universität wien

DISSERTATION

Titel der Dissertation

“Essays on Decisions:
The Impact of Information and Discounting”

Verfasserin

Magdalena Six, MSc.

angestrebter akademischer Grad

Doctor of Philosophy (PhD)

Wien, 2015

Studienkennzahl lt. Studienblatt:	A 094 151 146
Studienrichtung lt. Studienblatt:	Betriebswirtschaft
Betreuer:	Univ.-Prof. Dr. Franz Wirl

Acknowledgments

When I started my position as a PhD student at the chair of Prof. Wirl, I was equipped with a "box of mathematical tools", but I had no idea how to use them on managerial or economic problems. It is thanks to Prof. Wirl that I learned to understand the economic aspects behind the formulas. Prof. Wirl is a virtually inexhaustible source of real-world stories that illustrate economic and managerial concepts better than any textbook could do. It is only fair to say that if it was not for him, this PhD thesis would not exist, his door was always open for any kind of questions or discussions, and I am deeply grateful for all the time he took for me.

I would also like to thank my colleagues at the chair, Christoph Graf, David Hirschmann, and David Wozabal. I did not only learn a lot from them about how academic research works, they also made the time in between the academic research very enjoyable.

A further necessary (but not sufficient) condition for the development of this thesis was the existence of my boyfriend Silviu in my life. He did not only provide me with motivation, support and (physical) energy when needed, as proofreader he was also my most demanding critic. And last but not least I want to thank my parents, my sister and my friends for supporting me, listening to my problems and helping me in a variety of ways when needed.

Contents

1 Introduction **7**

1.1 Chapter 2: The Effect of Information on the Quality of Decisions (published) . . 9

1.2 Chapter 3: Information as Key Determinant in Switching the Electricity Suppliers for Austrian Households (submitted) 9

1.3 Chapter 4: Optimal Pollution Management When Discount Rates are Endogenous (submitted) 10

2 The Effect of Information on the Quality of Decisions **11**

2.1 Introduction 11

2.2 Concept and Hypotheses 13

2.2.1 Concept 13

2.2.2 Hypotheses 14

2.3 Method 15

2.3.1 The Iterative Improvement of the Information Set 15

2.3.2 The (Averaged) Relative Volume 17

2.4 Simulation 18

2.4.1 One Simulation Run 18

2.4.2 Parameter Settings 19

2.5 Results 20

2.5.1 Hypothesis H1 21

2.5.2 Hypothesis H2 21

2.5.3 Hypothesis H3 22

2.5.4 Hypothesis H4 23

2.6 Conclusion and Topics for Future Research 23

3 Information as Key Determinant in Switching the Electricity Supplier **27**

3.1 Introduction 27

3.2 Data and Descriptive Statistics 30

3.3 Estimating Binary Response Models 31

3.4 Variables 32

3.5 Estimation Results 33

3.5.1 Discussion and Policy Implications 36

3.5.2 Goodness of Fit Tests 38

3.6 Treating the Variable *Tariff Calculator* as Endogenous 39

3.6.1 Potential Endogeneity of the Tariff Calculator and Possible Instrumental Variables . . . 39
3.6.2 The Reference Case of a Fully Linear Model . . . 40
3.6.3 Binary Choice Model with an Endogenous Dummy Variable . . . 41
3.6.4 The Two-Stage Predictor Substitution Model . . . 41
3.6.5 Control Functions . . . 42
3.6.6 The Linear Probability Model . . . 43
3.6.7 Maximum Likelihood Estimation of a Bivariate Probit Model . . . 45
3.7 Propensity Score Matching . . . 47
3.7.1 The Theoretical Framework of PSM Estimators . . . 47
3.7.2 The PSM Estimation Results . . . 49
3.8 Final Remarks . . . 50
3.9 Appendix . . . 54
3.9.1 Alternative Probit Model Ensuring Comparability . . . 54

4 Optimal Pollution Management When Discount Rates are Endogenous 55
4.1 Introduction . . . 55
4.2 The Model . . . 58
4.3 The Reference Case of a Constant Discount Rate . . . 60
4.4 Optimal Policies . . . 60
4.4.1 Necessary Optimality Conditions . . . 60
4.4.2 Characteristics . . . 61
4.4.3 Steady State(s) . . . 61
4.4.4 Dynamics . . . 64
4.5 Examples . . . 65
4.5.1 Unique Steady State with High Damages and High Consumption . . . 66
4.5.2 Multiple Steady States . . . 67
4.6 Concluding Remark . . . 68
4.7 Appendix . . . 71
4.7.1 Transversality Conditions in Case of Endogenous Discounting. . . . 71
4.7.2 Sufficiency Conditions . . . 73
4.7.3 Shift of a Unique Steady State due to Discount Parameter b . . . 74
4.7.4 Consequences of a Discount Function Increasing with Pollution . . . 74

5 Summary 76

6 Curriculum Vitae 78

Chapter 1

Introduction

This cumulative PhD thesis consists of three articles. Two articles (1.1 and 1.2) revolve around the question how the quality of information influences the quality of decisions, once from a theoretical point of view, once from the viewpoint of a real-life person facing the decision to switch the electricity supplier or not. The third paper (1.3) covers the question if endogenous discounting is the right method to implement a decision maker's sustainability motive in a well-known stock externality problem such as global warming.

The order in which the papers are presented in the following chapters does not coincide with the chronological order in which they were written. Chronologically, after the theoretical work on the role of information, I delved into the concept of endogenous discounting within the framework of optimal control models. As the methodical pitfalls - overseen by other researchers working with endogenous discounting – became apparent, we presented them in a paper, but we saw no point in proceeding any further in this direction. That is why I returned in my third paper to the impact of different levels of information on the quality of decisions.

The first paper, presented in Chapter 2, deals with the effect of additional information on the quality of decisions by making use of simulation methods. We start from a scenario where a hypothetical decision maker has no information about the probabilities which would be needed to calculate expected utilities. From this starting point, we apply two methods developed by us which allow to constantly and gradually increase the decision maker's information level by systematically limiting the ranges of probabilities. We simulate probabilities lying within the defined ranges and compare the decisions stemming from the thereby gradually increased information to a reference scenario. In this reference scenario the decision maker has complete information about the probabilities of the occurrence of each state and can thus decide in line with expected utility theory. The mayor result of our simulation study is that the returns to scale on information decrease.

The second paper I worked on is presented in Chapter 4. It revolves around the question if endogenous discounting represents a way out of the old dilemma how to discount events occurring only far in the future: The problem with constant, empirically compatible discount rates is that they give too little weight to future generations. A discount rate decreasing with time such as hyperbolic discounting accounts for the mentioned problem, but leads to time-inconsistent optimal policies. A third alternative is endogenous discounting, which links the discount rate to another variable within the model, such as the environmental state, or the wealth level of a decision maker. Strictly speaking, the idea of endogenous discounting is not new, but has recently been rediscovered in papers such as Schumacher (2009, 20011) or Ayong Le Kama and Schubert (2007). The topic of endogenous discounting seemed to be worth to dig deeper into the theoretical concepts and potential fields of application. Is endogenous discounting via wealth the key to understand the real world observation that some societies seem to enjoy constant high wealth and high consumption levels whereas others are trapped with low wealth and low consumption levels, as indicated by Schumacher (2009)? Is endogenous discounting the proper tool to implement a sustainability motive in a model and does such a model set-up always lead to more sustainable outcomes, as claimed by Along Le Kama and Schubert (2007)? So as to answer these questions, we take a standard and simple stock externality problem, namely 'global warming', but switch from a constant to a state-dependent discount rate. Similar to Ayong Le Kama and Schubert, we implement a sustainability motive: the higher the level of pollution, the more patient decision makers become. Comparing the optimal policies to the standard model, the main differences are twofold: First, an endogenous discount rate, albeit always smaller than the exogenous one, can nevertheless imply higher stationary damages than in the standard model. Second, an endogenous discount rate can lead to multiple steady states. The first difference clearly contradicts the initial idea of implementing a sustainability motive. We show that these outcomes are not a special case due to an extraordinary parameter constellation, but are inherent to a model including an endogenously determined discount rate. In our work, we give a technical as well as an intuitively understandable explanation for the mentioned phenomena. After identifying and clarifying this fundamental problem of endogenous discounting, we saw no point in our original plan to apply endogenous discounting to more realistic and thus complexer models. In a complexer model-set up, the additional (and sometimes counterintuitive) influence of the endogenous discount rate on the optimal policies would be harder to identify, but it would still distort the optimal policies.

It was the time when it became apparent that we would not pursue a follow-up paper on endogenous discounting, when Jaqueline Wolf approached Prof. Wirl with an idea for her master thesis: She wanted to examine the reasons for the low switching rates of electricity suppliers of Austrian households. Since she worked at a survey institute, she was able to collect a workable amount of information, but her master thesis consisted mainly of a descriptive analysis of the collected answers and lacked an econometric model. From a conceptional point of view, one of the main questions of the survey was very similar to the research question of my first paper: How does better information influence the quality of a decision? The annual report of the energy think tank VaasaETT reveals that for the typical Austrian household it is the right decision (at least in monetary terms) to switch away from the local by-default contract to the cheapest offer. The average annual savings amounted to almost 100€ or almost 10% of the standard by-default electricity bill in 2012[1]. The level of information of a decision maker is described by the question if the respondent had used a tariff calculator (an online tool to compare electricity suppliers and their offers; it is provided by the Austrian regulatory authority E-control) or a similar tariff comparison tool. I grasped the opportunity and used the collected data myself to build an econometric model, which is able to highlight the factors influencing the decision maker's decision (not) to switch and the dominating role of information among these factors. The paper describing this econometric model is presented in Chapter 3. In the course of the modelling, I came across the question how to handle a (potentially) endogenous dummy variable in a binary choice setting. More specifically, the question came up how to instrument

[1] see VaasaETT 2013, *European Residential Energy Report 2013*.

the explanatory binary variable *tariff calculator* in a non-linear setting such as the estimation of a switching probability with the help of a logit or a probit model. In such a setting, the well known two-state-least-square estimation is not applicable. Surprisingly, this question has barely been treated in the literature before. This is the reason why I included two sections in which I describe different approaches how to handle an endogenous dummy variable in a binary choice setting. In Section 3.6 two instrumental variable approaches are presented, which also work in a non-linear setting. In Section 3.7 I describe a propensity score matching approach. The submitted journal version of the paper (see 1.2) differs from the version in this thesis in the way how the econometric theory is presented: In the journal version, Section 3.6 was shortened and delegated to the Appendix; the propensity score matching approach of Section 3.7 was not included at all. We have included Section 3.7 here as supplementary material.

1.1 Chapter 2: The Effect of Information on the Quality of Decisions (published)

This paper is a joint work with my PhD colleague Christoph Graf. It is published in the *Central European Journal of Operations Research*, Volume 22, Issue 4 (2014), Page 647-662. According to the latest *VHB-JOURQUAL* 2.1 (2011) ranking, the Central European Journal of Operations Research is ranked C.

I contributed to all parts of the paper except for the algorithm to test if a probability vector leads to the same/similar ranking of alternatives as the reference probability vector does (see Section 2.3.2 in Chapter 2).

1.2 Chapter 3: Information as Key Determinant in Switching the Electricity Suppliers for Austrian Households (submitted)

This paper is a joint work with my supervisor Franz Wirl and Jaqueline Wolf, a master student of Franz Wirl. A shortened version of the work presented in Chapter 3 was submitted for publication to the *Journal of Business Economics* on January 26, 2015. According to the latest *VHB-JOURQUAL* 2.1 (2011) ranking, the Journal of Business Economics is ranked B.

The data, on which the econometric model is based, was collected by Jaqueline Wolf in the course of her master thesis. Mrs. Wolf's master thesis consisted only of a descriptive analysis of the data and lacked an econometric model. Apart from the data collection and the descriptive statistics in Section 3.2, all parts of the paper originate from me. I acknowledge close supervision in writing up the results by my supervisor Franz Wirl. Furthermore, I especially acknowledge helpful discussions about the question how to handle a potentially endogenous dummy variable in a binary choice model with Prof. Junsoo Lee.

1.3 Chapter 4: Optimal Pollution Management When Discount Rates are Endogenous (submitted)

This paper is a joint work with my supervisor Franz Wirl. It was submitted for publication to the journal *Resource and Energy Economics* on December 12, 2013. The reviewers as well as the associated editor of the journal considered the scientific contribution of the paper to be potentially significant, but asked for a revision. A revised version of the paper was submitted on December 23, 2014. Resource and Energy Economics does not appear in the latest VHB-JOURQUAL 2.1 (2011) ranking, but has an impact factor of 1.404 according to 2013 Journal Citation Reports by Thomas Reuters (2014).

The paper has evolved in the course of a close cooperation between my supervisor Franz Wirl and me. I acknowledge close supervision and crucial guidance, especially concerning mathematical questions in optimal control theory, by Franz Wirl.

Chapter 2

The Effect of Information on the Quality of Decisions

We study the effect of additional information on the quality of decisions. We define the extreme case of complete information about probabilities as our reference scenario. There, decision makers (DMs) can use expected utility theory to evaluate the best alternative. Starting from the worst case – where DMs have no information at all about probabilities – we find a method of constantly increasing the information by systematically limiting the ranges of the probabilities. In our simulation-based study, we measure the effects of the constant increase in information by using different forms of relative volumes. We define these as the relative volumes of the gradually narrowing areas which lead to the same (or a similar) decision as with the probability in the reference scenario. Thus, the relative volumes account for the quality of information. Combining the quantity and quality of information, we find decreasing returns to scale on information, or in other words, the costs of gathering additional information increase with the level of information. Moreover, we show that more available alternatives influence the decision process negatively. Finally, we analyze the quality of decisions in processes where more states of nature are considered. We find that this degree of complexity in the decision process also has a negative influence on the quality of decisions.

2.1 Introduction

In decision making under uncertainty, decision makers (DMs) are confronted with several alternatives with different outcomes in various states of nature. What they do not know are the precise probabilities of the states of nature. For example, DMs may know only the intervals, in which the true probabilities lie, or they might have information about the rank order of the true probabilities. In order to incorporate the aforementioned kinds of information, we consider the case in which the available information about the true probabilities is represented by linear inequalities. We define the area in the probability space, which consists only of probability-vectors (p-vectors) fulfilling the inequalities, as an *information set*. Therefore, an information set represents the given level of information.

A vast amount of literature deals with the issue of incomplete information on probability distributions, or, conceptually equivalent, with incomplete information on attribute weights in multicriteria decision problems. Our work distinguishes itself from the existing literature in a crucial point. Instead of working with static incomplete information, we develop a method to systematically include additional information. Building on the work of Fishburn (1965), many studies concentrated on the DM's difficulties to find the best alternative in case of incomplete information. One of these static approaches is presented in Charnetski and Soland (1978), who opt to decide for the alternative maximizing a measure called *comparative hypervolume*. Essentially, we avail ourselves of the same measure, but use it under a dynamic setting. By the help of an exogenous mechanism, we analyze the outcomes of decisions when information about the probabilities becomes more precise. Thereby, we can examine the effect of additional information on the quality of decisions, i.e., does better information lead to better decisions? Do we need a disproportional increase in information in order to obtain an increase in decision quality at higher information levels? Moreover, we are interested in the impact of more complex decision processes on the quality of decisions. This can either be the case when DMs face more alternatives to choose from or when a decision under uncertainty consists of more potential outcomes.

The area of sensitivity analysis is based on the theory of uncertain probabilities in decision making. Its key objective is to study the sensitivity of an optimal decision subject to changes in the probabilities of the states of nature. Evans (1984) and Schneller and Sphicas (1985) determined the optimal alternative for a selected p-vector. On the basis of the selected vector, the authors analyzed the maximal distances one can travel in the probability space without a change in the optimal alternative. Furthermore, the authors built confidence spheres with maximal radius around the chosen p-vector. As opposed to the mentioned distance-based approach, the discussion of Starr's domain criteria (cf. Starr, 1962, 1966) in Schneller and Sphicas (1983) is a volume-based analysis. The volume-based approach specifies that in a decision under uncertainty, DMs should decide for the alternative whose area of optimality has the largest volume. The paper also describes how to incorporate beliefs about the probabilities in the form of linear inequalities in Starr's domain criteria.

In the related field of multicriteria decision analysis, weights that specify the DMs' trade-offs between the individual criteria are analyzed instead of probabilities. Dias and Clímaco developed a decision support tool dealing with incomplete information on attribute weights (Dias and Clímaco, 2000, 2005). Eiselt and Laporte (1992) extended Starr's domain criteria and applied it in the weight space of multicriteria decision problems. In-depth and additional research was undertaken in the field of multicriteria decision aid methods (c.f. Bana e Costa, 1986) and in the field of stochastic multicriteria acceptability analysis (c.f. Lahdelma et al., 1998; Lahdelma and Salminen, 2001; Lahdelma et al., 2003; Tervonen and Lahdelma, 2007).

Vetschera (2009) uses the volume of an area in weight space as a measure for the information of DMs in a negotiation process. Information refers to a DM's ability to assess the opponent's preferences, represented by constraints in the weight space. The approach described in Sarabando et al. (2012) resembles Vetschera (2009), but argues from a mediator's view. In a methodological setting similar to ours, Vetschera et al. (2010) studied the impact of the amount of case information on the results of a case-based multiple criteria sorting problem.

Studies working with specific areas in the probability or in the weight space usually choose between two approaches: The distance-based approach (e.g., Evans, 1984; Schneller and Sphicas, 1985) or the the volume-based approach (e.g., Schneller and Sphicas, 1983). In our simulation, we make use of the latter in two senses. First, the measure of the given level of information is defined as the volume of the respective information set. Second, also the measure of the quality of decisions refers to the volume of a specific area in the probability space. Such an area is defined by the property that it consists only of p-vectors inducing the "right" decision. When a DM without knowledge of the precise probabilities ranks a number of alternatives in the same way as if he knew the precise probabilities, we call this a right decision. Later on, we relax this rule and allow also for similar rankings.

The remainder we organize as follows. Section 2.2 explains the concept in detail and deals with certain hypotheses. In Section 2.3, we introduce the measure for the quality of information, the method to systematically generate a sequence of information sets of increasing quality, and a measure for the quality of decisions. Section 2.4 describes the framework of the simulation and Section 2.5 states the results. Finally, we conclude the study and give an outlook onto future research.

2.2 Concept and Hypotheses

2.2.1 Concept

Figure 2.1 illustrates the aforementioned method of increasing information. We always start on the left side, where we have no information about the precise probabilities, besides the usual assumptions[1]. We call the respective information set *starting information set.* On the right side we have the situation of complete information about the precise probabilities. This case is equivalent to the situation of decision making under risk, where DMs can apply expected utility theory to decide for an alternative. The p-vector under complete information is our "reference" p-vector, denoted as *true p-vector* $\mathbf{p}^*$.

Beginning from the starting information set, we gradually improve the information sets, i.e., the DM is given more precise information about $\mathbf{p}^*$ in each step. This leads to a sequence of information sets. The stepwise improvement of information is implemented by gradually limiting the ranges for the probabilities. When looking at the respective information sets, this means that the information sets contract around $\mathbf{p}^*$.

In order to measure the implications of better quality of information, the sequence of information sets has to fulfill an important condition. The relative increase in quality of information from one information set to the next has to be constant for the whole sequence of information sets.

A smaller area of an information set is equivalent to more precise information about the ranges of probabilities. Therefore, we define the measure for the quality of information as the volume of the respective information set.

Our concept to measure the quality of decisions given a certain level of information can be described as follows. First, we have to specify what we mean by making a right decision: a DM makes the right decision given a certain level of information, when he ranks a number of alternatives in the same way as if he knew $\mathbf{p}^*$. This is the reason why we first have to determine

[1] Probabilities are between zero and one and also sum up to one.

the expected utility ranking of alternatives induced by $\mathbf{p}^*$. The given level of information is represented by the respective information set. Therefore, we determine the area within an information set consisting only of p-vectors which lead to the same or a predefined ranking as $\mathbf{p}^*$. The quantitative measurement of the quality of decisions is then defined as the volume of the aforementioned area within an information set, divided by the volume of the respective information set. We call this measure the *relative volume* (rv). The rv can be interpreted as the probability to make the right decision with a given level of information.

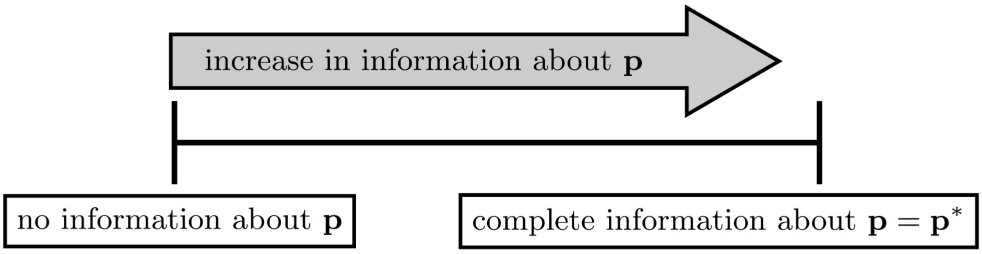


Figure 2.1: A dynamic approach of increasing information

2.2.2 Hypotheses

With the help of four hypotheses, we formulate the main research questions of this paper.

Since the information set's volume appears in the denominator of the rv, a smaller information set leads automatically to a higher rv – unless the nominator, the volume of the area of p-vectors inducing correct choices, does not decrease overproportionally. Our construction of the rv allows for such an overproportional decrease only in exceptional cases, and thus on average, the rv has to increase. Although this increase has its origin in our methodology, we state it as Hypothesis 1, since it serves as the basis for Hypothesis 2.

- **H1:** *The rv rises if the quality of information increases.*
  In average, the quality of decisions increases if the information about $\mathbf{p}^*$ becomes more precise.

- **H2:** *There are decreasing returns to scale on information.*
  With the improvement of information at a constant rate, we also expect the rv to increase, but at a decreasing rate. We argue that on a high level of information the marginal return of information will be smaller than on a lower information level.

- **H3:** *More alternatives induce lower rvs.*
  We postulate that a setup with more alternatives leads to lower rvs. The argument for this is simply combinatorial, i.e., more alternatives lead to more ranking possibilities. Thus, the area in which all p-vectors lead to the same or predefined ranking as $\mathbf{p}^*$ becomes smaller.

- **H4:** *A higher number of states leads to lower rvs.*
  In a more complex decision process, i.e., more states of nature, we conjecture lower rvs.

2.3 Method

In Section 2.3.1 we describe the method of generating a sequence of information sets, where the relative increase in information from one information set to the next is constant. The sequence of information sets represents the increasing quality of information. We make use of the sequences in Section 2.3.2, when we measure the influence of better information on the quality of decisions with the help of the rvs.

2.3.1 The Iterative Improvement of the Information Set

We consider decisions under risk where the outcome is uncertain and alternatives lead to different outcomes in different states of nature. In the following, we consider decision problems with a finite set of alternatives $A = \{A_1, \ldots, A_N\}$ and a finite number of K states. Let p_k be the probability that state k occurs for $k = 1, \ldots, K$, then $\mathbf{p} = (p_1, p_2, \ldots, p_K)$ is the p-vector of all states. We assume that the utilities of the alternatives in different states are known, but we do not know the precise probabilities, p_k.

The information about the probabilities is given in the form of linear inequalities. These inequalities define the information set P of possible p-vectors $\mathbf{p}$. If no other information about probabilities is available, P is at least characterized by the conditions that $\sum_{k=1}^{K} p_k = 1$ and $p_k \geq 0, \forall k$. We call this *information set* starting information set and label it as P_0. Moreover, let $\mathbf{p}^*$ denote the true p-vector. Throughout the paper, $\mathbf{p}^*$ is not known to the DMs. However, we generate values for $\mathbf{p}^*$ so that we have a reference for measuring the improvement of an information set.

The currently available information of a DM is represented through an information set. We start with no available information, i.e., in the information set P_0. Starting from P_0, our method iteratively generates the information sets $P_j, j > 0$. Depending on the number of states K, P_0 corresponds to the standard-$(K-1)$-simplex lying in the affine hyperplane defined by $\left\{(p_1, \ldots, p_K) \in \mathbb{R}^K \,|\, \sum_{k=1}^{K} p_k = 1\right\}$. P_0 also corresponds to the convex hull of its vertices $\{v_1, v_2, \ldots, v_K\}$, where $v_1 = (1, 0, \ldots, 0, 0), v_2 = (0, 1, \ldots, 0, 0), \ldots, v_K = (0, 0, \ldots, 0, 1)$.

Improving the quality of an information set is equivalent to finding linear inequalities, which cut away regions not containing the true p-vector $\mathbf{p}^*$. By iteratively adding new inequalities, we obtain a sequence of improving information sets, $\langle P_j \rangle_{j \geq 0}$. Thus, we know that P_{j+1} is always a part of P_j.

We measure the quality of an information set by its *volume* $V(P_j)$. Consequently, the sequence of volumes of improving information sets, $\langle V(P_j) \rangle_{j \geq 0}$, has to be decreasing, i.e., $V(P_{j+1}) \leq V(P_j)$.

Working towards hypotheses **H1** and **H2**, we develop an algorithm which generates – depending on P_j – in each iteration an information set P_{j+1}, so that the ratio of $V(P_{j+1})$ and $V(P_j)$ equals a constant rate for all $j \geq 0$. An important property of this method is that as j tends to infinity, the resulting information set P_∞ contains only the true parameter $\mathbf{p}^*$.

We use two different, but similar, algorithms which fulfill the above described properties – the *barycentric subdivision* (cf. Munkres, 1984) and a similar method defined by us, which we call the *alternative barycentric subdivision.* For both subdivisions, the steps to generate P_{j+1} out of P_j are the same for all j. Hence, we use P_1 and P_0 in illustrating the algorithms in order to simplify the notation. Let V_K equal the set of vertices $\{v_1, \ldots, v_K\}$ of P_0 as defined before and let $\mathcal{P}(V_K)$ be the powerset of V_K without the empty set. Further, we define the *barycenter*

B of a set of vertices $\{v_1, \ldots, v_m\}$ with $1 \leq m \leq K$ as

$$B(\{v_1, \ldots, v_m\}) = \frac{1}{m} \sum_{i=1}^{m} v_i.$$

The *barycentric subdivision* (bsd) of P_0 is a function defined on the set of vertices V_K, whose output is a set of $K!$ sets of vectors. The vectors of each set correspond to the vertices of a subpart of P_0, all subparts have the same volume, and the union of these subparts is P_0. To give a formal definition of the bsd, we first have to order the powerset $\mathcal{P}(V_K)$ by inclusion. Second, we have to find the $K!$ maximal chains $C_i = \{c_{i1}, \ldots, c_{iK}\}$, in the partial order of $\mathcal{P}(V_K)$. Third, we apply the barycenter function $\mathcal{B}(\cdot)$, defined as $\mathcal{B}(C_i) = \{B(c_{i1}), \ldots, B(c_{iK})\}$, to each maximal chain C_i. Then, the bsd is defined as the following set,

$$\text{bsd}(V_K) = \{\mathcal{B}(C_i), i = 1, \ldots, K!\},$$

(cf. Munkres, 1984). The subpart containing the true p-vector $\mathbf{p}^*$ becomes P_1.

By iteratively applying this procedure, we yield a sequence of information sets $\langle P_j \rangle_{j \geq 0}$ which fulfills that

$$V(P_{j+1}) = \frac{1}{K!} V(P_j). \tag{2.1}$$

The barycentric subdivision has the disadvantage that the volumes of the subparts become very small, even when working only with a small number of states K.

The *alternative barycentric subdivision* (a-bsd) of P_0 yields K subparts of equal volume. Let $B(V_K)$ be the barycenter of P_0 and let W_K^h be the set of vertices V_K without the h^{th} vertex v_h, $1 \leq h \leq K$. Then, the a-bsd is defined as

$$\text{a-bsd}(V_K) = \{\{W_K^h\} \cup \{B(V_K)\}, \quad h = 1, \ldots, K\}.$$

Each of these sets corresponds to the vertices of a subpart of P_0, the union of these subparts yields again P_0. As before, we declare the subpart which contains the true p-vector $\mathbf{p}^*$ as information set P_1.

The iterative application of the a-bsd yields a sequence of information sets $\langle P_j \rangle_{j \geq 0}$ which fulfills that

$$V(P_{j+1}) = \frac{1}{K} V(P_j). \tag{2.2}$$

The a-bsd results in a sequence of information sets that does not contract as regularly as the ones of the bsd. This follows from the property that a-bsd, in contrast to bsd, does not necessarily shrink the largest edge of an information set in each iteration step. For our purposes, however, the a-bsd is more appropriate since the volumes decrease also for a larger number of states only by the factor K^{-1}. The a-bsd implies a further advantage over the bsd when it comes to implementing the idea of information sets. In order to find the vertices of the next information set, the only necessary calculation using a-bsd is the determination of the barycenter of the current vertices. Whereas in case of the bsd, the calculation of the vertices of the next information set entails the systematical intersection of (hyper-)planes, which becomes very cumbersome in higher dimensions. Figure 2.2 shows the difference between the bsd and the a-bsd of P_0 for the case that $K = 3$.

Figure 2.3 shows the sequence of information sets $\langle P_j \rangle_{j=0,\ldots,4}$ depending on a fixed $\mathbf{p}^*$ (shown

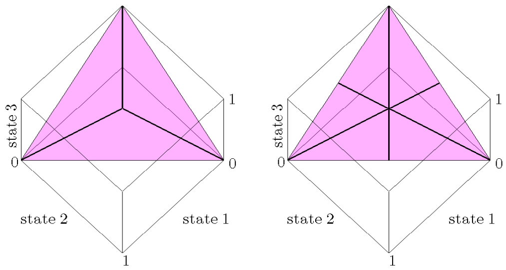


Figure 2.2: Alternative barycentric subdivision (left) versus barycentric subdivision (right) for $K = 3$

as the black point) for $K = 3$ in case of the a-bsd.

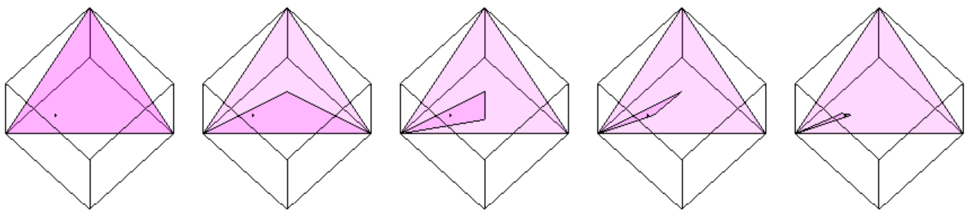


Figure 2.3: A sequence of information sets in case of the alternative barycentric subdivision

2.3.2 The (Averaged) Relative Volume

We can now develop a measure for the quality of decision depending on the quality of information. We take the same true p-vector $\mathbf{p}^*$, which was used to generate a sequence of information sets $\langle P_j \rangle_{j \geq 0}$. We generate a utility matrix for a set of N alternatives and K states of nature. For each alternative A_i, $i = 1, \ldots, N$, $\mathbf{p}^*$ is used to calculate the expected utility value, $\mathbb{E}_{p^*}[A_i]$. Arranging the expected utilities $\mathbb{E}_{p^*}[A_i]$ according to their value gives the "true" ranking of the alternatives.

So as to compare the actual ranking with the true ranking we distinguish between two methods. First, we consider only p-vectors in an information set which lead to exactly the same ranking as induced by $\mathbf{p}^*$. Second, we suggest a weaker matching of rankings. We pay attention only to the top-ranked alternatives induced by $\mathbf{p}^*$. In particular, we want an identical ranking for these top-level alternatives, while the remaining alternatives might be ranked differently. We call the thereby resulting measures rv^{exact} and $\text{rv}^{\text{top}T}$, where T specifies the number of alternatives in the top-level. We will discuss the exact definitions of the measures below.

We denote by M_j either an area which consists only of p-vectors inducing the same ranking as $\mathbf{p}^*$ or an area consisting only of p-vectors which lead at the top-level to the same ranking of the alternatives as $\mathbf{p}^*$ does. The rv of M_j is given by

$$rv(M_j) = \frac{V(M_j)}{V(P_j)}, \tag{2.3}$$

and can thus be interpreted as the probability of making the right decision. The rv depends on the specific position of $\mathbf{p}^*$ and on the specific values of the generated utility matrix. That is why we perform several runs with a different $\mathbf{p}^*$ and a different utility matrix in each run, and average over the rv of each run, so as to wipe out this effect.

Making use of Equation 2.1 in case of the bsd and Equation 2.2 in case of the a-bsd, we know the volume $V(P_j)$ of the j-th information set.

As described above, we work with two different methods to compare the induced rankings of alternatives. The procedure to calculate the volume of M_j, described below, is similar with both methods. The difference, arising in the top-ranked matching method, is that only the first T alternatives in the ranking are considered.

The area M_j is defined by several inequalities. First, the K inequalities defining the information set P_j are also valid for M_j. Second, the requirement that all the p-vectors in M_j have to induce the same ranking as $\mathbf{p}^*$, can also be expressed with the help of inequalities.

We first address the case of the exact matching. Let A'_i be the alternative on the i^{th} place in the ranking induced by $\mathbf{p}^*$, $i = 1, \ldots, N$. More formal, this means that $\mathbb{E}_{\mathbf{p}^*}[A'_1] \geq \mathbb{E}_{\mathbf{p}^*}[A'_2] \geq \cdots \geq \mathbb{E}_{\mathbf{p}^*}[A'_N]$. Then, $a'_{i,k}$ is the value of the alternative ranked on the i^{th} place, in state k, $k = 1, \ldots, K$. Thus, all p-vectors element of M_j fulfill

$$\sum_{k=1}^{K} a'_{i,k} p_k \geq \sum_{k=1}^{K} a'_{i+1,k} p_k, \qquad \forall i = 1, \ldots, N-1.$$

Plugging in $1 - \sum_{k=1}^{K-1} p_k$ for p_K and rearranging the inequalities, gives for all $i = 1, \ldots, N-1$

$$\sum_{k=1}^{K-1} (a'_{i,k} - a'_{i+1,k} - a'_{i,K} + a'_{i+1,K}) p_k \geq a'_{i+1,K} - a'_{i,K}.$$

Together with the inequalities $p_k \geq 0, \forall k = 1, \ldots, K-1$, we obtain a system of $(K) + (N - 1) + (K - 1) = N + 2K - 2$ inequalities defining the area M_j within P_j in dimension $K - 1$.

In case of the top-level matching, only the first T alternatives in the ranking have to be equal to the true ranking. The ranking of the alternatives on rank $T+1, \ldots, N$ need not be equal to the true ranking. Formally, this translates to the inequalities $\mathbb{E}_{\mathbf{p}^*}[A'_1] \geq \mathbb{E}_{\mathbf{p}^*}[A'_2] \geq \cdots \geq \mathbb{E}_{\mathbf{p}^*}[A_T]$. Moreover, an implied requirement is that $\mathbb{E}_{\mathbf{p}^*}[A'_{T+i}] \leq \mathbb{E}_{\mathbf{p}^*}[A'_T]$, $i = 1, \ldots, N-T$, which ensures that the expected values of the alternatives on ranks $T+1, \ldots, N$ are actually smaller than the ones of the top-ranked.

2.4 Simulation

In this section we examine our hypotheses defined in Section 2.2.2 using simulation–based methods. To calculate the volume of M_j from the inequalities we used the Matlab-package *multi-parametric toolbox (mpt)*, by Kvasnica et al. (2004). We summarize the framework of the simulation in Figure 2.4 and subsequently explain the simulation in more detail.

2.4.1 One Simulation Run

We describe each step of a simulation run in the following.

1. Generate a utility matrix for a set of N alternatives and K states of nature. The entries

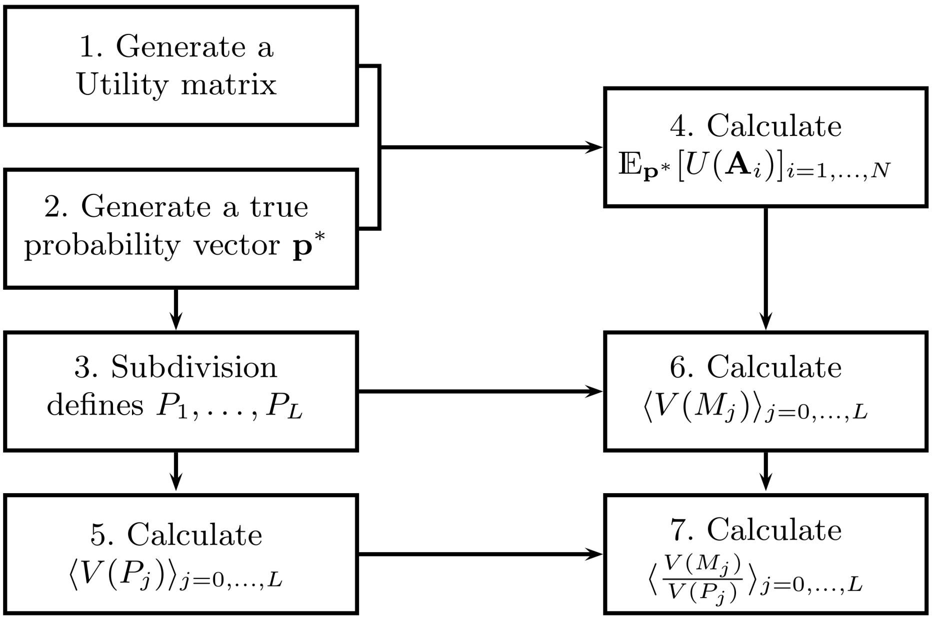


Figure 2.4: The simulation framework

of the utility matrix are randomly drawn and uniformly distributed on $[0, 1]$.

2. Generate a fixed true p-vector $\mathbf{p}^*$. We use the method developed by Rubinstein (1982) to ensure that $\mathbf{p}^*$ is randomly drawn and uniformly distributed on the simplex. There, an exponentially distributed p-vector with rate one is generated and then standardized to sum up to one.

3. Use an appropriate way of subdividing an information set, i.e., bsd or a-bsd. Depending on the true p-vector $\mathbf{p}^*$, generate, using one of these methods, the sequence of information sets $\langle P_j \rangle_{j=0,\ldots,L}$, where L is an arbitrarily chosen parameter.

4. Calculate the expected utility of each alternative using the true p-vector $\mathbf{p}^*$. Rank the expected utilities of alternatives induced by $\mathbf{p}^*$.

5. Calculate the volumes of each information set P_j, $j = 0, \ldots, L$ (independent of $\mathbf{p}^*$).

6. Calculate the volumes of each area M_j, within P_j, which consists only of p-vectors leading to the same (top–) ranking as $\mathbf{p}^*$.

7. Calculate the sequence of rvs $\langle V(M_j)/V(P_j)\rangle_{j=0,\ldots,L}$.

Performing the simulation run several times, R, and averaging over each rv results in a sequence of *averaged relative volumes* (arv) $\langle \sum_{r=1}^{R} rv_r(M_j)/R \rangle_{j=0,\ldots,L}$.

2.4.2 Parameter Settings

Table 2.1 summarizes the parameter settings. The maximum number of six states and five information sets is given by limitations of the simulation engine. More precisely, in the case of $K > 6$ or $L > 5$, the angle between the bounding hyperplanes of an information set can become so small that the volume of the information set is almost zero and cannot be computed anymore by *mpt*. The number of alternatives (N) we use in our simulation ranges from three to nine, with six alternatives in the standard case. This choice is justified by the following arguments. In case of very few alternatives, the probability is high that alternatives are ranked in the "true" order by chance and not due to the given information. On the other hand, if the number of alternatives is very large, it becomes very unlikely to obtain an exact matching, even with good information at hand. Thus, our chosen numbers of alternatives represent a balance between a low and a high number of alternatives.

In order to remove the effect of the generated $\mathbf{p}^*$ and the generated utility matrix, we repeat the simulation run described above 10,000 times with different utility matrices and different true p-vectors in each run. To test whether the repetitions were sufficient and to ensure thereby that our results are robust, we run the whole procedure again. No statistically significant differences were found between these runs. Thus, we conclude to have stable results.

Parameters	Standard	Other values
# of states (K)	3/4/5/6	–
# of alternatives (N)	6	3/9
# of information sets (L) (incl. P_0)	5	–
# of examined top-ranked alternatives (T)	3 (for $N = 6$)	–
Mode of subdivision	a-bsd	bsd (for $K = 3$)

Table 2.1: Summary of the simulation parameters

2.5 Results

Figure 2.5 shows the results for $K = 3$ and all other parameters chosen according to the standard parameter settings listed in Table 2.1. Note that on the x-axis we plot the increase of information compared to the information set P_0. In the case of a-bsd as mode of subdivision this yields a stepwise improvement by the factor $V(P_j)(V(P_{j+1}))^{-1} = K$. Using a logarithmic scale to base K on the x-axis, gives the number j of the corresponding information set, $j = 0, \ldots, 4$.

Table 2.2 shows the results of the arv$^{\text{exact}}$ and the arv$^{\text{top3}}$ for the standard parameter settings. In the following we discuss the attributes of the rvs referring to our four hypotheses.

K	3	4	5	6	3	4	5	6
	arv$^{\text{exact}}$				arv$^{\text{top3}}$			
P_0	0.1039	0.0617	0.0449	0.0349	0.1875	0.1344	0.1115	0.0936
P_1	0.1985	0.1118	0.0752	0.0547	0.3117	0.2092	0.1611	0.1287
P_2	0.3117	0.1786	0.1160	0.0815	0.4405	0.2952	0.2217	0.1708
P_3	0.4310	0.2588	0.1682	0.1150	0.5562	0.3868	0.2872	0.2185
P_4	0.5416	0.3425	0.2280	0.1554	0.6572	0.4763	0.3550	0.2708

Table 2.2: Values for arv$^{\text{exact}}$ and arv$^{\text{top3}}$, standard parameter setting

We also analyzed the rvs using the bsd as mode of subdivision. However, a direct comparison of the results using a-bsd and bsd as mode of subdivision is not possible. In case of the bsd, the quality of information increases by the factor $(K!)^j$, compared to an increase by K^j in case

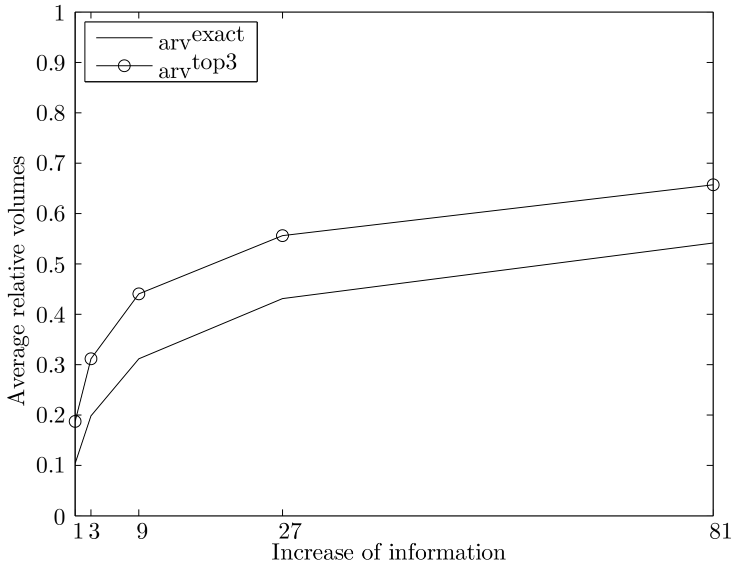


Figure 2.5: Comparison of $\mathrm{arv}^{\mathrm{exact}}$ and $\mathrm{arv}^{\mathrm{top3}}$ for $K = 3$, all other parameters chosen according to the standard setting

of the a-bsd. Hence, we cannot test whether one mode of subdivision increases the quality of decisions faster than the other.

2.5.1 Hypothesis H1

Hypothesis **H1** postulates an increase of the rvs in additional information. Table 2.2 shows that the $\mathrm{arv}^{\mathrm{exact}}$ and $\mathrm{arv}^{\mathrm{top3}}$ increase with the number of information sets. In Figure 2.5, we see for $K = 3$ that an increase in information lets the arvs rise. In order to test hypothesis **H1** we calculate the slopes of our piecewise linear functions and test whether the slopes are greater than zero or not. In particular, we perform one-sided t-tests. The null hypothesis states that the slopes are a random sample from a normal distribution with mean below or equal zero and unknown variance, whereas the alternative hypothesis states that the mean is larger than zero. Table 2.3 shows highly significant results, meaning that the null hypothesis can be rejected and therefore that the rvs increase with better information using the standard parameter setting. Hence, we confirm this hypothesis.

2.5.2 Hypothesis H2

Hypothesis **H2** concerns the curvature of $\mathrm{rv}^{\mathrm{exact}}$ and $\mathrm{rv}^{\mathrm{top3}}$. Taking into account the improvement of information we expect the indices to increase at diminishing rates (Figure 2.5 shows this for the average values). To test the curves' shapes we compute first differences of the slopes of the piecewise linear functions. Moreover, we test whether the first differences are on average greater or equal than zero against the alternative hypothesis that they are smaller than zero. The results are stated in Table 2.4. In all cases we can reject the null hypothesis. This means $\mathrm{rv}^{\mathrm{exact}}$ and $\mathrm{rv}^{\mathrm{top3}}$ have decreasing slopes. Therefore, hypothesis **H2**, which states that there are decreasing returns to scale on information, is confirmed.

	K	3	4	5	6	3	4	5	6
		$\mathrm{rv}^{\mathrm{exact}}$				$\mathrm{rv}^{\mathrm{top3}}$			
	Mean	0.0473	0.0167	0.0076	0.0040	0.0621	0.0250	0.0124	0.0070
s_1	t	93.5212	75.0463	65.1739	56.1527	90.227	83.9358	61.7798	54.3238
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	0.0189	0.0056	0.0020	0.0009	0.0215	0.0072	0.0030	0.0014
s_2	t	84.3525	70.4481	61.9935	55.4899	78.0223	66.6864	61.1373	54.1039
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	0.0066	0.0017	0.0005	0.0002	0.0064	0.0019	0.0007	0.0003
s_3	t	75.9681	66.6315	59.2062	53.3669	65.2112	62.2338	55.9185	51.1556
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	0.0020	0.0004	0.0001	0.0000	0.0019	0.0005	0.0001	0.0000
s_4	t	63.2037	61.9142	56.2533	51.5506	56.4932	56.5767	51.8509	48.9099
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

$s_p, \forall p = 1, \ldots, 4$: Slope of the rv between the steps of information

Table 2.3: One-sided t-tests of the slopes, standard parameter setting

	K	3	4	5	6	3	4	5	6
		$\mathrm{rv}^{\mathrm{exact}}$				$\mathrm{rv}^{\mathrm{top3}}$			
	Mean	−0.0284	−0.0111	−0.0055	−0.0031	−0.0407	−0.0178	−0.0094	−0.0056
$s_2 - s_1$	t	−53.1449	−49.5250	−48.2661	−44.0482	−53.4511	−50.0147	−45.9517	−43.0954
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	−0.0122	−0.0039	−0.0015	−0.0007	−0.0150	−0.0053	−0.0024	−0.0011
$s_3 - s_2$	t	−49.8279	−47.4183	−45.8938	−44.1966	−49.1579	−46.2319	−46.4913	−43.2306
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean	−0.0046	−0.0012	−0.0004	−0.0001	−0.0046	−0.0014	−0.0005	−0.0002
$s_4 - s_3$	t	−47.0844	−46.8965	−44.3933	−42.5289	−41.7030	−44.1477	−42.9052	−41.1170
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

$s_p, \forall p = 1, \ldots, 4$: Slope of the rv between the steps of information

Table 2.4: One-sided t-tests of the first differences of the slopes, standard parameter setting

2.5.3 Hypothesis H3

Table 2.5 shows the values of the $\mathrm{arv}^{\mathrm{exact}}$, running the simulation with different amounts of alternatives. We only concentrate on the $\mathrm{rv}^{\mathrm{exact}}$ since the $\mathrm{rv}^{\mathrm{top3}}$ is meaningless for the case $N = 3$. In order to show that the $\mathrm{rv}^{\mathrm{exact}}$ decreases when we consider more available alternatives, we perform one-sided t-tests. In particular, we compare the $\mathrm{rv}^{\mathrm{exact}}$ using 3 (6) alternatives versus 6 (9) alternatives. The results displayed in Table 2.6 are highly significant, meaning that the null hypothesis (the $\mathrm{rv}^{\mathrm{exact}}$ using 3 (6) alternatives is lower or equal than the averages using 6 (9) alternatives) can be rejected. Hence, we also confirm hypothesis **H3**.

K	3	4	5	6	3	4	5	6
	$\mathrm{arv}^{\mathrm{exact}}(N = 3)$				$\mathrm{arv}^{\mathrm{exact}}(N = 9)$			
P_0	0.4929	0.4521	0.4208	0.4014	0.0255	0.0095	0.0045	0.0027
P_1	0.6247	0.5442	0.4954	0.4569	0.0611	0.0232	0.0104	0.0058
P_2	0.7267	0.6251	0.5609	0.5110	0.1206	0.0483	0.0217	0.0115
P_3	0.8029	0.6965	0.6235	0.5624	0.2042	0.0880	0.0398	0.0210
P_4	0.8542	0.7554	0.6766	0.6105	0.3027	0.1403	0.0670	0.0354

Table 2.5: Results of $\mathrm{arv}^{\mathrm{exact}}$ for $N = 3$, and $N = 9$, all other parameters chosen according to the standard setting

	K	3	4	5	6	3	4	5	6
		$\text{rv}^{\text{exact}}(N=3) - \text{rv}^{\text{exact}}(N=6)$				$\text{rv}^{\text{exact}}(N=6) - \text{rv}^{\text{exact}}(N=9)$			
P_0	Mean	0.3890	0.3904	0.3759	0.3665	0.0784	0.0522	0.0404	0.0322
	t	139.9196	152.1096	154.1671	158.1796	81.0498	81.9203	80.0287	80.2076
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P_1	Mean	0.4262	0.4324	0.4202	0.4022	0.1374	0.0885	0.0648	0.0489
	t	129.9675	145.7406	151.4024	153.3706	78.7104	79.9712	79.9128	77.5489
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P_2	Mean	0.4151	0.4466	0.4449	0.4295	0.1911	0.1303	0.0943	0.0700
	t	115.0061	135.7107	145.5064	150.3952	75.1169	77.449	78.8527	76.8143
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P_3	Mean	0.3718	0.4377	0.4553	0.4474	0.2268	0.1708	0.1284	0.0940
	t	98.5106	122.3489	137.7429	145.8184	70.2775	73.4865	77.1057	76.1901
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
P_4	Mean	0.3126	0.4128	0.4486	0.4551	0.2389	0.2023	0.1609	0.1200
	t	81.7497	109.9185	127.2424	139.1587	63.4219	69.7484	74.3125	74.5458
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2.6: One-sided t-tests of the differences of results between different amounts of alternatives. Comparison only on rv^{exact}, all other parameters chosen according to the standard setting

2.5.4 Hypothesis H4

Our last hypothesis deals with the "role of dimension". We conjecture that more states of nature lead to lower rvs. The average values of the rvs represented in Table 2.2 suggest that this conjecture may be true. In order to clarify hypothesis **H4** – a higher number of states leads to lower rvs – we perform one-sided t-tests on the differences of results between different states of nature. Note that we can only compare the rvs of the information set P_0 because here the volumes of P_0 are equal for all dimensions. Table 2.7 reports these results. The t-tests confirm that the differences of the rvs between different states of nature decrease. Thus, we conclude that the simulation shows a "role of dimension" for the analyzed cases.

	K	$4-3$	$5-4$	$6-5$	$7-6$	$4-3$	$5-4$	$6-5$	$7-6$
		rv^{exact}				rv^{top3}			
P_0	Mean	-0.0422	-0.0168	-0.0100	-0.0049	-0.0531	-0.0229	-0.0179	-0.0081
	t	-37.6538	-21.0266	-15.5985	-9.1490	-28.0639	-14.5345	-13.0242	-6.6122
	p	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 2.7: One-sided t-tests of the differences of results between different states of nature. Comparison only on P_0, all other parameters chosen according to the standard setting

2.6 Conclusion and Topics for Future Research

In our simulation-based study we found an algorithm to limit probability ranges systematically. We used this method as input for studying the relation between additional information and the quality of decisions. We started in an information set with no information on the potential probability ranges of a predefined reference p-vector. Next, we applied our algorithm to limit these probability ranges, i.e., to increase information. We analyzed the effect of this additional information on the quality of decisions.

With this framework in mind we analyzed various factors influencing the quality of decisions. The results were the following. First, the quality of decisions increased when the information about the reference p-vector increased. Second, the quality of information increased but at diminishing rates. This means, that at higher information levels additional information is valued less than at lower information levels. Third, more complex decisions in form of more available alternatives for the DM lead to a lower quality of decisions. Fourth, we observed a certain “role of dimension”, meaning that more challenging decision processes, including more possible outcomes (higher number of states of nature), lead to a lower decision quality.

There are many possibilities to extend our work, or find an alternative setup to examine a similar problem. First, there may be other appropriate ways to limit the probability ranges in order to increase information systematically. One could use the complexer *barycentric subdivision* of a simplex instead of the *alternative barycentric subdivision* used by us. One could also think of an iterative bisection of the information set, where hyperplanes divide the information set into two parts of equal moment. Second, there are alternative measures for the quality of information. Instead of a volume-based approach, a distance-based approach comes to mind. For example, the radius of the largest (hyper–)sphere within an information set with the reference p-vector as center would also serve as an appropriate measure. And third, other measures of the quality of decision are possible. There exist many ways to compare the ranking of alternatives. For example, instead of working with an exact or a top3 ranking, one could take a predefined distance between rankings into consideration.

Acknowledgments

The authors would like to thank Rudolf Vetschera, David Wozabal, and four anonymous referees for their valuable comments.

Bibliography

Bana e Costa, C. A., 1986. A multicriteria decision aid methodology to deal with conflicting situations on the weights. European Journal of Operational Research 26 (1), 22–34.

Charnetski, J., Soland, R., 1978. Multiple-attribute decision making with partial information: The comparative hypervolume criterion. Naval Research Logistics Quarterly 25 (1), 279–288.

Dias, L. C., Clímaco, J. N., 2000. Additive aggregation with variable interdependent parameters: The vip analysis software. The Journal of the Operational Research Society 51 (9), 1070–1082.

Dias, L. C., Clímaco, J. N., 2005. Dealing with imprecise information in group multicriteria decisions: a methodology and a GDSS architecture. European Journal of Operational Research 160 (2), 291–307.

Eiselt, H. A., Laporte, G., 1992. The use of domains in multicriteria decision making. European Journal of Operations Research 61 (3), 292–298.

Evans, J. R., 1984. Sensitivity analysis in decision theory. Decision Sciences 15 (2), 239–247.

Fishburn, P. C., 1965. Analysis of decision with incomplete knowledge of probabilities. Operations Research 13 (2), 217–237.

Kvasnica, M., Grieder, P., Baotić, M., 2004. Multi-Parametric Toolbox (MPT).
URL `http://control.ee.ethz.ch/~mpt/`

Lahdelma, R., Hokkanen, J., Salminen, P., 1998. SMAA - Stochastic multiobjective acceptability analysis. European Journal of Operational Research 106 (1), 137–143.

Lahdelma, R., Miettinen, K., Salminen, P., 2003. Ordinal criteria in stochastic multicriteria acceptability analysis – SMAA. European Journal of Operational Research 147 (1), 117–127.

Lahdelma, R., Salminen, P., 2001. SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making. Operations Research 49 (3), 444–454.

Munkres, J., 1984. Elements of Algebraic Topology. Addison-Wesley.

Rubinstein, R. Y., 1982. Generating random vectors uniformly distributed inside and on the surface of different regions. European Journal of Operational Research 10 (2), 205–209.

Sarabando, P., Dias, L., Vetschera, R., 2012. Mediation with incomplete information: Approaches to suggest potential agreements. Group Decision and Negotiation 1, 1–37.

Schneller, G. O., Sphicas, G. P., 1983. Decision making under uncertainty: Starr's domain criterion. Theory and Decision 15, 321–336.

Schneller, G. O., Sphicas, G. P., 1985. Notes and Communications on sensitivity analysis in decision theory. Decision Sciences 16 (4), 399–409.

Starr, M. K., 1962. Product Design and Decision Theory. Prentice Hall.

Starr, M. K., 1966. A Discussion of Some Normative Criteria for Decision-Making Under Uncertainty. Industrial Management Review 8, 71–78.

Tervonen, T., Lahdelma, R., April 2007. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research 178 (2), 500–513.

Vetschera, R., 2009. Learning about preferences in electronic negotiations – a volume-based measurement method. European Journal of Operational Research 194 (1), 452–463.

Vetschera, R., Chen, Y., Hipel, K., Kilgour, D., 2010. Robustness and information levels in case-based multiple criteria sorting. European Journal of Operational Research 202 (3), 841–852.

Chapter 3

Information as Key Determinant in the Decision to Switch Electricity Suppliers of Austrian Households

Including an Econometric Analysis how to Handle an Endogenous Dummy Variable in a Binary Choice Model

In this paper, we investigate the drivers and obstacles to switch the electricity supplier for Austrian households. Since the liberalization of the market in 2001, the annual switching rates have remained at surprisingly low levels, leading to a loss of significant savings potential for Austrian households. The results of our exploratory study suggest that the knowledge about tariffs and electricity suppliers is the driving force for the decision to switch. The lack of information might also explain the low incentive effect of savings potentials. Thus, we suggest that for price competitive electricity suppliers, it might be more effective to familiarize customers with tariff structures and savings potentials than to advertise with "green electricity" as it is the common strategy at the moment.

Since the question occurs how to handle a potentially endogenous dummy variable in a binary choice model – a question barely treated in the literature – we include two sections covering this econometric problem.

3.1 Introduction

The Austrian energy market was completely liberalized in the year 2001 after the partial liberalization in 1999 just met the EU common market mandate. Before, the Austrian electricity business was divided between one nationwide organization (*Verbund*), nine provincial provider companies, five provincial capital companies and a range of other companies (local suppliers and power companies). Essentially, all of these companies were publicly owned (cf. Haberfellner (2002), Hofbauer (2006)) and held a (local) monopoly in their respective region.

Since liberalization, electricity supply is open to competition, whereas the areas of transmission and distribution of electricity remain regulated (Hofbauer (2006)). A report from 2002 by the Austrian regulatory authority E-control shows, that they expected a sharp rise of switching rates after the first phase of liberalization (Haberfellner (2002)). However, these expectations have not been met. Especially Austrian households tend to stay with the incumbent suppliers, a behavior that has persisted now for more than a decade. Switching rates among households for the years 2007–2012 range between 1.1% and 1.7%, with a slight upward trend in 2013 (E-Control Austria (2012), E-Control Austria (2013)).

A comparison to similar liberalized European electricity retail markets such as Germany or the Netherlands shows that their switching rates are several times higher than the ones in Austria[1]. For example, in the year 2011, the switching rate in Germany amounted to 7.8%, in the Netherlands to 9.7%. In a market such as Great Britain, where the liberalization process started earlier, the annual switching rate in 2011 reached 15.45% (Council of European Energy Regulators (2012)).

The annual report of the energy think-tank VaasaETT reveals that through this passive behavior, the Austrian households let a significant savings potential slip away: the average annual savings achievable by the typical Austrian household who switched away from their local by-default contract to the cheapest offer available amounted almost 100€ or almost 10% of the standard by-default electricity bill in 2012 (VaasaETT (2013)).

From a theoretical standpoint, two main concepts – one from the field of economics, one from the field of marketing - explain the low switching rates, (cf. Wieringa and Verhoef (2007)). Economic scholars tend to approach switching behavior from a switching costs perspective. Klemperer (1995) lists the following causes which can provoke switching costs: (a) need of compatibility with existing equipment, (b) searching and enforcement costs (transaction costs), (c) costs of learning to use the new brand, (d) uncertainty about the quality of the untested brand, (e) economic brand loyalty generated e.g. by loyalty programs. In the field of marketing, the related concept of inertia prevails. Whereas the theory about switching costs presumes a rational decision maker, who can weigh the benefits of switching against its costs, inertia is not assumed to be an explicit decision, but a habitual buying behavior. In a further approach, Wieringa and Verhoef (2007) focus on the characteristics of a liberalizing market, such as the electricity retail market: (a) customers are only familiar with the former monopolist and do not know the new market entrants, (b) customers are unaccustomed to the switching process, (c) the long-term relationships with the former monopolists promote commitment and trust, (d) a lack of experience in the market may prevent customers to assess a fair price. All of these characteristics can foster switching costs and inertia.

Several papers consider the explicit decision of households (not) to switch their electricity supplier. Similar to our exploratory study, Ek and Söderholm (2008) as well as Wieringa and Verhoef (2007) use an econometric framework to explain inter-household differences in supplier switching behavior. The work by Ek and Söderholm (2008) considers the Swedish electricity market. The authors identify consumers with higher incomes and higher electricity

[1] To enable an international comparison, we have to consider combined switching rates for households and small industry (including the commercial and the agriculture sector). Generally, these combined rates slightly exceed the households' rate alone, (cf. E-Control Austria (2013))

costs as those more likely to switch. The same holds true for consumers believing that the high prices reflect a lack of competition. Further findings show that difficulties in understanding the electricity bill as well as the benefits of switching decreases the probability to switch. This can be interpreted as an indicator that switching costs and inertia prevent consumers in the electricity market from switching. For the Dutch retail electricity market, the econometric work by Wieringa and Verhoef (2007) identifies relationship quality as well as perceived switching costs as the most important determinants of switching behavior.

In an empirical study, Wilson and Waddams Price (2010) examine the switching decisions of UK electricity consumers focusing on their ability to choose the best alternative supplier. Even if only those consumers are considered who stated that the price saving potential was their only motive to switch, the authors find that only 8–12% among them switched to the supplier offering the highest surplus. Furthermore, 17–32% among them appear to have lost surplus through their choice of suppliers. Whereas the former observation is in line with search cost literature, the latter contradicts the principles of rational decisons. Annala et al. (2013), studying the switching behavior of Finish electricity customers, reveals a clear discrepancy between the number of customers who would have benefited from switching and who actually switched supplier. The authors list search costs, bounded rationality, prospect theory and risk aversion as possible explanations for this discrepancy.

This paper investigates drivers and obstacles of changing the electricity supplier for Austrian households. Based on survey data, we try to find an answer to two questions with the help of an econometric model: First, we want to find the determinants of an actual switch of the supplier in the past and second, we investigate the determinants of current intentions to switch. There are a number of interesting questions. In particular, we are interested in how information (and acquisition thereof) affects switching behavior (actual in the past and intended in the future). The Austrian regulatory authority E-Control provides an online tool called *tariff calculator* to compare electricity tariffs and calculate saving potential[2]. Since our data includes a variable measuring if an online price comparison tool was used, it allows to draw conclusions on the impact of information. During the last years, several energy suppliers have entered the market who offer green electricity, some exclusively. These green suppliers often try to attract customers through their environmental responsibility. This raises the question whether and how far customers with a green attitude have switched their electricity supplier compared to less green-minded customers. We further examine if customers who have already undergone the switching procedure in the past, are those more likely to consider switching at the moment. Furthermore, we investigate the seemingly obvious linkage between (subjective and reported) price-sensitivity and switching. The low switching rates contrast those of mobile phone operators, which Austrians do change regularly. This difference in behavior is puzzling, given the similarities of a network service, the presumably even larger expenditures for electricity and potential savings from leaving the incumbent. Other questions concern the contribution of typical socio-economic factors on the decision to change one's electricity supplier.

The paper is organized as follows. In Section 3.2 we introduce the data and present some descriptive statistics. In Section 3.3 we describe the general structure of binary response models. The variables of our model are the topic of Section 3.4. We present and discuss the estimation

[2]Recently, a range of private internet-based tariff comparison sites have popped up offering consumers advice on choosing between suppliers. The web survey broadly asked if the E-control's tariff calculator or any comparable tool was used. In this paper, we subsume the consultation of any of these sites under "usage of a tariff calculator".

results in Section 3.5. In Section 3.6, we present methods how to deal with the potential endogeneity of the explanatory variable *tariff calculator*. Propensity score matching – an alternative to model the causal effect of information gain through the tariff calculator on the switching probability – is introduced in Section 3.7. We conclude the paper with final remarks in Section 3.8.

3.2 Data and Descriptive Statistics

The data set was obtained from a web survey conducted in July and August 2012. The target group comprised 2500 people, listed in the database of a Viennese market research enterprise. The recipients were contacted via e-mail and were asked to complete a questionnaire attached by a hyper link concerning the following questions: knowledge of the Austrian electricity market, attitudes toward sustainability, their effort to search and choose the supplier with the lowest price, expenditures on electricity and socio-demographic background. Furthermore, the recipients were asked if they had switched their electricity supplier (ES) in the last ten years, if they consider churning at the moment, and if they had switched their cell phone provider during the last five years. The questionnaire explicitly asked the respondents not to report a switch of the ES if it happened due to a relocation. 358 recipients returned the questionnaire, out of which 302 could be used, which is a usual and thus suitable return rate of 12.1%. Of these respondents, 29.4% stated that they had switched their ES during the last ten years[3], compared with 49.3% who had switched their cell phone provider during the last five years; 24.2% indicated their intention to churn the ES.

Comparing the percentage of households in our sample who have switched since liberalization (29.4%) with the aggregate churn rate until 2012 in Austria (almost 15%, E-Control Austria (2013)), self-selection is immanent. It seems reasonable to expect that people with a lack of interest in electricity market issues and/or limited or non-existent experience in this respect are over-represented among the non-respondents[4]. However, the results about the drivers of an active switching behavior among the respondents may still provide information for policy makers and electricity companies how to stimulate interest in switching among the whole customer base.

Table 3.1 lists reasons of respondents for churning (staying with their incumbent supplier respectively) during the past ten years, as well as reasons why the respondents consider switching at the moment. By far the most important motive to (consider to) switch is the price, but also sustainability motives are an issue. The quality of the service seems to play only a minor role, which makes sense as the quality does not depend on the particular supplier but on the maintenance and operation of the grid. Those respondents who did not (do not consider to) switch are either content with their current supplier, or consider switching as too costly or as too cumbersome.

The respondents were also asked how their electricity supply contracts were established. Only 25.8% answered that they had actively searched for a supplier, 13.6% did not remember, but the majority (60.6%) simply took over the ES from the previous tenant. This highlights that markets, which are fully liberalized after decades of regulation and no choice (regional

[3]The dependent variable *switched* concerns decision taking over a ten year period, whereas the explanatory variables describe current characteristics of households. This should be kept in mind when drawing conclusions based on the estimated cause-effect relationships.

[4]Ek and Söderholm (2008) report the same self-selection bias.

Table 3.1: Reasons for (not) having switched in the past and (not) considering to switch at the moment respectively (%)

	switched	considering to switch
Cost-saving opportunity	85.4	82.2
Environmental protection	41.6	41.2
Better service	5.6	9.6
Other	5.2	1.4
	not switched	*not* considering to switch
Satisfaction with current supplier	41.8	60.3
Switching costs are too high	31.5	22.7
No time	26.3	14.0
Potential hidden extra costs	24.0	17.0
Awaiting developm. in energy prices	14.1	15.7
No interest in topic	12.2	10.5
Several Bills	7.0	4.3
I was advised against switching	2.8	2.6
Other	10.3	6.6

Note: Multiple responses possible

monopolies), are subject to inertia for quite some time as customers (maybe except for the young) are unaccustomed to shop around.

3.3 Estimating Binary Response Models

We examine two decisions of households: if they have switched their electricity supplier in the last ten years (since the liberalization), and if they are thinking about switching at the moment. Approximately 29% of the respondants have switched their supplier in the past, whereas approximately 24% of the respondants indicate that they are thinking about switching at the moment.

Since the outcome variable in the switching model, *switched* (= 1 if respondent has switched, = 0 otherwise), and the outcome variable in the consideration model, *consideration* (= 1 if respondent considers to switch, = 0 otherwise), are both binary, we study binary response models of the form

$$P(y = 1|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta}) \equiv p(\mathbf{x}), \tag{3.1}$$

where $p(\mathbf{x})$ is the probability that the outcome variable y (*switched* or *consideration*) equals one given the vector of explanatory variables $\mathbf{x}$. G is a cumulative distribution function that maps $\mathbf{x}\boldsymbol{\beta} = \beta_0 + \sum_{i\geq 1} \beta_i x_i$ into the response probability (c.f. Wooldridge (2002)). In line with the related literature[5], this approach assumes that customers switch their supplier when their subjective net utility from switching is higher than from staying with the old supplier. Since the subjective net utility y^* cannot be meaningfully observed (y^* is a latent variable), the binary variable y is defined as follows,

$$y = 1[y^* > 0], \quad y^* = x\beta + e.$$

[5]Logit and probit regressions for modeling the switching behavior of electricity customers can also be found in Wieringa and Verhoef (2007), Ek and Söderholm (2008); Giulietti et al. (2005) uses a slightly more advanced bivariate probit model.

For a symmetric cumulative distribution function G, it follows that

$$P(y = 1|\mathbf{x}) = P(\mathbf{x}\boldsymbol{\beta} + e > 0|\mathbf{x}) = P(e > -\mathbf{x}\boldsymbol{\beta}|\mathbf{x}) = 1 - G(-\mathbf{x}\boldsymbol{\beta}) = G(\mathbf{x}\boldsymbol{\beta}),$$

which is exactly equation (3.1). Under the probit model, the error term e has a standard normal distribution, under the logit model, e has a standard logistic distribution, (cf. Wooldridge (2002)). Both versions were estimated and led to similar results, but for the sake of brevity and readability, we decided to report only one version. In order to maintain comparability with the bivariate probit model in Section 3.6.7, we chose to report the results of the probit model, the results for the logit model are available on request.

3.4 Variables

Potential explanatory variables are those listed in Table 3.2. They can be grouped into three categories. Economic variables include annual expenditure on electricity, the size of the household's living space, whether the respondent has used a tariff calculator to compare different ES, etc. Attitude variables describe how much importance respondents attach to sustainability, to the price and to the popularity of the ES. Sociodemographic variables address gender, age, and education. Since we want to measure if respondents, who have switched their ES in the last ten years, are more or less prone to switch again, the dependent variable *switched* of the switching model is used as an explanatory variable in the consideration model.

The questionnaire offered the respondents the possibility to choose between seven categories for expenditures on electricity: the first six categories described expenditure intervals in ascending order, the seventh category could be selected if the respondent did not know his expenditure level. This set up makes an implementation difficult. The *unknown expenditures* category prohibits the application of the usual midpoint method. We consider two different ways to address this issue. The first approach is to simply exclude the 26 respondents who did not know their expenditure levels from the estimation and to apply the usual midpoint method.

The second approach is econometrically more advanced. Here we pool the first six categories by twos into the categorical variables *Small, Medium* and *high expenditures*, whereas *Unknown Expenditures* is not changed. Then, we apply reversed Helmert coding, which allows us to compare every category with the mean of the subsequent categories. Thereby we are able to capture the effect of increasing expenditures and at the same time, we can measure the impact of *Unknown Expenditures* compared to the mean of all of those knowing their expenditure levels.

There are several reasons in favor of the first approach. First, *Unknown Expenditures* and *Tariff calculator* are not independent, if we include both in the estimation, we are running into the problem of collinearity. Second, the second approach makes the interpretation of the marginal effect difficult. Third, the first approach ensures a better comparability between a model treating all explanatory variables as exogenous (cf. Section 3.5) and a model taking into account that *tariff calculator* is potentially an endogenous variable (cf. Section 3.6). Therefore we choose the first approach. The results for the second approach are available on request (and do not differ significantly from the reported results).

The respondents could choose from several intervals concerning the size of their living space, which was subjected to the midpoint method, see Table 3.2 for further details. The same is true for the variable *age*.

The other questions - about attitude towards sustainability, price level and how important they consider popularity of their ES - allowed them to tick either "very important", "rather important", "rather unimportant" or "unimportant". The variable *popularity of the ES* was rather equally distributed among these four categories, whereas almost no one stated that the *price* or *sustainability* were "rather unimportant" or "unimportant" to them. Thus we decided to recode *price* and *sustainability* as binary variables, only measuring if the respondents consider them as "very important" or not[6]. In the case of *popularity of the ES* we decided to treat the four response options as four categories[7].

Since it makes no sense to assume that the intervals concerning the respondents' education are equally spaced, we treated *education* as a categorical variable.

3.5 Estimation Results

As mentioned and explained, the 26 respondents who did not know their expenditures levels are excluded in order to apply the midpoint method for the six remaining expenditure categories. Both estimation methods, logit and probit estimation, were applied. However, the discussion is confined to results obtained for the probit model due to its relation to the bivariate probit model applied in Section 3.6.7. The corresponding results for the switching model as well as the consideration model are summarized in Tables 3.3 and 3.4.

The probit regression coefficients give the change in the z-score (probit index) for a one unit change in the predictor variable (or for the change from zero to one for binary predictors respectively). The sign of the coefficients states if an increase (or the change from zero to one respectively) in the corresponding independent variable increases (positive sign) or decreases (negative sign) the probability to switch the ES. In order to interpret the estimated coefficients in terms of probabilities, we have to turn to the marginal effects of explanatory variables, which are reported in column four in Tables 3.3 and 3.4. The marginal effect of a specific independent variable is the partial derivative of the probability function, evaluated at the sample mean of all other independent variables. For binary variables, the marginal effect measures the change in the switching probability when the explanatory dummy variable goes from zero to one. For numerical variables, it measures the change in the switching probability for a one-unit change in the explanatory variable.

According to these estimates, it is the usage of a tariff calculator that has the largest effect on the decision to switch. That is good news to regulators and economists as information matters and lack thereof can explain a substantial degree of inaction. The coefficient of 1.14 of the z-score of the *tariff calculator* corresponds to the following probabilities: Keeping all other variables at their means, the probability to switch for tariff calculator-users is 46.9%, for non-users it is 11.1%. Thus the use of a tariff calculator quadruples the probability to switch, or expressed in absolute terms, it raises that probability by 35.8% (cf. Table 3.3, column four). The estimation

[6] A different model specification, treating the intervals either as four categories or as four numerical values did not lead to further insight.

[7] We also tried to treat them as four numerical values, but this specification would have obscured valuable information, see the estimation results in Section 3.5.

[8] As stated above, we implemented two approaches, differing in the exclusion of the 26 respondents not knowing their expenditures. The values in brackets refer to the subgroup of respondents knowing their expenditures, but do not differ significantly from the overall values.

[9] Was not included in the reported estimation results

[10] Was not included in the reported estimation results

[11] A different model specification with a finer subdivision of categories, differing between compulsory school and apprenticeship, was tested and did not lead to significantly different results.

Table 3.2: Variables included in the estimation and descriptive statistics[8]

Variable name	Coding and definitions	Mean	Std. dev.
Dependent Variable(s)			
Switched	1 for respondents reporting that they have changed the ES during the last ten years, 0 otherwise	0.29 (0.31)	0.45 (0.46)
Consideration	1 for respondents reporting that they are considering switching at the moment, 0 otherwise	0.24 (0.24)	0.42 (0.43)
Economic variables			
Switched	1 for respondents reporting that they have changed the ES during the last ten years, 0 otherwise	0.29 (0.31)	0.45 (0.46)
Tariff calculator	1 if respondents reported they had used a tariff calculator to compare different ES, 0 otherwise	0.46 (0.49)	0.5 (0.5)
Expenditure	Annual expenditures, six intervals (starting at "smaller than 200 €" up to "more than 1500€")	– (689)	– (370)
Property	1 if the respondent owns the apartment/house he lives in, 0 otherwise	0.55 (0.55)	0.50 (0.50)
Living space	Sqm of living space, five intervals (starting at "smaller than 45 sqm" up to "more than 120 sqm")	93 (93)	34 (35)
Mobile operator churn	1 if the respondent reported that he had switched the mobile operator in the last five years, 0 otherwise	0.49 (0.50)	0.50 (0.50)
Mobile tariff information	1 if respondents reporting of actively informing themselves about mobile tariffs, 0 otherwise	0.88 (0.90)	0.32 (0.30)
Known Expenditures[9]	1 for respondents who knew their annual electricity costs, 0 otherwise	0.91 (–)	0.28 (–)
Electric heating[10]	1 for electricity heated home, 0 otherwise	0.13 (0.13)	0.34 (0.34)
Attitudes towards			
Sustainability	1 for "very important", 0 otherwise (not "very important"),	0.45 (0.47)	0.50 (0.50)
Price	1 for "very important", 0 otherwise (not "very important"),	0.74 (0.74)	0.44 (0.43)
Popularity of the ES	Four categories from 1 for "very important" up to 4 for "unimportant"	2.75 (2.76)	0.95 (0.95)
Socio-demographic Var.			
Age	Respondent's self reported age interval, six intervals (starting at "younger than 20" up to "older than 65")	40 (40)	13 (13)
Gender	1 if respondent is female, 0 if respondent is male	0.58 (0.58)	0.49 (0.49)
Urban	1 if respondent lives in a city with more than 10.000 inhabitants, 0 otherwise	0.69 (0.67)	0.46 (0.47)
Education[11]	Category "Low" if the self-reported highest educational attainment is compulsory school or a completed apprenticeship	22.8% (23.2%)	
	Category "Medium" if the self-reported highest educational attainment is a secondary education ("Matura") and/or a subsequent course of lectures ("Kolleg")	35.4% (36.6%)	
	Category "High" if the self-reported highest educational attainment is a higher education institution (university or "Fachhochschule")	41.7% (40.2%)	

Table 3.3: Probit estimation results for the dependent variable *switched*.

Variables	Coeff.	Stand.Err.	Marg.Eff.	Stand.Err.
Constant	-2.544**	0.864	–	–
Tariff calculator	1.144**	0.199	0.358**	0.057
Expenditure	-0.000	0.000	-0.000	0.000
Property	-0.240	0.233	-0.078	0.076
Living space	0.002	0.004	0.001	0.001
Mobile operator churn	0.032	0.193	0.010	0.062
Mobile tariff information	-0.015	0.314	-0.005	0.101
Sustainability	-0.013	0.188	-0.004	0.060
Price	0.102	0.216	0.032	0.067
Popularity of the ES				
Categ. "rather important"	0.358	0.371	0.069	0.064
Categ. "rather unimportant"	1.014**	0.368	0.272**	0.076
Categ. "unimportant"	1.047**	0.378	0.284**	0.083
Age	0.006	0.008	0.002	0.003
Gender	0.302	0.196	0.095	0.060
Urban	0.982**	0.251	0.274**	0.059
Education				
Category "medium"	-0.532*	0.256	-0.184*	0.090
Category "high"	-0.532*	0.263	-0.184*	0.092
N	276		276	
Log-likelihood	-130.523		–	
$\chi^2_{(16)}$	81.399		–	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

also shows that the probability to switch is significantly higher in urban areas (36.7%) than in rural areas (9.3%). Furthermore, those who attach greater importance to the popularity of the ES switch significantly less often: the probability to switch is 8.3% among respondents who find the popularity of the ES "very important", 15.2% among the "rather important"-group, 35.5% among the "rather unimportant"-group and 36.7% among the "unimportant"-group[12]. The estimation further indicates that compared to the switching probability of the low education category, the probability for respondents with a medium or a high education is 18.4 percentage points smaller.

For the consideration model, we see again that those familiar with a tariff calculator have a higher probability to consider switching (28.5% vs. 15.5%), but the influence is not as strong as in the case of the real thing, switching. The variable with the highest coefficient (in absolute terms) is *switched* (-0.77), revealing an interesting observation: Those who have already switched have a significantly smaller probability to consider switching again (9.3% vs 28.9%). Again, when popularity is considered important, the probability to consider switching becomes smaller[13] (for the "very important" and "rather important" categories, switching probabilities are ca. 13%, for the "rather unimportant" and "unimportant" categories, the probabilities range between 26.5% and 29.7%). In contrast to the switching model, the attitude towards

[12]The differences between the switching probability of category "very important" and the three other categories are displayed in column four of Table 3.3

[13]Pooling the "very important" and "rather important" categories as well as the "rather unimportant" and "unimportant" categories under a different model specification leads to highly significant results.

Table 3.4: Probit estimation results for the dependent variable *consideration*.

Variables	Coeff.	Stand.Err.	Marg.Eff.	Stand.Err.
Intercept	-2.092*	0.827	–	–
Switched	-0.768**	0.233	-0.196**	0.051
Tariff calculator	0.448*	0.203	0.130*	0.059
Expenditures	0.000	0.000	0.000	0.000
Property	0.268	0.237	0.077	0.067
Living space	0.006	0.004	0.002	0.001
Mobile operator churn	0.078	0.190	0.023	0.055
Mobile tariff information	-0.125	0.311	-0.038	0.097
Sustainability	0.203	0.183	0.059	0.054
Price	0.425†	0.221	0.113*	0.053
Popularity of the ES				
Categ. "rather important"	-0.029	0.356	-0.006	0.079
Categ. "rather unimportant"	0.456	0.354	0.126	0.087
Categ. "unimportant"	0.549	0.362	0.157†	0.092
Age	-0.007	0.008	-0.002	0.002
Gender	-0.069	0.188	-0.020	0.055
Urban	0.276	0.235	0.077	0.063
Education				
Category "medium"	-0.274	0.257	-0.075	0.073
Category "high"	0.063	0.250	0.020	0.078
N	276		276	
Log-likelihood	-134.118		–	
$\chi^2_{(17)}$	37.7		–	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

the price level plays a role – the higher the self-reported price-sensitivity of the respondents is, the more they consider to switch. Since almost no respondents considered the price level as rather unimportant or unimportant, the variable *price* is binary. The probability to consider switching of those for whom the price level is very important is almost twice as high as of those who consider the price level as not very important (24.6%vs. 13.3%).

3.5.1 Discussion and Policy Implications

The significance and magnitude of the coefficient of the variable *tariff calculator* attests the effectiveness of the regulator's (E-control) efforts to ease the comparison of tariffs offered by different ES, as this leads to higher switching rates. Recently, also private internet-based price comparison sites have offered consumers advice on choosing between suppliers. In contrast to the tariff calculator offered by E-control, these private operators promote their services. This could be a reason for the most recent increase in switching rates in Austria ([14].

The higher switching rates in urban areas indicate that the former regional monopolies enjoy an even 'more quiet life', to quote Hick's famous statement about monopoly profits, in rural areas. The high impact of *popularity* can be explained by the regulatory past and inertia. Customers generally are familiar only with the former (regional) monopolist and have

[14] In a press release of the E-control from 10th of February, 2014, a switching rate of 1% for the first half 2013 is reported, c.f. E-control (2014). The all-year switching rate for 2013 has not been published yet.

little knowledge what to expect from new (regional) market entrants (cf. Wieringa and Verhoef (2007)). Further, the customers may ascribe the experienced past reliability to the incumbent supplier. In other words, some consumers may worry about supply reliability although this is of course unrelated to the choice of a particular supplier due to the nature of electricity networks. In this light, it is not surprising that respondents who attach high importance to *popularity* seem to be exactly those who are discouraged from switching due to their unfamiliarity of the new competitors on the electricity market.

It is surprising that neither *expenditure level* (which determines the absolute amount of savings potential), nor the subjective sensitivity to prices significantly influence the past switching behavior. Contrary to mobile phone tariffs, which are well promoted in public, the publicly barely advertised tariff structures complicate a comparison between ES (cf. Chamber of Labour (2005)). The independent variable *mobile operator churn* enables us to investigate if the group of respondents with a high switching rate in the mobile telephony market exhibit also a high(er) switching behavior in the electricity market. The estimation results do not indicate that there exists a group characterized by a higher switching behavior encompassing several markets. These findings coupled with the significant difference in switching rates between mobile phone (high) and electricity (low) services poses a puzzle as the potential savings are larger and electricity customers seem to be less bound to a supplier via a contract than mobile customers. There are a number of explanations but the most crucial one seems to be the one of information, which is compatible with the high contribution of the variable *tariff calculator*. Mobile phone services are heavily advertised focusing on tariff (and tariff structure) with nothing comparable undertaken by electricity suppliers (and if they do advertise, they try rather to establish a brand rather than to announce prices).

We tested several model specifications. It may be worth noting that the inclusion of *electric heating* as explanatory variable had no significant impact. It is conceivable that respondents have preciser knowledge about their heating system than about their expenditure levels and thus *electric heating* might capture the customers' potential for cost reductions better. However, *electric heating* and *expenditures* may be correlated. Hence, we tested a model specification, where only *electric heating* was included, but this did not reveal a significant influence either[15]. These observations do not correspond with the findings of Ek and Söderholm (2008) in the Swedish electricity market, where customers with an electric heating are on average more active switchers.

The estimation of the consideration model shows a highly significant influence of the independent variable *switched* on the dependent variable *consideration*. Those who have switched at least once have a significantly lower probability to consider switching (again). This can be interpreted in two ways. First, the respondents might believe after switching that their current supplier is still the best in the market and that this condition does not change. Second, the respondent may remember the switching procedure as cumbersome and thus do not consider switching again. Both interpretations do not raise hope that the switching rates in Austria will skyrocket in the near future. We find neither an impact of the respondents' attitude towards sustainability on their switching behavior in the past, nor on their intention to switch in the future. This result is of special interest, since many ES in Austria offer green or at least renewable (hydro) electricity. This characteristic is used for advertising purposes and also a list of

[15]So as to avoid collinearity problems, we decided to include only *expenditures* in the reported model.

	Observed Switched	
	0	1
Predicted Switched 0	168	34
1	22	52

Table 3.5: Switching model

	Observed Consid.	
	0	1
Predicted Consid. 0	200	50
1	9	17

Table 3.6: Consideration model

green ES is officially published[16]. Contrary to the switching model, the variable *price importance* influences the probability to consider switching. As expected, being more price-sensitive increases the probability to consider switching. This reveals a potential (but so far not pursued) advertising strategy for competing ES. Familiarizing customers more with the tariff structures and promoting potential savings – as it is common e.g. in the telecommunication market – could induce many customers to move from consideration to action.

3.5.2 Goodness of Fit Tests

By the help of a log-likelihood ratio test, we can reject the hypothesis that all coefficients of the explanatory variables are zero for both models on a 1% level. (the calculated test statistic $\chi^2_{16} = 81.40$ with 16 degrees of freedom for the switching model and $\chi^2_{17} = 37.70$ with 17 degrees of freedom for the consideration model). The Pearson χ^2 goodness-of-fit test is a test of the observed against the predicted number of responses using cells defined by the covariate patterns. For both models, the Pearson χ^2 statistic does not lead to a rejection of the null hypothesis, stating that the model fits the data reasonably well (switching model: $\chi^2(254) = 271.98$, $p-$value$= 0.2092$; consideration model: $\chi^2(253) = 269.04$, $p-$value$= 0.2332$). Since the number of covariate patterns is close to the number of observations (271 vs. 276 in both models), the applicability of the Pearson χ^2 test is questionable. As suggested in Hosmer and Lemeshow (2000), we regroup the data by ordering on the predicted probabilities and form $g = 12$ equally sized groups. The Hosmer-Lemeshow goodness-of-fit statistic $\hat{C}$ is obtained by calculating the Pearson χ^2 statistic from the $g \times 2$ table of observed and estimated expected frequencies (Hosmer and Lemeshow (2000)). The $\hat{C}$ statistic is well approximated by the χ^2 distribution with $g - 2$ degrees of freedom. In the switching (consideration) model, $\chi^2(10) = 11.08$ ($\chi^2(10) = 12.86$), with a corresponding $p-$value of 0.3512 (0.2317), thus also indicating a reasonable fit of both models.

The confusion matrices for both models are shown in Tables 3.6 and 3.6. If the predicted probability of switching (consideration) is higher than the cutoff value of 0.5, the predicted value of the dependent variable *switched (consideration)* is set equal to 1. The percentage of correct classifications by the model is 79.1% in case of the switching model and 78.6% in case of the consideration model. To put this in perspective, note that only 31.2% (24.3%) have switched (consider to switch), so if we simply predicted everyone to have not switched (to not consider to switch) we would get 68.8% (75.7%) correct.

Furthermore, the area under the ROC curve amounts to 0.82 for the churn model and to 0.73 for the consideration model.

Overall, the goodness of fit measures show a reasonably good fit for the switching model, but only a mediocre fit for the consideration model. The weak point of the consideration model is the low predictive power for those who intend to switch, which is also reflected by a

[16]see http://www.e-control.at/en/consumers/renewables/green-electricity-suppliers, last accessed November, 2nd 2014

small sensitivity rate of 25.37% (in contrast to 60.47% for the switching model). Hence, in the following section, we concentrate only on the switching model when we address the potential endogeneity of the *tariff calculator*.

3.6 Treating the Variable *Tariff Calculator* as Endogenous

3.6.1 Potential Endogeneity of the Tariff Calculator and Possible Instrumental Variables

Endogeneity arises owing to problems such as simultaneity between the predictor variable and the outcome or due to omitted variables. There is reason to suppose that the predictor variable *tariff calculator* is endogenous. First, the predictor *tariff calculator* and the outcome *switched* could be influenced by an unmeasured interest in switching (omitted variables). Second, the thought of switching may cause the usage of the tariff calculator and not vice versa as assumed in the model (simultaneity bias).

In fully linear models, the instrumental variable (IV) approach constitutes a popular and well-developed method for endogeneity problems (cf. Section 3.6.2). An instrumental variable is an observed variable that predicts the endogenous predictor variable, but is unassociated with the outcome, neither directly nor indirectly through unmeasured confounders, except via the effect of the predictor variable (cf. Wooldridge (2002)).

In our study, we want the variable *tariff calculator* to measure the level of information of the respondents. Thus, we need a variable that can serve as a predictor for the information level concerning the electricity market, but is unassociated with the variable *switched*. Our dataset also includes the information if and how the respondents monitor the developments in mobile tariffs (mapped by the variable *mobile phone information*). Further, the variable *mobile operator churn* measures if respondents had switched their mobile operator in the last five years. Both variables are potential candidates to serve as an instrument for the *tariff calculator*: First, both variables do not influence the dependent variable *switched* significantly (c.f. Section 3.5)[17]. There is, also from an intuitive point of view, no reason to believe that there is a causal relationship between behavior on the mobile churn market and the switching behavior of an ES other than the existence of cohorts of respondents who are characterized by a general higher information level.

Second, it makes sense to assume that respondents who monitor mobile tariff developments more closely are also those who inform themselves about tariffs on the electricity market by the help of a tariff calculator. One can also argue that those who have switched their mobile operator are those whose probability to use a tariff calculator is higher. It clarifies matters to run two separate probit estimations with the *tariff calculator* as dependent variable and one time *mobile phone information* and the other time *mobile operator churn* as explanatory variable (plus in both cases all other exogenous variables). The coefficient of the z-score of *mobile operator churn* (-0.13) is small, negative and signals even a negative relation between *mobile operator churn* and *tariff calculator*. Thus, *mobile operator churn* is dropped from the list of potential instruments. *Mobile phone information* is by far more promising. The

[17]Also regressing *switched* separately and individually on *mobile operator churn* and *mobile operator churn* does not show any significant influence.

coefficient of the z-score, 0.87, is significant on a 2% level, keeping the other covariates at their means, this relates to an absolute increase in the probability to use a tariff calculator of 31.4% (20.7% for respondents not actively informed about mobile tariffs vs. 52.1% for well-informed respondents).

There are two more details needing further clarification. Obviously, *mobile operator churn* and *mobile phone information* are not independent, and since an instrument should be independent from all explanatory variables other than the endogenous one, we drop the variable *mobile operator churn* completely. Second, as in Section 3.5, the variable *unknown expenditures* unnecessarily complicates the estimation process and is thus, just as in Section 3.5, excluded.

In order to examine the available methods and their applicability to our endogeneity problem, we first recapitulate in Section 3.6.2 how endogeneity is formally modeled and treated under a fully linear model with one endogenous explanatory variable (the linear case is based on Chapter 5, Wooldridge (2002)). In Section 3.6.3 we introduce several possibilities how (and how not) to treat an endogenous dummy variable in a non-linear model and present the respective estimation results.

3.6.2 The Reference Case of a Fully Linear Model

Let y denote the dependent variable, $\mathbf{x} = (1, x_1, \ldots, x_K)$ the vector of exogenous explanatory variables, x_e the endogenous explanatory variable, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)$ and β_e the coefficients to be estimated, z the instrumental variable and e and v the error terms. We confine ourselves to the case of one endogenous and one instrumental variable. Then,

$$y = \beta_e x_e + \mathbf{x}\boldsymbol{\beta} + e, \tag{3.2}$$

with $Cov(x_i, e) = 0$ for $i = 1, \ldots, K$ and $Cov(x_e, e) \neq 0$. In order to understand why the condition $Cov(x_e, e) \neq 0$ formally describes the idea of endogeneity, it helps to think of a variable which influences y and x_e, but which is omitted in equation (3.2). Since it is omitted in the estimation, its influence is only captured by the error term e; as a consequence $Cov(x_e, e) \neq 0$. Endogeneity leads to inconsistent estimates for all the coefficients β_e and $\boldsymbol{\beta}$.

A valid instrumental variable z has to fulfill two requirements. First,

$$Cov(z, e) = 0, \tag{3.3}$$

and second, z has to be partially correlated with x_e once the other exogenous variables have been netted out. Due to a lack of an unbiased estimator of e, the first requirement (3.3) can not be tested formally , so one has to rely on common sense. The second requirement can easily be tested by testing if α_z equals zero in equation (3.4).

When both requirements are fulfilled, we can estimate in a first stage the so-called auxiliary equation, which regresses the endogenous variable on the instrument and the explanatory variables,

$$x_e = \alpha_z z + \mathbf{x}\boldsymbol{\alpha} + v. \tag{3.4}$$

The estimation $\hat{x}_e$, determined by $\hat{x}_e = \hat{\alpha}_z z + \mathbf{x}\hat{\boldsymbol{\alpha}}$, is then substituted for x_e in the second stage regression,

$$y = \beta_e \hat{x}_e + \mathbf{x}\boldsymbol{\beta} + e. \tag{3.5}$$

By our assumptions, e is uncorrelated with all explanatory variables in (3.5), thus the ordinary least square method delivers consistent estimates for the coefficients β_e and $\boldsymbol{\beta}$. It is worth noting that the procedure seems to include two separate stages, but in order to get correct estimation errors, both stages have to be estimated simultaneously. Respective software packages, enabling the user to estimate both stages simultaneously, are available under all major statistical computing software.

3.6.3 Binary Choice Model with an Endogenous Dummy Variable

Extending the familiar latent variable approach for binary choice models leads to the following presentation of the counterpart of the linear equation (3.2), (cf. Wooldridge (2002), Miranda and Rabe-Hesketh (2006)),

$$y = I(\beta_e x_e + \mathbf{x}\boldsymbol{\beta} + e \geq 0), \tag{3.6}$$

where $I(\cdot)$ is the indicator function which equals one if the latent variable $\beta_e x_e + \mathbf{x}\boldsymbol{\beta} + e$ is positive and zero otherwise. As already noted in Section 3.3, the assumption on the error term e defines if a logit or a probit model is applied. Analogous to the linear model, x_e is endogenous if x_e and e are not independent. Under our set-up, the endogenous variable *tariff calculator* is a dummy, thus also the counterpart of equation (3.4) is of the form of a threshold crossing model, (c.f. Miranda and Rabe-Hesketh (2006)),

$$x_e = I(\alpha_z z + \mathbf{x}\boldsymbol{\alpha} + v \geq 0), \tag{3.7}$$

where, as before, $\alpha_z z + \mathbf{x}\boldsymbol{\alpha} + v$ represents a latent continuous variable, α_z, $\boldsymbol{\alpha}$ are the coefficients and v is a residual term.

Under our set-up, the dependent variable as well as the endogenous explanatory variable are both binary. Thus, the estimated equations for the dependent variable *switched* as well as for the endogenous variable *tariff calculator* both involve non-linear regressions. This raises the following questions concerning the transferability of the methods applied in linear models: Do the consistency properties of the two stage least square estimation also hold in the non-linear case? Without the help of a common statistical program, how should the simultaneous estimation of both stages be accomplished? And, perhaps most important of all, logit and probit models used in the first stage estimate the probability $E[x_e|z, \mathbf{x}] = P(x_e = 1|z, \mathbf{x})$. The estimated probability, taking values in $[0, 1]$, is clearly a qualitatively different kind of variable than the binary endogenous variable *tariff calculator*. Should this estimated probability be substituted for the endogenous binary variable in the second stage?

In the following sections, we present different approaches how to handle these questions. We start – quite unconventionally – with two methods, which seem rather intuitive and are commonly used for other non-linear endogeneity problems, but which do not work for our binary choice model (see Section 3.6.4 and 3.6.5). In Sections 3.6.6, 3.6.7 and 3.7, we present approaches which can be applied to our set-up, we summarize the pros and cons of each approach and we display the respective estimation results.

3.6.4 The Two-Stage Predictor Substitution Model

The two-stage predictor substitution (2SPS) method, so called by Terza et al. (2008), is the seemingly obvious mimicry of the two-step least square estimation in the linear model (cf.

Wooldridge (2002)). In a first step, the probabilities $P(x_e = 1|z, \mathbf{x}) = G(\alpha_z z + \mathbf{x}\boldsymbol{\alpha})$ are estimated, where G is defined as in Section 3.3. In a second step, we estimate β_e and $\boldsymbol{\beta}$ from the probit or logit regression of y on $\mathbf{x}$ and on $G(\hat{\alpha}_z z + \mathbf{x}\hat{\boldsymbol{\alpha}})$. This approach does not produce consistent estimators, as both Terza et al. (2008) and Wooldridge (2002) warn[18]. It is worth noting that the estimated probabilities $G(\alpha_z z + \mathbf{x}\boldsymbol{\alpha})$ from the first stage are also known under the name propensity scores. When dealing with endogenous variables, suggestions to include the propensity scores as a covariate in a multivariate regression model predicting the outcome, as can be found e.g. in Shah et al. (2005), should be treated with caution for the above mentioned reasons.

3.6.5 Control Functions

A number of variants of control functions exists, descriptions can be found for example in Wooldridge (2010). Terza et al. (2008) present a method called two stage residual inclusion (2SRI), which basically is a variant of control functions. They claim that, in contrast to the 2SPS method (see Section 3.6.4), the 2SRI method is generally statistically consistent for the broader class of nonlinear models. Lewbel et al. (2012) show that for the special case of an endogenous dummy variable, this claimed general consistency of the control function approach does not hold.

Intuitively, the 2SRI method works as follows: The first stage coincides with the first stage of the 2SPS method. Then, departing from 2SPS, the estimated auxiliary equation is solved for the estimated residuals. Instead of substituting the estimation for the endogenous variable, $\hat{x}_e$, (in our case the probabilities $G(\hat{\alpha}_z z + \mathbf{x}\hat{\boldsymbol{\alpha}})$) for the endogenous variable x_e in the second stage, 2SRI includes the estimated residuals from the first stage as additional covariates in the second stage. Thereby, the bias of the explanatory variable x_e, which is maintained as a covariate in the second stage estimation, should be corrected for. As shown in Terza et al. (2008), under a fully linear model, 2SRI, 2SPS and the two stage least squares estimation coincide.

In Lewbel et al. (2012), the authors show formally, why the control function approach fails on the problem of binary choice models with an endogenous explanatory variable. The following formal description is based on their work. The auxiliary equation of the first stage, describing the relationship between x_e and $(z, \mathbf{x})$, is now captured by the equation $G(x_e, z, \mathbf{x}, v) = 0$, where v is the error term. Lewbel et al. (2012) name three conditions for the control function approach to work. First, $G(x_e, z, \mathbf{x}, v) = 0$ has to be parametrically or non-parametrically specified and so can be estimated. Second, the equation $G(x_e, z, \mathbf{x}, v) = 0$ has to be solvable for the errors v. And third, there has to exist a function h and a "clean" error U such that $e = h(v, U)$ and U is independent of both x_e and v.

Suppose for illustration purposes that x_e is not a binary but a scalar. For a binary choice model, we still have that $y = I(\beta_e x_e + \mathbf{x}\boldsymbol{\beta} + e \geq 0)$. For a continuous variable x_e, the function G has the following form, $G(x_e, z, \mathbf{x}, v) = x_e - \alpha_z z - \mathbf{x}\boldsymbol{\alpha} - v = 0$. Further suppose that e and v are jointly normal with mean zero. Due to their jointly normal distribution, they can be linearly decomposed as $e = h(v, U) = \lambda v + U$, where U is independent of v and the constant λ depends on the covariance matrix of (e, v), (cf. Lewbel et al. (2012)). Solving for the error term in the

[18] Wooldridge (2002) gives the following explanation for the inconsistent estimation results: For the two-step procedure to work, we would have to have $P(y = 1|x_e, \mathbf{x}, z) = G(\beta_e G(\hat{\alpha}_z z + \mathbf{x}\boldsymbol{\alpha}) + \mathbf{x}\boldsymbol{\beta})$. But $P(y = 1|x_e, \mathbf{x}, z) = E[y|x_e, \mathbf{x}, z] = E[1[\mathbf{x}\boldsymbol{\beta} + \beta_e x_e + e > 0|x_e, \mathbf{x}, z]]$ and since the indicator function $1[\cdot]$ is nonlinear, we cannot pass the expected value through.

estimated function G and making use of h leads us to the equation $y = I(\beta_e x_e + \mathbf{x}\boldsymbol{\beta} + \lambda v + U \geq 0)$, which is just an ordinary probit model with independent normal error U. The included error v is simply treated as an additional regressor, and λ is the corresponding coefficient.

Now, let us come back to our actual problem of an endogenous binary variable. Unfortunately, in this case, G does not fulfill the required condition to be solvable for the error term v. For a binary endogenous variable, G has the following form, $G(x_e, z, \mathbf{x}, v) = x_e - I(\alpha_z z + \mathbf{x}\boldsymbol{\alpha} + v \geq 0) = 0$, which makes it impossible to solve for v given observations of x_e, z and $\mathbf{x}$.

Of course, as is also noted by Lewbel et al. (2012), one could always define a model G by $G(x_e, z, \mathbf{x}, v) = E[x_e|z, \mathbf{x}] + v$ (as looks familiar from the previous sections) and estimate v as the residual from the regression of x_e on z and $\mathbf{x}$. But since x_e is binary, the resulting v would not be independent of $(z, \mathbf{x})$, which would generally cause violations of the required assumptions regarding U.

To sum up, we have seen that control functions are limited to models with continuous endogenous variables and cannot be used when the endogenous variable x_e is binary.

3.6.6 The Linear Probability Model

The linear probability model (LPM) for binary response is an approximation for threshold crossing models such as logit or probit, which circumvents many of the aforementioned problems. In case of a model with no endogenous regressors, the corresponding equation of (3.1) has the simple following form (c.f. Wooldridge (2002)),

$$y = \mathbf{x}\boldsymbol{\beta} + e. \tag{3.8}$$

$$P(y|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \ldots, \beta_K x_K = \mathbf{x}\boldsymbol{\beta}. \tag{3.9}$$

Of course, several problems arise due to this simplification. First, heteroskedasticity is present, since the error term e must either equal $1 - \mathbf{x}\boldsymbol{\beta}$ or $-\mathbf{x}\boldsymbol{\beta}$, which are functions of all the elements of $\mathbf{x}$. As stated in Wooldridge (2002), using heteroskedasticity-robust standard errors is a possible way to deal with this problem. Second, with a LPM, the estimated probabilities can never have the typically S-shaped form of a distribution function. And third – this yields the most commonly recognized drawback of the LPM – the fitted probabilities can go below zero or above one (cf. Lewbel et al. (2012)). But, as stated in Wooldridge (2002), even with these weaknesses, the LPM often seems to provide good estimates of the partial effects on the response probabilities near the center of the distribution of $\mathbf{x}$.

In case of the presence of an endogenous binary regressor x_e, we can apply the approximation via the LPM in two stages, which coincides with the two stage least squares (TSLS) estimation as presented in Section 3.6.2. We estimate in a first stage the auxiliary regression

$$x_e = \alpha_z z + \mathbf{x}\boldsymbol{\alpha} + v. \tag{3.10}$$

The estimated $\hat{x}_e$, determined by $\hat{x}_e = \hat{\alpha}_z z + \mathbf{x}\hat{\boldsymbol{\alpha}}$, is then substituted for x_e in the second stage regression,

$$y = \beta_e \hat{x}_e + \mathbf{x}\boldsymbol{\beta} + e. \tag{3.11}$$

From Section 3.6.2 we know that the two stage least squares estimation delivers consistent estimates, but this comes at the cost of using a linear approximation instead of a threshold crossing model such as logit or probit. Lewbel et al. (2012) state that they do not know any evidence if the application of e.g. a probit model and ignoring the endogeneity is generally any worse or better than using the LPM on two stages.

Table 3.7: TSLS estimation results for the dependent variable *switched*, together with two "exogenous" reference models.

	Endogenous LPM (TSLS)		**Exogenous Probit**		**Exogenous LPM**	
Variables	**Coeff.**	**Rob.St.Err.**	**Marg.Eff.**	**St.Err.**	**Coeff.**	**Rob.St.Err.**
Constant	-0.201	0.188	–	–	-0.202	0.189
Tariff calculator	0.430	0.294	0.358**	0.055	0.323**	0.053
Expenditure	0.000	0.000	-0.000	0.000	0.000	0.000
Property	-0.065	0.066	-0.077	0.075	-0.074	0.064
Living space	0.000	0.001	0.001	0.001	0.000	0.001
Sustainability	-0.011	0.063	-0.005	0.060	0.002	0.052
Price	0.021	0.058	0.032	0.069	0.031	0.053
Popularity of the ES						
Cat. "rather imp."	0.085	0.097	0.068	0.064	0.085	0.095
Cat. "rather unimp."	0.268**	0.094	0.272**	0.076	0.264**	0.093
Cat. "unimp."	0.284**	0.101	0.284**	0.083	0.292**	0.096
Age	0.002	0.002	0.002	0.003	0.002	0.002
Gender	0.073	0.052	0.097	0.062	0.063	0.051
Urban	0.226*	0.093	0.314**	0.078	0.253**	0.060
Education						
Cat. "medium"	-0.124†	0.068	-0.182*	0.089	-0.123†	0.069
Cat. "high"	-0.134†	0.069	-0.183*	0.092	-0.127†	0.068
N	276		276		276	
R^2	0.245		–		0.257	
	$\chi^2_{(14)} = 61.415$		–		$F_{(14,261)} = 8.582$	

Significance levels: † : 10% $*$: 5% $**$: 1%

Table 3.7 displays the estimation results of three different models. In the first column ("Endogenous LPM"), we present the results of the above described linear approximation of the switching model by a TSLS estimation, where the endogenous variable *tariff calculator* was instrumented by *mobile tariff information.* The coefficients in this column give the partial effect on the probability to have switched due to a one unit change or due to the change from zero to one in the explanatory variables. The Stata option `vce(robust)` was used to deal with the heteroskedasticity problem, leading to robust standard errors. We compare the TSLS approach with two other models. In the second column ("Exogenous probit") we display the marginal probabilities from the probit model of Section 3.5[19]. Note that the numbers differ slightly from Table 3.3 because for comparison reasons, we did not include *mobile tariff information* or *mobile operator churn.* In the third column ("Exogenous LPM") we present the estimation results where we approximated also the exogenous probit model by a linear probability model (ordinary least square estimation (OLS)). Thereby we can examine the impact of a linear approximation without the additional change due to the instrumented endogenous variable. Again, *mobile tariff information* and *mobile operator churn* were excluded. So as to be able to perform endogeneity tests, we used also in the third model robust standard errors.

Comparing the three estimation results, it is immediately striking that there is a high degree of similarity between the estimated coefficients as well as between the marginal probabilities of the different models.

Approximating the exogenous probit model (second column) with the help of a LPM (third

[19]Table 3.10 in Appendix 3.9.1 lists the coefficients as well as the marginal probabilities of this alternative probit estimation.

column) results in almost the same marginal probabilities. Also the significance levels resemble each other, the same explanatory variables are significant in both models. This confirms that approximating a latent threshold model with the help of a linear probability model works surprisingly well.

When we turn now to the TSLS model with special focus on the coefficient of the instrumented *tariff calculator*, we see that the magnitude of the coefficients resembles the ones from the reference models. The particularly interesting coefficient of *tariff calculator* (0.43) is even slightly higher than in the exogenous reference models (0.32 for the LPM and 0.36 for the probit), but it is not significant any longer. This can possibly be ascribed to the potential weakness of the instrument *mobile phone information*.

A natural question to ask is whether *tariff calculator* is indeed endogenous or if we can stick to treating the variable as exogenous as done in Section 3.5. The advantages of an exogenous model are apparent: First, it avoids the approximation of a threshold crossing model by a LPM, and second, even if we switch to the linear approximation, OLS estimators are more efficient than the TSLS estimator. The test for endogeneity of a variable can only be performed after fitting an endogenous model. The procedure for linear models (OLS vs. TSLS) became known as "Wu-Hausman Test" and determines whether the differences between OLS and TSLS estimates are significant.

Since we requested a robust variance-covariance matrix, we cannot apply the Wu-Hausman test. Instead, we apply Wooldridge's robust score test and a robust regression-based test, both implemented in the Stata command `estat endogenous`. The null hypothesis in both tests is that the variable *tariff calculator* can be treated as exogenous. Neither of the two tests rejects the null hypothesis: The test statistic of Wooldridge's robust score test, $\chi^2(1)$ equals 0.146, with a $p-$value of 0.703, the test statistic of the regression-based test of exogeneity, $F(1, 260)$, equals 0.139 with a $p-$value of 0.710. Thus, at least under the approximation by the linear probability model, we can treat *tariff calculator* as exogenous.

3.6.7 Maximum Likelihood Estimation of a Bivariate Probit Model

The following formal description of the Maximum Likelihood (ML) estimation of a probit model with a binary endogenous regressor follows Winkelmann (2011). A more detailed derivation of the joint distribution of y and x_e can be found in Wooldridge (2002), p. 478.

We deal with the already well known following two equations,

$$y = I(\beta_e x_e + \mathbf{x}\boldsymbol{\beta} + e \geq 0), \tag{3.12}$$

$$x_e = I(\alpha_z z + \mathbf{x}\boldsymbol{\alpha} + v \geq 0), \tag{3.13}$$

where (e, v) is independent of $\mathbf{x}$ and z. In this section, we make the additional assumptions that the latent errors are distributed as bivariate normal with mean zero, each has unit variance, and $\rho = \text{Corr}(e, v)$. If $\rho = 0$, the model for y is the standard probit model, and there is no need for instrumenting the variable x_e as done in equation (3.13). If $\rho \neq 0$, then x_e and e are correlated, or in other words, x_e is endogenous.

To set up the bivariate probit model, we need to consider the four possible cases of the joint distribution of y and x_e (conditional on $\mathbf{x}$ and z):

$$\begin{aligned}
P(y=0, x_e=0|\mathbf{x}, z) &= P(e \leq -\beta_e x_e - \mathbf{x}\boldsymbol{\beta}, v \leq -\alpha_z z - \mathbf{x}\boldsymbol{\alpha}) \\
P(y=1, x_e=0|\mathbf{x}, z) &= P(e \geq -\beta_e x_e - \mathbf{x}\boldsymbol{\beta}, v \leq -\alpha_z z - \mathbf{x}\boldsymbol{\alpha}) \\
P(y=0, x_e=1|\mathbf{x}, z) &= P(e \leq -\beta_e x_e - \mathbf{x}\boldsymbol{\beta}, v \geq -\alpha_z z - \mathbf{x}\boldsymbol{\alpha}) \\
P(y=1, x_e=1|\mathbf{x}, z) &= P(e \geq -\beta_e x_e - \mathbf{x}\boldsymbol{\beta}, v \geq -\alpha_z z - \mathbf{x}\boldsymbol{\alpha})
\end{aligned}$$

We see that the joint distribution of e and v fully determines the joint distribution of y and x_e. In the bivariate probit model, it is assumed that e and v have the joint distribution function $F(e, v) = \Phi(e, v, \rho)$, where Φ is the cumulative density function of the bivariate standard normal probability distribution. This allows us to write the joint probability function of y and x_e compactly as

$$f(y, x_e|\mathbf{x}, z) = \Phi\big((2y-1)(\beta_e x_e + \mathbf{x}\boldsymbol{\beta}), (2x_e-1)(\alpha_z z + \mathbf{x}\boldsymbol{\alpha}), (2y-1)(2x_e-1)\rho\big). \tag{3.14}$$

Assuming an independent sample of n observations $(y_i, x_{ei}, \mathbf{x}_i, z_i)$, and abbreviating the parameters as $\zeta = (\boldsymbol{\beta}, \beta_e, \boldsymbol{\alpha}, \alpha_z, \rho)$, the likelihood function $L(\zeta; y, x_e, \mathbf{x}, z)$ has the following form,

$$L(\zeta; y, x_e, \mathbf{x}, z) = \prod_{i=1}^{n} f(y_i, x_{ei}|\mathbf{x}_i, z_i), \tag{3.15}$$

with $f(y_i, x_{ei}|\mathbf{x}_i, z_i)$ defined in equation (3.14).

The Stata command `biprobit` maximizes the likelihood function in (3.15) numerically. The estimation results for the dependent variable *switched*, the endogenous variable *tariff calculator* and the instrumental variable *mobile tariff information* are displayed in Table 3.8.

The right part of Table 3.8 describes the first stage estimation where the *tariff calculator* is regressed on the instrument *mobile tariff information* and on the exogenous variables. The left part shows the estimation results for the second stage, where the estimation of the instrumented variable from the first stage is substituted for the endogenous variable *tariff calculator*. We see on the right side of the table that the instrument has, as required, a significant influence on the use of the *tariff calculator*. There is also some interesting side information: The usage of a tariff calculator seems to be more common in urban areas. Table 3.8 also reports a slightly higher usage rate for respondents with a higher concern for sustainability. Since the tariff calculator by the E-control also lists information which electricity supplier offers green energy, this observation makes sense.

Let us turn now to the key result, the estimated coefficients of the second stage regression. Our main interest is a comparison between the z-scores in the biprobit model (left side of Table 3.8) with the z-scores of the probit model from Section 3.5, where all explanatory variables are treated as exogenous. For a precise comparison, Table 3.3 is only restrictedly suitable, as the estimation includes the variables *mobile tariff information* and *mobile operator churn*, which we excluded in the biprobit model. So as to ensure comparability, we included Table 3.10 in Appendix 3.9.1, which displays the results of a probit model without the variables *mobile tariff information* and *mobile operator churn*[20].

The comparison between the coefficients in Table 3.9.1 and Table 3.8 provides the same

[20]The marginal probabilities in Table 3.10 coincide with the values listed in the middle column of Table 3.7.

Table 3.8: Biprobit model for dependent variable *switched*.

Dep. variable *switched*			**Dep. variable *tariff calculator***		
Variables	**Coeff.**	**St.Err.**	**Variables**	**Coeff.**	**St.Err.**
Constant	-2.506**	0.777	Constant	-2.423**	0.727
Tariff calculator	1.444	1.045	Mobile tariff inf.	0.852**	0.288
Expenditure	0.000	0.000	Expenditure	0.000	0.000
Property	-0.205	0.263	Property	-0.219	0.211
Living Space	0.002	0.004	Living Space	0.005	0.003
Sustainability	-0.052	0.223	Sustainability	0.290†	0.164
Price	0.072	0.235	Price	0.239	0.186
Popularity			Popularity		
Cat. "rather imp."	0.352	0.365	Cat. "rather imp."	0.047	0.301
Cat. "rather unimp."	1.011**	0.367	Cat. "rather unimp."	-0.040	0.301
Cat. "unimportant"	1.006*	0.413	Cat. "unimportant"	0.225	0.313
Age	0.005	0.009	Age	0.010	0.007
Gender	0.327	0.209	Gender	-0.255	0.167
Urban	0.885*	0.450	Urban	0.793**	0.207
Education			Education		
Cat. "medium"	-0.524*	0.254	Cat. "medium"	0.087	0.220
Cat. "high"	-0.542*	0.260	Cat. "high"	0.239	0.224

N	276
Log-likelihood	-304.708
$\chi^2_{(28)}$	85

Significance levels : † : 10% * : 5% ** : 1%

insights as in Section 3.6.6: The magnitude and the significance levels are very similar; the magnitude of the instrumented endogenous variable *tariff calculator* is slightly greater than in the exogenous model, but it is not significant any longer.

Again, we want to test if endogeneity is present after all. The Wald test, included in the Stata command `biprobit`, tests the null hypothesis that the endogeneity parameter ρ equals zero. For our model, the null hypothesis cannot be rejected ($\chi^2(1) = 0.074$, p-value = 0.786).

In conclusion, both estimation methods available for binary choice models with an endogenous dummy variable indicate that the variable *tariff calculator* can be treated as exogenous. Thus, the probit model of Section 3.5 is sufficient.

3.7 Propensity Score Matching

3.7.1 The Theoretical Framework of PSM Estimators

Before we present the estimation results, we give a brief outline of the propensity score matching estimator in the general evaluation framework, which follows Caliendo and Kopeinig (2005).

In this section, we choose another approach to measure the influence of the *tariff calculator* on the switching decision, namely propensity score matching. This methods allows us to measure the causal effect of one single treatment on a response by comparing the response of persons, who have been treated, to the response of persons, who are similar to the first-mentioned with respect to all other covariates, but have not been treated.

In terms of our data, the treatment $T_i, T_i \in \{0,1\}$ is the usage of a tariff calculator of the $i-$th individual, and the response y_i is his switching decision. What we would like to measure is the so called treatment effect τ_i,

$$\tau_i = y_i(1) - y_i(0). \tag{3.16}$$

The treatment effect measures the effect of the treatment on the response of the i-th individual, or the difference in the switching decision under tariff-calculator usage and non-usage respectively. The fundamental evaluation problem arises because only one of the potential outcomes is observed for each individual i. The unobserved outcome is called counter-factual outcome. As a consequence it is impossible to observe the individual treatment effect τ_i and one has to concentrate on (population) average treatment effects. The most popular measure in this respect is the average treatment effect on the treated (ATET), which is defined as follows,

$$ATET = \mathbb{E}[y(1) - y(0)|T = 1] = \mathbb{E}[y(1)|T = 1] - \mathbb{E}[y(0)|T = 1]. \tag{3.17}$$

The second part of equation (3.17) is counter-factual and cannot be observed. Thus, one has to find a proper substitute for $\mathbb{E}[y(0)|T = 1]$ so as to be able to estimate the ATET. Using simply the mean outcome of untreated individuals, $\mathbb{E}[y(0)|T = 0]$, is not advisable, because in non-randomized studies it is most likely that components which determine the treatment decision also influence the response (self-selection bias). Thus, we have to invoke some identifying assumptions to handle this self-selection problem.

One possible identification strategy is the conditional independence assumption (see also Wooldridge (2002); in Rosenbaum and Rubin (1983), it is called *ignorability of treatment*): Given a set of observable covariates $\mathbf{x}$, which are not affected by treatment, the potential responses are independent of treatment assignment,

$$y(0), y(1) \perp T \quad | \mathbf{x}, \quad \forall \mathbf{x}. \tag{3.18}$$

This condition is also called *unconfoundedness*. The implication of unconfoundedness is that selection is only based on observable characteristics and, further, that all variables are observed that influence treatment assignment and potential responses simultaneously (Caliendo and Kopeinig (2005)). Conditioning on all relevant covariates is limited in case of a high dimensional covariate-vector. Thus, Rosenbaum and Rubin (1983) came up with the idea to use a so-called balancing score instead. They showed that if (3.18) holds, this implies that potential outcomes are also independent of treatment assignment conditional on the *balancing score*. Rosenbaum and Rubin (1983) further show that the propensity score $ps(\mathbf{x}) = P(T = 1|\mathbf{x})$ is one possible balancing score, thus the independence assumption can be written as,

$$y(0), y(1) \perp T \quad | ps(\mathbf{x}), \quad \forall \mathbf{x}. \tag{3.19}$$

Both the probit and the logit function are common to estimate the propensity score function.

Another condition has to be fulfilled: Namely, the overlap condition, which ensures that persons with the same $\mathbf{x}$ values have a positive probability of being both in the treatment and the control group. Therefore, it rules out the phenomenon of perfect predictability of treatment given $\mathbf{x}$ (cf. Caliendo and Kopeinig (2005)). The overlap condition has been tested and is fulfilled for our data set.

If conditional independence if fulfilled, it follows that

$$\mathbb{E}[y(0)|T=0, ps(\mathbf{x})] = \mathbb{E}[y(0)|T=1, ps(\mathbf{x})]. \tag{3.20}$$

We have now all the parts of the PSM estimator for the ATET,

$$ATET = \mathbb{E}\{\mathbb{E}[y(1)|ps(\mathbf{x}), T=1] - \mathbb{E}[y(0)|ps(\mathbf{x}), T=0]\}. \tag{3.21}$$

3.7.2 The PSM Estimation Results

Statistical software packages (such as Stata) offer a wide range of PSM estimators, differing in matching algorithms and parameter constellations. All PSM estimators have in common that they contrast the resonse of a treated individual with the resonse(s) of (a) non-treated individual(s), who are characterised by the same or a similar propensity score $ps(\mathbf{x})$ based on the covariates $\mathbf{x}$. There exists a range of possibilities how to empirically calculate the PSM estimators. The matching algorithms to choose from differ in the way how similarity is defined, how many non-treatment group members are considered, how weights assigned to matched individuals are handled, and many other. Caliendo and Kopeinig (2005) give a nice overview of matching algorithms, options associated with each algorithm and the pro and cons of each variant.

The general structure of the empirically estimated ATT can be presented as follows,

$$ATET = \frac{1}{N} \sum_{i \in \{T=1\}} \left[y(1)_i - \sum_j w(i,j) y(0)_j \right], \tag{3.22}$$

where every treated observation i is matched with j control observations and their outcomes $y(0)_j$ are weighted by $w(i,j)$, and where N is the number of matched couples [21].

We report the results of four different matching algorithms, all implemented as options in the Stata command `psmatch2` by Leuven and Sianesi (2003). First, *one-to-one nearest neighbor matching (1:1-NN)* selects for every treated observation one observation from the control group, which minimizes the distance $|ps_i(\mathbf{x}) - ps_j(\mathbf{x})|$. In this straightforward case $j = 1$ and $w(i,1) = 1, \forall i$. Second, we report the results of *five-nearest neighbors matching (5-NN)*, where every treated observation is matched with five nearest neighbors. As the third algorithm, we choose *radius matching*, where each treated observation i is matched with control observations j that fall within a specified radius r, $|ps_i(\mathbf{x}) - ps_j(\mathbf{x})| < r$. We use the default radius $r = 0.1$. The fourth method, *kernel matching*, matches every treated observation i with several control observations j, with weights $w(i,j)$ inversely proportional to the distance between the propensity score of the treated ($ps_i(\mathbf{x})$) and of the control observations ($ps_j(\mathbf{x})$). For all four algorithms, we choose the default option "with replacement", so that each control observation can be used as a match to several treated observations. For the sake of comparability, we exclude, as before, the respondents who did know their expenditure level.

[21] For a more precise description of the empirical form of 3.22 under different matching algorithms see Becker and Ichono (2002).

[22] Note that the standard errors under the Stata command `psmatch2` do not take into account that the propensity score is estimated. The Stata command `teffects psmatch`, building on Abadie and Imbens (2011), can handle this problem, but allows only for nearest neighbor matching. The respective standard error of the 1:1-NN (5-NN) is 0.080 (0.065). `teffects psmatch` also allows the estimation of a 95% confidence interval, which is for 1:1-NN $[0.237, 0.548]$ and for 5-NN $[0.172, 0.426]$.

Table 3.9: The ATET for several different matching algorithms

	Treated	**Controls**	**Difference (ATET)**	**Standard Error**[22]
Unmatched	0.496	0.135	0.362	0.052
1:1 Nearest neighbor matching	0.496	0.104	0.393	0.063
5-Nearest neighbor matching	0.496	0.197	0.299	0.057
Radius matching	0.496	0.159	0.338	0.055
Kernel matching	0.496	0.168	0.325	0.056

In the first row in Table 3.9 we simply see the relative frequencies with no estimation or matching applied: Out of 135 individuals who used a tariff calculator ($T = 1$), 67 switched the electricity supplier (relative frequency of 0.496). Out of 141 non users ($T = 0$), 19 switched, giving a relative frequency of 0.135. The estimated ATET in case of 1:1NN-matching seems significantly higher than the other estimated ATETs. A reason may be the option "with replacement". Potentially, a few non-treated observations who do not switch have propensity scores very similar to a large part of the treated observations and were matched overproportionally often. Choosing the option "noreplacement" reduces the estimated ATET for the 1:1NN-matching to 0.363. The other matching algorithms lead to more similar estimations between 0.299 and 0.338.

Recalling the results from the probit model from Section 3.5, the marginal probability of the *tariff calculator* was increased from 11.2% for non-users to 46.9% for users, leading to a difference of 35.9%. In our case, the ATET has a similar interpretation: It is the average difference in the switching probability between matched individuals, whereby the matched individuals are similar with respect to all covariates, but differ in their tariff calculator usage. The majority of the estimated values for the ATET resulting from different matching algorithms lies slightly below the above mentioned marginal probability (0.299 − 0.338), only the 1:1NN-matching results in a higher value (0.393). As mentioned above, this deviation may be caused by the already described potential problems with 1:1NN-matching.

Summarizing, both approaches to measure the effect of information – a multivariate probit model and the PSM model – lead to similar, though not identical outcomes. In both approaches, the tariff calculator-users have a significantly higher probability of switching, also after accounting for all other explanatory variables. Furthermore, the results of the PSM indicate that the probit model slightly overrates the influence of the *tariff calculator*. However, there are several factors making a direct comparison between the two methods difficult. In case of the probit model, the influence of the means of the other explanatory variables and the potential endogeneity of the *tariff calculator* could influence the results. In case of the PSM model, it is possible that the required assumption of conditional independence is not fulfilled, and the wide range of algorithms, options and parameters to choose from – all resulting in slightly different outcomes – make the estimated ATETs seem a bit arbitrary.

3.8 Final Remarks

This paper investigated the question which Austrian households exercised the option to choose a new electricity supplier after the complete market opening in 2001 and why. Compared to other countries, the prevailing small switching rates in Austria are puzzling. The standard answer to low switching rates - namely that regulatory regimes can cast a long shadow over a liberal regime - is insufficient since this holds also for countries with more lively electricity supplier markets.

Having data on the switching behavior on the electricity supplier as well as on the mobile phone market, it was especially interesting to inquire if there is a group of customizers displaying a higher switching activity encompassing both markets. Our data did not support this hypothesis.

However, our data clearly reveals that well–informed customizers have a significantly higher switching probability. The usage of a tariff calculator turned out as the most effective variable explaining actual changes of ES. Potential issues of endogeneity between *switching* and the *tariff calculator* entailed – due to the binary nature of these two variables – a cumbersome econometric exploration, but did not confirm the presence of endogeneity.

Contrary to our expectations, we did not find any significant influence of the *expenditure level* on the actual switching decision nor on the intention to switch. However, the self-reported price sensitivity seems to be positively correlated with the intention to switch. An explanation for this could be that customers are not aware of the tariff structures and the associated saving potential.

It seems to be a promising strategy for competing ES, who offer low-priced contracts, to advertise more with the tariff and the savings potentials. This might induce more customers to switch than the popular advertising strategy to emphasize the "green energy"-aspect.

Since the data collection in summer 2012, the switching rates in Austria have slightly increased. For future research it would thus be very interesting to further investigate how much of this increase was driven by better information on the range of ES and their tariffs. Such an investigation could confirm our strong presumption that the recent increase in switching rates has been induced by the emergence of a range of easily accessible and heavily promoted price and tariff comparison tools.

Acknowledgements

The authors acknowledge helpful discussions with Profs. Jungsoo Lee, Oliver Fabel, Rudolf Vetschera and Marcus Hudec. The data was collected by Jaqueline Wolf in the course of her master thesis. The extensive literature review in the master thesis by Robin Douglas Johnston, covering switching rates in Portugal, was also very helpful.

Bibliography

Abadie, A., Imbens, G., 2011. Bias-corrected matching estimators for average treatment effects. Journal of Business & Economic Statistics 29 (1), 1–11.

Annala, S., Viljainen, S., Tuunanen, J., 2013. Rationality of supplier switching in retail electricity markets. International Journal of Energy Sector Management 7/4, 459–477.

Becker, S., Ichono, A., 2002. Estimation of average treatment effects based on propensity scores. The Stata Journal 2/4, 358–377.

Caliendo, M., Kopeinig, S., 2005. Some practical guidance for the implementation of propensity score matching. IZA Discussion paper 1588, 1–29.

Chamber of Labour, 2005. Wettbewerbsbericht, Teil 1, Vienna. Kammer für Arbeit und Angestellte.

Council of European Energy Regulators, 2012. 2012 annual report of the council of european energy regulators. http://www.ceer.eu/portal/page/portal/EER_HOME/EER_PUBLICATIONS/ANNUAL_REPORTS.

E-control, 2014. E-control: Bisher höchste Wechselzahlen im ersten Halbjahr. URL http://www.e-control.at/de/presse/aktuelle-meldungen/wechselzahlen_2013_final

E-Control Austria, 2012. Statistikbroschüre 2012. http://www.e-control.at/de/publikationen/statistik-bericht.

E-Control Austria, 2013. Statistikbroschüre 2013. http://www.e-control.at/de/publikationen/statistik-bericht.

Ek, K., Söderholm, P., 2008. Households' switching behavior between electricity suppliers in Sweden. Utilities Policy 16, 254–261.

Giulietti, M., Price, C., Waterson, M., 2005. Consumer choice and competition policy: a study of uk energy markets. The Economic Journal 115/506, 949–968.

Haberfellner, M., 2002. Liberalisierung und Regulierung des österreichischen Strommarktes, working Paper E-Control Ltd, http://www.e-control.at/de/publikationen/workingpaper/working-papers-archive#1354.

Hofbauer, I., 2006. Liberalisation, privatisation and regulation in the Austrian electricity sector, country report on liberalisation and privatisation processes and forms of regulation, Forschungs und Beratungsstelle Arbeitswelt, `http://www.forba.at/data/downloads/file/204-FB%2016-06%20PIQUE%20Electricity%20Austria.pdf`.

Hosmer, D., Lemeshow, S., 2000. Applied Logistic Regression. No. 2. John Wiley&Sons, Inc.

Klemperer, P., 1995. Competition when customers have switching costs: an overview with applications to industrial organization. Review of Economic Studies 62, 515–539.

Leuven, E., Sianesi, B., 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. URL `version4.0.10`

Lewbel, A., Dong, Y., Yang, T., 2012. Comparing features of convenient estimates for binary choice models with endogenous regressors. Canadian Journal of Economics 45/3, 809–829.

Miranda, A., Rabe-Hesketh, S., 2006. Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal and count variables. The Stata Journal 6/3, 285–308.

Rosenbaum, P. R., Rubin, D. B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.

Shah, B. R., Laupacis, A., Hux, J. E., Austin, P. C., 2005. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. Journal of Clinical Epidemiology 58 (6), 550 – 559.

Terza, J., Basu, A., Rathouz, P., 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. Journal of Health Economics 27/3, 531–543.

VaasaETT, 2013. European residential energy report 2013. URL `http://www.vaasaett.com/wp-content/uploads/2013/05/European-Residential-Energy-Price-Report-2013_Final1.pdf,accessed12August2014`

W., G., 2008. Econometric Analysis. Prentice Hall.

Wieringa, J., Verhoef, P., 2007. Understanding costumer switching behavior in a liberalizing service market. Journal of Service Research 10/2, 174–186.

Wilson, C., Waddams Price, C., 2010. Do consumers switch to the best supplier. Oxford Economic Papers 62/4, 647–668.

Winkelmann, R., 2011. Copula bivariate probit models: with an application to medical expenditures, working paper no.29.

Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press.

Wooldridge, J., 2010. Econometric Analysis of Cross Section and Panel Data. The MIT Press.

3.9 Appendix

3.9.1 Alternative Probit Model Ensuring Comparability

Table 3.10 displays the estimation results for a small modification of the probit model of Section 3.5 (Table 3.3): The variables *mobile operator churn* and *mobile tariff information* are excluded to ensure comparability with the biprobit model of Section 3.6.7 (Table 3.8).

Table 3.10: Probit estimation results for the dependent variable *switched*.

Variables	**Coeff.**	**Stand.Err.**	**Marg.Eff.**	**Stand.Err.**
Constant	-2.527**	0.763	–	–
Tariff calculator	1.143**	0.194	0.358**	0.055
Expenditure	-0.000	0.000	-0.000	0.000
Property	-0.238	0.232	-0.077	0.075
Living space	0.002	0.004	0.001	0.001
Sustainability	-0.017	0.187	-0.005	0.060
Price	0.099	0.215	0.032	0.069
Popularity of the ES				
Categ. "rather important"	0.352	0.368	0.068	0.064
Categ. "rather unimportant"	1.012**	0.367	0.272**	0.076
Categ. "unimportant"	1.043**	0.377	0.284**	0.083
Age	0.006	0.008	0.002	0.003
Gender	0.302	0.196	0.097	0.062
Urban	0.979**	0.249	0.314**	0.078
Education				
Category "medium"	-0.527*	0.253	-0.182*	0.089
Category "high"	-0.530*	0.261	-0.183*	0.092
N	276		276	
Log-likelihood	-130.537		–	
$\chi^2_{(14)}$	81.371		–	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Chapter 4

Optimal Pollution Management When Discount Rates are Endogenous

This paper investigates the consequences of endogenous discounting related to a stock externality, like 'global warming'. This topic is of interest, especially since the Stern report and the ensuing public and academic debate have made clear how crucial discounting is for evaluating anti-global warming measures. We investigate the consequences of the assumption that discounting is endogenous, more precisely, decreases as the damage increases, i.e., decision makers become more patient when facing the environmental damages following their high consumption levels. This avoids time inconsistency, a pitfall when using hyperbolic discounting. However, it can lead to quite different and sometimes counterintuitive behavior in comparison to standard exogenous discounting: an endogenous discount rate, albeit always smaller than the exogenous one, (i) can nevertheless imply high stationary damages; (ii) can lead to multiple steady states.

4.1 Introduction

The discussion about the 'correct' discount rate applicable to long-run issues like global warming has received substantial attention in recent years. At the center of the debate are the Stern report (c.f. Stern (2006)) and its critique, most prominently by Nordhaus (2007). The standard approach, which is defended in Nordhaus (2007), assumes a constant and positive discount rate to value the stream of future benefits and costs[1]. This has been criticized when decisions have long-run consequences affecting also future generations. The case of global warming is typical for this phenomenon, because a currently emitted ton of CO2 stays for 200 years (or even forever) in the atmosphere. Worse, induced global warming may trigger the release of

[1]The argument in Nordhaus (2007) is in fact a little bit more subtle than our argument below since the author differentiates between the pure rate of time preference - the focus of our paper - and the rate of return on capital, which depends in addition on growth and the intertemporal rate of substitution.

carbon bound in the earth's surface, e.g., methane from the tundra, that induces a 'positive' feedback on global warming, see Winter (2012). The problem is that discounted utilitarianism with empirically compatible discount rates gives too little weight to the benefit of generations far in the future. Beltratti et al. (1993) illustrate this by discounting the world GDP over 200 years at an annual interest rate of 5%, which leads to a present value smaller than a few hundred dollars. On the other hand, using a zero discount rate implies that present generations sacrifice their consumption for future, presumably much richer, generations.

Several approaches have been suggested. The seminal paper Strotz (1955) assumes that agents discount[2] hyperbolically: near future returns are discounted at higher rates than those in the distant future. This approach has been extended in Laibson (1997) and applied to environmental problems, e.g. Li and Löfgren (2000). The disadvantage is that it leads to time inconsistent policies: what is distant from today will become a short term decision in the future, calling for a revision of the original plan. The inability of governments to commit adds to this time inconsistency, which suggests that such public policy recommendations are not implementable. In the case of global warming, this means that politicians can optimally promise high and unchanged consumption today coupled with much less consumption in the future. Yet as time unravels, they will be unable to carry out this commitment, because they find the above mentioned policy, namely delaying conservation, optimal at any future date too. This characterization is confirmed empirically. After all, the US scrapped their cap and trade problem, the French President Hollande wanted to cap the price of gasoline in 2012, yet they all 'committed' in Copenhagen to do everything in order to stabilize future temperature increase at 2^oC. The snag is to commit to a plan and start reducing consumption when the short term needs dictate high consumption today and thereby postponing conservation.

Chichilnisky (1996) proposes a combination of the sum of discounted utilities over time (the standard net present value (NPV)) and the steady state solution. Alvarez-Cuadrado and Van Long (2009) propose a modified Bentham-Rawls criterion by maximizing as in Chichilnisky (1996) the NPV coupled with the payoff of the generation which is worst off. Both proposals render a time consistent policy. Weitzman (2001) argues that high discount rates are appropriate for instant needs but less so for decisions with a long planning horizon. More precisely, discount rates close to zero are appropriate for problems like climate change. The difference to hyperbolic discounting is that constant discount rates, that vary not over time but with planning horizons, are applied. This characterization, which Weitzman (2001) based on a survey, is intuitive and psychologically convincing considering our choice about instant needs (say food, a glass of water, the next cigarette) and about choices with a larger planning horizon (say a car, or a house). A pitfall of Chichilnisky's as well as Weitzman's model is the somehow ad–hoc choice of the parameters by the decision maker. Actually, Weitzman's approach is very similar to an earlier suggestion by Heal, well described e.g. in Heal (2005), to discount analogously how sound is measured in decibel, i.e., at an exponential scale (Weitzman (2001) uses the more general Γ -distribution instead of the exponential distribution implied by Heal's approach). In a recent paper, Weitzman (2012) links the discount rate to the stochastic future growth rate. In an optimal and sustainable growth model with intergenerational equity, Endress et al. (2013) introduce distinct inter- and intra-generational discount rates.

A further alternative is endogenous discounting which links the discount rate to the state of

[2]Here and in all the following discussion discounting refers to the pure rate of time preference.

the world, e.g., income (rich people may discount less[3]) or the environment. The major advantage is that the resulting optimal programs are time consistent and that it allows to account for concern about global warming damages faced by future generations in an endogenous way. This approach of endogenous discounting looks, at least at first sight, more promising than the above discussed exogenous and ad hoc discounting rules a la Strotz and Weitzman. Many of the related papers use a flow variable (consumption) to determine the rate of discount endogenously (Epstein and Hynes (1983), Obstfeld (1990), Das (2003)). More recent papers introduce stock variables as the source of endogenous discounting. (Schumacher, 2009, 2011) suppose that the discount rate declines with respect to the wealth of a decision maker. The wealthier somebody is, the more he can afford to be patient and postpone consumption. Schumacher (2009) shows that endogenous discounting via wealth can lead to two stable steady states, one with high wealth and high consumption level, and one in which decision makers are trapped with low wealth and low consumption levels. To the best of our knowledge, Schumacher (2011) is the first to emphasize the role of the domain of the felicity function. In the case of exogenous discounting (including the hyperbolic variants) the sign of the felicity function is irrelevant for the determination of steady states. However, in the case of endogenous discounting the sign can change the qualitative behavior of the steady state. Yanase (2011) extends a neoclassical growth model by introducing a time preference rate depending on both, flow (consumption) and the state (environment). This kind of endogenous discounting within a competitive equilibrium can (i) lead to indeterminacy and (ii) can reverse the intended effect of a tighter environmental policy. Tsur and Zemel (2009) account for the risk of a discontinuous drop in output due to climate change and emissions. This implies a related kind of endogenous discounting and that greenhouse gas emissions should terminate at a finite time.

Ayong Le Kama and Schubert (2007) consider an endogenous discount rate depending on the state of the environment in an optimal growth model. Assuming that the discount rate increases with environmental quality reflects a sustainability motive: the smaller the environmental quality is, the more pressing environmental questions become and the more patient decision makers are. Restricting to CRRA utility functions, they show that endogenous discounting can support sustainability, i.e., a steady state with positive consumption and environmental quality, whereas the boundary solution - zero consumption and environmental quality - results for constant discounting.

In this paper, we follow the approach outlined in Schumacher (2011), apply it to a standard stock externality problem, namely 'global warming', and compare it with alternatives. The basic assumption is that discounting decreases as the damage increases, which may reflect a sustainability motive similar to Ayong Le Kama and Schubert (2007): The higher the (public) bad and thus the damages, the more patient decision makers become by attaching more weight (i.e., applying a lower discount rate) to the future. We return in Section 4.6 to Ayong Le Kama and Schubert (2007) so as to explain the different implications.

We will show that the implicitly chosen domain of the payoff function (or normalization) and not only its derivatives influence the steady state, the stability, and the intertemporal strategies. In particular, a small change from an exogenous to an endogenous discount rate leads to radical and sometimes counterintuitive differences in the qualitative behavior. First, we recall that

[3]Of course rich people may also be rich due to their patience to begin with. Clark (2007) stresses this (genetic) link between subjective discounting and economic growth since the industrial revolution, at least for Britain.

under exogenous discounting, decreasing the discount rate always leads to a steady state in which the level of consumption and the level of the bad become smaller. Intuitively, giving more weight to the future leads to smaller tolerable levels of the bad, which in turn restricts possible consumption. Counterintuitively, we show that an endogenous discount rate, which is always smaller than the exogenous, constant rate, can lead to a steady state in which the level of the bad as well as the optimal consumption policies are both higher than in the exogenous case.

The next section describes the model. Section 4.3 introduces the reference case of constant, exogenous discounting. In Section 4.4, we derive the optimal policies, the steady state(s), and the dynamics. Examples are the content of Section 4.5. The paper ends with final remarks in Section 4.6.

4.2 The Model

The analysis uses a standard, simple, and familiar stock externality problem that is used in many papers, in particular with reference to global warming or with environmental pollution in general. For example, a simplified reduced form of the DICE model, introduced e.g. in Nordhaus (1992), was used in Hoel (1992) for global warming and by Dockner and van Long (1993) for a dynamic tragedy of the commons. It is the issue of global warming that has so far received the most concern about proper discounting, culminating in the debate about the Stern report as mentioned in the introduction. In our analysis below, the only change to the standard models is that discounting is neither constant, nor exogenous, nor hyperbolic, but endogenous.

An infinitely-lived decision maker benefits from consumption c and suffers from a stock variable T, representing the rising temperature, or more generally, pollution. Pollution is a stock externality that accumulates with current 'emissions' that are linked to consumption (say of fossil fuels). We assume, for simplicity, that the dependency is linear. Consumption is normalized in units of emissions and the stock of pollution depreciates at a constant rate (δ)[4].

The extension of the standard model is the introduction of the additional state Θ, that replaces the usual discount term (rt). Θ is endogenously determined by the discount function $f(T)$ and therefore endogenously by the evolution of the state T. It is assumed that the discount rate decreases with respect to the pollution stock ($f' < 0$), such that the decision maker becomes more patient when facing higher pollution. As long as there is no stock of pollution, the usual constant (and presumably) high discount rate applies. However, when damages appear, the decision maker becomes more sensitive and puts more weight on the future. The decision maker can perfectly foresee the aforementioned reaction.

Summarizing, the corresponding decision maker has to solve the following optimal control problem:

$$\max_{\{c(t)\geq 0\}} \int_0^\infty \left[e^{-\Theta(t)}(u(c(t)) - D(T(t))) \right] \mathrm{d}t, \tag{4.1}$$

[4]Note that nonlinear (inversely U-shaped) depreciation functions can lead directly to multiple steady states, compare the familiar shallow lake models, Mäler et al. (2003), and Wagener (2003); yet the much claimed non-concavity (or even local convexity) of the Hamiltonian is not necessary, see Wirl (2004), Wirl and Feichtinger (2005).

subject to

$$\dot{T}(t) = c(t) - \delta T(t), \forall t, \qquad \text{with } \delta > 0, \tag{4.2}$$
$$T(0) = T_0 = 0,$$

and

$$\dot{\Theta}(t) = f(T(t)). \tag{4.3}$$

The above model setup seems incomplete because no initial condition $\Theta(0) = \Theta_0$ is specified. Therefore and at least at a first glance this suggests that one is free to choose Θ_0. Then, however, the corresponding shadow price is 0 at any initial date and time inconsistency emerges through the back door: at any restart, at say $t' > 0$, we can choose again the Θ_0 which maximizes our objective, irrespective at which level $\Theta(t')$ is. If instead Θ_0 reflects initial discounting, then $\Theta(0) = 0$ seems a natural choice, since constant discounting implies $\Theta(t) = rt$. However, it turns out that the introduction of the state Θ is an artificial construct, the value of which is irrelevant; this holds in general and not only for the particular model. In contrast, the discount rate function f is crucial.

Assumption 1. *The felicity function $u : \mathbb{R}_+ \to \mathbb{R}$ is at least twice differentiable and has the standard properties of $u'(c) > 0, u''(c) < 0, \forall c$. Moreover, we assume the Inada conditions, $\lim_{c\to 0} u'(c) = \infty$, $\lim_{c\to\infty} u'(c) = 0$ and that $\lim_{c\to\infty} u(c)$ exists.*

Assumption 2. *The damage function $D : \mathbb{R}_+ \to \mathbb{R}_+$ is at least twice differentiable and convex: $D'(T) \geq 0, D''(T) > 0$. Moreover, we assume again Inada-type conditions, $D'(0) = 0$, and $\lim_{T\to\infty} D'(T) = \infty$.*

Assumption 3. *$f(0) = r > 0, f > 0$ and $f' < 0$, $\forall T$, and $\lim_{T\to\infty} f(T) = 0$.*

The parameter r is the maximal discount rate, serving as a reference in the standard case of constant discount rates.[5]

The above assumptions about utility and damage functions are standard. The Inada conditions ensure interior solutions. The assumption that utility is bounded is needed for a technical reason, the proof that optimal paths cannot diverge. This assumption is plausible if u is measured by consumer surplus and is satisfied by most of the standard utility functions including the ones used in our examples. The focus on the pollution accumulation process explains the choice of the initial condition of pollution at $T_0 = 0$ and $c \geq 0$ renders emissions irreversible. The assumptions about the discount function f reflect the motive addressed above that the decision maker reduces the discount rate if confronted with larger pollution. At first sight, the psychological description of a farsighted decision maker foreseeing the evolution of his future patience seems a little bit far fetched. However, there are some reasonable and confirming examples: an adolescent starting to smoke presumably 'knows' how this habit will lower his future patience, or conversely, those abstaining from drug use know that this decision will increase their future patience. Raising children, in particular installing 'middle class' values, amounts to a great deal of raising the child's level of patience as it grows up. Hence, ascribing

[5] For the examples in Section 4.5 we use the following discount function:

$$f(T) = re^{-bT}, \ r, b > 0.$$

such farsightedness about future patience to an already anyway omnipotent, omniscient, and benevolent decision maker does not seem implausible, at least on theoretical grounds and for our application.

4.3 The Reference Case of a Constant Discount Rate

For purposes of comparison, we recapitulate the usual case of constant discounting before we analyze the endogenous case.

Assuming a constant discount rate, $\rho = f(0) = r$, the optimal policy (following a unique saddlepoint path) can be explicitly characterized (the superscript ^{x} identifies this case of exogenous and constant discounting),

$$\dot{c}^x = \frac{u'}{u''}\left(r + \delta - \frac{D'}{u'}\right).$$

The exogenous steady state $\bar{T}^x$ is determined by the intuitive condition

$$u'\left(\delta \bar{T}^{\mathrm{x}}\right) = \frac{D'\left(\bar{T}^{\mathrm{x}}\right)}{r+\delta}, \tag{4.4}$$

i.e., the marginal utility from consumption (which equals pollution evaluated at a steady state, $\bar{c}^{\mathrm{x}} = \delta \bar{T}^{\mathrm{x}}$) must equal the NPV of the induced marginal damage.

4.4 Optimal Policies

4.4.1 Necessary Optimality Conditions

The present value Hamiltonian,

$$\mathcal{H} = (u(c) - D(T))e^{-\Theta} + \lambda\,(c - \delta T) - \mu f(T),$$

is time autonomous. The necessary optimality conditions are,

$$\mathcal{H}_c = \lambda + u' e^{-\Theta} = 0, \tag{4.5}$$

$$\mathcal{H}_{cc} = u'' e^{-\Theta} \leq 0, \tag{4.6}$$

$$\dot{\lambda} = D'(T)e^{-\Theta} + \lambda\delta + \mu f'(T), \tag{4.7}$$

$$\dot{\mu} = -\left(u(c) - D(T)\right)e^{-\Theta}, \tag{4.8}$$

$$\lim_{t\to\infty} \mathcal{H}(t) = 0. \tag{4.9}$$

A two-dimensional control problem leads to two costates, thus the canonical equations system is four-dimensional. The additional costate $\mu(t)$ represents the implicit value of relaxing the constraint (4.3) by one unit. Therefore, $\mu(t)$ equals the change in net present value (viewed from $t = 0$) due to an infinitesimal change of the discount rate. Integrating equation (4.8),

$$\mu(t) = \int_t^\infty \left(u(c(\tau)) - D(T(\tau))\right) e^{-\Theta(\tau)} \mathrm{d}\tau. \tag{4.10}$$

confirms this interpretation. The transversality condition (4.9) is taken from Michel (1982) and is also used in Schumacher (2009) and in Ayong Le Kama and Schubert (2007). Strictly speaking, Michel's theorem assumes a constant discount rate. We derive in the Appendix that (4.9) holds also for endogenous discounting based on the usual transversality conditions,

$$\begin{aligned} \lim_{t\to\infty} \lambda\left(t\right) T\left(t\right) &= 0, & (4.11)\\ \lim_{t\to\infty} \mu\left(t\right) \Theta\left(t\right) &= 0. & (4.12) \end{aligned}$$

Equation (4.10) determines the equality between the costate and the (maximal) net present value computed from any date t. Since (4.12) requires that the costate μ converges to zero, the integral on the right hand side of (4.10) does not only exist but must converge to zero. The sufficiency conditions, or the lack thereof, are addressed in 4.7.2.

4.4.2 Characteristics

Differentiating the Hamiltonian maximizing condition (4.5) with respect to time and replacing the costate, $\lambda = -u'(c)e^{-\Theta}$, yields the following system of dynamic equations,

$$\dot{c} = \frac{u'(c)}{u''(c)}\left(f(T) + \delta - \frac{D'(T)}{u'(c)}\right) - \frac{1}{u''(c)e^{-\Theta}}\mu f'(T), \tag{4.13}$$

$$\dot{T} = c - \delta T, \tag{4.14}$$

$$\dot{\mu} = -(u(c) - D(T))e^{-\Theta}. \tag{4.15}$$

Since the control problem is autonomous, it follows that $\dot{\mathcal{H}} = \mathcal{H}_t = 0$. Hence, the Hamiltonian is constant and moreover equal to zero due to the transversality condition (4.9). Therefore, $\mathcal{H} = 0, \forall t$, and one can subsequently eliminate the costate,

$$\mu = \frac{(u(c) - D(T) - u'(c)\dot{T})e^{-\Theta}}{f(T)}. \tag{4.16}$$

Substituting (4.16) into (4.13), we arrive at the following system of two equations,

$$\dot{c} = \frac{u'(c)}{u''(c)}\left(f(T) + \delta - \frac{D'(T)}{u'(c)}\right) - \frac{1}{u''(c)}\frac{f'(T)}{f(T)}\left(u(c) - D(T) - \dot{T}u'(c)\right), \tag{4.17}$$

$$\dot{T} = c - \delta T, \tag{4.18}$$

which characterizes the entire solution. This reduction of dimensionality from four to two is surprising. The reason is, as mentioned in Section 4.2, that the additional state Θ is only a helpful construct and arithmetically cancels out (and with it, its shadow price), whereas the discount function f remains crucial.

4.4.3 Steady State(s)

In Proposition 1, we will show that endogenous discounting can lead to higher stationary damages than the exogenous counterpart, but before, we give an economically intuitive (heuristic) explanation: Let us assume the integrand in (4.1) is positive, then there exists an additional or countervailing incentive to move into the domain of low discount rates in order to increase the NPV. Contrary to the case of a constant discount rate, the state-dependent discount rate can now be (indirectly) influenced. Since the discount rate decreases with the level of pollution, there exists an incentive to reach high levels of pollution. In contrast, in case of a negative integrand, a high discount rate lowers the NPV loss and thus, ceteris paribus, a movement towards high discount rates is advisable. This results in an incentive to reach even smaller levels of pollution. As a consequence, contrary to the von Neumann-Morgenstern framework[6], the sign of the integrand becomes important. This sign will also play a significant role in the following technical part.

Defining

$$G(T) := -u'(\delta T)(f(T) + \delta) + D'(T) + \frac{f'(T)}{f(T)}(u(\delta T) - D(T)), \tag{4.19}$$

any root $\bar{T}^{\mathrm{n}}$ of G is a steady state of the system (4.17) – (4.18). This condition for any steady state under endogenous discounting can be re-written in terms of marginal benefit and marginal damage,

$$u'(\delta \bar{T}^{\mathrm{n}}) = \frac{D'(\bar{T}^{\mathrm{n}})}{f(\bar{T}^{\mathrm{n}}) + \delta} + \frac{f'(\bar{T}^{\mathrm{n}})}{f(\bar{T}^{\mathrm{n}})} \frac{u(\delta \bar{T}^{\mathrm{n}}) - D(\bar{T}^{\mathrm{n}})}{f(\bar{T}^{\mathrm{n}}) + \delta}. \tag{4.20}$$

The first term on the right hand side (rhs) of (4.20) is the traditional NPV of the marginal damage using the discount rate at the stationary level of pollution. The second term accounts for the induced change by multiplying the NPV of the stationary payoff $u - D$ (using again the discount rate at the stationary level) with the percentage change in the discount function. The additional second term in (4.20) raises questions about the existence of steady states, how they compare with conventional discounting, and about their stability properties, which are, however, the subject of the next subsection.

Proposition 1. *Given the assumptions 1 – 3, the system (4.17) and (4.18) has at least one steady state. As expected, endogenous discounting can lead to more conservation $\bar{T}^n < \bar{T}^x$. Contrary to expectations, and albeit using an endogenous discount function always smaller than the corresponding exogenous discount rate, $f(T) < r$, $\forall T$, also the opposite outcome, $\bar{T}^n > \bar{T}^x$, is possible.*

Proof. The limits $\lim_{T \to 0} G(T) = -\infty$ and $\lim_{T \to \infty} G(T) > 0$, which hold due to the Inada conditions, the other Assumptions 1 – 3, and the (at least C^1) continuity of G, ensure the existence of at least one steady state.

The next crucial question is what are the implications of endogenous discounting compared to the standard case of constant and exogenous discounting characterized by (4.4). The left hand sides of (4.4) and (4.20) are identical, representing marginal benefit from consumption (at a steady state). The first term on the rhs of (4.20) is the net present value of the marginal damage and is thus qualitatively identical to endogenous discounting. This term differs from the exogenous case only because $f < r$ for all $T > 0$. This effect causes, ceteris paribus, a decrease in the steady states compared to the usual case. The graph of $D'(\delta T)/(f(T) + \delta)$, drawn in Figure 4.1 with small dashes, illustrates the effect of the first term.

The second term in (4.20), the NPV of the stationary payoff times f'/f, is the one that can trigger the counterintuitive outcomes of less conservation due to endogenous discounting and even of multiple equilibria. A negative stationary payoff,

$$\pi(T) := u(\delta T) - D(T) < 0, \tag{4.21}$$

[6]There, the sign of the integrand is completely irrelevant since objectives are only defined up to positive affine transformations.

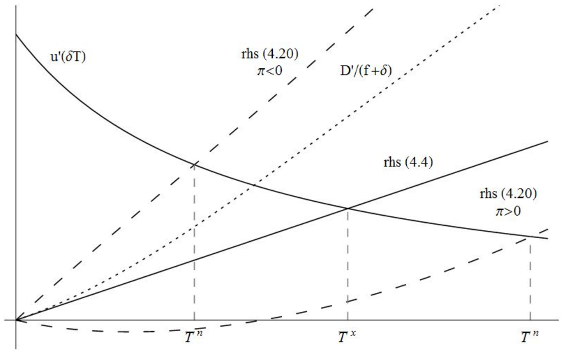


Figure 4.1: Marginal utility $u'(\delta T)$ together with the rhs of the steady state conditions in case of exogenous (cf. (4.4)) and endogenous (cf. (4.20)) discounting. The intersections determine the respective steady state. Additionally, the graph of the function $D'(\delta T)/(f(T)+\delta)$ is shown.

increases the rhs of (4.20) and thus implies a steady state that is smaller than the steady state reached under constant discounting. Therefore, negative payoffs at a steady state enhance the effect of the first term and ensure a fortiori less pollution (more conservation) compared with the usual case. Conversely, a positive payoff, $\pi(T) > 0$ for $T \in [\underline{T}, \overline{T}]$, lowers the rhs of (4.20) in this range. This diminishes and potentially even reverses the effect of lower discounting, $f < r$. The assumptions of concavity coupled with the Inada conditions imply that π is inversely U-shaped. Measuring benefits in terms of consumer surplus suggests positive stationary payoff $\pi(T) > 0$ at least for T small or more generally, for $T \in [\underline{T}, \overline{T}]$, and $\pi(T) < 0$ for $T > \overline{T}$, with a unique (static) maximum at $T^* : \delta u'(\delta T^*) = D'(T^*)$, $T^* \in [\underline{T}, \overline{T}]$. Therefore, depending on T^*, the rhs of (4.20) can be increased as well as decreased compared with the NPV of the marginal damage. This completes the proof of Proposition 1. $\square$

Figure 4.1 illustrates potential positions of steady states associated with the cases of exogenous and endogenous discounting by drawing $u'(\delta T)$ together with the rhs of (4.4) and two versions of (4.20), where the latter are drawn with long dashes. The two versions of (4.20) differ in the sign of the payoff (π) in (4.20). The curve with large dashes above all others represents a case where $\pi(T) < 0$, $\forall T$ and thus $\frac{f'}{f}(u - D) > 0$, $\forall T$, whereas the curve below all others requires a positive payoff, $\pi > 0$, for an interval $[\underline{T}, \overline{T}]$. Figure 4.1 illustrates that $\pi(T) > 0$, $T \in [\underline{T}, \overline{T}]$, can lead to a higher stationary pollution stock, $\bar{T}^n > \bar{T}^x$, whereas the case $\pi < 0, \forall T$, implies a more conservationist steady state $\bar{T}^n < \bar{T}^x$. It is worth noting that $\pi > 0$ for some interval $[\underline{T}, \overline{T}]$ is a necessary, but not a sufficient condition for $\bar{T}^n > \bar{T}^x$, whereas $\pi < 0, \forall T$, is a sufficient, but not a necessary condition for $\bar{T}^n < \bar{T}^x$. To summarize, the shift to the left (below exogenous discounting), equivalent to more conservation, is due to the combination of endogenous discounting and a negative payoff.

Switching from an exogenous discount rate to an endogenously determined discount function opens a wide area of potential, quite subtle problems. Schumacher (2011) emphasizes the role of the domain of the felicity function and comes to the conclusion that only negative felicity functions lead to intuitively correct steady states. The counterintuitively high steady states with a positive felicity function in our set up show that his argument falls short. From the steady state condition (4.20) we see that it is the interplay of the sign of the first derivative of the discount function, $f'(T)$, and the sign of the payoff $u(c) - D(T)$ (equivalent to the felicity function in Schumacher (2011)), which determines whether introducing a higher patience level leads to the expected result of less pollution (see also Remark 2 and 4.7.4). Furthermore, varying the discount function parameter b can lead to shifts in both directions, $\bar{T}^n \lesseqgtr \bar{T}^x$, for an example see 4.7.3. And finally, even if the higher patience leads to a smaller steady state level of pollution, $\bar{T}^n < \bar{T}^x$, one has to ask how much of this decrease is actually "environmentally" driven.

Corollary 2. *1. A necessary condition for multiple steady states is that $u(\delta T) < D(T)$ for $T \to 0$.*

2. Only an uneven number of steady states can exist.

Proof. Multiple steady states occur iff $G(T)$ in (4.19) has several roots, or put differently, if $u'(\delta T)(f(T)+\delta)$ and $D'(T) + f'/f(u(\delta T) - D(T))$ have several intersections. We have that $\lim_{c\to 0} u'(c) = \infty$ and $u'' < 0$, thus, multiple intersections can only occur, if $D'(T) + f'/f(u(\delta T) - D(T))$ is a falling function for a range of values of T. Since $D'(T)$ is monotonically increasing, and since $u(\delta T) - D(T)$ is inversely U-shaped, multiple intersections can only happen if $\lim_{T\to 0} u(\delta T) - D(T) < 0$.

The second part follows immediately from examining the function $G(T)$ in (4.19) under Assumptions 1 - 3. □

As in the case of a single steady state, all conceivable positions of the steady states under endogenous ($\bar{T}^{\mathrm{n}}_j, j = 1, 2\dots$) relative to exogenous discounting ($\bar{T}^{\mathrm{x}}$) are possible: $\bar{T}^{\mathrm{n}}_j \lesseqgtr \bar{T}^{\mathrm{x}}$, $j = 1, 2, 3, \dots$, as well as $\bar{T}^{\mathrm{n}}_i < \bar{T}^{\mathrm{x}} < Tn_j, j \neq i$. Using the same argument as for a single steady state, we see that $\pi(T) > 0$ for $T \in [\underline{T}, \overline{T}]$ is a necessary but not sufficient condition such that at least one steady state $\bar{T}^{\mathrm{n}}_j$ under endogenous discounting has a higher value than the corresponding steady state $\bar{T}^{\mathrm{x}}$, $\bar{T}^{\mathrm{n}}_j > \bar{T}^{\mathrm{x}}$.

4.4.4 Dynamics

Linearizing the dynamic system around a steady state, we use the following Jacobi matrix,

$$J = \begin{bmatrix} \frac{\partial \dot{c}}{\partial c} & \frac{\partial \dot{c}}{\partial T} \\ \\ 1 & -\delta \end{bmatrix}, \tag{4.22}$$

with

$$\frac{\partial \dot{c}}{\partial c} \mid_{\dot{c}=0,\dot{T}=0} = -u''(c)(f(T)+\delta) + \frac{f'(T)}{f(T)} u'(c) \tag{4.23}$$

$$\frac{\partial \dot{c}}{\partial T} \mid_{\dot{c}=0,\dot{T}=0} = D''(T) - u'(c)f'(T) - \frac{f'(T)}{f(T)} D'(T) \tag{4.24}$$

$$+ \left(\frac{f''(T)}{f(T)} - \frac{f'(T)^2}{f(T)^2} \right) (u(c) - D(T)). \tag{4.25}$$

Lemma 3.

The determinant of the Jacobian is linked to the slope of G, defined in (4.19), at a steady state as follows,

$$Det(J) \mid_{\dot{c}=0,\dot{T}=0} = -G'(\bar{T}^n).$$

Proposition 4. *Using the definition of G introduced in (4.19) again, then*

$$G'(\bar{T}^n) > 0 \Leftrightarrow \bar{T}^n \textit{ is saddlepoint stable.}$$

In other words, saddlepoint stability is equivalent to G crossing the abscissa from below.

Proof. Using Lemma 3 and the fact that a steady state is saddlepoint stable iff $Det(J) < 0$ proves the proposition. □

Proposition 5. *In case of a unique steady state, this steady state is saddlepoint stable. In case of multiple steady states, we enumerate them in ascending order starting with index* 1*. Then, all steady states with an uneven index are saddlepoint stable.*

Proof. The properties of G, in particular continuity and the limits, $\lim_{T\to 0} G(T) = -\infty$ and $\lim_{T\to 0} G(T) > 0$ (as already used in Section 4.4.3 to prove the existence of at least one steady state), determine also the stability property of the steady state(s). At the steady state with index 1, G cuts the abscissa from below and thus $G' > 0 \iff Det\,(J) < 0$; by continuity the second intersection with the abscissa, if existing, must be from above, thus $G' < 0$, and so on. □

Remark 1. The stability property of (a) steady state(s) with an even index is not clear since $Det\,(J) > 0$ per se is inconclusive. This is surprising, because in a standard optimal control model $Det\,(J) > 0$ is sufficient for instability. The reason is that there, both eigenvalues are positive (or have positive real parts), since the trace of the Jacobian equals the discount rate and is thus positive (the eigenvalues are symmetric around half of the discount rate). However, in this extended case it is not clear that the trace is positive. Therefore, an unstable steady state of the system (4.17) and (4.18) is only given iff

$$Det\,(J) > 0 \wedge \frac{\partial \dot{c}}{\partial c} - \delta > 0. \tag{4.26}$$

This is fulfilled in all our examples and holds presumably in general. Moreover, it confirms intuition since otherwise the necessary optimality conditions do not allow to determine the optimal one if all paths converged to the same steady state. However, arithmetically, the sign of the trace of the Jacobian in (4.22) remains ambiguous. This difficulty to sign the trace is due to the non-monotonicity of the isocline $\dot{c} = 0$, occurring even in simple standard examples such as a CRRA utility function (see e.g. Figure 4.4).

Remark 2. Even if we dropped $f'(T) < 0$ and assumed $f'(T) > 0$ instead as done in Yanase (2011), this would not change our (technical) results. More precisely, under certain parameter constellations, $\bar{T}^n \lesseqgtr \bar{T}^x$ as well as multiple steady states are possible. For more details and an example proving this claim, see Appendix 4.7.4.

4.5 Examples

The major objective of the following examples is to highlight unexpected consequences of endogenous discounting on the steady state(s) and on the dynamics. In the first example, we show that endogenous discounting can indeed lead to a steady state with higher damages and higher consumption than in the case with a constant and higher discount rate. In the second

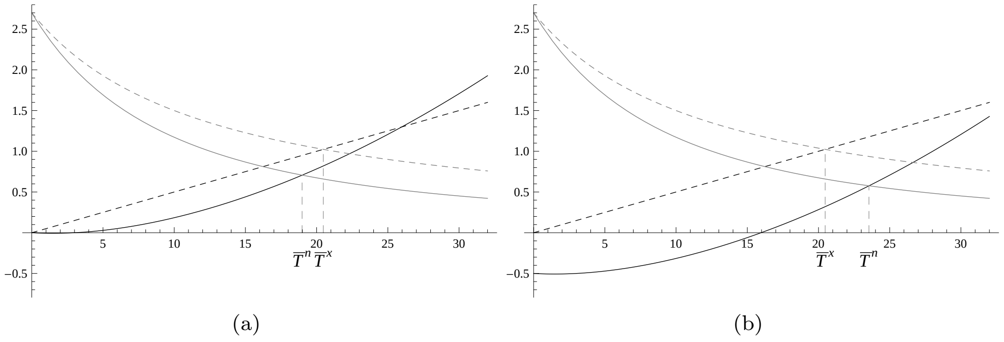


Figure 4.2: L^x, L^n (gray) and R^x, R^n (black) in case of exogenous discounting (dashed lines), as well as in case of endogenous discounting (solid lines), using different values for the parameter d. $a = 15, r = 0.1, \delta = 0.08, k = 0.05, b = 0.05, d = 0$ (a), $d = 10$ (b), respectively.

example, we prove the existence of multiple steady states due to endogenous discounting using a CRRA utility function.

For illustration and abbreviation purposes, we multiply the steady state conditions (4.4) and(4.20) by $(r + \delta)$ and $(f(\bar{T}) + \delta)$, respectively. The thereby resulting rearranged left hand sides of (4.4) and (4.20) are then labeled and defined as follows; $L^x = u'\left(\delta\bar{T}^{\mathrm{x}}\right)(\delta + r)$ and $L^n = u'\left(\delta\bar{T}^{\mathrm{n}}\right)(\delta + f(\bar{T}^{\mathrm{n}}))$ respectively. Likewise, the rearranged right hand sides are labeled and defined as $R^x = D'(\bar{T}^{\mathrm{x}})$ and $R^n = D'(\bar{T}^{\mathrm{n}}) + \frac{f'(\bar{T}^{\mathrm{n}})}{f(\bar{T}^{\mathrm{n}})}(u(\delta\bar{T}^{\mathrm{n}}) - D(\bar{T}^{\mathrm{n}}))$, respectively, where x stands for the exogenous and n for the endogenous case. The intersection of L^n and R^n determines the endogenous steady state $\bar{T}^{\mathrm{n}}$. Likewise, the intersection of L^x and R^x determines the exogenous steady state $\bar{T}^{\mathrm{x}}$.

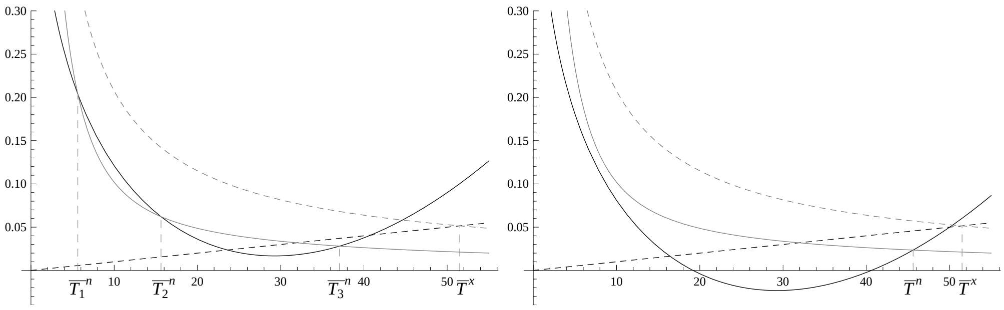
4.5.1 Unique Steady State with High Damages and High Consumption

Logarithmic utility, quadratic damages, and an exponential discount function serve as our first example that meets all Assumptions 1 – 3 except one of the two Inada-conditions,

$$u(c) = d + a\ln(1 + c), \quad D(T) = \frac{k}{2}T^2, \quad f(T) = re^{-bT};$$

$a > 0, k > 0, r > 0, b > 0$ and d are the model parameters.

The dashed lines in both plots in Figure 4.2 show the functions L^x and R^x for the exogenous discounting case, identical in both sides (Figure 4.2 (a) and (b)), whereas the solid lines illustrate the functions L^n and R^n for endogenous discounting. Varying the constant d shifts the steady state level $\bar{T}^{\mathrm{n}}$ either above (Figure 4.2 (b)) or beneath (Figure 4.2 (a)) the level $\bar{T}^{\mathrm{x}}$. Since $\bar{c} = \delta\bar{T}$ at the steady state, a larger pollution level is equivalent to a higher consumption level. Thus we have shown the existence of an endogenous steady state with higher consumption and higher pollution although the value of the discount function $f(T)$ is always smaller than the corresponding exogenous discount rate r.



(a) (b)

Figure 4.3: L^x, L^n(gray) and R^x, R^n (black) in case of exogenous discounting (dashed lines), as well as in case of endogenous discounting (solid lines), with varying parameter d. $\alpha = 0.85, r = 0.05, \delta = 0.035, k = 0.001, b = 0.2, d = -6.2$ (a), $d = -6.0$ (b), respectively.

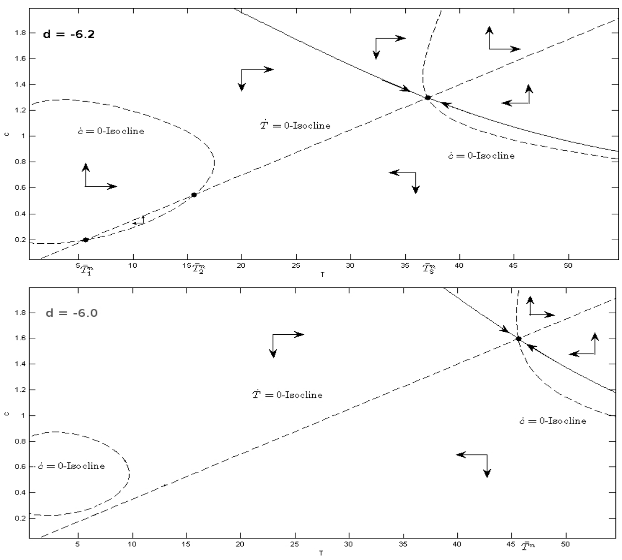


Figure 4.4: (T, c)-phase diagrams for the case of endogenous discounting, differing in the parameter d, functions and parameters as in Figure 4.3.

4.5.2 Multiple Steady States

We choose the familiar CRRA utility function, quadratic damages, and an exponential discount function, satisfying all Assumptions 1 – 3:

$$u(c) = d + \frac{1}{1-\alpha} c^{1-\alpha}, \quad D(T) = \frac{k}{2} T^2, \quad f(T) = re^{-bT},$$

where $0 < \alpha < 1, k > 0, r > 0, b > 0$ and d are the model parameters.

Figure 4.3 and Figure 4.4 illustrate the possible, but not necessary occurrence of multiple steady states when $D(0) > u(0)$. Whereas an intercept of $d = -6.0$ leads to a unique steady state (see Figure 4.3 (b)), the small change to $d = -6.2$ triggers multiple equilibria (see Figure 4.3 (a)).

4.6 Concluding Remark

We applied endogenous (more precisely state contingent) discounting to the problem of global warming, assuming that the discount rate declines as the damage increases. The objective is to investigate whether this reduction of discount rates over time induces a conservationist attitude (relative to a constant discount rate) without the pitfalls of hyperbolic discounting. In contrast to hyperbolic discounting, this approach leads to time consistent policies and moreover seems plausible from a psychological perspective: it represents the increasing patience of decision makers when they are confronted with the consequences of too much pollution.

However, the expectation that this fosters conservation does not necessarily materialize. A small shift from an exogenous, constant discount rate to a state-dependent discount rate can lead to unexpected and at least at first sight counterintuitive optimal policies. The sign of the instantaneous net payoff (benefit minus costs), which plays no role within the traditional von Neumann-Morgenstern framework, can determine whether it has the desired conservational effect. As a result, high levels of pollution exceeding their counterpart under constant discounting can be 'optimal'. Therefore, endogenous discounting does not represent – contrary to Ayong Le Kama and Schubert (2007) – a silver bullet for protecting the environment. Since the work by Ayong Le Kama and Schubert (2007) is similar in spirit – the discount rate decreases when environmental quality is low – but results in almost diametrically opposed conclusions, it is worth to have a closer look at the reasons thereof. Whereas we introduce a depreciating public bad (at rate δ), Ayong Le Kama and Schubert (2007) work with a recovering public good (at rate m), which enables growth in their model. In absolute terms, depreciation of pollution increases with the level of pollution stock in our model, whereas in Ayong Le Kama and Schubert (2007), regeneration grows with the level of the public good, giving another incentive to reach high levels of the public good. Ayong Le Kama and Schubert (2007) restrict their model to a utility function of the CRRA-type with a negative domain. Thereby, under exogenous discounting, a positive steady state is only possible for a razor-edge case, whereas endogenous discounting allows for a generic, positive steady state as long as the discount function lies within certain bounds. Ayong Le Kama and Schubert (2007) interpret this as a stabilization of the environmental quality due to endogenous discounting. Therefore, their specifications rule out that endogenous discounting can also lead to a steady state environmentally inferior compared to the exogenous case but this holds not in general as proven in this paper.

Furthermore, endogenous discounting can trigger multiple steady states, which can be ruled out under exogenous discounting (and our assumptions of a strictly concave and additive payoff function and a simple stock externality).

Aside from the possible counterintuitive effects from higher patience, the assumption of patience increasing with environmental awareness may be restricted to rich societies. Thus, a natural extension of our model would be to make the discount rate dependent on both, the pollution and the wealth level of a decision maker. This extension increases the state space, which makes the analysis of the system less manageable and is thus left for future research.

Bibliography

Alvarez-Cuadrado, F., Van Long, N., 2009. A mixed Bentham-Rawls criterion for intergenerational equity: Theory and implications. Journal of Environmental Economics and Management 58/2, 154 – 168.

Ayong Le Kama, A., Schubert, K., 2007. A note on the consequences of endogenous discounting depending on environmental quality. Macroeconomic Dynamics 11/2, 272–289.

Beltratti, A., Chichilnisky, G., Heal, G., 1993. Sustainable growth and the green golden rule. NBER working paper series 4430, 1–35.

Chichilnisky, G., 1996. An axiomatic approach to sustainable development. Social Choice and Welfare 13, 231–257.

Clark, G., 2007. A Farewell to Alms: A Brief Economic History of the World. Princeton University Press.

Das, M., 2003. Optimal growth with decreasing marginal impatience. Journal of Economic Dynamics & Control 27, 1881–1898.

Dockner, E., van Long, N., 1993. International pollution control: Cooperative versus non-cooperative strategies,. Journal of Environmental Economics and Management 25, 13–29.

Endress, L., Pongkijvorasin, S., Roumasset, J., Wada, C., 2013. Intergenerational equity with individual impatience in a model of optimal and sustainable growth. Environmental and Engery Economics preprint, xxx–xxx.

Epstein, L., Hynes, J., 1983. The rate of time preference and dynamic economic analysis. The Journal of Political Economy 91/4, 611–635.

Heal, G., 2005. Intertemporal welfare economics and the environment. In: Mäler, K., Vincent, J. (Eds.), Handbook of Environmental Economics. Elsevier B., pp. 1105–1145.

Hoel, M., 1992. Emission taxes in a dynamic international game of CO_2 emissions. In: Pethig, R. (Ed.), Conflicts and Cooperation in Managing Environmental Resources. Springer, Berlin-Heidelberg-New York-Tokyo, pp. 39–68.

Laibson, D., 1997. Golden eggs and hyberbolic discounting. The Quaterly Journal of Economics 112/2, 443–477.

Li, C., Löfgren, K., 2000. Economic growth, environmental quality and hyperbolic discounting. Dalarna University College Electronic Archive - DALEA, 1–22.

Mäler, K., Xepapadeas, A., de Zeeuw, A., 2003. The economics of shallow lakes. Environmental and Resource Economics 26, 603–624.

Michel, P., 1982. On the transversality condition in infinite horizon optimal problems. Econometrica 50/4, 975–985.

Nordhaus, W., 1992. An optimal transition path for slowing climate change. Science 258, 1315–1319.

Nordhaus, W., 2007. A review of the stern review on the economics of climate change. Journal of Economic Literature XLV, 686–702.

Obstfeld, M., 1990. Intertemporal dependence, impatience and dynamics. Journal of Monetary Economics 26, 45–75.

Schumacher, I., 2009. Endogenous discounting via wealth, twin-peaks and the role of technology. Economic Lettres 203, 78–80.

Schumacher, I., 2011. Endogenous discounting and the domain of the felicity function. Economic Modelling 28, 574–581.

Stern, N., 2006. Stern review on the economics of climate change. ARROW Discovery Service [http://search.arrow.edu.au/apps/ArrowUI/OAIHandler] (Australia).
URL `http://apo.org.au/?q=node/4420`

Strotz, R., 1955. Myopia and inconsistency in dynamic utility maximization. The Review of Economic Studies 23/3, 165–180.

Tsur, Y., Zemel, A., 2009. Endogenous discounting and climate policy. Environmental & Resource Economics 44, 507–520.

Wagener, F., 2003. Skiba points and heteroclinic bifurcations, with applications to the shallow lake system. Journal of Economic Dynamics and Control 27, 1533–1561.

Weitzman, M., 2001. Gamma discounting. The American Economic Review 91/1, 260–271.

Weitzman, M., 2012. The Ramsey discounting formula for a hidden-state stochastic growth process. Environmental Resource Economics 53, 209–321.

Winter, R., 2012. Innovation and the dynamics of global warming.

Wirl, F., 2004. Sustainable growth, renewable resources and pollution: thresholds and cycles. Journal of Economic Dynamics and Control 28, 1149–1157.

Wirl, F., Feichtinger, G., 2005. History dependence in concave economies. Journal of Economic Behavior and Organization 57, 390–407.

Yanase, A., 2011. Impatience, pollution, and indeterminacy. Journal of Economic Dynamics & Control 35, 1789–1799.

4.7 Appendix

4.7.1 Transversality Conditions in Case of Endogenous Discounting.

Strictly speaking, Michel's transversality condition (4.9) was only proven for a constant discount rate. To show that it is also applicable for endogenous discounting, we proceed in two steps. First, we show in Lemma 6 that pollution must remain finite along an optimal path. It is Lemma 6, where the assumption $\lim_{c\to\infty} u(c) < \infty$ is needed. Lemma 6 is then used in Proposition 7 to derive that $\lim_{t\to\infty} \mathcal{H}(t) = 0$ holds also in our case of endogenous discounting.

Lemma 6. *Given the assumptions about costs, benefits and optimal paths of the control problem (4.1)–(4.3), it follows that* $\underset{t\to\infty}{T}(t) < \infty$, *i.e., pollution cannot grow beyond any bound and must converge to a finite steady state.*

The economic intuition behind this lemma is obvious from the Inada conditions since divergence of T means that consumption increases forever. This adds only negligible incremental benefit ($u' \to 0$) but causes infinite undiscounted ($f' \to 0$) future damages ($D'/\delta \to \infty$). However, the arithmetic is slightly more complicated and it is the proof of this lemma that uses the assumption of a bounded utility.

Proof. Indirectly, assuming $T \to \infty$ requires $c > \delta T$ at least for T sufficiently large since otherwise a steady state will be reached. Therefore, $c \to \infty$ such that,

$$\lim_{t\to\infty} - e^{\Theta(t)}\lambda(t) = \lim_{t\to\infty} u'(c(t)) = 0, \tag{4.27}$$

due to the maximum principle (4.5). Hence,

$$\lim_{t\to\infty} \frac{\lambda}{e^{-\Theta}} \to 0. \tag{4.28}$$

If $\lim_{t\to\infty} \Theta < \infty$, then the costate differential equation (4.7) has the limit

$$\dot{\lambda} \to D' e^{-\Theta(t)} \to \infty$$

due to $\lambda \to 0$ and $\mu \to 0$ from the corresponding transversality conditions (4.11) and (4.12). This contradicts directly (4.28).

Therefore, $\Theta \to \infty$ such that the limit in (4.28) is of the 0/0-type since $\lambda \to 0$ due to (4.11). Application of the rule of De L'Hospital and substitution of (4.7) yields,

$$\lim_{t\to\infty} \frac{\lambda}{e^{-\Theta}} = \lim_{t\to\infty} - \frac{\dot{\lambda}}{e^{-\Theta} f} = \lim_{t\to\infty} - \frac{D' + e^{\Theta}\lambda\delta + e^{\Theta}\mu f'}{f}. \tag{4.29}$$

The denominator goes to zero: $f \to 0$ for $T \to \infty$, while the first term of the numerator, the marginal damage goes to infinity. The second term of the numerator goes to zero due to (4.27). Therefore the minimum requirement is $e^{\Theta}\mu f' \to -\infty$ such that the numerator converges to zero as the denominator. Since $f' \to 0$ for $T \to \infty$, the crucial term is the current value of the costate (using (4.10)),

$$e^{\Theta}\mu = \int_t^\infty (u(c(\tau)) - D(T(\tau)))\, e^{-(\Theta(\tau)-\Theta(t))} \mathrm{d}\tau < 0$$

Choosing t and thus T sufficiently large, the integrand is definitely negative since u is bounded but $D \to \infty$. Therefore the net present value of benefits minus costs discounted (at a discount rate close to zero, $\Theta(\tau) - \Theta(t) \leq f(t)(\tau - t)$ and $f(t) \approx 0$) to period t is negative. Hence, $\lim_{t\to\infty} e^{\Theta} \mu f' \geq 0$. Thus, the ratio on the right hand side of (4.29) diverges to minus infinity contradicting (4.27). $\square$

Proposition 7. *The transversality condition* $\lim_{t\to\infty} \mathcal{H}(t) = 0$ *holds also in case of endogenous discounting.*

Proof. Assuming that all initial values are given, thus also Θ_0, the only corresponding transversality conditions are those as stated in (4.11)–(4.12). If Θ_0 can be chosen freely, then the additional boundary condition, $\mu(0) = 0$, applies. If Θ_0 can be chosen within boundaries, e.g. $\Theta_0 \geq 0$, then additional Kuhn-Tucker multipliers arise.

From Lemma 6 we know that $\bar{T}^{\mathrm{n}} < \infty$ and thus $\bar{c}^{n} = \delta\bar{T}^{\mathrm{n}} < \infty$, then from the maximum principle (4.5),

$$\begin{aligned} \lim_{t\to\infty} \lambda(t) &= -u'\left(\delta\bar{T}^{\mathrm{n}}\right) e^{-\lim_{t\to\infty}\Theta(t)} \\ &\Longrightarrow \lim_{t\to\infty} \lambda(t)\, T(t) = -\bar{T}^{\mathrm{n}} u'\left(\delta\bar{T}^{\mathrm{n}}\right) e^{-\lim_{t\to\infty}\Theta(t)} \end{aligned}$$

and the limiting transversality conditions (4.11) holds iff

$$\lim_{t\to\infty} \Theta(t) = \infty, \tag{4.30}$$

which indeed follows from integrating the state equation,

$$\dot{\Theta}(t) = f(T(t)) \geq 0.$$

Combining (4.12) and (4.30) implies,

$$\lim_{t\to\infty} \mu(t) = 0. \tag{4.31}$$

Using all these transversality conditions, one obtains for the Hamiltonian,

$$\lim_{t\to\infty} \mathcal{H}(t) = (u - D) \underbrace{\lim_{t\to\infty} e^{-\Theta}}_{0} + \lambda \underbrace{[c - \delta T]}_{0} + \underbrace{\lim_{t\to\infty} \mu f}_{0} = 0 \tag{4.32}$$

$\square$

This shows that Schumacher's only transversality condition is an implication of the usual transversality conditions combined with the optimality conditions.

The problem is autonomous, thus $\mathcal{H}_t = 0$. Hence, knowing a single value of $\mathcal{H}$ determines the entire Hamiltonian for all t. Using the above limit (4.32) implies therefore,

$$\mathcal{H}(t) = 0, \forall t.$$

This result can be used to determine μ along an optimal path,

$$\begin{aligned} \mathcal{H} &= (u-D)e^{-\Theta} + \lambda\left[c-\delta T\right] + \mu f = 0 \\ \Longrightarrow & \quad \mu = -\frac{(u-D)e^{-\Theta} + \lambda\left[c-\delta T\right]}{f}, \end{aligned} \tag{4.33}$$

which in turn eliminates this costate.

4.7.2 Sufficiency Conditions

In order to guarantee that the necessary conditions are also sufficient, the standard procedure is to show that the Hamiltonian is a concave function in the control and the state variables. This implies a 3×3 Hessian matrix, and that the sufficient, but not necessary, condition for concavity is that the principal minors M_1, M_2, and M_3 have to alter in sign, starting with a negative determinant. In our set–up, the version with an exogenous discount rate fulfills this principal-minor-condition. In contrast, the version with an endogenous discount rate fails to guarantee concavity when using the principal minor criteria.

In case of an endogenous discount rate, the Hessian matrix M for the Hamiltonian

$$\mathcal{H} = (u(c) - D(T))e^{-\Theta} + \lambda\left[c - \delta T\right] - \mu f(T)$$

has the following form,

$$\begin{aligned} M &= \begin{bmatrix} \frac{\partial^2\mathcal{H}}{\partial^2 c} & \frac{\partial^2\mathcal{H}}{\partial c\partial T} & \frac{\partial^2\mathcal{H}}{\partial c\partial\Theta} \\ \frac{\partial^2\mathcal{H}}{\partial T\partial c} & \frac{\partial^2\mathcal{H}}{\partial^2 T} & \frac{\partial^2\mathcal{H}}{\partial T\partial\Theta} \\ \frac{\partial^2\mathcal{H}}{\partial\Theta\partial c} & \frac{\partial^2\mathcal{H}}{\partial\Theta\partial T} & \frac{\partial^2\mathcal{H}}{\partial^2\Theta} \end{bmatrix}, \\ &= \begin{bmatrix} u''(c)e^{-\Theta} & 0 & -u'(c)e^{-\Theta} \\ 0 & -D''(T)e^{-\Theta} - \mu f''(T) & D'(T)e^{-\Theta} \\ -u'(c)e^{-\Theta} & +D'(T)e^{-\Theta} & (u(c)-D(T))e^{-\Theta} \end{bmatrix}. \end{aligned}$$

The principal minors are given by

$$\begin{aligned} |M_1| &= u''(c)e^{-\Theta}, \\ |M_2| &= u''(c)e^{-\Theta}\left(-D''(T)e^{-\Theta} - \mu f''(T)\right), \\ |M_3| &= u''(c)e^{-\Theta}\left[\left(-D''(T)e^{-\Theta} - \mu f''(T)\right)\left(u(c) - D(T)\right)e^{-\Theta} - D'(T)^2 e^{-2\Theta}\right] \\ &\quad - u'(c)e^{-\Theta}\left[u'(c)e^{-\Theta}\left(-D''(T)e^{-\Theta} - \mu f''(T)\right)\right]. \end{aligned}$$

Due to the concavity of the utility function, we have that $|M_1| < 0$. The sign of $|M_2|$ depends on the sign of the costate μ, which in turn depends on the sign of π due to equation (4.10).

There is only one simple way to determine the sign of μ without knowing the specific functions, namely to use a negative utility function. In this case, π is always negative, thus $\mu < 0$. Assuming for the moment a negative μ, we have to make the further assumption $-D''(T)e^{-\Theta} < \mu f''(T)$ so as to guarantee that $|M_2|$ is positive. Further, for the most part, very strong assumptions are necessary to make sure that $|M_3| < 0$.

Also if μ were always positive, we would nevertheless run into problems. The condition $|M_2| > 0$ would always be met, but then, $|M_3| > 0$ for sure, and thus, concavity of the

Hamiltonian could not be confirmed.

Another possibility to show that the necessary optimality conditions are also sufficient are the Mangasarian conditions. These conditions require that the maximized Hamiltonian $(\mathcal{H}^0)$,

$$\mathcal{H}^0 = (u(u'^{-1}\left(-\lambda e^{\Theta}\right)) - D(T))e^{-\Theta} + \lambda\left[u'^{-1}\left(-\lambda e^{\Theta}\right) - \delta T\right] - \mu f(T)$$

is concave with respect to the states. The corresponding Hesse matrix is

$$H = \begin{bmatrix} \frac{\partial^2\mathcal{H}^0}{\partial^2 T} & \frac{\partial^2\mathcal{H}^0}{\partial T\partial\Theta} \\ \frac{\partial^2\mathcal{H}^0}{\partial T\partial\Theta} & \frac{\partial^2\mathcal{H}^0}{\partial^2\Theta} \end{bmatrix} = \begin{bmatrix} -e^{-\Theta}D'' & e^{-\Theta}D' \\ e^{-\Theta}D' & \frac{\partial^2\mathcal{H}}{\partial^2\Theta} \end{bmatrix},$$

where we calculate the term $\frac{\partial^2\mathcal{H}0}{\partial^2\Theta}$ below,

$$\frac{\partial\mathcal{H}^0}{\partial\Theta} = -e^{-\Theta}\left[u(u'^{-1}(-\lambda e^{-\Theta})) - D(T)\right] + \frac{\lambda 2e^{-\Theta}}{u''(u'^{-1}(-\lambda e^{-\Theta}))}(1-e^{-2\Theta}).$$

Thus we have that,

$$\begin{aligned}\frac{\partial^2\mathcal{H}^0}{\partial^2\Theta} =& e^{-\Theta}[u(u'^{-1}(-\lambda e^{-\Theta})) - D(T)] + \frac{3\lambda 2e^{-3\Theta}}{u''(u'^{-1}(-\lambda e^{-\Theta}))} \\ &+ \frac{u'''(u'^{-1}(-\lambda e^{-\Theta}))}{u''(u'^{-1}(-\lambda e^{-\Theta}))}\lambda 3e^{-2\Theta}(1-e^{-2\Theta}).\end{aligned}$$

Since the sign of the payoff $\pi = u - D$ changes for some values of c and T, we can not fully determine the sign of $\frac{\partial^2\mathcal{H}^0}{\partial^2\Theta}$, even if we make a further assumption about the domain of $u'''(c)$. Thus, the Mangasarian conditions are only fulfilled if

$$\frac{\partial^2\mathcal{H}^0}{\partial^2\Theta} < -\frac{D'(T)2e^{-2\Theta}}{D''(T)},$$

which can, but does not have to be the case.

4.7.3 Shift of a Unique Steady State due to Discount Parameter b

We use exactly the same functions and parameters as in Section 4.5.1, but instead of letting the intercept d vary, the shift happens due to a different b. Figure 4.5 illustrates the possible shifts, $\bar{T}^n \lesseqgtr \bar{T}^x$ due a variation of this parameter of the discount function.

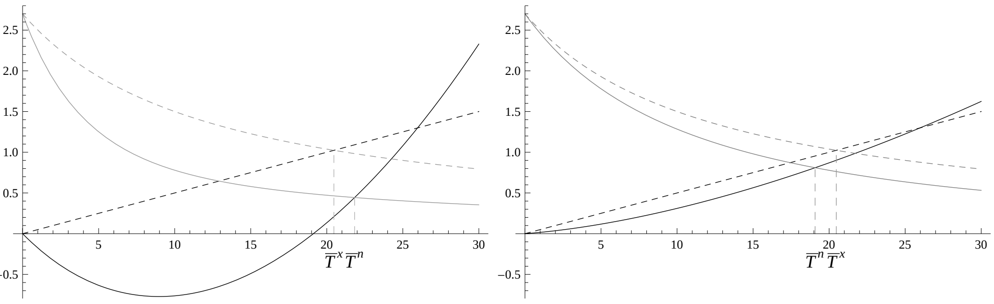
4.7.4 Consequences of a Discount Function Increasing with Pollution

The felicity function $u(c)$ and the damage function $D(T)$ are chosen as in Section 4.5.1, but the discount function, now increasing in T, differs[7]:

$$f(T) = r + \beta(1 - e^{-bT}),\ r, b, \beta > 0. \tag{4.34}$$

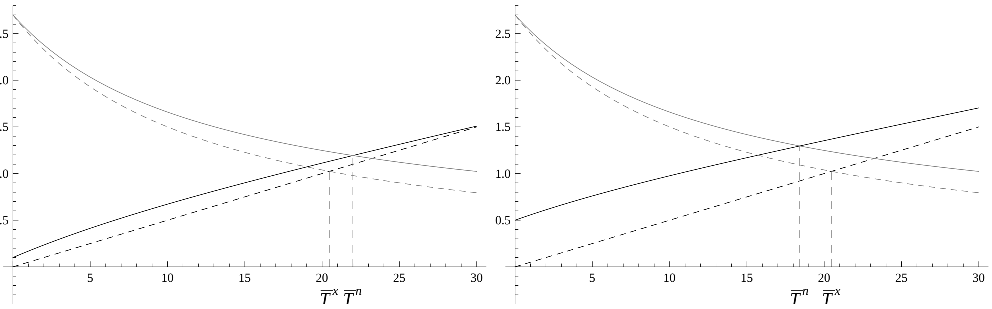
We add the exogenous discount rate r, so that we have a comparable situation as with a decreasing discount function: If there is no pollution, decision makers discount as in the exogenous case, $f(0) = r$, but this time, their impatience increases with higher pollution,

[7]Apart from the constant r, this discount function is equivalent to the one suggested in Ayong Le Kama and Schubert (2007).



(a) (b)

Figure 4.5: L^x, L^n (gray) and R^x, R^n (black) in case of exogenous discounting (dashed lines), as well as in case of endogenous discounting (solid lines), using different values for the parameter b. $a = 15, r = 0.1, \delta = 0.08, k = 0.05, d = 0, b = 0.2$ (a), $b = 0.03$ (b), respectively.



(a) (b)

Figure 4.6: L^x, L^n (gray) and R^x, R^n (black) in case of exogenous discounting (dashed lines), as well as in case of endogenous discounting (solid lines), using different values for the parameter d. $a = 15, r = 0.1, \delta = 0.08, k = 0.05, \beta = 0.2, b = 0.2, d = 5$ (a), $d = 25$ (b), respectively.

$f' > 0$. Technical results do not hinge on the added constant r. Since $f(T) > r$, $\forall T > 0$, L^n always lies above L^x, causing an effect that drives $\bar{T}^n$ to higher levels. But again, depending on the intercept value d, R^n can be shifted so high above R^x that the before mentioned effect is reversed and $\bar{T}^n < \bar{T}^x$. An example is shown in Figure 4.6.

Playing around with the parameters d and b makes it simple to construct examples proving that also multiple steady states can occur.

Chapter 5

Summary

This cumulative PhD thesis consists of three articles. Two articles revolve around the question how the quality of information influences the quality of decisions, once from a theoretical point of view, once from the viewpoint of a real-life person facing the decision to switch the electricity supplier or not. The third paper covers the question if endogenous discounting is the right method to implement a decision maker's sustainability motive in a well-known stock externality problem such as global warming.

The first paper deals with the effect of additional information on the quality of decisions by making use of simulation methods. We start from a scenario where a hypothetical decision maker has no information about the probabilities which would be needed to calculate expected utilities. We then apply two methods developed by us which allow us to constantly and gradually increase the decision maker's information level by systematically limiting the ranges of probabilities. We simulate probabilities lying within the defined ranges and compare the decisions stemming from the thereby gradually increased information to a reference scenario. In this reference scenario the decision maker knows the precise probabilities and can thus decide in line with expected utility theory. The mayor result of our simulation study is that the returns to scale on information decrease.

The second article uses an econometric model, based on survey data, to examine the factors that influence the decision of Austrian households to switch their electricity supplier or not. Thereby, the fundamental role of the respondents' information level about tariffs and competing electricity suppliers becomes apparent. In the course of the modeling the question came up how to handle an endogenous dummy variable in a binary choice setting. Since this question is barely treated in the literature, we describe how to treat this econometric problem in an extra section of the article.

In the third article, we apply endogenous discounting to the problem of global warming, assuming that the discount rate declines as the damage increases. The objective is to investigate whether this reduction of discount rates induces a conservationist attitude (relative to a constant discount rate) without the pitfalls of hyperbolic discounting. However, the expectation that this fosters conservation in the steady state does not necessarily materialize. A small shift from a constant to a state-dependent discount rate can lead to seemingly counterintuitive optimal policies. The sign of the instantaneous net payoff (benefits minus costs), which plays no role in case of an exogenous discount rate, can determine whether the desired conservational effect occurs. As a result, high levels of pollution exceeding their counterpart under constant discounting can be 'optimal' and multiple steady states can occur. In our work we clarify the reasons for these pitfalls of endogenous discounting technically as well as intuitively.

German Summary

Diese Dissertation besteht aus drei einzelnen Artikeln. Zwei dieser Artikel beschäftigen sich mit der Frage, welchen Einfluss die Qualität von Informationen auf die darauf basierenden Entscheidungen hat. Im ersten Artikel wird diese Frage theoretisch mithilfe einer Simulationsstudie abgehandelt, im zweiten Artikel wird die Frage geklärt, welche Rolle die Information über Stromtarife beim Wechselverhalten der österreichischen Haushalte einnimmt. Der dritte Artikel geht der Frage nach, ob sich endogenes Diskontieren bei Problemstellungen mit langen Planungshorizonten als Alternative zu exogenem bzw. zu hyperbolischem Diskontieren eignet.

Im ersten Artikel dieser Arbeit untersuchen wir, wie sich zusätzliche Information auf Qualität von Entscheidungen auswirkt. Im Konkreten geht es um die Information, welchen Wert die Eintrittswahrscheinlichkeiten von disjunkten Ereignissen annehmen können. Im Referenzszenario der vollständigen Information sind die Eintrittswahrscheinlichkeiten genau gegeben, in diesem Fall kann der Entscheidungsträger nach den Prinzipien der Erwartungsnutzentheorie entscheiden. In unserer Arbeit starten wir jedoch im gegenteiligen Fall – keinerlei Information über die Wahrscheinlichkeiten – und implementieren einen Algorithmus, der den möglichen Bereich, in dem die Wahrscheilichkeiten liegen können, schrittweise und konstant verkleinert. Dadurch haben wir die Möglichkeit zu beobachten, wie sich die schrittweise verbesserte Information auf die Qualität von Entscheidungen auswirkt. Das Hauptresultat dieser Studie ist, dass die Qualität der Entscheidungen durch zusätzliche Information zwar besser wird, allerdings mit abnehmender Grenzrate.

Im zweiten Teil der Arbeit werden mithilfe eines ökonometrischen Modells die Einflussfaktoren für den Stromanbieterwechsel von österreichischen Haushalten untersucht. Dabei zeigt sich, dass die Information über Tarife und konkurrierende Anbieter eine dominante Rolle bei der Entscheidung für einen Anbieterwechsel spielt. Die Arbeit beinhaltet auch einen ökonometrischen Teil, der die Handhabungsmöglichkeiten einer endogenen binären Erklärvariablen in einem diskreten Entscheidungsproblem erläutert.

Im dritten Artikel dieser Arbeit wenden wir endogenes Diskontieren auf das Kontrollproblem "Globale Erwärmung" an. Unsere Annahme lautet, dass die Diskontierungsrate fällt, wenn der Verschmutzungsgrad steigt, d.h., wenn der Entscheidungsträger mit den negativen Konsequenzen seines Handelns konfrontiert wird. Mit diesem Modell untersuchen wir, ob eine fallende, zustandsabhängige Diskontierungsrate zu weniger Verschmutzung (relativ zum Fall einer konstanten Diskontierungsrate) im stationären Zustand führt. In unserem Artikel zeigen wir, dass dies nicht der Fall sein muss, die Annahme einer kleiner werdenden Diskontierungsrate kann – auf den ersten Blick überraschend – zu größeren stationären Verschmutzungszuständen führen. Anders als im Fall konstanten Diskontierens bestimmt das Vorzeichen des Integranden (Nutzen minus Kosten) mit, ob die erwünschte Reduktion des Verschmutzunglevels eintritt. Der Wechsel von einer konstanten zu einer endogen bestimmten Diskontierungsrate macht sowohl das Aufteten von höheren Verschmutzungslevels (als im Fall einer konstanten Diskontierungsrate) als auch das Auftreten von multiplen stationären Zuständen möglich. In unserer Arbeit klären wir diese Fallstricke des endogenen Diskontierens sowohl auf technischer, als auch auf intuitiver Ebene.

Chapter 6

Curriculum Vitae

Name	Magdalena Six
Birth	1984 in Ried im Innkreis, Austria
Education	Since March 2010: PhD Management Program University of Vienna, Austria
	October 2003 – October 2009: MSc. in Applied Mathematics University of Salzburg, Austria
Publications	Graf, C., Six, M., 2014. The Effect of Information on the Quality of Decisions, *Central European Journal of Operations Research*, Volume 22, Issue 4 (2014), Page 647–662
Submitted	Six, M., Wirl, F., Wolf, J., Information as Key Determinant in Switching the Electricity Suppliers for Austrian Households, (under review for *Journal of Business Economics*)
	Six, M., Wirl, F., Optimal Pollution Management When Discount Rates are Endogenous, (under review for *Resource and Energy Economics*)