e.g. δ and η in the discounted utilitarian framework) are sometimes presented as *fait accomplis*, we believe that the debate on these matters within the community remains largely unresolved for a reason: this is the kind of thing about which reasonable people can reasonably disagree. Indeed, in some ways the debate about the appropriate values of normative parameters is arguably too narrow – should intertemporal social preferences even be expected discounted utilitarian, and if not which alternatives might be more attractive? There are many developments in decision theory and welfare economics – e.g. on the treatment of uncertainty – that are relevant to this question, but not often discussed in this context. The normative approach closes the door on one source of contention, but opens several windows on others.

Third, the zoo of potential options for making normative distributive judgments across time should not intimidate us. Rather, we should recognise the irreducible nature of disagreements on these issues, and seek methods for achieving consensus that respect individual views, but still provide practical tools for policy analysis. The methods of social choice – broadly conceived – hold promise for this endeavour, but there is much still to be done to flesh them out and apply them to this crucial problem. One quite robust finding of existing aggregation methods is that long-run risk-free discount rates should likely be very low. Indeed, three of the models we discussed – utilitarian aggregation, intergenerational Pareto, and non-dogmatic discounting – require that one or both of the normative parameters that enter the standard Ramsey rule should be as low as, or lower than, the lowest value that is recommended by any individual.[42] With so many arrows pointing in the same direction, the burden of proof required to overturn this recommendation is likely substantial.

References

D. Acemoglu (2008). *Introduction to Modern Economic Growth.* Princeton University Press.

A. Adams, et al. (2014). 'Consume now or later? Time inconsistency, collective choice, and revealed preference.'. *American Economic Review* **104**(12):4147–4183.

P. Antras, et al. (2017). 'Globalization, inequality, and welfare'. *Journal of International Economics* **108**:387–412.

K. Arrow, et al. (2013). 'Determining Benefits and Costs for Future Generations'. *Science* **341**(6144):349–350.

K. J. Arrow (1999). 'Discounting, morality, and gaming'. In P. R. Portney & J. P. Weyant (eds.), *Discounting and Intergenerational Equity.* Resources for the Future.

K. J. Arrow, et al. (2014). 'Should Governments Use a Declining Discount Rate in Project Analysis'. *Review of Environmental Economics and Policy* **8**(2):145–163.

K. J. Arrow, et al. (2003). 'Evaluating Projects and Assessing Sustainable Development in Imperfect Economies'. *Environmental and Resource Economics* **26**(4):647–685.

[42] This finding is anticipated in the prescient but somewhat controversial and informal analysis of Weitzman (1998). We have chosen to focus on rigorous frameworks rooted in formal welfare theory in this paper.

K. J. Arrow & M. Kurz (1970). *Public investment, the rate of return, and optimal fiscal policy*. RFF Press.

K. J. Arrow & R. C. Lind (1970). 'Uncertainty and the Evaluation of Public Investment Decisions'. *The American Economic Review* **60**(3):364–378.

G. B. Asheim (2010). 'Intergenerational Equity'. *Annual Review of Economics* **2**(1):197–222.

G. B. Asheim & W. Buchholz (2003). 'The malleability of undiscounted utilitarianism as a criterion for intergenerational justice'. *Economica* **70**:405–422.

D. K. Backus, et al. (2004). 'Exotic Preferences for Macroeconomsits'. *NBER Macroeconomics Annual* **19**(319-390).

L. Bakkensen & L. Barrage (2018). 'Flood risk belief heterogeneity and coastal home price dynamics: going under water?'. *NBER Working paper no. 23854* .

R. Bansal & A. Yaron (2005). 'Risks for the long run: A potential resolution of asset pricing puzzles'. *Journal of Finance* **59**(4):1481–1509.

L. Barrage (2018). 'Be careful what you calibrate for: Social discounting in general equilibrium'. *Journal of Public Economics* **160**:33–49.

W. J. Baumol (1968). 'On the Social Rate of Discount'. *American Economic Review* **58**(4):788–802.

S. Benartzi & R. Thaler (2007). 'Heuristics and Biases in Retirement Savings Behavior'. *Journal of Economic Perspectives* **21**(3):81–104.

B. D. Bernheim (1989). 'Intergenerational altruism, dynastic equilibria, and social welfare'. *Review of Economic Studies* **56**:119–128.

C. Blackorby & D. Donaldson (1990). 'A review article: The case against the use of the sum of compensating variations in cost-benefit analysis'. *Canadian Journal of Economics* **23**(3):471–494.

R. T. Boylan & R. D. McKelvey (1995). 'Voting over Economic Plans'. *The American Economic Review* **85**(4):860–871.

D. F. Bradford (1975). 'Constraints on Government Investment Opportunities and the Choice of Discount Rate'. *American Economic Review* **65**(5):887–899.

J. Broome (2014). *Climate Matters: Ethics in a Warming World*. W. W. Norton & Company, New York, NY London.

M. K. Brunnermeier, et al. (2014). 'A welfare criterion for models with distorted beliefs'. *Quarterly Journal of Economics* **129**(4):1753–1797.

G. A. Calvo & M. Obstfeld (1988). 'Optimal Time-Consistent Fiscal Policy with Finite Lifetimes'. *Econometrica* **56**(2):411–432.

J. Y. Campbell & J. H. Cochrane (1999). 'By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behaviour'. *Journal of Political Economy* **107**(2):205–251.

A. Caplin & J. Leahy (2004). 'The Social Discount Rate'. *Journal of Political Economy* **112**(6):1257–1268.

C. P. Chambers & F. Echenique (2018). 'On Multiple Discount Rates'. *Econometrica* **86**(4):1325–1346.

J. Cohen, et al. (2020). 'Measuring time preferences'. *Journal of Economic Literature* **58**(2).

F. Collard, et al. (2018). 'Ambiguity and the historical equity premium'. *Quantitative Economics* **9**(2):945–993.

M. L. Cropper, et al. (2014). 'Declining Discount Rates'. *American Economic Review: Papers & Proceedings* **104**(5):538–543.

P. Dasgupta (2008). 'Discounting climate change'. *Journal of Risk and Uncertainty* **37**(2):141–169.

P. Dasgupta, et al. (1972). *Guidelines for project evaluation*. United Nations, New York.

P. A. Diamond (1965). 'The Evaluation of Infinite Utility Streams'. *Econometrica* **33**(1):170–177.

A. Downs (1957). *An economic theory of democracy*. New York: Harper Collins.

J. Dreze & N. Stern (1987). 'The theory of cost-benefit analysis'. In A. Auerbach & M. Feldstein (eds.), *Handbook of Public Economics, vol. II*. North-Holland.

M. Drupp, et al. (2018). 'Discounting Disentangled'. *American Economic Journal: Economic Policy* **10**(4):109–134.

E. F. Fama & K. R. French (2004). 'The Capital Asset Pricing Model: Theory and Evidence'. *Journal of Economic Perspectives* **18**(3):25–46.

E. Farhi & I. Werning (2007). 'Inequality and Social Discounting'. *Journal of Political Economy* **115**(3):365–402.

M. S. Feldstein (1964). 'The Social Time Preference Discount Rate in Cost Benefit Analysis'. *The Economic Journal* **74**(294):360.

T. Feng & S. Ke (2018). 'Social Discounting and Intergenerational Pareto'. *Econometrica* **86**(5):1537–1567.

T. Feng, et al. (2020). 'Utilitarianism and Social Discounting with Countably Many Generations'. *Working paper, https://drive.google.com/file/d/1btQz_U4oi0xcJgKps-DqyYadnPv6t5MD/view* .

M. Fleurbaey & S. Zuber (2015). 'Discounting, risk and inequality: A general approach'. *Journal of Public Economics* **128**:34–49.

J. Geanakoplos (1990). 'An introduction to General Equilibrium with incomplete asset markets'. *Journal of Mathematical Economics* **19**:1–38.

J. D. Geanakoplos & H. M. Polemarchakis (1986). 'Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete'. In *Uncertainty, information, and communication: Essays in honor of Kenneth J. Arrow, Vol III*. Cambridge University Press.

P. Ghirardato (2002). 'Revisiting Savage in a Conditional World'. *Economic Theory* **20**(1):83–92.

J. Gierlinger & C. Gollier (2017). 'Do interest rates decline when there is ambiguity about growth?'. *Working Paper* .

S. Giglio, et al. (2015a). 'Very Long-Run Discount Rates'. *The Quarterly Journal of Economics* **130**(1):1–53.

S. Giglio, et al. (2015b). 'Climate Change and Long-Run Discount Rates: Evidence from Real Estate'. Working Paper 21767, National Bureau of Economic Research.

I. Gilboa, et al. (2014). 'No-Betting-Pareto Dominance'. *Econometrica* **82**(4):1405–1442.

I. Gilboa & D. Schmeidler (1989). 'Maxmin expected utility with non-unique prior'. *Journal of Mathematical Economics* **18**(2):141–153.

C. Gollier (2010). 'Ecological discounting'. *Journal of Economic Theory* **145**(2):812–829.

C. Gollier (2011). 'Le calcu du risque dans les investissements publics'. Tech. rep., Centre d'Analyse Strategique.

C. Gollier (2012). *Pricing the Planet's Future: The Economics of Discounting in an Uncertain World*. Princeton University Press.

C. Gollier (2018). *Ethical Asset Valuation and the Good Society*. Kenneth J. Arrow Lecture Series. Columbia University Press.

C. Gollier (2019). 'The welfare cost of ignoring the beta'. *Working Paper https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/by/gollier/forget4.pdf* .

C. Gollier & J. K. Hammitt (2014). 'The Long-Run Discount Rate Controversy'. *Annual Review of Resource Economics* **6**(1):273–295.

C. Gollier & R. Zeckhauser (2005). 'Aggregation of Heterogeneous Time Preferences'. *Journal of Political Economy* **113**(4):878–896.

W. Gorman (1953). 'Community Preference Fields'. *Econometrica* **21**(1):63–80.

H. Greaves (2017). 'Discounting for public policy: A survey'. *Economics and Philosophy* **33**(3):391–439.

B. Groom & C. Hepburn (2017). 'Reflections—Looking Back at Social Discounting Policy: The Influence of Papers, Presentations, Political Preconditions, and Personalities'. *Review of Environmental Economics and Policy* **11**(2):336–356.

B. Groom, et al. (2005). 'Declining Discount Rates: The Long and the Short of it'. *Environmental and Resource Economics* **32**(4):445–493.

O. Gueant, et al. (2012). 'Ecological intuition vesus economic "reason"'. *Journal of Public Economic Theory* **14**(2):245–272.

P. J. Hammond & H. Zank (2014). 'Rationality and Dynamic Consistency under Risk and Uncertainty'. In *Handbook of the Economics of Risk and Uncertainty*, vol. 1, pp. 41–97. Elsevier.

J. C. Harsanyi (1955). 'Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility'. *The Journal of Political Economy* **63**(4):309–321.

F. A. Hayek (1945). 'The Use of Knowledge in Society'. *The American Economic Review* **35**(4):519–530.

G. Heal (2005). 'Intertemporal Welfare Economics and the Environment'. In K. Mäler & J. R. Vincent (eds.), *Handbook of Environmental Economics*, vol. 3, pp. 1105–1145. Elsevier.

G. M. Heal & A. Millner (2014). 'Agreeing to disagree on climate policy'. *Proceedings of the National Academy of Sciences* **111**(10):3695–3698.

N. Hendren (2019). 'Efficient Welfare Weights'. *Working Paper, https://scholar.harvard.edu/files/hendren/files/eww.pdf* .

M. O. Jackson & L. Yariv (2015). 'Collective Dynamic Choice: The Necessity of Time Inconsistency'. *American Economic Journal: Microeconomics* **7**(4):150–78.

T. H. Johnsen & J. B. Donaldson (1985). 'The Structure of Intertemporal Preferences under Uncertainty and Time Consistent Plans'. *Econometrica* **53**(6):1451–1458.

E. Jouini, et al. (2010). 'Discounting and divergence of opinion'. *Journal of Economic Theory* **145**(2):830–859.

T. C. Koopmans (1960). 'Stationary Ordinal Utility and Impatience'. *Econometrica* **28**(2):287–309.

D. Laibson (1997). 'Golden Eggs and Hyperbolic Discounting'. *The Quarterly Journal of Economics* **112**(2):443–478.

R. C. Lind (ed.) (1982). *Discounting for Time and Risk in Energy Policy*. RFF Press.

I. M. D. Little & J. A. Mirrlees (1974). *Project appraisal and planning for developing countries*. Heinemann.

M. J. Machina (1989). 'Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty'. *Journal of Economic Literature* **27**(4):1622–1668.

S. A. Marglin (1963). 'The Social Rate of Discount and The Optimal Rate of Investment'. *The Quarterly Journal of Economics* **77**(1):95–111.

A. Mas-Colell, et al. (1995). *Microeconomic theory*. Oxford University Press, New York ; Oxford.

A. Millner (2013). 'On welfare frameworks and catastrophic climate risks'. *Journal of Environmental Economics and Management* **65**(2):310–325.

A. Millner (2020). 'Nondogmatic Social Discounting'. *American Economic Review* **110**(3):760–775.

A. Millner & G. Heal (2018a). 'Discounting by committee'. *Journal of Public Economics* **167**:91–104.

A. Millner & G. Heal (2018b). 'Time consistency and time Invariance in collective intertemporal choice'. *Journal of Economic Theory* **176**:158–169.

A. Millner & H. Ollivier (2016). 'Beliefs, Politics, and Environmental Policy'. *Review of Environmental Economics and Policy* **10**(2):226–244.

P. Mongin (2016). 'Spurious unanimity and the Pareto principle'. *Economics and Philosophy* **32**(3):511–532.

E. S. Phelps & R. A. Pollak (1968). 'On Second-Best National Saving and Game-Equilibrium Growth'. *The Review of Economic Studies* **35**(2):185–199.

P. R. Portney & J. P. Weyant (eds.) (1999). *Discounting and Intergenerational Equity*. RFF Press.

E. Quinet (2013). 'Cost benefit assessment of public investments'. Tech. rep., Commissariat general a la strategie et a la prospective.

F. P. Ramsey (1928). 'A Mathematical Theory of Saving'. *The Economic Journal* **38**(152):543–559.

K. L. Ricke & K. Caldeira (2014). 'Maximum warming occurs about one decade after a carbon dioxide emission'. *Environmental Research Letters* **9**(12):124002.

A. Sandmo & J. H. Dreze (1971). 'Discount Rates for Public Investment in Closed and Open Economies'. *Economica* **38**(152):395–412.

A. Sen (2010). *The idea of justice*. Penguin, London.

C. Severen, et al. (2018). 'A forward-looking Ricardian approach: Do land markets capilatize climate change forecasts?'. *Journal of Environmental Economics and Management* **89**:235–254.

H. Sidgwick (1874). *The method of ethics*. Macmillan and Co.

J. Skinner (2007). 'Are you sure you're saving enough for retirement?'. *Journal of Economic Perspectives* **21**(3):59–80.

S. Solomon, et al. (2009). 'Irreversible climate change due to carbon dioxide emissions'. *Proceedings of the National Academy of Sciences* **106**(6):1704–1709.

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

R. H. Strotz (1955). 'Myopia and Inconsistency in Dynamic Utility Maximization'. *The Review of Economic Studies* **23**(3):165.

C. P. Traeger (2011). 'Sustainability, limited substitutability, and non-constant social discount rates'. *Journal of Environmental Economics and Management* **62**(2):215–228.

C. P. Traeger (2014). 'Why uncertainty matters: discounting under intertemporal risk aversion and ambiguity'. *Economic Theory* **56**(3):627–664.

A. Tsyvinski & N. Werquin (2018). 'Generalized Compensation Principle'. *Working paper, https://economics.yale.edu/sites/default/files/files/Faculty/Tsyvinski/Compensation%2003%2022%202018.pdf* .

F. van der Ploeg (2020). 'Discounting and Climate Policy'. *CESifo Working Papers* (8441).

H. R. Varian (1982). 'The Nonparametric Approach to Demand Analysis'. *Econometrica* **50**(4):945–973.

H. R. Varian (1992). *Microeconomic Analysis*. W. W. Norton & Company, 3 edn.

M. L. Weitzman (1998). 'Why the Far-Distant Future Should Be Discounted at Its Lowest Possible Rate,'. *Journal of Environmental Economics and Management* **36**(3):201 – 208.