Non-paternalistic social discounting

By Antony Millner*

The long run social discount rate converts the long-term costs and benefits of public projects into present values. Its value depends on parameters of social preferences that capture social impatience and aversion to consumption inequalities. Experts disagree about these parameters, leading to substantial disagreements on the benefits of projects with long-term consequences. I show that if policy makers with diverse opinions on social preferences wish to avoid being paternalistic (i.e. imposing their opinions on others), they ***must*** *agree on these parameters. Addressing one common critique of normative social discounting (paternalism) could thus help to resolve another operational difficulty (disagreement).*

JEL: H43,D61,D90

Keywords: Social discount rate, heterogeneous preferences, non-paternalism, interdependent preferences, cost-benefit analysis.

The social discount rate (SDR) tells us how to convert the future consequences of public projects into present values, and is thus a critical input to

* Millner: London School of Economics and Political Science, Houghton St, London, WC2A 2AE, UK, a.millner@lse.ac.uk. I am grateful to Partha Dasgupta, Simone Galperti, Ben Groom, Geoff Heal, Derek Lemoine, Lucija Muehlenbachs, Frikk Nesje, Bruno Strulovici, the audiences of numerous seminars and conferences, four anonymous referees, and the editor, for helpful comments. I acknowledge financial support from the ESRC Centre for Climate Change Economics and Policy, and the Grantham Foundation for the Protection of the Environment.

public cost-benefit analysis. Small changes in its value can have an enormous effect on the net present values of public projects with long-run consequences such as infrastructure investments, climate change mitigation measures, and nuclear waste management (Arrow et al., 2013). Yet despite almost a century of economic research on intertemporal public decision-making (cf. Ramsey, 1928), opinion is still divided on how costs and benefits that occur more than a few decades in the future should be discounted.

Most estimates of social discount rates rely on one of two methodologies, commonly referred to as the 'positive' and 'normative' approaches (see e.g. Gollier and Hammitt, 2014). The difference between these approaches turns on whether one believes the economy to be at an intertemporal social optimum or not. The discount factor used to convert marginal changes in consumption in the future into present values should coincide with the social marginal rate of substitution between consumption today and consumption in the future. If the economy is at a social optimum, the marginal rate of substitution is equal to the marginal rate of transformation, which is in turn related to market interest rates in a competitive equilibrium (see e.g. Gollier, 2012). Proponents of the positive approach believe that social discount rates should be chosen to reflect risk-free market interest rates, thus implicitly assuming that the economy is at an optimum. An often claimed advantage of this approach is that it does not require judgements about social preferences, as interest rates already reflect the preferences of market participants. The positive approach is thus often seen as democratic and non-paternalistic, but relies on strong optimality assumptions.

By contrast, proponents of the normative approach are unwilling to accept that the economy is at a social optimum, citing e.g. incomplete markets, externalities, and divergences between individualistic preferences exhibited in the marketplace and ethically motivated social preferences.[1] Market imperfections are seen as particularly salient for long-run social discount rates. Gollier and Hammitt (2014) argue that 'the positive approach cannot be applied for time horizons exceeding 20 or 30 years, because there are no safe assets traded on markets with such large maturities', while Gollier (2012) observes that '[inefficiencies due to] the existence of overlapping generations imply that...the interest rate observed on financial markets should not be used...to evaluate public policies impacting several generations.'

Since the normative approach to social discounting assigns no special welfare significance to market interest rates it must work directly with social marginal rates of substitution. This requires a choice for the intertemporal social welfare function. In a seminal paper, Ramsey (1928) showed that if social preferences are discounted utilitarian, the social discount rate $r(s)$ at maturity s must be chosen to be

$$r(s) = \rho + \eta g(s) \tag{1}$$

where ρ is the Pure Rate of Social Time Preference (PRSTP), η is the elasticity of the marginal social utility of consumption, and $g(s)$ is the compound annual consumption growth rate between today and year s.[2] Implementing this formula in practice requires choices for the ethical parameters ρ and η,

[1]Arrow, Dasgupta and Mäler (2003), for example, state that 'using market observables to infer social welfare can be misleading in imperfect economies. That we may have to be explicit about welfare parameters...in order to estimate marginal rates of substitution in imperfect economies is not an argument for pretending that the economies in question are not imperfect after all.'

[2]The formula (1) assumes consumption growth is deterministic; it receives additional terms if it is uncertain (Gollier, 2012). η may depend on consumption in general, but is constant if social utilities are iso-elastic.

which measure social impatience and aversion to intertemporal consumption inequalities respectively, as well as a forecast of consumption growth. Two related criticisms are often leveled at this approach. First, informed commentators *disagree* on the appropriate values of ρ and η (Drupp et al., 2015), so whose opinion should count? Second, unlike the positive approach, the normative approach is *paternalistic*, since it requires policy makers to impose their opinion on the appropriate values of ρ and η on everyone, whether they share this opinion or not.[3]

This paper pursues a 'third way' for setting social discount rates. I develop a model of policy makers' opinions on intertemporal social welfare that combines features of both the positive and normative approaches. The model is rooted in the normative approach, and thus does not require us to assume that the economy is at a social optimum. However, in common with the positive approach, it takes the problem of paternalism seriously, and adopts non-paternalism as a desirable feature of social evaluation. Decision makers in the model may have any opinion on social utility functions, and may discount future social wellbeing in any way they deem appropriate. However, they may *not* impose their views on social preferences on others – the heterogeneity in opinions on social evaluations in the future is recognized, and respected. The central result of the paper is that addressing the paternalistic critique of the normative approach also helps to resolve disagreements about welfare parameters. Non-paternalistic policy makers *must* agree on the welfare parameters to use when calculating long run so-

[3]Disagreements about these welfare parameters largely explain the substantial differences between the climate policy recommendations of Stern (2007) and Nordhaus (2007), for example. Nordhaus views Stern's analysis, which is based on a normative approach to intertemporal social choice, as paternalistic, criticizing him for taking 'the lofty vantage point of the world social planner, perhaps stoking the dying embers of the British Empire.'

cial discount rates, despite arbitrary disagreements about the constituents of intertemporal social welfare functions. A calibration of the model to data on experts' opinions on the parameters ρ and η suggests that adopting non-paternalism as a principle of social evaluation could also substantially reduce disagreements about social discount rates at shorter maturities.

Formally, the model adapts existing models of non-paternalistic intertemporal preferences (Ray, 1987; Saez-Marti and Weibull, 2005; Galperti and Strulovici, 2017), extending them to account for heterogeneous preferences, and reinterpreting them in the context of social cost-benefit analysis. These models, and mine, correspond to a version of Bergstrom's (1999) model of interdependent preferences with an infinite sequence of forward looking agents. Unlike these papers, which focus on intergenerational altruism, I am concerned with opinion formation amongst a group of decision makers who are tasked with recommending a term structure of social discount rates for cost-benefit analysis. Weitzman (2001) and Freeman and Groom (2015) study the aggregation of heterogeneous opinions on social discount rates, and Gollier and Zeckhauser (2005) and Heal and Millner (2014) study utilitarian aggregation of time preferences. This paper is not concerned with aggregation, but rather focuses on the normative opinions of individual policy makers. Finally, the model bears a family resemblance to the *ad hoc* model of belief updating in DeGroot (1974). Unlike this work however, I focus on the formation of opinions on intertemporal social preferences, and the preferences I study are a direct consequence of requiring that social evaluations be non-paternalistic.

I. Model

A. A simple example

The essential features of the model can be illustrated in a simple example. Suppose that opinion on intertemporal social preferences is split into two camps, or types, indexed by $i \in \{1,2\}$. At time τ representatives of each camp are asked for their opinion on the term structure of social discount rates for cost-benefit analysis. That is, for each annual maturity s each type must specify a discount rate $r^i(s)$, from which their opinion on the net present value (NPV) of public projects with sequences of annual payoffs $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ will be computed thus:

$$NPV^i(\boldsymbol{\pi}) = \sum_{s=0}^{\infty} \pi_s e^{-r^i(s)s}. \tag{2}$$

For the sake of simplicity in this example assume that both types care only about society's current consumption utility and the discounted value of next year's social wellbeing. The two types *disagree* about the social utility of consumption and the discount factor on future wellbeing, and insist on computing next year's social wellbeing using their preferred social preferences. Denote type i at time τ's opinion on social wellbeing by V^i_τ, her opinion on the utility of consumption by $U^i(c)$, and her opinion on the appropriate discount factor by $\beta_i \in (0,1)$. Thus, types' opinions on social preferences can be written as

$$V^i_\tau = U^i(c_\tau) + \beta_i V^i_{\tau+1}, i \in \{1,2\}. \tag{3}$$

It is clear that these preferences have the following equivalent representation:

$$V_\tau^i = \sum_{s=0}^{\infty} (\beta_i)^s U^i(c_{\tau+s}). \tag{4}$$

If project payoffs π_s are small relative to consumption levels, type i should choose the discount factor $e^{-r^i(s)s}$ in (2) to reflect the current value of a marginal change in consumption s years from now (see e.g. Dasgupta, Sen and Marglin, 1972; Gollier, 2012):

$$e^{-r^i(s)s} = MRS_s^i = (\beta_i)^s \frac{(U^i)'(c_{\tau+s})}{(U^i)'(c_\tau)},$$

where MRS_s^i denotes type i's opinion on the Marginal Rate of Substitution between consumption at times $\tau + s$ and τ. Clearly, there is no possibility of the two types agreeing on any part of the term structure $r^i(s)$ in this case.

Now suppose that instead of insisting on imposing their own opinions on social preferences on others, the two types recognize the plurality of opinions that exists in the next year. If current types wish to avoid paternalism towards types that do not share their views next year their social preferences might take the following form:

$$V_\tau^1 = U^1(c_\tau) + \beta_1(w_1 V_{\tau+1}^1 + (1 - w_1)V_{\tau+1}^2), \tag{5}$$

$$V_\tau^2 = U^2(c_\tau) + \beta_2((1 - w_2)V_{\tau+1}^1 + w_2 V_{\tau+1}^2) \tag{6}$$

where $w_i \in (0, 1)$ is the weight type i assigns to type i opinions in the next year. If they are feeling democratic types might choose the w_i to match the weight of opinion on social preference i, but this is not required for the analysis. Note that types with such non-paternalistic social preferences can still express any view on social impatience through their choice of the discount factor β_i, and are also free to choose any social utility function they deem appropriate. So long as $w_i < 1$, these preferences are immune to a charge of paternalism towards types in the next year.

To analyze this coupled system of preferences, define

$$\vec{V}_\tau = \begin{pmatrix} V_\tau^1 \\ V_\tau^2 \end{pmatrix}; \vec{U}_\tau = \begin{pmatrix} U^1(c_\tau) \\ U^2(c_\tau) \end{pmatrix}; \mathbf{F} = \begin{pmatrix} \beta_1 w_1 & \beta_1(1-w_1) \\ \beta_2(1-w_2) & \beta_2 w_2 \end{pmatrix}$$

Then we can write the system (5–6) as:

$$\vec{V}_\tau = \vec{U}_\tau + \mathbf{F}\vec{V}_{\tau+1} = \sum_{s=0}^{\infty} \mathbf{F}^s \vec{U}_{\tau+s}. \tag{7}$$

Current types' attitudes to consumption changes in the distant future depend on the behaviour of $\mathbf{F}^s$ for large s. If preferences are non-paternalistic (i.e. $w_i < 1$), the matrix $\mathbf{F}$ is strictly positive. The Perron-Frobenious theorem (see e.g. Sternberg (2014)) then tells us that there is a matrix $\mathbf{A}$, with elements $a^{ij} > 0$, such that

$$\lim_{s\to\infty} \frac{\mathbf{F}^s}{\mu^s} = \mathbf{A}, \tag{8}$$

where $\mu \in (0,1)$ is the largest eigenvalue of $\mathbf{F}$. Thus when s is large both types' weights on future utilities are proportional to a common factor μ^s.

To understand the intuition for this result notice that current types at τ only care about utilities at future times $\tau+1, \tau+2, \ldots$ indirectly through a mixture $\mathbf{F}$ of the preferences of types at $\tau+1$. Types at $\tau+1$ in turn only care about utilities at times $\tau+2, \tau+3, \ldots$ through a mixture $\mathbf{F}$ of the preferences of types at time $\tau + 2$. Iterating, we see that current types' preferences over utilities at time $\tau + s$ depend on iterating the preferences of types at $\tau + s$ backwards to τ, passing through the preferences of types at times $\tau + s - 1, \tau + s - 2, \ldots, \tau + 1$. With each step back in this iteration the discount factors of different types are mixed by the matrix $\mathbf{F}$. As the number of mixing operations grows (i.e. as s increases), types' discount factors become homogenized. For large s the process of repeated mixing of discount factors converges, and both types' long run utility weights are proportional to a common discount factor μ^s which reflects the long run effect of repeatedly mixing preferences with $\mathbf{F}$.

Substituting (8) into (7) we see that according to type i, the marginal rate of substitution between consumption at τ and consumption at distant future times $\tau + s$ is

$$MRS_s^i = \frac{\mu^s[a^{i1}(U^1)'(c_{\tau+s}) + a^{i2}(U^2)'(c_{\tau+s})]}{(U^i)'(c_\tau)}. \tag{9}$$

With a few calculations we can simplify this expression further. Denote the long run growth rate of consumption by g, i.e. $c_{\tau+s} = e^{gs}c_\tau$ for large s. In addition, define the long run PRSTP $\rho = -\ln \mu$, and assume for simplicity that utility functions are iso-elastic, i.e. $(U^i)'(c) = c^{-\eta_i}$ where without loss of generality we take $\eta_2 \geq \eta_1 \geq 0$. Since $(U^i)'(c_{\tau+s}) \propto e^{-g\eta_i s}$ for large s, MRS_s^i is dominated by the smallest (largest) value of η if $g > 0$ ($g < 0$). Substituting these definitions and assumptions into (9), it is straightforward to see that for *both* types $i \in \{1, 2\}$,

$$e^{-r^i(s)s} = MRS_s^i \propto \begin{cases} e^{-(\rho+\eta_1 g)s} & \text{if } g > 0 \\ e^{-(\rho+\eta_2 g)s} & \text{if } g < 0 \end{cases} \tag{10}$$

when s is large. Thus, although types may have arbitrary disagreements about the welfare parameters β_i, η_i and the weights w_i, they both agree that

$$r^i(s) \to \begin{cases} \rho + \eta_1 g & \text{if } g > 0 \\ \rho + \eta_2 g & \text{if } g < 0 \end{cases}$$

when s is large. The interdependence between non-paternalistic preferences resolves disagreements about the welfare parameters that should be used to compute long run social discount rates.

B. *General model*

This finding can be extended to a more general model in which there is an arbitrary number of types with idiosyncratic opinions on how to compute intertemporal social preferences. Moreover, each type may care about social wellbeing in all future years, and not just in the next year as in the example above. Although the mathematics is more complex in the general model, the intuition for the main finding that non-paternalistic types agree on the parameters of the long-run social discount rate is similar to that in the example above.

Assume that there are N types, indexed by i, in each year τ. As before I denote type i at time τ's opinion on social wellbeing by V_τ^i, and let i's opinion on the social utility function be $U^i(c)$. Types' preferences are non-paternalistic, i.e. they internalize future opinions on social wellbeing. In addition, social preferences are assumed to be forward looking and time separable in utilities,[4] and the distribution of opinions on social preferences

[4]The wellbeing measure V_τ^i is time separable in utilities if and only if it is a linear function of $V_{\tau+s}^j$ for all $j = 1 \ldots N, s \in \mathbb{N}$. See Galperti and Strulovici (2017).

is assumed to be stationary over time. The most general model that meets these assumptions is:

$$V^i_\tau = U^i(c_\tau) + \sum_{s=1}^{\infty}\sum_{j=1}^{N} f^{ij}_s V^j_{\tau+s}, \tag{11}$$

where $f^{ij}_s \geq 0$ is the weight type i at time τ assigns to type j's conception of social wellbeing at time $\tau + s$.[5] Lemma 1 in the Appendix shows that (11) defines a unique set of preferences, which are non-decreasing in all utilities, if and only if

$$\max_i \left\{ \sum_{s=1}^{\infty}\sum_{j=1}^{N} f^{ij}_s \right\} < 1. \tag{12}$$

I assume this condition from now on. We will say that preferences are *fully non-paternalistic* if $f^{ij}_s > 0$ for all $i, j = 1 \ldots N, s = 1 \ldots \infty$.

As in the simple example above, V^i_τ has an equivalent representation in terms of sums of future utilities which may be determined by solving the infinite system of equations (11) (see appendix for details). We write the solution of this system as

$$V^i_\tau = \sum_{s=0}^{\infty}\sum_{j=1}^{N} a^{ij}_s U^j(c_{\tau+s}), \tag{13}$$

where a^{ij}_s is the weight type i in period τ gives to type j's preferred measure of social utility in period $\tau + s$. Type i's opinion on the social discount rate

[5]There is no difficulty allowing type i at time τ to place positive weight on V^j_τ for $j \neq i$ (see the appendix). I have ruled this out for simplicity by assuming types are only non-paternalistic towards the future. This yields a model that is a natural analogue of familiar forward-looking models of intertemporal choice.

at maturity s in this model is given by

$$r^i(s) = -\frac{1}{s}\ln\left(\frac{1}{(U^i)'(c_\tau)}\sum_{j=1}^{N} a_s^{ij}(U^j)'(c_{\tau+s})\right). \tag{14}$$

As before, the social discount rate is the compound annual rate of decline of the marginal rate of substitution between consumption at time $\tau + s$ and consumption in the present.

Define the elasticity of type i's marginal social utility function as

$$\eta^i(c) = -c\frac{(U^i)''(c)}{(U^i)'(c)}. \tag{15}$$

If $\eta^i(c)$ is uniformly larger than $\eta^j(c)$, type i is more averse to intertemporal consumption inequalities than type j. I assume that $\eta^i(c) \geq 0$, is bounded for all c, and that $\lim_{c\to\infty}\eta^i(c) > 0$; $\lim_{c\to 0}\eta^i(c) > 0$ for all i. In addition, define the long run growth rate of consumption to be

$$g = \lim_{s\to\infty}\frac{1}{s}\ln\left(\frac{c_{\tau+s}}{c_\tau}\right) \tag{16}$$

and let

$$\hat{\eta} = \begin{cases} \min_i\left\{\lim_{c\to\infty}\eta^i(c)\right\} & \text{if } g > 0 \\ \max_i\left\{\lim_{c\to 0}\eta^i(c)\right\} & \text{if } g < 0. \end{cases} \tag{17}$$

With these definitions in place the main result can be stated.

PROPOSITION 1: *If the social preferences (11) are fully non-paternalistic all types agree on the long run social discount rate:*

$$\forall i = 1 \ldots N,\ \lim_{s\to\infty} r^i(s) = \hat{\rho} + \hat{\eta}g. \tag{18}$$

$\hat{\rho} = -\lim_{M\to\infty} \ln \mu(M)$, *where* $\mu(M) \in (0,1)$ *is the largest eigenvalue of an* $NM \times NM$ *matrix constructed from the weights* $f_1^{ij}, \ldots, f_M^{ij}$.

Thus, despite arbitrary disagreements about how to discount future social wellbeings, how to compute social utilities, and how much weight to give to different types' opinions on social preferences, non-paternalistic types must agree on the welfare parameters that enter the long-run social discount rate.[6] The proof of this result provides details of how the sequence of eigenvalues $\mu(M)$ can be computed, and shows that $\lim_{M\to\infty} \mu(M)$ exists. It also shows that full non-paternalism is a substantially stronger condition than is required for the result to hold. It is sufficient for each type to place positive weight on some other type in some future period, in such a way that if we look far enough ahead, all types' preferences influence each other. Type i need not place positive weight on type j directly – they could influence one another through the preferences of several intermediate types, only some of which they care about directly. Full non-paternalism is however the normatively relevant case, as only then will types avoid paternalism across all types and all times.

Proposition 1 provides a simple characterization of the consensus long run elasticity of marginal social utility $\hat{\eta}$. The consensus long run PRSTP $\hat{\rho}$ is, however, a much more complex quantity, which depends on the full set of intertemporal weights f_s^{ij}. The appendix provides further discussion of $\hat{\rho}$, and some comparative statics results in special cases of the model. We will

[6]The formula (18) can be extended to the case where consumption growth is uncertain. For example, if consumption growth rates are i.i.d. and normally distributed with mean m and variance σ^2, $\eta^i(c) = \eta^i$ is constant, and $\eta^i < 2m/\sigma^2$ for all i (as is empirically the case for common calibrations of m, σ^2), one can show that $\lim_{s\to\infty} r^i(s) = \hat{\rho} + \min_i\{\eta^i m - \frac{1}{2}(\eta^i)^2\sigma^2\}$. I focus on the deterministic case for simplicity. Extending the model to account for uncertainty (see e.g. Gollier (2012)) is relatively straightforward, and will not be pursued here.

content ourselves with describing two intuitive properties of $\hat{\rho}$ here.

PROPOSITION 2: *1) $\hat{\rho}$ is decreasing in f_s^{ij} for all i, j, s.*

2) Suppose that

$$f_s^{ij} = f_s^{ij}(\epsilon) = \begin{cases} f_s^{ii} & j = i \\ h_s^{ij}(\epsilon) & j \neq i. \end{cases}$$

where the functions $h_s^{ij}(\epsilon)$ are continuous, $h_s^{ij}(\epsilon) > 0$ for $\epsilon > 0$, and $h_s^{ij}(0) = 0$. Let $\hat{\rho}_i$ be type i's idiosyncratic long-run PRSTP when $\epsilon = 0$, and let $\hat{\rho}(\epsilon)$ be the consensus long-run PRSTP when $\epsilon > 0$. Then

$$\lim_{\epsilon \to 0^+} \hat{\rho}(\epsilon) = \min_i \hat{\rho}_i. \tag{19}$$

The first part of the proposition is intuitive – any increase in f_s^{ij} increases the weight type i places on future wellbeings. Since all types' preferences depend on type i's preferences, all types are less impatient if f_s^{ij} increases. Thus the consensus long-run PRSTP decreases if f_s^{ij} increases. The second part of the proposition shows that if all types assign arbitrarily small, but positive, weight to the opinions of other types, they should agree to discount distant utilities at the rate that is the *lowest* of all of their paternalistic PRSTPs. To understand the intuition for this finding, note that although type i places arbitrarily small weight on preferences that do not coincide with her own as $\epsilon \to 0$, each type's preferences still enter into her social evaluation V_τ^i for all $\epsilon > 0$. When $\epsilon = 0$ type j's paternalistic weights on future utilities decline like $e^{-\hat{\rho}^j s}$ as $s \to \infty$. Thus the type with the lowest value of $\hat{\rho}^j$ will place exponentially more weight on distant future utilities than any more impatient type as $s \to \infty$ when $\epsilon = 0$. Since the most patient type's preferences are part of each type's preferences for $\epsilon > 0$, continuity of preferences in ϵ requires that the consensus long run PRSTP is given by the most patient type's paternalistic PRSTP as $\epsilon \to 0$.

Part 2 of Proposition 2 has something of the flavour of related findings on the aggregation of opinions on uncertain interest rates (Weitzman, 2001; Freeman and Groom, 2015), and on the aggregation of pure time preferences in standard paternalistic models of collective intertemporal choice (e.g. Gollier and Zeckhauser, 2005). In each of these cases averaging over a distribution of discount factors leads to a 'certainty equivalent' discount rate, or a representative discount rate, that declines to the lowest rate as the time horizon tends to infinity. Proposition 2 however differs from these results as it pertains to the preferences of each individual type, rather than an average across preferences or real discount rates. *Each type* should want to discount distant future utilities using the discount rate (19) if she is minimally non-paternalistic. There is no possibility of disagreement about the long run.[7]

[7]Gollier and Zeckhauser (2005) demonstrate that social planners that aggregate the time preferences of a group of individuals in a utilitarian manner always discount distant future utilities using the PRSTP of the most patient individual in the group, regardless of their choice of aggregation weights. A crucial difference between my model and theirs is that planners are only free to choose aggregation weights in their model – they have no personal view on the PRSTP or social utility functions. In my model however each type has her own views on social utility functions and how to discount future wellbeings. This makes the model better adapted to the question at hand, as it is these welfare judgements that experts actually disagree about, and not the aggregation weights in a utilitarian objective function. In addition, their model is still intertemporally paternalistic, as preferences are defined directly in terms of utilities rather than future social wellbeings. Finally, the result that the long-run consensus PRSTP is determined by the most patient type only occurs in a very special case of my model, showing that it yields results that are materially different from utilitarian preference aggregation in general.

II. Consequences for cost-benefit analysis

While Proposition 1 emphasizes the emergence of a consensus on the welfare parameters that enter the long run social discount rate when types are non-paternalistic, this result implies a more general phenomenon that has relevance for cost benefit analysis. As (2) shows, calculations of the net present value of public projects depend on the full term structure of social discount rates $r(s)$. Since non-paternalistic types' opinions on $r^i(s)$ converge completely as $s \to \infty$, their opinions on $r^i(s)$ must also exhibit partial convergence at finite maturities. Thus non-paternalism could reduce disagreement about project NPVs by acting through the entire term structure of the social discount rate. In this section I illustrate the effect of non-paternalism on cost-benefit analysis of public projects using some simple numerical examples. These examples also serve to demonstrate how quickly opinions on social discount rates can converge as a function of maturity.

To facilitate this analysis I will work with data on the distribution of economists' opinions on the appropriate values of the welfare parameters that enter social discounting formulae collected by Drupp et al. (2015). Although there is no deep reason why economists' opinions on welfare parameters should be seen as representative of the distribution of considered views, they do arguably have an advantage in understanding the quantitative implications of different recommendations for cost benefit analysis. Rawls (1971), in his notion of 'reflective equilibrium', argues that this is an essential feature of good normative reasoning. For our purposes these economists' opinions merely provide an interesting and informed distribution of views on these matters. Calibrating our model to their responses allows us to demonstrate how their opinions on social discount rates *might* change *if* they accepted non-paternalism as a desirable principle of social evaluation.

The Drupp et al. (2015) survey contains 173 complete responses from scholars who have published papers on social discounting. The 5-95% ranges of opinions on the PRSTP and elasticity of marginal social utility amongst the respondents were [0,3.85%/yr] and [0.2,3] respectively.[8] A full description of the survey data is provided in the online appendix. To map the survey data into my model I use a model of non-paternalistic social preferences in which the intertemporal wellbeing weights f_s^{ij} in (11) take the following form:

$$f_s^{ij} = \begin{cases} x\gamma\alpha_i^s & i = j \\ \frac{1-x}{N-1}\gamma\alpha_i^s & i \neq j \end{cases} \tag{20}$$

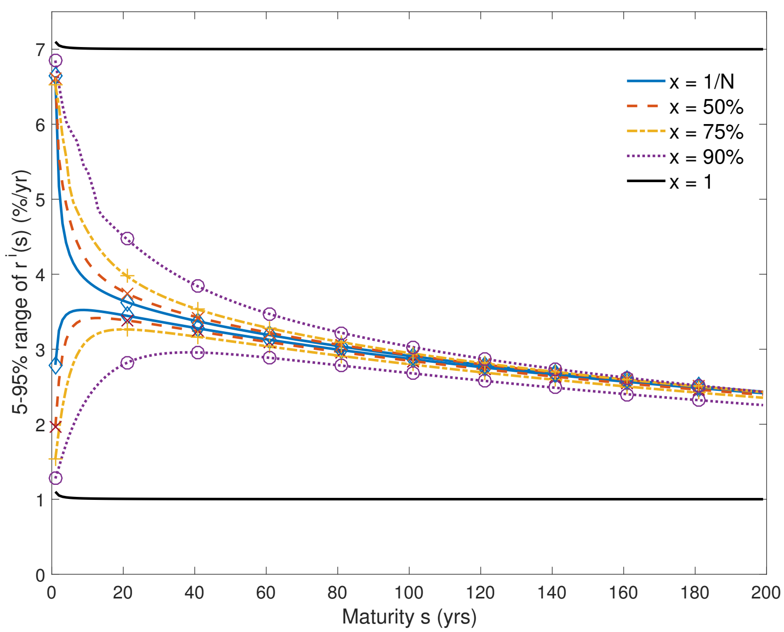
where $x \in [1/N, 1]$. When $x = 1/N$ in this model all types place equal weight on all future types' opinions in all future years (but may discount the future in any way they please). For $x > 1/N$ types give their own opinions a larger weight x in future periods, with the remaining weight $1 - x$ distributed equally between all other types. When $x = 1$, the model reduces to a set of N paternalistic social preferences, and there is no consensus on long-run social discount rates. I calibrate the values of the parameters γ, α_i so that when $x = 1$ types' preferences reduce to discounted exponential time preferences with a PRSTP that is consistent with respondents' reported values. Social utility functions $U^i(c)$ are taken to be iso-elastic, with the elasticity of marginal utility calibrated to respondents' reported opinions.

[8]There is no statistically significant correlation between the reported values of these parameters.

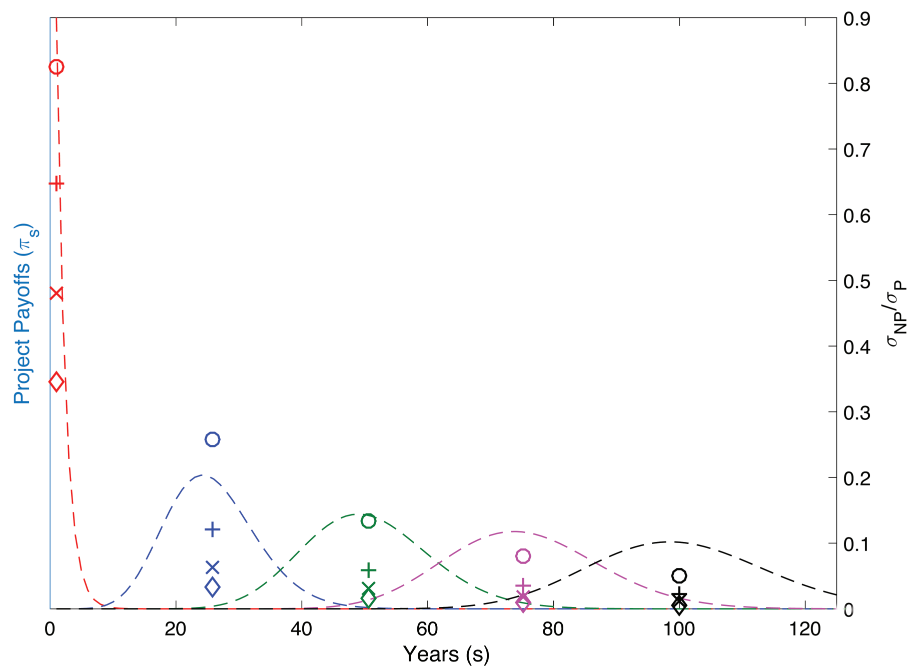
Given this calibration we can compute types' opinions on the term structure of the social discount rate $r^i(s)$ using equation (14), for different values of the parameter $x \in [1/N, 1]$. The values of types' utility weights a_s^{ij} in (14) are obtained by solving the preference system (11) given the calibrated specification of f_s^{ij} in (20).

Figure 1a depicts the results of this calibration exercise, assuming a constant consumption growth rate of 2%/yr. The figures show that if non-paternalism were adopted as a principle of social evaluation disagreements over the appropriate social discount rate $r(s)$ should reduce dramatically as maturity s increases. If types are democratic, and assign each type equal weight in future periods (i.e. $x = 1/N$), the 5-95% range of opinions about the discount rate to apply on payoffs that occur 50 years hence (i.e. $r^i(50)$) is $3.2 - 3.3$%/yr. The corresponding range of opinions in a paternalistic model (i.e. $x = 1$) is $1 - 7$%/yr. Thus if preferences are non-paternalistic and democratic the range of opinions on the social discount rate shrinks by a factor of 60 for maturities of 50 years. When types are less democratic, and assign large weight to their own views, the 5-95% range of opinions on $r^i(50)$ expands, but even for $x = 90\%$ it shrinks by more than a factor of 8 compared to the paternalistic case. For maturities longer than 50 years, the reduction in the range of opinions is even more dramatic.

Of course, non-paternalistic types still disagree about values of the social discount rate at short maturities. Nevertheless, non-paternalism may still substantially reduce the range of opinions on project NPVs, even for projects whose payoffs occur mainly in the relatively near term. This is illustrated in Figure 1b. The figure depicts five hypothetical project payoff sequences $\boldsymbol{\pi}$. Each project is assumed to cost 1 unit of consumption today, and to yield a



(a) Simulated 5-95% range for types' opinions on social discount rates $r^i(s)$. Curves marked with $\circ, +, \times, \diamond$ denote ranges when $x = 90\%, 75\%, 50\%, 1/N$ respectively in (20), while the solid black curve denotes the 5-95% range when $x = 1$, i.e. when preferences are paternalistic. Consumption growth is assumed to be a constant 2%/yr.



(b) Reduction in disagreement about project NPVs as a consequence of non-paternalism. Each curve in the figure denotes a hypothetical time sequence of project payoffs. The markers centered on each curve denote the values of σ_{NP}/σ_P, defined in (21), for this payoff sequence. $\circ, +, \times, \diamond$ denote values of σ_{NP}/σ_P when $x = 90\%, 75\%, 50\%, 1/N$ respectively in (20).

Figure 1. : Consequences of non-paternalism for cost-benefit analysis.

sequence of future benefits π_s, depicted by the dashed curves in the figure, where $\sum_{s=1}^{\infty}\pi_s = 2$. Thus the undiscounted NPV of each project is 1. To quantify the reduction in disagreement about project NPVs as a function of the parameter x in (20), let $\sigma(y^i)$ denote the standard deviation of data y^i, and compute the following ratio for each project $\boldsymbol{\pi}$:

$$\frac{\sigma_{NP}}{\sigma_P}(\boldsymbol{\pi};x) = \frac{\sigma(NPV^i(\boldsymbol{\pi};x))}{\sigma(NPV^i(\boldsymbol{\pi};1))} \tag{21}$$

where $NPV^i(\boldsymbol{\pi};x)$ is the net present value of $\boldsymbol{\pi}$ according to expert i when the weight placed on own opinions in (20) is x. The markers placed on top of each dashed payoff sequence in Figure 1b denote the values of σ_{NP}/σ_P for that project, for $x = 90\%, 75\%, 50\%, 1/N$. The figure shows that even for the project on the far left whose payoffs are strongly concentrated in the near term, the spread of opinions on NPVs is reduced by approximately two thirds relative to the paternalistic case if types are non-paternalistic and democratic (i.e. $x = 1/N$). Reductions in disagreements are more modest if types favour their own preferences ($x > 1/N$), but increase strongly as payoffs move further into the future. For the project on the far right, whose payoffs largely occur further than 60 years in future, disagreements are reduced by approximately a factor of 20 even if types assign 90% weight to their own preferences.

III. Conclusion

This paper introduced a model of policy makers' normative opinions on intertemporal social preferences. Decision makers are permitted any normative view on social impatience and aversion to intertemporal consumption inequalities, but they cannot impose their views on those who disagree with them – they are non-paternalistic. The key finding is that non-paternalism helps to resolve disagreements about the welfare parameters that determine long-run social discount rates. While the normative approach to social discounting is commonly seen as irredeemably paternalistic, this paper shows that not only may it be made non-paternalistic, but doing so generates consensus values of the welfare parameters that enter the long-run social discount rate.

Although non-paternalism is most effective at reducing disagreements about the net present value of 'long run' projects, the problems with the positive approach to social discounting are most acute for precisely these long maturities, as discussed in the introduction. The model developed here provides a normative method for setting long-run discount rates that retains an often claimed advantage of the positive approach (non-paternalism), but does not require the strong optimality assumptions it relies. Happily, the model works best at achieving consensus precisely where the positive approach is most problematic. Addressing the paternalistic critique of the normative approach to social discounting could thus help to resolve a long-standing issue with the application of this method: disagreements about the values of welfare parameters.

REFERENCES

Arrow, Kenneth J., Partha Dasgupta, and Karl-Göran Mäler. 2003. "Evaluating Projects and Assessing Sustainable Development in Imperfect Economies." *Environmental and Resource Economics*, 26(4): 647–685.

Arrow, K., M. Cropper, C. Gollier, B. Groom, G. Heal, R. Newell, W. Nordhaus, R. Pindyck, W. Pizer, P. Portney, T. Sterner,

R. S. J. Tol, and M. Weitzman. 2013. “Determining Benefits and Costs for Future Generations.” *Science*, 341(6144): 349–350.

Bergstrom, Theodore C. 1999. “Systems of Benevolent Utility Functions.” *Journal of Public Economic Theory*, 1(1): 71–100.

Dasgupta, Partha, Amartya Sen, and Stephen A Marglin. 1972. *Guidelines for project evaluation.* New York:United Nations.

DeGroot, Morris H. 1974. “Reaching a Consensus.” *Journal of the American Statistical Association*, 69(345): 118.

Drupp, Moritz, Mark Freeman, Ben Groom, and Frikk Nesje. 2015. “Discounting Disentangled.” *Grantham Research Institute on Climate Change and the Environment Working Paper 172.*

Freeman, Mark C., and Ben Groom. 2015. “Positively Gamma Discounting: Combining the Opinions of Experts on the Social Discount Rate.” *The Economic Journal*, 125(585): 1015–1024.

Friedland, Shmuel. 1981. “Convex spectral functions.” *Linear and Multilinear Algebra*, 9(4): 299–316.

Galperti, Simone, and Bruno Strulovici. 2017. “A Theory of Intergenerational Altruism.” *Econometrica (forthcoming).*

Gollier, Christian. 2012. *Pricing the Planet’s Future: The Economics of Discounting in an Uncertain World.* Princeton University Press.

Gollier, Christian, and James K. Hammitt. 2014. “The Long-Run Discount Rate Controversy.” *Annual Review of Resource Economics*, 6(1): 273–295.

Gollier, Christian, and Richard Zeckhauser. 2005. "Aggregation of Heterogeneous Time Preferences." *Journal of Political Economy*, 113(4): 878–896.

Heal, Geoffrey M., and Antony Millner. 2014. "Agreeing to disagree on climate policy." *Proceedings of the National Academy of Sciences*, 111(10): 3695–3698.

Nordhaus, William D. 2007. "A Review of the Stern Review on the Economics of Climate Change." *Journal of Economic Literature*, 45(3): 686–702.

Prelec, Drazen. 2004. "Decreasing Impatience: A Criterion for Non-stationary Time Preference and "Hyperbolic" Discounting." *Scandinavian Journal of Economics*, 106(3): 511–532.

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

Rawls, John. 1971. *A theory of justice.* Harvard University Press.

Ray, Debraj. 1987. "Nonpaternalistic intergenerational altruism." *Journal of Economic Theory*, 41(1): 112–132.

Saez-Marti, Maria, and Jörgen W. Weibull. 2005. "Discounting and altruism to future decision-makers." *Journal of Economic Theory*, 122(2): 254–266.

Sternberg, Shlomo. 2014. *Dynamical Systems.* Dover Publications.

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

Weitzman, Martin L. 2001. "Gamma Discounting." *The American Economic Review*, 91(1): 260–271.

FOR ONLINE PUBLICATION

LEMMA 1: *The preference system (11) defines a unique set of preferences, which are non-decreasing in all utilities, if only if*

$$\max_i \left\{ \sum_{s=1}^{\infty} \sum_{j=1}^{N} f_s^{ij} < 1 \right\}.$$

PROOF:

The system of preferences (11) can be written as a single matrix equation as follows:

$$\begin{pmatrix} V_\tau^1 \\ \vdots \\ V_\tau^N \\ V_{\tau+1}^1 \\ \vdots \\ V_{\tau+1}^N \\ \vdots \end{pmatrix} = \begin{pmatrix} U^1(c_\tau) \\ \vdots \\ U^N(c_\tau) \\ U^1(c_{\tau+1}) \\ \vdots \\ U^N(c_{\tau+1}) \\ \vdots \end{pmatrix} + \begin{pmatrix} \vec{0}_N & f_1^{11} & \dots & f_1^{1N} & f_2^{11} & \dots & f_2^{1N} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vec{0}_N & f_1^{N1} & \dots & f_1^{NN} & f_2^{N1} & \dots & f_2^{NN} & \dots \\ \vec{0}_N & \vec{0}_N & f_1^{11} & \dots & f_1^{1N} & f_2^{11} & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vec{0}_N & \vec{0}_N & f_1^{N1} & \dots & f_1^{NN} & f_2^{N1} & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} V_\tau^1 \\ \vdots \\ V_\tau^N \\ V_{\tau+1}^1 \\ \vdots \\ V_{\tau+1}^N \\ \vdots \end{pmatrix}$$

where $\vec{0}_N$ is an $1 \times N$ vector of zeros. Letting $\vec{X}_\tau$ denote the vector on the left hand side of this expression, $\mathbf{\Lambda}$ the infinite dimensional square matrix on the right hand side, and $\vec{U}_\tau$ denote the vector of Us on the right hand side, we have

$$\vec{X}_\tau = \vec{U}_\tau + \mathbf{\Lambda}\vec{X}_\tau$$
$$\Rightarrow \vec{X}_\tau = (\mathbf{1}_\infty - \mathbf{\Lambda})^{-1}\vec{U}_\tau,$$

where $\mathbf{1}_\infty$ is the infinite dimensional identity matrix.

In general infinite dimensional matrices do not have unique inverses. However, Lemma 1 in Bergstrom (1999) shows that $(\mathbf{1}_\infty - \mathbf{\Lambda})^{-1}$ exists, is unique, and has non-negative elements if and only if $\mathbf{1}_\infty - \mathbf{\Lambda}$ is a dominant diagonal matrix. A matrix $\mathbf{B}$ is dominant diagonal iff its elements B_{ij} satisfy $|B_{ii}| > \sum_{j\neq i} |B_{ij}|$ for all i. Thus, $\mathbf{1}_\infty - \mathbf{\Lambda}$ is dominant diagonal iff $\sum_{s=1}^{\infty} \sum_{j=1}^{N} f_s^{ij} < 1$ for all i.

Proof of Proposition 1

We prove a more general version of the result in Proposition 1. The proof has two main steps. First we find conditions under which all types' utility weights a_s^{ij} are proportional to a common discount factor $\hat{\mu}^s$ for large s. We then show that when these conditions are satisfied all types opinions on the long-run social discount rate will converge.

STEP 1:

Begin by defining the sequence of $N \times N$ matrices

$$\mathbf{F}_s := \begin{pmatrix} f_s^{11} & f_s^{12} & \dots & f_s^{1N} \\ f_s^{21} & f_s^{22} & \dots & f_s^{2N} \\ \vdots & \vdots & \vdots & \vdots \\ f_s^{N1} & f_s^{N2} & \dots & f_s^{NN} \end{pmatrix} \tag{A1}$$

and the sequences of $N \times 1$ vectors

$$\vec{V}_\tau = \begin{pmatrix} V_\tau^1 \\ V_\tau^2 \\ \vdots \\ V_\tau^N \end{pmatrix}, \quad \vec{U}_\tau = \begin{pmatrix} U^1(c_\tau) \\ U^2(c_\tau) \\ \vdots \\ U^N(c_\tau) \end{pmatrix}. \tag{A2}$$

Our general model (11) can be written as:

$$\vec{V}_\tau = \vec{U}_\tau + \sum_{s=1}^{\infty} \mathbf{F}_s \vec{V}_{\tau+s}. \tag{A3}$$

We seek an equivalent representation of this system of the form

$$\vec{V}_\tau := \sum_{s=0}^{\infty} \mathbf{A}_s \vec{U}_{\tau+s}, \tag{A4}$$

where $\mathbf{A}_s$ is a sequence of $N \times N$ matrices of the form,

$$\mathbf{A}_s := \begin{pmatrix} a_s^{11} & a_s^{12} & \dots & a_s^{1N} \\ a_s^{21} & a_s^{22} & \dots & a_s^{2N} \\ \vdots & \vdots & \vdots & \vdots \\ a_s^{N1} & a_s^{N2} & \dots & a_s^{NN} \end{pmatrix} \tag{A5}$$

where a_s^{ij} is the weight type i at time τ assigns to consumption utility according to type j at time $\tau + s$, i.e. $U^j(c_{\tau+s})$.

We now prove the following:

PROPOSITION 3: *Assume that the condition (12) is satisfied, and that $f_s^{ii} > 0$ for all $i = 1 \dots N$, $s = 1 \dots \infty$. Construct a directed graph G with N nodes labelled $1, 2, \dots, N$. Draw an edge from node i to node $j \neq i$ iff $f_s^{ij} > 0$ for at least one $s \geq 1$. If G contains a directed cycle of length N, then there exists a $\hat{\mu} \in (0,1)$ such that*

$$\lim_{s\to\infty} \frac{a_s^{ij}}{\hat{\mu}^s} = K_{ij} > 0$$

where the K_{ij} are finite constants.

Notice that if preferences are fully non-paternalistic the graph G in the statement of this proposition is complete (i.e. all edges exist), and the directed cycle condition is satisfied. However, the directed cycle condition itself is considerably weaker than full non-paternalism.

PROOF:

Substitute (A4) into (A3) to find

$$\sum_{s=0}^{\infty} \mathbf{A}_s \vec{U}_{\tau+s} = \vec{U}_\tau + \sum_{p=1}^{\infty} \mathbf{F}_p \left(\sum_{q=0}^{\infty} \mathbf{A}_q \vec{U}_{\tau+p+q} \right) \tag{A6}$$

Equating coefficients of $\vec{U}_{\tau+s}$ in this expression, we see that $\mathbf{A}_s$ must satisfy

$$\mathbf{A}_0 = \mathbf{1}_N \tag{A7}$$

$$\mathbf{A}_s = \sum_{p=1}^{s} \mathbf{F}_p \mathbf{A}_{s-p} \text{ for } s > 0. \tag{A8}$$

where $\mathbf{1}_N$ is the $N \times N$ identity matrix. The solution of this recurrence relation determines the utility weights a_s^{ij}. It will be convenient to split this matrix recurrence relation into a set of N vector recurrence relations as follows. Let $\vec{A}_s^j$ be the j-th column vector of $\mathbf{A}_s$, i.e.

$$\vec{A}_s^j = \begin{pmatrix} a_s^{1j} \\ a_s^{2j} \\ \vdots \\ a_s^{Nj} \end{pmatrix}. \tag{A9}$$

Define $\vec{e}^j$ to be the unit vector with elements

$$(\vec{e}^j)_i = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{A10}$$

Then (A8) is equivalent to the N vector recurrence relations

$$\begin{aligned} \vec{A}_0^j &= \vec{e}^j \\ \vec{A}_s^j &= \sum_{p=1}^{s} \mathbf{F}_p \vec{A}_{s-p}^j \text{ for } s > 0. \end{aligned} \tag{A11}$$

for $j = 1 \ldots N$.

The proof now has the following steps. We consider finite order models, i.e. $\mathbf{F}_{M'} = 0$ for all M' greater than some finite M. We show that if a certain augmented matrix constructed from the matrices $\mathbf{F}_1, \ldots, \mathbf{F}_M$ is *primitive*, all types will have a common long-run utility discount factor. A square matrix $\mathbf{B}$ is primitive if there exists an integer $k > 0$ such that $\mathbf{B}^k > 0$. We then extend this result to infinite order models by taking an appropriate limit of finite order models. Finally, we show that primitivity of the required matrices in the infinite order case is ensured by the graph theoretic condition in the statement of the proposition.

Begin with the finite order case. Let $M = \max\{s | \exists i, j \; f_s^{ij} > 0\} < \infty$. In this case, for all $s > M$, (A11) reduces to

$$\vec{A}_s^j = \sum_{p=1}^{M} \mathbf{F}_p \vec{A}_{s-p}^j. \tag{A12}$$

Define the $NM \times NM$ matrix

$$\boldsymbol{\Phi}_M = \begin{pmatrix} \mathbf{F}_1 & \mathbf{F}_2 & \dots & \mathbf{F}_{M-1} & \mathbf{F}_M \\ \mathbf{1}_N & 0 & \dots & 0 & 0 \\ 0 & \mathbf{1}_N & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{1}_N & 0 \end{pmatrix} \tag{A13}$$

where $\mathbf{1}_N$ is the $N \times N$ identity matrix. In addition, define the 'stacked' vector

$$\vec{Y}^j_s = \begin{pmatrix} \vec{A}^j_s \\ \vec{A}^j_{s-1} \\ \vdots \\ \vec{A}^j_{s-M+1} \end{pmatrix} \tag{A14}$$

Then we can rewrite the Mth order recurrence (A12) as a first order recurrence as follows:

$$\begin{aligned} \vec{Y}^j_s &= \boldsymbol{\Phi}_M \vec{Y}^j_{s-1} \\ \Rightarrow \vec{Y}^j_{M+s} &= (\boldsymbol{\Phi}_M)^s \vec{Y}^j_M. \end{aligned} \tag{A15}$$

We now assume that $\boldsymbol{\Phi}_M$ is a primitive matrix. By the Perron-Frobenius theorem for primitive matrices (Sternberg, 2014), this implies

1) $\boldsymbol{\Phi}_M$ has a positive eigenvalue, which we label as $\mu(M)$.

2) All other eigenvalues of $\boldsymbol{\Phi}_M$ have complex modulus strictly less than $\mu(M)$.

3) There exists a matrix $\mathbf{C} > 0$ such that

$$\lim_{s\to\infty} \frac{\mathbf{\Phi}_M^s}{[\mu(M)]^s} = \mathbf{C}$$

4) $\mu(M)$ increases when any element of $\mathbf{\Phi}_M$ increases.

5)

$$\mu(M) < \max_i \sum_j \phi_{ij}. \tag{A16}$$

where ϕ_{ij} is the ijth element of $\mathbf{\Phi}_M$.

Since the first N elements of $\vec{Y}_s^j$ coincide with a_s^{ij}, the third of these conclusions implies that

$$\forall i, j, \lim_{s\to\infty} \frac{a_s^{ij}}{[\mu(M)]^s} = \mathbf{C}\vec{Y}_M^j > 0. \tag{A17}$$

To bound the value of $\mu(M)$, note that from point 5 of the Perron-Frobenius theorem in (A16), and the definition of $\mathbf{\Phi}_M$ in (A13), we have

$$\mu(M) < \max_i \left\{ \sum_{s=1}^{M} \sum_{j=1}^{N} f_s^{ij} \right\} \tag{A18}$$

Thus, if

$$\sum_{s=1}^{\infty} \sum_{j=1}^{N} f_s^{ij} < 1 \tag{A19}$$

for all i, $\mu(M) < 1$, and hence $\lim_{s\to\infty} a_s^{ij} = 0$. Thus (12) guarantees that the preferences (11) are complete (i.e. finite on bounded utility streams) for all finite M. This concludes the finite M case.

We now extend this result to the case of infinite M. Assume that there exists an $M' > 0$ such that the matrix $\mathbf{\Phi}_M$, defined in (A13), is primitive for all $M > M'$. For $M > M'$, define

$$\vec{V}_\tau(M) = \vec{U}_\tau + \sum_{s=1}^{M} \mathbf{F}_s \vec{V}_{\tau+s}(M) \tag{A20}$$

and let

$$\hat{\vec{V}}_\tau = \lim_{M\to\infty} \vec{V}_\tau(M). \tag{A21}$$

Define the equivalent representations of these preferences by

$$\vec{V}_\tau(M) = \sum_{s=0}^{\infty} \mathbf{A}_s(M)\vec{U}_{\tau+s} \tag{A22}$$

$$\hat{\vec{V}}_\tau = \sum_{s=0}^{\infty} \hat{\mathbf{A}}_s \vec{U}_{\tau+s} \tag{A23}$$

In addition, let $\mu(M)$ be the Perron-Frobenius eigenvalue of $\mathbf{\Phi}_M$. We begin by proving that:

LEMMA 2:

$$\hat{\mu} := \lim_{M\to\infty} \mu(M) \textit{ exists.} \tag{A24}$$

PROOF:

Consider the eigenvalue $\mu(M+1)$, where $M > M'$. This is the Perron-Frobenius eigenvalue of $\mathbf{\Phi}_{M+1}$. The M-th order preferences $\vec{V}_\tau(M)$ are equivalent to an $M+1$th order model, with $\mathbf{F}_{M+1} = 0$. The matrix $\mathbf{\Phi}_M$, which controls the asymptotic behavior of $\vec{V}_\tau(M)$ can thus be thought of as an $N \times (M+1)$ matrix, where the last M rows and columns are zeros. Call this matrix $\tilde{\mathbf{\Phi}}_{M+1}$. The matrix $\mathbf{\Phi}_{M+1}$, associated with the asymptotic behavior of $\vec{V}_\tau(M+1)$, has entries that are strictly larger than than those of $\tilde{\mathbf{\Phi}}_{M+1}$ in at least some elements. Thus, by point 4 in our statement of the Perron-Frobenius theorem, $\mu(M+1) > \mu(M)$. We also know that $\mu(M) < 1$ for all M. Since the sequence $\mu(M)$ is increasing and bounded above, the monotone convergence theorem implies that $\hat{\mu}$ exists.

We have thus proved that if the matrices $\mathbf{\Phi}_M$ are primitive for $M > M'$,

$$\lim_{M\to\infty} \lim_{s\to\infty} \frac{a_s^{ij}(M)}{a_s^{ij}(M)} = \lim_{M\to\infty} \mu(M) = \hat{\mu}. \tag{A25}$$

Note that since (A16) and (A19) are strict inequalities, $\hat{\mu} < 1$. We now wish to know whether it is also true that:

$$\lim_{s\to\infty} \lim_{M\to\infty} \frac{a_{s+1}^{ij}(M)}{a_s^{ij}(M)} = \hat{\mu}. \tag{A26}$$

That is, can we change the order of the limits in (A25)? For limit operations to be interchangeable we require the sequence of functions they operate on to be uniformly convergent. The functions in question here are $V_\tau^i(M)$ and $\hat{V}_\tau^i$, which we can think of as linear functions from the infinite dimensional space $\mathbb{R}^\infty \times \mathbb{R}^N = \{(\vec{U}_\tau, \vec{U}_{\tau+1}, \vec{U}_{\tau+2}, \ldots)\}$ to $\mathbb{R}$. If the sequence of functions $V_\tau^i(M)$ converges uniformly to $\hat{V}_\tau^i$ on any bounded subset of $\mathbb{R}^\infty \times \mathbb{R}^N$, then (A26) will be satisfied. We now prove a second lemma:

LEMMA 3: *Let B be a compact subset of $\mathbb{R}^\infty \times \mathbb{R}^N$, and assume that (12) is satisfied. Then $V_\tau^i(M)$ converges uniformly to $\hat{V}_\tau^i$ on B.*

PROOF:

Equation (A11) shows that for all $s \leq M$, $a^{ij}_{\tau+s}(M) = \hat{a}^{ij}_{\tau+s}$. Let $\bar{U} = \max_j\{\sup_s\{U^j(c_{\tau+s})\}\}$ be the largest component of any $\vec{U} \in B$. For any $\vec{U} \in B$,

$$\begin{aligned}\sup_{\vec{U}\in B}\left|V^i_\tau(M) - \hat{V}^i_\tau\right| &= \sup_{\vec{U}\in B}\left|\sum_{s=1}^{\infty}\sum_{j=1}^{N} a^{ij}_{\tau+M+s}(M)U^j(c_{\tau+M+s}) - \sum_{s=1}^{\infty}\sum_{j=1}^{N} \hat{a}^{ij}_{\tau+M+s}U^j(c_{\tau+M+s})\right| \\ &\leq \sum_{s=1}^{\infty}\sum_{j=1}^{N}\left[\left|a^{ij}_{\tau+M+s}(M)\right| + \left|\hat{a}^{ij}_{\tau+M+s}\right|\right]\bar{U}\end{aligned}$$

By Lemma 1, $\hat{\mu} < 1$ also implies $\mu(M) < 1$ for all M, so we know that $\lim_{M\to\infty} a^{ij}_{\tau+M+s}(M) = 0 = \lim_{M\to\infty} \hat{a}^{ij}_{\tau+M+s}$ for all i, j. Thus

$$\lim_{M\to\infty}\sup_{\vec{U}\in B}\left|V^i_\tau(M) - \hat{V}^i_\tau\right| = 0.$$

Hence $V^i_\tau(M)$ converges uniformly to $\hat{V}^i_\tau$.

This concludes the infinite order case.

The final step of the proof is to show that if the graph G, defined in the statement of the proposition, has a directed cycle of length N, then there exists an $M' > 0$ such that for all $M > M'$ the matrix $\mathbf{\Phi}_M$ is primitive. We demonstrate this using a graphical argument.

Consider an aribtrary $R \times R$ matrix B_{ij}, and form a directed graph $H(B)$ on nodes $1 \ldots R$, where there is an edge from node i to node j iff $B_{ij} > 0$. The matrix B_{ij} is primitive if there exists an integer $k \geq 1$ such that there is a path of length k from each node i to every other node j in $H(B)$. If $H(B)$ is *strongly connected*, i.e. there exists a path from every node to every other node, then a sufficient condition for B_{ij} to be primitive is if there exists at least one node that is connected to itself.

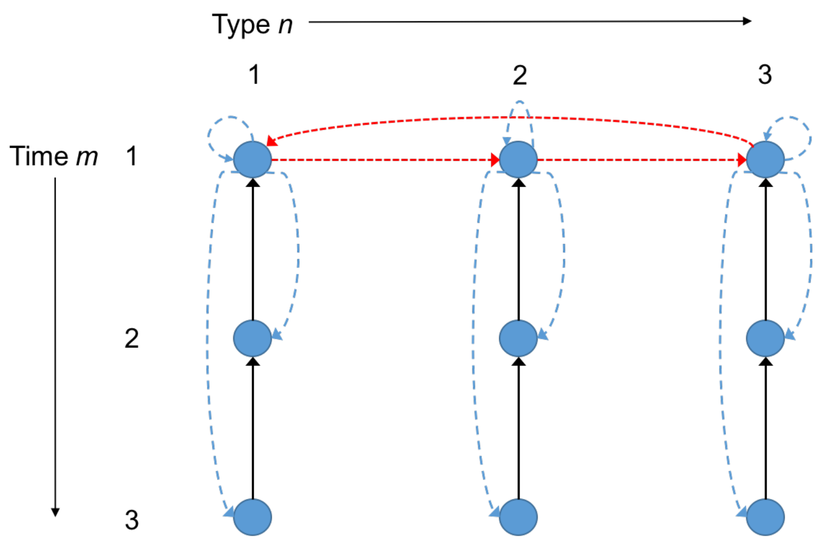


Figure A1. : The directed graph $H(\mathbf{\Phi}_3)$ associated with the matrix in our example. The vertical black edges arise from the identity matrices in the definition of $\mathbf{\Phi}_M$ (see (A13)). The dashed blue edges arise from $f_s^{ii} > 0$, and the dashed red edges from $f_1^{12}, f_1^{23}, f_1^{31} > 0$.

Now consider our $NM \times NM$ matrices $\mathbf{\Phi}_M$. To construct the directed graph $H(\mathbf{\Phi}_M)$ associated with $\mathbf{\Phi}_M$ in a convenient form, follow the following procedure: Construct an $M \times N$ grid of nodes (where N is the number of types), with node (m, n) representing type n at time $\tau + m$. For all $m > 1, n$, construct a directed edge from node (m, n) to node $(m - 1, n)$. In addition, construct a directed edge from node $(1, n)$ to node (m', n') if $f_{m'}^{nn'} > 0$.

As an example, take the case $M = N = 3$, i.e. a third order model with three types. In this case $\mathbf{\Phi}_M$ is a 9×9 matrix. Assume that $f_s^{ii} > 0$ for all $i, s = 1 \ldots 3$, that $f_1^{12}, f_1^{23}, f_1^{31} > 0$, and that $f_s^{ij} = 0$ otherwise. Figure A1 represents the directed graph associated with the matrix $\mathbf{\Phi}_3$ in this case.

Examination of the figure shows that since $f_s^{ii} > 0$, each of the 'column' subgraphs $\{(m, 1)\}, \{(m, 2)\}, \{(m, 3)\}, m = 1 \ldots 3$ is strongly connected.

Moreover, the cycle between columns (the red dashed edges) connects the columns to each other, and causes the entire graph to be strongly connected. Since each node in the first row is connected to itself, the matrix $\mathbf{\Phi}_3$ in this example is regular.

Returning to the general case, suppose that $f_s^{ii} > 0$ for all i and s. From the example in Figure A1 it is clear that this implies that for each fixed i the subgraph $\{(m, i)|m = 1 \ldots \infty\}$ is strongly connected, with each of the nodes $(1, i)$ connected to itself. Thus, if there is a directed cycle between all of the 'columns' of the graph $H(\mathbf{\Phi}_{M'})$ for some M', then for all $M > M'$, $H(\mathbf{\Phi}_M)$ is strongly connected, and contains nodes that are connected to themselves. Hence for all $M > M'$, $\mathbf{\Phi}_M$ is a primitive matrix. This concludes the proof.

STEP 2:

We now show that when the conditions of Proposition 3 are satisfied, all types will agree on the long run social discount rate, and we compute an explicit formula for this consensus discount rate.

Begin by defining

$$\hat{\rho} = -\ln \hat{\mu}.$$

When the conditions of Proposition 3 hold we know that

$$a_s^{ij} \sim K_{ij}(s) e^{-\hat{\rho} s} \tag{A27}$$

where $\sim$ denotes $s \to \infty$ asymptotic behaviour, and the multiplicative factors $K_{ij}(s)$ satisfy $\frac{1}{s} \lim_{s\to\infty} \ln K_{ij}(s) = 0$.

Now integrate the definition of $\eta^j(c)$ in (15) to find

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

Make the change of variables $x = c_\tau e^{gs'}$ in the integral in the exponent (recall that g is the long run consumption growth rate), and evaluate $(U^j)'(c)$ at $c = c_\tau e^{gs}$ to find

$$(U^j)'(c_\tau e^{gs}) = \exp\left(-g\int_0^s \eta^j(c_\tau e^{gs'})ds'\right)$$

Defining

$$\hat{\eta^j} = \begin{cases} \lim_{c\to\infty}\eta^j(c) & g>0 \\ \lim_{c\to 0}\eta^j(c) & g<0 \end{cases}$$

we see that the $s\to\infty$ asymptotic behaviour of marginal utility is given by

$$(U^j)'(c_\tau e^{gs}) \sim L_j(s)e^{-g\hat{\eta^j}s} \tag{A28}$$

where $\frac{1}{s}\lim_{s\to\infty}\ln L_j(s) = 0$. Combining (A27) and (A28), we find

$$\begin{aligned} r^i(s) &= -\frac{1}{s}\ln\left(\frac{1}{(U^i)'(c_\tau)}\sum_{j=1}^{N} a_s^{ij}(U^j)'(c_{\tau+s})\right) \\ &\sim -\frac{1}{s}\ln\left(\sum_j K_{ij}(s)L_j(s)e^{-\hat{\rho}s}e^{-\eta^j gs}\right) \\ &\sim \hat{\rho} - \frac{1}{s}\ln\left(\sum_j K_{ij}(s)L_j(s)e^{-\eta^j gs}\right) \end{aligned}$$

Define $\tilde{K}_{ij}(s) = K_{ij}(s)L_j(s)$, and let q be the index of the type with the lowest (highest) value of $\hat{\eta}^j$ when $g > 0$ ($g < 0$). Then

$$\sum_j K_{ij}(s)L_j(s)e^{-\eta^j gs} = \sum_j \tilde{K}_{ij}(s)e^{-\eta^j gs}$$
$$= \tilde{K}_{iq}(s)e^{-\eta^q gs}\left(1 + \sum_{j\neq q}\frac{\tilde{K}_{ij}(s)}{\tilde{K}_{iq}(s)}e^{-(\eta^j-\eta^q)gs}\right)$$

Since $\eta^j - \eta^q > 0$ for all $j \neq q$ when $g > 0$, and $\eta^j - \eta^q < 0$ for all $j \neq q$ when $g < 0$,

$$\sum_j K_{ij}(s)L_j(s)e^{-\eta^j gs} \sim \tilde{K}_{iq}(s)e^{-\hat{\eta}gs},$$

where $\hat{\eta}$ is given by (17). Thus

$$r^i(s) \sim \hat{\rho} - \frac{1}{s}\ln\left(\tilde{K}_{iq}(s)e^{-\hat{\eta}gs}\right)$$
$$\Rightarrow \lim_{s\to\infty} r^i(s) = \hat{\rho} + \hat{\eta}g.$$

Non-paternalism towards current decision-makers

The model developed in the main body of the paper assumes that non-paternalism is exclusively forward-looking. It is straightforward to extend the model to account for non-paternalism towards current decision-makers too.

Consider social preferences of the form

$$V_\tau^i = U^i(c_\tau) + \sum_{j=1}^N q^{ij}V_\tau^j + \sum_{s=1}^\infty\sum_{j=1}^N f_s^{ij}V_{\tau+s}^j,$$

where $q^{ii} = 0$ and $q^{ij} \geq 0$ for $j \neq i$. Let $\mathbf{Q}$ be the $N\times N$ matrix of coefficients q^{ij}, and let $\mathbf{1}_N$ be the $N \times N$ identity matrix. We can write this preference system using the vector notation of Proposition 3 as

$$\vec{V}_\tau = \vec{U}_\tau + \mathbf{Q}\vec{V}_\tau + \sum_{s=1}^{\infty} \mathbf{F}_s \vec{V}_{\tau+s}. \tag{B1}$$

If $(\mathbf{1}_N - \mathbf{Q})^{-1}$ exists and has positive entries, we find

$$\vec{V}_\tau = (\mathbf{1}_N - \mathbf{Q})^{-1}\vec{U}_\tau + \sum_{s=1}^{\infty}[(\mathbf{1}_N - \mathbf{Q})^{-1}\mathbf{F}_s]\vec{V}_{\tau+s}.$$

This system is of exactly the same form as the preferences studied in Proposition 3, and so all the results go through. A sufficient condition for $(\mathbf{1}_N - \mathbf{Q})^{-1}$ to exist and be positive is if $(\mathbf{1}_N - \mathbf{Q})$ is strictly dominant diagonal, i.e. $\max_i\{\sum_j q_{ij}\} < 1$.

A word of caution is in order however. In general, if $\max_i\{\sum_j q_{ij}\} < 1$ the row sums of $(\mathbf{1}_N - \mathbf{Q})^{-1}$ will exceed 1, so it is not guaranteed that the analogous condition to (12), which ensures that preferences are well defined, will always be satisfied in this model if $\max_i\{\sum_{s,j} f_s^{ij}\} < 1$. This condition must be modified to require that the row sums of $\sum_{s=1}^{\infty}(\mathbf{1}_N - \mathbf{Q})^{-1}\mathbf{F}_s$ not exceed 1 for any row.

I neglect $\mathbf{Q}$ in the body of the paper because it is difficult to interpret in the context of intertemporal social decision-making, and it is unclear how to calibrate its values in the empirical application in Section II.

Proof of Proposition 2

Part 1 of the proposition is immediate from point 4 in our statement of the Perron-Frobenius theorem in Proposition 3. Part 2 of the proposition follows from the fact that the eigenvalues of a matrix are continuous in its entries. Consider a set of N paternalistic models, in which each type assigns weight only to its own preferences in future periods. This set of models can be represented as a single model with N types where $f_s^{ij} = 0$ if $j \neq i$. As in the proof of Proposition 3, begin by considering a model of finite order M. Equation (A15) shows that the asymptotic behaviour of such a model can be described by first order difference equations of the form:

$$\vec{Y}_s^j = \mathbf{\Phi}_M^0 \vec{Y}_{s-1}^j. \tag{C1}$$

In this case however, the matrix $\mathbf{\Phi}_M^0$, defined in (A13), is reducible. The largest eigenvalue of $\mathbf{\Phi}_M^0$ is the rate of decline of the utility weights of the most patient type in the long run. As $M \to \infty$, the set of eigenvalues of $\mathbf{\Phi}_M^0$ contains $\hat{\mu}_1^i$, the long run utility discount factor of model i, and all eigenvalues of $\mathbf{\Phi}_M^0$ are less than or equal to $\max_i\{\hat{\mu}_1^i\}$.

Now consider the continuous set of models with weights $f_s^{ij}(\epsilon)$, where $\epsilon > 0$. Let $\mathbf{\Phi}_M(\epsilon)$ be the corresponding $\mathbf{\Phi}_M$ matrix for this set of models, where by assumption $\lim_{\epsilon \to 0^+} \mathbf{\Phi}_M(\epsilon) = \mathbf{\Phi}_M^0$. The consensus long run discount factor in model ϵ of order M, denoted $\mu_1(\epsilon, M)$ is the largest eigenvalue of $\mathbf{\Phi}_M(\epsilon)$. Define

$$\hat{\mu}_1(\epsilon) = \lim_{M \to \infty} \mu_1(M, \epsilon).$$

We know that this limit exists, due to the proof of Proposition 3. Since the matrix $\mathbf{\Phi}_M(\epsilon)$ is continuous in $\epsilon > 0$, and in the limit as $M \to \infty$ the largest eigenvalue of $\mathbf{\Phi}_M(0) = \mathbf{\Phi}_M^0$ is equal to $\max_i\{\hat{\mu}_1^i\}$, we must have

$$\lim_{\epsilon \to 0^+} \hat{\mu}_1(\epsilon) = \max_i\{\hat{\mu}_1^i\}.$$

Since $\hat{\rho}(\epsilon) = -\ln \hat{\mu}_1(\epsilon)$ by definition, the result follows.

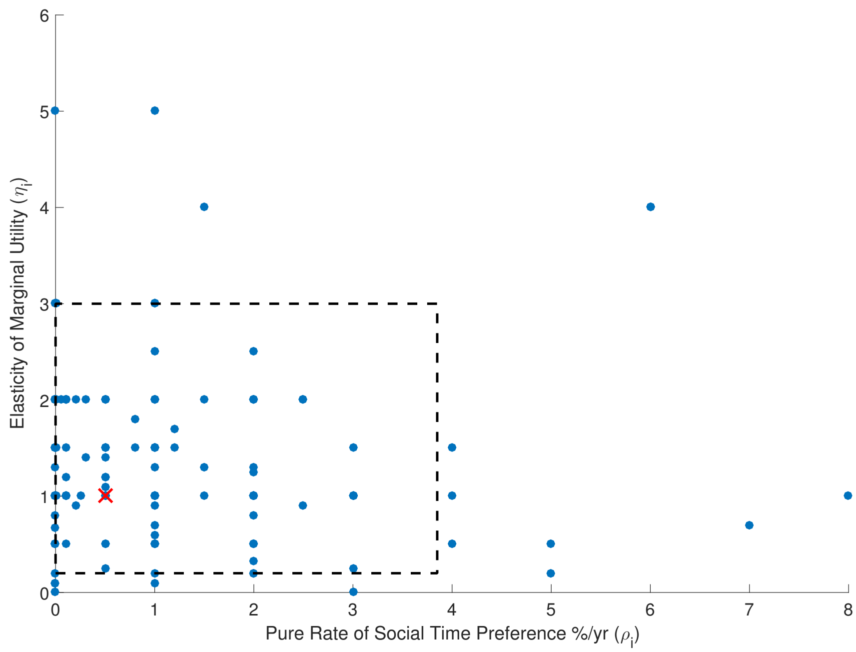


Figure D1. : Experts' recommended values for the pure rate of social time preference (ρ_i), and the elasticity of marginal utility (η_i) for appraisal of long run public projects, from the Drupp et al. (2015) survey. 173 responses were recorded. The dashed box depicts data points that fall inside the $5-95\%$ ranges of both parameters. The red cross indicates the location of the median values of ρ_i and η_i.

Details of calibration

The data I use to calibrate the model and generate the results in Figures 1a and 1b are taken from a recent survey by Drupp et al. (2015). They surveyed expert economists who have published papers on social discounting, asking for their opinions on, amongst other things, the appropriate values of the pure rate of social time preference and the elasticity of marginal social utility. The distribution of respondents' views on these two parameters is plotted in Figure D1.

To calibrate the values of γ, α_i in (20), I exploit the fact that when $x = 1$ the model reduces to a set of N homogeneous exponential models of non-paternalistic intertemporal preferences. Such models have been studied by e.g Saez-Marti and Weibull (2005); Galperti and Strulovici (2017). The latter authors in particular provide an axiomatic characterization of homogeneous exponential non-paternalistic preferences.

When $x = 1$, decision makers' social preferences can be represented as follows:

$$V_\tau^i = U^i(c_\tau) + \gamma \sum_{s=1}^{\infty} (\alpha_i)^s V_{\tau+s}^i, \tag{D1}$$

where $\alpha \in (0, 1)$ and $\gamma \in (0, \frac{1-\alpha}{\alpha})$. It is straightforward to show (see e.g. Galperti and Strulovici, 2017) that these preferences have the following equivalent representation:

$$V_\tau^i = U^i(c_\tau) + \sum_{s=1}^{\infty} \beta(\delta_i)^s U^i(c_{\tau+s}), \text{ where } \beta = \frac{\gamma}{\gamma+1}, \ \delta_i = (1+\gamma)\alpha_i. \tag{D2}$$

Notice that if we make the change of variables $\alpha_i = \tilde{\delta}_i/(1+\gamma)$ in (D1), and then take the limit as $\gamma \to \infty$, equation (D2) implies that the resulting preferences have the representation

$$V_\tau^i = \sum_{s=0}^{\infty} \tilde{\delta}_i^{\,s} U^i(c_{\tau+s}). \tag{D3}$$

Thus in this limit the exponential homogeneous non-paternalistic model has an equivalent representation as discounted utilitarian time preferences.

To calibrate the discount factors α_i, I assume that the data in Figure D1 correspond to paternalistic views on the long run PRSTP (i.e. $x = 1$).

Equation (D2) tells us that in this case type i's long run utility discount factor is $(1+\gamma)\alpha_i$. I will treat γ as a free parameter of the model, and thus calibrate α_i so that

$$\alpha_i = \frac{e^{-\rho_i}}{1+\gamma}, \tag{D4}$$

where ρ_i is the observed value of type i's opinion on the utility discount rate. Using this calibration methodology I can simulate the model for different values of the free parameter γ. From (D2) the discount factor of type i at $s = 1$ is given by

$$\frac{1}{1+\gamma^{-1}} e^{-\rho_i} \approx e^{-(\gamma^{-1}+\rho_i)}$$

when γ^{-1} is small. Thus e.g. $\gamma = 100$ corresponds to an additional 1% discount rate on the immediate future, over and above the long run discount rate ρ_i. To make the model a close approximation to discounted utilitarianism when $x = 1$, but also ensure that all types place positive weight on all future wellbeing measures (which requires γ be finite), I pick γ^{-1} to be small, but non-zero, i.e. $\gamma^{-1} = 0.1\%$.

In addition, I assume that types believe the social utility function is isoelastic, i.e.

$$U^i(c) = \frac{c^{1-\eta_i}}{1-\eta_i} \tag{D5}$$

for some $\eta_i > 0$. This implies that the elasticity of marginal social utility is constant and equal to η_i, and I simply calibrate η_i to be each respondent's preferred value of this elasticity.

Comparative statics of the consensus long-run PRSTP

It is naturally of interest to ask how the consensus long run PRSTP $\hat{\rho}$ depends on the intertemporal wellbeing weights f_s^{ij}. Unfortunately strong comparative statics results on this question are likely out of reach. Technically, we need to understand how the spectral radius (i.e. largest eigenvalue) of the matrices $\mathbf{\Phi}_M$ from Proposition 3 behaves when we spread out or contract the distribution of weights f_s^{ij}. In order to sign the effect of a spread in the weights we require something akin to a convexity property for the spectral radius. Unfortunately, it is known that the spectral radius of a matrix is a convex function of its diagonal elements, but not of the off-diagonal elements (Friedland, 1981).[9]

This section describes a special case of the model in which clean comparative statics are possible.

Assume that the intertemporal weights f_s^{ij} take the following symmetric form:

$$f_s^{ij} = \begin{cases} g(s,\lambda_i)x_s & j = i \\ g(s,\lambda_i)\frac{1-x_s}{N-1} & j \neq i \end{cases} \tag{E1}$$

where $x_s \in [1/N, 1)$ for all $s = 1 \ldots \infty$, and $\sum_{s=1}^{\infty} g(s,\lambda) < 1$ for all $\lambda \in I \subset \mathbb{R}^+$. In this model the time dependence of types' intertemporal weights f_s^{ij} has a common functional form, given by a discount function $g(s,\lambda)$ on wellbeings s years in the future, where $\lambda > 0$ is a parameter. Variations in types' attitudes to time are solely due to differences in their values of λ.

[9]Similarly, it is not possible to sign the effect of premultiplying $\mathbf{\Phi}_M$ by a doubly stochastic matrix, as the spectral radius of a product of two matrices is not sub-multiplicative in general. Gelfand's formula shows that the spectral radius of a matrix product is sub-multiplicative if the matrices in question commute, but this is not much use for our purposes.

This model is a generalization of the model defined in (20), which we used in Section II of the paper.

Let $\vec{\lambda} = (\lambda_1, \ldots, \lambda_N)$ be the vector of types' λ parameters in the model (E1), and let $X = (x_1, x_2, \ldots)$ be the sequence of values of x_s. Finally, let $\hat{\rho}(\vec{\lambda}, X)$ be the consensus long-run PRSTP in a model characterized by the parameters $\vec{\lambda}, X$.

PROPOSITION 4: *Assume that the discount function $g(s, \lambda)$ is strictly log convex in λ. Then if the parameter vector $\vec{\lambda}^A$ majorizes $\vec{\lambda}^B$,*

$$\hat{\rho}(\vec{\lambda}^A, X) < \hat{\rho}(\vec{\lambda}^B, X).$$

In words, this result says that if the discount function $g(s, \lambda)$ is log convex in λ, and policy makers in group A disagree more about the parameter λ than policy makers in group B, the consensus long-run PRSTP will be *lower* in group A than in group B.

I will provide some interpretation of the log-convexity condition in examples below, but first we turn to the proof.

PROOF:

The proof relies on the following result due to Kingman (1961):

LEMMA 4: *Let $b_{ij}(\theta) > 0$ be the elements of a non-negative matrix $\mathbf{B}$, where $\theta \in \mathbb{R}$ is a parameter. If $b_{ij}(\theta)$ is log-convex in θ for all i, j, the spectral radius of $\mathbf{B}$ is a log-convex function of θ.*

We will employ the usual trick of working with finite order models first (i.e. setting f_s^{ij} to zero for $s > M$), and taking a limit as $M \to \infty$ at the end. The consensus long run PRSTP in a model of order M is determined by the largest eigenvalue of $\mathbf{\Phi}_M$, defined in (A13). Denote this eigenvalue by $\hat{\mu}_M(\vec{\lambda})$, where I have suppressed the dependence on the parameters X for simplicity.

Now consider a parametric family of matrices of the form $\mathbf{\Phi}_M(\theta)$, where the matrices in this family are constructed by analogy with (A13), and where the intertemporal weights are of the form

$$f_s^{ij}(\theta) = \begin{cases} g(s, \lambda_i\theta)x_s & j = i \\ g(s, \lambda_i\theta)\frac{1-x_s}{N-1} & j \neq i \end{cases} \tag{E2}$$

The matrix we are actually interested in corresponds to $\theta = 1$. Let $\hat{\mu}_M(\vec{\lambda}, \theta)$ be the spectral radius of $\mathbf{\Phi}_M(\theta)$. Since λ_i only enters the matrix elements of $\mathbf{\Phi}_M(\theta)$ as a product with θ, the spectral radius can only depend on λ_i through such products, i.e.

$$\hat{\mu}_M(\vec{\lambda}, \theta) = Z(\lambda_1\theta, \lambda_2\theta, \ldots, \lambda_N\theta)$$

for some unknown function Z.

Since $g(s, \lambda)$ is log convex in λ by assumption, the elements of $\mathbf{\Phi}_M(\theta)$ are log convex functions of θ. Thus by Kingman's result the spectral radius $\hat{\mu}_M(\vec{\lambda}, \theta)$ is also log convex in θ. We now show that this implies that $\log Z(\lambda_1\theta, \ldots, \lambda_N\theta)$ is a convex function of $\vec{\lambda}$ for all θ, and in particular for $\theta = 1$:

Let $u_i = \lambda_i\theta$, so that $Z = Z(u_1, u_2, \ldots, u_N)$

$$\frac{d \log Z}{d\theta} = \frac{\sum_i \frac{\partial Z}{\partial u_i}\lambda_i}{Z} \tag{E3}$$

$$\frac{d^2 \log Z}{d\theta^2} = \frac{\sum_i \lambda_i \sum_j \left(\frac{\partial Z}{\partial u_i \partial u_j}\lambda_j\right) - \sum_{i,j} \frac{\partial Z}{\partial u_i}\frac{\partial Z}{\partial u_j}\lambda_i\lambda_j}{Z^2} \tag{E4}$$

Now since

$$\frac{\partial Z}{\partial u_i} = \frac{\partial Z}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial u_i} = \frac{\partial Z}{\partial \lambda_i}\frac{1}{\theta}$$

we have

$$\frac{d^2 \log Z}{d\theta^2} \propto \sum_{i,j} \frac{\lambda_i \lambda_j}{\theta^2}\left(\frac{\partial^2 Z}{\partial \lambda_i \partial \lambda_j} - \frac{\partial Z}{\partial \lambda_i}\frac{\partial Z}{\partial \lambda_j}\right) \tag{E5}$$

$$\propto \sum_{i,j} \frac{\lambda_i \lambda_j}{\theta^2}\left(\frac{\partial^2 \log Z}{\partial \lambda_i \partial \lambda_j}\right) > 0. \tag{E6}$$

where the last inequality follows from the log-convexity of Z in θ. Since this inequality holds for all values of λ_i, it must be true that

$$\frac{\partial^2 \log Z}{\partial \lambda_i \partial \lambda_j} > 0$$

for all i, j. Thus, setting $\theta = 1$, we conclude that the spectral radius of our original matrix, $\hat{\mu}_M(\vec{\lambda})$, is a log convex function of $\vec{\lambda}$.

The final step of the proof is to observe that because of the symmetry of the set of intertemporal weights in (E1) the spectral radius must be a symmetric function of $\vec{\lambda}$, i.e. any permutation of the elements of $\vec{\lambda}$ will leave the spectral radius unchanged. Since $\hat{\mu}_M(\vec{\lambda})$ is a log convex, symmetric function of $\vec{\lambda}$, its log is Schur-convex. Since $\hat{\mu}_M(\vec{\lambda}) = e^{-\hat{\rho}_M(\vec{\lambda}, X)}$, this implies that $\hat{\rho}_M(\vec{\lambda}, X)$ is Schur-concave in $\vec{\lambda}$. Thus by the properties of Schur-concave functions, if $\vec{\lambda}^A$ majorizes $\vec{\lambda}^B$ we must have

$$\hat{\rho}_M(\vec{\lambda}^A, X) < \hat{\rho}_M(\vec{\lambda}^B, X).$$

The final result follows by taking the limit as $M \to \infty$.

As an initial example of the application of this result, consider a model in which the discount function declines exponentially, i.e.

$$g(s,\lambda) = (1+\lambda)^{-s}$$

In this case $\log g(s,\lambda) = -s\log(1+\lambda)$, which is strictly convex in λ. Thus the result applies – more disagreement about the 'social wellbeing discount rate' λ *decreases* the consensus long run PRSTP.

We can extend this finding to a more general class of models by assuming that $g(s,\lambda) = \tilde{g}(\lambda s)$, i.e. the parameter λ acts to rescale the time variable s. Following Prelec (2004) we will say that $\tilde{g}(s)$ exhibits *decreasing impatience* if $\log \tilde{g}(s)$ is a convex function of s for $s > 0$. Discount functions that exhibit decreasing impatience have the form $\tilde{g}(s) = e^{-h(s)}$ where $h(s)$ is a concave function. The rate of increase of $h(s)$ (which measures impatience) slows as the time horizon s increases.

COROLLARY 1: *Assume that $\tilde{g}(s)$ exhibits decreasing impatience, and that the parameter vector $\vec{\lambda}^A$ majorizes $\vec{\lambda}^B$. Then*

$$\hat{\rho}(\vec{\lambda}^A, X) < \hat{\rho}(\vec{\lambda}^B, X).$$

Thus, for example, in a hyperbolic model (see e.g. Prelec, 2004) we would have

$$\text{(E7)} \qquad \tilde{g}(s) = (1+s)^{-(1+p)} \Rightarrow g(s,\lambda) = \tilde{g}(\lambda s) = (1+\lambda s)^{-(1+p)}$$

where $p > 0$ is a parameter. $\tilde{g}(s)$ is log convex in s, so more disagreement about λ reduces the consensus PRSTP in this model.