On the welfare economics of climate change

Francis Dennig
Oriel College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2013

To Sasha, Claire and Edgar,
whose complete ignorance of the contents of this thesis belies the importance of the contribution they have made to it.

Acknowledgements

First and foremost I would like to thank my supervisor Professor John Quah for his patience, tolerance and support. Without his guidance, this thesis would certainly not have seen the light of day. Many friends and colleagues deserve to be mentioned, but one stands out by his presence in the following pages. Bassel Tarbush has helped me tackle big conceptual problems as well as minor technical issues, and made the process much less tedious than it otherwise would have been. Other people in my vicinity, students and staff of the economics department and outside of it, have endured this process with varying degrees of fortitude, and I thank them for their patience. Finally, I must thank the ESRC, for this work was supported by the Economic and Social Research Council [grant number ES/I903887/1].

On the welfare economics of climate change

Francis Dennig

Oriel College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2013

The three constituent chapters of this thesis tackle independent, self-contained research questions, all concerning welfare economics in general and its application to climate change policy in particular.

Climate change is a policy problem for which the costs and benefits are distributed unequally across space and time, as well as one involving a high degree of uncertainty. Therefore, cost-benefit analysis of climate policy ought to be based on a welfare function that is sufficiently sophisticated to incorporate the three dimensions of aggregation: time, risk and space. Chapter 1 is an axiomatic treatment of a stylised model in which all three dimensions appear. The main result is a functional representation of the social welfare function for policy assessment in such situations.

Chapter 2 is a numerical mitigation policy analysis. I modify William Nordhaus' RICE-2010 model by replacing his social welfare function with one that allows for different degrees of inequality aversion along the regional and inter-temporal dimension. I find that, holding the inter-temporal coefficient of inequality aversion fixed, performing the optimisation with a greater degree of regional inequality reduces the optimal carbon tax relative to treating the world as a single aggregate consumer.

In Chapter 3 I analyse climate policy from the point of view of intergenerational transfers. I propose a system of transfers that allows future generations to compensate the current one for its mitigation effort and demonstrate the effects in an OLG model. When the marginal benefit to a – possibly distant – future generation is greater than the cost of compensating the current generation for its abatement effort, a Pareto improvement is possible by a combination of mitigation policy and transfer payments. I show that under very general assumptions the business-as-usual outcome is Pareto dominated by such policies and derive the conditions for the set of climate policies that are *not* dominated thus.

Contents

Chapter 1

A social objective with explicit risk and time dimensions

Chapter abstract

The social choice literature lacks an analysis of welfare functions which explicitly incorporates distinct periods of consumption as well as uncertainty in outcomes. This paper provides an axiomatic treatment of social choice with separate risk, time and interpersonal dimensions. The main result establishes assumptions under which the representing welfare function treats the three dimensions of aggregation separately.

1.1 Introduction

Some situations require decisions at a social level which involve trade-offs along both risk and time dimensions, as well as an aggregation across individuals. Starting with Koopmans (1960) the literature contains various axiomatic analyses of social time preference and similarly, since Harsanyi (1955) there has been a substantial amount of work analysing social choice under risk. The aim of this paper is to derive an axiomatic welfare representation of social choice in the presence of all three dimensions of aggregation which acknowledges that different parameters may govern the respective trade-offs.

Broome (1991) provides a framework encompassing all three dimensions, but does so in a way which conflates risk and time preferences. In the context of rational individual choice, Selden (1978) and Kreps and Porteus (1978) take the view that time and risk represent two separate dimensions of aggregation. We propose the same approach in the context of collective choice.

Models of intertemporal choice such as that in Koopmans (1960) are often applied to society as a whole without explicit modeling of the intratemporal trade-offs between contemporaneous individuals. We model two time-periods and make explicit the assumptions which allow us to separate the within period aggregation from the between period aggregation.

Harsanyi (1955) considers lotteries with outcomes affecting many contemporary individuals. His representation theorem yields a social objective that is necessarily affine in the individual risk preferences. The result is obtained by requiring that the social preference be sensitive to the individual *ex-ante* rational risk preferences as well as demanding that the social preference itself satisfy the expected utility hypothesis *ex-post*. The result is a social preference that is additive in utilities along the risk dimension as well as the interpersonal dimension. Broome (1991) provides a similar result under the inclusion of a time dimension. Blackorby, Donaldson, and Mongin (2004) and Zuber (2009) provide extensions, with similar results, in which the sure thing principle is abandoned for weaker rationality criteria both in the ex-ante and ex-post conditions.

Fleurbaey (2010) avoids the affine utilitarian result by weakening the ex-ante condition. He only requires the social objective to respect unanimity when individuals face no risk, or when there is risk, but no inequality across individuals. By giving up on the full ex-ante Pareto principle he is able to derive an expectational representation of the social objective which departs from the additive form along the interpersonal dimension. This allows for a degree of inequality aversion in the social preference that is not possible in Harsanyi's representation. The weakening of the Pareto principle is justified by arguing that sensitivity is retained to preferences over certain, which carry the objective information required for social trade-offs, while the preferences in face of uncertainty are made under ignorance and are interpreted as subjective assessments of risk at best.

As done in Hammond (1983) we assume that *consequentialist utilities on outcomes* are the primitive concept relevant to social choice rather than preferences over risky or temporally separated prospects. Therefore we cannot appeal to any *ex-ante* conditions at all. We model a social decision concerning a group of individuals spanning two periods, with certainty in outcomes during the first period and uncertainty in outcomes during the second period. Assuming the existence of consequentialist utilities in outcomes that are measurable and interpersonally comparable in the sense understood by Sen (1970), we employ the results on the existence of a social welfare function from d'Aspremont and Gevers (1977) and Roberts (1980).

We propose axioms particular to the configuration of our problem which allow us to pin down the structure of the welfare function in relation to the three dimensions of aggregation. The final representation has the feature that amongst outcomes which have no uncertainty as well as no variation in utilities across individuals it is simply a separable two-period welfare function, with one argument in each period. Its interpretation is simply that of the social time preference. When there is no variation across individuals, but uncertainty in the second period outcomes, we have that the social planner's preference takes the form of the ordinal certainty equivalent representation described in Selden (1978). When comparing arbitrarily distributed utility vectors, the welfare function computes first the equally distributed equivalent across individuals and passes these as arguments to the risk and time aggregators described above. More formally, let $\vec{v}_1 \in \mathbb{R}^N$ be a vector of period 1 utilities and $\vec{v}_\omega \in \mathbb{R}^N$ be vectors of period 2 utilities in each state of the world $\omega \in \Omega$. Let $e_1 : \mathbb{R}^N \to \mathbb{R}$ be the equally distributed equivalent utility in period 1 and $e_2 : \mathbb{R}^N \to \mathbb{R}$ that in period 2. If $CE : \mathbb{R}^{|\Omega|} \to \mathbb{R}$ is the certainty equivalent function for state space Ω and $T : \mathbb{R}^2 \to \mathbb{R}$ the social time preference, the social welfare function we derive has the general form

$$W = T\left\{e_1(\vec{v}_1), CE\left[e_2(\vec{v}_\omega)_{\omega\in\Omega}\right]\right\}$$

Under the additional assumption of separability across time this yields the more familiar looking

$$W = u\{e_1(\vec{v}_1)\} + \frac{1}{1+\rho} u\{CE\left[e_2(\vec{v}_\omega)_{\omega\in\Omega}\right]\}$$

The chapter is organised as follows. Section 1.2 describes the social choice framework. Section 1.3 derives the expectational form of the risk aggregation and Section 1.4 shows the separability of the interpersonal aggregation from the risk aggregation in the second period. Finally, Section 1.5 combines those results with a weak condition on the separability of choice across time to yield the full representation theorem.

1.2 The framework and social preference

Let $\mathcal{N} = \{1, 2, \ldots, N\}$ be a set of individuals. We consider a set of outcomes, Y_n, attributable to individual $n \in \mathcal{N}$. Loosely speaking, these are to be understood as atemporal and devoid of uncertainty.[1] This is, of course, a simplification. As discussed in Savage (1954) in the context of uncertainty, restricting the set of such outcomes to completely certain consequences would impose too complex a structure on the uncertain choice. Similarly, dividing personal outcomes into fully atemporal, discrete events would make choice among such objects unrealistic. The idea here is that the time and risk aspects which are salient to the decision are abstracted from, leaving 'consequences' over which the consequentialist utilities are subsequently defined.[2] Consumption over a generation is one of the examples we have in mind. It is over these consequences that individual welfare is understood for the purposes of this paper. Denote by $h_n : Y_n \to \mathbb{R}$ the functions for which $h_n(y)$ is the *consequentialist utility* of individual $n \in \mathcal{N}$ from the consequence $y \in Y_n$. Notice that at this stage there is no notion of intertemporal or uncertain choice; these objects are defined independently of that.

Let Y_n^1 be the set of possible outcomes to individual n in period 1 and Y_n^2 the possible outcomes in period 2. We will assume that $|Y_n^i| < \infty$ for $i = 1, 2$ and $n \in \mathcal{N}$. As is done in Selden (1978) in the context of individual decision-making, we will consider only two periods with certain outcomes in the first and uncertain outcomes in the second. Denote by $\mathbf{Y_1} = \prod_{n\in\mathcal{N}} Y_n^1$ and $\mathbf{Y_2} = \prod_{n\in\mathcal{N}} Y_n^2$ the sets of outcomes to all individuals in periods 1 and 2 respectively.[3] Let Ω be a finite state space of cardinality $|\Omega| = K \geq 3$ with a uniform distribution[4] and let $\mathcal{M}(\mathbf{Y_2})$ be the set of

[1] *A priori* there is no restriction on the nature of Y_n. Neither topological, nor conceptual.

[2] Chapter 5 in Savage (1954) discusses the formal requirements for this to be possible

[3] The discussion assumes throughout that the same individuals live through both periods and that it is possible to identify them across these, but the problem of social choice across two non-overlapping generations is mathematically indistinguishable.

[4] Notice that this is not a restriction in any practical sense, since any discrete distribution with

functions from Ω to $\mathbf{Y_2}$. Define the set of social alternatives as

$$X = \mathbf{Y_1} \times \mathcal{M}(\mathbf{Y_2}) \tag{1.1}$$

An element $x \in X$ contains a vector of certain outcomes for all individuals in period 1 and a vector of state-contingent outcomes for all individuals in period 2. Extend the state space by an additional index denoting period 1 to describe the time/state space jointly by one index set,

$$\Theta = \{1\} \cup \Omega. \tag{1.2}$$

For $\theta \in \Theta$ let $Y_n^\theta = Y_n^1$ and $h_n^\theta = h_n^1$ for $\theta = 1$ and $Y_n^\theta = Y_n^2$ and $h_n^\theta = h_n^2$ for $\theta \in \Omega$. Denote by $x(n, \theta) \in Y_n^\theta$ the outcome alternative $x \in X$ assigns to individual $n \in \mathcal{N}$ in period/state $\theta \in \Theta$. A given array of consequentialist welfare functions

$$\{h_n^i : Y_n^i \to \mathbb{R}\}_{n \in \mathcal{N}, i \in \{1,2\}}$$

defines a function $\vec{u} : X \times \Theta \to \mathbb{R}^N$ by setting, for all $x \in X$, $n \in \mathcal{N}$ and $\theta \in \Theta$,

$$u_n(x, \theta) = h_n^\theta(x(n, \theta))$$

Thus defined, $u_n(x, \theta)$ is the consequentialist utility at alternative $x \in X$ of individual $n \in \mathcal{N}$ in period/state $\theta \in \Theta$. The function $\vec{u}$ contains all the welfare information of the social choice problem given the welfare functions $\{h_n^i\}_{n \in \mathcal{N}, i \in \{1,2\}}$.[5] Define $\mathcal{U}$ to be the set of all such functions from $X \times \Theta$ to $\mathbb{R}^N$ which can be generated by a composition with *some* array $\{h_n^i\}_{n \in \mathcal{N}, i \in \{1,2\}}$. It is easy to see that $\mathcal{U}$ is equivalent to

rational weights can be reproduced by such a state space.

[5]Depending on the measurability of the functions $\{h_n^i\}_{n \in \mathcal{N}, i \in \{1,2\}}$ a welfare representation may or may not be possible. We will assume that the welfare is sufficiently measurable and comparable across individuals (see Roberts (1980)).

the space of all real vector valued functions on $X \times \Theta$ that satisfy the restriction

$$x(n,\theta) = x(n,\vartheta) \implies u_n(x,\theta) = u_n(x,\vartheta)$$

The *social choice functional* is a mapping $f : \mathcal{D} \to \mathcal{R}(X)$ where $\mathcal{D}$ is some subset of $\mathcal{U}$ and $\mathcal{R}(X)$ is the set of *preferences* on X. A preference on X is taken to mean a complete, reflexive and transitive binary relation on X. To every $\vec{u} \in \mathcal{U}$, which contains all the consequentialist welfare information relevant to the social decision, f assigns a preference on X, representing the society's ranking of the social alternatives X. Given such a ranking, it is always possible to find a most preferred alternative under the restriction to a finite set of alternatives $A \subset X$.

The following are the Arrovian conditions adapted to the current framework.

Axiom 1 (U: Unrestricted domain)**.** $\mathcal{D} = \mathcal{U}$

This is the requirement that the functional be able to aggregate any function $\vec{u} \in \mathcal{U}$ representing individual welfares.

Axiom 2 (I: Independence of irrelevant alternatives)**.** *For any $\vec{u}, \vec{u}' \in \mathcal{D}$, and $A \subset X$, if $\vec{u}(x,\cdot) = \vec{u}'(x,\cdot)$, for all $x \in A$, then $f(\vec{u})$ and $f(\vec{u}')$ coincide on A.*

This is the multi-profile condition which requires the welfare functional to yield the same order on a subset of alternatives for two different lists of individual welfares provided the two lists of welfares are the same for every individual in every period and state for all the alternatives in the subset. Put differently, the social ranking between any set of alternatives may only depend on the welfare information at said alternatives.

Axiom 3 (P: Pareto principle for every period and state)**.** *For all $x, z \in X$, $\vec{u} \in \mathcal{D}$, if $\vec{u}(x,\theta) \geq \vec{u}(z,\theta)$ for all $\theta \in \Theta$, then xRz, where $R = f(\vec{u})$. If, furthermore, there exists $n \in \mathcal{N}$ and $\vartheta \in \Theta$ such that $u_n(x,\vartheta) > u_n(z,\vartheta)$, then xPz, where P is the asymmetric part of R.*

As this only restricts the collective choice in the cases where there is unanimity across all individuals in every period and state of the world, this constitutes a departure from and substantial weakening of the *ex-ante* based Pareto principle as in Harsanyi (1955) or even Blackorby, Donaldson, and Mongin (2004) and Zuber (2009). In the same way as Axioms 4 and 5 in Fleurbaey (2010), (P) is open to the criticism that the criterion does not restrict the social objective from overriding a decision which would be made unanimously by the individuals given their individual risk and time preferences. The condition only imposes a restriction when one alternative is uniformly better than the other in terms of consequentialist utilities.

Axioms 1 to 3 are used to derive a social ordering on Euclidian space from the welfare functional. They are conditions that will be understood to hold throughout the chapter, as without them, it is not guaranteed that the social preference can be represented by such an ordering on Euclidean space.

Define $\mathcal{I} = \mathcal{N} \times \Theta$. Then $I := |\mathcal{I}| = N \cdot (1 + K)$ and for each $x \in X$ and $u \in \mathcal{U}$

$$\vec{u}(x, \cdot) \in \mathbb{R}^I \tag{1.3}$$

That is, the welfare outcomes *at a particular alternative* $x \in X$ can be seen as an I-dimensional real vector. We will denote a typical element by $\vec{\mathbf{v}} \in \mathbb{R}^I$. The following proposition is due to d'Aspremont and Gevers (1977).

Proposition 1. *If a social welfare functional f satisfies Axioms 1, 2 and 3 then there exists a preference $\succsim$ on $\mathbb{R}^I$ such that for $x, z \in X$ and $\vec{u} \in \mathcal{U}$, if $R = f(\vec{u})$, $\vec{\mathbf{v}} = \vec{u}(x, \cdot)$ and $\vec{\mathbf{w}} = \vec{u}(y, \cdot)$ then*

$$xRz \iff \vec{\mathbf{v}} \succsim \vec{\mathbf{w}}$$

1.2.1 Social preference

Proposition 1 postulates the existence of a *social preference* on Euclidean space which summarises the aggregation procedure contained in the social choice functional f. It serves as a justification for the consequentialist approach which regards the utility outcome of an alternative as the *only* relevant information from a normative standpoint. More crucially, it also reduces the problem mathematically to an analysis of orderings on Euclidean space. The remaining axioms and results will be stated in terms of this social preference $\succsim$ on $\mathbb{R}^I$, while the corresponding conditions on f, along with the proof of their equivalence to the axioms below – contingent on (U), (P), and (I) – are left to the Appendix. Conditions on the social preference will be denoted by an asterisk, while those on the social choice functional drop the asterisk. Continuity (C)* below illustrates this.

For any $\vec{\mathbf{v}} \in \mathbb{R}^I$ define the lower and upper contour sets by

$$L(\vec{\mathbf{v}}) = \left\{ \vec{\mathbf{w}} \in \mathbb{R}^I \middle| \, \vec{\mathbf{v}} \succsim \vec{\mathbf{w}} \right\} \tag{1.4}$$

$$U(\vec{\mathbf{v}}) = \left\{ \vec{\mathbf{w}} \in \mathbb{R}^I \middle| \, \vec{\mathbf{w}} \succsim \vec{\mathbf{v}} \right\} \tag{1.5}$$

Axiom 4 (C*: Continuity)**.** *For all $\vec{\mathbf{v}} \in \mathbb{R}^I$, $L(\vec{\mathbf{v}})$ and $U(\vec{\mathbf{v}})$ are closed sets.*

Continuity is a technical assumption and will guarantee that the hypothesis of Debreu's representation theorem[6] is satisfied for the preference $\succsim$, thereby ensuring the existence of a continuous welfare function $W : \mathbb{R}^I \to \mathbb{R}$ representing $\succsim$ in that for all $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in \mathbb{R}^I$

$$\vec{\mathbf{v}} \succsim \vec{\mathbf{w}} \iff W(\vec{\mathbf{v}}) \geq W(\vec{\mathbf{w}}). \tag{1.6}$$

[6] Debreu (1954)

1.3 The separation theorem

In Subsection 1.3.1 we provide some notation and establish some mathematical properties of preferences defined over Cartesian products of sets. They are presented below without any economic intuition, but will be used repeatedly in the remaining sections of the chapter. In Subsection 1.3.2 we present the axioms regarding rational social choice under uncertainty. Essentially we propose Savage's Sure Thing Principle to the conditional period 2 choice on the welfare vectors that have no variation across individuals. We then apply a result from Debreu (1960) to derive a functional representation for the social choice amongst such outcomes. Proposition 2 provides for the existence of a social risk preference on the second period choice under uncertainty *conditional on the first period outcome.* That is to say that the aggregation across the risk dimension in the second period takes place *before* any aggregation across the two time periods. The resulting representation is analogous to the Ordinal Certainty Equivalent derived in Selden (1978) and for the restriction to two periods, the functional form is also analogous to the representation in Kreps and Porteus (1978) and Epstein and Zin (1989).

1.3.1 Preferences on subspaces

Let $\mathbf{A} = \prod_{j \in J} A_j$ be a Cartesian product of a finite number of sets $\{A_j\}_{j \in J}$ and $\succsim$ a preference on $\mathbf{A}$. Denote by $\mathbf{a} = (a_j)_{j \in J} \in \mathbf{A}$ a representative element therein. We will write (a_j) when it is clear from the context what the index set is. For any subset $L \subset J$ denote the projection of $\mathbf{a}$ onto the subspace $\prod_{j \in L} A_j$ by $\mathbf{a}_L = (a_j)_{j \in L}$. Given the preference $\succsim$, a subset $L \subset J$ and element $\mathbf{b} = (b_j)_{j \in L} \in \prod_{j \in L} A_j$, a preference $\overset{\mathbf{b}}{\succsim}_{J \setminus L}$ is induced on the projection $\prod_{j \in J \setminus L} A_j$ by

$$\text{for } \mathbf{x}, \mathbf{y} \in \prod_{j \in J \setminus L} A_j, \quad \mathbf{x} \overset{\mathbf{b}}{\succsim}_{J \setminus L} \mathbf{y} \iff (\mathbf{x}, \mathbf{b}) \succsim (\mathbf{y}, \mathbf{b}) \tag{1.7}$$

Let $M = J \backslash L$. If $\overset{\mathbf{b}}{\succsim}_M$ is independent of the particular element $\mathbf{b} \in \prod_{j \in L} A_j$, then the induced preference on $\prod_{j \in M} A_j$ is said to be *independent* and we can write $\succsim_M$ without the dependence on $\mathbf{b}$. If the induced preference is independent for every subset $M \subset I$, then the induced preference is said to be *strongly separable.*

If $A_j = A_i$ for all $i, j \in J$, then, with a slight abuse of notation, it is possible to write $\mathbf{A} = A^{|J|}$. In this case, a preference $\succsim_A$ is induced on A as follows: $a \succsim_A b$ for $a, b \in A$ if $[a] \succsim [b]$ where $[a] = (a, a, \ldots, a) \in \mathbf{A}$ and $[b] = (b, b, \ldots, b) \in \mathbf{A}$ are the elements of $\mathbf{A}$ in which the same value is replicated in each entry.

1.3.2 Rationality of social preference

Recall that $\mathbb{R}^I$ can be identified with the space of utility outcomes of the social decision problem described in Section 1.2. An element $\vec{\mathbf{v}} = (\vec{v}_\theta)_{\theta \in \Theta} \in \mathbb{R}^I$ represents the utilities of each individual in each period or state of the world where $\vec{v}_\theta \in \mathbb{R}^N$ represents the vector of utilities of N agents in the period/state $\theta \in \Theta$. Recall that $\Theta = \{1\} \cup \Omega$.

The following is a symmetry condition, justified by the fact that state space Ω has a uniform distribution. It states that the social preference should be unaffected by a reshuffling of the states in which outcomes occur.

Axiom 5 (SI*: State Independence)**.** *Let σ be a permutation of Θ which maps 1 onto itself. For all $\vec{\mathbf{v}} = (\vec{v}_\theta)_{\theta \in \Theta}, \vec{\mathbf{w}} = (\vec{w}_\theta)_{\theta \in \Theta} \in \mathbb{R}^I$ with $\vec{v}_\theta, \vec{w}_\theta \in \mathbb{R}^N$, if for all $\theta \in \Theta$,*

$$\vec{w}_\theta = \vec{v}_{\sigma(\theta)} \implies \vec{\mathbf{v}} \backsim \vec{\mathbf{w}}$$

Let $\vec{s} \in \mathbb{R}^N$ be any vector of period 1 utilities. Then denote by $\overset{\vec{s}}{\succsim}_\Omega$ the preference on $\mathbb{R}^{N \times K}$ representing the social preference on the uncertain utility vectors in period 2 conditional on the period 1 utilities being $\vec{s} \in \mathbb{R}^N$. Notice that this is a preference over the space of risky prospects for N agents, which we identify with $\mathbb{R}^{N \times K}$. If amongst period 2 prospects you consider only those that assign the same utility to all agents in period 2 (possibly a different amount in each state of the world) you get a preference $\overset{\vec{s}}{\succsim}_{\mathbb{R}^K}$ on $\mathbb{R}^K$. This is the space of lotteries over equal utility vectors in period 2. Let $\mathbb{R}^K$ be $\mathbf{a} = (a_\omega)_{\omega \in \Omega} \in \mathbb{R}^K$ be a typical element therein.

Axiom 6 (S*: Separability across states)**.** *For any vector of period 1 utilities $\vec{s} \in \mathbb{R}^N$, the conditional preference $\overset{\vec{s}}{\succsim}_{\mathbb{R}^K}$ on $\mathbb{R}^K$ (i.e. the conditional preference on those period 2 prospects that assign all agents the same utility) is strongly separable in the sense defined in Subsection 1.3.1.*

To understand Axiom 6 it is important to keep in mind that it is a condition on the second period choice under uncertainty, and it only imposes a restriction on the cases in which the decision is stripped of its interpersonal dimension. The condition says that the uncertain choice satisfies the sure thing principle, or rather, Savage's condition P2 in this special case. This is the rationality condition which ensures that collective choice is consistent in the face of uncertainty.[7]

Conditions (C)*, (SI)* and (S)* allow is to apply a result on separability of preference representations due to Debreu (1960) to characterise the conditional period 2 risk preference induced by the social preference. We state the result as a corollary to Debreu's Lemma.

Proposition 2. *Suppose the social preference $\succsim$ on $\mathbb{R}^I$ satisfies Axioms 4, 5 and 6 as well as being based on a social welfare functional satisfying Axioms 1, 2 and 3. Then for every vector of period 1 utilities $\vec{s} \in \mathbb{R}^N$ there exists a function*

$$V[\vec{s}] : \mathbb{R} \to \mathbb{R} \tag{1.8}$$

[7]Axiom 6 is essentially an extension of Assumption 4 in Selden (1978) to collective choice. It states that the sure thing principle applies intratemporally, conditioning on previous periods. Bommier, Chassagnon, and Le Grand (2012) provide a criticism of such an approach, which rests essentially on a condition which applies the sure thing principle to dynamic prospects.

that is unique up to an increasing affine transformation, such that the expectation thereof, defined by

$$E[\vec{s}](\mathbf{a}) := \frac{1}{K}\sum_{\omega\in\Omega} V[\vec{s}](a_\omega) \tag{1.9}$$

on $\mathbf{a} = (a_\omega)_{\omega\in\Omega} \in \mathbb{R}^K$*, represents the induced preference* $\overset{\vec{s}}{\succsim}_{\mathbb{R}^K}$ *on those second period outcomes at which all agents have the same utility.*

The following is due to Debreu (1960).

Lemma 1. *Let* $\succsim$ *be a strongly separable and continuous preference on a finite Cartesian product* $\mathbf{A} = \prod_{j\in J} A_j$ *of connected and separable*[8] *spaces* A_j*,* $j \in J$*. Suppose the induced orders on more than two single projections* A_i, A_k*, for* $i \neq k$ *are not trivial. Then there exist* $|J|$ *continuous functions* $\xi_j : A_j \to \mathbb{R}$*, which are unique up to a common increasing affine transformation, such that for all* $\mathbf{x} = (x_j)_{j\in J}, \mathbf{y} = (y_j)_{j\in J} \in \mathbf{A}$*,*

$$\mathbf{x} \succsim \mathbf{y} \iff \sum_{j\in J} \xi_j(x_j) \geq \sum_{j\in J} \xi_j(y_j)$$

Proof of Proposition 2. The social preference $\succsim$ is continuous by (C)*. Therefore, the induced preference $\overset{\vec{s}}{\succsim}_{\mathbb{R}^K}$ is continuous as well. By (S)*, it is strongly separable. For each $\omega \in \Omega$ identify $\mathbb{R}_\omega \equiv \mathbb{R}$. Then $\mathbb{R}^K = \prod_{\omega\in\Omega} \mathbb{R}_\omega$. Let $\mathbf{a} = (a_\omega)_{\omega\in\Omega}, \mathbf{b} = (b_\omega)_{\omega\in\Omega} \in \mathbb{R}^K$ be such that there exists exactly one $\vartheta \in \Omega$ such that $a_\vartheta > b_\vartheta$ and $a_\omega = b_\omega$ for all $\omega \neq \vartheta$. Then, by the (P), it must be that $\mathbf{a} \overset{\vec{s}}{\succ}_{\mathbb{R}^K} \mathbf{b}$ and therefore, the induced order on the projection $\mathbb{R}_\vartheta$ is not trivial. This argument is valid for all $\vartheta \in \Omega$ and since $|\Omega| = K \geq 3$, the preference $\overset{\vec{s}}{\succsim}_{\mathbb{R}^K}$ satisfies the conditions of Lemma 1. Therefore, for all $\omega \in \Omega$ there exist $\xi_\omega[\vec{s}] : \mathbb{R} \to \mathbb{R}$ such that, for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^K$

$$\mathbf{a} \overset{\vec{s}}{\succsim}_{\mathbb{R}^K} \mathbf{b} \iff \sum_{\omega\in\Omega} \xi_\omega[\vec{s}](a_\omega) \geq \sum_{\omega\in\Omega} \xi_\omega[\vec{s}](b_\omega)$$

By (SI)*, the ranking may not be reversed if the elements of $\mathbf{a}$ are permuted. Therefore

[8] These are topological requirements satisfied by Euclidean space.

Table 1.1: Inequality and risk

	Heads		Tails	
	X	Y	X	Y
Ann	1	1	1	0
Bob	0	0	0	1

$\xi_\omega[\vec{s}] = \xi_\vartheta[\vec{s}] = \xi[\vec{s}]$ for all $\omega, \vartheta \in \Omega$. After an appropriate linear transformation we get the desired representation, i.e.

$$\mathbf{a} \overset{\vec{s}}{\succsim}_{\mathbb{R}^K} \mathbf{b} \iff \frac{1}{K}\sum_{\omega\in\Omega} V[\vec{s}](a_\omega) \geq \frac{1}{K}\sum_{\omega\in\Omega} V[\vec{s}](b_\omega)$$

□

$V[\vec{s}] : \mathbb{R} \to \mathbb{R}$ contains the social risk attitude in that it is a von Neumann-Morgenstern function on the utilities of the agents *when all agents have the same utility*. As with an individual von Neumann-Morgenstern utility on income, the function's curvature embodies the implied attitude to risk in the argument. For example, concavity in $V[\vec{s}](\cdot)$ would imply social aversion to risk in equally distributed utility.

1.4 Individuals and risk

The previous section provided a representation of the risk preferences of the social planner when comparing alternatives that do not have any variation along the dimension of individuals. Intuitively, the social planner would aggregate along the dimension of individuals, and simply pass that aggregate as an argument to the risk preference derived in the previous section. The conditions we have so far do not guarantee that this is possible. In fact, consider the following situation. The social planner must choose between alternatives X and Y which yield identical outcomes in the period 1, and the outcomes described in Table 1.1 in period 2. Alternative X

yields utility of one for Ann in each state of the world, and zero for Bob. Alternative Y gives one to Ann and zero to Bob in state Heads, and a reversal of fortunes in state tails. The conditions so far do not require the social planner to be indifferent between X and Y. Axiom 7 below imposes just such a restriction. Combined with the result of the previous section, this allows us to conclude that the representation of the social preference in the second period is the expectation of the equally distributed equivalent. That is, it takes the form $\mathbb{E}\left[e_2(\vec{v}_\omega)\right]$.

1.4.1 Weak dominance and the equally distributed equivalent

As throughout Section 1.3 let $\vec{s} \in \mathbb{R}^N$ denote a vector of period 1 utilities. As explained previously, conditional on $\vec{s}$ the social preference induces a preference $\overset{\vec{s}}{\succsim}_\Omega$ on the uncertain and prospects across the individuals in period 2. Mathematically these prospects are identified with the space $\mathbb{R}^{N\times K}$. We will denote a typical element therein by $\vec{\mathfrak{v}} \in \mathbb{R}^{N\times K}$. In Section 1.3 we analysed the restriction of that preference on the space of uncertain prospects at which all agents have the same utility, identified with $\mathbb{R}^K$. Now consider the restriction on the space of certain outcomes (no variation across states) with arbitrary variation in utility across agents. We can identify this space with $\mathbb{R}^N$ and will denote the restriction of the preference to this subset by $\overset{\vec{s}}{\succsim}_{\mathbb{R}^N}$. More formally, for any $\vec{v}, \vec{w} \in \mathbb{R}^N$, define

$$\vec{\mathbf{v}} = (\vec{s}, \overbrace{\vec{v}, \vec{v}, \ldots, \vec{v}}^{K \text{ times}}) \in \mathbb{R}^I$$

$$\vec{\mathbf{w}} = (\vec{s}, \underbrace{\vec{w}, \vec{,}, \ldots, \vec{w}}_{K \text{ times}}) \in \mathbb{R}^I$$

Then $\overset{\vec{s}}{\succsim}_{\mathbb{R}^N}$ is defined by

$$\vec{v} \overset{\vec{s}}{\succsim}_{\mathbb{R}^N} \vec{w} \iff \vec{\mathbf{v}} \succsim \vec{\mathbf{w}} \tag{1.10}$$

The following condition, due to Fleurbaey (2010) relates the preference $\overset{\vec{s}}{\succsim}_{\mathbb{R}^N}$ on $\mathbb{R}^N$ to $\overset{\vec{s}}{\succsim}_{\Omega}$ on $\mathbb{R}^{N\times K}$.

Axiom 7 (W*: Weak Dominance). *For all period 1 utility vectors $\vec{s} \in \mathbb{R}^N$, and second period utility arrays $\vec{\mathfrak{v}} = (\vec{v}_\omega)_{\omega\in\Omega}, \vec{\mathfrak{w}} = (\vec{w}_\omega)_{\omega\in\Omega} \in \mathbb{R}^{N\times K}$, if $\vec{v}_\omega \overset{\vec{s}}{\succsim}_{\mathbb{R}^N} \vec{w}_\omega$ for all $\omega \in \Omega$, then $\vec{\mathfrak{v}} \overset{\vec{s}}{\succsim}_{\Omega} \vec{\mathfrak{w}}$.*

The preference $\overset{\vec{s}}{\succsim}_{\mathbb{R}^N}$ ranks period 2 utility vectors under certainty, whereas $\overset{\vec{s}}{\succsim}_{\Omega}$ ranks period 2 utility vectors with potentially different outcomes across states of the world as well as across individuals. Condition (W)* says that if two such arrays $\vec{\mathfrak{v}} = (\vec{v}_\omega)_{\omega\in\Omega}, \vec{\mathfrak{w}} = (\vec{w}_\omega)_{\omega\in\Omega} \in \mathbb{R}^{N\times K}$ have the property that the state by state comparison according to the purely interpersonal aggregation under certainty $\overset{\vec{s}}{\succsim}_{\mathbb{R}^N}$ uniformly ranks one vector over the other, then the ranking must extend to the comparison between the uncertain vectors themselves according to $\overset{\vec{s}}{\succsim}_{\Omega}$. This is the axiom which guarantees that the interpersonal aggregation can be performed independently of the risk aggregation.[9]

We describe the interpersonal aggregation in period 2 under certainty by the *equally distributed equivalent.* For any vector of period 1 utilities, $\vec{s} \in \mathbb{R}$, and vector of period 2 utilities under certainty, $\vec{v} \in \mathbb{R}^N$, let the equally distributed equivalent utility of $\vec{v}$ conditional on $\vec{s}$ be defined by $e[\vec{s}](\vec{v})$, such that

$$\vec{v} \overset{\vec{s}}{\sim}_{\mathbb{R}^N} \left(e[\vec{s}](\vec{v}), \ldots, e[\vec{s}](\vec{v})\right)^\top \tag{1.11}$$

[9]It is also responsible for the state by state rationality which underlies the objection by Diamond (1967) to social choice rules that take an expectational form. The following representation is therefore also subject to such criticism. The essence of Diamond's criticism is that it ought to be *ex-ante* fairer to choose an alternative which allocates a single prize to two individuals with equal probability of winning, than to allocate the same prize to a particular individual with certainty. Valid cases are made against this proposal in Broome (1991) and Fleurbaey (2010), making a reiteration here unnecessary. We just add the further point that the *ex-ante* fairness implied by Diamond's example requires uncertainty to be truly objective in the sense that both possible worlds still have to be possible outcomes at the time the social decision is being made. Any concept of decision uncertainty as a subjective assessment of ignorance over an actually deterministic outcome would invalidate the underlying logic since the decision would determine the outcome with certainty anyway, even if the social planner doesn't know the outcome. Hence, the appeal of the example depends strongly on the particular view of the nature of uncertainty.

That is, conditional on the period 1 utility vector $\vec{s}$, the social planner is indifferent between the certain period 2 utility vector $\vec{v}$ and a a certain outcome in which each individual in the second period gets the utility level $e[\vec{s}](\vec{v})$.

Proposition 3. *Suppose the social preference satisfies Axioms 4, 5, 6 and 7. Then, for every vector of first period utilities* $\vec{s} \in \mathbb{R}^N$ *the function*

$$\frac{1}{K} \sum_{\omega \in \Omega} V[\vec{s}] \left(e[\vec{s}](\vec{v}_\omega) \right) \tag{1.12}$$

on $\vec{\mathfrak{v}} = (\vec{v}_\omega)_{\omega \in \Omega} \in \mathbb{R}^{N \times K}$ *represents the induced preference* $\overset{\vec{s}}{\succsim}_\Omega$ *on the uncertain period 2 utility vectors, where* $V[\vec{s}] : \mathbb{R} \to \mathbb{R}$ *is defined in Proposition 2 and* $e[\vec{s}] : \mathbb{R}^N \to \mathbb{R}$ *is defined by* (1.11).

Proof. Fix $\vec{s} \in \mathbb{R}^N$ and consider $\vec{\mathfrak{v}} = (\vec{v}_\omega)_{\omega \in \Omega}, \vec{\mathfrak{p}} = (\vec{p}_\omega)_{\omega \in \Omega} \in \mathbb{R}^{N \times K}$ such that

$$\vec{\mathfrak{v}} \overset{\vec{s}}{\succsim}_\Omega \vec{\mathfrak{p}}. \tag{1.13}$$

Define the arrays $\vec{\mathfrak{w}} = (\vec{w}_\omega)_{\omega \in \Omega}, \vec{\mathfrak{q}} = (\vec{q}_\omega)_{\omega \in \Omega} \in \mathbb{R}^{N \times K}$ by setting

$$\vec{w}_\omega = (e[\vec{s}](\vec{v}_\omega), e[\vec{s}](\vec{v}_\omega), \ldots, e[\vec{s}](\vec{v}_\omega))^\top \tag{1.14}$$

$$\vec{q}_\omega = (e[\vec{s}](\vec{p}_\omega), e[\vec{s}](\vec{p}_\omega), \ldots, e[\vec{s}](\vec{p}_\omega))^\top \tag{1.15}$$

for all $\omega \in \Omega$. That is, for each $\omega \in \Omega$, $\vec{w}_\omega \in \mathbb{R}^N$ is the equally distributed vector for which the social planner would be indifferent to $\vec{v}_\omega \in \mathbb{R}^N$, if both were certain outcomes in period 2. Similarly, the social planner is indifferent between $\vec{q}_\omega$ with certainty and $\vec{p}_\omega$ with certainty. Therefore, by (W)*

$$\vec{\mathfrak{v}} \overset{\vec{s}}{\sim}_\Omega \vec{\mathfrak{w}} \quad \text{and} \quad \vec{\mathfrak{p}} \overset{\vec{s}}{\sim}_\Omega \vec{\mathfrak{q}} \tag{1.16}$$

and thus, by hypothesis and transitivity of preferences, $\vec{\mathfrak{w}} \overset{\vec{s}}{\succsim}_\Omega \vec{\mathfrak{q}}$. By construction, $\vec{\mathfrak{w}}$ and $\vec{\mathfrak{q}}$ have no variation across the interpersonal dimension so, defining $\mathbf{w} = (w_\omega)_{\omega\in\Omega}, \mathbf{q} = (q_\omega)_{\omega\in\Omega} \in \mathbb{R}^K$ by

$$w_\omega = e[\vec{s}](\vec{v}_\omega) \quad \text{and} \quad q_\omega = e[\vec{s}](\vec{p}_\omega) \tag{1.17}$$

we get that $\mathbf{w} \overset{\vec{s}}{\succsim}_{\mathbb{R}^K} \mathbf{q}$. By (C)*, (SI)* and (S)*, the social preference $\succsim$ satisfies the hypothesis of Proposition 2. Therefore we know that the preference $\overset{\vec{s}}{\succsim}_{\mathbb{R}^K}$ can be represented by the expectation defined by (1.9), so that

$$\frac{1}{K}\sum_{\omega\in\Omega} V[\vec{s}]\left(e[\vec{s}](\vec{v}_\omega)\right) \geq \frac{1}{K}\sum_{\omega\in\Omega} V[\vec{s}]\left(e[\vec{s}](\vec{p}_\omega)\right) \tag{1.18}$$

Hence, (1.13) implies (1.18). By the same argument we can show that $\vec{\mathfrak{v}} \overset{\vec{s}}{\succ}_\Omega \vec{\mathfrak{p}}$ implies

$$\frac{1}{K}\sum_{\omega\in\Omega} V[\vec{s}]\left(e[\vec{s}](\vec{v}_\omega)\right) > \frac{1}{K}\sum_{\omega\in\Omega} V[\vec{s}]\left(e[\vec{s}](\vec{p}_\omega)\right) \tag{1.19}$$

which shows that (1.12) represents the preference $\overset{\vec{s}}{\succsim}_\Omega$ on $\mathbb{R}^{N\times K}$ for every $\vec{s} \in \mathbb{R}^N$. □

1.5 The final representation

The discussion so far has dealt exclusively with the conditional social preference on the set of period 2 outcomes, $\mathbb{R}^{N\times K}$. So far, we have a result on the period two social choice under uncertainty. This section finally incorporates period 1 and the preference across the two time periods.

We propose two different separability conditions – both weaker than full separability – and derive a representation of social choice that can be said to separate the time aggregation from the risk and individual aggregation. However, the weak separability conditions we propose leaves open the possibility that the interpersonal aggregation in the second period may depend explicitly on the achieved distribution in the first period, i.e., there is not full separability between the two periods. To illustrate the difference, consider the following welfare function

$$W = u[e_1(\vec{v}_1)] + \frac{1}{1+\rho} u[CE\left(e_2(\vec{v}_\omega)\right)] \tag{1.20}$$

where $CE : \mathbb{R}^K \to \mathbb{R}$ is the certainty equivalent and $e_1 : \mathbb{R}^N \to \mathbb{R}$ and $e_2 : \mathbb{R}^N \to \mathbb{R}$ are the equally distributed equivalents in periods 1 and 2. The case in which $e_1 \equiv e_2$ is certainly admissible under the representation derived below, but not the only possibility. In fact, the representation below allows for the equally distributed equivalents to differ in both periods, and for the second period equally distributed equivalent to depend on the distribution of welfare in the first period. Such an approach may be desirable from an equity point of view if the same individuals are alive in both periods. Then the social planner may want to put priority in the second period on those individuals who achieved a lower utility in the first period. An example of such a welfare function (admissible under the axioms below) would be one in which

$$e_1 = \phi^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\phi(v_n)\right)$$

and

$$e_2 = \phi^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\left[\phi(v_n)\frac{\sum_{m\neq n} w_m}{(N-1)\sum w_m}\right]\right)$$

where v_n denotes the period 2 utility outcome to individual n and w_n the period 1 utility outcome. The weights

$$\frac{\sum_{m\neq n} w_m}{(N-1)\sum w_m}$$

have the feature that they drop out when there is no inequality in either the first period or the second period, but when there is unequal distribution they favour those individuals who had a less favourable outcome in the first period.

As elaborated at the end of the section, a condition of full separability would make such a feature inadmissible.

1.5.1 Risk and time

Recall that Proposition 2 establishes a social risk attitude contained in the function $V[\vec{s}] : \mathbb{R}^K \to \mathbb{R}$. Define the *certainty equivalent* with respect to $V[\vec{s}]$ of an equally distributed uncertain prospect $\mathbf{a} = (a_\omega)_{\omega\in\Omega} \in \mathbb{R}^K$ by

$$CE[\vec{s}]\left((a_\omega)_{\omega\in\Omega}\right) := V[\vec{s}]^{-1}\left(\frac{1}{K}\sum_{\omega\in\Omega} V[\vec{s}](a_\omega)\right) \tag{1.21}$$

As $V[\vec{s}](\cdot)$ is continuous and strictly increasing, the certainty equivalent is unique and well defined.

Recall that we have been identifying the space of period 2 uncertain outcomes across individuals with $\mathbb{R}^{N\times K}$. Within this space, denote by U_2 the space of outcomes in which all agents get the same utility in every state of the world.

Axiom 8 (I1*: Independence of first period in case of equal certain distribution)**.** *For two vectors $\vec{\mathfrak{v}}, \vec{\mathfrak{w}} \in \mathbb{R}^{N\times K}$, let $\overset{\vec{\mathfrak{v}}}{\succsim}_1$ and $\overset{\vec{\mathfrak{w}}}{\succsim}_1$ be the preferences on the period 1 utility vectors induced by the social preference by fixing $\vec{\mathfrak{v}}$ and $\vec{\mathfrak{w}}$ respectively as the period 2 outcomes. If $\vec{\mathfrak{v}}, \vec{\mathfrak{w}} \in U_2$ then $\overset{\vec{\mathfrak{v}}}{\succsim}_1$ and $\overset{\vec{\mathfrak{w}}}{\succsim}_1$ must be the same.*

Condition (I1)* states that if there is no uncertainty and equally distributed utility amongst all individuals in the second period, then the preference describing the interpersonal trade-off in the first period should not depend on the amount of welfare everyone is getting in the second period. Denote this preference $\succsim_1$ and define

$$e_1 : \mathbb{R}^N \to \mathbb{R} \tag{1.22}$$

to be its equally distributed equivalent representation in analogy to (1.11).

We can now state a complete representation theorem for the social preference $\succsim$ on $\mathbb{R}^I$.

Theorem 1. *Suppose the social preference satisfies Axioms 4, 5, 6, 7 and 8. Let* $e_1 : \mathbb{R}^N \to \mathbb{R}$ *is the period 1 equally distributed equivalent defined by* (1.22)*,* $e : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ *the conditional period 2 equally distributed equivalent defined by* (1.11) *and* $CE : \mathbb{R}^N \times \mathbb{R}^K \to \mathbb{R}$ *the conditional period 2 certainty equivalent defined by* (1.21)*. Then there exists a function* $T : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$*, such that, for all* $\vec{\mathbf{v}} = (\vec{v}_\theta)_{\theta \in \Theta} \in \mathbb{R}^I$

$$W(\vec{\mathbf{v}}) = T\left\{e_1(\vec{v}_1), CE[\vec{v}_1]\left(e[\vec{v}_1](\vec{v}_\omega)_{\omega \in \Omega}\right)\right\} \tag{1.23}$$

The function $T : \mathbb{R}^2 \to \mathbb{R}$ can be said to reflect the collective intertemporal trade-off between certain and equally distributed utility outcomes in the two periods.

Remark. Recall from Proposition 1 that Axioms 1, 2 and 3 ensure that the more general study of social welfare functionals can be reduced to that of social preferences on Euclidean space. Starting with Section 1.3 we have taken these three axioms to hold implicitly and stated conditions and results directly on the social preference, rather than the social welfare functional. In particular, Theorems 1 and 2 implicitly assume that the social preference under consideration is based on a social welfare functional that satisfies Axioms 1, 2 and 3.

Proof. Recall that $U_2 \subset \mathbb{R}^{N \times K}$ is the one dimensional subspace in which all entries are identical. Define the canonical inclusion $l : \mathbb{R} \to U_2$ by

$$l(a) = \vec{\mathbf{a}} = (a, a, \ldots, a) \in \mathbb{R}^{N \times K} \tag{1.24}$$

This function takes a real number and assigns each agent that many utils in each state of the world. Define $\kappa : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}^I$ by

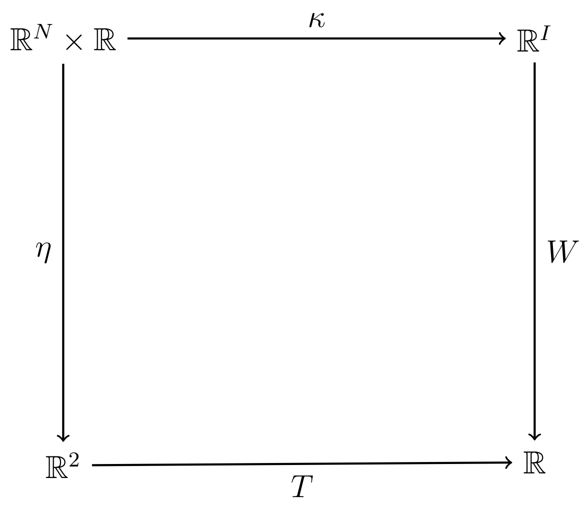


Figure 1.1: Commutative diagram

$$(\vec{v}, a) \mapsto (\vec{v}, l(a)) \in \mathbb{R}^I \tag{1.25}$$

and $\eta : \mathbb{R}^N \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ by

$$(\vec{v}, a) \mapsto (e_1(\vec{v}), a) \in \mathbb{R} \times \mathbb{R}. \tag{1.26}$$

Also, recall that $W : \mathbb{R}^I \to \mathbb{R}$ is the welfare function representing the social preference $\succsim$ defined by (2.22). By (I1)* we have that if $\eta(\vec{v}, a) = \eta(\vec{w}, b)$ then $\kappa(\vec{v}, a) \backsim \kappa(\vec{w}, b)$ and therefore, by (2.22), $W(\kappa(\vec{v}, a)) = W(\kappa(\vec{w}, b))$. Therefore there exists a function $T : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ which makes the diagram in Figure 1.1 commute. That is, for all $\vec{v} \in \mathbb{R}^N$ and $a \in \mathbb{R}$

$$W(\kappa(\vec{v}, a)) = T\left\{e_1(\vec{v}), a\right\} \tag{1.27}$$

Consider an element $\vec{\mathbf{v}} = (\vec{v}_\theta)_{\theta \in \Theta} \in \mathbb{R}^I$ that is in the image of κ. That is, there exist $\vec{v} \in \mathbb{R}^N$ and $a \in \mathbb{R}$ such that $\vec{\mathbf{v}} = (\vec{v}, l(a))$. Since $l(a)$ is equally distributed across both the agents and states in the second period it must be the case that

$$a = CE[\vec{v}_1]\left(e[\vec{v}_1](\vec{v}_\omega)_{\omega \in \Omega}\right) \tag{1.28}$$

and hence by (1.27)

$$W(\vec{\mathbf{v}}) = T\left\{e_1(\vec{v}_1), CE[\vec{v}_1]\left(e[\vec{v}_1](\vec{v}_\omega)_{\omega\in\Omega}\right)\right\} \tag{1.29}$$

It remains to be shown that, for any $\vec{\mathbf{w}} = (\vec{w})_{\theta\in\Theta} \in \mathbb{R}^I$ for which $\vec{w}_1 = \vec{v}_1$ and

$$\vec{\mathbf{w}} \backsim \vec{\mathbf{v}} \tag{1.30}$$

it is also the case that

$$CE[\vec{v}_1]\left(e[\vec{v}_1](\vec{w}_\omega)_{\omega\in\Omega}\right) = CE[\vec{v}_1]\left(e[\vec{v}_1](\vec{v}_\omega)_{\omega\in\Omega}\right) \tag{1.31}$$

Since $\vec{w}_1 = \vec{v}_1$,

$$\vec{\mathbf{w}} \backsim \vec{\mathbf{v}} \iff (\vec{w}_\omega)_{\omega\in\Omega} \overset{\vec{v}_1}{\backsim}_\Omega (\vec{v}_\omega)_{\omega\in\Omega} \tag{1.32}$$

By (1.12)

$$(\vec{w}_\omega)_{\omega\in\Omega} \overset{\vec{v}_1}{\backsim}_\Omega (\vec{v}_\omega)_{\omega\in\Omega} \iff \frac{1}{K}\sum_{\omega\in\Omega} V[\vec{v}_1]\left(e[\vec{v}_1](\vec{v}_\omega)\right) = \frac{1}{K}\sum_{\omega\in\Omega} V[\vec{v}_1]\left(e[\vec{v}_1](\vec{w}_\omega)\right) \tag{1.33}$$

By (1.21), it must be the case that (1.35) holds. This completes the proof. □

This concludes our primary result identifying the conditions necessary for the interpersonal, intertemporal and risk aggregations to be separately identifiable. As indicated by the notation, the period 1 interpersonal aggregation is taken to be independent of the following period, but both the interpersonal and the risk aggregations in period 2 may depend essentially arbitrarily on the first period welfare vector. In order to allow for some degree of interaction between the interpersonal distributions of welfare between the two periods, this feature seems appropriate for the second period equally distributed equivalent $e[\cdot]$. However, it is not altogether clear why the social risk attitude should be allowed to depend on the particular distribution of welfare in the first period. The following strengthening of (I1)* eliminates that feature.

Recall that in Section 1.3 we had identified the space of period 2 outcomes along which there was no variation across individuals, only across states of the world, with $\mathbb{R}^K \subset \mathbb{R}^{N \times K}$.

Axiom 9 (I1'*: Independence of first period in case of equal distribution state by state)**.** *For two vectors $\vec{\mathfrak{v}}, \vec{\mathfrak{w}} \in \mathbb{R}^{N \times K}$, let $\overset{\vec{\mathfrak{v}}}{\succsim}_1$ and $\overset{\vec{\mathfrak{w}}}{\succsim}_1$ the preferences on the period 1 utility vectors induced by the social preference by fixing $\vec{\mathfrak{v}}$ and $\vec{\mathfrak{w}}$ respectively as the period 2 outcomes. If $\vec{\mathfrak{v}}, \vec{\mathfrak{w}} \in \mathbb{R}^K \subset \mathbb{R}^{N \times K}$ then $\overset{\vec{\mathfrak{v}}}{\succsim}_1$ and $\overset{\vec{\mathfrak{w}}}{\succsim}_1$ must be the same.*

We can now state a modification of Theorem 1.

Theorem 2. *Suppose the social preference satisfies Axioms 4, 5, 6, 7 and 9, as well as being based on a social welfare functional satisfying Axioms 1, 2 and 3. Let $e_1 : \mathbb{R}^N \to \mathbb{R}$ is the period 1 equally distributed equivalent defined by* (1.22)*, $e : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ the conditional period 2 equally distributed equivalent defined by* (1.11) *and $CE : \mathbb{R}^N \times \mathbb{R}^K \to \mathbb{R}$ the conditional period 2 certainty equivalent defined by* (1.21)*. Then there exists a function $T : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, such that, for all $\vec{\mathbf{v}} = (\vec{v}_\theta)_{\theta \in \Theta} \in \mathbb{R}^I$*

$$W(\vec{\mathbf{v}}) = T\left\{e_1(\vec{v}_1), CE[\vec{v}_1]\left(e[\vec{v}_1](\vec{v}_\omega)_{\omega \in \Omega}\right)\right\} \tag{1.34}$$

Furthermore, there exists a function $\mathcal{CE} : \mathbb{R} \times \mathbb{R}^K \to \mathbb{R}$ such that, for all $(\vec{v}, \mathbf{a}) \in \mathbb{R}^N \times \mathbb{R}^K$

$$\mathcal{CE}[e_1(\vec{v})](\mathbf{a}) = CE[\vec{v}](\mathbf{a}) \tag{1.35}$$

In addition to the conclusion of Theorem 1, Theorem 2 states that the dependence of the risk attitude on the first period welfare is reduced to a dependence on the first period interpersonal aggregate. That is, the interpersonal distribution of welfare in the first period has no effect on the second period risk attitude.

Proof. The hypothesis is stronger than that of Theorem 1, so the implications therein must hold. That is, a function of the form

$$W = T\left\{e_1(\vec{v}_1), CE[\vec{v}_1]\left(e[\vec{v}_1](\vec{v}_\omega)_{\omega\in\Omega}\right)\right\} \tag{1.36}$$

represents the social preference.

Consider the preference $\succsim_K$ on $\mathbb{R}^N \times \mathbb{R}^K$ induced by the social preference on the restriction to $\mathbb{R}^N \times \mathbb{R}^K \subset \mathbb{R}^I$. Denote a typical element therein by $(\vec{v}, \mathbf{a}) \in \mathbb{R}^N \times \mathbb{R}^K$. By an analogous argument to the one used to establish the commutativity of Figure 1.1 the function $(\vec{v}, \mathbf{a}) \mapsto T[e_1(\vec{v}), CE[\vec{v}](\mathbf{a})]$ must represent $\succsim_K$. Fixing $\mathbf{a} = \mathbf{1}$, the function

$$e_1'(\vec{v}) = t[e_1(\vec{v}), CE[\vec{v}](\mathbf{1})]$$

must, like e_1, represent $\succsim_1$ (see (1.22)). By Theorem 3 in Krantz, Luce, Suppes, and Tversky (1971)

$$t[e_1(\vec{v}), CE[\vec{v}](\mathbf{1})] = \phi\{e_1(\vec{v})\}$$

for some increasing function $\phi : \mathbb{R} \to \mathbb{R}$ and therefore $CE[\cdot](\mathbf{1})$ cannot depend on any aspect in its argument other than that captured by the image of the function $e_1(\cdot)$. Thus, it must be the case that there exists a function $\mathcal{CE} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that

$$\mathcal{CE}[e_1(\vec{s})](\mathbf{a}) = CE[\vec{s}](\mathbf{a}) \tag{1.37}$$

□

This proposition yields a characterisation of the welfare function in which the dependence of the second period interpersonal aggregation, $e[\cdot] : \mathbb{R}^N \to \mathbb{R}$, on its first argument remains arbitrary, but the risk function $V[\cdot] : \mathbb{R} \to \mathbb{R}$ or, more specifically, its certainty equivalent form, may only depend on the aggregate of the period 1 utility vector.

Both (I1)* or (I1')* are weaker than full separability across the time dimension.

Assuming such full separability would further restrict the representation to

$$W = u\{e_1(\vec{v}_1)\} + \frac{1}{1+\rho} u\{CE\left(e_2(\vec{v}_\omega)_{\omega\in\Omega}\right)\}$$

Recall that

$$CE := V^{-1}\left(\mathbb{E}\left[V(e_2(\vec{v}_\omega))\right]\right)$$

Thus, if $V \equiv u$ we get the representation

$$W = u\{e_1(\vec{v}_1)\} + \frac{1}{1+\rho}\mathbb{E}\left[u\left\{e_2(\vec{v}_\omega)\right\}\right]$$

So when the social attitude to risk is identical to its preference for intergenerational inequality, we recover the standard representation.

1.6 Conclusion

The paper fills a gap in the literature, providing a representation result for social decisions in which the time and risk dimensions are important. The result yields a social objective which clearly separates the aggregation across the three dimensions, as well as some minor functional restrictions. We have derived the result for a set of individuals living across two periods, but it is easily reinterpreted for uncertain choice affecting two distinct non-overlapping generations, as is the case in cost-benefit analysis of climate change. For such an objective to be useful in practice a discussion needs to take place concerning the actual parameter values. The literature discussing parameters in social decisions to date has been based on revealed preferences. To a certain extent, that is not possible in this context because individual consumption/savings decisions - which inform the current debate - are not an adequate basis for the parameterisation of an intergenerational trade-off. In some sense the project this paper embarks upon is overly ambitious, since the parameters involved cannot be known for sure and the framework is not able to handle ignorance of the decision parameters. However, other widely used approaches suffer from the same deficiency in addition to conflating dimensions of aggregation by using the same choice parameters for different aspects of the decision.

1.7 Appendix

Axiom 10 (Continuity[10]). *(C): For all* $\vec{\mathbf{v}} \in \mathbb{R}^I$ *and* $x, z \in X$*, the sets*

$$\{\vec{\mathbf{w}} | \vec{u}(x,\cdot) = \vec{\mathbf{v}}, \vec{u}(y,\cdot) = \vec{\mathbf{w}} \;\;\&\;\; xf(\vec{u})y \text{ for some } \vec{u} \in \mathcal{U}\} \tag{1.38}$$

$$\{\vec{\mathbf{w}} | \vec{u}(x,\cdot) = \vec{\mathbf{v}}, \vec{u}(y,\cdot) = \vec{\mathbf{w}} \;\;\&\;\; yf(\vec{u})x \text{ for some } \vec{u} \in \mathcal{U}\} \tag{1.39}$$

are closed.

Axiom 11 (State Independence). *(SI): For any permutation* σ *of* Ω*, if for* $\vec{u}, \vec{u}' \in \mathcal{U}$ $\vec{u}(x,\theta) = \vec{u}'(x,\sigma(\theta))$ *for all* $x \in X$*,* $\theta \in \Theta$*, then* $f(\vec{u}) \equiv f(\vec{u}')$*, where*

$$\sigma(\theta) = \begin{cases} 1, & \text{if } \theta = 1 \\ \sigma(\omega), & \text{if } \theta = \omega \end{cases}$$

The last few conditions require the following preliminaries. For $\vec{u}, \vec{u}' \in \mathcal{U}$, consider the following statements:

$$\vec{u}(\cdot,\theta) = \vec{u}'(\cdot,\theta) \tag{1.40}$$

and, for all $x, z \in X$

$$\vec{u}(x,\vartheta) = \vec{u}(z,\vartheta) \tag{1.41}$$

$$\vec{u}'(x,\vartheta) = \vec{u}'(z,\vartheta) \tag{1.42}$$

[10]Due to Maskin (1978).

Axiom 12 (Separability across states). *(S): For all* $\vec{u}, \vec{u}' \in \mathcal{U}$*, if there exists a subset* $E \subset \Omega$ *such that* (1.40) *holds for all* $\theta \in \{1\} \cup E$ *and* (1.41) *and* (1.42) *hold for* $\vartheta \in \Theta \setminus E$*, then* $f(\vec{u}) \equiv f(\vec{u}')$*.*

Axiom 13 (Independence of first period in case of equal certain distribution). *(I1): For all* $\vec{u}, \vec{u}' \in \mathcal{U}$ *such that* (1.40) *holds for* $\theta = 1$, (1.41) *and* (1.42) *hold for all* $\vartheta \in \Omega$ *and* $u_n(x, \vartheta) = u_m(x, \vartheta')$ *for all* $n, m \in \mathcal{N}$ *and* $\vartheta, \vartheta' \in \Omega$*,* $f(\vec{u}) \equiv f(\vec{u}')$*.*

Axiom 14 (Independence of first period in case of equal distribution state by state). *(I1'): For all* $\vec{u}, \vec{u}' \in \mathcal{U}$ *such that* (1.40) *holds for* $\theta = 1$, (1.41) *and* (1.42) *hold for all* $\vartheta \in \Omega$ *and* $u_n(x, \vartheta) = u_m(x, \vartheta)$ *for all* $n, m \in \mathcal{N}$ *and* $\vartheta \in \Omega$*,* $f(\vec{u}) \equiv f(\vec{u}')$*.*

Theorem 3. *Conditional on (U), (P) and (I), the conditions (C),(SI), (S), (I1) and (I1') are equivalent to (C)*,(SI)*, (S)*, (I1)* and (I1')* respectively.*

Proof. Recall that Lemma 1, due to d'Aspremont and Gevers (1977), yields an order $\succsim$ of $\mathbb{R}^I$ such that for $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in \mathbb{R}^I$, $\vec{\mathbf{v}} \succsim \vec{\mathbf{w}}$ if and only if there exist $x, z \in X$ and $\vec{u} \in \mathcal{U}$ such that

$$\vec{u}(x, \cdot) = \vec{\mathbf{v}}, \quad \vec{u}(z, \cdot) = \vec{\mathbf{w}} \tag{1.43}$$

and

$$x f(\vec{u}) z \tag{1.44}$$

- $(C) \iff (C)^*$: For $\vec{\mathbf{v}} \in \mathbb{R}^I$, consider the set defined in (1.38) and an element $\vec{\mathbf{w}}$ therein. By definition (1.43) and (1.44) hold, and hence $\vec{\mathbf{w}} \in L(\vec{\mathbf{v}})$. Alternatively, assume $\vec{\mathbf{w}} \in L(\vec{\mathbf{v}})$. By (U), there exist $x, z \in X$ and $\vec{u} \in \mathcal{U}$ such that (1.43) and (1.44) hold, and hence $\vec{\mathbf{w}}$ is in the set defined in (1.38). Since they are one and the same set, if one is closed, so it the other. The analogous argument applies to $U(\vec{\mathbf{v}})$ and its corresponding set.

- $(S) \implies (S)^*$: Fix $\vec{v} \in \mathbb{R}_1^N$ along with a subset $E \subset \Omega$ and vectors $\vec{\mathbf{v}} = (\vec{v}_\theta), \vec{\mathbf{w}} = (\vec{w}_\theta) \in \mathbb{R}^I$, which are equal to each other on $F = \Omega \backslash E$ as well as equal to $\vec{v}$ on $\{1\}$. That is, $\vec{\mathbf{v}}_F = \vec{\mathbf{w}}_F$ and $\vec{v}_1 = \vec{w}_1 = \vec{v}$. Without loss of generality, let $\vec{\mathbf{v}} \succsim \vec{\mathbf{w}}$. Consequently, $\vec{\mathbf{v}}_\Omega \overset{\vec{v}}{\succsim}_\Omega \vec{\mathbf{w}}_\Omega$ and there exist an individual welfare function $\vec{u} \in \mathcal{U}$ and $x, z \in X$ such that $\vec{u}(x, \theta) = \vec{v}_\theta$, $\vec{u}(z, \theta) = \vec{w}_\theta$ for all $\theta \in \Theta$ and $xf(\vec{u})z$. Now consider $\vec{u}' \in \mathcal{U}$ such that the antecedent of (S) is satisfied for $\vec{u}$ and $\vec{u}'$ and $E \subset \Omega$. Define $\vec{\mathbf{v}}' = (\vec{v}'_\theta) \in \mathbb{R}^I$ and $\vec{\mathbf{w}}' = (\vec{w}'_\theta) \in \mathbb{R}^I$ by $\vec{v}'_\theta = \vec{u}'(x, \theta)$ and $\vec{w}'_\theta = \vec{u}'(z, \theta)$ for all $\theta \in \Theta$. We then have that $\vec{\mathbf{v}}' \succsim \vec{\mathbf{w}}'$, $\vec{v}'_1 = \vec{w}'_1$, $\vec{\mathbf{v}}'_E = \vec{\mathbf{v}}_E$, $\vec{\mathbf{w}}'_E = \vec{\mathbf{w}}_E$ and $\vec{\mathbf{v}}'_F = \vec{\mathbf{w}}'_F$. We can conclude that $\vec{\mathbf{v}}_E$ is ranked above $\vec{\mathbf{w}}_E$ by the order induced on $\prod_{\theta \in E} \mathbb{R}^N_\theta$ by $\overset{\vec{v}}{\succsim}_\Omega$, independent of what the value of $\vec{\mathbf{v}}_F = \vec{\mathbf{w}}_F$ is. Since $\vec{\mathbf{v}}_E$ and $\vec{\mathbf{w}}_E$ were chosen arbitrarily, as was E, then entire order is independent for every subset E of Ω and $\overset{\vec{v}}{\succsim}_\Omega$ is strongly separable.

- $(S) \Longleftarrow (S)^*$: Let $\vec{u}, \vec{u}' \in \mathcal{U}$ satisfy the antecedent of (S) for some $E \subset \Omega$. Let the vectors $\vec{\mathbf{v}}, \vec{\mathbf{w}}, \vec{\mathbf{v}}'$ and $\vec{\mathbf{w}}' \in \mathbb{R}^I$ be equal to $\vec{u}(x, \cdot), \vec{u}(z, \cdot), \vec{u}'(x, \cdot)$ and $\vec{u}'(z, \cdot)$ respectively, for some $x, z \in X$. It is clear that $\vec{v}_1 = \vec{w}_1 = \vec{v}'_1 = \vec{w}'_1$. Assume, without loss of generality, that $xf(\vec{u})z$ and hence $\vec{\mathbf{v}} \succsim \vec{\mathbf{w}}$. $\vec{\mathbf{v}}_F = \vec{\mathbf{w}}_F$ for $F = \Omega \setminus E$ while $\vec{\mathbf{v}}_E = \vec{\mathbf{v}}'_E$, $\vec{\mathbf{w}}_E = \vec{\mathbf{w}}'_E$ and $\vec{\mathbf{v}}'_F = \vec{\mathbf{w}}'_F$ so, by strong separability of $\overset{\vec{v}_1}{\succsim}_\Omega$, $\vec{\mathbf{v}}' \succsim \vec{\mathbf{w}}'$ and therefore $xf(\vec{u}')z$. This holds for every pair $x, z \in X$ and hence $f(\vec{u}) \equiv f(\vec{u}')$.

- $(I1) \implies (I1)^*$: Consider vectors $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in \mathbb{R}^N \times \mathbb{R}_2 \subset \mathbb{R}^I$ for which $\vec{\mathbf{v}}_\Omega = \vec{\mathbf{w}}_\Omega \in \mathbb{R}_2$. Assuming $\vec{\mathbf{v}} \succsim \vec{\mathbf{w}}$ we have that $\vec{v}_1 \overset{\vec{\mathbf{v}}_\Omega}{\succsim}_1 \vec{w}_1$. Let $\vec{u} \in \mathcal{U}$ be such that $\vec{u}(x, \cdot) = \vec{\mathbf{v}}$ and $\vec{u}(z, \cdot) = \vec{\mathbf{w}}$, and choose $\vec{u}' \in \mathcal{U}$ such that the antecedent of (I1) holds for $\vec{u}$ and $\vec{u}'$ and consequently, $xf(\vec{u}')z$. Again, let $\vec{\mathbf{v}}' = \vec{u}'(x, \cdot)$ and $\vec{\mathbf{w}} = \vec{u}(z, \cdot)$. Since (1.40) holds for $\theta = 1$, we have that

$$\vec{v}_1 = \vec{v}'_1 \text{ and } \vec{w}_1 = \vec{w}'_1 \tag{1.45}$$

Since $\vec{\mathbf{v}}' \succsim \vec{\mathbf{w}}'$, (1.42) holds for all $\vartheta \in \Omega$, and $\vec{\mathbf{v}}'_\Omega = \vec{\mathbf{w}}'_\Omega \in \mathbb{R}_2$ we have that $\vec{v}_1 \overset{\vec{\mathbf{v}}_\Omega}{\succsim}_1 \vec{w}_1$ is equivalent to $\vec{v}_1 \overset{\vec{\mathbf{w}}_\Omega}{\succsim}_1 \vec{w}_1$ whenever both $\vec{\mathbf{v}}_\Omega$ and $\vec{\mathbf{w}}_\Omega$ are in $\mathbb{R}_2 \subset \prod_{\omega \in \Omega} \mathbb{R}^N_\omega$.

Since the argument is symmetric, and $\vec{v}_1$ and $\vec{w}_1$ were chosen arbitrarily, we get that $\overset{\mathfrak{v}}{\succsim}_1$ and $\overset{\mathfrak{w}}{\succsim}_1$ must be the same whenever $\mathfrak{v}, \mathfrak{w} \in \mathbb{R}_2$.

- $(I1) \Longleftarrow (I1)^*$: The method is analogous to the one used to proof that $(S) \Longleftarrow (S)^*$.
- $(I1') \iff (I1')^*$: This follows almost immediately from the proof that $(I1) \iff (I1)^*$.
- $(SI) \iff (SI)^*$: For this standard result see d'Aspremont and Gevers (1977).

□

Bibliography

Arrow, K. J. (1951): "Social choice and individual values.," *Cowles Commission Mongraph No. 12.*

Blackorby, C., D. Donaldson, and P. Mongin (2004): "Social Aggregation Without the Expected Utility Hypothesis," Working Paper hal-00242932, HAL.

Bommier, A., A. Chassagnon, and F. Le Grand (2012): "Comparative risk aversion: A formal approach with applications to saving behavior," *Journal of Economic Theory*, 147(4), 1614–1641.

Broome, J. (1991): *Weighing goods.* Basil Blackwell, Oxford.

d'Aspremont, C., and L. Gevers (1977): "Equity and the Informational Basis of Collective Choice," *The Review of Economic Studies*, 44(2), 199–209.

Debreu, G. (1954): "Representation of a preference ordering by a numerical function," *Decision processes*, pp. 159–165.

Debreu, G. (1960): "Topological methods in cardinal utility theory," *Mathematical methods in the social sciences*, 1959, 16–26.

Diamond, P. A. (1967): "Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment," *The Journal of Political Economy*, 75(5), 765–766.

Epstein, L. G., and S. E. Zin (1989): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework," *Econometrica: Journal of the Econometric Society*, p. 937969.

Fleurbaey, M. (2010): "Assessing Risky Social Situations," *The Journal of Political Economy*, 118(4), 649–680.

Gorman, W. M. (1968): "The structure of utility functions," *The Review of Economic Studies*, 35(4), 367–390.

Hammond, P. (1983): "Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty.," in *Social Choice and Welfare*, ed. by P. K. Pattanaik, and M. Salles, Contributions to economic analysis, pp. 175–205. North-Holland Publishing Company.

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

Koopmans, T. C. (1960): "Stationary ordinal utility and impatience," *Econometrica*, 28, 287–309.

Krantz, D. H., R. D. Luce, P. Suppes, and A. Tversky (1971): *Foundations of measurement: Additive and polynomial representations*, vol. 1. Academic Press.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46(1), 185–200.

Maskin, E. (1978): "A theorem on utilitarianism," *The Review of Economic Studies*, 45(1), 93–96.

Roberts, K. W. (1980): "Interpersonal comparability and social choice theory," *The Review of Economic Studies*, 47(2), 421–439.

Savage, L. J. (1954): "The foundations of statistics," *New York*.

Selden, L. (1978): "A new representation of preferences over "certain x uncertain" consumption pairs: The "ordinal certainty equivalent" hypothesis," *Econometrica*, 46(5), 1045–1060.

Sen, A. K. (1970): *Collective choice and social welfare*. Holden-Day, Inc., San Francisco.

Zuber, S. (2009): "Harsanyi's Theorem without the Sure-Thing Principle," Discussion paper, mimeo.

Chapter 2

Inequality in Climate Change: a modification of RICE

Chapter abstract

I modify Nordhaus' RICE-2010 model with the aim of performing the policy optimisation in such a way that the effects of regional inequality are accounted for as well as intergenerational inequality. I introduce a social objective that allows for different degrees of inequality aversion along the regional dimension. Comparing along this dimension, I find that a full account of the effects of policy on regional inequality reduces the optimal carbon tax relative to treating the world as an aggregate consumer. Furthermore, the way in which the social objective is introduced allows for a separation of key normative parameters such as the discount rate and coefficient of inequality aversion from those used to model the savings decision, thereby allowing us to apply normative parameters that are different from Nordhaus'.

2.1 Introduction

2.1.1 Integrated assessment with regional inequality

Integrated assessment models (IAMs) weigh damages and benefits of greenhouse gas (GHG) emissions to determine *optimal* climate policy. The prescriptions are optimal to a consequentialist global social planner. The analogous method for policy appraisal at a national level has been common since the 1970s, but, despite a short-lived academic debate, distributional concerns have been largely omitted in its implementation (See Mirrlees (1978)). The methodology prescribes that the social planner's objective be concave across the temporal dimension but linear across the spatial dimension. The inequality aversion across time is not matched by inequality aversion across space. Linearity across space makes the methodology much simpler, and complementary policies may be implemented to attain spatial redistribution if, indeed, the policy in question raises any distributional concerns at all. Little and Mirrlees (1974) contains further justifications for this simplification. In his early comprehensive economic analysis of global warming, Schelling (1992) highlights, amongst other things, that the effects of climate change will be unequally distributed over an already unequal world. To this we can add that the costs of averting climate change will most likely be unequally distributed as well. Furthermore, the commonly valid reasons for ignoring distributional issues in national and subnational project appraisal are not equally valid in assessing global climate policy.

This paper modifies William Nordhaus' regionally disaggregated RICE model to determine optimal emission policy which takes full account of regional inequality. Regional inequality aversion is introduced to climate modeling by Azar and Sterner (1996) through the use of equity weights. They calculate the social cost of carbon (SCC) for a doubling of global CO_2 concentrations in a stylised two-region model.

Fankhauser, Tol, and Pearce (1997) calculate the equity weighted SCC for the same path but for a greater number of regions, based on estimates from the Second Assessment report of the IPCC. Hope (2008) uses the PAGE2002 model (see Hope (2006)) to apply equity weighting to the calculation of the SCC for the baseline scenario A2 from the IPCC SRES. Most recently, Anthoff, Hepburn, and Tol (2009) use the FUND model and apply equity weighting to calculate the SCC for scenarios SRES A1b, A2, B1, B2 and the FUND default scenario. All but Hope (2008) compare the outcomes with and without equity weights, while leaving the discount rate and the implicit intergenerational equity weighting therein intact. With the exception of the specification with a very high pure rate of time preference in Anthoff, Hepburn, and Tol (2009), equity weighting uniformly increases the social cost of carbon relative to the counterfactual.

Tol (2001) uses the FUND model to calculate the optimal emissions path as we do, rather than just the social cost of carbon. That means that the impact of the policy is taken into account in the assessment, rather than just the impact of a marginal unit of emission. In addition to being based on a different model, Tol's analysis distinguishes itself from the following in that it's social planner's objective consists of an inequity averse transformation applied to the already intertemporally aggregated regional utilities. Such a welfare function is problematic because it depends on the normalisation chosen for the regional utilities and it does not treat hypothetically equally well of, contemporaneous individuals in different regions equally.[1]

[1] To see this, consider a simplified two-region {A,B}, two-period {1,2} model. For a given scenario individuals in the first period differ in their consumption, say $c_{A1} \neq c_{B1}$, and individuals in the second period have identical consumption, say $c_{A2} = c_{B2}$. Without loss of generality, an example – without discounting – of the objective used by Tol (2001) would be

$$W = \sqrt{\log(c_{A1}c_{A2})} + \sqrt{\log(c_{B1}c_{B2})}$$

One would expect that a marginal reduction in the consumption of A2 would be socially equivalent to the same marginal reduction in the consumption of B2, both being equally well off, but as can be seen by differentiating W with respect to those consumptions, this is generically not the case.

2.1.2 Behavioural objective

The original RICE model also uses a social objective that is not suited for an analysis focusing on regional inequality. In the model, every region i is endowed with an objective function over its own consumption stream similar to that of the Ramsey agent's,[2] which governs savings behaviour:

$$\mathcal{L}_i = \sum_{t=1}^{T} \left(\frac{1}{1+\varrho} \right)^t L_{it} u(c_{it}); \quad u(x) = \frac{x^{1-\gamma}}{1-\gamma} \tag{2.1}$$

This objective is *also* taken to be the standard whereby region i assesses the effects of climate change policy to itself. The RICE methodology combines these into a global social objective over the consumption streams of all regions, $\vec{\mathbf{c}} = (c_{it})_{i \leq I, t \leq T}$, whereby the optimal carbon tax is determined:

$$W(\vec{\mathbf{c}}) = \sum_{t=1}^{T} \left(\frac{1}{1+\varrho} \right)^t \sum_{i=1}^{I} \nu_{it} L_{it} u(c_{it}) \tag{2.2}$$

where the weights ν_{it} are known as *time varying Negishi weights*. The weights are constructed so that the intergenerational trade-off is not affected, but the trade-off between regions is altered by putting more welfare weight on more affluent regions. So despite exhibiting concavity in both dimensions, the welfare function in the original RICE model is not adequate for an assessment focusing on distribution along the spatial dimension *as well as* the temporal dimension.

A further consequence of the approach in RICE is that the observed savings behaviour determines the parameter range for the policy assessment. The pure rate of time preference, ϱ and the elasticity of intertemporal substitution, γ, are codetermined so that observed savings behaviour is modeled as optimal at historic interest rates. Different versions of RICE use different value pairs, but they have always

[2]The original Ramsey model does not include a term for population size since the representative agent is modeling individual rather than social behaviour.

been chosen to fit the Ramsey equation. Since (2.2) is simply an amalgamation of the Ramsey objectives (2.1), the same parameters that govern savings decision also determine mitigation policy.

Kaplow, Moyer, and Weisbach (2010) make the point that the savings decision and the decision of how to deal with the climate externality are distinct and need not be governed by the same objective functions or parameters. As noted in Hepburn and Beckerman (2007) the former concerns a single individual at different points in time, whereas the latter concerns different individuals in different regions and different generations: unless, as pointed out in Schelling (1995), we are actually dealing with a single immortal agent. This is at the root of the debate regarding the parameter values to use for welfare analysis. The authors of the Stern Review choose normatively appealing parameters for their welfare function, which would not result in realistic savings rates if the same parameters were also used to determine these endogenously.[3] In a globally aggregated context, Below (2011) uses an approximation to Nordhaus' aggregate DICE model due to Golosov, Hassler, Krusell, and Tsyvinski (2011) to distinguish the social pure rate of time preference used for intergenerational choice from the behavioural pure rate of time preference used for personal saving decisions of individuals. This allows him to calculate saving rates that are realistic, while at the same time allowing him to independently choose the social time preference based on normative considerations.

2.1.3 Social objective

This analysis takes two additional steps. It builds on RICE, rather than DICE, allowing us to look at the regional effects. Furthermore, it separates the social and positive objectives completely. That is, it keeps objective (2.1) to accurately describe saving behaviour, but replaces the social objective (2.2) with a completely different objec-

[3] The PAGE2002 model used for the assessment in the Stern Review does not endogenise the savings decision.

tive function, allowing the independent choice of interregional and intergenerational inequality aversion as well as the discount rate. Atkinson, Dietz, Helgeson, Hepburn, and Saelen (2009) note that in a model with regional disaggregation the elasticity parameter must perform the two[4] separate tasks of quantifying aversion to inequality across the spatial as well as the temporal dimension. The normative view is taken here that the two should, in fact, be the same, but the following social objective is used, which allows for an *a priori* distinction between the two, so that comparisons can be made between different levels of inequality aversion along one dimension while keeping the parameter along the other dimension fixed:

$$W(\vec{\mathbf{c}}) = \sum_{t=1}^{T} \left(\frac{1}{1+\rho}\right)^{t} \phi\left[\check{c}_t\right] \cdot \sum_{i=1}^{I} L_{it} \tag{2.3}$$

For every generation t, the argument of ϕ is the *equally distributed equivalent consumption* of that generation and takes the shape

$$\check{c}_t = \psi^{-1}\left(\frac{\sum_{i=1}^{I} L_{it}\psi(c_{it})}{\sum_{i=1}^{I} L_{it}}\right) \tag{2.4}$$

Taking the functions ϕ and ψ to be of constant elasticities η and ξ respectively, you get the functional form

$$\phi(c_{it}) = \frac{c_{it}^{1-\eta}}{1-\eta}$$

and similarly for ψ. In this case the objective (2.3) consists of three separate *normative* determinants of choice: the parameters ρ, η and ξ. The first is the well known social pure rate time preference. The other two are the coefficients of the social preference for equal distribution of consumption between distinct generations and within generations. In absence of inequality within generations, i.e., when there is no variation in consumption across the index i in any period t, (2.4) simply yields the

[4] When the models include risk, there are actually three sources of aversion to unequal distribution that the single parameter must specify.

average – equally distributed – consumption in every period. In that case it is easy to interpret the function ϕ as the utility from consumption attributed to individuals by the social planner for the purpose of trading-off consumption across generations. Combined with the discount factor, this yields a social welfare function characterised by impatience, as axiomatised in Koopmans (1960), *for the case of no inequality within generations.*

The function ψ represents the utility from consumption attributed to individuals by the social planner for the purpose of trading off consumption between contemporaries. If ψ is linear, i.e. $\xi = 0$, then $\check{c}_t$ is simply the average global consumption. Using such a parameter value in a disaggregated model is equivalent to considering only global aggregate consumption without regard for distributional issues between contemporaries. If ψ is equal to ϕ, i.e. $\xi = \eta$, then (2.3) becomes a function in which the intergenerational trade-off and the intra-generational trade-off become the same.

The modification of RICE is run as follows. The social planner sets the path of globally uniform carbon taxes. These carbon taxes are taken as given by the individual regions and they modify their behaviour with respect to GHG emissions accordingly. Each region abates optimally given the tax, i.e., by such an amount that the marginal cost of an additional unit is equal to the tax. Given this, and the effects the regions' actions have on the climate and the economy, every region optimises its saving behaviour according to the objective (2.1). The optimal tax path is that which maximises (2.3) taking into account the effects this has on regional behaviour.

As highlighted in the debate sparked by the Stern Review,[5] climate policy prescriptions are not only sensitive to the discount rate, but also to the elasticity of intergenerational inequality. Like the discount rate, a lower elasticity has the effect of inducing higher optimal carbon taxes. Put differently, taxes increase intergenerational inequality. This is because mitigation makes future generations better off at

[5]See Nordhaus (2007), Dasgupta (2007) and Dietz (2008).

the expense of the current mitigating generations, and the former are assumed to be wealthier than the latter. The results below reflect that intuition over the entire range of elasticities studied.

2.1.4 Major findings

The separation of the contemporaneous aggregation from the intergenerational one in the social objective (2.3) makes it possible to determine the sensitivity of the optimal carbon taxes to different degrees of inequality aversion along this dimension as well. In magnitude, the effect of varying this parameter is smaller than the equivalent variation of intertemporal inequality aversion. The direction is determined by the following effects. As demonstrated by the results applying equity weights to the calculation of the social cost of carbon, the poorer Global South stands to gain more from mitigation as the climate damages they stand to avoid are larger, so greater carbon taxes reduce inequality in the distant future. However, the forgone productivity due to mitigation via a globally uniform carbon tax is also estimated to be greater in the Global South, increasing inequality during the mitigating present and near future. I assume that no additional inter-regional redistributive measures take place; neither to redistribute the costs of mitigation, nor those of climate damages, making it necessary to account for the effect of both on regional distribution. The net effect of including regional inequality aversion is to *lower* optimal taxes relative to the optimisation without inequality aversion. For example, at a coefficient of intergenerational inequality aversion of $\eta = 1.5$, increasing the coefficient of regional inequality aversion from $\xi = 0$ to $\xi = 1.5$ reduces the optimal carbon tax by 20% in the first period.

If the parameters are chosen to match the the ones chosen in RICE-2010, i.e. $\rho = 1.5\%$, and both elasticities equal to 1.5, objectives (2.2) and (2.3) become equivalent in their approach to the intergenerational trade-off, but remain distinct in their approach to the intragenerational trade-off, due to the welfare weights ν_{it} that feature in (2.2). Despite this disparity, the results are remarkably similar. Both policies start at around \$37 per tC and rise constantly to reach full mitigation in the same period, differing no more than 5% for the entire 21st century and not significantly more thenceforth. This is surprising because it was expected that there would be a significant difference between the optimisation with Negishi weights and the one without.

One of the benefits of this approach is that the ethical parameters need not be determined by the same restrictions as the positive parameters, allowing their choice to be guided by other principles. In choosing the coefficient η, I follow Cline (1992) and the intuition provided by an example from Dasgupta (2008), which yields the value $\eta = 1.5$. Recall that setting the coefficient ξ to zero is equivalent to treating the whole world as an aggregate consumer; just the approach this analysis sets out to avoid. Equivalent treatment of the inter- and intratemporal trade-offs would require equal values for η and ξ, making both equal to 1.5. Based on the anonymity principle, moderated by a small and constant risk of extinction, I use an extremely low discount rate of 0.1%, as proposed by the Stern Review. With these parameters the optimisation yields an optimal tax policy that starts at \$113 per tC, rising steadily to reach full mitigation by the end of this century. Such a policy amounts to significantly stronger mitigation than proposed by RICE (See Nordhaus (2010)).

2.1.5 Structure of the analysis

The analysis is structured into a description of the original RICE model (Section 2.2), the modifications made to it, in particular the parameterisation of the social objective (Section 2.3), the analysis of the optimal policies (Section 2.4) and concluding remarks (Section 2.5).

2.2 RICE

2.2.1 The model

Since 1999 William Nordhaus has incrementally developed three versions of the RICE model resulting in RICE-2010, the version this analysis is based on. The model consists of twelve autarkic, neoclassical, single-good production economies representing the twelve amalgamated regions[6] from 2005 to 2605 in ten year time steps. This section explains the model, but a more detailed description is relegated to the Appendix.

The consumption-saving decision is endogenised and chosen, region by region, by a representative agent with the following felicity function

$$u(x) = \frac{x^{1-\gamma}}{1-\gamma}; \quad \gamma = 1.5 \tag{2.5}$$

and a behavioural pure rate of time preference of $\varrho = 1.5\%$ per year. The joint choice of ϱ and γ reflects the view that the representative agent chosen to model behaviour ought to be consistent with observed behaviour according to the Ramsey model, which states that the proportion of investment out of global output is determined by a singe representative agent solving an infinite horizon optimal control problem. The solution to the model is described by the Ramsey equation, (2.6), in which the parameters ϱ and γ determine the relationship between the consumption growth rate and the interest rate, r_t:

$$r_t = \varrho + \frac{\dot{L}_t}{L_t} + \gamma \frac{\dot{c}_t}{c_t} \tag{2.6}$$

This is the result of a variational argument, the essence of which is that if equality in (2.6) doesn't hold, the representative agent would always be able to make a gain by changing the investment path such that it does. Given that interest rates and

[6]The regions are USA, OECD-Europe, Japan, Russia, non-Russia Eurasia, China, India, Middle East Africa, Latin America, Other High Income countries, non-OECD Asia.

consumption growth rates are observed,[7] (2.6) can be used to pin down a relationship between the behavioural parameters ϱ and γ.

The populations are assumed to grow exogenously, at slightly different rates, and eventually stabilise. Denoting region i's population in period t by L_{ti} the representative agent has the following objective

$$\mathcal{L}_i = \sum_{t=1}^{T} \left(\frac{1}{1+\varrho} \right)^t L_{it} u(c_{it}) \tag{2.7}$$

defined on region i's per capita consumption stream $(c_{i1}, c_{i2}, \ldots, c_{iT})$.

Gross output for region i in period t is simply the product of capital, K_{it}, and labour, L_{it}, of the neoclassical production function.[8]

$$Y_{it} = F_{it}(K_{it}, L_{it}) \tag{2.8}$$

Emissions of a region are proportional to gross output. The mitigation rate, μ_{it}, is the proportion of emissions reduced relative to the amount that would be emitted if fossil fuel consumption were determined only by its factor productivity and relative scarcity: economic determinants summarised in the exogenous emissions to output ratio, σ_{it}.

$$E_{it} = \sigma_{it}(1 - \mu_{it})Y_{it} \tag{2.9}$$

Total period t emissions are the sum of the regional emissions plus the *exogenous* amount of emissions due to land use changes, EL_t.

$$E_t = \sum_{i=1}^{12} E_{it} + EL_t \tag{2.10}$$

Y_{it} being the gross output, the *net* output takes into account mitigation costs and

[7] The population growth rate is also observed.

[8] Note that population is exogenous, so the second argument is not a choice variable.

climate damages:[9]

$$Q_{it} = \frac{1 - \Lambda_{it}}{1 + D_{it}} Y_{it} \tag{2.11}$$

Here the numerator quantifies the cost of mitigation - as a proportion of output - to region i in period t. The denominator quantifies the cost of the climate damages to region i in period t.

The mitigation cost as a proportion of gross output is defined by the convex function of the mitigation rate,

$$\Lambda_{it} = \theta_{it}^1 \mu_{it}^{\theta_2} \tag{2.12}$$

where $\theta_2 = 2.8$ and θ_{it}^1 is an exogenous parameter, decreasing in time at an exogenous rate and calibrated in 2005 to equate the marginal cost of the last unit of mitigation with the estimated price of a completely green backstop technology.

The regional damage functions are quadratic in atmospheric global mean temperatures above the preindustrial average:[10]

$$D_{it} = \alpha_i^0 + \alpha_i^1 T_{at} + \alpha_i^2 T_{at}^2 \tag{2.13}$$

The temperature, which is the argument of the damage function,[11] is determined by a very simple climate module, modeling the greenhouse effect linking atmospheric concentrations of greenhouse gases to global temperature.

The RICE methodology consists of solving the Ramsey savings problem for the 12 regions given equations (2.5) to (2.13). The optimal savings rates, s_{it}^* are determined *in absence of mitigation*, i.e. $\mu_{it} = 0$, $\forall i, t$. This is known as the *baseline* run. Given

[9] This is actually the functional form used in previous RICE models. RICE 2010 has replaces $1/(1 + D)$ with $(1 + D^{10} - D)/(1 + D^{10})$. I have implemented the old specification below.

[10] RICE-2010 adds an additional damage term due to sea level rise, which I have not implemented.

[11] The presumption here is that the extent of global warming, as it is expected to affect the climate, can be adequately summarised by one statistic: the global temperature averaged over a year.

these savings rates, the baseline consumptions are defined by

$$c^*_{it} = \frac{(1 - s^*_{it})Q_{it}}{L_{it}}. \tag{2.14}$$

This baseline is then used to determine the relative weights of the welfare function, defined as the inverse of the marginal utility of consumption *at the baseline consumption level*:

$$\nu_{it} = \frac{u'(c^*_{it})^{-1}}{\sum_{j=1}^{I} u'(c^*_{jt})^{-1}} \tag{2.15}$$

These ν_{it} are labeled *time-varying Negishi* weights, as they are similar to weights resulting from a procedure described in Negishi (1960). Given these weights, the mitigation policy is chosen according to the following welfare function:

$$W^{Nord} = \sum_{t=1}^{T} \left(\frac{1}{1+\varrho}\right)^t \sum_{i=1}^{I} \nu_{it} L_{it} u(c_{it}) \tag{2.16}$$

Mitigation policy consists of choosing non-zero mitigation levels μ_{it}. RICE focuses on policies such that the resulting marginal cost of mitigation is equalised across regions. Given a carbon tax τ_t in period t, the mitigation rates for every region i are taken to be:

$$\mu_{it} = \left(\frac{\tau_t \sigma_{it}}{\theta^1_{it}\theta_2}\right)^{\frac{1}{\theta_2 - 1}} \tag{2.17}$$

These are determined under the rationale that given the tax rate τ_t, region i will choose that mitigation rate which maximises its economic output gross of climate damages and net of mitigation costs

$$(1 - \theta^1_{it}\mu_{it}^{\theta_2})Y_{it} - \tau_t(1 - \mu_{it})\sigma_{it}Y_{it} \tag{2.18}$$

The argmax to (2.18) is (2.17).[12]

[12]This is an approximation, because there ought to be an additional term due to damages in (2.18), which is ignored, presumably for numerical efficiency.

2.2.2 Global tax and second best

In a world of differentiated abatement technologies, equalising the marginal abatement cost across regions (via a uniform tax or other means) has the effect of minimising the total cost for a give level of abatement. Given the possibility of lump-sum transfers, equal marginal abatement costs implied by (2.17) are optimal. Chichilnisky and Heal (1994) show in a static model that this can be far from optimal in when inter-regional lump-sum transfers are not possible. Anthoff (2011) expands this result to a dynamic model and uses the FUND model to find the second best policy with differentiated marginal abatement costs under the restriction of no lump-sum transfers. He finds that despite the large regional variation in carbon taxes at the optimum, the total abatement effort is *lower* when regional inequality aversion is included, relative to the globally aggregated analysis. This mirrors our result highlighted in Figure 2.7, even though our analysis imposes globally uniform abatement costs.

The reason for this restriction is that any policy in which the marginal abatement costs differ across regions will lead to distortions that are unlikely to be sustainable. Differential carbon taxes will result in the relocation of industry, and an allocation of carbon credits that strongly benefits poorer regions will lead to capital outflows that are equally politically unsustainable. Therefore, if one is to consider the second best in which the political economy of the negotiation process does not allow for international transfers, the equal marginal abatement cost assumption seems like a reasonable benchmark.

2.3 The changes to RICE

2.3.1 Identification of the changes

The complete description of the economy, the emissions and the climate module are left unchanged. All the functions and parameters describing these, as well as the behavioural parameters in (2.7) are kept just as in RICE-2010. The damage function is changed slightly in that I omit the damages due to sea-level rise[13] and the functional form is that of previous RICE models rather than that of RICE-2010.[14]

The main change is that the welfare function (2.16) is replaced with an objective that has a more general specification for the trade-off between regions, while omitting the Negishi weights.

2.3.2 Decision structure and sequence

Recall that per-capita consumption is related to other variables by

$$c_{it} = \frac{(1 - s_{it})Q_{it}}{L_{it}} \tag{2.19}$$

Define $\mathcal{L}_i(\vec{s}_i, \vec{\mathbf{s}}_{-i}, \vec{\tau})$ as the decision utility (2.7) of the representative agent in region i when the carbon tax path is $\vec{\tau}$, region i's savings rate stream is $\vec{s}_i$ and the other regions' savings rate streams are $\vec{\mathbf{s}}_{-i} = (\vec{s}_1, \ldots, \vec{s}_{i-1}, \vec{s}_{i+1}, \ldots, \vec{s}_{12})$. To see that $\mathcal{L}_i(\vec{s}_i, \vec{\mathbf{s}}_{-i}, \vec{\tau})$ is well defined we must establish that its arguments completely determine the per capita consumption stream of region i, since that is the only argument of (2.7). By (2.17) the tax path determines the mitigation rate streams. The combined saving rates, $(\vec{s}_i, \vec{\mathbf{s}}_{-i})$ determine global gross output which, in turn, determines emissions and through the

[13]The sea level rise (SLR) module was introduced in RICE-2010 without full documentation describing the exact implementation. The supplement to Nordhaus (2010) indicates that the difference between previous versions that don't include the (SLR) module are very small.

[14]At the damage levels attained the difference between the two specifications is less than half a percent, so I took the old specification as sufficient.

carbon and temperature cycles, the damages. The damage and mitigation streams determine net output, which yields the consumption streams as the proportion that is not saved.

It is important to notice that, given a fixed tax path, any region's utility depends on the savings rate streams of all the other regions. This is because the externality affects all regions through the increased temperature, which is determined jointly by the emissions of all regions, and these depend on regional outputs which, in turn, depend on the capital accumulated through the saving rates. Therefore, the choice of saving rates is strategic. Given the decision utilities $\mathcal{L}_i(\vec{s}_i, \vec{\mathbf{s}}_{-i}, \vec{\tau})$ for every tax path $\vec{\tau}$, we can define the saving rate streams that constitute Nash Equilibrium strategies by $\vec{\mathbf{s}}^*(\vec{\tau}) = (\vec{s}_1^*(\vec{\tau}), \ldots, \vec{s}_{12}^*(\vec{\tau}))$ such that for all regions i

$$\mathcal{L}_i(\vec{s}_i^*(\vec{\tau}), \vec{s}_{-i}^*(\vec{\tau}), \vec{\tau}) \geq \mathcal{L}_i(\vec{r}, \vec{s}_{-i}^*(\vec{\tau}), \vec{\tau}), \quad \forall \vec{r} \in [0,1]^T \tag{2.20}$$

This is a standard fixed-point problem for the set of saving strategies of the regions, parameterised by the tax vector.

Having defined how the saving rate streams are determined as a function of the tax paths, we can write down the social planner's problem:

$$\begin{aligned} \max_{\vec{\tau}} W_{\phi,\psi}(\vec{\mathbf{c}}) \quad \text{s.t.} \\ c_{it} = \frac{(1 - s_{it}^*(\vec{\tau}))Q_{it}}{L_{it}} \end{aligned} \tag{2.21}$$

The approach taken here is in the spirit of a principle-agent problem. The social planner chooses a tax such that the social objective is maximised, under the constraint that the tax will affect the outcome, not just directly, but also indirectly in its effect on saving behaviour. This is seen as sensible as the policy implementing authority sets the tax-schedule (or emissions cap) and then the economy adjusts to the constraint.

2.3.3 The welfare function

Based on a results from **?** and Fleurbaey (2010), I adopt a functional form for the objective of a nested aggregation which has the property that it reduces to the canonical form of the outer aggregation when the inner aggregation is trivially distributed. In our context this means that the welfare function collapses to the standard representative agent's objective for intertemporal aggregation whenever the regional variation is trivial. When the regional distribution is not trivial, the objective function has sufficient flexibility to treat the parameters determining intertemporal and interregional aggregations differently.

If $\vec{\mathbf{c}} = (c_{it})_{i \leq I, t \leq T}$ is the array of consumption streams of all regions as described in the preceding section, the social objective used for the remaining analysis takes the general form

$$W_{\phi,\psi}(\vec{\mathbf{c}}) = \sum_{t=1}^{T} \left(\frac{1}{1+\rho} \right)^{t} \phi \left[\psi^{-1} \left(\frac{\sum_{i=1}^{I} L_{it} \psi(c_{it})}{\sum_{i=1}^{I} L_{it}} \right) \right] \cdot \sum_{i=1}^{I} L_{it} \tag{2.22}$$

where $\phi : \mathbb{R}^{+} \to \mathbb{R}$ and $\psi : \mathbb{R}^{+} \to \mathbb{R}$ are the concave tranformations encapsulating inequality aversion across generations in the absence of inequality within generations and within generations in the absence of inequality across generations respectively.

It is quite standard to isolate the argument of ϕ, known as the *equally distributed equivalent consumption*[15] (EDEC) in the respective period:

$$\check{c}_{t,\psi} = \psi^{-1} \left(\frac{\sum_{i=1}^{I} L_{it} \psi(c_{it})}{\sum_{i=1}^{I} L_{it}} \right) \tag{2.23}$$

Notice that this is a generalisation of the simple population weighted average con-

[15] As introduced in Atkinson (1970)

sumption in period t, which is its special case when ψ is the identity:

$$\langle c_t \rangle := \frac{\sum_{i=1}^{I} L_{it} c_{it}}{\sum_{i=1}^{I} L_{it}}$$

With this definition, the social objective takes the more recognisable form

$$W_{\phi,\psi}(\vec{\mathbf{c}}) = \sum_{t=1}^{T} \left(\frac{1}{1+\rho} \right)^t \phi(\check{c}_{t,\psi}) \cdot \sum_{i=1}^{I} L_{it} \tag{2.24}$$

So when ψ is the identity, our functional form looks like a standard intertemporal choice function in which the welfare attained by a given generation is reflected in that generation's average income. If ψ is strictly concave, which implies strict aversion to inequality, it will be the case that $\check{c}_{t,\psi} < \langle c_t \rangle$ unless $(c_{1t}, c_{2t}, \ldots, c_{It})$ is equally distributed, i.e., $c_{it} = c_{jt}$, $\forall i, j$. Finally, if the social aversion to inequality between generations and within generations is identical, i.e. $\psi \equiv \phi$, then (2.22) yields

$$W_{\phi,\phi}(\vec{\mathbf{c}}) = \sum_{t=1}^{T} \sum_{i=1}^{I} \left(\frac{1}{1+\rho} \right)^t L_{it} \phi(c_{it}) \tag{2.25}$$

which, if $\rho = 1.5\%$ and $\phi(x) = u(x) = -\frac{x^{-\frac{1}{2}}}{2}$ is identical to (2.16), bar the Negishi weights, ν_{it}.

2.3.4 Numerical strategy

In order to solve the fixed-point problem (2.20) an iterative numerical strategy is employed. The strategy consists of decoupling the effects of the regions' choices on each other in each iteration by fixing the temperature path $(T_{at})_{t=1}^{60}$. We start with an initial guess for the temperature *path*. Given fixed tax and temperature paths in each iteration the decisions of the individual regions' representative agents simply become non-strategic Ramsey saving decisions with the net production functions:[16]

$$\mathcal{F}_{it} = \frac{1 - \Lambda_{it}}{1 + D_{it}} F_{it}(K_{it}, L_{it}) \tag{2.26}$$

Given these, Ramsey savings rate streams are determined *for every region independently* according to the positive utilities (2.7). These saving rate streams, along with the tax path, completely determine all the equations of motion, and in particular the temperature flow. This process yields a new guess at the path of atmospheric temperature $(T'_{at})_{t=1}^{60}$, for which the iteration can be re-run. This defines a fixed-point iteration, that is terminated when the difference between two consecutive temperature paths *and* saving rate streams are negligible. This results in savings paths that are optimal, given the temperature path that results from the savings paths. That is, savings rates from which no region would deviate given that all regions have chosen those savings rates. The solution to the fixed point.

The motivation for this particular procedure is the realisation[17] that the Ramsey saving rates are remarkably insensitive to the changes in total factor productivity allowed for by (2.26) and range of temperatures involved.

The numerical approach to solving the principal-agent problem (2.21) is also iterative and similar to the strategy outlined above. Rather than nesting the solution algorithm to (2.20) inside the optimisation (2.21), the two are performed sequentially. The problem (2.20) is solved for an initial guess for the tax path, $\vec{\tau}$. Taking the resulting savings rate streams, $\vec{\mathbf{s}}^*(\vec{\tau})$, as given, (2.21) is solved taking into account only the direct effect of the carbon tax on net output, via (2.26), through the mitigation channel and the damage channel.[18] This results in a new estimate for the optimal tax, $\vec{\tau}'$, which is used to repeat the iteration. The process converges numerically very quickly and is terminated at the desired accuracy.

[16]Recall that the μ_{it} are completely determined by the tax path $\vec{\tau}$.
[17]Implicit in approach taken by Nordhaus in the latest, Excel, version of RICE.
[18]But not the savings channel, as these are taken as exogenous in every iteration.

2.4 Optimal Taxes

2.4.1 Inequality aversion

In the context of social welfare functions the following are known as functions of *constant relative inequality aversion.*[19]

$$\phi(x) = \frac{x^{1-\eta}}{1-\eta}; \quad \psi(y) = \frac{y^{1-\xi}}{1-\xi} \tag{2.27}$$

The numerical optimisation is run with these functional forms, where the parameters η and ξ encapsulate the degree of inequality aversion of the intergenerational and intragenerational trade-off respectively.

The welfare function (2.16) used by Nordhaus also uses (2.27) as the functional form for the felicity function. However, the parameter values are chosen so that (2.7) accurately describes savings behaviour. As explained in Section 2.2, I retain (2.7) with its values for the behavioural pure rate of time preference, $\varrho = 1.5\%$, and elasticity of intertemporal substitution, $\gamma = 1.5$, but remain free to choose the parameters ρ, η and ξ in (2.22) independently. In the following the focus is specifically on the variation in η and ξ, while fixing the social behavioural rate of time preference, ρ. Increasing the social pure rate of the time preferences has the well documented effect of reducing optimal taxes and this is borne out in additional model runs with this functional specification as well. However, I take the view that not much distinguishes the present generations from the unborn ones whose welfare would be discounted by a higher value for ρ, and set $\rho = 0.1\%$ according the the rationale of the Stern review: a small but constant probability of the extinction of humanity.

To understand the aversion to inequality inherent in the parameter value range, I borrow a hypothetical trade-off from Dasgupta (2008). Let ζ refer to either η or ξ.

[19] See Atkinson (1970)

Consider two individuals, A and B, with consumption flows of \$360 and \$36,000 per annum, respectively. Restricting ourselves to the functional form (2.27), we have the following trade-offs at different parameter values:

I At $\zeta = 0$, the social objective would consider a 100% reduction in individual A's consumption normatively equivalent to a 1% reduction in B's consumption.

II At $\zeta = 1$, a 1% reduction to A is equivalent to a 1% reduction to B.

III At $\zeta = 1.5$, a 1% reduction to A is equivalent to a 10% reduction to B.

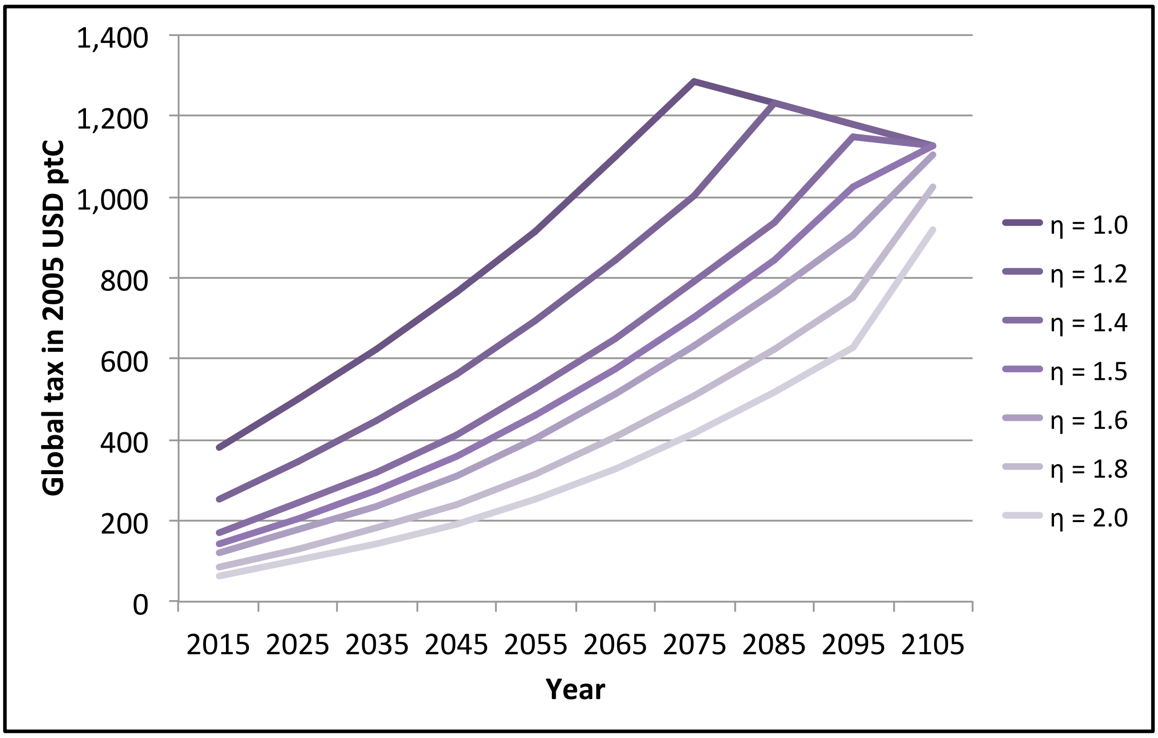
IV At $\zeta = 2$, a 1% reduction to A is equivalent to a 50% reduction to B.

Case I amounts to no aversion to inequality. Case II is probably at the lower end of reasonable inequality aversion and Case IV is probably at the upper end. I find a value of 1.5 most defensible, but acknowledge that reasonable people may disagree. However, my reading of the example above makes a value outside of the range $[1, 2]$ is plainly unreasonable.

2.4.2 Intergenerational inequality

Setting $\xi = 0$ amounts to treating the world as an aggregate consumer in every generation, ignoring contemporaneous inequality while accounting for inequality between generations at the *distinct* value of η. This is essentially what globally aggregated models do. Figure 2.1 plots the optimal taxes of such a case for different values of η. All tax paths are increasing. At a flatter path delaying mitigation would be beneficial and at a steeper path there would be a gain from bringing mitigation forward. The decreasing cost of mitigation, relative affluence of the future over the present and the rate at which carbon is removed from the atmosphere by the carbon cycle all combine to yield the slope. The descending trend, starting in 2075, which all tax paths eventually join, is the exogenously assumed decreasing price path of a fully

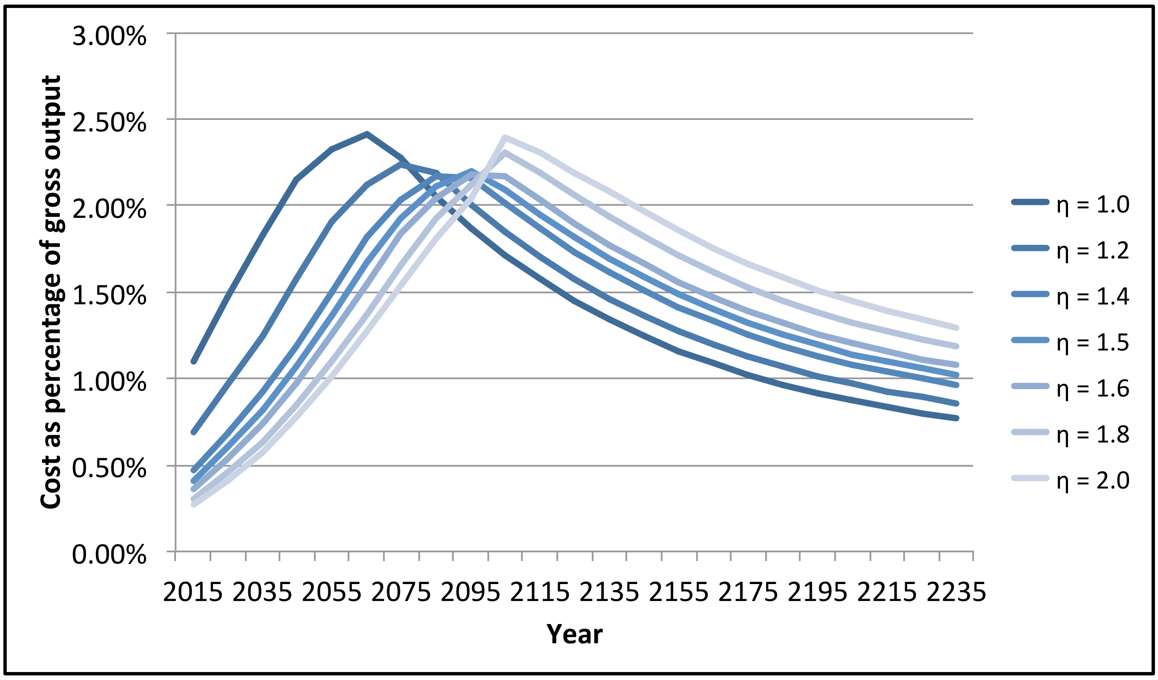
Figure 2.1: Optimal tax paths when $\xi = 0$ (in 2005 USD ptC)



green backstop, i.e. the implicit tax of full mitigation. The kinks in the paths before reaching the backstop are boundary effects of the numerical method.

The tax paths rise at similar slopes, and the 2015 tax ranges from \$63 to \$380 per tC. This is equivalent to global mitigation rates ranging from 20% to 55% *in 2015*. It is clear that the degree of intergenerational inequality aversion has a huge impact on the prescribed policy. Over the range [1,2] there is an unambiguous effect whereby lower intergenerational inequality aversion requires a stronger policy response in form of uniformly higher carbon taxes. Qualitatively, this effect is documented (See Dietz, Hope, Stern, and Zenghelis (2007) for similar results in PAGE) and well understood. Figure 2.2 plots the time paths of the net deflating effect on *global* output – the output-weighted average of the fraction in equation (2.26) representing both mitigation costs and climate damages – for the policies plotted in Figure 2.1. The more conservationist mitigation policies associated with lower degrees of inequality

Figure 2.2: Net Costs on Gross Output



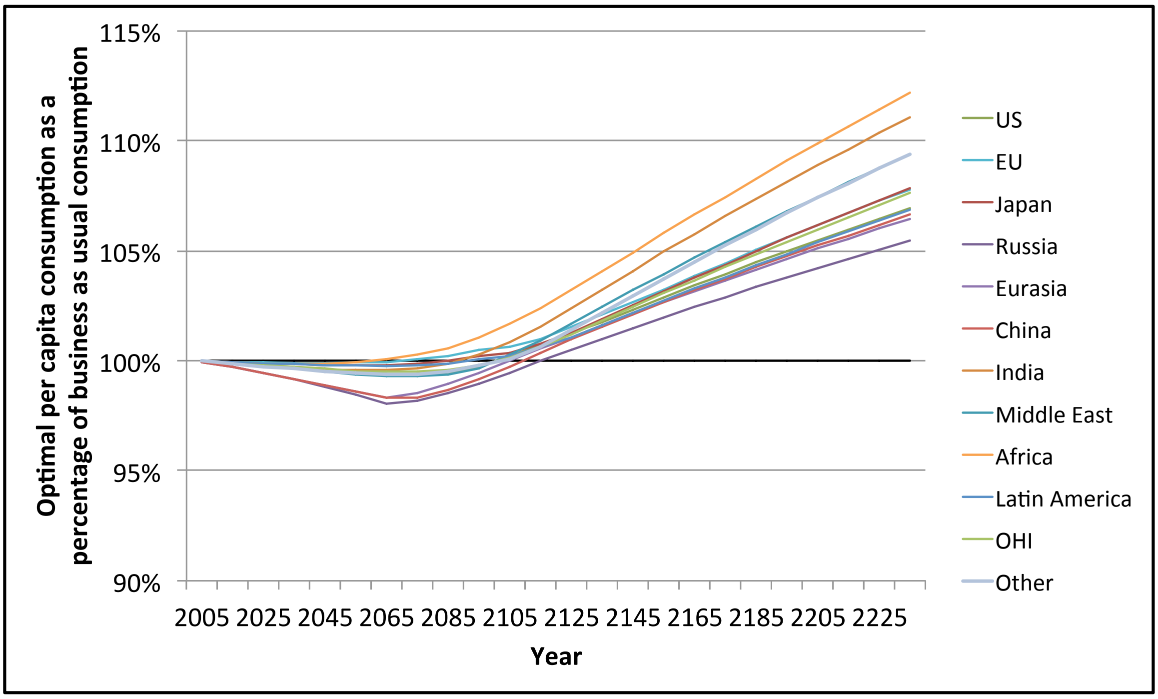
aversion incur higher (mitigation) costs early on for the benefit of lower (climate) damages to later generations. Since the assumed growth in global consumption due to exogenous productivity growth is qualitatively unaffected[20] future generations are inexorably wealthier. Stronger mitigation policy favours them at the expense of the poorer mitigating generations, so the more one cares about inequality along this dimension, the less stringent is the prescribed policy.

Figure 2.3 shows how the optimal consumption[21] compares to the business-as-usual consumption paths.

[20]See Dietz and Asheim (2012) for a stochastic extension of DICE in which the possibility of negative consumption growth due to climate effects is acknowledged.

[21]The policy chosen is the optimum to the welfare function with parameters $\rho = 0.1\%$, $\eta = 1.5$ and $\xi = 1.5$.

Figure 2.3: Optimal capita consumption as a proportion of BAU consumption

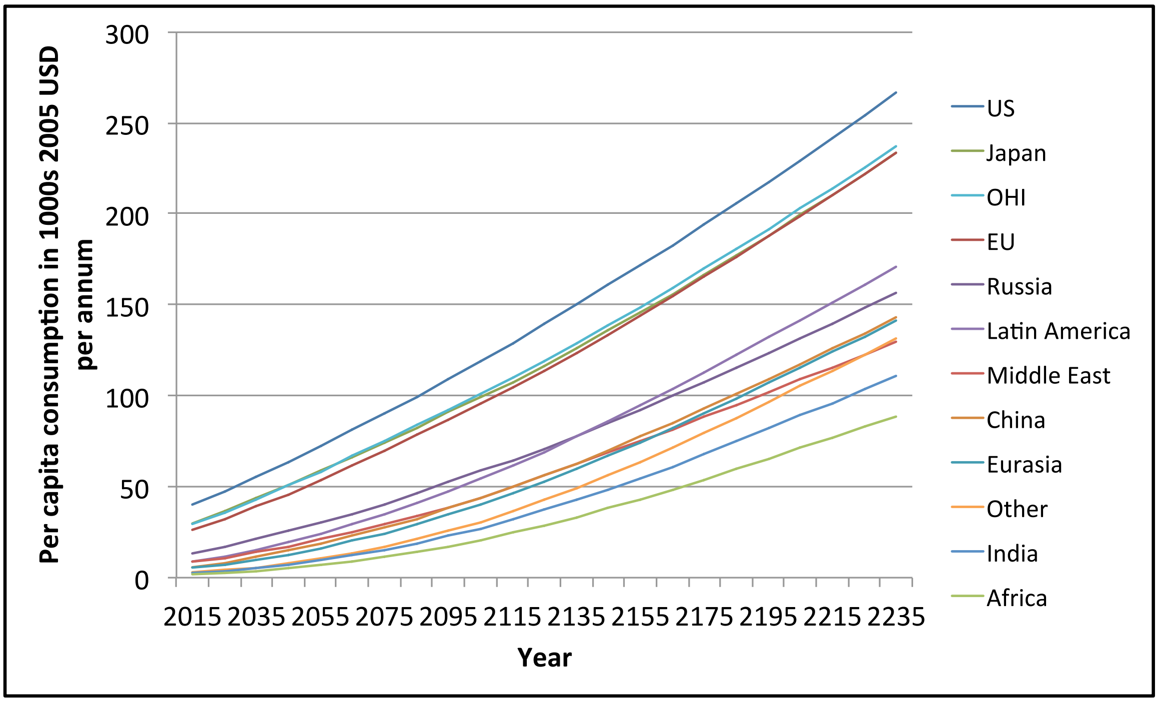


2.4.3 Intragenerational inequality

At zero intragenerational inequality aversion regional distribution is irrelevant to the social planner and the intergenerational trade-off in global consumption via the effect on global output seen in Figure 2.2 is essentially the only determinant of optimal climate policy. When a regionally disaggregated model is not available, the $\xi = 0$ approximation is often deemed sufficient for policy appraisal. In their seminal *Project Appraisal and Planning*, Little and Mirlees justify such an approach of separating the intertemporal and contemporaneous decisions and appealing to further redistributive measures for the purpose of contemporaneous redistribution:

"It (spatial disaggregation) would be unnecessary if it were true that other instruments of policy [...] could achieve as much equality as was desirable [...] more efficiently"

Figure 2.4: Per capita consumption (in 1000s 2005 USD per annum)



Such other instruments of policy are unlikely to be available at a global level. The authors of the Stern review concede that much, but were not able to model the regional distribution due to time constraints. Relying on an analysis from Nordhaus and Boyer (2000) they provide a back-of-the-envelope calculation which estimates that the global analysis without inequality aversion across regions underestimates the cost of climate change by 20% (Dietz (2008)). The rationale underlying the analysis in Nordhaus and Boyer (2000) is that climate damages will be more severe in poorer regions and that this unequally distributed cost leads to the global model underestimating actual welfare costs. Azar and Sterner (1996), Fankhauser, Tol, and Pearce (1997), Hope (2008) and Anthoff, Hepburn, and Tol (2009) arrive at qualitatively similar conclusions in their calculations of the social cost of carbon. Figure 2.4 shows the paths of per capita consumption of for the twelve different regions for a particular mitigation policy.[22] Figure 2.5 plots the estimated damages of

[22]The policy chosen is the optimum to the welfare function with parameters $\rho = 0.1\%$, $\eta = 1.5$

a 3-degree temperature increase. Together, the two graphs provide a similar intuition. In general, one can say that poorer regions will suffer more severely from the effects of climate change.[23]

However, the distribution of climate costs is not the full story, since the mitigation costs are also expected to be distributed unequally, which is not taken into account in a calculation of the social cost of carbon. As explained in Section 2.2.2 a globally uniform tax is used rather than differentiated marginal mitigation costs as suggested by Anthoff (2011). As explained above, this second best approach is taken here because differentiated carbon prices would presumably lead to inter-regional transfers that seem politically infeasible.

Substituting (2.17) into (2.12) yields the following distribution of mitigation costs for a given carbon tax τ_t:

$$\Lambda^*_{it} = \theta^1_{it} \left(\frac{\tau_t \sigma_{it}}{\theta^1_{it} \theta_2} \right)^{\frac{\theta_2}{\theta_2 - 1}} \tag{2.28}$$

Figure 2.6 plots the distribution of costs for a 100 USD carbon tax in 2015.[24] Despite the fact that the two poorest regions, Africa and India are spared particularly large costs, the distribution can hardly be seen to favour the poorer regions in general. It is shown below that in a well-defined sense, globally uniform taxes have the effect of increasing inequality *in the period of taxation*.

Given the way in which the equally distributed equivalent enters the welfare function, there is a canonical way to measure period by period spacial inequality in this context. Underlying every inequality measure is a welfare function (Dalton (1920)) and Atkinson proposed a particularly intuitive one based on the equally distributed equivalent consumption (Atkinson (1970)). Recall that $\check{c}_{t,\psi}$ denotes the equally dis-

and $\xi = 1.5$. However, the effect of climate policy on this scale is so small that the qualitative features of the consumption paths don't change.

[23]The specification of the damage function is quadratic, so the relative magnitudes do not remain the same for *all* temperatures. However, figure 2.5 is representative for the range of temperature outcomes considered.

[24]For different periods the absolute magnitudes decrease, but the relative magnitudes stay the same.

Figure 2.5: Damage Cost of a 3 degree temperature increase

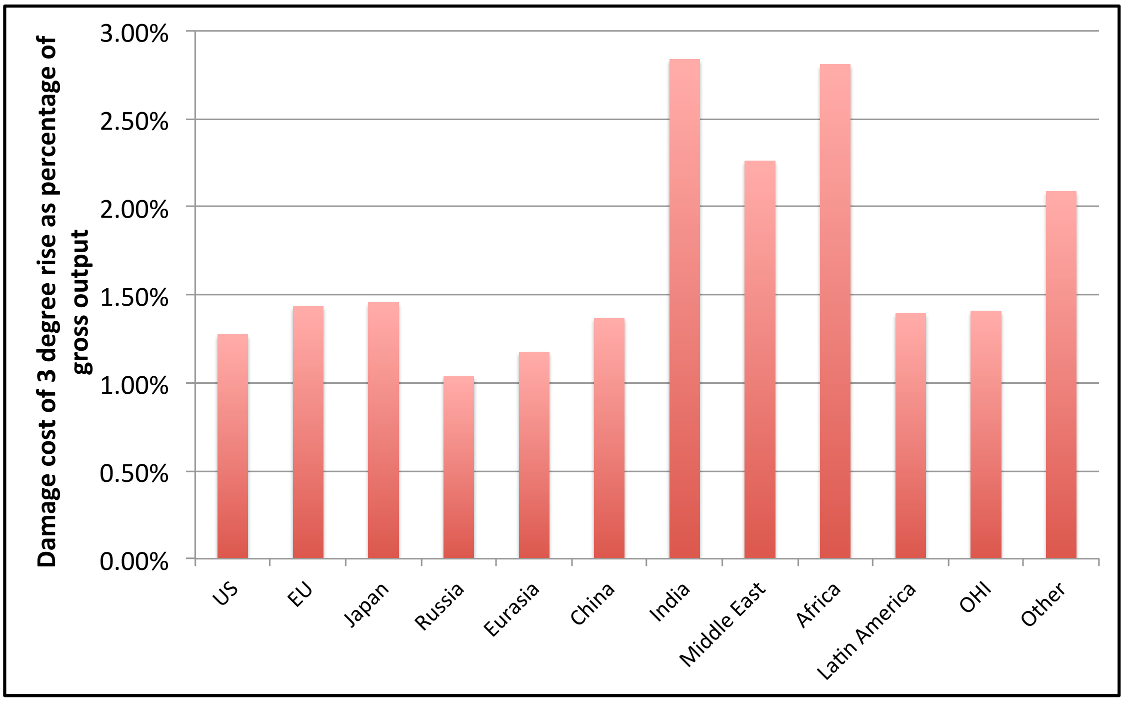
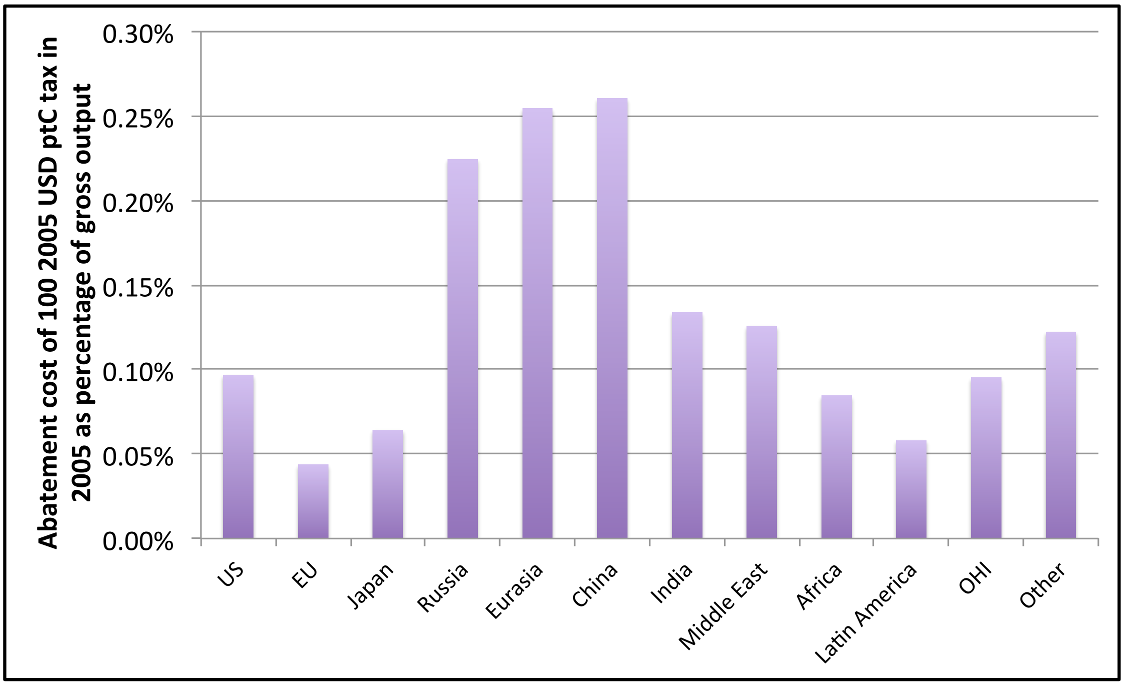


Figure 2.6: Abatement Cost of a 100 USD carbon tax



tributed equivalent consumption and $\langle \vec{c}_t \rangle$ the average consumption of a generation t. Following Atkinson, define the index measuring inequality in generation t by

$$I_{t,\psi} = 1 - \frac{\check{c}_{t,\psi}}{\langle c_t \rangle}. \tag{2.29}$$

Consider the optimal outcome at $\eta = 1.5$ and $\xi = 0$. If we want to know whether inequality increases or decreases with marginal changes in the tax rate, we need to choose the coefficient of inequality aversion with respect to which inequality is to be measured. There is no good reason not to choose $\xi = 1.5$. At this coefficient of inequality aversion for ψ, the inequality measure (2.29) behaves as follows at the optimum to $\eta = 1.5, \xi = 0$; a marginal increase in any component of the tax vector[25] increases the inequality measure significantly in the period of the tax increase, moderately in the subsequent period, and *reduces* inequality in all subsequent periods.[26] I repeated this exercise for all optima plotted in Figure 2.1 and the same pattern holds.[27] Marginal changes in carbon taxes invariably increase inequality for two periods and reduce inequality during the subsequent periods in which there is a reduction in climate damages.

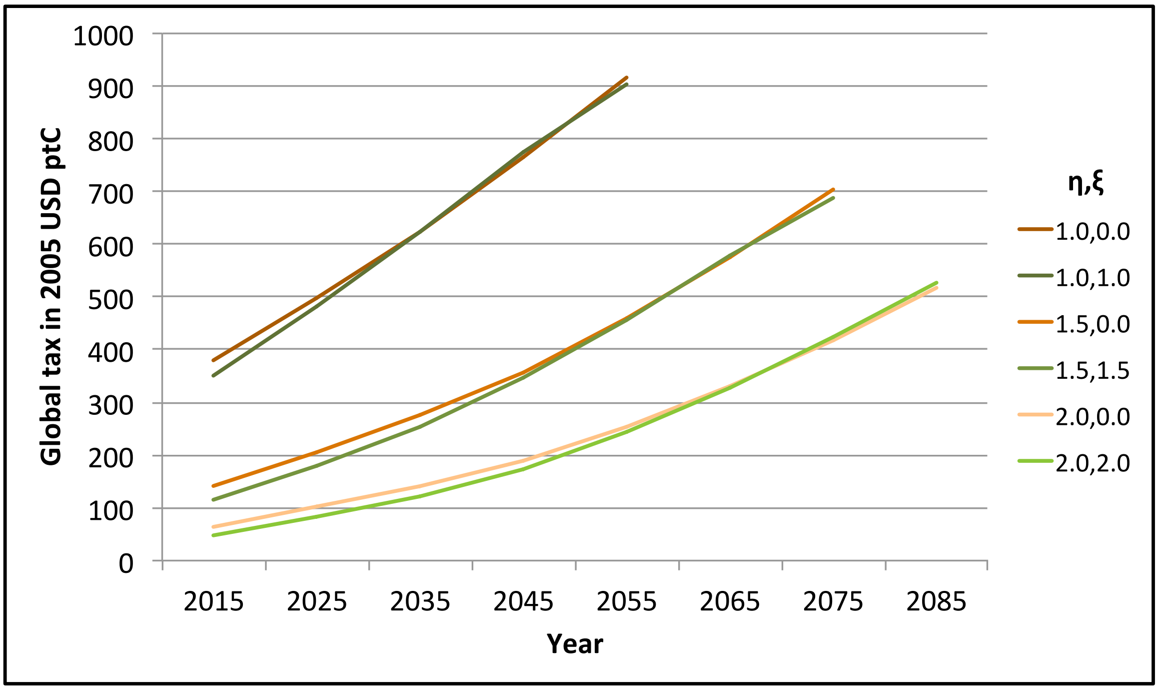
Taking into account only the inequality increasing effect of climate damages on future periods, studies calculating the social cost of carbon conclude that it increases as the equity weights associated with ξ are introduced. Nordhaus and Boyer (2000) similarly estimate that the optimal tax ought to increase as ξ is increased. However, if the inequality increasing effect of mitigation through *globally uniform carbon taxes* is also taken into account, the net effect of higher intratemporal inequality aversion on optimal taxes is different. Figure 2.7 plots the two optimal taxes, for $\xi = 0$ and $\xi = \eta$ for representative values of η in the selected range, [1,2]. The graphs show that

[25] That is, any component before full mitigation.

[26] Into the very distant future from the period of the marginal change, the effect becomes quite erratic, but not particularly large.

[27] The inequality measure used is varied to correspond to the value of η of the respective optimum.

Figure 2.7: Optimal tax paths (in 2005 USD ptC)



for all parameters in the range, the inclusion of regional inequality aversion *reduces the initial* optimal taxes, and therefore the mitigation levels, relative to the globally aggregated analysis. These results stand in stark contrast to previous studies, and are driven by the fact that uniform carbon taxes increase inequality in the mitigating periods: an effect that was not previously considered.

As the tax policies reach full mitigation, the optima taking into account regional inequality catch up or even overtake the corresponding optima treating the globe as an aggregate. The intuition behind this is that the effect of climate damages on inequality starts dominating in these periods. However, in the periods immediately preceding full mitigation, not plotted in Figure 2.7, the policies which account for regional inequality have a kink. This is presumably due to boundary effects, making it difficult to conclude much about the relative magnitudes in those periods.

2.4.4 Optimal policy

As stated earlier, the set of welfare parameter values proposed here are $\rho = 0.1\%$, $\eta = \xi = 1.5$. At these values, the optimal tax path, in 2005 USD per tC is

$$\vec{\tau}^*_{1.5,1.5} = (113, 180, 254, 343, 458, 576, 688, 791, 841\ldots)$$

and reaches full mitigation at the beginning of the 22nd century. If regional inequality were not taken into account, by setting $\xi = 0$, the optimal tax would be greater for most of the time horizon, as can be seen in Figure 2.7, with a first period tax of \$141 per tC.

Using a globally aggregated version of the PAGE model, the Stern review calculates the social cost of carbon of the business as usual SRES A2 path at \$350 per tC. The models' outputs are not fully comparable, but the ethical parameters used by Stern can be interpreted as equivalent to $\rho = 0.1\%$, $\eta = 1.0$ and $\xi = 0$ in this context. They estimate that taking into account regional inequality would result in climate change costing 20% more.[28] The set of parameters that describe such an evaluation in this model would be $\rho = 0.1\%$, $\eta = 1.0$ and $\xi = 1.0$. The optimal tax paths to both sets of parameters are plotted in Figure 2.7. The results are comparable in orders of magnitude to Stern's, however, with the qualitative difference that, the inclusion of regional inequality aversion *reduces* the tax rate rather than increasing it.

Finally, from a normative standpoint, the difference to Nordhaus' original specification is twofold. The intergenerational trade-off is modified only by a different choice of discount rate - 0.1% rather than 1.5% - while the coefficient of intergenerational inequality aversion remains the same at 1.5. The intragenerational trade-off retains the same coefficient of inequality aversion, but ignores the Negishi weights.

[28]This does *not* simply equate to a 20% increase in the social cost of carbon, since there are non-linearities involved.

The result is a uniformly greater carbon tax than Nordhaus', which starts at $37 per tC as opposed to $113 per tC.

The main driver for this is the different choice in discount rate, while the removal of the Negishi weights has only a very small effect on the optimal tax. This can be seen by optimising with respect to our specification but with Nordhaus' discount rate. The result is an optimal tax that also starts at $37 per tC, and stays very close to Nordhaus' optimum throughout.

2.5 Conclusion

This paper introduces a regionally disaggregated social objective to optimal mitigation analysis within the RICE-2010 model. This modification allows for two things. The analysis of the effects of spatial as well as temporal inequality aversion on optimal policy, as well as an independent, *normative* choice of the welfare determining parameter values that does not affect the description of the endogenous savings decision. That higher levels or intergenerational inequality aversion reduce the optimal mitigation effort is well known and confirmed in this framework.

Given additional policy tools at a global level such as lump-sum transfers, it would be unnecessary to assess mitigation policy with a spatially inequality averse welfare function, since distributional objectives could be achieved by other means. However, such transfers are highly unlikely for several reasons, not least the political economy of climate policy. It is also assumed that any distribution of abatement costs that deviates from that which would result from a uniform carbon tax will not be politically sustainable. This is certainly an extreme assumption, as some transfers may be expected to take place to assist poorer regions both in their abatement efforts as well as their adaptation efforts and the distortionary effects of slightly differing carbon taxes are likely to be tolerated. However, these are unlikely to be in the same orders of magnitude – percentages of GDP – as the full abatement costs and climate damages, making the assumption a reasonable benchmark.

Under this assumption, climate policy affects inequality via two channels, in opposing directions. Mitigation increases inequality through the unequal distribution of abatement costs, and decreases inequality as it reduces climate damages, which are unequally distributed as well. The former dominates at the chosen parameter values but the net effect is small, mostly due to the fact that two effects push in opposite directions. The analysis in Anthoff (2011) allows for regionally different taxes, and also concludes that the disaggregated optimum results in *lower* optimal mitigation levels.

An important issue relating to disaggregated models raised by other studies,[29] but not touched upon here, is the level of spatial resolution. The resolution determines which individuals get lumped together and treated as a homogenous group, and if these are not expected to be homogenous, too coarse a resolution will introduce additional inaccuracy in the outcome. The functional form of the social objective used above can be used to increase the accuracy without raising the the level of resolution of the model. Instead of estimating the effects of climate and policy on the consumption of the regions in the model, it could be possible to estimate the effects on equally distributed equivalent consumption, thus taking into account *intraregional* distribution without having to explicitly model at a finer resolution. The equally distributed equivalent consumption can be passed to the welfare function as an argument resulting in an account of social welfare that is consistent in its approach to spatial and temporal disaggregation. Replace the consumption flows c_{it} with equally distributed equivalents in the argument to in (2.25) to see this.

The social objective also lends itself to a natural extension along the risk dimension in which the parameters determining social risk preference are *a priori* different

[29] See Tol (2001)

from those determining inequality preferences. Seeing as uncertainty is a prevalent feature in climate change, it would be a natural extension to look at the effect of risk preference parameters on optimal policy and contrast it to the effect of inequality preference parameters studied above.

2.6 Appendix: RICE

The equations governing the RICE model are:

$$\mathcal{L}_i = \sum_{t=1}^{60} \left(\frac{1}{1+\varrho}\right)^{10t} L_{it} u(c_{it}) \tag{2.30}$$

$$F_{it}(K,L) = A_{it} K^{\gamma} L^{1-\gamma} \tag{2.31}$$

$$Y_{it} = F_{it}(K_{it}, L_{it}) \tag{2.32}$$

$$E_{it} = \sigma_{it}(1-\mu_{it})Y_{it} \tag{2.33}$$

$$E_t = \sum_{i=1}^{12} E_{it} + EL_t \tag{2.34}$$

$$\sum_{t=1}^{60}\sum_{i=1}^{12} E_{it} \leq R \tag{2.35}$$

$$Q_{it} = \frac{1-\Lambda_{it}(\mu_{it})}{1+D_{it}(T_{at})} Y_{it} \tag{2.36}$$

$$\Lambda_{it}(\mu) = \theta^1_{it} \mu_{it}^{\theta_2} \tag{2.37}$$

$$D_i(T_{at}) = \alpha^1_i T_{at} + \alpha^2_i T^2_{at} \tag{2.38}$$

$$K_{it+1} = 10 s_{it} Q_{it} + (1-\delta)^{10} K_{it} \tag{2.39}$$

$$\begin{pmatrix} M_{at} \\ M_{bt} \\ M_{ct} \end{pmatrix} = \begin{pmatrix} m_{aa} & m_{ba} & 0 \\ m_{ab} & m_{bb} & m_{cb} \\ 0 & m_{bc} & m_{cc} \end{pmatrix} \cdot \begin{pmatrix} M_{at-1} \\ M_{bt-1} \\ M_{ct-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \cdot E_t \tag{2.40}$$

$$\begin{pmatrix} T_{at} \\ T_{bt} \end{pmatrix} = \begin{pmatrix} t_{aa} & t_{ba} \\ t_{ab} & t_{bb} \end{pmatrix} \cdot \begin{pmatrix} T_{at-1} \\ T_{bt-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \cdot F_t\left(\frac{M_{at}+M_{at+1}}{2}\right) \tag{2.41}$$

$$F_t(x) = \xi_2 \left(\xi_1 \log_2\left(\frac{x}{M_{aPI}}\right) + F\mathrm{exo}_t\right) \tag{2.42}$$

- The savings rates are determined endogenously by a *finitely* lived representative agent with decision utility (2.30).
- The production function is Cobb-Douglas in capital and labour, where labour

is given by the exogenous population and capital is determined by the savings rate.

- The gross total factor productivities, A_{it} are calibrated to the 2005 economy for A_{i1} and pushed forward under the assumption of unrelenting, albeit declining TFP growth and regional convergence.

  TFP of the USA region is given by the growth process

$$A_{1t} = A_{11} \exp\{10 \cdot G^A_{1t}\} \tag{2.43}$$

$$G^A_{1t} = \sum_{s=1}^{t} g^A_{1t} \tag{2.44}$$

$$g^A_{1t} = (1-\gamma)\left[g^A_l + \exp(-\delta^A)^{t-1}(g^y_{11} - g^A_l)\right] \tag{2.45}$$

  TFP of all other regions is given by the convergence process to the USA TFP

$$A_{it} = A_{i,t-1} \exp\{10 \cdot g^A_{it}\} \tag{2.46}$$

$$g^A_{it} = g^A_{1t} + (1-\lambda)^{t-2}\lambda \cdot x_i \tag{2.47}$$

$$x_i = \log\left\{\frac{y_{11}}{y_{i1}}\right\} + \log\{\kappa^A_i\} + 10 \cdot (g^y_{11} - g^y_{i1}) \tag{2.48}$$

- Equation (2.32) defines the gross output - the projected output *potential* in absence of climate change costs or mitigation.

- In absence of mitigation, emissions are an inexorable by product of production, at an exogenous emissions to output ratio: (2.33)

- The emissions to output ratio, σ_{i1}, is calculated as the realised ratio between emissions and output in 2005, and projected forward at an exogenously declining rate. For future time periods, the emissions to output ratio is projected to decline exogenously as[30]

$$\sigma_{it} = \sigma_{i1} \cdot \exp\{10 \cdot G^{\sigma}_{it}\} \tag{2.49}$$

$$G^{\sigma}_{it} = (t-1)g^{\sigma}_{l} + (g^{\sigma}_{iH} - g^{\sigma}_{l}) \sum_{j=1}^{t-1} (1-\delta^{\sigma})^{j} \tag{2.50}$$

- The total change in atmospheric carbon consists of the sum of industrial emissions plus exogenously assumed emissions due to land use change, which are assume positive only in Africa, Latin America and Asia, and declining at an exogenous rate. (2.34)

- The entirety of carbon resources is assumed to be finite at quantity $R = 6000$ Gtons. The supply side of carbon resources is not explicitly modeled. Instead, as the resource approaches exhaustion (in the baseline run) a high-order polynomial mimics the exponential rise in relative price predicted by the Hotelling rule. More precisely, define

$$H_t = 100 \cdot \left(\frac{CumE_t}{R}\right)^{12} \tag{2.51}$$

as the *pseudo* Hotelling rent in period t, where $CumE_t$ are the cumulative emissions at date t and R is the total amount of the resource available. The the amount of the resource used at date t is determined by (2.33) with the mitigation rate μ_{it} given by equation (2.17) by replacing the tax τ_t with the sum of the tax plus the pseudo Hotelling rent $\tau_t + H_t$. That is

$$\mu_{it} = \left(\frac{(\tau_t + H_t)\sigma_{it}}{\theta^{1}_{it}\theta_2}\right)^{\frac{1}{\theta_2 - 1}} \tag{2.52}$$

The renewable backstop comes in when $\mu_{it} = 1$, or equivalently when the 'cost'

[30]The 2015 emissions to output ratios are adjusted to from (2.50) to reflect more accurate predictions at the time RICE 2010 was designed.

$\tau_t + H_t$ is equal to the price of the backstop. (See equation (2.53) below)

- Net output consists of the potential gross output deflated by the dual costs of abatement and climate damages: (2.36)

- The coefficients on the abatement cost function are given exogenously by

$$\theta_{it} = \frac{pb_{it}\sigma_{it}}{\theta_2} \tag{2.53}$$

  where pb_{it} is the price of the renewable backstop. With this parameterisation is is clear (from (2.52) and (2.53)) that when $\tau_t + H_t = pb_{it}$ region i mitigates fully, i.e. reverts completely to the renewable backstop technology.

- The price of the renewable backstop is given exogenously by

$$pb_{it} = \begin{cases} \Theta pb_{i1} + (1 - \bar{\delta}^p)(pb_{i,t-1} - \Theta pb_{i0}) & \text{if } t < 2250 \\ \underline{\delta}^p pb_{i,t-1} & \text{if } t \geq 2250 \end{cases} \tag{2.54}$$

- Climate damages depend exclusively on the global mean temperature rise over pre-industrial levels: (2.38)

- Capital accumulation has the slightly odd feature that annual investment cumulates arithmetically, but annual capital depreciation cumulates geometrically. The result is that this process overstates capital accumulation, and therefore growth, relative to both the purely arithmetic and purely geometric processes: (2.39)

- Carbon flows at exogenously determined rates between the three 'reservoirs' of atmosphere, upper oceans and lower oceans in a process that will move towards equilibrium once no additional carbon is pumped into the atmosphere: (2.40)

- Similarly, the oceans provide a cooler reservoir for temperature, absorbing some of the additional heat from the radiative forcing of the greenhouse effect: (2.41)

- Equation (2.42) contains the greenhouse effect. An increase in the atmospheric concentration of carbon increases the radiative forcing of the atmosphere, leading to an increase in its temperature. Part of the forcing is due to the emissions from industrial use and land use change, and part of it is additional and exogenously assumed to be

$$Fexo_t = \begin{cases} Fexo2000 + 0.1 \cdot (Fexo2100 - Fexo2000) \cdot (t-1) & \text{if } t \leq 2100 \\ Fexo2100 & \text{if } t > 2100 \end{cases} \tag{2.55}$$

- It is important to reemphasize that the 12 regions are considered autarkic from an economic perspective. That is to say they have access to the production technology given by (2.31) and the regional parameters in Table 2.2 and optimise over their consumption savings decision without the possibility to trade.

- The only link across the regions is through the climate externality, whereby the emission of each region affects global radiative forcing, which increases global mean temperatures and affects the regions through the damage functions given by (2.38) and the parameters in Table 2.2.

Table 2.2: Parameters of RICE 2010

Parameter	Symbol	Value	Unit
Behavioural discount rate	ρ	1.5%	rate per year
Felicity function	$u(c)$	$-2c^{-\frac{1}{2}}$	utils per year
Population	L_{it}	UN estimates	millions
Capital Share	γ	0.3	—

Parameter	Symbol	Value	Unit
2005 Capital	$K_{1,1}$	22.8	Trillion 2005 USD
	$K_{2,1}$	23.3	Trillion 2005 USD
	$K_{3,1}$	7.1	Trillion 2005 USD
	$K_{4,1}$	2.8	Trillion 2005 USD
	$K_{5,1}$	1.4	Trillion 2005 USD
	$K_{6,1}$	9.3	Trillion 2005 USD
	$K_{7,1}$	4.1	Trillion 2005 USD
	$K_{8,1}$	5.4	Trillion 2005 USD
	$K_{9,1}$	2.1	Trillion 2005 USD
	$K_{10,1}$	7.7	Trillion 2005 USD
	$K_{11,1}$	6.7	Trillion 2005 USD
	$K_{12,1}$	4.4	Trillion 2005 USD
2005 TFP	$A_{1,1}$	11.4	residual
	$A_{2,1}$	8.4	residual
	$A_{3,1}$	9.1	residual
	$A_{4,1}$	4.9	residual
	$A_{5,1}$	2.7	residual
	$A_{6,1}$	2.3	residual
	$A_{7,1}$	1.5	residual
	$A_{8,1}$	3.9	residual
	$A_{9,1}$	1.3	residual
	$A_{10,1}$	3.7	residual
	$A_{11,1}$	9.0	residual
	$A_{12,1}$	1.8	residual
2005 GDP/c	$y_{1,1}$	41.8	1000s 2005 USD
	$y_{2,1}$	26.6	1000s 2005 USD

Parameter	Symbol	Value	Unit
	$y_{3,1}$	30.3	1000s 2005 USD
	$y_{4,1}$	11.9	1000s 2005 USD
	$y_{5,1}$	5.2	1000s 2005 USD
	$y_{6,1}$	4.1	1000s 2005 USD
	$y_{7,1}$	2.2	1000s 2005 USD
	$y_{8,1}$	8.4	1000s 2005 USD
	$y_{9,1}$	1.7	1000s 2005 USD
	$y_{10,1}$	8.2	1000s 2005 USD
	$y_{11,1}$	29.7	1000s 2005 USD
	$y_{12,1}$	2.8	1000s 2005 USD
2005 GDP/c growth	$g^y_{1,1}$	1.5%	rate per year
	$g^y_{2,1}$	1.6%	rate per year
	$g^y_{3,1}$	1.4%	rate per year
	$g^y_{4,1}$	2.6%	rate per year
	$g^y_{5,1}$	2.5%	rate per year
	$g^y_{6,1}$	7.1%	rate per year
	$g^y_{7,1}$	4.6%	rate per year
	$g^y_{8,1}$	2.4%	rate per year
	$g^y_{9,1}$	3.7%	rate per year
	$g^y_{10,1}$	2.9%	rate per year
	$g^y_{11,1}$	1.9%	rate per year
	$g^y_{12,1}$	2.8%	rate per year
TFP growth decline rate	δ^A	0.1	rate per decade
Long term TFP growth rate	g^A_l	0.33%	rate per year
TFP convergence rate	λ	0.1	rate per decade
TFP convergence ratio	κ^A_2	0.9	—

Parameter	Symbol	Value	Unit
	κ_3^A	0.9	—
	κ_4^A	0.6	—
	κ_5^A	0.6	—
	κ_6^A	0.6	—
	κ_7^A	0.5	—
	κ_8^A	0.5	—
	κ_9^A	0.4	—
	κ_{10}^A	0.7	—
	κ_{11}^A	0.9	—
	κ_{12}^A	0.6	—
2005 carbon intensity	$\sigma_{1,1}$	0.13	tC/1000 2005 USD
	$\sigma_{1,2}$	0.09	tC/1000 2005 USD
	$\sigma_{1,3}$	0.10	tC/1000 2005 USD
	$\sigma_{1,4}$	0.25	tC/1000 2005 USD
	$\sigma_{1,5}$	0.32	tC/1000 2005 USD
	$\sigma_{1,6}$	0.30	tC/1000 2005 USD
	$\sigma_{1,7}$	0.17	tC/1000 2005 USD
	$\sigma_{1,8}$	0.17	tC/1000 2005 USD
	$\sigma_{1,9}$	0.15	tC/1000 2005 USD
	$\sigma_{10,1}$	0.09	tC/1000 2005 USD
	$\sigma_{11,1}$	0.14	tC/1000 2005 USD
	$\sigma_{12,1}$	0.14	tC/1000 2005 USD
Long term σ growth rate	g_l^σ	-0.25%	rate per year
Historic σ growth rate	g_{1H}^σ	-1.8%	rate per year
	g_{2H}^σ	-1.5%	rate per year
	g_{3H}^σ	-1.8%	rate per year

Parameter	Symbol	Value	Unit
	g^{σ}_{4H}	-1.7%	rate per year
	g^{σ}_{5H}	-2.4%	rate per year
	g^{σ}_{6H}	-2.5%	rate per year
	g^{σ}_{7H}	-2.2%	rate per year
	g^{σ}_{8H}	-1.8%	rate per year
	g^{σ}_{9H}	-2.2%	rate per year
	g^{σ}_{10H}	-1.5%	rate per year
	g^{σ}_{11H}	-1.9%	rate per year
	g^{σ}_{12H}	-1.9%	rate per year
σ growth decline rate	δ^{σ}	0.1	rate per decade
Total carbon resources	R	6000	Giga Tons C
Exponent on abatement cost	θ_2	2.8	—
2005 price of backstop	$pb_{1,1}$	1.1	1000 2005 USD/tC
	$pb_{2,1}$	1.8	1000 2005 USD/tC
	$pb_{3,1}$	1.8	1000 2005 USD/tC
	$pb_{4,1}$	0.8	1000 2005 USD/tC
	$pb_{5,1}$	0.8	1000 2005 USD/tC
	$pb_{6,1}$	0.9	1000 2005 USD/tC
	$pb_{7,1}$	1.4	1000 2005 USD/tC
	$pb_{8,1}$	1.3	1000 2005 USD/tC
	$pb_{9,1}$	1.4	1000 2005 USD/tC
	$pb_{10,1}$	1.6	1000 2005 USD/tC
	$pb_{11,1}$	1.4	1000 2005 USD/tC
	$pb_{12,1}$	1.5	1000 2005 USD/tC
Initial to final pb ratio	Θ	0.1	—
Initial rate of pb decline	$\bar{\delta}^{p}$	0.05	rate per decade

Parameter	Symbol	Value	Unit
Final rate of pb decline	$\underline{\delta}^p$	0.5	rate per decade
Linear coefficient on damage	α_1^1	0	—
	α_2^1	0	—
	α_3^1	0	—
	α_4^1	0	—
	α_5^1	0	—
	α_6^1	0.0785	—
	α_7^1	0.4386	—
	α_8^1	0.2780	—
	α_9^1	0.3410	—
	α_{10}^1	0.0609	—
	α_{11}^1	0	—
	α_{12}^1	0.1755	—
Quadratic coefficient on damage	α_1^2	0.1414	—
	α_2^2	0.1591	—
	α_3^2	0.1617	—
	α_4^2	0.1151	—
	α_5^2	0.1305	—
	α_6^2	0.1259	—
	α_7^2	0.1689	—
	α_8^2	0.1586	—
	α_9^2	0.1983	—
	α_{10}^2	0.1345	—
	α_{11}^2	0.1564	—
	α_{12}^2	0.1734	—
Depreciation rate	δ	0.1	rate per year

Parameter	Symbol	Value	Unit
Carbon cycle coefficients	m_{aa}	88	percent per decade
	m_{ab}	12	percent per decade
	m_{ba}	4.704	percent per decade
	m_{bb}	94.7960	percent per decade
	m_{bc}	0.5	percent per decade
	m_{cb}	0.075	percent per decade
	m_{cc}	99.925	percent per decade
Temperature flow coefficients	t_{aa}	68.85	percent per decade
	t_{ab}	6.45	percent per decade
	t_{ba}	5	percent per decade
	t_{bb}	95	percent per decade
Speed of adjustment	ξ_2	0.208	rate per decade
Climate sensitivity	ξ_1	3.8	C° per 2X CO_2
Preindustrial atmospheric carbon	M_{aPI}	596.4	Giga Tons C
Exogenous forcing in 2000	$Fexo2000$	-0.06	—
Exogenous forcing in 2100	$Fexo2100$	0.3	—

Table 2.1: Regions in RICE 2010

Index	Region
1	USA
2	OECD Europe
3	Japan
4	Russia
5	Non-Russia Eurasia
6	China
7	India
8	Middle East
9	Africa
10	Latin America
11	Other High Income Countries
12	Other Non-OECD Asia

Bibliography

Anthoff, D. (2011): "Optimal Global Dynamic Carbon Abatement," *University of California, Berkeley, Working Paper.*

Anthoff, D., C. Hepburn, and R. S. Tol (2009): "Equity weighting and the marginal damage costs of climate change," *Ecological Economics*, 68(3), 836–849.

Atkinson, A. B. (1970): "On the measurement of inequality," *Journal of economic theory*, 2(3), 244–263.

Atkinson, G., S. Dietz, J. Helgeson, C. Hepburn, and H. Saelen (2009): "Siblings, Not Triplets: Social Preferences for Risk, Inequality and Time in Discounting Climate Change," *Economics: The Open-Access, Open-Assessment E-Journal.*

Azar, C., and T. Sterner (1996): "Discounting and distributional considerations in the context of global warming," *Ecological Economics*, 19(2), 169–184.

Below, D. (2011): "Optimal carbon taxes with social and private discounting," *PhD dissertation.*

Chichilnisky, G., and G. Heal (1994): "Who should abate carbon emissions?: An international viewpoint," *Economics Letters*, 44(4), 443–449.

Cline, W. R. (1992): *The economics of global warming*. Peterson Institute.

Dalton, H. (1920): "The measurement of the inequality of incomes," *The Economic Journal*, 30(119), 348–361.

Dasgupta, P. (2007): "Commentary: The Stern review's economics of climate change," *National Institute Economic Review*, 199(1), 4–7.

——— (2008): "Discounting climate change," *Journal of Risk and Uncertainty*, 37(2), 141–169.

Dietz, S. (2008): "A long-run target for climate policy: the Stern Review and its critics," *Unpublished report.*

Dietz, S., and G. B. Asheim (2012): "Climate policy under sustainable discounted utilitarianism," *Journal of environmental economics and management*, 63(3), 321–335.

Dietz, S., C. Hope, N. Stern, and D. Zenghelis (2007): "Reflections on the Stern Review," *World Economics*, 8(1), 121–168.

Fankhauser, S., R. S. J. Tol, and D. W. Pearce (1997): "The aggregation of climate change damages: a welfare theoretic approach," *Environmental and Resource Economics*, 10(3), 249–266.

Fleurbaey, M. (2010): "Assessing Risky Social Situations," *The Journal of Political Economy*, 118(4), 649–680.

Golosov, M., J. Hassler, P. Krusell, and A. Tsyvinski (2011): “Optimal taxes on fossil fuel in general equilibrium,” Discussion paper, National Bureau of Economic Research.

Hampicke, U. (2011): “Climate change economics and discounted utilitarianism,” *Ecological Economics*, 72(C), 45–52.

Hepburn, C. J., and W. Beckerman (2007): “Ethics of the discount rate in the Stern Review on the economics of climate change,” *World Economics*, 8(1).

Hope, C. (2006): “The marginal impact of CO2 from PAGE2002: An integrated assessment model incorporating the IPCC’s five reasons for concern,” *Integrated assessment*, 6(1), 19–56.

Hope, C. (2008): “Discount rates, equity weights and the social cost of carbon,” *Energy Economics*, 30(3), 1011–1019.

Hotelling, H. (1931): “The Economics of Exhaustible Resources,” *Journal of Political Economy*, 39(2), 137–175.

Kaplow, L., E. Moyer, and D. A. Weisbach (2010): “The social evaluation of intergenerational policies and its application to integrated assessment models of climate change,” *The BE Journal of Economic Analysis & Policy*, 10(2).

Koopmans, T. C. (1960): “Stationary ordinal utility and impatience,” *Econometrica*, 28, 287–309.

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

Mirrlees, J. (1978): “Social benefit-cost analysis and the distribution of income,” *World Development*, 6(2), 131–138.

NEGISHI, T. (1960): "Welfare economics and existence of an equilibrium for a competitive economy," *Metroeconomica*, 12(2-3), 9297.

NORDHAUS, W., AND J. BOYER (2000): *Warming the world: economic models of climate change*. MITPress, Cambridge, MA, USA.

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

NORDHAUS, W. D. (2010): "Economic aspects of global warming in a post-Copenhagen environment," *Proceedings of the National Academy of Sciences*, 107(26), 11721–11726.

NORDHAUS, W. D., AND Z. YANG (1996): "A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies," *American Economic Review*, 86(4), 741–65.

SCHELLING, T. C. (1992): "Some economics of global warming," *The American Economic Review*, 28, 1–14.

SCHELLING, T. C. (1995): "Intergenerational discounting," *Energy policy*, 23(4-5), 395401.

STERN, N. (2006): "Stern Review on the Economics of Climate Change," .

TOL, R. S. (2001): "Equitable cost-benefit analysis of climate change policies," *Ecological Economics*, 36(1), 71–85.

Chapter 3

Consuming more and polluting less today: intergenerationally efficient climate policy

Chapter abstract

Climate policy benefits future generations at the expense of current ones. We propose a system of transfers that allow future generations to compensate the current one for its mitigation effort and demonstrate the effects in an OLG model. When the marginal benefit to a - possibly distant - future generation is greater than the cost of compensating the current generation for its abatement effort, a Pareto improvement is possible by a combination of mitigation policy and transfer payments. We show that under very general assumptions the business-as-usual outcome is Pareto dominated by such policies and derive the conditions for the set of climate policies that are *not* dominated thus.

This chapter is based on joint research with David von Below and Niko Jaakkola

3.1 Introduction

Most climate policy assessments model greenhouse gases (GHGs) as an externality the emission of which is beneficial at the date it is emitted, while its contribution to the atmospheric stock causes economic damages at all future dates. Policy appraisal under the infinitely lived representative agent paradigm will propose emission abatement to the point where the abatement cost equals the sum of the benefits from the reduction in damages. Since almost all of these benefits accrue after the individuals who incurred the abatement cost have passed away, the abating generation will incur a net loss from such a policy. Alternatively, by treating the affected explicitly as a sequence of generations it is possible to quantify the effect of policy on the welfare of each generation separately and thereby design Pareto improving policies, i.e. ones that make each generation better off.

In the context of a representative agent model, Rezai, Foley, and Taylor (2012) and Foley (2007) propose that abatement effort come combined with a reduction in savings. If the reduction in savings is chosen correctly the consumption in the period during which abatement takes place is not reduced and the benefit to subsequent generations is positive as the lower capital stock is compensated for by reduced climate damages. However, in such a model it is unclear whether a single agent lives across periods, in which case it is not necessary to ensure that consumption is greater in every period as long as the agent's total welfare is increased, or if separate generations live just for one period, in which case the the savings behaviour is unmodeled since agents living for a single period would presumably not save of their own volition.

3.1.1 Main results

In this paper we use an overlapping-generations (OLG) model to explicitly consider the welfare of different generations while maintaining decentralised investment decisions. Emission abatement is interpreted as a public policy prescription which restricts current production possibilities in form of an abatement cost and expands future ones by reducing the accumulation of GHGs and thereby the damages. Capital savings on the other hand are interpreted as the result of the private inter-temporal consumption smoothing of *finitely* lived agents. Current abatement policy creates value well beyond the finite lives of the agents alive at the date of abatement. Therefore, abatement policy will be desirable to agents in the distant future, even once the current ones have done what is optimal amongst themselves. If these distant generations could somehow compensate the current ones for some additional abatement effort, and this compensation has a lower marginal cost *to the compensators* than the marginal benefit of an additional unit of abatement, a Pareto improvement is possible.

We achieve just such a compensation by taking advantage of the contemporaneity of old and young in an OLG model. The young could compensate the old at a date t, in return for costly abatement policy undertaken at $t-1$. This is essentially a pay-as-you-go pension. Furthermore, the young at $t+1$ could compensate the old at $t+1$, who themselves are the young at t and could pass on the compensation to the old at t for even more abatement at $t-1$. This system of reverse transfers can be achieved for an arbitrary number of generations.

Being a pay-as-you-go transfer, such policies have well-known deleterious effects on the incentives to save and therefore on the capital accumulation process. Taking this into account we are able to demonstrate that even for arbitrarily distant generations it is beneficial to enter a contract with the current one (and all the intermediate ones who must act as conduits) in which the current generation abates in return for a pension at a later date. What is more, we can show that once a set of adjacent generations have exhausted all the mutual gains from such contracts, the possibility of adding one yet more distant generation to the contract results in Pareto improvements over the best that the subset of generations could achieve.

We characterise the set of policies that dominate the business-as-usual policy as well as the set of Pareto efficient policies. The condition for efficiency amongst any subset of adjacent generations takes a form that is similar to the Samuelson rule, where efficiency is determined by the policy level at which marginal cost (of abatement) is equal to the sum of the marginal benefits. However, because the beneficiaries are separated in time from the abaters, the compensation mechanism (the sequence of pay-as-you-go transfers) is itself costly, resulting in a modification of the rule whereby the marginal cost of compensating the abater *including the cost of compensation* is equalised to the marginal benefit. We derive an easily interpretable condition for this and use it to establish that such abatement-pension policies can provide significant mutual improvements in welfare.

3.1.2 Literature

Previous literature has covered related issues. As mentioned above Rezai, Foley, and Taylor (2012) and Foley (2007) look at a related question in a representative agent model. Bovenberg and Heijdra (1998) look at Pareto improving mitigation in a Blanchard-Yaari OLG model in which investment is decentralised and the compensation of mitigating generations is achieved contemporaneously by adjustment of public debt. For this mechanism to be possible the initial level of public debt and taxation to service it must be sufficiently high so that its removal is sufficient to compensate the abaters for the cost they incur. Else the government would have to borrow on foreign capital markets to achieve the local Pareto improvement. We model a closed (global) economy but do not face the same constraint as in our model the compensation happens at a later date by a pay-as-you-go mechanism.

Gerlagh and Keyzer (2001) expand a pure exchange OLG model based on Gale (1973) by adding a productive non-renewable natural resource. The main difference to our model is that they abstract from capital accumulation thus removing one dimension of the intergenerational trade-off. Within this model they analyse how different property right arrangements impact the welfare of the generations involved. Comparing zero extraction of the natural resource, ownership of the first generation, and ownership by a trust fund in which all generations have a stake they find that the trust-fund Pareto dominates zero-extraction and leads to a better steady state than ownership by the first generation. Treating the atmosphere as the complement to the fossil fuel resource, our model has a similar interpretation. However, our result shows that the inclusion of future generations in an 'agreement' for how to use the resource can be superior even than the business as usual.

Karp and Rezai (2012) consider long-lived capital stocks in an OLG setup. The productivity of capital depends on the state of the environment, as well as current emissions. They show that an improvement in the future state of the environment, by increaseing future productivity, leads to appreciation in the value of the physical assets held by the current old relative to the wage of the young. Climate policy then induces a cost on both generations alive at the date in question, but since the young must acquire the appreciated assets from the old, the old may achieve a net benefit with the entire cost of abatement falling on the young. Thus, a Pareto improving policy would involve the current beneficiaries of the climate policy– the old who hold the appreciated capital – to compensate the current young. The future generations benefit as well but their marginal benefit is not exhausted to achiever greater abatement policy. The main difference to our paper is that we model the capital stock as endogenous and therefore rule out the form of appreciation of the capital stock relative to the consumption good. The result is that we require transfers going from young to old, rather than from old to young, as well as being able to incorporate the benefit to distant future generations into the current abatement decision.

Howarth and Norgaard (1992) work in a general equilibrium OLG framework similar to ours, but only consider sustainability of consumption paths under social welfare maximisation, and do not consider indexation of the intergenerational transfers to the state of the environment. Instead of sustainability issues, we are more interested in policies which take the economy to points on the Pareto frontier which dominate the business-as-usual outcome. John and Pecchenino (1994) develop an OLG model with two assets, physical and natural capital, both of which create a positive externality on the subsequent generation. The share of output that is invested is exogenous, but the relative shares going into physical and natural capital are made endogenous. They find that the allocation of investment between the two capital stocks may be inefficient, since the agents don't live infinite lives. We endogenise the consumption savings decision and depart from the assumption that insufficient abatement (investment in natural capital) is taking place, because current generations don't take into account the benefit to future ones. We then characterise what the efficient levels of abatement would look like, if the beneficiaries were able to provide the correct compensation to the abaters.

3.1.3 Organisation of the paper

In Section 3.2 we outline the OLG framework and the nature of climate externality. In Section 3.3 we state the main result of the paper, Theorem 4, and discuss its implications. In Section 3.4 we assume specific functional forms for the production and utility functions which allow us to explicitly solve OLG model for the equilibrium prices and quantities. This explicit solution is used throughout Sections 3.5 to 3.7 to demonstrate the intuition of the result, namely that Pareto improvements are possible by linking the magnitudes of pay-as-you-go pensions to the abatement cost. We show that including more generations into such an intergenerational contract in which pensions are used to compensate for the mitigation effort yields greater improvements.

That is to say, accounting for the marginal damage to $N+1$ generations immediately succeeding the abatement policy in question results in greater abatement and a Pareto improvement over the policy that only accounts for first N succeeding generations. Finally, in Section 3.8 we generalise our results to more general functional forms than the ones assumed in Section 3.4. Theorem 6, which is the main result of the section, is the complete statement of Theorem 4.

3.2 The model

Our model consists of a production economy augmented by a persistent externality on output given by the atmospheric stock of greenhouse gases (GHGs). Capital is saved by overlapping generations (OLG) of consumers who live for two periods (see Diamond (1965)). The carbon stock is accumulated as a byproduct of economic output the reduction of which is costly. Denoting gross output by F^t, the cost of emissions reduction (*abatement*) by Φ^t, and the economic damage from climate change by Ψ^t, our model is simply a modification of Diamond's OLG model in which net economic output is given by

$$Y^t = F^t - \Phi^t - \Psi^t. \tag{3.1}$$

3.2.1 Production and damage

In each period t, following date t, the economy is endowed with a gross productive capacity net of depreciation represented by a function, F^t, that has constant returns to scale (CRS) in physical capital and labour. More formally,

$$(K^t, L^t) \mapsto F^t(K^t, L^t). \tag{3.2}$$

For all functions we will omit the arguments when no confusion is possible as to what values these take. When a function is subscripted by an argument it denotes the derivative with respect to the argument in question.

Industrial emissions are modeled as an additional output of production, the quantity of which can be reduced by a costly abatement effort. The GHGs emitted during period t are given by

$$E^t = (\bar{e}^t - a^t) \cdot F^t, \tag{3.3}$$

where $\bar{e}^t$ is an exogenous parameter referred to as the business-as-usual *emission intensity* or *emission-to-output ratio*, and a^t is the abatement effort at date t. The actual emission-to-output ratio is $e^t = \bar{e}^t - a^t$. The cost of abatement during period t is given by

$$\Phi^t = B^t(a^t) \cdot F^t \tag{3.4}$$

where $B^t(a^t)$ is the *abatement cost per unit of output.* This is similar to the approach taken in the integrated assessment model RICE-2010. By varying $\bar{e}^t$ and the functional form of B^t this formulation is sufficiently flexible for the realistic modeling of the productivity, cost of extraction, substitutability and time value of fossil fuel energy inputs.

Greenhouse gas emissions add to the stock of atmospheric carbon. A proportion of the stock dissipates over time and the actual atmospheric stock is presumed to cause climate change and thereby an externality as a damage to *future output.* Following Golosov, Hassler, Krusell, and Tsyvinski (2011), the stock of carbon in the atmosphere at date t is modeled as

$$S^t = \sum_{i=1}^{t-T_0} (1 - d^i) E^{t-i}, \tag{3.5}$$

where T_0 is the date at which industrial GHG emission began, and the *dissipation parameters* have the properties $d^i \in (0, 1)$ and $d^i \leq d^{i+1}$. The interpretation of these is as follows. Of the emissions in period t a proportion $(1 - d^i)$ remains in the atmosphere in period $t+i$. The damage at date t is

$$\Psi^t = D^t(S^t) \cdot L^t, \tag{3.6}$$

where $D^t(S_t)$ is the *per capita climate damage*. At this stage it may be useful to point out that there are several ways in which damages can be modeled in such a framework. In a decentralised economy the labour wage is set to the marginal product of labour and the interest rate (return on capital) to the marginal product of capital. If the damage enters as proportional to labour, the climate damages result in a loss only to the wage, as can be seen from equations (3.11) and (3.12). Conversely, if the damages enter as proportional to capital, rather than labour, the damages result in a loss on the return on capital. The damage can be modeled as proportional to *any* CRS function of labour and capital without disrupting the decentralised logic that labour and capital shares exhaust output. Damages could thus be proportional to output itself, or any other function aggregating labour and capital in a CRS manner. The result would be that loss terms would appear both in the wage and interest rates ((3.11) and (3.12)) in some proportion. That is to say the way in which the damage function enters net output determines how the loss is distributed amongst the different parts of the population.

The choice of damage function also determines how the magnitude of the effect of a given stock of carbon changes over time. In this regard, modeling damages as proportional to capital or output may seem more desirable than to population. The magnitude of the damage at a given stock of atmospheric carbon will depend on what is affected (rather than who bears the loss). The cost of the loss of coral reefs and UNESCO world heritage sites can reasonably be thought of as proportional to the amount of people who will be around to enjoy them. There are, of course, damages that are more likely to be proportional to capital.

The choice to model damages as proportional to population is based on the incidence of the loss. Having the loss accrue exclusively to the wage earners without affecting the capital return simplifies the accounting of damages as each generation is only affected once, during the period it earns its wages. This allows for a clearer exposition of the main features, at the possible cost of misrepresenting the magnitude. If we relax the assumption of perfect competition, and impose that the complete loss – whether the damage affects capital productivity or labour productivity – falls on the wage share and none on the capital share (as in (3.11) and (3.12)), then the results of this chapter extend to any specification of the damage function.

3.2.2 Net output

Under assumptions (3.4) and (3.6) net economic output becomes

$$Y^t(K^t, L^t, S^t, a^t) = (1 - B^t(a^t))F^t(K^t, L^t) - D^t(S^t)L^t \tag{3.7}$$

The damage functions D^t and abatement costs B^t are taken to satisfy the following assumptions.

$$D^t_S(\cdot) > 0; \quad D^t_{SS}(\cdot) \geq 0 \tag{3.8}$$

$$B^t_a(\cdot) \geq 0; \quad B^t_{aa}(\cdot) > 0 \tag{3.9}$$

$$B^t(0) = B^t_a(0) = 0; \quad \lim_{a \to \bar{e}^t} B^t_a(a) = \infty \tag{3.10}$$

Conditions (3.8) state that the damage is increasing and the marginal damage is weakly increasing as a function of the carbon stock. The convexity assumption is not strictly necessary, but simplifies the exposition as the alternative *necessary* condition is more involved and has a less intuitive interpretation. In (3.9) we assume that abatement cost is weakly positive and *stricly* convex as a function of abatement. Furthermore, at zero abatement both the cost and the marginal cost are assumed zero. The presumption underlying the latter is that $\bar{e}^t$ is set at just the emission level that is considered optimal by the competitive factor markets at date t, and thus the first unit of emissions reduction is virtually costless. Condition (3.10) is akin to an Inada condition and guarantees that the first order conditions are satisfied in the interior, $a^t \in (0, \bar{e}^t)$.

3.2.3 Firms

Economic output is produced by firms that are modeled as perfectly competitive and profit maximising. This yields the well-known result that the *wage* and *interest rates* are equated to the marginal products of labour and capital respectively:

$$w^t = \left[1 - B^t(a^t)\right] F_L^t(K^t, L^t) - D^t(S^t) \tag{3.11}$$

$$r^t = \left[1 - B^t(a^t)\right] F_K^t(K^t, L^t). \tag{3.12}$$

The net production function (equation (3.7)) has constant returns to scale in capital and labour which ensures that output is exhausted by the labour and capital shares, i.e.

$$Y^t = W^t + R^t := w^t L^t + r^t K^t.$$

We refer to W^t as the total wage and R^t as the *capital rent*.

3.2.4 Overlapping generations of consumers

Consumers live for two periods, *youth* and *retirement*, and are grouped together into homogenous generations consisting of one unit of population – or labour. The generation born at date t is denoted by $\mathcal{G}_t$. It derives utility from consumption goods and saves capital to transfer them between youth and retirement. The savings of generation $\mathcal{G}_t$ constitute the entire capital stock at date $t+1$. Denoting $\mathcal{G}_t$'s youth and retirement consumption by C^{1t} and C^{2t+1} its maximisation problem can be written as

$$\max_{C^{1t},C^{2t+1}} U(C^{1t}, C^{2t+1}) : \tag{3.13}$$

$$C^{1t} + K^{t+1} = M^t \tag{3.14}$$

$$C^{2t+1} = \left[1 + r^{t+1}\right] K^{t+1} + Z^{t+1} \tag{3.15}$$

where the endowments of consumption goods during youth and retirement are denoted by M^t and Z^{t+1} and the rate at which wealth can be transferred between the two periods is given by the growth factor $(1 + r^{t+1})$.

Generations are born without assets and earn a wage during youth. Absent inter-generational transfers the entire wage is either consumed during youth or saved for retirement when the entire principal plus the capital rent is consumed; there are no bequests. In period t generation $\mathcal{G}_t$ pays $\mathcal{G}_{t-1}$ a (possibly nil) pension denoted P^t. Including these transfers the endowments in the budget constraints (3.14) and (3.15) become

$$M^t = w^t - P^t \tag{3.16}$$

$$Z^{t+1} = P^{t+1}. \tag{3.17}$$

In addition to the endowments M^t and Z^{t+1} the capital rent $R^{t+1} = r^{t+1}K^{t+1}$ is the remaining source of wealth to $\mathcal{G}_t$.

Consider abatement at date m. As we explain in greater detail later, this will have a detrimental effect on the welfare of the generations alive during period m, and a positive effect on all future generations. The former can be seen by the effect abatement has in reducing the wage of $\mathcal{G}_m$ (equation (3.11)) and the capital rent of $\mathcal{G}_{m-1}$ (equation (3.12)). The latter is a consequence of the fact that such an effort will reduce the carbon stock *in all future periods* and thereby the damages in all periods $t \geq m+1$. Notice that climate damages enter only in the wage and not in the capital rent (equations (3.11) and (3.12)). Therefore, even though $\mathcal{G}_m$ will be alive at $m+1$, when the first benefits from reduced damages appear, it will *not* benefit from the abatement at date m. The benefits will only accrue to generations from $\mathcal{G}_{m+1}$ onwards.

As a function of the wealth variables, the solution to optimisation (3.13) yields a *savings function* such that $K^{t+1} = s(M^t, Z^{t+1}, r^{t+1})$. Recall that given the factor market equilibrium (equations (3.11) and (3.12)) and pension transfers (equations (3.16) and (3.17)) the three arguments of the the savings function are given by

$$M^t = \left[1 - B^t(a^t)\right] F_L^t - D^t(S^t) - P^t, \tag{3.18}$$

$$Z^{t+1} = P^{t+1}, \tag{3.19}$$

$$r^{t+1} = \left[1 - B^{t+1}(a^{t+1})\right] F_K^{t+1}(K^{t+1}, L^t). \tag{3.20}$$

The equilibrium capital is therefore determined as the fixed point K_*^{t+1}:

$$K_*^{t+1} = s\left(M^t, Z^{t+1}, \left[1 - B^{t+1}\right] F_K^{t+1}(K_*^{t+1}, L^t)\right). \tag{3.21}$$

How these variables affect the welfare of $\mathcal{G}_t$ is measured by the *value function*, which is defined as the utility evaluated at the consumption levels that maximise (3.13), i.e.

$$V^t = U(M^t - K_*^{t+1}, K_*^{t+1} + r^{t+1} K_*^{t+1} + Z^{t+1}). \tag{3.22}$$

where the equilibrium capital, K_*^{t+1}, is defined by (3.21).

Defining the equilibrium capital rent, $R_*^{t+1} = r^{t+1} K_*^{t+1}$, we have that M^t, Z^{t+1}, and R_*^{t+1} are the sources of wealth of $\mathcal{G}_t$, therefore changes in the policies that increase these variables will be beneficial to $\mathcal{G}_t$'s welfare and changes that reduce them will be detrimental. The relative magnitudes of any gains or losses are essentially the content of the value function (3.22). However, due to the general equilibrium adjustment of capital – equation (3.21) – the relationship is not a straightforward one. The effect of policies on welfare will depend jointly on (3.18), (3.19), (3.20), (3.21) and (3.22).

3.3 Main result

Denote the initial period by m. Consider a policy vector

$$\mathcal{P} = (a^m, P^m, a^{m+1}, P^{m+1}, a^{m+2}, P^{m+2}, \ldots, a^{m+N}, P^{m+N}, \ldots) \in \mathbb{R}^\infty \tag{3.23}$$

At any date t, the policies a^t and P^t are implemented by the generations alive, $\mathcal{G}_{t-1}$ and $\mathcal{G}_t$. Therefore, any complete policy vector $\mathcal{P}$ must be implemented as an agreement or contract between all generations, starting at date m. This raises issues of credibility of such a 'commitment' policy from which we currently abstract. In the following we will analyse policy vectors, which improve the welfare of all generations, conditional on the credibility of the policy. Extensions analysing the strategic aspect of such policies are under consideration.

Assume that policies before date m were zero. A given history $(\ldots, K^{m-1}, K^m)$ combined with a vector $\mathcal{P}$ completely determines the equilibrium evolution of capital, $\{K^t\}_{t>m}$, and carbon, $\{S^t\}_{t>m}$ (See (3.5) and (3.21)). By (3.22) this also determines the sequence of values $\{V^t\}_{t\geq m}$. We will write $V^t(\mathcal{P})$ to denote the dependence of these on the policy, while suppressing the dependence on the history.

Loosely speaking our main result states that any policy $\mathcal{P}_N$ which which reverts to the zero policy $(a^t = 0, P^t = 0)$ from date $N+1$ onwards can be Pareto improved upon by one which includes non-zero policies at date $N+1$. More formally,

Definition 1. *A policy $\mathcal{P}$ is efficient between the set of generations $\{\mathcal{G}_{m+i} : i =$*

$0, 1, \ldots, N\}$ *if there is no policy* $\mathcal{Q}$ *such that*

$$V^t(\mathcal{Q}) \geq V^t(\mathcal{P}), \quad \forall t = m+1, m+2, \ldots, m+N$$

with a strict inequality for at least one t.

Theorem 4. *Suppose*

$$\mathcal{P}_N = (a_N^m, P_N^m, a_N^{m+1}, P_N^{m+1}, a_N^{m+2}, P_N^{m+2}, \ldots, a_N^{m+N}, P_N^{m+N}, \ldots) \in \mathbb{R}_+^\infty$$

is efficient between $\{\mathcal{G}_{m+i} : i = 0, 1, \ldots, N\}$ *and that* $P_N^t = 0$ *for all* $t \geq m + N$. *Then there is a policy*

$$\mathcal{P}_{N+1} = (a_{N+1}^m, P_{N+1}^m, a_{N+1}^{m+1}, P_{N+1}^{m+1}, a_{N+1}^{m+2}, P_{N+1}^{m+2}, \ldots, a_{N+1}^{m+N}, P_{N+1}^{m+N}, \ldots)$$

such that

1. $a_{N+1}^m > a_N^m$
2. $P_{N+1}^t > P_N^t$ *for all* $m \leq t \leq N+1$
3. $V^t(\mathcal{P}_{N+1}) \geq V^t(\mathcal{P}_N)$ *for all* $m \leq t \leq N+1$

where strict inequality holds in 3 for at least one generation.

According to point *3*, policy $\mathcal{P}_{N+1}$ Pareto dominates $\mathcal{P}_N$ for the generations under consideration. Thus, any policy $\mathcal{P}_N$ that is efficient between the first N generations is Pareto dominated by a policy $\mathcal{P}_{N+1}$ which includes the $(N+1)$'th generation. The improvement involves a higher initial abatement level (point *1*, $a_{N+1}^m > a_N^m$), and greater pension transfers in between all generations from the current one to the $(N+1)$'th (point *2*, $P_{N+1}^t > P_N^t$). Since the damages are persistent, considering efficiency between any subset of generations will no exhaust all gains. The theorem shows that such additional gains can be distributed in such a way that the very first generation incurring the abatement cost is sufficiently compensated.

Consider the *business-as-usual* policy, denoted by $\mathcal{P}_0^m \equiv \vec{0}$. This is the policy for which both abatement and pensions are always zero.[1] Note that $\mathcal{P}_0^m$ is efficient between the generations alive during period m. This is because for the generations alive at date m, $\mathcal{G}_{m-1}$ and $\mathcal{G}_m$, abatement is costly at no benefit as the benefit only starts to accrue to the next, currently unborn, generation, $\mathcal{G}_{m+1}$. Applied to $\mathcal{P}_0^m$, the theorem states that a Pareto improvement is possible by increasing the abatement level in period m, a^m and the pension levels P^m and P^{m+1}. The intuition behind this is that since $\mathcal{G}_{m+1}$ benefits from abatement at date m, it is possible to find a pension P^{m+1} it would be willing to pay which is sufficiently high to compensate $\mathcal{G}_{m-1}$ and $\mathcal{G}_m$ for the abatement costs they incur.

What is remarkable about our result is that the same logic holds for any future beneficiary of *current* abatement policy, no matter how distant. Provided the right sequence of intermediate pensions are implemented, Pareto improvements can be achieved by redistributing part of the benefit of the future beneficiary back to the current generation incurring the abatement cost.

Since policies that are not efficient between a subset of generations, $\{\mathcal{G}_{m+i} : i = 0, 1, \ldots, N\}$, are dominated by an efficient one, and for any efficient one, we can find a policy that is superior to it by including a further generation into the agreement, the theorem shows that *any* policy can be dominated by one which includes further generations into the agreement.

Furthermore, we derive a condition, similar to the Samuelson rule which determines whether a given policy is efficient for the inclusion of a given number, N, of generations. The condition tells us whether efficiency holds and the direction in which

[1]Pensions may actually be at some positive level reflecting the fact that pay-as-you-go pensions exist independently of abatement policy. The essential feature of the business as usual policy is not altered by that, since only differences in pensions matter.

policies must be adjusted if efficiency doesn't hold. This makes a numerical policy assessment based on a calibration of the model easy to implement.

3.4 Specific functional forms

In this section we will make assumptions about the functional forms of the production and utility functions which allow us to get a simple and intuitive expression for the value function – equation (3.22) – as well as an analytical description of the endogenous capital accumulation process which highlights the key equilibrium effects – equation (3.21). These assumptions will be taken to hold in the analysis of Sections 3.6 through 3.7 before we state the more general result in Section 3.8.

3.4.1 Leontief utility and logarithmic production

Definition 2. *The Leontief utility for the two-period inter-temporal consumption problem is defined by*

$$(C^1, C^2) \mapsto \min\{\beta C^1, C^2\}. \tag{3.24}$$

The parameter β quantifies the relative preference of second-period over first-period consumption.

The solution to the optimisation (3.13) with Leontief utility is given by the savings function

$$K^{t+1} = \frac{\beta M^t - Z^{t+1}}{1 + \beta + r^{t+1}} \tag{3.25}$$

Setting $\beta C^{1t} = C^{2t+1}$ and solving (3.14) and (3.15) for K^{t+1} as a function of M^t, Z^{t+1} and r^{t+1} establishes this.

Definition 3. *The logarithmic production function in capital and labour inputs is defined by*

$$(K^t, L^t) \mapsto A^t L^t \ln\left(\frac{K^t}{L^t}\right) \tag{3.26}$$

Note that the function has constant returns to scale in capital and labour.

As is shown in the Appendix, in the acceptable domain ($K \geq L \cdot e$) the function has the requisite properties of decreasing and convex isoquants in (L, K)-space. That is, given the right choice of units, it is a well-behaved production function. Notice that

$$F_K^t = \frac{A^t L^t}{K^t} \tag{3.27}$$

and therefore the gross capital rent $F_K^{t+1} K^{t+1} = A^{t+1} L^{t+1}$. That is, the capital rent is independent of the amount of capital saved.[2] The term $R_*^{t+1} = r^{t+1} K_*^{t+1}$ in (3.22) is therefore *independent of the equilibrium capital* K_*^{t+1}. We can therefore drop the asterisk subscript in R^{t+1}. This fact, along with the simple savings function resulting from the Leontief utility provides for the great simplification of the value and capital accumulation equation that results from Proposition 4.

Proposition 4. *Suppose $U(C^{1t}, C^{2t+1})$ is Leontieff, $F^t(K^t, L^t)$ is logarithmic and*

[2]In this respect the logarithmic production technology can be seen as the limit of the Cobb-Douglas production technology with vanishing capital share. Recall the Euler identity for homogenous functions, $F(K, L) = F_L \cdot L + F_K \cdot K$. By differentiating both sides with respect to K, we get that for every CRS production function:

$$F_K = \frac{\partial}{\partial K}(F_L L) + \frac{\partial}{\partial K}(F_K K), \tag{3.28}$$

i.e. the increase in output due to an increase in the amount of productive capital – the productivity of capital – is shared between the wage and the capital rent. In our notation equation (3.28) becomes $r^t = W_K^t + R_K^t$. For the Cobb-Douglas function with capital share α we have that

$$W_K^t = (1 - \alpha) r^t; \quad R_K^t = \alpha r^t$$

For the logarithmic production function it is as if $\alpha = 0$:

$$W_K^t = r^t; \quad R_K^t = 0$$

Another way to see that the second order properties mimic a Cobb-Douglas function with α going to zero is the limit below:

$$L \ln\left(\frac{K}{L}\right) = \lim_{\alpha \to 0} \frac{K^\alpha L^{1-\alpha} - L}{\alpha}$$

$K^t \geq L^t \cdot e$. Then the equilibrium interest rate is

$$r^{t+1} = \frac{[1+\beta]R^{t+1}}{\beta M^t - Z^{t+1} - R^{t+1}} \tag{3.29}$$

and the equilibrium consumption and savings of $\mathcal{G}_t$ are given by

$$C^{1t} = \frac{M^t + Z^{t+1} + R^{t+1}}{1+\beta} = \frac{C^{2t+1}}{\beta} \tag{3.30}$$

$$K_*^{t+1} = \frac{\beta M^t - Z^{t+1} - R^{t+1}}{1+\beta} \tag{3.31}$$

where

$$M^t = \left[1 - B^t(a^t)\right] F_L^t(K^t) - D^t(S^t) - P^t \tag{3.32}$$

$$Z^{t+1} = P^{t+1} \tag{3.33}$$

$$R^{t+1} = \left[1 - B^{t+1}(a^{t+1})\right] A^{t+1} \tag{3.34}$$

Proof. Note that (for $L^t = 1$) by (3.12) and (3.27)

$$r^{t+1} = \left[1 - B^{t+1}\right] \frac{A^{t+1}}{K^{t+1}} = \frac{R^{t+1}}{K^{t+1}} \tag{3.35}$$

Substituting (3.35) into the Leontief savings function (3.25) yields

$$K_*^{t+1} = \frac{\beta M^t - Z^{t+1}}{1 + \beta + R^{t+1}/K_*^{t+1}} \tag{3.36}$$

Therefore

$$K_*^{t+1}\,[1+\beta] + R^{t+1} = \beta M^t - Z^{t+1} \tag{3.37}$$

and thus,

$$K_*^{t+1} = \frac{\beta M^t - Z^{t+1} - R^{t+1}}{1+\beta}$$

This establishes (3.31). The interest rate is shown to be (3.29) by substituting the equilibrium capital into (3.35) and the consumptions are shown to be (3.30) by substituting the equilibrium capital into the budgets (3.14) and (3.15). □

With the analytic expression for the equilibrium capital (3.31) we will now drop the asterisk subscript and in subsequent sections and refer to the equilibrium capital simply by K^{t+1}. Furthermore, having assumed that the population is constant across time and fixed at $L^t = 1$, we will drop the population argument in the production function and its derivatives.

3.5 Value, policies, and states

This section summarises all the effects that policy and state variables have on the welfare of different generations as well as the effect on the evolution of the capital stock. These relationships are used repeatedly throughout the subsequent sections.

3.5.1 The effect of policy on different generations' welfare

In equilibrium $\mathcal{G}_t$'s youth and retirement consumptions satisfy $C_{2t+1} = \beta C_{1t}$ and therefore its equilibrium utility simplifies to

$$U(C^{1t}, C^{2t+1}) = \min\{\beta C^{1t}, C^{2t+1}\} = \beta C^{1t}.$$

Thus, the value of $\mathcal{G}_t$ is simply the equilibrium retirement consumption. For notational convenience we will renormalise the utility function by $(1+\beta)/\beta$ so by equation (3.30) we get the value

$$V^t = M^t + Z^{t+1} + R^{t+1}.$$

Replacing for M^t, Z^{t+1} and R^{t+1} with (3.32), (3.33) and (3.34) we get the value as a function of the state and policy variables $(K^t, S^t, a^t, P^t, a^{t+1}, P^{t+1})$:

$$V^t = \left[1 - B^t(a^t)\right] F_L^t(K^t) - D^t(S^t) - P^t + P^{t+1} + \left[1 - B^{t+1}(a^{t+1})\right] A^{t+1} \tag{3.38}$$

Consider the direct effects that changing policy variables in period m has on the welfare of $\mathcal{G}_m$ and $\mathcal{G}_{m-1}$. Abatement in period m has a negative effect on the wage of $\mathcal{G}_m$ and on the capital rent of $\mathcal{G}_{m-1}$ thus reducing their wealth and thereby welfare:

$$V_{a^m}^m = -B_a^m F_L^m; \quad V_{a^m}^{m-1} = -B_a^m A^m \tag{3.39}$$

A pension transfer P^m simply reduces the welfare of $\mathcal{G}_m$ and increases that of $\mathcal{G}_{m-1}$ by the same unit:

$$V_{P^m}^m = -1; \qquad V_{P^m}^{m-1} = 1 \tag{3.40}$$

3.5.2 The effect of policy on capital accumulation

As in (3.38) replace for M^t, Z^{t+1} and R^{t+1} in (3.31) to get

$$K^{t+1} = \frac{\beta\left(\left[1 - B^t(a^t)\right] F_L^t(K^t) - D^t(S^t) - P^t\right) - P^{t+1} - \left[1 - B^{t+1}(a^{t+1})\right] A^{t+1}}{1+\beta} \tag{3.41}$$

Consider the effect of abatement on the capital stock. If anticipated in period $m-1$ we have that

$$K_{a^m}^m = \frac{B_a^m A^m}{1+\beta}. \tag{3.42}$$

That is, a promise of future abatement increases the the incentive to save for $\mathcal{G}_{m-1}$ as the abatement reduces its capital rent ($R^m = [1 - B^m(a^m)]A^m$). A pension has the opposite effect, as it increases $\mathcal{G}_{m-1}$'s retirement budget, thus reducing the incentive to save:

$$K^m_{P^m} = \frac{-1}{1+\beta} \tag{3.43}$$

The effect a^m and P^m on K^{m+1} is negative as both reduce the endowment M^m.

$$K^{m+1}_{a^m} = -\frac{\beta}{1+\beta} B^m_a F^m_L \tag{3.44}$$

and

$$K^{m+1}_{P^m} = -\frac{\beta}{1+\beta} \tag{3.45}$$

3.5.3 The value of capital

The effect of a change in the capital stock at date m on the value of $\mathcal{G}_m$ is given by

$$V^m_K = \left[1 - B^t\right] F^t_{LK} \tag{3.46}$$

Since $F^m_{LK} = F^m_L$ for the logarithmic production function, by (3.12), (3.46) becomes

$$V^m_K = r^m \tag{3.47}$$

3.6 Abatement at date m

Consider the decision to abate at date m. As can be seen from the equations (3.39), a^m has a negative effect on $\mathcal{G}_m$ and $\mathcal{G}_{m-1}$, the two generations alive during that period. The benefits due to a reduction in the carbon stock only appear in the values of $\{\mathcal{G}_t; t > m\}$. Thus, it is not in the economic interest of those alive in period m to choose any positive level of abatement.

However, since $\mathcal{G}_m$ is alive in period $m+1$, it can take compensation from $\mathcal{G}_{m+1}$ in form of a pension P^{m+1} to cover the abatement cost (to $\mathcal{G}_m$ and $\mathcal{G}_{m-1}$). Without the dynamic features of the problem the Samuelson rule would state that Pareto improvements are possible if the marginal abatement costs to $\mathcal{G}_{m-1}$ and $\mathcal{G}_m$ are less than the marginal damage avoided by $\mathcal{G}_{m+1}$. However, such an improvement would require a side-payment, or it wouldn't be a Pareto improvement at all. Because the abaters and beneficiaries are separated in time additional effects come into play, and therefore efficiency is *not* characterised by marginal cost being equal to marginal benefit. In this section we propose the pay-as-you-go pension transfer as the mechanism for the side payment and characterise the efficient policies as those upon which no Pareto improvements are possible, *given that the side-payments are themselves costly as they have a negative effect on the accumulation of capital.* We establish that under general assumptions the business-as-usual policy of zero abatement is dominated by a set of policies involving non-zero abatement and pension levels.

3.6.1 Pareto improving abatement

Throughout the paper the notation dx will be used to denote a sufficiently small change in the variable x so that the effects on the related variables may be described by the first order approximation. In this section we determine the conditions on the relative magnitudes of such small changes in the policies a^m, P^m and P^{m+1} such that the result is an improvement to the welfares of $\mathcal{G}_{m-1}, \mathcal{G}_m$ and $\mathcal{G}_{m+1}$. Any point at which such a joint improvement is *not* possible will be said to be on the *efficiency frontier.*

3.6.1.1 Indifference of $\mathcal{G}_{m-1}$ and insensitivity of K^m

To better focus on the improvements possible to future generations we will restrict ourselves to policies that leave $\mathcal{G}_{m-1}$ indifferent. Such policies have the effect of leaving the incentives to save of $\mathcal{G}_{m-1}$ unchanged and therefore have no effect on the capital stock K^m *even if the policies were anticipated.* For this purpose, we will fix the pension P^m at

$$P^m = B^m(a^m)A^m \tag{3.48}$$

When this is the case the value of $\mathcal{G}_{m-1}$ (see equation (3.38)) is

$$V^{m-1} = F_L^{m-1} - D^{m-1} - P^{m-1} + A^m + \overbrace{B^m(a^m)A^m}^{P^m} - B^m(a^m)A^m$$

Thus, the cost of the abatement effort is compensated in full by the pension. Furthermore, if both a^m and P^m are anticipated and (3.48) holds, the effect on $\mathcal{G}_{m-1}$'s savings behaviour is neutralised. From equation (3.41) we get that

$$K^m = \frac{\beta\left(F_L^{m-1} - D^{m-1} - P^{m-1}\right) - A^m - B^m(a^m)A^m + B^m(a^m)A^m}{1+\beta}$$

3.6.1.2 The welfare of $\mathcal{G}_m$

The share of the abatement cost that accrues to $\mathcal{G}_m$ is proportional to the gross wage $B^m(a^m)F_L^m$. Setting the pension that $\mathcal{G}_m$ must pay $P^m = B^m(a^m)A^m$ results in a total cost to $\mathcal{G}_m$ of

$$B^m(a^m)[F_L^m + A^m] = B^m(a^m)[F_L^m + A^m] = B^m(a^m)F^m$$

Setting the pension that $\mathcal{G}_m$ pays $\mathcal{G}_{m-1}$ to just the level that makes $\mathcal{G}_{m-1}$ indifferent results in a combined (pension and abatement) cost to $\mathcal{G}_m$ that is exactly equal to the total abatement cost $\Phi^m = B^m(a^m)F^m$ (see (3.4)). Taking (3.48) to hold as an identity throughout the remaining analysis, we can rewrite the value of $\mathcal{G}_m$ to explicitly reflect this

$$V^m = F_L^m(K^m) - D^m(S^m) - B^m(a^m)F^m(K^m) + P^{m+1} + [1 - B^{m+1}(a^{m+1})]A^{m+1} \tag{3.49}$$

Thus, the effect of $\mathrm{d}a^m$ and $\mathrm{d}P^{m+1}$ on the welfare of $\mathcal{G}_m$ is

$$\mathrm{d}V^m = -\mathrm{d}a^m B_a^m F^m + \mathrm{d}P^{m+1}. \tag{3.50}$$

Therefore, $\mathrm{d}V^m \geq 0$ if and only if

$$B_a^m F^m \mathrm{d}a^m \leq \mathrm{d}P^{m+1}. \tag{3.51}$$

By a minor abuse of terminology will will refer to the curves in $(a^m, P^{m+1}) - space$ along which $\mathcal{G}_m$ has constant utility as *indifference curves*. The slope of these is given by

$$\left.\frac{dP}{da}\right|_{U^m} := B_a^m(a^m)F^m \tag{3.52}$$

Notice that the magnitude of the slope (3.52) depends on the level of abatement a^m, but *not* on the pension level P^{m+1}. This is because the marginal cost for which $\mathcal{G}_m$ must be compensated is an increasing function of a^m, but independent of the pension level. By (3.51), any (sufficiently small) *increase* in policy $(\mathrm{d}a^m, \mathrm{d}P^{m+1})$ such that

$$\mathrm{d}a^m \cdot \left.\frac{dP}{da}\right|_{U^m} < \mathrm{d}P^{m+1} \tag{3.53}$$

will lead to a welfare level for $\mathcal{G}_m$ that is greater than U_m.

3.6.1.3 The welfare of $\mathcal{G}_{m+1}$

Using equation (3.38), the change in value to $\mathcal{G}_{m+1}$ is given by

$$\mathrm{d}V^{m+1} = (1 - d^1)F^m D_S^{m+1}\mathrm{d}a^m - \mathrm{d}P^{m+1} + V_K^{m+1}\mathrm{d}K^{m+1} \tag{3.54}$$

The first term in $\mathrm{d}V^{m+1}$ contains the beneficial effect of abatement at date m on the welfare of $\mathcal{G}_{m+1}$: the reduction of damages through the lower carbon stock. The second term is simply the direct cost of the pension it must pay as a reduction of its wealth. In addition to those effects, *both* a^m and P^{m+1} have a negative effect on capital accumulation. By (3.43) and (3.44) we get that

$$\mathrm{d}K^{m+1} = \frac{-\beta B_a^m F^m \mathrm{d}a^m - \mathrm{d}P^{m+1}}{1+\beta} \tag{3.55}$$

The value to $\mathcal{G}_{m+1}$ of an additional unit of capital $\mathrm{d}K^{m+1}$ is $V_K^{m+1} = r^{m+1}$ (see (3.47)). Thus (3.54) becomes

$$\mathrm{d}V^{m+1} = (1-d^1)F^m D_S^{m+1}\mathrm{d}a^m - \mathrm{d}P^{m+1} + r^{m+1}\left[\frac{-\beta B_a^m F^m \mathrm{d}a^m - \mathrm{d}P^{m+1}}{1+\beta}\right]$$

and $\mathrm{d}V^{m+1} \geq 0$ if and only if

$$\mathrm{d}P^{m+1}\left[1 + \frac{r^{m+1}}{1+\beta}\right] \leq \mathrm{d}a^m\left[(1-d^1)F^m D_S^{m+1} - \frac{\beta r^{m+1} B_a^m F^m}{1+\beta}\right] \tag{3.56}$$

The left-hand-side of (3.56) contains the twofold negative effect on the welfare of $\mathcal{G}_{m+1}$ of paying a pension $\mathrm{d}P^{m+1}$. In addition to the direct cost of the pension, such an anticipated transfer of income in retirement will have the effect of reducing the previous generation's capital savings, and thus result in a lower capital stock at $m+1$.[3] The right-hand-side contains the benefit to $\mathcal{G}_{m+1}$ from abatement due to reduced climate damages as well as the negative effect costly abatement policy has on the incentives to save and therefore the capital accumulation of $\mathcal{G}_m$.

In analogy to (3.52), the slope in $(a^m, P^{m+1}) - space$ of $\mathcal{G}_{m+1}$'s indifference curves is given by

$$\left.\frac{dP}{da}\right|_{U^{m+1}} := \frac{(1-d^1)F^m D_S^{m+1} - \frac{\beta r^{m+1} B_a^m F^m}{1+\beta}}{1 + \frac{r^{m+1}}{1+\beta}} \tag{3.57}$$

[3]This is the well known effect that pay as you go pensions have on the capital accumulation process.

By (3.56) any (sufficiently small) *increase* in the policies $(\mathrm{d}a^m, \mathrm{d}P^{m+1})$ such that

$$\mathrm{d}P^{m+1} < \mathrm{d}a^m \cdot \left.\frac{dP}{da}\right|_{\bar{U}^{m+1}} \tag{3.58}$$

leads to an outcome in which the welfare level of $\mathcal{G}_{m+1}$ is greater than U^{m+1}. Thus, by (3.53) and (3.58) if

$$\left.\frac{dP}{da}\right|_{\bar{U}^m} < \left.\frac{dP}{da}\right|_{\bar{U}^{m+1}} \tag{3.59}$$

there exists an increase in the policies (a^m, P^{m+1}) that constitutes an improvement to both $\mathcal{G}_m$ and $\mathcal{G}_{m+1}$.[4]

Proposition 5. *Suppose that the gross production function is logarithmic and each generation has Leontief utility. Then there exist policies* $(\mathrm{d}a^m, \mathrm{d}P^m, \mathrm{d}P^{m+1})$ *such that*

$$\mathrm{d}a^m > 0 \tag{3.60}$$

$$\mathrm{d}P^m = B_a^m A^m \mathrm{d}a^m \tag{3.61}$$

$$0 < \mathrm{d}P^{m+1} < \frac{(1-d^1)F^m D_S^{m+1}(S^{m+1})}{1+\frac{r^{m+1}}{1+\beta}}\mathrm{d}a^m \tag{3.62}$$

and if $\mathrm{d}a^m$ *is sufficiently small, such policies will keep* $\mathcal{G}_{m-1}$ *indifferent to and make* $\mathcal{G}_m$ *and* $\mathcal{G}_{m+1}$ *strictly better off than the business-as-usual (BAU) policy* $a^m = 0, P^m = 0, P^{m+1} = 0$.

Furthermore, under such a policy the capital stock K^{m+2} *is greater than and the carbon stock* S^{m+2} *is lower than under the (BAU), thus providing greater economic opportunities for the following generation.*

Proof. Firstly, every term in the fraction on the right hand side of (3.62) is positive, so policies satisfying (3.60), (3.61) and (3.62) exist. Denote $\mathcal{G}_m$ and $\mathcal{G}_{m+1}$'s utilities at the BAU by $\bar{U}^m$ and $\bar{U}^{m+1}$ respectively. Recall that at BAU the marginal abatement

[4] If the strict inequality in (3.59) goes the other way, there is a *decrease* in that is a mutual improvement.

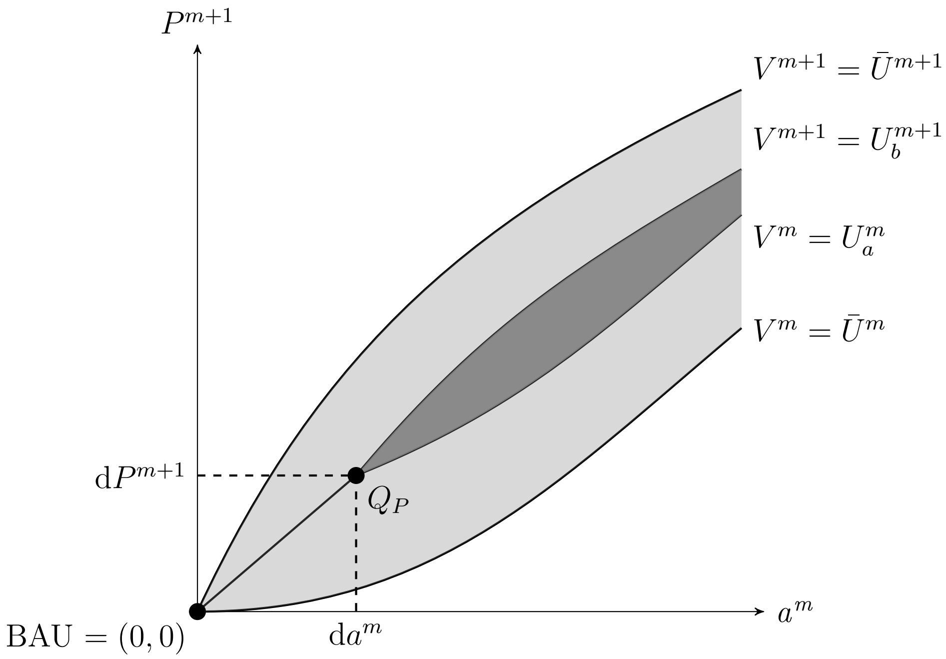


Figure 3.1: Pareto improvements

cost $B_a(0) = 0$. Therefore by (3.52) and (3.57)

$$\left.\frac{dP}{da}\right|_{\bar{U}^m} = 0$$

and

$$\left.\frac{dP}{da}\right|_{\bar{U}^{m+1}} = \frac{(1-d^1)F^m D_S^{m+1}(S^{m+1})}{1+\frac{r^{m+1}}{1+\beta}} > 0$$

and consequently

$$\left.\frac{dP}{da}\right|_{\bar{U}^m} < \left.\frac{dP}{da}\right|_{\bar{U}^{m+1}}$$

Policies $(\mathrm{d}a^m, \mathrm{d}P^{m+1})$ for which (3.60) and (3.62) hold satisfy

$$\left.\frac{dP}{da}\right|_{\bar{U}^m} < \frac{\mathrm{d}P^{m+1}}{\mathrm{d}a^m} < \left.\frac{dP}{da}\right|_{\bar{U}^{m+1}}$$

and by (3.53) and (3.58) are therefore strict improvements to $\mathcal{G}_m$ and $\mathcal{G}_{m+1}$. The condition (3.61) ensures that $\mathcal{G}_{m-1}$ is indifferent.

The policies in question have a twofold effect on the carbon stock S^{m+2}. By (3.3) and (3.5) we have that the change in the carbon stock relative to the business as usual is

$$\mathrm{d}S^{m+2} = -(1-d^2)F^m \mathrm{d}a^m + (\bar{e}^{m+1} - a^{m+1})F_K^{m+1}\mathrm{d}K^{m+1} \tag{3.63}$$

Since $\mathrm{d}K^{m+1}$ is given by (3.55) is must be negative and therefore $\mathrm{d}S^{m+2} < 0$. To see that the change in capital stock K^{m+2} is positive, first note that, due to (3.38) and (3.41) we have that, for all t

$$K^{t+1} = \frac{\beta}{1+\beta}V^t - P^{t+1} - \left[1 - B^{t+1}(a^{m+2})\right] A^{t+1} \tag{3.64}$$

Replacing $t = m+1$ and noting that $\mathrm{d}P^{m+2} = \mathrm{d}a^{m+2} = 0$ we get that

$$\mathrm{d}K^{m+2} = \mathrm{d}V^{m+1}\beta/(1+\beta)$$

Since $\mathrm{d}V^{m+1} > 0$ due to the fact that the policies were (strictly) Pareto improving, we get that $\mathrm{d}K^{m+2} > 0$. □

Proposition 5 is illustrated in Figure 3.1. At the BAU the slope of the indifference curve of $\mathcal{G}_m$ is zero and the slope of the indifference curve of $\mathcal{G}_{m+1}$ is strictly positive. The slope of the segment BAU-Q_P is strictly in between the two, and therefore Q_P is inside the lightly grey shaded area, which contains all possible mutual improvements to $\mathcal{G}_m$ and $\mathcal{G}_{m+1}$. In the figure the slopes of the indifference curve that go through Q_P are drawn so that $\mathcal{G}_{m+1}$'s is steeper than $\mathcal{G}_m$'s. By the same logic the dark grey shaded area consists of the improvements over Q_P.

The indifference curves are drawn as convex and concave respectively. That this is in fact the case is shown in the Appendix.

3.6.2 Efficiency between $\mathcal{G}_{m-1}, \mathcal{G}_m$ and $\mathcal{G}_{m+1}$

From the preceding discussion it is clear that the points in $(a^m, P^{m+1}) - space$ at which the slopes of the indifference curves are equal, i.e.

$$\left.\frac{dP}{da}\right|_{U^m} = \left.\frac{dP}{da}\right|_{U^{m+1}}$$

no improvement to both $\mathcal{G}_m$ and $\mathcal{G}_{m+1}$ is possible.[5] By equations (3.52) and (3.57) this is the case when

$$\left[1 + r^{m+1}(a^m, P^{m+1})\right] B_a^m(a^m)F^m - (1 - d^1)F^m D_S^{m+1}(S^{m+1}(a^m)) = 0 \qquad (3.65)$$

Condition (3.65) contains explicitly all the dependences on the policies a^m and P^{m+1}.[6] This makes it clear that the condition defines a locus in $(a^m, P^{m+1}) - space$, which we draw as the dotted line going through in Q^* Figure 3.2. One way of understanding the locus is to consider a point Q_0 at which the welfare of $\mathcal{G}_m$ is U_0 and

$$\left.\frac{dP}{da}\right|_{U^m} < \left.\frac{dP}{da}\right|_{U^{m+1}}.$$

Moving north-eastwards along $\mathcal{G}_m$'s indifference curve defined by ensures that $\mathcal{G}_m$ remains indifferent. So long as the indifference curves of $\mathcal{G}_{m+1}$ are steeper than the indifference curve of $\mathcal{G}_m$ they are being crossed *in the direction of its preference* and $\mathcal{G}_{m+1}$'s welfare is being increased, i.e. $\bar{U}^{m+1} < U_1^{m+1} < U_2^{m+1}$. At Q^* the indifference curves are tangent and the mutual gains are exhausted. At that point (3.65) holds.

The efficiency locus is drawn with a negative *finite* slope. That it must be so is easily shown by implicitly differentiating (3.65). The logic behind the result is the following. The efficiency locus is determined by the amount of abatement that $\mathcal{G}_{m+1}$

[5]Note that $\mathcal{G}_{m-1}$ is being kept indifferent throughout by (3.48).

[6]The dependence on P^m is contained in a^m and the condition that $P^m = B^m(a^m)A^m$.

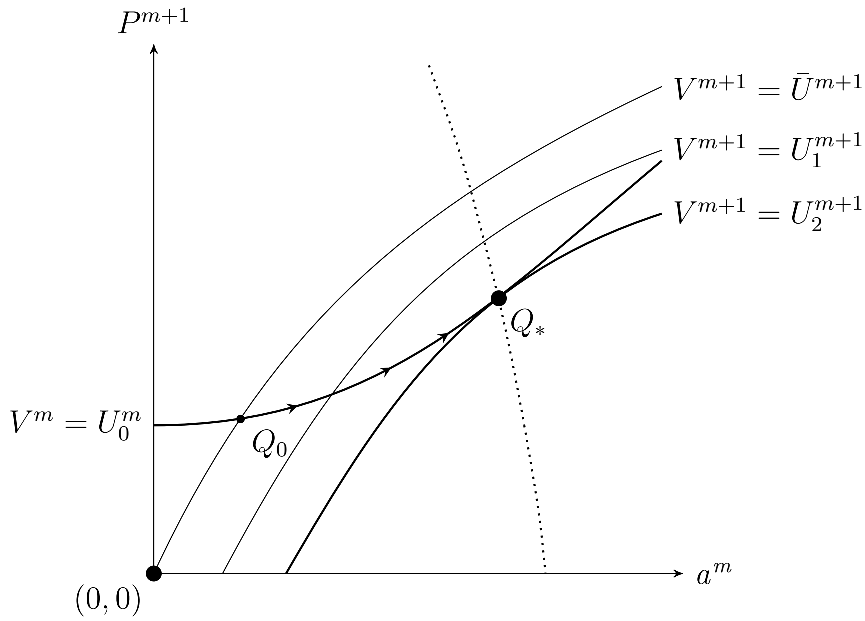


Figure 3.2: Efficiency frontier

can compensate $\mathcal{G}_m$ for, given that the compensation itself has an additional negative effect through the reduction of the $\mathcal{G}_m$'s incentive to save and the resulting capital accumulated at date $m+1$. This additional cost of the compensation mechanism is proportional to the interest rate at $m+1$, since the interest rate determines the marginal value of capital. Since the equilibrium interest rate depends positively on the pension P^{m+1} (see equation (3.29)), compensation for an additional unit of abatement is greater at a higher pension level, and therefore the efficiency frontier has a lower abatement level at higher pension levels. The intuition is illustrated graphically in Figure 3.3. Consider a point $Q_\circ$ directly under Q_*. Both points have the same abatement level, but $P_*^{m+1} > P_\circ^{m+1}$. Therefore $r_*^{m+1} > r_\circ^{m+1}$ (see equation (3.29)). All the other terms in (3.65) are independent of P^{m+1} and therefore the left hand side of condition (3.65) must be less than zero at $Q_\circ$; the lower pension reduces the equilibrium interest rate which reduces the compensation cost to $\mathcal{G}_{m+1}$. More abatement can be compensated for, so the intersection of the frontier with the indifference curve of $\mathcal{G}_m$ going through $Q_\circ$ is to the northeast of $Q_\circ$. The above is summarised in

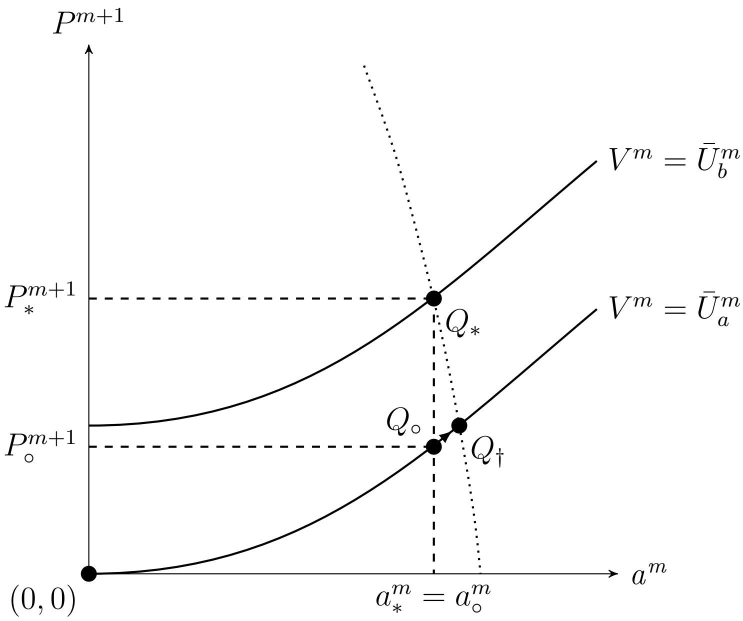


Figure 3.3: Higher cost at higher pension

Proposition 6.

Proposition 6. *Suppose that the gross production function is logarithmic and the each generation has Leontief utility. Then the efficiency frontier between* $\mathcal{G}_{m-1}, \mathcal{G}_m$ *and* $\mathcal{G}_{m+1}$ *is given by the monotonically decreasing locus in* (a^m, P^{m+1}) *space defined by* (3.65)*.*

Remark. The left hand side of (3.65) embodies the cost to $\mathcal{G}_{m+1}$ of a unit of abatement conditional on paying $\mathcal{G}_m$ a pension that keeps it indifferent, accounting for the direct benefit and cost of abatement and the pension, as well as the indirect cost both abatement and pensions induce through the negative effect on the equilibrium capital stock at date $m+1$. When it is negative, it is not a cost, but a net benefit to $\mathcal{G}_{t+1}$.

Figure 3.4 illustrates the results in Propositions 5 and 6. The grey shaded area consists of all the policy pairs that constitute Pareto improvements over the business-as-usual policy and the dotted line represents the efficiency frontier. The policies denoted by $Q^\dagger$ and $Q^\ddagger$ represent those policies on the frontier which make $\mathcal{G}_m$ and

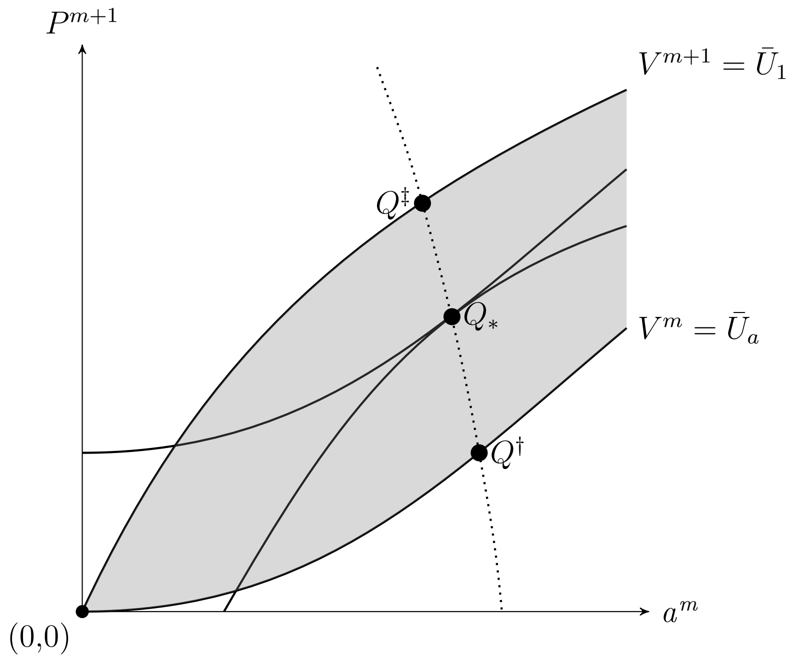


Figure 3.4: Pareto improvements and the frontier

$\mathcal{G}_{m+1}$ respectively indifferent to the business-as-usual. The policy Q_* represents an outcome at which the surplus is shared between those two generations.

3.7 Even more abatement at date m

Throughout this section, when not otherwise stated we will be assuming that any Pareto improvements that are possible are those achievable by modifying *only* the abatement at date m and the pension payouts in subsequent periods. *The abatement levels at all future dates $t > m$ are assumed fixed at exogenous levels.*

This section generalises the results of Section 3.6 to an arbitrary number of generations, $N \in \mathbb{N}$. In Subsection 3.7.1 we derive the conditions that define the locus policies that are efficient between the set of generations $\{\mathcal{G}_{m+i} : i = 0, 1, \ldots, N\}$. In Subsection 3.7.2 we use these conditions to establish that policy vectors that are efficient between $\{\mathcal{G}_{m+i} : i = 0, 1, \ldots, N\}$ can be (Pareto) improved upon by the inclusion of $\mathcal{G}_{m+N+1}$ and a change in policy involving a greater level of abatement at date m and a positive pension P^{m+N+1}.

3.7.1 The N-pension frontier

We now consider the welfare of $\mathcal{G}_{m+2}$ in addition to that of $\mathcal{G}_{m-1}, \mathcal{G}_m$ and $\mathcal{G}_{m+1}$. Condition (3.65) defines policies at which mutual gains are exhausted from changes in policies a^m, P^m and P^{m+1}. We now allow for a further pension P^{m+2} to be paid to $\mathcal{G}_{m+1}$ by $\mathcal{G}_{m+2}$. Equation (3.50) still describes the change in $\mathcal{G}_m$'s value from the policy changes considered. By (3.38) the change in $\mathcal{G}_{m+1}$'s value differs from (3.54) by the inclusion of $\mathrm{d}P^{m+2}$,

$$\mathrm{d}V^{m+1} = D_S^{m+1}(1-d^1)F^m \mathrm{d}a^m - \mathrm{d}P^{m+1} + r^{m+1}\mathrm{d}K^{m+1} + \mathrm{d}P^{m+2}. \tag{3.66}$$

The change in the value of $\mathcal{G}_{m+2}$ is essentially analogous to that of $\mathcal{G}_{m+1}$ with the difference that there are two mechanisms whereby the carbon stock at $m+2$ is lower. From (3.3) and (3.5) it is easy to see that the direct effect on S^{m+2} from abatement in at date m is $(1-d^2)F^m$. There is an additional effect on S^{m+2} given by the fact that the capital stock K^{m+1} is changed, which (by (3.3)) leads to a change in E^{m+1} given by

$$E_K^{m+1}\mathrm{d}K^{m+1} := (\bar{e}^{m+1} - a^{m+1})F_K^{m+1}\mathrm{d}K^{m+1}, \tag{3.67}$$

where we denote by E_K^t the increase in period t emissions due to a unit increase in the stock of productive capital at date t. Thus, the total effect on the welfare of $\mathcal{G}_{m+2}$ is given by

$$\mathrm{d}V^{m+2} = D_S^{m+2}[(1-d^2)F^m \mathrm{d}a^m - E_K^{m+1}\mathrm{d}K^{m+1}] - \mathrm{d}P^{m+2} + r^{m+2}\mathrm{d}K^{m+2}. \tag{3.68}$$

By (3.50), $\mathcal{G}_m$ is made indifferent if the policies are such that

$$\mathrm{d}P^{m+1} = B_a^m F^m \mathrm{d}a^m \tag{3.69}$$

Notice that, by (3.55) this yields $\mathrm{d}K^{m+1} = -\mathrm{d}a^m B_a^m F^m = -\mathrm{d}P^{m+1}$. Conditional on $\mathcal{G}_m$ being kept indifferent then, by (3.66), $\mathcal{G}_{m+1}$ is made indifferent if

$$\mathrm{d}P^{m+2} = [(1 + r^{m+1})B_a^m F^m - D_S^{m+1}(1 - d^1)F^m]\mathrm{d}a^m \tag{3.70}$$

So when (3.69) and (3.70) hold we have that $\mathcal{G}_m$ *and* $\mathcal{G}_{m+1}$ are indifferent. The key step in the generalisation of Proposition 6 to $\mathcal{G}_{m+2}$ and more generations is the realisation that in this case, by (3.38) and (3.41), for all t,

$$K^{t+1} = \frac{\beta}{1+\beta}V^t - P^{t+1} - \left[1 - B^{t+1}(a^{m+2})\right] A^{t+1}. \tag{3.71}$$

Since $\mathrm{d}a^{m+2} = 0$ by assumption and $\mathrm{d}V^{m+1} = 0$ by construction (see (3.70)) we get that

$$\mathrm{d}K^{m+2} = -\mathrm{d}P^{m+2} \tag{3.72}$$

Define

$$H_1^m = (1 + r^{m+1})B_a^m F^m - D_S^{m+1}(1 - d^1)F^m$$

and

$$H_2^m = (1 + r^{m+2})H_1^m - D_S^{m+2}\left[(1 - d^2)F^m + (1 - d^1)E_K^{m+1} B_a^m F^m\right]$$

Then by (3.69), (3.70), (3.72) and (3.68) the policy changes that keep $\mathcal{G}_m$ and $\mathcal{G}_{m+1}$ indifferent increase the welfare of $\mathcal{G}_{m+2}$ if

$$H_2^m \mathrm{d}a^m \leq 0 \tag{3.73}$$

Notice that at the frontier between $\mathcal{G}_m$ and $\mathcal{G}_{m+1}$, $H_1^m = 0$ (see equation (3.65)). When $H_1^m > 0$ the marginal cost to $\mathcal{G}_{m+1}$ to compensating $\mathcal{G}_m$ for a unit of abatement is greater than the marginal benefit it experiences via the reduced damages. Thus, when $H_1^m > 0$ too much abatement has taken place from the point of view of just $\mathcal{G}_m$ and $\mathcal{G}_{m+1}$ and the policies must be to the East of the frontier drawn in Figure 3.4. For those two generations to be kept indifferent to an additional unit of abatement when policies are past the frontier, $\mathcal{G}_{m+1}$ requires a pension transfer $\mathrm{d}P^{m+2} = H_1^m \mathrm{d}a^m$ (see equation (3.70)). The net cost of such a policy to $\mathcal{G}_{m+2}$, accounting for the negative effect of the capital stock is $(1 + r^{m+2})H_1^m \mathrm{d}a^m$ (see equation (3.72)). The benefit to $\mathcal{G}_{m+2}$ stems from the lower damages resulting from the reduced carbon stock:

$$D_S^{m+2}\left[(1-d^2)F^m + (1-d^1)E_K^{m+1}B_a^m F^m\right]$$

When the cost equals the benefit (to $\mathcal{G}_{m+2}$) efficiency is achieved, i.e. at $H_2^m = 0$.[7]

For the generalisation to the inclusion of $\{\mathcal{G}_{m+i} : i = 0, 1, \dots, N\}$ for arbitrary $N \in \mathbb{N}$ we introduce the following notation. Denote by

$$\mathcal{P}^m = (a^t, P^{m+1}, P^{m+2}, \dots, P^{m+N}, 0, 0, \dots) \in \mathbb{R}_+^{\mathbb{N}}$$

a vector of policies in question.[8]

[7]Notice that $H_1^m = 0$ (equation (3.65)) implicitly defined the Pareto frontier in $(a^m, P^{m+1}) - space$. Similarly, $H_2^m = 0$ implicitly defines the points in $(a^m, P^{m+1}, P^{m+2}) - space$ that are efficient between $\mathcal{G}_m, \mathcal{G}_{m+1}$ and $\mathcal{G}_{m+2}$. In addition to the dependence on a^m and P^{m+1} within H_1^m outlined in Figure 3.3, E_K^{m+1} depends on a^m and P^{m+1} and r^{m+1} depends on a^m, P^{m+1} and P^{m+2}.

[8]Recall that the abatement levels in all future periods are assumed fixed at some non-negative level.

Lemma 2. *Define the functions $H_i^m(\mathcal{P}^m)$ recursively by*

$$\begin{aligned}
H_0^m(\mathcal{P}^m) &= B_a(a^m)F^m \\
H_1^m(\mathcal{P}^m) &= (1+r^{m+1})H_0^m F^m - D_S^{m+1}(1-d^1)F^m \\
H_i^m(\mathcal{P}^m) &= (1+r^{m+i})H_{i-1}^m(\mathcal{P}^m) - D_S^{m+i}\left[(1-d^i)F^m + \sum_{j=1}^{i-1}(1-d^j)E_K^{m+i-j}H_{i-1-j}^m\right]
\end{aligned} \tag{3.74}$$

Suppose the gross production function is logarithmic and each generation has Leontief preferences. Then

$$H_N^m(\mathcal{P}^m) = 0 \tag{3.75}$$

is a necessary condition for the $\mathcal{P}^m$ to be on the efficiency frontier between the generations $\mathcal{G}_m, \mathcal{G}_{m+1}, \ldots \mathcal{G}_{m+N}$

The formal proof of Lemma 2 is in the Appendix.

3.7.2 Pareto improvements

Denote by $V^t(\mathcal{P}^m)$ the welfare of generation $\mathcal{G}_t$ under the policy $\mathcal{P}^m$. Define the binary relation $\gg$ on Euclidian space by

$$a \gg b \iff a_i > b_i \quad \forall i = 1, 2, \ldots, |a| \tag{3.76}$$

With this we can state our main result.

Theorem 5. *Suppose that the gross production function is logarithmic and each generation has Leontief preferences. Then, if $\mathcal{P}_N^m = (a_N^m, P_N^{m+1}, \ldots, P_N^{m+N}, 0)$ is on the* *efficiency frontier between* $\{\mathcal{G}_{m+i}, i = 0, 1, \ldots, N\}$*, there is a policy*

$$\mathcal{P}^m_{N+1} = (a^m_{N+1}, P^{m+1}_{N+1}, \ldots, P^{m+N}_{N+1}, P^{m+N+1}_{N+1})$$

such that

$$\mathcal{P}^m_{N+1} \gg \mathcal{P}^m_N \tag{3.77}$$

and

$$V^t(\mathcal{P}^m_{N+1}) \geq V^t(\mathcal{P}^m_N),\ \forall t \geq m \tag{3.78}$$

with strict inequality for at least one $t \in \{m, m+1, m+2, \ldots, m+N+1\}$*. Furthermore, under* $\mathcal{P}^m_{N+1}$ *the capital stock* K^{m+N+2} *is greater than and the carbon stock* S^{m+N+2} *is lower than under* $\mathcal{P}^m_N$*, thus endowing* $\mathcal{G}_{m+N+2}$ *with greater economic possibilities.*

The theorem states that any outcome that is on the efficiency frontier between $\{\mathcal{G}_{m+i} : i = 0, 1, \ldots, N\}$ can be improved upon by the inclusion of $\mathcal{G}_{m+N+1}$.

Proof. By Lemma 2 any efficient point $\mathcal{P}^m_N$ must satisfy $H^m_N(\mathcal{P}^m_N) = 0$. If that is the case, then $H^m_{N-1}(\mathcal{P}^m_N) > 0$ and in fact $H^m_i(\mathcal{P}^m_N) > 0$ for all $i = 0, 1, \ldots, N-1$. Consider the policy change

$$\mathrm{d}a^m > 0, \tag{3.79}$$

$$\mathrm{d}P^{m+i} = H^m_i(\mathcal{P}^m_N)\mathrm{d}a^m \quad \text{for } i = 0, 1, 2, \ldots, N-1 \tag{3.80}$$

Such a sequence of policies changes will make $\{\mathcal{G}_{m+j} : j = 0, 1, \ldots, i-1\}$ indifferent. Since $H^m_N(\mathcal{P}^m_N) = 0$ a pension $\mathrm{d}P^{m+N+1}$ such that

$$0 < \mathrm{d}P^{m+N+1}$$

will will lead to a strict improvement in the welfare of $\mathcal{G}_{m+N}$ (given (3.79) and (3.80)).

Furthermore

$$H_{N+1}^{m}(\mathcal{P}_{N}^{m}) = (1+r^{m+N+1})H_{N}^{m}(\mathcal{P}_{i}^{m}) \tag{3.81}$$

$$- D_{S}^{m+N+1}\left[(1-d^{N+1})F^{m} + \sum_{j=1}^{N}(1-d^{j})E_{K}^{m+N+1-j}H_{N-j}^{m}\right] \tag{3.82}$$

$$= - D_{S}^{m+N+1}\left[(1-d^{N+1})F^{m} + \sum_{j=1}^{N}(1-d^{j})E_{K}^{m+N+1-j}H_{N-j}^{m}\right] < 0 \tag{3.83}$$

So there is a strictly positive benefit to $\mathcal{G}_{m+N+1}$ to the policies defined by (3.79) and (3.80). Thus, $\mathrm{d}P^{m+N+1}$ such that

$$\mathrm{d}P^{m+N+1} < D_{S}^{m+N+1}\left[(1-d^{N+1})F^{m} + \sum_{j=1}^{N}(1-d^{j})E_{K}^{m+N+1-j}H_{N-j}^{m}\right]\mathrm{d}a^{m} \tag{3.84}$$

will result in a strict improvement in the welfare of $\mathcal{G}_{m+N+1}$. Thus, defining

$$\mathrm{d}\mathcal{P} = (\mathrm{d}a^{m}, \mathrm{d}P^{m+1}, \ldots, \mathrm{d}P^{m+N+1}) \tag{3.85}$$

for some $\mathrm{d}P^{m+N+1} > 0$ that satisfies (3.84) and

$$\mathcal{P}_{N+1}^{m} = \mathcal{P}_{N}^{m} + \mathrm{d}\mathcal{P} \tag{3.86}$$

we have that $\mathcal{P}_{N+1}^{m} \gg \mathcal{P}_{N}^{m}$ and $V^{t}(\mathcal{P}_{N+1}^{m}) = V^{t}(\mathcal{P}_{N}^{m})$ for $t = m, m+1, \ldots, m+N-1$ and $V^{t}(\mathcal{P}_{N+1}^{m}) > V^{t}(\mathcal{P}_{N}^{m})$ for $t = m+N, m+N+1$.

To see that the capital stock is greater under $\mathcal{P}_{N+1}^{m}$ than under $\mathcal{P}_{N}^{m}$, notice that, by (3.64), it must be the case that

$$\mathrm{d}K^{m+N+2} = \frac{\beta}{1+\beta}\mathrm{d}V^{m+N+1} - \mathrm{d}P^{m+N+2} \tag{3.87}$$

So if the change in $\mathcal{G}_{m+N+1}$'s welfare is positive and there is no change in the pension level it receives, the change in the capital it accumulates in equilibrium must also be positive. That S^{m+N+2} decreases is as simple consequence of (3.3) and (3.5) and the fact that $\mathrm{d}a^m > 0$ and $\mathrm{d}K^{m+i} < 0$ for $i = 1, 2, \ldots, m + N + 1$. The latter is a consequence of (3.87) and the fact that $\mathrm{d}P^{m+i} > 0$ for $i = 1, 2, \ldots, m + N + 1$. This establishes the claim. □

3.8 The general result

The assumptions of logarithmic production and Leontief utility allow for a very simple, albeit highly stylised solution to the inter-temporal choice problem. They are relaxed in this section. In what follows we will assume that the utility function to be a general monotonically increasing and twice differentiable function of consumption: $U(C_{1t}, C_{2t+1})$. As mentioned in Subsection 3.2.4, the solution to the inter-temporal optimisation problem (3.13) yields a savings function

$$(M^t, Z^{t+1}, r^{t+1}) \mapsto s(M^t, Z^{t+1}, r^{t+1}) = K^{t+1} \tag{3.88}$$

We make no assumptions on the gross production function other than that is has constant returns to scale. The wage and interest rates are then give by (3.11) and (3.12), and we will write

$$w_K^t = (1 - B^t)F_{LK}^t \quad \text{and} \quad r_K^t = (1 - B^t)F_{KK}^t$$

for their derivatives with respect to the capital stock.

3.8.1 Policies and frontier

We can now state the generalisation of Lemma 2.

Lemma 3. *Suppose the gross production function has constant returns to scale in capital and labour and the each generation has a twice differentiable utility function resulting in a savings function* (3.88) *as the solution to each generations inter-temporal consumption problem. Define the functions* $G_i^m(\mathcal{P}_t)$ *recursively by*

$$G_0^m = (1 + r^{m+1})T^m B_a^m(a^m)F^m \tag{3.89}$$

$$G_1^m = (1 + r^{m+2})T^{m+1}\left[R^m G_0^m - D_S^{m+1}(S^{m+1})(1 - d^1)F^m\right] \tag{3.90}$$

$$G_i^m = (1 + r^{m+i+1})T^{m+i}\left[R^{m+i-1}G_{i-1}^m - D_S^{m+i}\frac{\partial S^{m+i}}{\partial a^m}\right] \tag{3.91}$$

with

$$J^t = \left(1 + \frac{r_K^{t+1}K^{t+1}s_M^t}{(1 + r^{t+1})(1 - r_K^{t+1}s_r^t)}\right) \tag{3.92}$$

$$N^t = \left(1 + \frac{r_K^{t+1}K^{t+1}s_Z^t}{1 - r_K^{t+1}s_r^t}\right) \tag{3.93}$$

$$Q^t = \frac{(s_M^{t-1} - s_Z^t)}{(1 + r^t)(1 - r_K^t s_r^t) + r_K^t K^t s_M^{t-1}} \tag{3.94}$$

$$R^t = 1 + w_K^{t+1}Q^t \tag{3.95}$$

$$T^t = \frac{J^t}{N^t} \tag{3.96}$$

$$\frac{\partial S^t}{\partial a^m} = (1 - d^{t-m})F^m + \sum_{j=1}^{t-m-1}(1 - d^j)E_K^{t-j}G_{t-m-j-1}^m Q^{t-j-1} \tag{3.97}$$

Then, provided the policies at date m *are unanticipated by* $\mathcal{G}_{m-1}$*,*

$$G_N^m(\mathcal{P}^m) = 0 \tag{3.98}$$

is a necessary condition for efficiency between $\mathcal{G}_m, \mathcal{G}_{m+1}, \ldots \mathcal{G}_{m+N}$

Note that apart from the factors (3.92) to (3.96), the functions G_i^m in Lemma 3 are identical to the functions H_i^m in Lemma 2. In particular, the qualitative dependence of G_{i+1}^m on G_i^m is the same as that of H_{i+1}^m on H_i^m. Since this is the only feature required for the proof of Theorem 5, we can state its generalisation without further proof.

Theorem 6. *Suppose the gross production function has constant returns to scale in capital and labour and the each generation has a twice differentiable utility function resulting in a savings function* (3.88) *as the solution to each generations intertemporal consumption problem. Suppose further that the policies implemented at date m were unanticipated by $\mathcal{G}_{m-1}$ at date $m-1$. Then, if $\mathcal{P}_N^m = (a_N^m, P_N^{m+1}, \ldots, P_N^{m+N}, 0)$ is on the efficiency frontier between $\{\mathcal{G}_{m+i}, i = 0, 1, \ldots, N\}$, there is a policy*

$$\mathcal{P}_{N+1}^m = (a_{N+1}^m, P_{N+1}^{m+1}, \ldots, P_{N+1}^{m+N}, P_{N+1}^{m+N+1})$$

such that

$$\mathcal{P}_{N+1}^m \gg \mathcal{P}_N^m \tag{3.99}$$

and

$$V^t(\mathcal{P}_{N+1}^m) \geq V^t(\mathcal{P}_N^m),\ \forall t \geq m \tag{3.100}$$

with strict inequality for at least one $t \in \{m, m+1, m+2, \ldots, m+N+1\}$. Furthermore, under $\mathcal{P}_{N+1}^m$ the capital stock K^{m+N+2} is greater than and the carbon stock S^{m+N+2} is lower than under $\mathcal{P}_N^m$.

3.8.2 Qualitative and quantitative difference

Notice that the relative magnitude of the points on the Pareto frontiers defined by (3.75) and (3.98) is somewhat determined by the product of the coefficients $R^t \cdot T^t$. If $R^m \cdot T^m = 1$ the frontiers defined by (3.75) and (3.98) for $N = 1$ are *identical.* If that were the case, the different simplifications assumed in Section 3.4 would cancel out exactly.

For $T^m R^m \leq 1$ it is the case that the frontier defined by (3.75) has uniformly greater abatement levels than (3.98). This is essentially because it is less costly to $\mathcal{G}_{m+1}$ to compensate $\mathcal{G}_m$ for any unit of abatement. The converse is also true, i.e. if $T^m R^m \geq 1$ the frontier defined by (3.75) has uniformly lower abatement levels than (3.98).

The frontiers for $N > 1$ are not as easy to characterise. This is because the frontier is partially determined by the emissions that are reduced by virtue of the reduction in capital stock at the intermediate dates $t = m+1, m+2, \ldots, m+N-1$ – the summation on the right hand side of (3.97). If $T^m R^m \leq 1$ and $Q^t \leq 1$ the effect on the capital stock is greater in equation (3.74) than in (3.91) so even though the cost of compensation is less in the general case than in the simplified case, the benefit is also slightly lower. However, if the coefficients E_K^t for $t = m+1, m+2, \ldots, m+N-1$ are sufficiently small, one can also conclude that the frontiers defined by (3.75) have a uniformly greater abatement level than the ones defined by (3.98). In this case, the simpler model (3.75) would underestimate the amount of desirable abatement (at any given level of pensions).

Testing the range of values that can be taken by the product $R^t \cdot T^t$ with Mathematica we can conclude that if the production and decision parameters are restricted to the domains specified below

- Capital share: $\alpha \in [0, 1]$
- Generational interest rate: $r^t \in (0, 4)$
- First period wealth effect: $s_M^{t-1} \in (0, 1)$
- Second period wealth effect: $s_Z^t \in (-1, 0)$
- Interest elasticity of savings: $s_r^t \frac{r^t}{K^t} =: \mathcal{E}_{s,r} \in (-1, 1)$

the product must be in the domain $T^t \cdot R^t \in (0, 1)$. For the following specific values deemed as reasonable by the authors

- $\alpha = 0.3$
- $r_t = 1$
- $s_M^{t-1} = -s_Z^t = 0.4$
- $\mathcal{E}_{s,r}^t = 0.01$

we have that

$$T^t R^t \approx 0.89$$

and

$$Q^t \approx 0.46$$

For those parameter values it would mean that the approximation introduced in Section 3.4 underestimates the efficient levels of abatement for the contracts involving only two generations. Whether or not the contracts involving more generations under or overestimate the abatement levels depends on the magnitude of the parameters $E_K^t = e^t F_K^t$, i.e. the product of the future emission intensities with the gross productivity of capital.

Either way, the approximation will presumably result in slightly conservative prescriptions of the same order of magnitude as the more complete model for any reasonable parameters for the production technology and savings propensities.

3.9 Conclusion

We have derived the conditions for a policy vector involving abatement in a given period and compensating intergenerational transfers in all subsequent periods to be efficient in the sense that any further benefits to future generations from current abatement are smaller than the cost of compensating the abaters for the cost they incur. We use this result to establish that Pareto improvements are possible if the benefits to any generation have not been included into the agreement.

The theoretical exposition is self-contained and complete. This being said, further work is necessary to exhaust the full potential of this research project. It would be desirable to include population growth into the analysis. The damages have been modeled as proportional to the population rather than the more usual form that makes them proportional to output. As discussed in Section 3.2.1 this feature makes our claim of the existence of Pareto improving abatement more difficult to establish, as damages on output would create an incentive for every generation to abate simply for the benefit of increasing their capital rent, which would be affected by damages if they are modeled as proportional to output. Thus, adding the dependance on output will not yield any new further intuition, but with a view of calibrating the model to one of the widespread integrated assessment models it would be useful to model the damages in the standard way. Such a calibration would allow us to establish the orders of magnitude of the pensions involved and help determine the political feasibility of such a policy.

From a theoretical point of view there is one extension in particular that may be worth pursuing. We have used used endowment transfers (pay-as-you-go pensions) as the compensation mechanism to achieve the Pareto improvements. These have a well-known theoretical property of disincentivising capital accumulation, a feature which determines the location of the efficiency frontier. It may be possible to achieve further gains if the compensation mechanism used is a subsidy on the capital returns, since these could easily achieve the same amount of compensation at possibly nil disincentive to the savings decision and therefore capital accumulation.

Finally, the strategic credibility of such intergenerational contracts must be analysed. There is an existing literature looking at this in OLG models. Most notably, Rangel (2003) looks at a link between social security and environmental services in a strategic setting. Our models differ significantly, but an analysis similar to his could yield interesting results in our framework as well.

3.10 Appendix: proofs and mathematical detail

Lemma 4. *In the domain* $\mathcal{D} = \{(L, K) : L > 0, K > L \cdot e\}$ *the isoquants of*

$$F(K, L) = L \ln\left(\frac{K}{L}\right)$$

are downward sloping and convex.

Proof. By the implicit function theorem the slopes of the isoquants are given by

$$\frac{\partial K}{\partial L} = -\frac{F_L}{F_K} = -\frac{K}{L}\left[\ln\left(\frac{K}{L}\right) - 1\right] \tag{3.101}$$

which is negative in $\mathcal{D}$. Notice that

$$L\frac{\partial K}{\partial L} = -K\left[\ln\left(\frac{K}{L}\right) - 1\right] \tag{3.102}$$

Differentiating (3.102) with respect to L you get the left hand side

$$\text{LHS} = \frac{\partial K}{\partial L} + L\frac{\partial^2 K}{\partial L^2}$$

and right hand side

$$\begin{aligned} \text{RHS} = &-\frac{\partial K}{\partial L}\left[\ln\left(\frac{K}{L}\right) - 1\right] - K\left[\frac{\partial K}{\partial L}\frac{1}{K} - \frac{1}{L}\right] \\ = &-\frac{\partial K}{\partial L}\left[-\frac{\partial K}{\partial L}\frac{L}{K}\right] - \frac{\partial K}{\partial L} + \frac{K}{L} \end{aligned}$$

Equating LHS and RHS yields

$$
\begin{aligned}
L\frac{\partial^2 K}{\partial L^2} &= \frac{L}{K}\left[\left(\frac{\partial K}{\partial L}\right)^2 - 2\frac{K}{L}\frac{\partial K}{\partial L} + \left(\frac{K}{L}\right)^2\right] \\
&= \frac{L}{K}\left[\frac{\partial K}{\partial L} - \frac{K}{L}\right]^2 \geq 0
\end{aligned}
$$

Thus the isoquants are convex, which completes the proof. □

Lemma 5. *The indifference curves (in $(a^m, P^{m+1}) - space$) of $\mathcal{G}_m$ are convex and within the domain for which they are increasing those of $\mathcal{G}_{m+1}$ are concave.*

Proof. The slope of the indifference curves of $\mathcal{G}_m$ is given by

$$
\left.\frac{dP}{da}\right|_{U^m} = B_a^m(a^m)F^m \tag{3.103}
$$

The derivative of (3.103) with respect to a^m is

$$
\left.\frac{d^2P}{da^2}\right|_{U^m} = B_{aa}^m F^m > 0
$$

The slope of the indifference curves of $\mathcal{G}_{m+1}$ is given by

$$
\left.\frac{dP}{da}\right|_{U^{m+1}} = \frac{(1-d^1)F^m D_S^{m+1} - \frac{\beta r^{m+1} B_a^m F^m}{1+\beta}}{1 + \frac{r^{m+1}}{1+\beta}} \tag{3.104}
$$

The derivative of (3.104) with respect to a^m is

$$\left.\frac{d^2P}{da^2}\right|_{U^{m+1}} = \frac{-(1-d^1)(F^m)^2 D_{SS}^{m+1} - \frac{\beta r^{m+1}}{1+\beta} B_{aa}^m F^m}{1+\frac{r^{m+1}}{1+\beta}} \tag{3.105}$$

$$- \frac{(1-d^1)F^m D_S^{m+1} - \frac{\beta r^{m+1} B_a^m F^m}{1+\beta}}{1+\frac{r^{m+1}}{1+\beta}} \frac{1+\frac{r_K^{m+1}}{1+\beta}\left[K_{a^m}^{m+1} + K_{P^{m+1}}^{m+1} \left.\frac{dP}{da}\right|_{U^{m+1}}\right]}{1+\frac{r^{m+1}}{1+\beta}} \tag{3.106}$$

$$= \frac{-(1-d^1)(F^m)^2 D_{SS}^{m+1} - \frac{\beta r^{m+1}}{1+\beta} B_{aa}^m F^m}{1+\frac{r^{m+1}}{1+\beta}} \tag{3.107}$$

$$- \left.\frac{dP}{da}\right|_{U^{m+1}} \frac{1+\frac{r_K^{m+1}}{1+\beta}\left[K_{a^m}^{m+1} + K_{P^{m+1}}^{m+1} \left.\frac{dP}{da}\right|_{U^{m+1}}\right]}{1+\frac{r^{m+1}}{1+\beta}} \tag{3.108}$$

Since $r_K^{m+1} < 0$ and $K_{a^m}^{m+1} > 0$, $K_{P^{m+1}}^{m+1} > 0$

$$\left.\frac{dP}{da}\right|_{U^{m+1}} \geq 0$$

is sufficient for

$$\left.\frac{d^2P}{da^2}\right|_{U^{m+1}} < 0$$

□

Lemma 6. *Under the assumptions of Lemma 4 the effect on the welfare of generation $\mathcal{G}_t$ from small changes in past policies up to and including a^{t-1} and P^{t+1} is given by*

$$\mathrm{d}V^t = D_S^t \sum_{i=1}^{t-m} (1-d^i) F^{t-i} \mathrm{d}a^{t-i} + \mathrm{d}P^{t+1} + r^t \frac{\beta}{1+\beta} \mathrm{d}V^{t-1} - (1+r^t)\mathrm{d}P^t \tag{3.109}$$
$$- D_S^t \sum_{i=1}^{t-m-1} (1-d^i) E_K^{t-i} \left(\frac{\beta}{1+\beta} \mathrm{d}V^{t-i-1} - \mathrm{d}P^{t-i} + A^{t-i} B_a^{t-i} \mathrm{d}a^{t-i} \right)$$

where period m is the first in which there is a change in the abatement policy.

Proof. The value V^t is a function of $K^t, S^t, a^t, a^{t+1}, P^t$ and P^{t+1} so

$$\mathrm{d}V^t = \frac{\partial V^t}{\partial K^t}\mathrm{d}K^t + \frac{\partial V^t}{\partial S^t}\mathrm{d}S^t + \frac{\partial V^t}{\partial a^t}\mathrm{d}a^t + \frac{\partial V^t}{\partial a^{t+1}}\mathrm{d}a^{t+1} + \frac{\partial V^t}{\partial P^t}\mathrm{d}P^t + \frac{\partial V^t}{\partial P^{t+1}}\mathrm{d}P^{t+1} \quad (3.110)$$

We are assuming $\mathrm{d}a^t = \mathrm{d}a^{t+1} = 0$

$$\frac{\partial V^t}{\partial P^t} = -1 \quad \text{and} \quad \frac{\partial V^t}{\partial P^{t+1}} = 1 \quad (3.111)$$

$$\frac{\partial V^t}{\partial S^t} = -D_S^t \quad \text{and} \quad \frac{\partial V^t}{\partial K^t} = r^t \quad (3.112)$$

It remains to be shown that

$$\mathrm{d}S^t = -\sum_{i=1}^{t-m}(1-d^i)F^{t-i}\mathrm{d}a^{t-i} + \sum_{i=1}^{t-m-1}(1-d^i)E_K^{t-i}\mathrm{d}K^{t-i} \quad (3.113)$$

and

$$\mathrm{d}K^t = \frac{\beta}{1+\beta}\mathrm{d}V^{t-1} - \mathrm{d}P^t + A^t B_a^t \mathrm{d}a^t \quad (3.114)$$

(3.113) is a direct consequence of (3.5). To see that (3.114) holds note that by combining (3.31) with (3.38) you get that

$$K^t = \frac{\beta}{1+\beta}V^{t-1} - P^t - A^t(1 - B^t(a^t)) \quad (3.115)$$

The conclusion follows directly. □

Proof of Lemma 2. Consider the vector of policy changes

$$\mathrm{d}\mathcal{P}^m = (\mathrm{d}a^m, \mathrm{d}P^{m+1}, \ldots, \mathrm{d}P^{m+N})$$

Since only abatement in period m is being considered, the hypothesis of Lemma 6 holds for $t > m$, and by (3.109)

$$\begin{aligned}\mathrm{d}V^t = D_S^t(1-d^{t-m})F^m\mathrm{d}a^m + \mathrm{d}P_{t+1} + r_t\frac{\beta}{1+\beta}\mathrm{d}V^{t-1} - (1+r_t)\mathrm{d}P_t \\ - D_S^t\sum_{i=1}^{t-m-1}(1-d^i)E_K^{t-i}\left(\frac{\beta}{1+\beta}\mathrm{d}V^{t-i-1} - \mathrm{d}P^{t-i}\right)\end{aligned}$$

Consider policies that leave $\{\mathcal{G}_i; i = m, m+1, \ldots, t-1\}$ indifferent. That is $\mathrm{d}V^i = 0$ for all $i = m, m+1, \ldots, t-1$. The welfare of $\mathcal{G}_t$ then changes by

$$\mathrm{d}V^t = D_S^t(1-d^{t-m})F^m\mathrm{d}a^m + \mathrm{d}P_{t+1} - (1+r_t)\mathrm{d}P_t + D_S^t\sum_{i=1}^{t-m-1}(1-d^i)E_K^{t-i}\mathrm{d}P^{t-i}$$

Such a policy will leave $\mathcal{G}_t$ indifferent if

$$\mathrm{d}P^{t+1} = (1+r^t)\mathrm{d}P^t + D_S^t(1-d^{t-m})F^m\mathrm{d}a^m + D_S^t\sum_{i=1}^{t-m-1}(1-d^i)E_K^{t-i}\mathrm{d}P^{t-i}$$

Therefore, when

$$\frac{\mathrm{d}P^{t+1}}{\mathrm{d}a^m} = (1+r^t)\frac{\mathrm{d}P^t}{\mathrm{d}a^m} - D_S^t(1-d^{t-m})F^m - D_S^t\sum_{i=1}^{t-m-1}(1-d^i)E_K^{t-i}\frac{\mathrm{d}P^{t-i}}{\mathrm{d}a^m} \tag{3.116}$$

it is the case that $\{\mathcal{G}_i; i = m, m+1, \ldots, t\}$ are all indifferent to the policy $\mathrm{d}\mathcal{P}$. Since $\mathcal{G}_{m+N+1}$ is *not* included, $\mathrm{d}P^{m+N+1} = 0$ by hypothesis. Thus, for $\mathcal{G}_N$ to be indifferent to the policy

$$(1+r^t)\frac{\mathrm{d}P^t}{\mathrm{d}a^m} - D_S^t(1-d^{t-m})F^m - D_S^t\sum_{i=1}^{t-m-1}(1-d^i)E_K^{t-i}\frac{\mathrm{d}P^{t-i}}{\mathrm{d}a^m} \stackrel{!}{=} 0$$

Thus, defining

$$H_i^m := \frac{\mathrm{d}P^{m+i+1}}{\mathrm{d}a^m}$$

and replacing for H_i^m in (3.116) yields the definitions in the hypothesis and the condition of the conclusion. □

Proof of Theorem 3. In this more general setting, the value function and equilibrium capital are defined by equations (3.18) to (3.22). Thus, the general form of the derivatives of the value function and the capital accumulation equation – equations (3.39) to (3.47) – are

$$V^m_{a^m} = -B^m_a F^m_L u'(C_{1m}) + K^{m+1} r^{m+1}_K K^{m+1}_{a^m} \beta u'(C_{2m+1}) \tag{3.117}$$

$$V^m_{P^m} = -u'(C_{1m}) + K^{m+1} r^{m+1}_K K^{m+1}_{P^m} \beta u'(C_{2m+1}) \tag{3.118}$$

$$V^{m-1}_{a^m} = -B^m_a F^m_K K^m \beta u'(C_{2m}) + K^m r^m_K K^m_{a^m} \beta u'(C_{2m}) \tag{3.119}$$

$$V^{m-1}_{P^m} = \beta u'(C_{2m}) + K^m r^m_K K^m_{P^m} \beta u'(C_{2m}). \tag{3.120}$$

The derivatives (3.135) and (3.136) are correct provided the respective change in policy was anticipated and thus allowed for an adjustment in the savings rate. If the change is unanticipated we get

$$V^{m-1}_{a^m} = -B^m_a F^m_K K^m \beta u'(C_{2m}) \tag{3.121}$$

$$V^{m-1}_{P^m} = \beta u'(C_{2m}). \tag{3.122}$$

The dependance of the values on the state variables is given by

$$V^m_{K^m} = w^m_K \left[u'(C_{1m}) - K^{m+1} r^{m+1}_K K^{m+1}_{P^m} \beta u'(C_{2m+1})\right] \tag{3.123}$$

$$V^m_{S^m} = -D^m_S \left[u'(C_{1m}) - K^{m+1} r^{m+1}_K K^{m+1}_{P^m} \beta u'(C_{2m+1})\right] \tag{3.124}$$

The first order effects on future capital savings are given by

$$K^{m+1}_{a^m} = \frac{-B^m_a F^m_L s^m_M}{1 - r^{m+1}_K s^{m+1}_r} \tag{3.125}$$

$$K^{m+1}_{P^m} = \frac{-s^m_M}{1 - r^{m+1}_K s^{m+1}_r} \tag{3.126}$$

and the effect on the current capital stock, if the the policy change was anticipated, is given by

$$K^m_{a^m} = \frac{-B^m_a F^m_K s^m_r}{1 - r^m_K s^m_r} \tag{3.127}$$

$$K^m_{P^m} = \frac{s_Z}{1 - r^m_K s^m_r} \tag{3.128}$$

Due to the fact that the Euler equation must hold in equilibrium for every generation we have

$$u'(C_{1t}) = (1 + r^{t+1})\beta u'(C_{2t+1}). \tag{3.129}$$

We will divide the derivatives of the value function of $\mathcal{G}_t$ by $u'(C_{1t})$ in order to get the change in value in first period consumption units. Using the Euler equation (3.129), the coefficient $\beta u'(C_{2t+1}$ will get replaced by $(1 + r^{t+1})^{-1}$. Replacing (3.125) through (3.128) into (3.133) through (3.136), defining

$$J^m = 1 - (1 + r^{m+1})^{-1} K^{m+1} r^{m+1}_K K^{m+1}_{P^m} \tag{3.130}$$

$$N^m = 1 + K^{m+1} r^{m+1}_K K^{m+1}_{P^{m+1}} \tag{3.131}$$

$$\Lambda^m = 1 + \frac{r^m_K s^m_r}{1 - r^m_K s^m_r} \tag{3.132}$$

and with a minor abuse of notation we get that the derivatives of the value function in first period consumption units become

$$V^m_{a^m} = -B^m_a F^m_L J^m \tag{3.133}$$

$$V^m_{P^m} = -J^m \tag{3.134}$$

$$V^{m-1}_{a^m} = -\frac{B^m_a F^m_K K^m}{1 + r^m} \Lambda^m \tag{3.135}$$

$$V^{m-1}_{P^m} = \frac{N^{m-1}}{1 + r^m}, \tag{3.136}$$

$$V_{a^m}^{m-1} = -B_a^m F_K^m K^m (1 + r^m)^{-1} \tag{3.137}$$

$$V_{P^m}^{m-1} = (1 + r^m)^{-1} \tag{3.138}$$

and

$$V_{K^m}^m = w_K^m J^m \tag{3.139}$$

$$V_{S^m}^m = -D_S^m J^m \tag{3.140}$$

Consider, as in Proposition 2, the vector of policy changes

$$\mathrm{d}\mathcal{P}^m = (\mathrm{d}a^m, \mathrm{d}P^{m+1}, \ldots, \mathrm{d}P^{m+N})$$

with the additional assumption that $\mathrm{d}P^{m+N+1} = 0$. The welfare change of $\mathcal{G}_{m-1}$ is given by

$$\mathrm{d}V^{m-1} = V_{a^m}^{m-1}\mathrm{d}a^m + V_{P^m}^{m-1}\mathrm{d}P^m \tag{3.141}$$

If the policy changes were unanticipated when $\mathcal{G}_{m-1}$'s savings decision was made, the condition on $\mathrm{d}a^m$ and $\mathrm{d}P^m$ that ensures $\mathcal{G}_{m-1}$ is indifferent to the business as usual is

$$\mathrm{d}P^m = B_a^m F_K^m K^m \mathrm{d}a^m \tag{3.142}$$

The condition for the changes to be welfare improving to $\mathcal{G}_m$,

$$\mathrm{d}V^m = V_{a^m}^m \mathrm{d}a^m + V_{P^m}^m \mathrm{d}P^m + V_{P^{m+1}}^m \mathrm{d}P^{m+1} \geq 0, \tag{3.143}$$

becomes

$$\mathrm{d}V^m = V_{a^m}^m \mathrm{d}a^m + V_{P^m}^m B_a^m F_K^m K^m \mathrm{d}a^m + V_{P^{m+1}}^m \mathrm{d}P^{m+1} \geq 0, \tag{3.144}$$

when $\mathrm{d}P^m$ is replaced with (3.142). At the time $\mathcal{G}_m$ makes it's savings decision the policy changes are assumed to be known and therefore $\mathrm{d}V^m \geq 0$ becomes

$$(1 + r^{m+1})^{-1} N^m \mathrm{d}P^{m+1} \geq B_a^m F^m J^m \mathrm{d}a^m \tag{3.145}$$

The condition for any future generation $\mathcal{G}_t$, for $t > m$ to remain (at least) indifferent is

$$\mathrm{d}V^t = V_{S^t}^t \mathrm{d}S^t + V_{K^t}^t \mathrm{d}K^t + V_{P^t}^t \mathrm{d}P^t + V_{P^{t+1}}^t \mathrm{d}P^{t+1} = 0 \tag{3.146}$$

Since abatement only happens during period m, for all $t > m$

$$\mathrm{d}V^{t-1} = \frac{s_M^{t-1} - s_Z^t}{1 + r^t s_M^{t-1}} \mathrm{d}P^t + \left[\frac{1 - r_K^t s_r^t}{s_M^{t-1}} + \frac{K^t r_K^t}{1 + r^t} \right] \mathrm{d}K^t \tag{3.147}$$

To see this notice that

$$V_{a^{t-1}}^{t-1} \mathrm{d}a^{t-1} - \left[\frac{1 - r_K^t s_r^t}{s_M^{t-1}} + \frac{K^t r_K^t}{1 + r^t} \right] K_{a^{t-1}}^t \mathrm{d}a^{t-1} = 0 \tag{3.148}$$

$$V_{P^{t-1}}^{t-1} \mathrm{d}P^{t-1} - \left[\frac{1 - r_K^t s_r^t}{s_M^{t-1}} + \frac{K^t r_K^t}{1 + r^t} \right] K_{P^{t-1}}^t \mathrm{d}P^{t-1} = 0 \tag{3.149}$$

and

$$V_{P^t}^{t-1} \mathrm{d}P^t - \left[\frac{1 - r_K^t s_r^t}{s_M^{t-1}} + \frac{K^t r_K^t}{1 + r^t} \right] K_{P^t}^t \mathrm{d}P^t = \left[\frac{1}{1 + r^t} - \frac{s_Z^t}{s_M^{t-1}} \right] \mathrm{d}P^t = \frac{s_M^{t-1} - s_Z^t}{(1 + r^t) s_M^{t-1}} \mathrm{d}P^t \tag{3.150}$$

Combining (3.148), (3.149) and (3.150) yields (3.147). Defining

$$Q^t = \frac{s_M^{t-1} - s_Z^t}{(1 + r^t)(1 - r_K^t s_r^t) + K^t r_K^t s_M^{t-1}} \tag{3.151}$$

we can rewrite (3.147) into

$$dK^t = Q^t \left[\frac{(1+r^t)s_M^{t-1}}{(s_M^{t-1} - s_Z^t)} dV^{t-1} - dP^t \right] \tag{3.152}$$

By (3.146) the condition that $dV^t \geq 0$ *conditional* on $dV^{t-1} = 0$ can be rewritten to

$$\frac{N^t}{1+r^{t+1}} dP^{t+1} = J^t \left[\left(1 + w_K^t Q^t\right) dP^t + D_S^t dS^t \right] \tag{3.153}$$

Finally, recall that, by (3.113)

$$dS^t = -(1-d^{t-m}) F^m da^m + \sum_{j=1}^{i-1} (1-d^j) E_K^{t-j} dK^{t-j} \tag{3.154}$$

Substituting (3.152) conditional on $dV^t = 0$ for $t - m = 0, 1, \ldots, i-1$ into (3.154) yields

$$dS^t = -(1-d^{t-m}) F^m da^m - \sum_{j=1}^{i-1} (1-d^j) E_K^{t-j} Q^{t-j} dP^{t-j} \tag{3.155}$$

Thus, when

$$\frac{N^t}{1+r^{t+1}} dP^{t+1} = J^t \left[\left(1 + w_K^t Q^t\right) dP^t - D_S^t \left((1-d^{t-m}) F^m da^m + \sum_{j=1}^{i-1} (1-d^j) E_K^{t-j} Q^{t-j} dP^{t-j} \right) \right] \tag{3.156}$$

holds for $t = m+1, \ldots, m+N-1$, the generations $\{\mathcal{G}_t, t = m, 1, 2, \ldots, m+N-1\}$ are indifferent to $d\mathcal{P}$. Since $dP^{m+N+1} = 0$, the condition for $\mathcal{G}_{m+N}$ to be indifferent is

$$\left(1 + w_K^t Q^t\right) dP^t - D_S^t \left((1-d^{t-m}) F^m da^m + \sum_{j=1}^{i-1} (1-d^j) E_K^{t-j} Q^{t-j} dP^{t-j} \right) = 0 \tag{3.157}$$

Defining G_i^m as in the hypothesis, conditions (3.156) and (3.157) yield the conclusion.

□

Bibliography

BOVENBERG, A., AND B. J. HEIJDRA (1998): "Environmental tax policy and intergenerational distribution," *Journal of Public Economics*, 67(1), 1–24.

DIAMOND, P. A. (1965): "National debt in a neoclassical growth model," *The American Economic Review*, 55(5), 1126–1150.

FOLEY, D. K. (2007): "The economic fundamentals of global warming," *October (New York, Department of Economics, New School for Social Research).*

GALE, D. (1973): "Pure exchange equilibrium of dynamic economic models," *Journal of Economic Theory*, 6(1), 12–36.

GERLAGH, R., AND M. A. KEYZER (2001): "Sustainability and the intergenerational distribution of natural resource entitlements," *Journal of Public Economics*, 79(2), 315–341.

GOLOSOV, M., J. HASSLER, P. KRUSELL, AND A. TSYVINSKI (2011): "Optimal taxes on fossil fuel in general equilibrium," Discussion paper, National Bureau of Economic Research.

HOWARTH, R. B., AND R. B. NORGAARD (1992): "Environmental Valuation under Sustainable Development," *The American Economic Review*, 82(2), pp. 473–477.

JOHN, A., AND R. PECCHENINO (1994): "An Overlapping Generations Model of Growth and the Environment," *The Economic Journal*, 104(427), pp. 1393–1410.

KARP, L., AND A. REZAI (2012): "The Political economy of environmental policy with overlapping generations," CUDARE Working Paper 1128.

RANGEL, A. (2003): "Forward and Backward Intergenerational Goods: Why Is Social Security Good for the Environment?," *American Economic Review*, 93(3), 813–834.

REZAI, A., D. K. FOLEY, AND L. TAYLOR (2012): "Global warming and economic externalities," *Economic Theory*, 49, 329–351.