Chapter 2
Foundations of the Theoretical Analysis

In this part of the study a theoretical model is developed to study the transition from an established polluting to a new clean energy technology. The model encompasses two distinctive features. First, the social rate of time preference is assumed to stay below the private. Second, the creation of new productive capital is supposed to exhibit a time-to-build property. This chapter introduces to the theoretical issue of this study and describes its relationship and contribution to the two respective strands of literature. Section 2.1 discusses the crucial assumption of the split of social and private time-preference rates as treated in the economic discounting debate thus far, states open issues and finally justifies the treatment of the assumption in the subsequent model analysis. Section 2.2 motivates the taking up of the modeling framework of time-lagged capital theory and shows how it is further developed in this study. Section 2.3 explains the analytical structure of the model developed in chapter 3.

2.1 Social versus Private Time Preferences

The assumption of the social rate of time preference staying below the private builds on a long-standing debate in economics on discounting in general and the correct determination and relationship of social and private rates of time preference in particular. While already in the 1960s and 1970s wide unanimity could be reached on the relationship of the crucial rates in the *first-best* case on the one hand, referring to the basic discounted utility model, as well as on the empirical validity of *positive* discounting in general on the other,[1] the 'dark jungles of the second best' (Baumol 1968) with respect to discounting have been remaining contentious. In more recent

[1] Early authors, such as, e.g., Fisher (1930), Pigou (1952) and Ramsey (1928), still being hesitant on ethical grounds, positive discounting is today recognised especially as an expression of revealed dynamic preferences and used as a standard assumption in economic models (Arrow and Kurz 1970, Caplin and Leahy 2004). Moreover, there are, of course, well recognised exceptions where the discount rate is zero or negative (e.g., Dasgupta 2001).

C. Heinzel, *Distorted Time Preferences and Structural Change in the Energy Industry*, Sustainability and Innovation, DOI 10.1007/978-3-7908-2183-3_2,

years the debate has even gained in complexity due to new empirical objections and the rise of problems with particular long-run effects, such as, e.g., anthropogenic climate change. A complementary debate in finance discusses why the rates of return on private securities are so high (equity premium puzzle) and those on state securities so low (risk-free rate puzzle).

This section introduces to the state of the economic discounting debate (subsection 2.1.1) and states three reasons for the social and private time-preference rates to diverge particularly relevant for investments in the energy sector (subsection 2.1.2). Subsection 2.1.3 summarises the findings from the literature survey and describes the treatment of the assumption in the model below as well as the contribution of the analysis.

2.1.1 The Economic Discounting Debate: First-best Benchmark, Second-best Cases, and Some More Recent Issues

This subsection, first, introduces basic concepts and restates, as a benchmark for the further discussion, the first-best case referring to Ramsey's (1928) basic model. Then some of the major general reasons for social and private rates of time preference to differ in terms of second-best arguments referring to the standard framework as well as policy implications that have been considered are introduced. Finally empirical objections put forward against the discounted-utility model with constant discount rate and empirical support for the split of time-preference rates, as well as theoretical implications are outlined and their importance for the present analysis is discussed.[2]

2.1.1.1 Equal Social and Private Rates of Time Preference in the First-best World

In general, the determination and relationship of social and private rates of time preference may be described with reference to a certain configuration of eight basic concepts:[3]

1. the private rate of pure time preference, ρ_p
2. the private rate of time preference, δ_p
3. the (private) consumption rate of interest, *CRI*
4. the private rate of return to investment, or marginal productivity of capital, i_p

---

[2] As regards the first-best case the account is based on Groom et al. (2005), Lind (1982) and Barro and Sala-i-Martin (2004: ch. 2). More complete accounts of the general debate and its current state can be found, e.g., in Lind et al. (1982), Portney and Weyant (1999), Frederick et al. (2002) and Groom et al. (2005).

[3] The following list could be further enlarged, e.g., by the private and social marginal rates of intertemporal substitution in consumption and the opportunity cost of public investment.

5. the market interest rate, i
6. the social rate of return to investment, i_s
7. the social rate of time preference, δ
8. the social rate of pure time preference, ρ

Note that the rates of pure time preference and of time preference are distinguished by the numeraire referred to. The rate of pure time preference represents the intertemporal weights placed on utility, the rate of time preference those on consumption. The rate of pure time preference is also called utility discount rate.

In the following the relationship between these rates in the first-best world is explained departing from the Ramsey growth model. The Ramsey model determines, for units of consumption as the numeraire, the social discount rate endogenously from the optimal saving, consumption and production decisions over time of an infinitely lived representative agent. For simplicity, the following brief exposition abstracts from depreciation, population growth and technological changes.

In the Ramsey model social welfare is represented by the intertemporal sum of the utility of a representative agent as follows:[4]

$$W(c(t)) = \int_0^\infty U(c(t)) \exp[-\rho t]\, dt \;. \tag{2.1}$$

Capital $k(t)$ is supposed to generate output $f(k(t))$, which can be devoted to consumption or investment subject to the following equation of motion of the capital stock,[5]

$$\frac{dk(t)}{dt} = f(k(t)) - c(t) \;. \tag{2.2}$$

Maximising welfare (2.1) subject to equation (2.2) yields the Euler equation:

$$U'(c(t))f'(k(t)) + U''(c(t))\frac{dc(t)}{dt} - \rho U'(c(t)) = 0 \;. \tag{2.3}$$

Substituting, in equation (2.3), the social rate of return to investment $r(t)$ for the social marginal productivity of capital $f'(k(t))$, and introducing the (per capita) consumption growth rate $g(t) = \frac{\frac{dc(t)}{dt}}{c(t)}$ as well as the elasticity of the marginal utility of consumption $\theta(t) = -\frac{U''(t)}{U'(t)}c(t)$, yields, when simplified, the familiar Ramsey rule:

$$i_s = \rho + \theta g = \delta \;. \tag{2.4}$$

The Ramsey rule (2.4) shows that in the centralised economy on the optimal path the social planner will choose consumption and savings such that the social rate of time preference equals the social rate of return to investment, $\delta = i_s$. Note that,

[4] The instantaneous utility function U is assumed to be twice differentiable with $U' > 0$, $U'' \leq 0$, and to satisfy Inada conditions, $\lim_{c\to 0} U' = \infty$, $\lim_{c\to\infty} U' = 0$.

[5] Instantaneous output f is assumed to be twice differentiable, with $f' > 0$, $f'' \leq 0$, and to satisfy Inada conditions, $\lim_{k\to 0} f' = \infty$, $\lim_{k\to\infty} f' = 0$.

by definition, the sum of the social rate of pure time preference ρ and the desire to smooth growing wealth over time θg is equal to the social rate of time preference δ.[6]

The private rates become relevant when the decentralised economy is considered. Therefore, assume a representative consumer, a representative firm and a public sector to interact on perfectly competitive markets for consumer and capital goods as well as financial capital under conditions of full rationality, complete certainty and in the absence of transaction costs and further externalities or distortions. Then the (private) consumption rate of interest, *CRI*, i.e. the rate at which consumers are willing to exchange consumption now to consumption in the future (equivalent to the private marginal rate of intertemporal substitution in consumption), thus coincides with the private rate of time preference, δ_p. The latter derives as $\delta_p = \rho_p + \theta g$. Moreover, in particular in the case of a representative consumer or of n identical individuals, as arising from aggregation over the individual(s), the social rate of time preference, δ, and the social rate of time pure preference, ρ, equal the respective private rates, δ_p and ρ_p.

Furthermore, the private consumption rate of interest, *CRI*, or private rate of time preference, δ_p, is equal to the market rate of interest, i, for, inequality would offer the possibility of arbitrage between the purchase of consumption goods and saving. Similarly, the firms' marginal rate of return to (private) investment, i_p, equals the market rate of interest, i, for, inequality would offer the possibility of arbitrage between investment in capital goods and investment on the financial market. Their equality to the market interest rate i implies furthermore that the four rates coincide, $CRI = \delta_p = i = i_p$. Due to the absence of distortions, also the social rate of return to investment, i_s, equals the private rate of return to investment, i_p, such that rates (2)–(7) coincide.

Note that public investments can only be carried out at the expense of resources (i.e., related to an equal opportunity cost) in the private sector such that their financing necessarily crowds out either private consumption, private investment, or a mixture of the two. Therefore, as an outcome of the Ramsey model, both i_s, i, and δ represent appropriate candidates for social discount rates to evaluate public investment projects, as long as costs and benefits are measured in consumption equivalents. The *CRI* is often used to measure δ, using observed rates of return on savings. The social utility discount rate, ρ, is the appropriate discount rate, if costs and benefits are measured in utility.

Thus, in general, in the first-best world in a decentralised economy with growth all rates (1)-(8) coincide apart from ρ and ρ_p. If there is no growth (i.e., $\frac{dc(t)}{dt} = g(t) = 0\ \forall t$), also the identical social and private rates of pure time preference, $\rho = \rho_p$, are equal to the other rates.

[6] $\theta(t)$ is a measure of the curvature of the utility function and equivalent to the coefficient of relative risk aversion. It represents preferences for smoothing consumption over time.

2.1.1.2 Second-best Issues and Policy Implications as Considered Thus Far

In the literature a multitude of reasons has been discussed for the social and private time-preference rates to differ (e.g., Baumol 1968, Lind 1982, Luckert and Adamowicz 1993, Stiglitz 1982, Tirole 1981). Particular attention has been paid to distortionary taxation, distortionary public investment and policies, imperfect competition, externalities in production, imperfect information, as well as intergenerational distributional concerns (i.e. altruism).[7] Especially with respect to the first four of these categories of reasons the existence of distortions in the real world is well established. For these four cases also the means to their avoidance or remedying are clear, at least theoretically. Distortionary taxation and distortionary public investment and policies are just to be avoided, imperfect competition necessitates an appropriate market regulation, production externalities are to be internalised. The respective measures are thus not necessarily directed directly to a correction of the rates of return or discount rates in question.

With respect to the issue of the present study, the broad debate which has been led in economics reveals, however, two lacunae. First, while a large part of the debate has been focusing on the public sector and the correct determination of the social discount rate to value public investment projects using cost-benefit analysis, it has not reflected, from a private-sector perspective, on the welfare-theoretic implications for private investment projects (e.g., to wit, Arrow and Kurz 1970, Lind et al. 1982, Arrow et al. 1995, Portney and Weyant 1999, Weitzman 2001, Groom et al. 2005). Second, abstracting from the four mentioned categories of reasons there remains a series of reasons which are particularly relevant in the present context for which there is, thus far, no optimal policy treatment as well established as for the first four cases. Three examples for such further reasons will be considered in subsection 2.1.2.

Among the exceptions, which have alluded welfare-theoretic implications for private investment projects, has been especially Hirshleifer (1966). In a discussion of the appropriate relationship of public and private discount rates he alluded the possibility of a subsidy for equilibrating different risk-pooling abilities between the public and private sectors and within the private sector. He concluded that such a subsidy is only justified if some kind of market imperfection hinders an efficient pooling of risks. His position has been supported by Baumol (1968). Reexamining Hirshleifer's discussion, Arrow and Lind (1970) have pointed out that Hirshleifer assumes identical costs of risk-bearing in the private and public sectors. They show that the costs of risk-bearing are negligible when risks are publicly borne. They underline that, therefore, the case that a public investment with a lower expected return than a given private investment is superior to the private alternative need not

[7] The case of distortionary taxation is particularly impressive, as, e.g., already Baumol's (1968) classic example showed: For corporation taxes of 0.5, income taxes of 0.25, and the social rate of time preference δ amounting to 0.06, firms, when investing, must pay dividends to shareholders such that they obtain a return of 0.06. Thus, shareholders must earn a pre-tax profit of 0.08 (for $0.08 * (1-0.25) = 0.06$), while firms must earn 0.16 (for $0.16*(1-0.5) = 0.08$). As a consequence, $i_p = 0.16$, whereas $\delta = 0.06$.

be a cause for concern. They are critical with direct subsidies to encourage more private investment, as they will not alter costs of risk-bearing but may encourage investments which are inefficient when the risk-bearing costs are taken into account. To achieve the desired result, they rather recommend policies which complete the system of markets for insurances relevant for private investments.

More recently, Grant and Quiggin (2003) have taken up the issue of the debate between Hirshleifer and Arrow–Lind in the context of the modern equity-premium debate.[8] They show that in the presence of financial-market imperfections the ability of superior risk-spreading via the taxation system may offset differences in technical efficiency between public and private projects. They discuss two approaches to stay with the advantages of risk-spreading without incurring the cost of undertaking projects with lower expected returns. First, they consider, similar to Arrow and Lind (1970), the possibility of a government offer of insurance against losses associated with cyclical fluctuations. They recognise that, while a voluntary public insurance would face the same adverse-selection problems as a privately provided insurance, a compulsory public insurance with the necessary degree of regulation and oversight to mitigate moral hazard problems could overcome the latter problems. At the same time the required regulation and oversight would, however, amount to a partial or complete public ownership of the project. In conclusion they find it more useful to consider a policy of public holdings of equity in private enterprises. Second, they consider the design of a set of payments to and from individuals that could mimic the risk-spreading benefits of a given public investment project without incurring any efficiency losses. For their single-period multi-state model a tax-financed unemployment benefit paid only in the recession state would have the appropriate properties. They point out that such a policy raises, however, difficulties of its own, notably if the labor supply response is taken into account.

With respect to environmental liability law, Endres et al. (2007) have recently analysed in a simple two-period partial-equilibrium model the robustness of the effect of diverging social and private discount rates on the abatement level in the two periods and the investment into abatement-technology improvements, and thus induced innovation, under different liability rules. They stay, however, with a welfare comparison and do neither reflect on the causes nor on policy implications of the split of discount rates.

2.1.1.3 Empirical Objections, Experimental Support, and Their Importance for the Present Study

While the traditional literature on project evaluation has usually been assuming agents to discount the future exponentially at one constant rate (e.g., Arrow and Kurz 1970, Lind et al. 1982), in the last decades the empirical foundations of the discounted-utility model have increasingly been investigated and questioned. Objections have been raised, e.g., with respect to the constancy of the discount rate

[8] Grant and Quiggin's (2003) basic model is detailed in subsection 2.1.2.2.

over time, its unicity in different cases of intertemporal choices, and the axiom of intertemporal independence (Frederick et al. 2002). Among the anomalies observed the variability of the discount rate over time, in terms of *hyperbolic discounting*, has attracted the most attention (e.g., Loewenstein and Prelec 1992, Laibson 1997, Harris and Laibson 2001, Groom et al. 2005, Winkler 2006). In the following it is first dealt with hyperbolic discounting and its theoretical implications. Then a number of experimental studies is considered which have directly investigated the relationship of social and private time-preference rates. Finally the importance of this research for the present study is discussed.

Hyperbolic discounting means that agents apply higher discount rates to near-term returns than to returns in the distant future. Thus, from the present perspective, t_0, a certain project evaluated at a discount rate D_0 in t_0 is discounted at some rate $D_1 < D_0$ in $t_1 > t_0$. Discounting at a declining discount rate implies in general a *time-inconsistent* behavior, in the sense that the same project, the investment in which was found optimal in t_0 for $t_1 > t_0$, is not judged so anymore in t_1 for that moment (Heal 1998, Winkler 2006).[9] Cropper and Laibson (1999) analyse the implications of hyperbolic preferences for private investment choices and public policy. Striving for time-consistent plans for a consumer with hyperbolic preferences, they formalise the agent's behavior as a game played by the consumer's different temporal selves. In the case of a rational, finite-lived consumer, the game has a unique subgame perfect equilibrium which may be characterised by an Euler equation similar to that of the Ramsey model (Arrow 1999). The consumption plan of the hyperbolic consumer characterising this equilibrium is thus observationally equivalent to the consumption path of a consumer who discounts the future exponentially. Along the optimal path the consumption rate of interest should always equal the rate of return on capital, $CRI = i_p$. The equilibrium of this game is, however, not Pareto efficient. For, consumers would be better off at any time, if each of them saved more. In absence of a commitment mechanism this will not occur (Phelps and Pollak 1968). Thus, government can induce Pareto improvement by subsidising the private rate of return on capital, or, equivalently, by lowering the required rate of return on investment. Note that this result arises irrespective of whether the government anticipates its own hyperbolic behavior or not. Rather, its optimal intervention brings the economy back to a path equivalent to one of exponential discounting.

There exists furthermore a number of experimental studies which have directly investigated the relationship of social and private rates of time preference. Thus, e.g., Pope and Perry (1989) conducted an experiment on management strategies for soil-erosion control. They show that a sample of business and natural-science students would apply a significantly higher discount rate, if a depletable natural resource was, hypothetically, under their own control, than they would prefer, if it was publicly managed. Similarly, Luckert and Adamowicz (1993) find in an experimental setting that students of a first year physiology course, abstracting from risk and inflation, recommend to use a lower discount rate, if a renewable natural resource,

---

[9] By contrast, in the case of exponential discounting with a constant discount rate, if at some time $t_0 < t_2$ investment in a project in t_2 has been found optimal, it will still be found optimal, when the project is reevaluated, at any other time $t_1 > t_0$, implying *time-consistent* behavior.

such as a forest, is publicly managed than if it is under private control. Lazaro et al. (2001) find that law students reveal higher rates of time preference when evaluating both future health benefits and monetary gains at an individual level than at a social level. While Pope and Perry (1989) do not reflect on theoretical reasons for their findings, the two latter papers refer for theoretical support to the so-called shizophrenic behavior approach as described, e.g., by Sen (1961, 1967) and Marglin (1963). According to this approach, individuals tend to apply higher rates of time preference in individual than in social decision contexts depending on whether they feel to act as an economic agent or as a social citizen, respectively.[10]

As regards the relationship of social and private time-preference rates, both (private) hyperbolic discounting and the experimental studies lead to the clear result of a private rate exceeding the social. Hyperbolic discounting can constitute a further reason for government intervention. However, it is neither specially linked to the energy sector, nor would its consideration change the analytical results of this study. Therefore, it will not be further considered in the model analysis in chapter 3. The schizophrenic behavior approach, as it does not explain the social rate as a result of aggregation of private rates, questions, in effect, the individualistic foundations of Paretian welfare theory, on which the basic argument of this study relies. Therefore, this approach will not be further considered here either.

2.1.2 Three Reasons Particularly Relevant for the Present Analysis

This subsection considers three more recent arguments for the social rate of time preference to stay below the private, which are, first, particularly relevant in the case of structural change in the energy industry under conditions of anthropogenic climate change and liberalised markets, and, second, abstract from the standard cases of distortionary taxation, distortionary public investment, imperfect competition, and production externalities. They both indicate the current heterogeneity of approaches and the complexity of aspects that matter. The first is *environment-specific* and focusses on the welfare and growth implications of the long-run environmental impact related to energy generation in a world of certainty. It goes back to Weitzman (1994). The second specially relates to *private investment financing* and refers to differing risk-premium claims on the financial market by individuals and the government, respectively. Based on a well-known observation in the finance literature, the long-standing theoretical economic argument has only recently particularly been substantiated by Grant and Quiggin (2003). The third relates to *long-run uncertainty*. It explains a declining social discount rate as an outcome of uncertainty over the growth rate in the long run and goes back to Gollier (2002a). In contrast to the first, the second and the third reason do not necessarily relate to environmental considerations.

---

[10] More recently, implicitly supporting this approach, the terms *homo oeconomicus* for an individual maximising private utility, and *homo politicus* for an individual who seeks to maximise social welfare have been introduced into the literature (e.g., Faber et al. 1997, 2002, Nyborg 2000).

2.1.2.1 The "Environmental" Discount Rate According to Weitzman (1994)

Weitzman (1994) develops a rationale for how environmental externalities influence the social discount rate, both immediately and over time. The focus of his argument is on 'environmental spending', Ψ. By environmental spending he means that GDP fraction which is spent on environmental improvement. Ψ constitutes a social cost external to the production process. To state his argument he departs from the stylised fact that in a world of environmental transition environmental spending increasingly grows important. His two theses are (1) that enduring environmental spending makes the social discount rate stay below the (undistorted) private discount rate, and (2) that the increasing environmental spending over time implies an intertemporally declining social discount rate.

His reasoning is as follows. Weitzman departs from a closed model economy in which (homogeneous) national income, Y, is either spent for consumption, C, gross investment, I, or environmental spending, Ψ:

$$Y(t) = f(k(t)) = C(t) + I(t) + \Psi(t)\ , \tag{2.5}$$

where f is the production technology, k the aggregate capital stock. He defines the relation between environmental spending, Ψ, and environmental damage, Θ, both expressed as a proportion of income, as:

$$\frac{\Theta}{Y} = \Gamma(\frac{\Psi}{Y})\ , \tag{2.6}$$

where Γ is a twice continuously differentiable function with constant returns to scale and $\Gamma_\Psi \equiv \frac{\partial \Gamma}{\partial \Psi} < 0$ and $\Gamma_{\Psi\Psi} \equiv \frac{\partial^2 \Gamma}{\partial \Psi^2} > 0$. Increased investment at time t by a marginal reduction in consumption, holding environmental spending constant, the following equation derives for the private rate of return on capital, i_p:

$$\frac{\partial Y}{\partial k} = f'(k) = i_p\ . \tag{2.7}$$

In order to prove his first thesis Weitzman claims environmental damage to be maintained at some initial level, $\overline{\Theta}$. Given equation (2.6) this can only be achieved by a marginal increase in environmental expenditure, $\Psi' = \frac{d\Psi}{dY}$, diverted from each unit of incremental output, $\frac{\partial Y}{\partial k}$. Hence, the social rate of return to investment derives as the rate of return in terms of output minus the rate of increase in expenditure required to maintain environmental standards:

$$i_s = \frac{\partial Y}{\partial k} - \Psi' \frac{\partial Y}{\partial k} = i_p[1 - \Psi']\ . \tag{2.8}$$

Taking the total derivative of equation (2.6) with $\Theta = \overline{\Theta}$ with respect to Y, solving for Ψ', and inserting in equation (2.8) yields:

$$i_s = i_p\Big[1 - Z\Big(1 + \frac{1}{\Lambda}\Big)\Big] , \tag{2.9}$$

where $Z \equiv \frac{\Psi}{Y}$ is the proportion of national income spent on environmental improvement, $\Lambda \equiv -Z\frac{\Gamma_\Psi}{\Gamma}$ the elasticity of environmental improvement (i.e. reducing Θ) with respect to environmental expenditure or the ease with which environmental damage can be reduced.

This analysis has two implications for the social discount rate. First, according to equation (2.9), for all $Z, \Lambda > 0$ the social rate of discount is lower than the private, $i_s < i_p$. Second, as i_s negatively depends on Z, for rising Z over time the socially efficient discount rate will decline.

Thus, according to Weitzman's analysis, the existence of consumption externalities reduces the level of the social rate of return below the private when society must divide the marginal return from investment between consumption and environmental protection. Moreover, the socially efficient discount rate will decline over time if the proportion of income spent on environmental goods, Z, increases, which holds if the elasticity of environmental improvement declines over time, or, similarly, if, with positive growth, environmental resources are luxury goods, as he maintains, too.

Weitzman's analysis has been criticised on a number of counts, notably for his reduced-form analytical set up without explicit modeling of preferences, environmental goods and externalities (Groom et al. 2005). While the assumption that some arbitrary environmental standard, $\overline{\Theta}$, must be maintained captures these effects generally, the substraction of environmental expenditures from the private rate of return in equation (2.8) remains rather ad hoc. Arrow et al. (1995) question moreover the interpretation of the derived i_s as a proper discount rate for its lack of an explicit conversion of the environmental benefits into equivalents of produced consumption.

However, despite the apparent further research need, it may be retained for the present analysis that, under fairly general conditions as those from which Weitzman (1994) departs, (negative) environmental externalities may induce the social rate of return to stay below the private. Moreover, it will decrease over time if the proportion of income spent on environmental goods, Z, increases. Anthropogenic climate change may be seen as a case in point for Weitzman's stylised fact of the positive and increasing environmental spending.

2.1.2.2 Diverging Social and Private Risk-premia on the Financial Market as Explained by Grant and Quiggin (2003)

In the finance literature, it is a well known empirical fact that there exists a (negative) spread between government and private security rates of any maturity. Thus, both the rates of return on equity capital (e.g., stocks) and on debt capital (e.g., corporate bonds with maturities of 10 years and more) exceed the rates of return on state securities with respective maturities. The issue has been discussed in a broad literature in both economics and finance particularly following the statement of the so-called equity premium puzzle and the risk-free rate puzzle (Mehra and Prescott

1985, Weil 1989). They state that standard theory, assuming complete asset markets and costless trading, can neither fully explain the observed high spread between the rates of return on government bills and private stocks, nor, why the government rates, usually dubbed 'risk-free', are so low as compared to historical data.[11] The discussions on the analytical puzzles add to an earlier literature in economics on uncertainty and financial-market distortions, with contributions by authors such as Sen (1961), Samuelson (1964), Hirshleifer (1966), Baumol (1968), Arrow and Lind (1970), and Stiglitz (1982), newly reflected, in part, e.g., by Hanley (1992) and Grant and Quiggin (2003). It has, apart from distortionary taxation, particularly been referring to differing risk-pooling and risk-spreading capacities at the individual and social levels. The more recent literature has moreover considered, e.g., idiosyncratic and uninsurable income risks, agent heterogeneity, borrowing constraints, and transaction costs. To date, however, none of the puzzles has been fully solved Kocherlakota 1996, Mehra and Prescott 2003.

In the following a theoretical argument provided by Grant and Quiggin (2003) is introduced. It substantiates the earlier arguments concerning risk pooling and spreading in the context of the debate on the equity premium puzzle. Grant and Quiggin formalise a statement by Mankiw (1986). Mankiw argued in an atemporal model with two dates that financial markets will not spread risks perfectly, if individuals, who are, *ex ante*, homogeneously exposed to risk, are, *ex post*, heterogeneously affected by catastrophes. Grant and Quiggin consider a framework where problems of *adverse selection* prevent individuals from insuring against systematic risk in labour income, modelled as the return to human capital. To state their argument, they study the formation of equilibrium asset prices in a model with two dates, $t = 0$, 1, using the state-claim approach. They assume a continuum of firms J uniformly distributed over the interval $[0,1]$, and a continuum of individuals I uniformly distributed over the rectangle $[0,2] \times [0,1]$. Individuals are assumed as risk-averse expected-utility maximisers with concave von Neumann-Morgenstern utility indices, u_i. Each individual is supposed to be endowed with human capital of quality hi and to be employed by firm j. More particularly, Grant and Quiggin define for each $i \in [0,2]$ the set of individuals who have human capital of quality hi as $H_i \equiv \{(i', j) : i' = i\}$ and assume that every individual $(i, j) \in H_i$ has the same preference relation over contingent consumption. There are two types of firms, 'risky' and 'safe', depending on whether they will shut down in recession or not. The share $\iota \in [0,1]$ of firms is assumed as risky, the remainder as safe. The type of a firm is supposed to be uncertain for anyone in the economy until uncertainty is resolved in $t = 1$. The model is normalised so that in the boom event both aggregate income and the return to human capital equal 1.Thus, in aggregate, the total endowment of

[11] Thus, for U.S. data from 1889 to 1978 (2000), the average annual stock return amounted to 6.98 (7.9)%, while that on T-bills were at 0.8 (1.0)%, giving rise to an equity premium of 6.18 (6.9) percentage points. By contrast, the Mehra and Prescott (1985) model predicted a stock return of 14.1, a risk-free rate of 12.7%, and, thus, in particular an average equity premium of 1.4 percentage points. Qualitatively equivalent findings hold also for France, Germany, Japan, and the U.K. which together account for more than 85% of capitalised global equity value (Mehra and Prescott 2003).

human capital for the economy is h. Further crucial model parameters are the riskiness of returns to physical (i.e., non-human) capital relative to the returns to human capital, ν, and the probability of recession, π.

In the model three states of the world arise. To the two global states, 'boom' (B) and 'recession' (R), add, for the case of recession, the two individual states of 'job' (j) and 'job loss' (nj):

(Bi) boom, occurring with probability $1-\pi$,
(Ri_j) recession *without* job loss, occurring with probability $\pi(1-\iota)$,
(Ri_{nj}) recession *with* job loss, occurring with probability $\pi\iota$.

In the boom state, any firm generates revenue equivalent to $h+\frac{(1-h)(1-\nu\iota)}{1-\iota}$ units of the single consumption good, if it is safe, and $h+(1-h)\nu$, if it is risky, with $\nu > 0$ by assumption. In both cases the wage paid to employee i by firm l is hi. The payout to non-employees is $\frac{(1-h)(1-\iota)}{1-\nu\iota}$ in the case of a safe firm, and $(1-h)\nu$ in the case of a risky firm. In the recession state, safe firms generate the same income as in the boom state, paying the same amount to their employees and non-employee claimants. However, the risky firms vanish and seize to generate income.

With respect to the financial market the model contains two types of securities, bonds and equity shares. The bond is assumed to be risk-free, in zero net supply, and to pay one unit of consumption in $t=1$ in every state of the world. Equity is the total sum of the firms' state-contingent non-wage or non-human capital payments. Equity supply is assumed to be fixed at one share. In the boom state, equity pays $(1-h)$, in the recession state $(1-h)(1-\nu\iota)$ units of the single consumption good. Each individual is supposed to be endowed with i units of equity and zero units of bonds. Moreover, the authors assume that trade in securities takes place before individuals have information about the index of the firm that employs them. No trade is possible after this information has become known.

For their analysis, Grant and Quiggin then consider two kinds of personal insurance contracts, with a positive payout in state Ri_1 and a negative payout in the other two states. In both cases, the insurance contracts must be entered into before individuals have information about the index of the firm which employs them. However, in the first case, the insurance contracts may not be renegotiated subsequently. As a result, idiosyncratic risk is *fully pooled*, i.e., for each $i \in [0,2]$ every individual in the set H_i will pool the returns to the individual's human capital and, thus, receive the (deterministic) return $(1-\iota)hi$ in the global recession event. In the second case, individuals are free to withdraw from the contracts after receiving private information about the index of their firm. Then, as a result of *adverse selection*, only individuals who find themselves employed by a risky firm will adhere to insurance contracts. Thus, in this case *no pooling* of idiosyncratic risk takes place.

To derive the equity risk premium for the two cases, the authors first find, relying on standard theory, that for any securities market equilibrium and any individual i there exists a supporting vector (p_{Bi},p_{Ri_j},$p_{Ri_{nj}}$) of state-contingent claim prices, where each price ratio equates the marginal rate of substitution between the respective state-contingent consumption claims. Moreover, for u_i strictly concave, the supporting price vector for individual i is unique. If the price b of a bond with payoff

(1,1,1) is normalised to 1, the supporting state-contingent claims price vector may, thus, be derived as the solution of the following system of equations:

$$\begin{aligned} p_{Bi} + p_{Ri_j} + p_{Ri_{nj}} &= 1 \ , \\ \frac{p_{Ri_j}}{p_{Bi}} &= \frac{\pi(1-\iota)u'(y_{Ri_j})}{(1-\pi)u'(y_{Bi})} \ , \\ \frac{p_{Ri_{nj}}}{p_{Bi}} &= \frac{\pi\iota u'(y_{Ri_{nj}})}{(1-\pi)u'(y_{Bi})} \ , \end{aligned} \tag{2.10}$$

with (y_{Bi},y_{Ri_j},$y_{Ri_{nj}}$) being the income vector of individual i. Since the boom state is the same for all individuals, and the set of securities spans the global boom and recession states, there exists a unique p_B in equilibrium, satisfying for all i

$$p_B = \frac{(1-\pi)u'(y_{Bi})}{(1-\pi)u'(y_{Bi}) + \pi(1-\iota)u'(y_{Ri_j}) + \pi\iota u'(y_{Ri_{nj}})} \ , \tag{2.11}$$

which implies, from the normalisation $b = 1$, that for all i

$$p_{Ri_j} + p_{Ri_{nj}} = 1 - p_B \ . \tag{2.12}$$

With $b = 1$, moreover, the price of a unit of equity may be expressed as

$$\begin{aligned} p_e &= (1-h)p_B + (1-h)(1-\nu\iota)(1-p_B) \\ &= (1-h)[1-\nu\iota(1-p_B)] \ . \end{aligned} \tag{2.13}$$

Since the expected return to a bond is 1 and that to a unit of equity is $(1-h)(1-\nu\iota\pi)$, the risk premium for equity, q, expressed in proportional terms, is then given by

$$q = \frac{\nu\iota[(1-\pi)-p_B]}{1-\nu\iota(1-p_B)} \ . \tag{2.14}$$

In order to characterise the equilibrium price for a claim on a unit of income in the boom state, p_B (and, hence, p_e and q), first, for the case of full pooling of risk among each class H_i of individuals, second, for the case with no pooling of idiosyncratic risk, they consider the following portfolio budget constraint for individual i in $t = 0$:

$$p_e\eta_i + b_i \ \leq \ p_e i \ , \tag{2.15}$$

where η_i is the individual's holding of equity, b_i the individual's holding of bonds.

In the first case, where all individuals are able to pool their idiosyncratic risky returns to their human capital in the recession state, the common return vector of individuals i to human capital is $[hi, (1-\iota hi)]$. Given the portfolio (η_i, b_i), this gives rise to the income vector

$$\begin{aligned} y_{Bi}^{p} &= hi + \eta_i(1-h) + b_i \ , \\ y_{Ri_j}^{p} &= (1-\iota)hi + \eta_i(1-h)(1-\nu\iota) + b_i \ , \\ y_{Ri_{nj}}^{p} &= (1-\iota)hi + \eta_i(1-h)(1-\nu\iota) + b_i \ . \end{aligned} \tag{2.16}$$

Inserting the income vector (2.16) in equation (2.11) for the price of a claim to consumption in the boom state gives:

$$p_B^p = \frac{(1-\pi)u_i'[hi + \eta_i(1-h) + b_i]}{(1-\pi)u_i'[hi + \eta_i(1-h) + b_i] + \pi u_i'[(1-\iota)hi + \eta_i(1-h)(1-\nu\iota) + b_i]} \ . \tag{2.17}$$

In the second case, without risk pooling, the income vector is

$$\begin{aligned} y_{Bi}^{np} &= hi + \eta_i(1-h) + b_i \ , \\ y_{Ri_j}^{np} &= hi + \eta_i(1-h)(1-\nu\iota) + b_i \ , \\ y_{Ri_{nj}}^{np} &= \eta_i(1-h)(1-\nu\iota) + b_i \ . \end{aligned} \tag{2.18}$$

Hence, the price of a claim to consumption in the boom state has to satisfy for all individuals i

$$p_B^{np} = \frac{(1-\pi)u_i'[hi + \eta_i(1-h) + b_i]}{E[u'(y)]} \ , \tag{2.19}$$

where

$$\begin{aligned} E[u'(y)] &= (1-\pi)u'[hi + \eta_i(1-h) + b_i] \\ &+ \pi(1-\iota)u'[hi + \eta_i(1-h)(1-\nu\iota) + b_i] \\ &+ \pi\iota u'[\eta_i(1-h)(1-\nu\iota) + b_i] \ . \end{aligned}$$

Assuming prudent behavior in the sense of Kimball (1990), i.e. marginal utility, u_i', strictly convex for all i, they show that, for given choices of (η_i, b_i), the price of a claim to consumption in the boom state in the case without risk pooling must exceed that of the case with risk pooling, $p_B^{np} > p_B^p$, which, according to equation (2.13), this also holds for the equity prices, $p_e^{np} > p_e^p$, and, thus, by equation (2.14), for the equity premia, $q_e^{np} > q_e^p$. That is, the absence of the capacity to pool idiosyncratic risk leads to an enhanced risk premium for equity.

Grant and Quiggin derive their result in an atemporal model with two dates. Further distortions absent, they show that an equity (risk) premium arises as a result of a market failure stemming from adverse selection, when an insurance contract against idiosyncratic income risk may be renegotiated after uncertainty is resolved. When risk-averse individuals can, thus, not fully spread the risk associated with recession, they are less willing to hold equity than they would be in a world of perfect financial markets.

The static model Grant and Quiggin (2003) analyse does not permit to consider the possibility of smoothing consumption over time through borrowing, lending, or asset liquidation. One objection raised by Kocherlakota (1996) in this respect against Grant and Quiggin's predecessor model by Mankiw (1986) is that difficulties associated with the absence of insurance markets could be overcome by intertemporal consumption-smoothing. However, Kocherlakota's argument is based on the assumption that individuals can borrow and lend freely at the bond rate. Thus, Grant and Quiggin maintain that, since governments can borrow and lend freely at the bond rate in their model, a dynamic analogue of it, where individuals face idiosyncratic human capital risks that are correlated with systemic risks to equity that unfold through time, would yield results similar to those derived in the static context.

The importance of Grant and Quiggin's argument in the context of the present analysis is obvious. In liberalised markets, (private) utilities rely upon the private financial market to finance investments in new technologies. Any negative distortion of it, as induced, e.g., by uninsurable risks, tends to hamper private investment. Grant and Quiggin's suggestions to increase social efficiency have briefly been considered in subsection 2.1.1.2. The treatment of the distortions in the conditions of structural technological change in the energy sector proposed in section 3.4 is generally in vein with their second suggestion.

2.1.2.3 Declining Social Discount Rate in View of an Uncertain Future According to Gollier (2002a)

Gollier (2002a) studies the effect of uncertain growth in the long run on the social rate of time preference, δ, considering particularly the importance of risk preferences in this context. His motivation is that financial markets hardly provide an optimal guideline for investing in technologies that prevent long-lasting risks to occur, as liquid financial instruments with such maturities do not exist.[12] Instrumental to his model is the concept of *prudence*, as formalised by Kimball (1990). An agent is prudent if his willingness to save increases in the face of an increase in his future income risk or, technically, if the third derivative of his utility function is positive. Gollier shows that prudence leads, first, to a social discount rate that is smaller than in the case of certain growth, and, second, moreover to one that declines over time.[13]

In order to state his argument, Gollier departs from the framework of a 'tree economy' (Lucas 1978) in which each individual is endowed with some productive capital, a 'tree', generating a number of 'fruits' in each period. There is no possibility to plant new trees. The number of fruits, i.e. the income or, equivalently, consumption, is assumed to grow in period t at the uncertain and exogenous rate $\widetilde{g}_t$, the $\widetilde{g}_t$ being independently and identically distributed over time for all t. For simplicity, he concentrates on a two-period model with three dates, $t = 0, 1, 2$. He defines $c = c_0$ as consumption in $t = 0$ and $\widetilde{c}_t = c_{t-1}(1+\widetilde{g}_t)$ as consumption at dates $t = 1, 2$. Seen

[12] Time horizons, e.g., of government bonds generally do not exceed 40 years, whereas relevant time horizons associated with global climate change extend, at least, over some 100 years.

[13] With respect to the shape of the socially efficient discount rate, Weitzman (1998) derived a similar result studying the effect of uncertainty of the social rate of return to capital, i_s, on the social discount rate, assuming risk neutral agents. His analysis omits, however, the explicit treatment of risk preferences. See Groom et al. (2005: 460–465) for a discussion.

from $t = 0$, the $\widetilde{c}_t$ for $t = 1, 2$ are uncertain. In order to describe the effect of uncertain growth on the *short*-term behaviour of the discount rate, he first considers only the first period. To disentangle the attitudes towards time and risk, Gollier assumes for the social planner Kreps–Porteus–Selden preferences (Kreps and Porteus 1978, Selden 1979). He represents them by defining the certainty equivalent consumption m in $t = 1$, as usual under expected utility theory:

$$v(m) = Ev(\widetilde{c}_1) \,, \tag{2.20}$$

and intertemporal social welfare as:

$$u(c) + (1+\rho)^{-1}u(m) \,, \tag{2.21}$$

where the utility functions v and u are assumed to be increasing and weakly concave in their argument, and ρ is the social rate of pure preference. For $u \equiv v$, from this representation the intertemporal time-separable expected utility model derives, with the classical objective function,

$$v(c) + (1+\rho)^{-1}Ev(\widetilde{c}_1) \,. \tag{2.22}$$

In this economy the only possibility is to invest in some prevention effort to improve future individual crops. To evaluate the effect of the uncertain growth on the socially optimal discount rate, Gollier considers when society should invest in a marginal project which costs a small x ($\Delta c = -x$) in $t = 0$ and brings a sure benefit $(1+r)x$ in period $t = 1$. This sure benefit increases the certainty equivalent consumption in $t = 1$ by

$$\left.\frac{\partial m}{\partial x}\right|_{x=0} = \frac{Ev'(\widetilde{c}_1)}{v'(m)}(1+r) \,. \tag{2.23}$$

The social welfare function (2.21) is increased by this investment if:

$$-u'(c) + (1+\rho)^{-1}\, u'(m)\left.\frac{\partial m}{\partial x}\right|_{x=0} \;\geq\; 0 \tag{2.24}$$

or, equivalently,

$$r \;\geq\; \frac{u'(c)}{(1+\rho)^{-1}\, u'(m)}\frac{v'(m)}{Ev'(\widetilde{c}_1)} - 1 \;=\; \delta_1 \,, \tag{2.25}$$

where δ_1 is the social rate of time preference in period 1, here equivalent to socially efficient discount rate, corresponding to the equilibrium risk-free rate in this exchange economy.

Gollier now compares δ_1 and δ_1^c, the socially optimal discount rate in an economy with a sure consumption level equal to $E\widetilde{c}_1$ in $t = 1$. As without uncertainty risk aversion does not matter, the certainty equivalent in such an economy is then $m = E\widetilde{c}_1$, such that

$$\delta_1^c = \frac{u'(c)}{(1+\rho)^{-1}\, u'(E\widetilde{c}_1)} - 1 \;. \tag{2.26}$$

From conditions (2.25) and (2.26) for time-separable expected utility preferences with $u \equiv v$ follows:

$$\delta_1 \;\leq\; \delta_1^c \;\Leftrightarrow\; Ev'(\widetilde{c}_1) \;\geq\; v'(E\widetilde{c}_1)\;. \tag{2.27}$$

As the right-hand side constitutes Jenson's inequality, equivalence (2.27) holds for any c and any distribution of $\widetilde{c}_1$, if and only if the marginal utility of consumption, v', is convex. Thus, uncertainty in growth will reduce the discount rate, or, equivalently, precautionary saving will occur, if and only if individuals behave in a prudent way ($v''' > 0$).

In order to describe the effect of uncertain growth on the *long*-term behaviour of the discount rate, Gollier extends the analysis to the second period, with $(1+\widetilde{g}_1)(1+\widetilde{g}_2)-1$ being the growth rate over the two periods. While with $u \equiv v$, according to condition (2.25) the socially optimal (gross) rate to discount costs and benefits occurring at date $t = 1$ is

$$1+\delta_1 = \frac{u'(c)}{(1+\rho)^{-1}E[u'(c(1+\widetilde{g}_1))]}\;, \tag{2.28}$$

the socially optimal discount rate per period for a cash flow occurring at $t = 2$ derives as:

$$(1+\delta_2)^2 = \frac{u'(c)}{(1+\rho)^{-2}E\big[u'(c(1+\widetilde{g}_1)(1+\widetilde{g}_2))\big]}\;. \tag{2.29}$$

Gollier notes that δ_2 is not equal to δ_1, except in the case of the isoelastic utility function, with $v'(c) = c^{\phi}$ for some constant relative risk aversion $\phi > 0$, as can easily be seen departing from conditions (2.28) and (2.29), which yield for $v'(c) = c^{\phi}$ with $\phi > 0$:

$$\begin{aligned}(1+\delta_1)^2 &= [(1+\rho)^{-1}E(1+\widetilde{g}_1)^{-\phi}]^{-2}\\ &= (1+\rho)^2[E(1+\widetilde{g}_1)^{-\phi}E(1+g_2)^{-\phi}]^{-1}\\ &= (1+\rho)^2[E((1+\widetilde{g}_1)(1+g_2))^{-\phi}]^{-1} \;=\; (1+\delta_2)^2\;,\end{aligned} \tag{2.30}$$

implying that $\delta_2 = \delta_1$. Thus, for constant relative risk aversion the wealth effect of a longer time horizon just compensates the precautionary effect.

Using conditions (2.29) and (2.30), Gollier then shows that for the consumption growth rate nonnegative almost surely and a three times differentiable utility function v, the following equivalence holds,

$$\delta_2 \;\leq\; \delta_1 \;\Leftrightarrow\; v'(c)E\,v'[c(1+\widetilde{g}_1)(1+\widetilde{g}_2)] \;\geq\; E\,v'[c(1+\widetilde{g}_1)]v'[c(1+\widetilde{g}_2)]\;, \tag{2.31}$$

if there is, in addition to the former assumptions, decreasing relative risk aversion.[14] Thus, in the case of an almost surely positive growth rate per period, prudent

[14] Gollier (2002a: 158) provides the following proof: Let function h: $R_+^2 \to R$, $h(x_1,x_2) \equiv v'(cx_1,x_2)$, be log supermodular, i.e. for all $(x_1,x_2),(x_1',x_2') \in R_+^2$:

behavior and decreasing relative risk aversion are sufficient, first, for the socially efficient discount rate to stay below that in the case of certain growth, and, second, for the long-term discount rate to be smaller than the short-term one. The latter result also extends on indeterminate, and thus in particular infinite, time horizons (Gollier 2002a,b). That is, with increasing time horizon, the socially efficient discount rate declines.

Gollier (2002a) provides a theoretically most rigorous rationale for declining discount rates. His analysis implies potentially testable propositions derived from expected utility theory. The formal economic foundation of the determination of long-term discount rates avoids ad-hoc adjustments of the discount rate. The complexity of the analysis depends on the assumptions concerning the probability distribution of growth and the intertemporal relationships. For convenience, although unrealistic, Gollier (2002a) assumes that the growth shocks are i.i.d. He, thus, avoids the complications associated with the analysis of serially correlated shocks. At the same time, to determine the trajectory of the social discount rate based on his analysis knowledge about the aversion to consumption fluctuations over time, the pure time-preference rate, and the degree of relative risk aversion is necessary. In addition, the probability distribution of growth has to be characterised in some way. Thus, an actual empirical application of his approach would be associated with high informational requirements.

The importance of Gollier's study for the present analysis coincides with its motivation. Gollier's aim is to determine the socially optimal discount rate for public investment projects that entail costs and benefits in the very long run, thus, the time horizons of which extend far beyond the longest maturity of any security available. His focus is on the effect of uncertain (consumption) growth. While he does not treat a specific source of uncertainty, it is clear that anthropogenic climate change may constitute an important cause. Prudent behavior, which is crucial for his argument, occurs as a rational strategy in the face of uncertainty. The factual inability of financial markets to reflect, and, in particular, to insure against, such long-term risks

$$h[min(x_1,x_1'),min(x_2,x_2')]\,h[max(x_1,x_1'),max(x_2,x_2')] \;\geq\; h(x_1,x_2)\,h(x_1',x_2')\ .$$

For $x_1 = x_2' = 1$ and all $x_1', x_2 > 1$, the log supermodularity of h implies:

$$h(1,1)\,h(x_1',x_2) \;\geq\; h(1,x_2)\,h(x_1',1)\ ,$$

which is, for all $g_1 = x_1' - 1$ and $g_2 = x_2 - 1$ with $g_1, g_2 > 0$, equivalent to

$$v'(c)\,v'[c(1+g_1)(1+g_2)] \;\geq\; v'[c(1+g_1)]\,v'[c(1+g_2)]\ .$$

Given an almost surely positive growth rate per period, the log supermodularity of h is sufficient to guarantee that the latter condition holds almost everywhere. Taking the expectation of this inequality directly yields inequality (2.31), which implies in turn that $\delta_2 < \delta_1$. For a three times differentiable utility function v, the log supermodularity of h means, furthermore, that the cross derivative of log h is positive, which is equivalent to require that relative risk aversion is decreasing.

□

constitutes another case for welfare-enhancing government intervention. Therefore, Gollier's analysis implies another second-best case relevant for the present analysis.

2.1.3 Summary and Treatment of the Assumption in the Model

This subsection summarises the findings from the analysis of the discounting literature, states open issues, and then describes the treatment of the assumption in the model below and the contribution of the analysis.

2.1.3.1 Summary of the Findings from the Analysis of Discounting Literature

In this section, as a benchmark for the further discussion, first, the first-best case was restated with reference to Ramsey's (1928) basic model. It establishes the equality of the above mentioned rates (2)–(7), including, in the case of the absence of growth, rates (1) and (8), and thus of the social and private rates of (pure) time preference. Then deviations from first-best conditions were considered. It was shown that there exists in economics a series of well recognised reasons for the social and private rates of time preference to differ in real world. They include distortionary taxation, distortionary public investment, imperfect competition, and production externalities. They generally imply the social rate to stay below the private. Moreover it was shown that (private) hyperbolic discounting may lead to the same conclusion of a socially efficient rate of time preference staying in general below the private, and that also experimental evidence supports this divergence. In addition, the corresponding (negative) spread between government and private security rates of any maturity is a well established fact in finance. Economists have thus been advancing over Baumol's (1968) 'dark jungles' insofar as there is now wide unanimity that and how in general social and private rates of time preference differ.

Economists also widely agree about optimal policy treatments for the four mentioned, well established causes for the split of time-preference rates, at least in theory. Thus, while the welfare-theoretic consequences of diverging time-preference rates for *private* investments have never systematically been addressed, in this respect these four causes do not constitute an issue.

In subsection 2.1.2, however, three reasons were introduced, which are clearly relevant for investments in the energy sector but also clearly go beyond the four standard cases. As put forward by Weitzman (1994), Grant and Quiggin (2003) and Gollier (2002a), they relate, respectively, to increasing environmental externalities over time as well as adverse-selection problems or uninsurable long-run risks as inducing financial-market distortions. The arguments were set out in detail in order, first, to illustrate the current heterogeneity of approaches to state causes for the split discount rates, and, second, to further detail to some extent the complexity of the aspects that matter.

Among the three contributions only Grant and Quiggin (2003) consider possibilities of internalisation with respect to the second-best case they deal with. The two approaches they discuss – insurance provision and (redistributive) subsidies (subsection 2.1.1) – stand for two major policy options. Their considerations point, however, at the same time to significant transaction-cost, incentive, and feasibility issues which may be generally related to the implementation of such policies. Weitzman (1994) and Gollier (2002a) do not reflect on second-best implications. Given that the discount-rate distortion considered by Weitzman arises only due to environmental externalities, only the case of incomplete environmental policies, as possibly occuring in practice, would let it persist as a second-best issue. In this case a direct correction of the discount rate might occur as a useful complementary second-best policy. The reduced-form set-up of his model does, however, not allow for a respective extension of the analysis. To cope with distortions from uninsurable long-run risks as pointed out by Gollier also the two policy approaches considered by Grant and Quiggin constitute possible general options.

Taking into account in addition the three contributions considered in subsection 2.1.2, six further points can be retained with respect to second-best issues relating to the split social and private time-preference rates.[15] First, neither the causes of the split nor policy implications associated with it or the welfare and policy implications as directly induced by it have, thus far, been subject to a systematic analysis. Second, to explain the causes of the split at present no unique or encompassing framework is available. Rather a multitude of different models prevails. Third, the variety of models and explanations does not only show the particular complexity of relevant aspects to be taken into account, but indicates in particular that the split is in general induced by several causes. Fourth, the split in itself does *not* constitute a market failure, but may be – though not necessarily, e.g., in the case of distortionary policies – induced by an underlying one. It is, however, in general the sign of the existence of a welfare-decreasing distortion in the economy. Fifth, there exists a number of reasons, such as transaction costs or incentive issues, for which the split in itself, rather than an underlying market failure directly, may be the point of reference for a welfare-enhancing policy intervention. Finally, there is a number of cases of market failure which are clearly relevant for investments in the energy industry under the condition of anthropogenic climate change and for which no optimal policy treatment is established.

2.1.3.2 Treatment of the Assumption in the Model

The present study is interested in the welfare-theoretic implications of the split of social and private time-preference rates for the conditions of structural technological change in the energy sector. In line with this focus the subsequent model analysis just departs, like Arrow and Kurz (1970: 116), from differing time preferences of

[15] Note that, though necessarily incomplete, the analysis of the literature has been including the major surveys as well as a number of key contributions in detail. The following statements are consistent with the whole of the literature consulted.

the representative consumer and the society, assuming in particular the private rate of time preference to exceed the social (section 3.1). In view of the multitude of different reasons relevant for the split and the to a good part still open state of discussion with respect to causes, an endogenous explanation of the split is avoided. Thus, e.g., the public sector (apart from the regulator), financial markets, eventual imperfections of the commodity markets, production externalities, and uncertainty remain exogenous to the analysis. Moreover, the specification of the environmental externality is such that it does not influence the private time preferences. This proceeding allows for the *most general* analysis of the welfare and policy implications of this assumption.

In the light of the above literature analysis the claim of generality needs, however, a qualification in two respects. First, the welfare implications of the split, whatever they are, occur always, irrespective of the particular causes of the split. Second, a policy intervention aiming at the correction of the *consequences* of distorted time-preference rates is only justified, if and only if the split is induced by an underlying market failure which cannot directly or differently be remedied. Thus, also for the most general study of policy implications of the split, the analysis is to be restricted, on welfare-theoretic grounds, to only the respective subset of all causes.

Accordingly, the investigation below provides a general clarification with respect, first, to the welfare implications of the split for private investments irrespective of its causes, and, second, to its policy implications for a certain class of causes. However, it does not shed further light on particular causes of the split, their implications for the determination of the social discount rate or policy implications directly related to the causes.

2.2 Time-Lagged Capital Theory

Structural change is usually defined as a change in the relative composition of an economy's capital stocks. In this study, as applied to the energy sector, it is assumed to be characterised by a shift in the relative use of different energy technologies over time. It thus usually occurs, e.g., when an old, established energy technology is partially or fully replaced by a new one. In the energy industry this kind of technological change is particularly marked by two features. First, power plants are particularly long-lived and cost-intensive capital goods. That is, as compared to other private investment projects, their construction necessitates a particularly large reallocation of resources from other opportunities in the economy, which, of course, will only be made if it seems economically profitable. Therefore, the current conditions of investment and replacement are instrumental to determine whether and when structural change will occur. Second, the construction period of a new power plant is often particularly long as compared to other investment projects. Economically speaking, there is a substantial *time lag* – to be designated by σ in the following – between the costs of investment and the benefits of production of new capital goods.

The kind of economic modeling that exactly captures these two features is *time-lagged capital theory*. In chapter 3 it is used to analyse the transition from an established polluting to a new clean energy technology. This section briefly highlights the historical evolution of time-lagged capital theory and introduces the basic model of the particular strand which is taken up in this study (subsection 2.2.1). Subsection 2.2.2 describes the contributions made to its further development in this study.

2.2.1 Its Evolution and the Basic Neo-Austrian Three-process Model

The idea of time-lagged capital accumulation goes back to the Austrian school of economics, especially to the work of von Böhm-Bawerk ([1889]1921).[16] It was revived by the neo-Austrian capital theory in the 1970s. Von Weizsäcker (1971), Hicks (1973), and Faber (1979) stand for three major approaches in this vein. With Kydland and Prescott (1982) the time-to-build feature became, moreover, prominent in the modern macroeconomics real business cycle literature.

The modeling framework used in the following analysis stands in the tradition of the third of the three strands of neo-Austrian capital theory.[17] This approach is basically characterised by its basic three-process model. It studies the conditions of innovation of a new production technique, T_2, which is only to be produced, into a system with an already existing technique, T_1 (Faber and Proops 1991, Winkler 2005). A *technique* is defined as the minimal combination of production processes necessary to produce the consumption good from non-produced inputs. The two techniques are assumed to produce a homogeneous consumption good $c(t) = c_1(t) + c_2(t)$. Labor is the only primary input. It is usually supposed to be given in a fixed amount $\bar{l}$ at all times t. Thus, while T_1 produces $c_1(t)$ from labor $l_1(t)$ alone, T_2 comprises an additional capital good production process in order to generate $c_2(t)$ from labor $l_2(t)$ and capital $k(t)$.[18] However, capital good production is assumed to be *time-lagged* by one period, such that the equation of motion of the capital stock is

$$k(t) = \frac{l_3(t-1)}{\lambda_3} + (1-\gamma)k(t-1) , \tag{2.32}$$

[16] Austrian economists were particularly interested in the issues of time and change in the economy. Schumpeter (1934, 1939), e.g., particularly dealt with structural technological change as an aspect of evolutionary change in the economic process.

[17] Further important contributions to this strand include, e.g., Bernholz et al. (1978), Faber (1986), Faber and Proops (1991, 1998), Faber et al. (1995, 1999), Stephan (1995), and Winkler (2003, 2005). See Winkler (2003: ch. 2) for a survey of the development of capital theory in general and its neo-Austrian variant in particular.

[18] By convention, all three production processes have usually been assumed to be linear-limitational. In equilibrium, labor is efficiently allocated to the three processes, such that $\bar{l} = l_1(t) + l_2(t) + l_3(t)$, where $l_3(t)$ is the amount of labor used in the capital good process.

where λ_3 is labor coefficient of capital good production, γ the deterioration rate of capital. Both are assumed to be constant and exogenously given.

Crucial concepts for the statement of the necessary and sufficient investment conditions are those of the *roundaboutness* and *superiority* of techniques. A technique (T_2 in the present case) is said to be more *roundabout* than another technique (T_1 in the present case), if and only if it needs more time to produce the same amount of $c(t)$ as the other technique (T_1 in the present case). A technique is called *superior* to another one, if and only if for a given endowment of non-produced inputs the maximal producible sum of consumption goods within the given time horizon τ is larger for that technique than for the other one.

Suppose a representative consumer who maximises her aggregate utility from consumption over time with a positive rate of time preference $\rho > 0$. Then, obviously, superiority of T_2 over T_1 is a necessary condition for the innovation of T_2 to be optimal. For sufficiency, T_2 has to yield a higher intertemporal utility to the representative consumer within the given time horizon τ than T_1. This is satisfied for a representative consumer with rate of time preference $\rho > 0$ and given positive time horizon τ, if her instantaneous utility function $U(c(t))$ is concave, i.e. $U_c > 0$, $U_{cc} < 0$.

Initially formulated for finite time horizons the modeling framework was extended to infinite time horizons by Stephan (1983, 1985, 1995). More recently Winkler (2003, 2005, 2008) provided a series of further refinements and extensions. He proved that the necessary and sufficient conditions for innovation of the new technique can be derived from the first-order conditions of the respective intertemporal optimisation problem. He systematically analysed the effect from joint production of the consumption or the capital good and a flow or stock pollutant, respectively, on the necessary and sufficient conditions for innovation. Moreover, he considered, as means of pollution treatment, the cases of complete abatement and the imposition of an emission standard as well as, separately, the negative valuation of environmental degradation by emissions entering into the utility function in addition to consumption. In no case he considers market interactions. He formulated and solved the model in continuous time using optimal-control techniques and calculated optimal paths of the model variables by use of numerical methods. Finally, he considered a version of the model with the consumption-good sector modeled by a Cobb-Douglas function in which two new clean energy technologies compete for their introduction in a system with an established polluting energy technology.

While the analysis in this framework has usually been carried out for a centralised economy, Stephan (1995) has discussed prices and price systems in finite- and infinite-horizon models of the neo-Austrian kind in discrete time. He proves, however, only the existence and Pareto optimality of a general equilibrium in an economy with markets only for non-durable goods. The modeling framework has, thus far, not yet been used for the explicit analysis of market failures in a decentralised economy with a complete market system. Nor have in this framework ever the effects from diverging social and private rates of time preference or the interplay between different kinds of technological change been analysed.

2.2.2 New Extensions

The model set up in the next chapter takes up (a slightly modified version of) the basic three-process model with infinite time horizon and adds, in a new combination, different elements developed in a general way by Winkler (2003, 2005). As compared to the basic neo-Austrian three-process model the consumption-good sector is substituted for an energy sector. For convenience, utility is assumed to directly derive from energy consumption. The model is formulated in continuous time. It takes into account joint production of energy and emissions. The latter are supposed to be negatively valued by the consumer.

In addition to that, for the analysis aimed at in the present study the modeling framework is methodically further extended in three new ways:

1. By introducing a complete market system, it is now opened to the explicit analysis of market failures, and their remedying, in the standard framework of Paretian welfare economics.
2. The specification of an abatement function allows to analyse the interplay of two kinds of technological change, *gradual*, i.e. the (gradual) refinement of a technology by enactment of an end-of-pipe abatement technology, and *structural*.[19]
3. In addition to the social, private time preferences are considered which may differ from the social.

On the level of modeling these three elements systematically extend the former framework of this kind of time-lagged capital theory. All of them are essential for the analysis in chapter 3, the results of which are particularly driven by the emission externality, the split of social and private rates of time preference and the time-lagged structure of capital accumulation.

2.3 Analytical Structure of the Model in Chapter 3

The model developed in chapter 3 to study the conditions of structural change in the energy industry under the conditions (i)–(iv) as specified in section 1.1 is an intertemporal general equilibrium model with infinite time horizon formulated in continuous time. The infinite time horizon has been chosen in order to analytically concentrate on the effects related to the (constant) rate of time preference as the only other remaining time-related parameter. The agents – the regulator in the cases of the centralised and the regulated decentralised economy, the representative Ramsey consumer and two representative firms in the cases of the unregulated and regulated decentralised economy – are supposed to maximise their intertemporal welfare or their profits, respectively, under certain constraints subject to the standard

[19] The term gradual has only been adopted in this study in order to terminologically distinguish this kind of technological change from the notion of structural technological change.

assumptions of full rationality and complete certainty.[20] On the production side a vertically integrated time-lagged production system is assumed composed of an energy and an investment sector.

From the model necessary and sufficient conditions for investment in the new technology and for partial and full replacement of the established technology are derived, both at the social optimum and in the unregulated market equilibrium. To derive these conditions the stationary states occurring in the optimal development of the economy are exploited. Two stationary states, both at the social optimum and in the unregulated competitive market equilibrium, arise as a consequence of the linear or linear-limitational functions assumed in the production sectors, which provide the respective corner solutions.[21] At the social optimum, in addition, a third stationary state occurs associated with an interior equilibrium.

The method used to solve the agents' intertemporal optimisation problems is *time-lagged optimal control theory*. In general, optimal control theory provides the optimal paths of the control variables (i.e., those variables the actor can influence) which maximise a goal functional (here, the intertemporal welfare function of the regulator or the representative consumer, respectively) subject to a given set of stock variables and their dynamics and a set of further restrictions. The understanding of the solutions requires some familiarity with optimal control theory and the theory of ordinary differential equations.[22] At places requiring a particular knowledge, e.g., with respect to the time-lagged nature of the problem analysed, respective references are indicated. All results presented have been derived analytically, i.e. without use of numerical methods. In particular, the question under study did not require the consideration of the optimal paths of variables out of the stationary states. However, in general, as soon as the investigation had been extended to the explicit analysis of such optimal paths the use of numerical methods would have been necessary.

2.4 Conclusion

In this chapter the literature behind the two distinctive features in the following model analysis was considered and the relationship and contribution of this study to it was described, first with respect to the crucial assumption of the split of social and private time-preference rates, then with respect to time-lagged capital theory. At the same time it was introduced to the theoretical issue of this study. Finally the analytical structure of the model at the basis of the analysis in the following chapter was explained. In the next chapter the theoretical model of structural technological change in the energy sector is introduced and analysed.

---

[20] The assumptions of exponential (e.g., versus hyperbolic) discounting and certainty, and their relationship to the subsequent analysis have been discussed in section 2.1.

[21] The rationale for these specific functional forms as well as further assumptions concerning the energy industry and their analytical implications are discussed in section 4.2.

[22] Useful general treatments of optimal control theory can be found, e.g., in Chiang 1992, Feichtinger and Hartl 1986,Gandolfo 1996 and Kamien and Schwartz (1991).

http://www.springer.com/978-3-7908-2182-6

Distorted Time Preferences and Structural Change in the Energy Industry
A Theoretical and Applied Environmental-Economic Analysis
Heinzel, C.
2009, XVI, 166 p., Hardcover
ISBN: 978-3-7908-2182-6
A product of Physica-Verlag Heidelberg