Heterogeneous Ability, Life Expectancy, and Social Security: Four Essays

Dissertation

zur Erlangung des Grades
Doktor der Wirtschaftswissenschaften (Dr. rer. pol.)
am Fachbereich Wirtschaftswissenschaften
der Universität Konstanz

vorgelegt von:

Stefan Hupfeld
Mangoldstraße 11
78462 Konstanz

Tag der mündlichen Prüfung: 09. Juni 2008

Prüfungskommission:
Prof. Dr. Winfried Pohlmeier, Universität Konstanz (Vorsitzender)
Prof. Dr. Friedrich Breyer, Universität Konstanz (1. Referent)
Prof. Dr. Ralf Brüggemann, Universität Konstanz (2. Referent)

Für Verena.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This thesis owes to a lot of people: fellow researchers, teachers, mentors, and friends. First of all, Prof. Dr. Friedrich Breyer was my teacher and mentor through the complete time of my studies in economics, on the doctoral level, but also before on the diploma level, as a teacher in several lectures, seminars, and as the supervisor of my diploma thesis, which already laid the path to some of the work included here. The broad question of this thesis was developed in joint work with him, while I always had the freedom to define in detail how to answer the main and subsequent issues.

Through the complete time of my doctoral studies, this research would not have been possible without funding of the DFG, namely via the research group *Heterogeneous Labor: Positive and Normative Aspects of the Skill Structure of Labor* (grant BR 740/15-2), which I gratefully acknowledge. The FDZ–RV of the German Pension System (Deutsche Rentenversicherung Bund) provided the data I use here, and especially Dr. Ralf Himmelreicher, Tatjana Mika, and Michael Stegman were always very helpful in collecting and working with the data.

A lot of the methods applied here, either empirical or theoretical, were taught in courses of the *Doctoral Programme in Quantitative Economics and Finance* at the University of Konstanz. The application of non–parametric methods is fundamentally influenced by the course in Microeconometrics by Prof. Dr. Winfried Pohlmeier, and Prof. Dr. Oliver Fabel gave me a concise and elaborate introduction in the analysis of principal–agent phenomena.

Normann Lorenz, working at the chair of Prof. Breyer, helped to improve this thesis on various levels and on various details. He especially sharpened a lot of the methodological arguments in the empirical chapters. Prof. Dr. Ralf Brüggemann kindly agreed to be the second referee of this thesis and gave helpful hints on the empirical chapters. Furthermore, colleagues, teachers, and discussants at workshops and conferences deserve thanks (in alphabetical order): Leo Kaas, Mathias Kifmann, Walter Krämer, Joaquim Oliveira Martins, Nick Netzer, Regina Riphahn, Harris Schlesinger, Thomas Stratmann, Rolf Tschernig, the participants of the FNA Doctoral Colloquium, the RGS Doctoral Conference, and the BGPE Conference, two anonymous referees appointed by the *Journal of Health Economics*, and two anonymous referees appointed by the *Journal of Population Economics*.

Last, but certainly not least, I thank my parents for ongoing support during my doctoral studies. Verena: for proof–reading the manuscript various times, but more important for being her; without her, the completion of the thesis might have been impossible.

Introduction

Though the present thesis is not intended as a monograph, its respective chapters are interrelated: The major research question is whether the German public pension system (or potentially any annuity–based pension system) causes redistribution between income groups. Especially the relationship between life expectancy and income, usually perceived to be increasing, gives rise to the conjecture that poor individuals subsidize the pensions of the rich. I analyze the income–life expectancy relationship theoretically and empirically in Chapters 2 and 4, respectively. The redistributive mechanism is empirically analyzed in Chapter 3. Chapter 5 proposes a policy instrument which is capable of neutralizing the redistributive effects.

The thesis consists of a literature overview in Chapter 1, and of four self–contained essays in Chapters 2 through 5. In Chapter 2 (*To Work or to Work Out: A Moral–Hazard Interpretation of Labor Supply, Retirement, and Investments in Longevity*), I analyze a moral–hazard model with multiple tasks, in which a single agent engages in three different activities labelled work, delayed retirement, and workout (interpreted as investments in longevity). The latter imposes higher effort costs on weekly labor supply, but increases possibilities for life time labor supply. Work–out does not affect aggregate output and does therefore not increase the benefit of the principal. Second best incentives for work–out are U–shaped in the agent's ability, and so is the effort level supplied by the agent.

In Chapter 3 (*Rich and Healthy—better than Poor and Sick? An Empirical Analysis of Income, Health, and the Duration of the Pension Benefit Spell*, currently under revision for the *Journal of Health Economics*), I analyze the relationship between duration of the pension benefit spell and pension benefit claims from the German public pension system, with a special emphasis on differential results with respect to health. This relationship is crucial and causal for a potential structural pattern of redistribution between different income and health groups, induced by the public pension system. Evidence for such redistribution from poor to rich is present for most of the specifications in my analysis. The most adequate specification is partially–linear, does therefore not impose any parametrical restrictions between duration and benefit claims, and allows for potential endogeneity. The relationship I extract is clearly positive. Additionally I find that the income gradient is steeper for pensioners in bad health, meaning that redistribution from less able to more able individuals is more pronounced the worse the health status is.

Subsequently, in Chapter 4 (*Non–Monotonicity in the Longevity–Income Relationship*, currently under revision for the *Journal of Population Economics*), I show that for major sub–groups of pensioners in the public pension system in Germany who died between 1994 and 2005, the relationship between income and life expectancy is non–monotonous. This cannot be explained by anomalies in the data or as an artifact of the estimation technique, and so I provide a fundamental conjecture based on different elasticities of labor supply over the income distribution. Finally, in Chapter 5 (*On the Fairness of Early Retirement Provisions*,[1] joint with Friedrich Breyer), we discuss several notions of 'fairness' of early retirement provisions in pay–as–you–go financed public pension systems and we claim that the 'right' notion of fairness depends upon the objectives pursued in the design of pension systems. We point out the problems attached to the extreme positions 'efficiency' and 'welfare maximization' and propose a more modest concept of equity called 'distributive neutrality', which is based on the notion that the ratio between total benefits and total contributions to the pension system should not depend systematically on the individual's ability. By applying this concept to the German retirement benefit formula and taking empirically estimated relationships between duration of the benefit spell and income into account, we show that at the present discount rate of 3.6 per cent per year there is systematic redistribution from low to high earners, which would be attenuated if the discount rate were raised. This seemingly paradoxical finding is due to the fact that in our data set, there is a negative relationship between earnings and retirement age.

[1] An earlier version of this paper is available as CESIfo Working Paper No. 2078 (2007).

Einleitung

Obgleich die vorliegende Dissertation nicht als Monographie angelegt ist, sind die einzelnen Kapitel doch verbunden: Die grundlegende Forschungsfrage lautet, ob die gesetzliche Rentenversicherung in Deutschland (oder möglicherweise jedes auf Annuitäten basierende Rentensystem) Umverteilung zwischen Einkommensgruppen verursacht. Besonders die Beziehung zwischen Einkommen und Lebenserwartung, die üblicherweise als steigend angenommen wird, begründet die Vermutung, dass arme Rentner die Renten der Reichen subventionieren. Diese Beziehung untersuche ich theoretisch und empirisch in den Kapiteln 2 und 4. Der umverteilende Mechanismus wird in Kapitel 3 empirisch untersucht. In Kapitel 5 wird ein Politikinstrument vorgeschlagen, welches in der Lage ist, die Umverteilung zu neutralisieren.

Die vorliegende Dissertation enthält eine Literaturübersicht in Kapitel 1 und vier eigenständige Aufsätze in den Kapiteln 2 bis 5. In Kapitel 2 (*To Work or to Work Out: A Moral–Hazard Interpretation of Labor Supply, Retirement, and Investments in Longevity*) analysiere ich ein Moral–Hazard Modell mit mehreren Aktivitäten, in welchem ein einzelner Akteur drei verschiedene Handlungen vornehmen kann, die ich Arbeit, verzögerten Renteneintritt und Training (*Work, Delayed Retirement* und *Work Out*) nenne. Die letztere Aktivität verursacht höhere Nutzeneinschränkungen beim wöchentlichen Arbeitsangebot, verbessert aber die Möglichkeiten einer längeren Lebensarbeitszeit. Training beeinflusst die aggregierte Wertschöpfung nicht, weshalb der Prinzipal kein direktes Augenmerk darauf legt. Zweitbeste Anreize für *Work Out* sind eine U–förmige Funktion der Fähigkeiten des Akteurs, genau wie sein Anstrengungsniveau.

In Kapitel 3 (*Rich and Healthy—better than Poor and Sick? An Empirical Analysis of Income, Health, and the Duration of the Pension Benefit Spell*, zur Zeit in Revision für das *Journal of Health Economics*) analysiere ich den Zusammenhang zwischen Rentenbezugsdauer und -ansprüchen in der deutschen gesetzlichen Rentenversicherung unter besonderer Berücksichtigung des Gesundheitszustandes. Diese Beziehung ist grundlegend für strukturelle Umverteilung zwischen Gruppen mit verschiedenen Einkommen und Gesundheitszuständen, die auf das Rentensystem zurückzuführen wäre. Die meisten Spezifikationen der Analyse finden einen Beleg für Umverteilung von Arm zu Reich. Die geeignetste Spezifikation ist partiell–linear, beinhaltet daher keine parametrischen Beschränkungen für den Zusammenhang zwischen Bezugsdauer und Rentenansprüchen und berücksichtigt mögliche Endogenitätsprobleme. Der gefundene Zusammenhang ist eindeutig positiv. Darüber hinaus erweist sich der Einfluss des Einkommens als stärker, je schlechter der Gesundheitszustand des Rentners ist, was bedeutet, dass die Umverteilung von Arm zu Reich ein größeres Ausmaß annimmt.

Danach zeige ich in Kapitel 4 (*Non–Monotonicity in the Longevity–Income Relationship*, zur Zeit in Revision für das *Journal of Population Economics*), dass der Zusammenhang zwischen Einkommen und Lebenserwartung für bedeutende Gruppen der Rentner, die im deutschen Rentensystem zwischen 1994 und 2005 verstorben sind, nicht monoton ist. Dieses Ergebnis ist weder auf Anomalien des Datensatzes zurückzuführen, noch ist es ein Artefakt der Schätzmethode. Daher stelle ich eine theoretische Vermutung an, die auf unterschiedlichen Arbeitsangebotselastizitäten beruht und den gefunden Zusammenhang erklären kann.

Schließlich diskutieren wir in Kapitel 5 (*On the Fairness of Early Retirement Provisions*,[2] gemeinsam mit Friedrich Breyer) verschiedene Auffassungen von *Fairness* bezüglich der Abschläge bei vorgezogenem Renteneintritt in umlagefinazierten Rentensystemen. Wir stellen fest, dass die richtige Auffassung von Fairness von den Zielen abhängt, die mit dem Rentensystem erreicht werden sollen. Wir zeigen das Problem bezüglich der extremen Positionen *Effizienz* und *Wohlfahrtsmaximierung* und schlagen ein maßvolleres Gleichheitskonzept vor, welches wir *distributive Neutralität* nennen. Es basiert auf der Vorstellung, dass das Verhältnis der gesamten Auszahlungen zu den gesamten Beiträgen zum Rentensystem nicht systematisch von den individuellen Fähigkeiten abhängen sollte. Wir wenden dieses Prinzip auf die deutsche Rentenformel an, und unter Berücksichtigung von Schätzungen des Zusammenhangs zwischen durchschnittlichem Jahreseinkommen und der Rentenbezugsdauer zeigen wir, dass die aktuellen Rentenabschläge für Frührente von 3.6% pro Jahr systematische Umverteilung von Arm zu Reich verursachen. Diese Umverteilung würde durch höhere Abschläge abgemildert. Dieses scheinbar paradoxe Ergebnis liegt daran, dass in unserem Datensatz der Zusammenhang zwischen Einkommen und Renteneintrittsalter negativ ist.

[2] Eine frühere Version dieses Papiers steht als CESIfo Working Paper No. 2078 (2007) zur Verfügung.

Chapter 1

A Brief Literature Overview

1.1 Introduction

This Chapter serves as a brief overview of the existing literature on the topics in the present thesis. Whenever possible, recent references are cited, together with seminal contributions which laid the foundation to complete strands of the literature. Section 1.2 provides a very quick overview of the theoretical and empirical methods I discuss and apply. Sections 1.3, 1.4, and 1.5 summarize theoretical and empirical literature on the relationship of life expectancy and income, on retirement behavior, and on redistributional effects and optimality of public pension systems.

1.2 Methodology

1.2.1 Mechanism Design and Optimal Taxation

Theoretical contributions frequently approach the question of optimality of pension system with tools provided by the theory of optimal taxation. Seminal papers on optimal taxation are certainly Mirrlees (1971) and Diamond and Mirrlees (1971a) and (1971b), whereas e.g. Myles (2002) provides a textbook coverage. The underlying problem is that a government would like to base taxation and redistribution on at least one unobservable characteristic, such as ability (or, in the present case, also health or life expectancy). This direct instrument is not feasible, so the tax system has to condition on observables and takes incentive compatibility constraints explicitly into account, in order to prevent individuals from mimicking other types. Yet, the formal results derived by models of optimal taxation are limited in the sense that they (a) rely on an arbitrary welfare function, and (b) even qualitative results with relevance for policy can often be derived only if one assumes explicit functional forms, as Diamond (1998) claims.

Closely related to optimal taxation is principal–agent theory I apply in Chapter 2. In principle, two separate phenomena arise due to unobservability, either hidden action (moral–hazard) or hidden information (adverse selection), but only the former is relevant in Chapter 2. For a textbook on standard moral–hazard models, refer to Bolton and Dewatripont (2005) or Salanié (1997). A similar moral–hazard effect is subject to the theory of insurance demand, applied to insurance against longevity, e.g. in Davies and Kuhn (1992) and Philipson and Becker (1998). In a model with one agent and three different activities (effort levels are observable to different degrees), I apply the solution mechanism proposed by Holmstrom and Milgrom (1991), which relies on linear reward mechanisms derived in Holmstrom and Milgrom (1987). The analysis and solution to the general question when to implement high–powered incentives, when to invest in monitoring, and how to assign ownership (of which the moral–hazard problem in Chapter 2 covers the question of high–powered incentives) is found in Holmstrom and Milgrom (1994). A major drawback of the proposed solution mechanism is its reliance on the first–order approach, hence on the assumption that incentive compatibility constraints can be replaced by first–order conditions of the agent's optimization problem. Although in Chapter 2 the validity of this assumption is satisfied, this is not the general case; see especially Grossman and Hart (1983), Rogerson (1985), and Jewitt (1988) on the justification and validity of the first order approach and on alternatives.

1.2.2 Non–Parametric Econometrics

Parametric estimation methods such as OLS impose a structure on the relationship that is rather to be discovered. If this structure, e.g. linearity, does not capture the true model, parametric models can be "quite misleading", as DiNardo and Tobias (2001) argue in their survey article. Such a structure has been imposed in the vast majority of empirical papers on the relationship between income and longevity or income and retirement age, and more complex relationships may be hidden behind the assumption of linearity. The functional form of the relationship discovered by non–parametric methods is supplied by the data alone, and not by the model. However, this freedom (despite smoothness of the fitted function) comes at the loss of easily interpretable parameters.

Cleveland (1974) introduced local regression as a means to interpret scatter–plots. In principle, a function is not fitted to the whole data, but at each observation separately, taking the respective remaining data to the left and right of this point into account with declining weights. The functional form of the kernel determines these weights, but the choice of the kernel itself has only a minor impact, yet, a bandwidth has to be determined, which governs how fast the weights decrease with the distance to the current observation. If the bandwidth is too large, the estimate will be too smooth, in the limiting case the fit is only a constant. Otherwise, a small bandwidth fits a rougher function and is more sensible to outliers; the limiting case here is an exact replication of the original data. The most recent and comprehensive textbook which also covers several bandwidth–selection methods is certainly Li and Racine (2007), but see also Loader (2004) for an overview. Yatchew (1998) and Fan (2000) provide survey articles and promote non–parametric methods.

If more than one regressor enters the model, the econometrician faces the curse of dimensionality, meaning that in order to reach the same approximation error, the number of observations has to increase over–proportionally with the number of regressors. From a practical point of view, non–parametric estimation in high dimensions is also computational burdensome, especially with huge data sets as the one analyzed in this thesis. Semi–parametric methods may balance the need for functional freedom and the requirement of more regressors. In principle, semi–parametric methods allow some regressors to enter fully non–parametrically, while others are bound to follow a parametric structure. For surveys of semi–parametric methods, see Robinson (1988) and also Yatchew (1998), while Yatchew (2003) provides a textbook treatment. Seminal applications of semi–parametric and non–parametric methods are e.g. Engle et al. (1986) on electricity sales, and Blundell and Duncan (1998) and Blundell et al. (1998) with a semi–parametric derivation of Engel–curves.

1.3 Income, Health, and Life Expectancy

1.3.1 Theoretical Considerations

The seminal economic paper on health is Grossman (1972), who introduces the concept of health capital. The original model is discussed extensively, e.g. in the textbook by Breyer et al. (2005) and in Grossman (2000). In the basic model, a household invests in health capital and benefits from the 'interest' the stock of health capital pays, namely from improved possibilities to be productive and from increased utility. Age at death is only implicitly defined in this model—death occurs once the stock of health capital falls short of a certain threshold. On the question whether age at death in the Grossman model (as the planning horizon) is under direct control of the invidual, see especially Ried (1998) and in addition Grossman (2000) for a short discussion.[1] Under some assumptions (namely, stressing the investment property of health capital, and hence its positive impact on future productivity), the model produces a positive relation between productivity and health, and therefore between income and life expectancy. Muurinen (1982) merges the alternative interpretations of health being either a consumption good or an investment good and describes them as extreme cases.

There exist several extensions with respect to uncertainty of health–related behavior, such as Dardanoni and Wagstaff (1987), Selden (1993), or Chang (1996). With uncertainty, the nature of 'health' (whether it is a normal good or not) depends on the risk attitude of the individual, and Chang (1996) shows that under certain assumptions, health can be inferior.

[1] The answer to this is rather methodological and far from trivial, because the optimal paths of all other control variables (such as consumption) are eventually determined by the age at death. Another theoretical approach cited below (Ehrlich and Chuma 1990) therefore separates the demand for health from the demand for longevity.

Ehrlich and Chuma (1990) provide a second seminal contribution and explicitly discriminate the demand for health from the demand for longevity. This approach yields different results with respect to the impact of initial wealth as opposed to transitory income: In Ehrlich and Chuma (1990), initial wealth determines the demand for health as well as longevity, while in Grossman (1972), initial endowments do not matter in the investment interpretation with certainty. Adding uncertainty to their model, Ehrlich (2000, 2001) explicitly derives not only the demand for longevity, but also for insurance and self–insurance. As a link to Section 1.4, retirement may serve as a means of investment in health capital; see Wolfe (1985) for a theoretical treatment of this topic.

1.3.2 Empirical Evidence

Literature on the empirical relationship between socio–economic status, health, and life expectancy (or mortality) abounds.[2] To reference especially recent contributions, I begin with Cutler et al. (2006), who find that income is positively related to life expectancy, both within a country and across country averages. For specific countries, this is confirmed e.g. by Attanasio and Emmerson (2003) for Great Britain or Deaton and Paxson (2004) for the United States, whereas for Germany, a clearly positive relation is found by Reil-Held (2000) and more recently by von Gaudecker and Scholz (2007). Yet, most of the cited contributions rely on parametrical assumptions (except of von Gaudecker and Scholz 2007), and the finding that individuals with the highest income not only outlive the poorest, but that this relation also holds for the comparison of all other income levels may rest on exactly these assumptions.

There is little existing literature on potentially negative impacts; even despite increasing alcohol consumption of the rich, Ettner (1996) still finds a positive net impact of income on health (which is confirmed by Banks et al. 2006, who show that the positive income–longevity relation survives after controlling for risk factors such as drinking and smoking). Others even claim that a major fraction of mortality differences can be attributed to socio–economic status alone instead of other factors, see e.g. Menchik (1993): He finds that differential life expectancy of blacks and whites in the United States almost vanishes once income enters as an additional control. Deaton (2006) challenges this result and argues that—at least on an international level—institutions in general and especially those that foster education drive a lot of the positive correlation between income and health, and on

[2]For the effect of income inequality on health or mortality—a topic not considered in this thesis—refer to the seminal work by Wilkinson (1996) and a recent empirical study by Gerdtham and Johannesson (2004), who cannot corroborate an impact of income–inequality on mortality. See also Jones (2006) for a survey of health and wealth inequality in Europe.

the individual level, Duggan et al. (2007) show that mortality differences between men and women remain large, independent of income. See also Soares (2007) for factors other than income that positively affect longevity, especially on institutions regarding education and health care.

So, in summary there are some empirical issues left open. Evidence for a positive association between health/longevity and wealth/income abounds, but the mechanism is still ambiguous. Some differences in health, which are associated with certain individual characteristics, can be attributed to differences in income (blacks and whites), while others differences (between men and women) survive despite controlling for socio–economic status.

The quantitative effect of income on life expectancy is disputed and depends on the country, the method and the data set. As a benchmark, Duggan et al. (2007) find a difference of two to three years between life expectancy of the lowest and the highest income group in the United States, whereas for the case of Germany, Reil-Held (2000) and von Gaudecker and Scholz (2007) agree on a gain of approximately six years for the highest income group.

It is still unclear whether the discovered relationship is causal; and if so, whether it is causal from income to health and mortality or vice versa. Cutler et al. (2006) downplay the direct impact of income on mortality and also stress the importance of education and technological progress. Education is a (potentially neglected) background variable, which is already present in the basic Grossman (1972) model of health capital. On the positive impact of education on health status, see Adams et al. (2003) and the recent contribution by Kiuila and Mieszkowski (2007), who show that the income gradient as well as the education gradient almost vanishes for the oldest–old; they argue that ultimately biological factors independent of income or education govern health status and mortality.

On the causal impact of income or wealth on health see also Meer et al. (2003) and Lindahl (2005), who instrument socio–economic status in order to account for reverse causality. But their studies draw opposite conclusions: The authors of the former find no evidence for a causal path from income to health, while the latter does. Smith (1998) and Adams et al. (2003) also apply econometric methods that explicitly consider the endogenous nature of health incidences and socio–economic variables, and the latter find that acute diseases are less likely to be caused by income or wealth, while causality cannot be ruled out for the case of chronic states of ill–health.[3] The German reunification may provide a quasi–natural experiment allowing to disentangle the effect of income on health alone, and Frijters et al. (2005)

[3] On methodological aspects of causality e.g. in the health–wealth nexus, see Granger (1969) and Ribeiro (2002), and especially the comments in Adda et al. (2003), Poterba (2003), Heckman (2003), Mealli and Rubin (2003), Robins (2003), Hausman (2003), Geweke (2003), and Florens (2003).

show in a study utilizing this event that at least satisfaction with health increases with income, yet only to a small extent. Endogeneity between health or mortality and socio–economic status may not only be due to reverse causality, but also a result of common background variables other than education. One major determinant of both socio–economic status and health is the inheritance of at least one of both, either genetical or via education, or in the case of wealth, via bequests. See especially Case et al. (2002), Currie and Stabile (2003), Currie et al. (2007), and Case et al. (2007), who discuss the impact of the income gradient on health at different ages in the United States, Canada, and England. But even when endowment at childhood is taken into account, endogeneity may still survive: Datar et al. (2007) argue that background risk in mortality affects parental investment in children, such that initial endowments in either health or wealth affect adult outcomes and are themselves endogenous.

1.4 The Retirement Decision

1.4.1 Theoretical Considerations

The foundation of the theoretical analysis of the retirement decision was laid by Feldstein (1974), Diamond and Mirrlees (1978), and Sheshinski (1978), who analyze (and estimate) the effect of a pension system on savings and, even more important, on the retirement decision. In absence of any early retirement discounts, Sheshinski (1978) derives that a public pension scheme induces early retirement, but with positive discounts (or, as the author denotes, with a return on postponed retirement), this result can be mitigated. Comparing this to more recent findings on the 'right' size of early retirement discounts, the result is easily compatible with the notion of Boersch-Supan (2000) and Boersch-Supan (2004) that discounts should be equal to the interest rate in order to leave incentives to retire unaffected by the pension system, if labor supply is intended to be undistorted.

Bloom et al. (2007) provide a recent theoretical contribution linking retirement age to life expectancy. They explicitly model the effect of increasing life expectancy and decreasing morbidity and find that increased lifetime has a substitution effect and an income effect, and although the latter dominates with respect to the retirement decision, retirement is delayed only under–proportionally. A similar argument is brought up by Kalemli-Ozcan and Weil (2002), yet, their case includes a discontinuity: If mortality is high (and life expectancy is low), individuals optimally refrain from planning to retire and work until the end of their life, without accumulating wealth intended to spend during the retirement period. Only if mortality decreases, retirement becomes a favored option, because with declining mortality, the per–period risk of a loss of the accumulated wealth decreases as well.

A further contribution by Bloom et al. (2005) indicates that initial wages are negatively related to retirement age, but the growth rate of income has a positive impact. The first is explained by the income effect of higher wages on labor supply in general, which also emerges for the retirement decision. I deliberately neglect forced retirement in the thesis; a recent theoretical contribution on this topic is Fleischhauer (2007), who finds that the retirement decision especially of low–ability workers is actually made by firms, and Dorn and Sousa-Poza (2005) provide an empirical corroboration.

1.4.2 Empirical Evidence

Empirically demanding is the fact that at least self–reported health is endogenous with respect to labor force participation, and that not all health incidences play the same role for the decision whether to participate in the labor force or not. Considering the case of Germany explicitly, Boersch-Supan and Jürges (2006) argue that early retirement is most likely a reaction to health incidences, because early retirees suffer from less well-being than normal retirees, yet, their status improves due to the retirement decision (which is in line with the theoretical reasoning applied by Wolfe 1985). Life expectancy itself affects retirement positively, as Waldron (2001) shows for the United States, which is contested by the empirical test in Bloom et al. (2006), who show that subjective probabilities of survival have no impact on the retirement decision in the United States (although longevity is a main driving variable in the theoretical model by Bloom et al. 2005). For a short survey on the positive effect of health levels and health changes on retirement behavior, refer to Jones (2006). In addition, see Banks et al. (2006), it is not only own health, but also the health status of other household members that shapes the retirement decision. See also the survey of Lindeboom (2006) on the reverse impact of health incidences on the retirement decision; in summary, health may positively affect labor supply, but also the demand for leisure.

Evidence on the impact of income on retirement age is controversial, see Berkel and Boersch-Supan (2004) versus Schils (2005) with opposing results with respect to the impact of income or wealth on retirement age. In general, the impact of wealth or income on early retirement depends on the specific pathway into retirement, as e.g. Banks et al. (2006) show for the UK; individuals at the bottom of the income distribution have often left the labor force already into unemployment before retirement.

Discounts on early retirement directly influence the implicit tax on work beyond the earliest retirement age and affect the expected income after retirement.

For an overview of these implicit tax rates, refer to Butrica et al. (2006), who calculate that in the United States the implicit tax on prolonged work even exceeds 50%, depending on age and income of the individual. Beckmann (2000) estimates the implicit tax for Germany—on a lifetime basis— to be in the neighborhood of 55% to 65%. On the effect of early retirement discounts, Hanel and Riphahn (2006) find that Swiss women facing higher early retirement discounts actually postpone retirement. The effect on men is not significantly different from zero, which is in line with the finding that men have a lower wage elasticity of labor supply, see e.g. Heckman (1995) or van Soest (1995). On simulated outcomes of policy reforms in Germany, refer to Berkel and Boersch-Supan (2004), who claim that an increasing discount factor or legal retirement age (together with the earliest possible retirement age) yields delayed retirement for both men and women. In the Netherlands, the transition from a pay–as–you–go pension system without discounts for early retirement to an actuarially fair funded system (including an adjustment of discounts) led to delayed retirement (see Euwals et al. 2006).

As a matter of course, there are also legal restrictions that influence retirement behavior, the most prominent being the choice of the earliest retirement age (besides discounts which adjust retirement benefits). On the impact of the earliest retirement age, see e.g. Baker and Benjamin (1999), who analyze the reaction of older Canadian workers to the introduction of an early retirement scheme. Interestingly, total labor supply under the new legislation was hardly affected, as those who claimed benefit payments were individuals who already had chosen to stay out of the labor force anyhow, even without early retirement schemes.

In most countries, legislation imposes a 'normal' retirement age, e.g. 65, and an earliest possible retirement age of e.g. 62 (see especially Gruber and Wise 1998, Gustman and Steinmeier 2001, and Mulligan and Sala–i–Martin 2004 on international comparisons and mutual properties of public pension systems). Gustman and Steinmeier (2005) provide a convincing explanation for the phenomenon that the empirical distribution of the individual retirement age is bimodal, with peaks at 62 and 65, respectively. Time preferences—in relation to early retirement discounts—govern at least a fraction of the retirement behavior and are distributed over a wider support than the one provided by a public pension system, such that all individuals who preferred to retire earlier than 62 concentrate around this age, and all those who would retire after 65 gather exactly there.

1.5 Redistribution and Optimality

1.5.1 Theory and Empirical Evidence for Redistribution

Simonovits (2006) show that an actuarially fair pension system is "far from being neutral". The same argument is brought forward by Cremer et al. (2006b) who show that a pension system mimicking the laissez–fair solution is not optimal, and from a political point view, not feasible either.[4] The driving force behind redistribution is heterogeneous life expectancy and the following heterogeneous sum of benefits from a pension system. A World Bank report (World Bank 1994, pp. 130 and all the references cited therein) contains early empirical confirmations of this conjecture: The regressive effect of an annuity–based public pension system may be so large to overcompensate progressive elements in income tax schemes and, if present, in the pension system.[5] Redistribution induced by differential mortality could, however, be overcome by corresponding differences in the retirement age—see e.g. Barnay (2007) with a study on France that highlights the potential benefit of induced retirement behavior to offset redistribution (Barnay 2007, however, conditions the redistribution on different professions and educational levels instead of income).

Even on private markets for annuities we see that individual perception of own life expectancy—especially of remaining life expectancy after retirement age—affects decisions with respect to the valuation of annuities. For a given retirement age, the return from an averaged annuity is smaller, the shorter the remaining life expectancy is. Turra and Mitchell (2004) show that retirees in worse health prefer a lump–sum payment over the annuity.[6] Similarly, Salm (2007) argues that eventually the sheer presence of an annuity may increase life expectancy, although he cannot clearly distinguish the effect of the pension from the effect of income or wealth. Both findings, in turn, are a partial empirical corroboration of adverse selection on annuity markets, as predicted by the theoretical contribution by Wolfe (1983). See also Brown (2007), who points at the phenomenon that the majority of private annuities are sold to healthy individuals, although—on complete markets—a fairly

[4]A recent contribution on the political economy of pension systems is Borck (2007). Browning (1975) provides the seminal model on the political economy of pension systems. Besides this, the political economics of social security security are not covered here.

[5]For the United Kingdom, see Creedy et al. (1993); for the Netherlands, see Nelissen (1995) and Nelissen (1998); for Italy, see Kostoris Padoa Schioppa (1990); and for Sweden, see Stahlberg (1989). See Hurd and Shoven (1986), who provide the first study of the redistributional effect of the US public pension system. The authors find that—despite the progressive elements in the US pension system—the ratio between benefits and contributions are almost the same for different wealth groups, because the progressive elements are offset by mortality differences.

[6]This results holds even if unfair annuities would still be optimal compared to a lump–sum payment to individuals in worse health.

priced contract would be to the benefit of all individuals, despite their health status or income.[7] Annuities and their (mandatory) public provision explicitly follow an additional objective, namely the insurance against longevity—in which sense they are potentially welfare–improving, see Sheshinski and Weiss (1981) on this issue.[8]

Simonovits and Esö (2003) and Simonovits (2006) derive a second–best optimal pension scheme, based on private information on life expectancies, and they show—depending on the welfare function—that redistribution from short–lived pensioners to long–lived pensioners may even be optimal, hence that neutrality of the pension system with respect to different income groups may be inferior to the second–best.

The impact of socio–economic factors on duration of the pension benefit spell in Germany has been analyzed by Lauterbach et al. (2006). They find that duration of the pension benefit spell is indeed higher for individuals with higher income. However, the death cohort they observe is rather small, and benefit payments before the age of 65 (and therefore the possibility to balance shorter life expectancy) are neglected. Another empirical contribution with regard to heterogeneous social security outcomes by different subgroups of the population in the United States has been made by Liu and Rettenmaier (2003), who confirm Hurd and Shoven (1986) in the sense that shorter life expectancy of blacks yields a lower return from social security even despite progressive elements in the U.S. public pension system. Yet, other groups with—on average—slightly lower life expectancy like the low educated offset their disadvantage and can directly benefit from the progressive elements in the benefit formula. The German pension system even lacks such progressive elements. Brown (2003) argues that if redistribution is measured not financially, but adjusted for utility, redistribution is still present, but diminished, and he also finds that despite its redistributional effects to the advantage of either rich or longer–lived individuals, uniformly priced annuitization is still optimal.

For an international overview with respect to fairness or neutrality of pension systems, and for concise definitions of these concepts, refer to Queisser and Whitehouse (2006).

1.5.2 Optimality

Compare Yaari (1965) on the optimality of full annuitization if annuities are fairly priced regardless of the size of any early retirement discounts, and see Davidoff

[7]Brown (2007) provides additional behavioral conjectures, whereas Diamond (2004) claims that the absence of a complete market for annuities serves as an explanation for the existence of a public pension system.

[8]Compare Fehr and Habermann (2007), who simulate circumstances under which mandatory annuitization may decrease welfare for future generations due to foregone bequests.

et al. (2005) on circumstances under which full annuitization may not be optimal—even if markets are complete—e.g. if individuals have bequest motives or the return of the annuity is smaller than the return on non–annuitized assets. If markets are incomplete (that is, if annuities are not necessarily fairly priced, or if not all consumption paths are available), liquidity and flexibility of the annuity contract determines the optimal degree of annuitization.

Then, with symmetric information or homogeneous individuals, annuities provided by the pension system should not distort labor supply and hence the retirement decision. Boersch-Supan (2000, 2004) derives the optimal early retirement discounts for this case, which are actuarially fair and match the interest rate.

Yet, if information on life expectancy or productivity is asymmetric and therefore not available for the designer of the pension system, a certain amount of redistribution may be desirable, depending on the social welfare objective. This approach utilizes optimal taxation, where some actions or characteristics are unobservable. In general, models of optimal taxation yield age–dependent taxes, see e.g. Erosa and Gervais (2002). The fact that age–dependent taxation can be implemented via a mandatory public pension system with a certain benefit formula is addressed by Wrede (1999), and Kifmann (2008) derives an optimal benefit formula, interacting with a general income tax scheme.

Cremer and Pestieau (1996) analyze the interaction of a tax system with social security (modeled as an insurance against an arbitrarily defined loss), and both, social security as well as the general tax system, should optimally depend on ability, the individual risk of occurring a loss, and the correlation of both. Optimal taxation is in general based on (labor supply) elasticities of the taxed individuals, hence on the feasibility of evasive behavior (see Saez 2001 on a thorough analysis of the role of elasticities). In Kifmann (2008), the optimal tax scheme depends on elasticities, augmented with income levels and income inequality. As Fenge et al. (2006) show, also the implicit tax rates from the pension system should follow an age–profile, which is—regardless of optimality—the case in the German pension system, because contributions are treated independently of age (see Lindbeck and Persson 2003 for a general overview).

In a model with differential life expectancy and different productivity levels, Diamond (2003) shows that individuals above a certain threshold of life expectancy *or* productivity can exploit the age–dependent tax system, because the government has to keep incentives for individuals with more intermediate characteristics. Still, as Brown (2003) explicitly claims, full annuitization is optimal for all individuals, no matter whether their life expectancy is specifically high or low. Also applying optimal taxation, Cremer et al. (2004) analyze the interaction of health and productivity with respect to the resulting optimal tax scheme and find that delayed retirement is (relatively) taxed in order to subsidize the respective 'weaker' type and to enable him to retire earlier. Late retirement in their setting is still preferred by individuals with high productivity or good health, which are assumed to be positively correlated. In a subsequent study, Cremer et al. (2007) propose disability testing to elicit individual health status in order to infer on an optimal tax scheme (and optimal retirement) in interaction with disability insurance. Healthy individuals with high disutility of labor should be distinguished from disabled individuals, which either justifies earlier retirement for the disabled or higher benefits from the disability insurance. Due to the relationship between income, life expectancy and health, Henriet and Rochet (2006) even argue that intentional redistribution from rich to poor can be based on public health insurance, if the rich cannot augment their insurance coverage on private markets.

On the question whether age–dependent taxation before retirement is feasible or not, see Bovenberg and Sorensen (2006), who raise the point that an optimal tax scheme should not condition on annual income, but on lifetime income, and at least for the case of labor income, the latter is observable at retirement age.

Breyer and Kifmann (2002) analyze the effect of the adjustment of discounts for early retirees on the implicit tax rate—however, heterogeneous life expectancy does not matter in their analysis. Its separate effect on the implicit tax of a pension system is analyzed in Breyer and Kifmann (2004). Sheshinski (2003) provides an explicit treatment for the adjustment of retirement benefits if retirement age is delayed even beyond the legal requirements (here, disutility from labor is unobservable to the government), and Fenge et al. (2006) derive a factual and an optimal life–cycle profile of the implicit tax imposed by a social security system. They argue that actual as well as optimal implicit taxation differs even despite absent differences in duration under the benefit spell.

Obviously, life expectancy and health status are interrelated. A normative approach to social security might therefore need to consider the interaction of two systems of social security, the public pension scheme and public health insurance. With exogenous health, this is already incorporated in Cremer et al. (2004), and Cremer et al. (2006a) propose refinements with explicit consideration of endogenized health and derive that subsidies for health expenditures are second–best optimal. A further strand of the literature explicitly considers the interaction of income and longevity with education for the optimal design of pension systems; see Lau and Poutvaara (2006), Gorski et al. (2007), or Ferreira and Pessoa (2007) for potential distortions of the educational decision imposed by the pension system.

Simonovits (2004) offers an approach which combines distributive neutrality and second–best taxation. Induced retirement age in his solution varies over–proportionally with life expectancy, because shorter–lived individuals have to retire very early and suffer therefore from neutral early retirement discounts which are high enough to deter the longer–lived individuals from mimicking the shorter–lived ones. However, since the second–best is further constrained by neutrality, the solution is potentially Pareto–inferior to the unconstrained second–best.

References of Chapter 1

ADAMS, P., M. D. HURD, D. MCFADDEN, A. MERRILL, AND T. RIBEIRO (2003): "Healthy, Wealthy and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status," *Journal of Econometrics*, 112, 3–56.

ADDA, J., T. CHANDOLA, AND M. MARMOT (2003): "Socio–Economic Status and Health: Causality and Pathways," *Journal of Econometrics*, 112, 57–63.

ATTANASIO, O. P. AND C. EMMERSON (2003): "Mortality, Health Status, and Wealth," *Journal of the European Economic Association*, 1, 821–850.

BAKER, M. AND D. BENJAMIN (1999): "Early Retirement Provisions and the Labor Force Behavior of Older Men: Evidence from Canada," *Journal of Labor Economics*, 17, 724–756.

BANKS, J., M. MARMOT, Z. OLDFIELD, AND J. P. SMITH (2006): "The SES Health Gradient on Both Sides of the Atlantic," *NBER Working Paper*, No. 12674.

BARNAY, T. (2007): "Redistributive Impact of Differential Mortality in the French Pay-as-You-Go System," *Geneva Papers on Risk and Insurance: Issues and Practice*, 32, 570–582.

BECKMANN, K. (2000): "A Note on the Tax Rate Implicit in Contributions to Pay–As–You–Go Public Pension System," *FinanzArchiv*, 57, 63–76.

BERKEL, B. AND A. BOERSCH-SUPAN (2004): "Pension Reform in Germany: The Impact on Retirement Decisions," *FinanzArchiv*, 60, 393–421.

BLOOM, D. E., D. CANNING, AND M. MOORE (2005): "The Effect of Improvements in Health and Longevity on Optimal Retirement and Saving," *Harvard Initiative of Global Health Working Paper Series*, No. 2.

——— (2007): "A Theory of Retirement," *NBER Working Paper*, No. 13630.

BLOOM, D. E., D. CANNING, M. MOORE, AND Y. SONG (2006): "The Effect of Subjective Survival Probabilities on Retirement and Wealth in the United States," *NBER Working Paper*, No. 12688.

BLUNDELL, R. AND A. DUNCAN (1998): "Kernel Regression in Empirical Microeconomics," *Journal of Human Resources*, 33, 62–87.

BLUNDELL, R., A. DUNCAN, AND K. PENDAKUR (1998): "Semiparametric Estimation and Consumer Demand," *Journal of Applied Econometrics*, 13, 435–461.

BOERSCH-SUPAN, A. (2000): "A Model under Siege: A Case Study of the German Retirement Insurance System," *Economic Journal*, 110, 24–45.

——— (2004): "Faire Abschläge in der gesetzlichen Rentenversicherung," *Sozialer Fortschritt*, 53, 258–261.

BOERSCH-SUPAN, A. AND H. JÜRGES (2006): "Early Retirement, Social Security and Well-Being in Germany," *NBER Working Paper*, No. 12303.

BOLTON, P. AND M. DEWATRIPONT (2005): *Contract Theory*, Cambridge, MA.: The MIT Press.

BORCK, R. (2007): "On the Choice of Public Pensions When Income and Life Expectancy Are Correlated," *Journal of Public Economic Theory*, 9, 711–725.

BOVENBERG, A. L. AND P. B. SORENSEN (2006): "Optimal Taxation and Social Insurance in a Lifetime Perspective," *CESIfo Working Paper*, No. 1690.

BREYER, F. AND M. KIFMANN (2002): "Incentives to retire later – a solution to the social security crisis?" *Journal of Pension Economics and Finance*, 1, 111–130.

——— (2004): "The German Retirement Benefit Formula: Drawbacks and Alternatives," *FinanzArchiv*, 60, 63–82.

BREYER, F., P. ZWEIFEL, AND M. KIFMANN (2005): *Gesundheitsökonomik*, Berlin: Springer.

BROWN, J. R. (2003): "Redistribution and Insurance: Mandatory Annuitization with Mortality Heterogeneity," *Journal of Risk and Insurance*, 70, 17–41.

——— (2007): "Rational and Behavioral Persepctives on the Role of Annuities in Retirement Planning," *NBER Working Paper*, No. 13537.

BROWNING, E. K. (1975): "Why the Social Insurance Budget Is Too Large in a Democracy," *Economic Inquiry*, 13, 373–388.

BUTRICA, B. A., R. W. JOHNSON, K. E. SMITH, AND C. E. STEUERLE (2006): "The Implicit Tax on Work at Older Ages," *National Tax Journal*, LVIV, 211–234.

CASE, A., D. LUBOTSKY, AND C. PAXSON (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92, 1308–1334.

——— (2007): "The Income Gradient in Children's Health: A Comment on Currie, Shields and Wheatley Price," *NBER Working Paper*, No. 13495.

CHANG, F.-R. (1996): "Uncertainty and Investment in Health," *Journal of Health Economics*, 15, 369–376.

CLEVELAND, W. S. (1974): "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

CREEDY, J., R. DISNEY, AND E. WHITEHOUSE (1993): "The Earnings-Related State Pension, Indexation and Lifetime Redistribution in the U.K." *Review of Income and Wealth*, 39, 257–278.

CREMER, H., J.-M. LOZACHMEUR, AND P. PESTIEAU (2004): "Social Security, Retirement Age and Optimal Income Taxation," *Journal of Public Economics*, 88, 2259–2281.

——— (2006a): "Retirement Age and Health Expenditures," *Annales d'Economie et de Statistique*, 167–186.

——— (2006b): "Social Security and Retirement Decision: A Positive and Normative Approach," *CEPR Discussion Paper*, No. 5542.

——— (2007): "Disability Testing and Retirement," *B.E. Journals in Economic Analysis and Policy: Contributions to Economic Analysis and Policy*, 7, 1–32.

CREMER, H. AND P. PESTIEAU (1996): "Redistributive Taxation and Social Insurance," *International Tax and Public Finance*, 3, 281–295.

CURRIE, A., A. SHIELDS, AND W. PRICE (2007): "The ChildHealth/Family Income Gradient: Evidence from England." *Journal of Health Economics*, 26, 213–232.

CURRIE, J. AND M. STABILE (2003): "Socioeconomic Status and Health: Why is the Relationship Stronger for Older Children?" *American Economic Review*, 93, 1813–1823.

CUTLER, D., A. DEATON, AND A. LLERAS-MUNEY (2006): "The Determinants of Mortality," *Journal of Economic Perspectives*, 20, 97–120.

DARDANONI, V. AND A. WAGSTAFF (1987): "Uncertainty, Inequalities in Health and the Demand for Health," *Journal of Health Economics*, 6, 283–290.

DATAR, A., A. GHOSH, AND N. SOOD (2007): "Mortality Risks, Health Endowments, and Parental Investments in Infancy: Evidence from Rural India," *NBER Working Paper*, No. 13649.

DAVIDOFF, T., J. R. BROWN, AND P. A. DIAMOND (2005): "Annuities and Individual Welfare," *American Economic Review*, 95, 1573–1590.

DAVIES, J. B. AND P. KUHN (1992): "Social Security, Longevity, and Moral Hazard," *Journal of Public Economics*, 49, 91–106.

DEATON, A. (2006): "Global Patterns of Income and Health: Facts, Interpretations, and Policies," *NBER Working Paper*, No. 12735.

DEATON, A. AND C. PAXSON (2004): "Mortality, Education, Income, and Inequality among American Cohorts," *NBER Working Paper*, No. 7140.

DIAMOND, P. A. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95.

——— (2003): *Taxation, Incomplete Markets, and Social Security*, Cambridge, MA.: The MIT Press.

——— (2004): "Social Security," *American Economic Review*, 94, 1–24.

DIAMOND, P. A. AND J. A. MIRRLEES (1971a): "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review*, 61, 8–27.

——— (1971b): "Optimal Taxation and Public Production II: Tax Rules," *American Economic Review*, 61, 261–278.

——— (1978): "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics*, 10, 295–336.

DINARDO, J. AND J. L. TOBIAS (2001): "Nonparametric Density and Regression Estimation," *Journal of Economic Perspectives*, 15, 11–28.

DORN, D. AND A. SOUSA-POZA (2005): "Early Retirement: Free Choice or Forced Decision?" *CESIfo Working Paper*, No. 1542.

DUGGAN, J. E., R. DILLINGHAM, AND J. S. GREENLEES (2007): "Mortality and Lifetime Income: Evidence from U.S. Social Security Records," *IMF Working Paper*, No. 07/15.

EHRLICH, I. (2000): "Uncertain Lifetime, Life Protection, and the Value of Life Saving," *Journal of Health Economics*, 19, 341–67.

——— (2001): "Erratum to 'Uncertain Lifetime, Life Protection, and the Value of Life Saving'," *Journal of Health Economics*, 20, 459–460.

EHRLICH, I. AND H. CHUMA (1990): "A Model of the Demand for Longevity and the Value of Life Extension," *Journal of Political Economy*, 98, 761–782.

ENGLE, R. F., C. W. J. GRANGER, J. RICE, AND A. WEISS (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310–320.

EROSA, A. AND M. GERVAIS (2002): "Optimal Taxation in Life-Cycle Economies," *Journal of Economic Theory*, 105, 338–369.

ETTNER, S. L. (1996): "New Evidence on the Relationshio Between Income and Health," *Journal of Health Economics*, 15, 67–85.

EUWALS, R., D. VAN VUUREN, AND R. WOLTHOFF (2006): "Early Retirement Behavior in the Netherlands—Evidence from a Policy Reform," *CEPR Discussion Paper*, No. 5596.

FAN, J. (2000): "Prospects of Nonparametric Modeling," *Journal of the American Statistical Association*, 95, 1296–1300.

FEHR, H. AND C. HABERMANN (2007): "Welfare Effects of Life Annuities: Some Clarifications," *Economics Letters (forthcoming)*.

FELDSTEIN, M. S. (1974): "Social Security, Induced Retirement, and Aggregate Capital Accumulation," *Journal of Political Economy*, 82, 905–926.

FENGE, R., S. UEBELMESSER, AND M. WERDING (2006): "On the Optimum Timimg of Implicit Social Secrity Taxes Over the Life Cycle," *FinanzArchiv*, 62, 68–107.

FERREIRA, P. C. AND S. D. A. PESSOA (2007): "The Effects of Longevity and Distortions on Education and Retirement," *Review of Economic Dynamics*, 10, 472–493.

FLEISCHHAUER, K.-J. (2007): "The Effects of Pension Reform on Retirement and Human Capital Formation," *University of St. Gallen Discussion Paper*, 2007–03.

FLORENS, J.-P. (2003): "Some Technical Issues in Defining Causality," *Journal of Econometrics*, 112, 127–128.

FRIJTERS, P., J. P. HAISKEN-DENEW, AND M. A. SHIELDS (2005): "The Causal Effect of Income on Health: Evidence from German Reunification," *Journal of Health Economics*, 24, 997–1017.

GERDTHAM, U.-G. AND M. JOHANNESSON (2004): "Absolute Income, Relative Income, Income Inequality, and Mortality," *Journal of Human Resources*, 39, 228–247.

GEWEKE, J. (2003): "Econometric Issues in Using the AHEAD Panel," *Journal of Econometrics*, 112, 115–120.

GORSKI, M., T. KRIEGER, AND T. LANGE (2007): "Pensions, Education and Life Expectancy," *Working Paper of the DFG Research Group Heterogeneous Labor*, No. 07/04.

GRANGER, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–438.

GROSSMAN, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80, 223–255.

——— (2000): "The Human Capital Model," in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, New York and Oxford: Elsevier Science, North-Holland, vol. 1A, 347–408.

GROSSMAN, S. J. AND O. D. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

GRUBER, J. AND D. WISE (1998): "Social Security and Retirement: An International Comparison," *American Economic Review*, 88, 158–163.

GUSTMAN, A. L. AND T. L. STEINMEIER (2001): "Retirement Outcomes in the Health and Retirement Study," *Social Security Bulletin*, 64, 3–17.

——— (2005): "The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth," *Journal of Public Economics*, 89, 441–463.

HANEL, B. AND R. RIPHAHN (2006): "Financial Incentives and the Timing of Retirement: Evidence from Switzerland," *IZA Discussion Paper*, No. 2492.

HAUSMAN, J. A. (2003): "Triangular Structural Model Specification and Estimation with Application to Causality," *Journal of Econometrics*, 112, 107–113.

HECKMAN, J. (2003): "Conditioning, Causality and Policy Analysis," *Journal of Econometrics*, 112, 73–78.

HECKMAN, J. J. (1995): "What Has Been Learned About Labor Supply in the Past Twenty Years?" *American Economic Review*, 83, 116–121.

HENRIET, D. AND J.-C. ROCHET (2006): "Is Public Health Insurance an Appropriate Instrument for Redistribution?" *Annales d'Economie et de Statistique*, 61–88.

HOLMSTROM, B. AND P. MILGROM (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303–328.

——— (1991): "Multitask Principle-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization*, 7, 24–51.

——— (1994): "The Firm as an Incentive System," *American Economic Review*, 84, 972–991.

HURD, M. D. AND J. B. SHOVEN (1986): "The Distributional Impact of Social Security," *NBER Working Paper*, No. 1155.

JEWITT, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56, 1177–1190.

JONES, A. M. A. (2006): "Health and Wealth: Empirical Findings and Political Consequences," *Perspektiven der Wirtschaftspolitik*, 7, 93–112.

KALEMLI-OZCAN, S. AND D. N. WEIL (2002): "Mortality Change, the Uncertainty Effect, and Retirement," *NBER Working Paper*, No. 8742.

KIFMANN, M. (2008): "Age-Dependent Taxation and the Optimal Retirement benefit Formula," *German Economic Review (forthcoming)*.

KIUILA, O. AND P. MIESZKOWSKI (2007): "The Effects of Income, Education and Age on Health," *Health Economics*, 16, 781–798.

KOSTORIS PADOA SCHIOPPA, F. (1990): "Undesirable Redistributions in the Retirement-Public Pension Schemes: The Italian Case Study," *CEPR Discussion Paper*, No. 463.

LAU, M. I. AND P. POUTVAARA (2006): "Social Security Incentives and Human Capital Investment," *Finnish Economic Papers*, 19, 16–24.

LAUTERBACH, K., M. LÜNGEN, B. STOLLENWERK, A. GERBER, AND G. KLEVER-DEICHERT (2006): "Zum Zusammenhang zwischen Einkommen und Lebenserwartung," *Studien zu Gesundheit, Medizin und Gesellschaft, Köln*, No. 01.

LI, Q. AND J. S. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*, Princeton: Princeton University Press.

LINDAHL, M. (2005): "Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income," *Journal of Human Resources*, 40, 144–168.

LINDBECK, A. AND M. PERSSON (2003): "The Gains from Pension Reform," *Journal of Economic Literature*, 41, 74–112.

LINDEBOOM, M. (2006): "Health and Work of Older Workers," in *The Elgar Companion to Health Economics*, ed. by A. M. Jones, Cheltenham, U.K. and Northampton, Mass., 26–35.

LIU, L. AND A. J. RETTENMAIER (2003): "Social Security Outcomes by Racial and Education Groups," *Southern Economic Journal*, 69, 842–864.

LOADER, C. (2004): "Smoothing: Local Regression Techniques," in *Handbook of Computational Statistics*, ed. by J. Gentle, W. Härdle, and Y. Mori, Springer.

MEALLI, F. AND D. B. RUBIN (2003): "Assumptions Allowing the Estimation of Direct Causal Effects," *Journal of Econometrics*, 112, 79–87.

MEER, J., D. L. MILLER, AND H. S. ROSEN (2003): "Exploring the Health-Wealth Nexus," *Journal of Health Economics*, 22, 713–730.

MENCHIK, P. L. (1993): "Economic Status as a Determinant of Mortality Among Black and White Older Men: Does Poverty Kill?" *Population Studies*, 47, 427–436.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 38, 175–208.

MULLIGAN, C. B. AND X. SALA–I–MARTIN (2004): "Internationally Common Features of Public Old-Age Pensions, and their Implications for Models of the Public Sector," *Advances in Economic Analyses & Policy*, 4.

MUURINEN, J.-M. (1982): "Demand for Health," *Journal of Health Economics*, 1, 5–28.

MYLES, G. D. (2002): *Public Economics*, Cambridge: Cambridge University Press.

NELISSEN, J. H. M. (1995): "Lifetime Income Redistribution by the Old-Age State Pension in the Netherlands," *Journal of Public Economics*, 58, 429–451.

——— (1998): "Annual versus Lifetime Income Redistribution by Social Security," *Journal of Public Economics*, 68, 223–249.

PHILIPSON, T. J. AND G. S. BECKER (1998): "Old-Age Longevity and Mortality-Contingent Claims," *Journal of Political Economy*, 106, 551–573.

POTERBA, J. M. (2003): “Some Observations on Health Status and Economic Status,” *Journal of Econometrics*, 112, 65–67.

QUEISSER, M. AND E. WHITEHOUSE (2006): “Neutral or Fair? Actuarial Concepts and Pension-System Design,” *OECD Social, Employment and Migration Working Papers*, No. 40.

REIL-HELD, A. (2000): “Einkommen und Sterblichkeit in Deutschland: Leben Reiche länger?” *Beiträge zur angewandten Wirtschaftsforschung*, 580-00.

RIBEIRO, T. (2002): “Discrete Dynamic Models: A HMM Approach to Estimation and Forecasting Using Panel Survey Sata,” *UC Berkeley Working Paper*.

RIED, M. (1998): “Comparative Dynamic Analysis of the Full Grossman Model,” *Journal of Health Economics*, 17, 383–426.

ROBINS, J. M. (2003): “General Methodological Considerations,” *Journal of Econometrics*, 112, 89–106.

ROBINSON, P. M. (1988): “Semiparametric Econometrics: A Survey,” *Journal of Applied Econometrics*, 3, 35–51.

ROGERSON, W. P. (1985): “The First-Order Approach to Prinicpial-Agent Problems,” *Econometrica*, 53, 1357–1368.

SAEZ, E. (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68, 205–229.

SALANIÉ, B. (1997): *The Economics of Contracts—A Primer*, Cambridge: MIT Press.

SALM, M. (2007): “The Effect of Penions on Longevity: Evidence from Union Army Veterans,” *IZA Discussion Paper*, No. 2668.

SCHILS, T. (2005): *Early Retirement Patterns in Europe: A comparative Panel Study*, Amsterdam: Dutch University Press.

SELDEN, T. (1993): “Uncertainty and Health Care Spending by the Poor: The Health Capital Model Revisited,” *Journal of Health Economics*, 12, 109–115.

SHESHINSKI, E. (1978): “A Model of Social Security and Retirement Decisions,” *Journal of Public Economics*, 10, 337–360.

——— (2003): “Optimum Delayed Retirement Credit,” *CESifo Working Paper*, No. 889.

SHESHINSKI, E. AND Y. WEISS (1981): "Uncertainty and Optimal Social Security Systems," *Quarterly Journal of Economics*, 96, 189–206.

SIMONOVITS, A. (2004): "Designing Benefit Rules for Flexible Retirement with or without Redistribution," *CESifo Working Paper*, No. 1370.

——— (2006): "Optimal Design of Pension Rule with Flexible Retirement: The Two-Type Case," *Journal of Economics (Zeitschrift fur Nationalokonomie)*, 89, 197–222.

SIMONOVITS, A. AND P. ESÖ (2003): "Designing Optimal Benefit rules for Flexible Retirement," *Journal of Pension Economics and Finance*, 2, 273–293.

SMITH, J. P. (1998): "Socioeconomic Status and Health," *American Economic Review*, 88, 192–196.

SOARES, R. R. (2007): "On the Determinants of Mortality Reductions in the Developing World," *Population and Development Review*, 33, 247–287.

STAHLBERG, A.-C. (1989): "Redistribution Effects of Social Policy in a Lifetime Analytical Framework," in *The Political Economy of Social Security*, ed. by B. A. Gustafsson and N. A. Klevmarken, Contributions to Economic Analysis, no. 179, 51–65.

TURRA, C. M. AND O. S. MITCHELL (2004): "The Impact of Health Status and Out–of–Pocket Medical Expenditures on Annuity Valuation," *PARC Working Paper Series*, No. 04–02.

VAN SOEST, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach," *Journal of Human Resources*, 30, 63–88.

VON GAUDECKER, H.-M. AND R. D. SCHOLZ (2007): "Differential Mortality by Lifetime Earnings in Germany," *Demographic Research*, 17, 83–108.

WALDRON, H. (2001): "Links between Early Retirement and Mortality," *ORES Working Paper (U.S. Social Security Administration)*, No. 93.

WILKINSON, R. (1996): *Unhealthy Societies: The Afflictions of Inequality*, London: Routledge.

WOLFE, J. R. (1983): "Perceived Longevity and Early Retirement," *Review of Economics and Statistics*, 65, 544–551.

——— (1985): "A Model of Declining Health and Retirement," *Journal of Political Economy*, 93, 1258–1267.

WORLD BANK (1994): *Averting the Old Age Crisis—Policies to Protect the Old and Promote Growth*, New York: Oxford University Press.

WREDE, M. (1999): "Pareto Efficient Pay–As–You–Go Pension Systems With Multiperiod Lives," *Jahrbücher für Nationalökonomie und Statistik*, 219, 494–503.

YAARI, M. E. (1965): "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer," *Review of Economic Studies*, 32, 137–150.

YATCHEW, A. (1998): "Nonparametric Regression Techniques in Economics," *Journal of Economic Literature*, 36, 669–721.

——— (2003): *Semiparametric Regression for the Applied Econometrician*, Cambridge: Cambridge University Press.

Chapter 2

To Work or to Work Out: A Moral Hazard Interpretation of Labor Supply, Retirement, and Investments in Longevity

2.1 Introduction

Recent empirical findings shed doubt on the monotonous link between ability (as measured by income or wealth) and life expectancy. Though they do not explicitly explain it, von Gaudecker and Scholz (2007) and Clark (2007, p. 98), discover a U–shaped relationship (the former for Germany between 1993 and 2003 and the latter for England in the 17th century), whereas Chapter 4 specifically addresses the downward sloping part of the non–monotonous link. However, this empirical phenomenon has not been satisfactorily explained by theory yet. In the following essay, the main result will include U–shaped incentives for investments in longevity and therefore U–shaped effort in this activity, which is derived within a principal–agent framework.

In my model, an agent is capable of engaging in three different activities, one of which is concerned with investments in longevity (*work out*), and the others being weekly or instantaneous labor supply (*work*) and life–time labor supply (*retirement*). However, effort levels are not perfectly observable, such that the government (the principal) can only condition taxes (or subsidies) on perturbed signals. The principal can levy taxes e.g. on work. These taxes (which may be negative) shift the revenue or outcome from activities between the agent and the principal, and at the same time allocate risk associated with this activity. The government is not strictly benevolent; it has own interests instead, which can be interpreted as the maximization of GDP, measured by output from *work* and length of the working life of the agent. However, although longevity is not valued in its own right by the principal, it is always indirectly considered in the tax system due to the interaction with *work* and *retirement*. Since the observation of the agent's effort levels is distorted, only a second best solution is feasible. I apply the solution mechanism for multi–task moral–hazard situations proposed by Holmstrom and Milgrom (1991), and derive comparative statics of the optimal incentives with respect to the agent's ability.

Several questions in this essay have already been addressed separately in different settings, but not together: Wolfe (1983) models an agent who has two choices, namely retirement age and consumption of leisure, which resembles the activities *work* and *retirement* in my specification. However, both activities are assumed to be perfectly observable, such that no moral hazard can arise. Instead, Wolfe (1983) finds adverse selection, since the agent chooses her retirement age (early or postponed) based on her life expectancy, which is not observed by the principal of the pension system. Individuals with high life expectancy benefit from a pension system which is actuarially fair for the *average* pensioner only. A similar reasoning is applied by Diamond (2003), who shows that agents above a certain threshold of life expectancy or productivity can exploit the tax/pension system, because the principal has to keep incentives for agents with more intermediate characteristics. Also in an approach applying optimal taxation, Cremer et al. (2004) analyze the interaction of health and productivity with respect to the resulting optimal tax scheme and find that delayed retirement is (relatively) taxed in order to subsidize the respective 'weaker' type, in order to enable him to retire earlier. Late retirement in their setting is still preferred by individuals with high productivity or good health, which are assumed to be positively correlated. Here, life expectancy is still exogenous (and equal across all types of individuals), and retirement age is assumed to by fully observable.

Introducing moral–hazard, Davies and Kuhn (1992) endogenize life expectancy by explicitly modelling longevity–enhancing behavior, however, as opposed to the models of optimal taxation and in contrast to the model I present, income is assumed to be exogenous. The possible trade–off between consumption and investments in longevity is derived from the budget constraint and the utility decreasing or increasing effect of the investment. The authors find moral–hazard, namely excessive investments in longevity, inflicting damage on *private* providers of annuities who, however, do not cease to exist, despite moral–hazard.[1] Under the presence of private or public old–age annuities, Philipson and Becker (1998) find a similar moral–hazard effect, where longevity is also endogenous. The authors additionally track the interaction with public health insurance, which potentially accelerates the moral–hazard effect, as no market prices govern the scarcity of longevity–enhancing means. In the following analysis, incentives for *work out* can be interpreted as a public health insurance system.

Endogenized life expectancy relates to Grossman (1972) and Grossman (2000), where investments in health capital consume time and mitigate instantaneous consumption, but allow higher productivity in the future and potentially delay the time at which the stock of health capital will be consumed. See additionally Wolfe (1985) for a treatment of this topic, who takes retirement explicitly into account as a special means of investment in one's health capital. Yet, in his setting, he fixes life expectancy, leaving an open issue to analyze. The main contribution of Ehrlich and Chuma (1990) is the explicit discrimination of the demand for *health* from the demand for *longevity*, the latter being directly dependent on the stock of health capital in Grossman (1972) and Grossman (2000). My analysis differs in one main aspect from Grossman (1972), Grossman (2000), or Ehrlich and Chuma (1990), since I neither apply the notion of health capital, nor use any initial endowments in capital, health-related or physical.

[1] Intuitively, Davies and Kuhn (1992) can also prove that mandatory annuities larger than those offered and bought on the private market reduce utility. They argue that a second best mandatory annuity scheme would have to be actuarially unfair.

The essay is organized as follows: In the following Section 2.2, I present the model, namely the interests of the agent and the principal, the technology that transforms inputs of the agent into output for the principal, the perturbations, and the tax system (the incentives). In Section 2.3, I solve the model and present comparative statics of the incentives with respect to a variation of the exogenous parameters, whereas Section 2.4 concludes and gives an outlook on further possibilities to enhance the presented model.

2.2 The Model

2.2.1 The Agent and the Principal

The representative agent behaves as if he were maximizing the following utility function, which is additively separable in monetary compensation t and effort q (time indices τ are omitted):

$$\widetilde{U} = \int_{0}^{q_2} \left[u\left[t(q)\right] - \widetilde{c}(q)\right] d\tau + \int_{q_2}^{\widetilde{f}(q_3)} \left[u\left[t(q)\right] - \widetilde{c}(q)\right] d\tau, \tag{2.1}$$

with

$$q = (q_1, q_2, q_3). \tag{2.2}$$

I refer to the elements of the activity vector q as

q_1: Effort spent on weekly (instantaneous) labor supply or labor intensity (*work*),

q_2: Effort spent on life time labor supply or number of years worked (*retirement*), and

q_3: Effort spent on investments in own longevity (*work out*).

Here, utility depends positively (and concave) on payments made to the agent t and negatively on effort costs $\widetilde{c}$, which are convex in effort levels. The monetary compensation to the agent depends indirectly on his effort levels, is constant over time, and is directly transformed into a consumption good. The first term of Equation (2.1) denotes utility and effort costs during his working life, which ends with retirement q_2, whereas the second term denotes utility from retirement until the day the agent dies. Although q_2 as integral bound vanishes, *retirement* still plays an indirect role, because of its influence on consumption and effort costs. Life expectancy is an increasing function $\widetilde{f}$ of the total *work out* the agent undertakes. If individual discounting and interest rate are zero and if capital markets are perfect, the agent will choose perfectly constant paths of the control variables q_1 and q_3, and life expectancy is a function of the instantaneous level of *work out*. The utility function therefore simplifies to

$$\widetilde{U} \quad = \quad f(q_3)\left[u\left[t(q)\right] - \widetilde{c}(q)\right], \tag{2.3}$$

where the payment net of taxes t are now either life-time consumption or average instantaneous consumption. I now assume that the properties of $f(q_3)$ can be at least approximately captured in the transformed cost function $c(q)$—which is not necessarily strictly convex in q_3 anymore—and in transformed utility from consumption. Finally, assuming exponential utility from consumption (with r being the constant parameter of absolute risk aversion), the reduced form objective function as applied from here on is

$$U(t, q) \quad = \quad -e^{-r \cdot q_3 t} - c(q). \tag{2.4}$$

In general, the properties of the cost function $c(q)$ can be interpreted as abilities innate to the agent. Due to the interaction of the activities and the private benefits from *work out*, the cost function is not necessarily convex in all arguments. The first and third activity are substitutes in the total cost function $c(q)$, such that private investments in longevity reduce ceteris paribus the possibilities (or willingness) to increase the labor intensity. Both activities, *work* and *work out*, are competing for the agent's weekly time endowment. However, the second and third activity are complements in the sense that investments in longevity ceteris paribus reduce the marginal cost of delayed retirement (or prolonged work life). At the same time, delayed retirement makes it more worthwhile to invest in longevity. Formally, these assumptions are summarized in the following cost function,

$$c(q) \quad = \quad \frac{c_1}{2a}q_1^2 + \frac{c_2}{2a}q_2^2 + \frac{c_3}{2a}q_3^2 + \frac{c_{13}}{a}q_1 q_3 + a c_{23} q_2 q_3, \tag{2.5}$$

which has the desired properties once all parameters are strictly positive, except c_{23}, which is strictly negative, such that the cross derivative of $c(q)$ with respect to q_2 and q_3 is negative. All c_i and c_{ij} parameters measure effort costs associated with the activities, and the common productivity parameter a reduces effort costs wherever possible. Via this specification, all third derivatives of $c(q)$ are zero, and I fully describe the agent by her ability a.

The risk-neutral principal (the government) maximizes

$$W = Q(q) - t[x(q)], \tag{2.6}$$

where Q is aggregate output from labor supply (as a measure for GDP, e.g.), x is a signal about the effort levels, and t is the net payment to the agent, conditional on the observed signals. Here, aggregate output will only depend on q_1 and q_2, such that the principal is only indirectly concerned with life expectancy of the agent. Such a behavior is still consistent with a government that may tax unhealthy behavior and subsidize healthy behavior. However, the government does drive neither of both incentives to their extremes, as *purely* healthy and longevity–enhancing activities may interfere with other goals (which is, in this case, depicted by the cost function). Additionally, an agent may even enjoy unhealthy activities, and the principal will not take away the freedom of doing so.

2.2.2 The Technologies

Assume that input supplied q is transformed into output received by

$$Q = Q(q_1, q_2). \tag{2.7}$$

Total output Q represents the aggregate of weekly labor supply and life–time labor supply, hence $Q = Q(q_1 \cdot q_2)$. However, in the neighborhood of the equilibrium, I approximate this function by a linearization around the equilibrium values of q_1 and q_2. I denote the vector of all $\partial Q/\partial q_i$ by w_i, and output takes the specific form

$$Q = w_1 q_1 + w_2 q_2 = w' q, \tag{2.8}$$
$$w = (w_1, w_2, 0). \tag{2.9}$$

The vector w can be interpreted as the *technology* which transforms input into output, such that Q resembles a production function with constant returns to scale around the equilibrium. The respective weights can also be understood as the principal's preferences, which, however, does not matter for the following analysis.

The principal observes effort levels q only with error, and the signal vector is

$$x = q + \theta, \tag{2.10}$$

where $\theta = (\theta_1, \theta_2, \theta_3)$ is distributed according to a three–dimensional normal distribution with the variance-covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}. \tag{2.11}$$

The model is consistent with $\sigma_i^2 \to \infty$ $(i = 1, 3)$, which corresponds to completely unobservable effort levels in activity i or a completely uninformative signal. The length of the work life, however, is fully observable to the principal, and all covariances are zero. This is not as restrictive as it may seem, since zero covariances only restrict the *observability*, and not possible interactions in the agent's abilities, which are solely expressed by the cross–derivatives of the cost function. The perturbations can be interpreted in two different ways: First, they impose a risk on the agent; hence the activity itself might be referred to as being risky. Second, the perturbations prevent observability and therefore possibilities to control, reward, or punish the agent. The perturbation term might be distributed according to any other distribution function with a single–peaked and non–monotonous density, the explicit results, however, are derived under the assumptions of a normal distribution.

2.2.3 The Tax/Pension System

The set of feasible tax systems is limited to a set of affine functions,[2] such that

$$t[x(q)] = \alpha' x(q) + \beta. \tag{2.12}$$

Note that $(1 - \alpha_i)$ is a tax rate on activity i. The respective tax (or subsidy) rates can only be conditioned on the observations x, and not on output supplied q directly. In general, a tax system could include the reduction of the allowed activities to $q = (q_1, q_2)$. Such a policy is ruled out as not being feasible.

[2]Which is not restrictive, as affine functions appear to be optimal under the conditions that apply here, see Holmstrom and Milgrom (1987).

For the case of *work*, $1-\alpha_1$ is equivalent to a payroll tax on weekly labor supply, as long as α_1 is strictly between zero and one; following this, $\alpha_1 = 1$ denotes the absence of a tax on work, while $\alpha_1 > 1$ denotes a subsidy. Whenever $\alpha_2 \in (0,1)$, there is a positive (*implicit*) tax on prolonged work beyond the absolute minimum, imposed by a pension system which is assumed to be mandatory. Without loss of generality, I normalize the minimum life–time labor supply to zero, which is possible due to the linearization of aggregate output Q. If $\alpha_2 = 1$, this implies actuarial fairness of the pension system with regard to the length of a career, whereas $\alpha_2 > 1$ is again a subsidy, over–proportionally encouraging late retirement. A further interpretation of the aggregate of α_1 and α_2 is that taxation of *work* depends on age; see e.g. Diamond (2003) for an application of the concept of age–dependent taxation.

For the case of *work out*, the interpretation of α_3 as a tax on this activity is not as straight forward as above. But assume that it is possible to levy a tax on the consumption of a longevity enhancing good. The nature of this good determines the degree of observability of the signal x_3; consulting a doctor or buying vitamins might be relatively easy to observe, while exercising at home is not. Any $\alpha_3 \in (0,1)$ indicates that only part of the amount of effort spent on *work out* becomes effective, hence that this activity is taxed. A subsidy is also possible, and even a negative α_3 can be thought of: Longevity enhancing behavior is not only discouraged, but the agent is actually invited to spend negative effort on q_3, namely to stay away from the doctor's and to consume some potato chips in front of the television. Finally, the policy parameter β is a constant basic income (or flat tax, depending on the sign) for the agent, independent of his actions, which ensures that the agent's participation constraint can always be fulfilled.

2.3 The Solution

2.3.1 The First Order Approach

In the general single–agent and single–task moral–hazard model, two conditions are always sufficient for the validity of the first order approach, namely the monotone likelihood ratio property (MLRP) and the convexity of distribution function condition (CDFC).[3] However, these conditions are generally too strong to be applicable in the multi–task moral–hazard model, such that in the specific case analyzed here, another set of conditions is sufficient, as Holmstrom and Milgrom (1991) show. Together with the assumption of normally distributed θ and exponential utility, the term

[3] For a textbook coverage, see e.g. Bolton and Dewatripont (2005, pp. 142).

$$\left(\frac{\partial c(q)}{\partial q}\right)' \Sigma \frac{\partial c(q)}{\partial q} \tag{2.13}$$

has to be convex in q.[4] Under the given assumptions, this expression factors out to

$$\left(\frac{c_1}{a} q_1 + \frac{c_{13}}{a} q_3\right)^2 \sigma_1^2 + \left(\frac{c_{13}}{a} q_1 + a c_{23} q_2 + \frac{c_3}{a} q_3\right)^2 \sigma_3^2. \tag{2.14}$$

The determinant of the Hessian of the above expression is zero, hence the Hessian is positive semi–definite and expression (2.13) is convex in q. Applying the exponential form for U, utility maximization is equivalent to the maximization of the certainty equivalent U_{CE} of U,

$$U_{\text{CE}} \;=\; \alpha' E(x) + \beta - c(q) - \frac{1}{2} r \alpha' \Sigma \alpha \tag{2.15}$$

over the choice of q for a given α and β. In general, I have the following incentive compatibility constraint,

$$q^* \;=\; \arg\max_q U_{\text{CE}}, \tag{2.16}$$

which simplifies under validity of the first order approach and the assumptions on $c(q)$ to to the explicit effort levels

$$q^* \;=\; \Phi^{-1} \left(\begin{array}{c} a c_2 \beta (c_{13} - a\alpha_1) \\ a^2 c_1 \beta (a c_{23} - \alpha_2) \\ -a c_1 c_2 \beta \end{array}\right), \tag{2.17}$$

with

$$\begin{aligned} \Phi \;=\; & c_{13}^2 c_2 + a^4 c_1 c_{23}^2 - 2 a c_{13} c_2 \alpha_1 + a^2 c_2 \alpha_1^2 \\ & + c_1 \left[a^2 \alpha_2 (\alpha_2 - 2 a c_{23}) - c_2 (c_3 - 2 a \alpha_3)\right]. \end{aligned} \tag{2.18}$$

[4] MLRP and CDFC are actually not satisfied together with the imposed assumptions—while MLRP is violated by $\partial Q / \partial q_3 = 0$, CDFC does not hold, because the normal distribution is not convex. The problem is addressed by Jewitt (1988), who shows how the relaxation of one assumption has to be accompanied by stricter versions of the others.

In general, the agent reacts to changes in incentive payments as shown in Table 2.1. The reaction of effort levels to the (indirect) incentives for the respective other activities depends on the substitutive or complementary nature of these activities. It is positive for the incentives for *retirement* and *work out*, as one of these two activities alleviates the effort costs for the other. The reaction to all other indirect incentives is negative, because the respective activities are substitutes to each other, either direct (*work* and *work out*) or indirect (*work* and *retirement*).

A special focus of this paper is on the variation of q_3 with respect to ability a. The effect cannot be signed independently of a: For small a, the effect is negative, whereas for large a, I find a positive slope, such that the function $q_3(a, \cdot)$, holding everything else constant, is U–shaped. Interpretation of the partial effects of the exogenous parameters on *ceteris paribus* effort levels (for a given set of incentives) has to take into account that 'small' c_{23} actually means $|c_{23}|$ being large, hence the interaction of *retirement* and *work out* being strong. If, on the other hand, c_{23} is large (close to zero), this is only compatible with negative incentives for *retirement*, hence with $\alpha_2 < 0$, which, however, is possible under the given assumptions (see Section 2.3.3); c_{13} being small means that c_{13} fulfills $c_1 c_3 > c_{13}^2$, which corresponds to the further condition in Table 2.1 of c_1 and c_3 being large.[5]

Except of the derivatives of q_3 with respect to c_1, c_2, and c_3, none of the effects can be uniquely signed. The main determinants of ambiguous signs is the relative size of the cross derivatives of the cost function, hence of c_{13} and c_{23}. This effect is due to their augmenting and reducing property in the cost function: If, for example, both cross derivatives are small, effort in activity *work* is decreasing in all c_i and decreasing in c_{23}. Naïvely, effort levels should decline in direct effort costs. Here, also the indirect effects play a role, such that q_2^* reacts positively on all c_i, because if effort costs are high, the best the agent can do, is to delay *retirement*, in order to increase payments and to exploit the cost reducing interaction with *work out*.

If the agent spends high effort on *work*, this increases the effort costs for *work out*. However, high costs for *work out* do not hurt so much, if the interaction of *work out* with *retirement* is strong. So once c_{23} is already small (indicating a strong interaction), a further decrease away from zero increases the effort level for *work*.

2.3.2 First Best Solution

The principal is risk neutral and the agent has exponential utility, hence maximization of joint surplus $E(W) + U(t, q)$ is equivalent to the maximization of

[5] Otherwise, some solutions to the model are exclusively complex and not real. This assumption will continue to hold throughout the rest of the paper.

Sign of $\partial q^* / \partial x$

	exogenous parameters x								
	α_1	α_2	α_3	a	c_1	c_2	c_3	c_{13}	c_{23}
q_1^*	$+$ if c_2, c_3 large	$-$ if c_{13}; c_{23} small; large	$+$ if c_{13} small	$+/-$	$-$ if c_{13}; c_{23} small; large	$-$ if c_{13} small	$-$ if c_{13} small	$+/-$	$+$ if c_{13}; c_{23} small; large
q_2^*	$-$ if c_{13}; c_{23} small; large	$+$ if c_2, c_3 large	$-$ if c_{23} large	$+/-$	$+$ if c_{23} large	$+$ if c_{23}, c_1, c_3 large	$+$ if c_{23} large	$+$ if c_{13}; c_{23} small; large	$+/-$
q_3^*	$+$ if c_{13} small	$-$ if c_{23} large	$+$ always	$+$ if a large	$-$ always	$-$ always	$-$ always	$-$ if c_{13} small	$+$ if c_{23} large

Table 2.1: Comparative Statics of Effort Levels

$$Q + U_{\text{CE}} = Q - c(q) - \frac{1}{2} r\alpha'\Sigma, \tag{2.19}$$

where U_{CE} is again the certainty equivalent of $U(t, q)$. The first best solution is reached once all activities are deterministically translated into observed output, hence when all σ_i^2 are zero. The incentives α^* are then given by

$$\begin{pmatrix} \alpha_1^* \\ \alpha_2^* \\ \alpha_3^* \end{pmatrix} = \begin{pmatrix} w_1 \\ w_2 \\ 0 \end{pmatrix}. \tag{2.20}$$

The incentives are simply the weight the principal assigns to the respective activities. The optimal effort levels of the first best solution are given once the general incentives α_i in Equation (2.17) are replaced by the weights in Equation (2.20).

2.3.3 Second Best Solution

Now, allowing for strictly positive elements in Σ, the second best solution is again reached by maximization of the joint surplus:

$$\max_{\alpha} E(W) + U_{CE} \tag{2.21}$$

I refrain from the lump sum payment β, which I only use in order to fulfill the participation constraint $U \geq \overline{U}$. The vector of marginal tax rates is given by

$$\alpha^* = \left(I + rc''\Sigma\right)^{-1} w' \tag{2.22}$$

with $\alpha^* = (\alpha_1^* \quad \alpha_2^* \quad \alpha_1^*)'$, I being the 3×3 identity matrix, c'' the 3×3 matrix of second derivatives of $c(q)$, and w and Σ as defined in Equations (2.9) and (2.11), respectively. Solving (2.22) explicitly, I have

$$\alpha^* = \begin{pmatrix} \Psi^{-1} a w_1 \left(a + c_3 r \sigma_3^2\right) \\ \Psi^{-1} a^2 w_1 c_{13} c_{23} r^2 \sigma_1^2 \sigma_3^2 + w_2 \\ -\Psi^{-1} a w_1 c_{13} r \sigma_1^2 \end{pmatrix}, \tag{2.23}$$

with

$$\Psi \;=\; a\left(a + c_1 r\sigma_1^2\right) + r\sigma_3^2\left[ac_3 + r\sigma_1^2\left(c_1c_3 - c_{13}^2\right)\right]. \tag{2.24}$$

In any case, α_1^* has always a different sign than α_3^*. The sign of α_2^* depends on the relative strength of the exogenous parameters—especially on the weight w_2 and the absolute size of c_{23}. Comparison of all α_i^* from the first best with the second best solution does not yield a clear answer and depends on the relative size of respective parameters, so incentives for *work*, *retirement*, and *work out* may be larger or smaller and can even be negative in all cases (but not at the same time).[6]

Since each of the optimal effort levels q depend on all α_i, and the transition from first best to second best does not allow to sign all α_i uniquely, the change of the optimal effort levels is also ambiguous. For the case of either *work* or *work out* being unobservable (hence, for $\sigma_1^2 \to \infty$ or $\sigma_3^2 \to \infty$), see Appendix 2.A.

2.3.4 Comparative Statics

See Table 2.2 for a summary of comparative statics with respect to the exogenous parameters, computed under the already known assumption of c_{13} being small.

SIGN OF $\partial\alpha^*/\partial x$

	exogenous parameters x										
	w_1	w_2	σ_1^2	σ_3^2	r	a	c_1	c_2	c_3	c_{13}	c_{23}
α_1^*	$+$	0	$-$	$+$	$-$	$+$	$-$	0	$-$	$+$	0
α_2^*	$-$	$+$	$-$	$-$	$-$	$-$	$+$	0	$+$	$-$	$+$
α_3^*	$-$	0	$-$	$+$	$+/-$	$+/-$	$+$	0	$+$	$-$	0

Table 2.2: Comparative Statics of Incentives

With the exception of the derivatives of α_3^* with respect to risk aversion r and ability a, all other derivatives of α_i $(i = \{1, 2, 3\})$ are uniquely signed. The weights have the predictable influence on incentives, namely that the higher the weight on a specific activity, the higher is the direct incentive. If another activity interferes, incentives for this alternative activity may decrease. Variance or risk in the observation of *work* reduces incentives for all activities in order to reduce the agent's exposure to that risk. Higher σ_3^2 leads to a lower reward for delayed *retirement* and

[6]Negative incentives arise for *work* if Ψ is negative, hence if c_{13} is large, and vice versa for *work out*.

a higher reward for *work*, because the principal wants the agent to shift effort from one activity to the other.

The higher the effort costs for *work*, the lower are the respective incentives, because the principal urges the agent to shift attention from *work* to delayed *retirement*. This ceteris paribus result is already described in general by Holmstrom and Milgrom (1991). Yet, the effort costs associated with delayed *retirement* have no influence on incentives, as this effort is perfectly observable and therefore enforceable. Even though the direct benefits of *work out* do not accrue to the principal, incentives change with the increased effort costs associated with it; the fraction of output from *work* that is distributed to the agent can be lower, while the reward for delayed *retirement* has to increase.

If the interaction between *work* and *work out* is strong, the principal is willing to pay more for *work*, while the payments for *work out* are even more discouraging. Yet, the interaction of *retirement* and *work out* does not change incentives for the latter, as *retirement* is perfectly observable and does not need high–powered indirect incentives. For larger c_{23} (meaning, for c_{23} being closer to zero), incentives for delayed *retirement* increase.

Now, focus on the important relationship between incentives and ability a. Incentives for *work* increase in ability, such that more able agents are encouraged to work more (e.g., by lower taxation for work), but incentives for delayed *retirement* actually decrease in a, which corresponds to higher implicit taxation or a progressive pension system. This relates to the theoretical findings of Bommier et al. (2005), who derive that the implicit tax on delayed retirement should be increasing in the retirement age. Finally, the relationship between incentives for *work out* and ability is not monotone. For small a, incentives decrease in a, whereas for large a, incentives increase. The threshold[7] is at

$$\widetilde{a} \;=\; \sqrt{\left(c_1 c_3 - c_{13}^2\right) r^2 \sigma_1^2 \sigma_3^2}, \tag{2.25}$$

meaning that the relationship is U–shaped in the positive domain of a and as long as all σ_i^2 are finite.[8] Since actual effort levels correspond monotonously to incentives, investments in longevity and ability exert a U–shaped relationship as well. This phenomenon can be summarized and explained as follows: An agent with low ability is encouraged to *work out* and to delay *retirement*, but not to *work*. This least able agent is not encouraged to both q_1 and q_3 at the same time, because of the

[7]The same threshold $\widetilde{a}$ from Equation (2.25) governs the sign of $\partial\alpha_3^*/\partial r$, for small a this derivative is greater than zero and vice versa. In terms of the parameter of risk aversion itself, incentives for *work out* increase for very risk averse agents, while they decrease if the agent is less risk averse.

[8]Note that while c_{13} is large, the threshold $\widetilde{a}$ is not in the real domain, which is ruled out.

cost augmenting interaction of both. He is better to be encouraged to the activities which interact as complements. A very able agent is encouraged to *work* and to *work out*, but not to delay *retirement*, which is not necessary, because the most able agent will choose relatively late *retirement* anyhow according to the interaction with his high effort in *work out*. But due to his high ability, it is worthwhile to encourage him to *work* as well. Eventually, an agent with intermediate ability is not encouraged to *work out* and faces moderate incentives for the other activities; incentives for q_3 are necessary only as a means to encourage either q_1 or q_2 indirectly. Here, incentives both facilitate *work* and delayed *retirement*, such that the aggregate indirect effects of *work out* on effort costs at least partially cancels out—incentives for *work out* are not effective in promoting the other activities.

The result that incentives for *work out* or investments in longevity as well as associated effort levels are not monotone, but U–shaped in the agent's ability relates to empirical findings e.g. by von Gaudecker and Scholz (2007), Clark (2007), and Chapter 4 of the present thesis. The downward sloping area of the function that maps life expectancy to income may not be an artifact of the respective data sets or the methodology, but expression of the underlying economic mechanism outlined in this essay.

2.4 Conclusion

In a Holmstrom and Milgrom (1991) moral–hazard model with multiple activities, I model an agent who spends effort on *work*, *retirement*, and *work out*. The effort costs of these activities interact, such that *work out* increases costs for *work*, but decreases costs associated with delayed *retirement*. Under certain assumptions on the parameters, the second best optimal incentives for *work out* are U–shaped in the agent's ability, and so is the optimal effort level in this activity. High incentives for *work out* for the least able encourage the agent to exploit the cost–reducing properties of the cost function and come with high incentives for the complement, namely delayed *retirement*. The most able agent faces high–powered incentives for *work out* mainly due to their indirect effect on delayed *retirement*, which allows to reduce incentives for this activity. In order to simultaneously promote *work* for the most able, incentives for it are high as well.

Of course, this model only serves as a first step from which further research on the potentially non–monotone relation between ability and longevity could depart. One direction includes the extension of multiple agents, either homogeneous or heterogeneous in ability. Risk will play a minor role in such a setting, once a relative incentive scheme is feasible for the principal (relative incentives in a nationwide tax system may seem unusual, but the public pension system in Germany e.g., conditions the earned benefit claims explicitly on the *relative* position of each agent among all others). Another direction may refine the labor supply decision of the agent and explicitly incorporate income effects, which work against a monotonously increasing labor supply, either weekly or with respect to retirement (see Chapter 4 for a brief conjecture on the effect of heterogeneous elasticities of labor supply).

2.A Appendix: Optimal Incentives with Infinite Variances

If either one of the perturbed activities *work* or *work out* cannot be observed at all, optimal incentives are given by

$$\alpha^*|_{\sigma_1^2\to\infty} = \begin{pmatrix} 0 \\ \Psi_1^{-1}a^2 w_1 c_{13} c_{23} r\sigma_3^2 + w_2 \\ -\Psi_1^{-1} a w_1 c_{13} \end{pmatrix} \tag{2.26}$$

with $\Psi_1 = ac_1 + r\sigma_3^2\left(c_1c_3 - c_{13}^2\right)$, and

$$\alpha^*|_{\sigma_3^2\to\infty} = \begin{pmatrix} \Psi_3^{-1} a w_1 c_3 \\ \Psi_3^{-1}a^2 w_1 c_{13} c_{23} r\sigma_1^2 + w_2 \\ 0 \end{pmatrix} \tag{2.27}$$

with $\Psi_3 = ac_3 + r\sigma_1^2\left(c_1c_3 - c_{13}^2\right)$. Obviously, incentives cannot be conditioned on unobservable effort levels, but, more importantly, if one of the effort levels becomes unobservable, incentives for the other activities change. Focus on the latter, $\sigma_3^2 \to \infty$, which is more plausible (it is easier to observe and therefore to tax instantaneous labor supply than potentially private activities that enhance longevity): Assuming that risk aversion is larger than one, and σ_i^2 being not smaller than one, I have $\Psi_3 < \Psi$. However, the denominator of $\alpha_i^*|_{\sigma_3^2\to\infty}$, $i = 1, 2$ is also smaller in the case of unobservable *work out*, such that the change of α_i^* is not uniquely signed.

References of Chapter 2

BOLTON, P. AND M. DEWATRIPONT (2005): *Contract Theory*, Cambridge, MA.: The MIT Press.

BOMMIER, A., M.-L. LEROUX, AND J.-M. LOZACHMEUR (2005): "Varying Life Expectancy and Social Security," *mimeo*.

CLARK, G. (2007): *A Farewell to Alms*, Princeton: Princeton University Press.

CREMER, H., J.-M. LOZACHMEUR, AND P. PESTIEAU (2004): "Social Security, Retirement Age and Optimal Income Taxation," *Journal of Public Economics*, 88, 2259–2281.

DAVIES, J. B. AND P. KUHN (1992): "Social Security, Longevity, and Moral Hazard," *Journal of Public Economics*, 49, 91–106.

DIAMOND, P. A. (2003): *Taxation, Incomplete Markets, and Social Security*, Cambridge, MA.: The MIT Press.

EHRLICH, I. AND H. CHUMA (1990): "A Model of the Demand for Longevity and the Value of Life Extension," *Journal of Political Economy*, 98, 761–782.

GROSSMAN, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80, 223–255.

——— (2000): "The Human Capital Model," in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, New York and Oxford: Elsevier Science, North-Holland, vol. 1A, 347–408.

HOLMSTROM, B. AND P. MILGROM (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303–328.

——— (1991): "Multitask Principle-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization*, 7, 24–51.

JEWITT, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56, 1177–1190.

PHILIPSON, T. J. AND G. S. BECKER (1998): "Old-Age Longevity and Mortality-Contingent Claims," *Journal of Political Economy*, 106, 551–573.

VON GAUDECKER, H.-M. AND R. D. SCHOLZ (2007): "Differential Mortality by Lifetime Earnings in Germany," *Demographic Research*, 17, 83–108.

WOLFE, J. R. (1983): "Perceived Longevity and Early Retirement," *Review of Economics and Statistics*, 65, 544–551.

——— (1985): "A Model of Declining Health and Retirement," *Journal of Political Economy*, 93, 1258–1267.

Chapter 3

Rich and Healthy—Better than Poor and Sick? An Empirical Analysis of Income, Health, and the Duration of the Pension Benefit Spell

3.1 Introduction

The length of the period between retirement and death (the duration of the benefit spell) is the crucial determinant for the rate of return from the German public pension system. Contributions are conditioned on income, whereas monthly benefits are paid according to the amount of benefit claims and the current value of these claims. The benefit claims are measured in points, and one point corresponds to the claims earned by the contributions based on one year of the average income. Since this point value is equal for all pensioners in each year, the individual monthly pension is proportional to the total amount of individual contributions. Yet, people tend to receive their pension for different periods, such that the total amount of benefits is not necessarily proportional to the contributions. Differences occur for two reasons, namely due to differences in the retirement age and differences in life expectancy.

The key question is whether differences in duration (and hence, in the internal rate of return from the pension system) are systematic. If they are not, the public pension simply fulfils its task as an insurance against longevity, and individual deviation from the average rate of return is random. If, however, the duration under the benefit spell varies systematically with income, e. g., then a potential for unintended redistribution among different income groups arises.

Especially life expectancy is usually perceived to be increasing in socio–economic factors such as income. Recent empirical findings at least partially corroborate this conjecture, among them von Gaudecker and Scholz (2007).[1] Other things—especially retirement age—equal, this finding alone would be an indication for redistribution from the bottom to the top of the income distribution. Additionally, retirement age might be a function of income as well. Evidence in this case is controversial, see Berkel and Boersch-Supan (2004) versus Schils (2005) with opposing results with respect to the impact of income or wealth on retirement age. The impact of socio–economic factors on the aggregate of both, namely on the duration, has been examined by Lauterbach et al. (2006). They find that duration is indeed higher for individuals with higher income. Still, the death cohort they observe is rather small, and benefit payments before the age of 65 are neglected. In contrast to this analysis, I propose to include any kind of pension benefits, especially those for early retirees and disability pensions. Another recent empirical contribution

[1] However, in their seminal paper, Adams et al. (2003) find no causal link of socio-economic status on mortality. Instrumenting income or wealth in order to account for endogeneity, there exist two opposing results with respect to the impact of income on health, which is closely related to mortality. Meer et al. (2003) find only a weak impact of income on health, whereas Lindahl (2005) does find a significantly positive impact. Even in the study by von Gaudecker and Scholz (2007), it is debatable whether the relationship between income and mortality is really monotonously increasing over the whole support of the income distribution.

with regard to heterogeneous social security outcomes by different subgroups of the population in the United States has been made by Liu and Rettenmaier (2003), who find that the shorter life expectancy of blacks yields a lower return from social security even despite progressive elements in the U.S. social security, which are not in place in the German retirement benefit formula.

Theoretical literature on duration as a function of income being the driving force behind heterogeneous implicit taxation in pension systems is mainly based on models of optimal taxation. Diamond (2003, pp. 87) derives an optimal pension system (or age–dependent tax system) that takes exogenous differences in life expectancy into account, and concludes that different implicit tax rates (or, in other words, different rates of return[2]) are actually second best optimal. Breyer and Kifmann (2002) analyze the effect of the adjustment of discounts for early retirees on the implicit tax rate—however, heterogeneous life expectancy does not matter in their analysis. Its separate effect on the implicit tax of a pension system is analyzed in Breyer and Kifmann (2004). Fenge et al. (2006) derive a factual and an optimal life–cycle profile of the implicit tax imposed by a social security system and find that actual as well as optimal implicit taxation differs even despite absent differences in duration under the benefit spell. Butrica et al. (2006) propose an exact measure of the size of the implicit tax exerted on prolonged work and find that the implicit tax in the United States is indeed smaller for individuals with a life expectancy above average, yet, their measure is not directly transferable to Germany, as general income tax and health insurance are also taken into account.

Standard theory on the effect of income on longevity is mainly contained in Grossman (2000), who summarizes his own earlier work on health capital as well as contributions and extensions by subsequent authors. Though life expectancy is in general not a direct choice variable in these models, death is usually assumed to occur once health capital depreciates below a certain threshold and is therefore (positively) affected by income or initial wealth. There exists a comparably vast strand of theoretical literature with regard of the retirement decision, effectively beginning with Feldstein (1974), Diamond and Mirrlees (1978), and Sheshinski (1978), who estimate and analyze the effect of a pension system on savings and, even more important, the retirement decision. A recent theoretical contribution by Bloom et al. (2005) indicates that initial wages are negatively related to retirement age, but the growth rate of income has a positive impact. The first is explained by the income effect of higher wages on labor supply in general, which is also important for the retirement decision. Although longevity is a main driving variable in their theoretical

[2]Both, internal rate of return and implicit tax measure the same phenomenon, the implicit tax being the difference between the rate of return on the capital market and the internal rate of return of the pension system.

model, an empirical test of this model in Bloom et al. (2006) shows that subjective probabilities of survival have no impact on the retirement decision in the United States.

In order to overcome parametrical structure, I apply fully non–parametric and semi–parametric methods. This is especially useful as neither theory nor earlier empirical research gives a guideline for a general pattern between benefit claims and duration of the benefit spell. On questions of redistribution or social security, there have been very few applications of non–parametric methods: I am not aware of any non–parametric analysis with respect to redistribution or the duration of the pension benefit spell. Meghir and Whitehouse (1997), Antolin and Scarpetta (1998), and Bratberg et al. (2004) analyze retirement behavior non–parametrically, and Meghir and Whitehouse (1997) find that in the UK, income has a positive impact on delayed retirement. All studies find that their respective measures of social security wealth (hence, the financial incentives the pension system provides) actually influences the retirement decision.

On the non–parametric estimation of mortality, see e.g. Denton et al. (2005), however, their analysis—used to forecast mortality—does not discriminate between different income groups. Portrait et al. (2001) and Fan et al. (2004) analyze the effects of mortality and health status or health care, respectively. While Portrait et al. (2001) estimate remaining life expectancies conditional on certain health incidences, Fan et al. (2004) estimate the relationship between income and health care expenditures, which they find to be highly non–linear. Although the mentioned approaches are all non–parametric, there are still fundamental differences to the methodology which I apply here (local regression), as e.g. Antolin and Scarpetta (1998) or Felipe et al. (2001) estimate a hazard–model, and Portrait et al. (2001) apply an index model.

In this paper, I find that the answer to my key conjecture of income–induced differences in duration is sensitive to the sub–group under consideration and to the question whether the impact of benefit claims on duration should be controlled for other influences in order to infer on redistribution. I propose to restrict the relevant group to individuals with at least 25 years of contribution in order to guarantee that benefit claims are a good proxy for life time income. Applying non–parametric and semi–parametric methods, I find evidence for duration to be increasing in benefit claims. The shape and level of the analyzed relationship depends additionally on the health status of the individual, which is measured by the time spent in ill–health or rehabilitation. In general, people in ill–health benefit from a longer benefit spell, but the positive impact of income is stronger for people in worse health. The most adequate specification accounts for endogeneity and produces a clearly positive impact of benefit claims on duration. So, on average there is evidence for redistribution from poor to rich in the pension system. A comparison of these results to standard (W)LS and 2SLS regressions indicates that the income gradient is always significant, although its predictive power is low.

The remaining paper is organized as follows. In section 3.2, I introduce the data set. Section 3.3 gives a short overview of the descriptive statistics, whereas section 3.4 introduces the econometric method, which is mainly based on locally linear regression. Section 3.5 discusses the results, and in Section 3.6 I conclude.

3.2 The Variables

The variables used in this analysis stem from a data set with pension discontinuations from 1994 to 2005, published by the Federation of German Pension Insurance Institutes (Deutsche Rentenversicherung Bund), see FDZ-RV (2007). It contains a 10% sample of all discontinued public pensions due to the death of the beneficiary, which amounts to roughly 828,000 observations. However, each observation corresponds to a pension, and not to an individual retiree, who can benefit from more than one pension. This is the case for individuals who receive a pension due to the death of a spouse before they are eligible for their own pension, or individuals who receive a disability pension which is transformed to an old–age pension. Correcting this, I am left with 209,751 observations.[3] The most important variables are the sum of pension benefit claims, the length of the work life, and the duration under the benefit spell. All used variables are described in detail below:

Benefit claims: The sum of pension benefit claims, measured in points. One point corresponds to one year of contributions based on the average income of those who contribute to the social insurance schemes.

Years of contribution: The number of years in which own contributions have been paid.

Duration: Constructed by subtracting the year of entry into the first pension from the year of death. The first pension may already be the old–age pension, but it also encompasses all kinds of disability pensions. For redistributional issues, the total length of the benefit spell (and therefore my notion of duration) is the relevant measure.

Sex: Sex of the beneficiary, though the major part of this paper is only concerned with male pensioners.

[3] Additionally, I exclude individuals where I cannot observe certain important characteristics, such as the date of birth, nationality, residence, retirement age, benefit claims, or the years of contribution.

Residence: The place of residence, aggregated to West Germany and East Germany (Berlin adds to West Germany).

Months in unemployment: Time spent in unemployment, in months.

Months in ill–health: Time spent in ill health (as long as this time is relevant for the pension system), in months. This also covers time spent in rehabilitation treatments.

Type of pension: Either old–age (the standard case) or pension paid due to a reduction of the earnings capacity (for readability, from here on denoted as *disability* pension, although from an administrative point of view, this term is not correct). At the legal retirement age, disability pensions are transformed into old–age pensions.

I have to take all other variables which are contained in the data set with care. They are only reliable when they have been used for the calculation of benefits, otherwise they are unreliable or missing, which is e.g. the case for marital status. Other variables are not reliable for a similar reason: The number of children is actually important for the benefit calculation, as once the individual has interrupted his working life, some benefit claims can be earned by the upbringing of children as a counterbalance to the lost years of work. However, only one of the parents can make use of this feature, such that this variable does not measure the 'physical' number of children.

In most local regressions, I do not include the whole set of observations. Though some results are reported for the whole sample as well, this is not necessarily the relevant one. Since income per year or working life (or the earnings capacity) cannot be observed directly, it has to be ensured that the benefit claims are a good substitute. In the simplest case, namely when an individual has worked his whole career and contributed to the public pension system, benefit claims are a linear transformation of income.[4] This still holds even if the individual under observations had suffered from periods of unemployment and (longer lasting) illness or if he or she had raised children. The measure is then slightly diluted, but still a good proxy for income. The close relationship between total income and benefit claims however is not guaranteed once the individual had been self–employed or had worked as a public servant for some time in his career. During these times, usually no contributions are paid, as membership in the public pension system is not mandatory (or even possible) anymore.

[4]Up to a certain income, beyond which contributions (and therefore claims) are capped. The maximum contributions are based (in 2008) on a monthly gross income of EUR 5300 and are adjusted on a yearly basis.

From a methodological point of view it has to be stressed that the basis of my analysis are death cohorts, and not birth cohorts, because the set of the latter is not self–contained and might lead to a bias, as there are only very few birth cohorts of which all members have died and of which total duration under the benefit spell has therefore been realized already. Despite an existing measure for remaining life expectancy (and therefore for expected duration) conditional on age, this measure is not applicable, because it only captures *average* remaining life expectancy, disregarding the differences among income groups.

3.3 Weighting Scheme and Descriptive Statistics

The sample I use suffers from a selection bias. Since I observe a death cohort (though a rather large one), life expectancies are biased downwards, and therefore duration of the benefit spell may also be biased. Life expectancy has been increasing with the year of birth.[5] The present sample only partially accounts for this increase, because especially individuals from younger birth cohorts (whose ex ante life expectancy should be higher) only appear in the sample if they died relatively young.

The approach to correcting the selection bias is the following. As the relationship between increased life expectancy and year of birth is empirically linear,[6] a linear weighting function, which decreases with the year of birth, corrects the bias. The choice parameter is the slope of the weighting function, while the intercept serves as a normalizing constant that limits the range of the potential slopes in order to ensure the non–negativity constraint. If b denotes the year of birth (normalized to zero for the earliest birth cohort), the weighting function ω has the following form, with s being the slope parameter:

$$\omega(b) \quad = \quad 1 - s \cdot b \tag{3.1}$$

With the intercept set to one, s can vary between zero (hence, a weight of one for all birth cohorts) and .0103, which just ensures that the weight for the latest birth cohort is still positive. I select the weighting function which minimizes the difference between the weighted average life expectancy in my data and the exogenously known life expectancy, which is the case once I set $s = .0103$.

Standard descriptive statistics for the variables of interest (weighted as well as

[5] See Chapter 4 for a detailed treatment on the problem of the present selection bias and its solution.

[6] See e.g. Statistisches Bundesamt (2007, p. 54) and Human Mortality Database (2005), own calculations.

unweighted) in each data set can be found in Tables 3.1 and 3.2. If the sample is restricted to individuals with at least 25 years of contributions, I expect the average amount of total benefit claims to increase, which it does. But more interesting, the average claims–per–year measure is only slightly affected by the restriction, which indicates that I can attribute the major fraction of the increase in total claims to the sheer number of additional work years alone.

Realized duration, however, differs from what is officially reported, which requires an explanation. My sub–sample yields (at maximum) an average of 8.82 years for total duration, whereas the average duration for *all* who died in the respective years was higher, namely 17.57 years. This is due to the same selection effect I introduce above, severed by legislation: A major legislative change in 1992 (essentially the introduction of early retirement discounts) affects the calculation of pension benefit claims at retirement age. Among others, the variable 'years of contribution' was adjusted, such that for any retirement before 1992, benefit claims were based on a different measure of years of contribution, which is not feasible anymore. Public pension administration based the pensions of all individuals who retired after 1992 on the new measure, which is included in the data set. The same holds for months in ill–health or unemployment, which are also recorded contingent on the year of retirement only. Yet, years of contribution are relevant and play a double role in this analysis. First, they serve as additional control, but second and more important, they are necessary to compute average claims per year, which in turn are necessary to compute an instrument for the potentially endogenous variable 'benefit claims'.

The selection effect works as follows: With the exception of those who died receiving a disability pension, the individuals I include in most specifications retired into old–age pensions 1992 or later. With deaths that occurred between 1994 and 2005, this restriction imposes an upper bound on realized life expectancy, which cannot be fully overcome by the weighting scheme I propose. Based on the otherwise excellent administrative data, a complete analysis with all types of pensions, but without weighting will have to be postponed at least two decades. Despite the bias, the findings of the analysis remain valid: In a locally linear regression of duration on total benefit claims which does not utilize years of contribution in one way or the other (hence, benefit claims are not instrumented), I include the complete sample. This yields to results which are—compared to the restricted samples—only different in *level*, but not in the shape of the relationship between duration and benefit claims. Furthermore, the sub–group I can observe is already decisive for major channels of redistribution which are relevant for all other individuals as well, as e.g. the interaction between benefit claims and the type of pension has its natural focus on those who died early, especially before the transformation of the disability pension into an old–age pension.

3.4 Econometric Method

3.4.1 Locally Linear Regression

In this analysis, I impose no restriction on the shape, and hence no parametric form of the relationship. Instead, I estimate a fully non-parametric function:

$$y_i = f(x_i) + \epsilon_i \tag{3.2}$$

I use locally linear regression as the method to fit y on x.[7] At each pair of observations (x_i, y_i), I fit a linear relation around this pair, using the neighboring observations with a kernel weight decreasing in the distance to x_i. This method is proven to be efficient and is especially not biased at the left and right boundaries, where fewer observations are found, see e.g. Fan and Gijbels (2003, pp.60). The weights I apply are based on the Gaussian kernel, such that the whole range of x and y is used for each local regression, however, with differing weights. For my application, the slight efficiency loss of the Gaussian kernel as compared to the Epanechnikov kernel (see Mittelhammer et al. 2000, p.,606) is more than outweighed by its computational advantages.

For the application of the kernel it remains to be determined how fast the weights decrease.[8] I apply the plug–in method proposed by Loader (2004) and choose the bandwidth h to be

$$h = \left(\frac{\sigma^2 (b-a)^2 \int K(v)^2 dv}{n \left(\int v^2 K(v) dv \right)^2 \int m''(x)^2 dx} \right)^{1/5}, \tag{3.3}$$

where σ^2 is the error variance, $m''(x)$ is the second derivative of the estimated function, and a and b are the lower and upper bounds of x. Using a first stage (or pilot) estimate, I obtain an estimate of the error variance by

[7] In principle, this method goes back to Cleveland (1974). On implementation and selection of the smoothing parameter, refer to the more recent Loader (2004).

[8] For a finite kernel, e. g. the triangular or rectangular kernel, this choice corresponds to the choice of the distance around (x_i, y_i) that determines the included observations for each local regression.

$$\widehat{\sigma^2} = \frac{1}{n - 2\nu_1 + \nu_2} \sum_i [y_i - m(x_i)]^2 \,, \tag{3.4}$$

with ν_1 and ν_2 as adjustment for the degrees of freedom (see again Loader 2004 for the computation). If the first stage estimate involves a local quadratic fit instead of a local linear regression, the elements of $m''(x)$ are the respective coefficients which measure the impact of the local quadratic term. It remains open to pick a bandwidth for the pilot estimate; in this analysis, I apply Silverman's Rule–of–Thumb, hence $h_{\text{pilot}} = 1.06\sigma_x n^{-1/5}$, where σ_x is the standard deviation of the regressor.[9]

3.4.2 Multi–Variate Locally Linear Regression

In principle, the regressor x is not restricted to be a scalar. Any regressor matrix of the dimension $n \times k$ can be implemented. In the case of two regressors, the result is a surface of fitted $m(x_{i1}, x_{i2})$ values above the (x_1, x_2) plane, which allows to extract as partial results all conditional moments, namely the functions $m(x_{i1}|x_{i2})$ and vice versa. Though this can (in theory) easily be extended to $k > 2$, two restrictions arise: First, a diagrammatic illustration of $m(x_{i1}, x_{i2}, \cdots, x_{ik})$ is no longer feasible. Secondly, the so-called curse of dimensionality arises, which states that the number of observations has to increase more than proportionally with each additional regressor if the same degree of precision[10] is desired. Generally, none of the $x_j, x \in (1, \ldots, k)$ should be a constant, as the level of $m(x_{ij})$ is determined via the *local* regressions. The bandwidth choice draws on Yang and Tschernig (1999), with some simplifications: I choose $h = 1.06\overline{\sigma_x} n^{-1/(4+d)}$, with $\overline{\sigma_x}$ being the average over the sample standard deviations of the regressors, and d denoting the number of regressors.[11]

[9] Note that Silverman's Rule–of–Thumb and Loader's proposal are related to each other; where Silverman's ROT relies on distributional assumptions with respect to the data, Loader (2004) replaces these assumptions by their sample counterparts (with measures for the variance and the skewness of the data and the kernel). The `GAUSS 7.0` code of this procedure I have written and applied is available upon request. In order to speed up computation, the final estimation of the conditional moment vector $E(y|x)$ is performed on an equally spaced grid of 50 points on the total range of x, whereas for each *local* regression the complete x and y vectors are used. The computation of the plug–in bandwidth, however, requires that the complete set of observations is used in a locally quadratic regression. In terms of CPU time, cross–validation is much slower, and in addition, the determination of the smoothing parameter by plug–in methods is more stable (see Fan 2000 for a short overview), which is corroborated by experiments with smaller sub–samples of the data set on hand, where the cross–validated bandwidth varies by the factor five.

[10] See Yatchew (2003), pp. 17. 'Degree of precision' is inversely defined by the approximation error, which has the order of magnitude of $1/n^{1/k}$.

[11] In the multi–variate case, the final estimation is performed on a grid of 25 equally spaced data points on *each* x vector. The local regression on each grid point, however, uses again the whole sample.

3.4.3 Partially Linear Regression

Accounting for the trade–off between imposed structure and the necessity of additional control variables, I also apply a partially linear model. Let x_1 be the regressor I want to analyze non–parametrically. I then denote all other regressors except x_1 which enter parametrically by x_{-1}. The model has the following form:

$$y_i = f(x_{1i}) + x'_{-1i}\beta + \epsilon_i \tag{3.5}$$

The parameter vector β is unknown (just as the function f). In order to approach an estimation technique, I rewrite the partially linear model in terms of expectations, conditional on x_1:

$$E(y_i|x_{1i}) \quad = \quad f(x_{1i}) + E(x_{-1i}|x_{1i})'\beta \tag{3.6}$$

I estimate these conditional expectations non–parametrically, i. e. by fitting a local polynomial. Denote the estimates by

$$\begin{aligned} E\widehat{(y_i|x_{1i})} &=: m_y(x_{1i}) \\ E(\widehat{x_{-1i}|x}_{1,i}) &=: m_x(x_{1i}). \end{aligned} \tag{3.7}$$

The partially linear model in terms of conditional expectations of Equation 3.6 is then

$$y_i - m_y(x_{1i}) = [x_{-1i} - m_x(x_{1i})]'\beta + \epsilon_i, \tag{3.8}$$

and β can be estimated by least squares. Denoting the estimate $\widehat{\beta}$ and using Equations 3.5, 3.6 and the Definition 3.7, I finally obtain an estimate for $f(x_{1i})$ by

$$\widehat{f}(x_{1i}) = m_y(x_{1i}) - m_x(x_{1i})'\widehat{\beta} \tag{3.9}$$

However, note that the elements of the partially linear model can only be identified under two restrictions,[12] namely

[12]See e. g. Pagan and Ullah (1999, p. 198).

$$E(\epsilon_i | x_{1i}, x_{-1i}) = 0 \tag{3.10}$$

and the absence of a constant in the parametric regressor vector x_{-1i}. The first condition will be violated once y_i and x_{1i} are endogenous variables. The latter is due to the fact that $f(x_{1i})$ is left unspecified, such that any constant term in x_{-1i} can not be distinguished from a shift of $f(x_{1i})$.

3.4.4 Endogeneity

Total benefit claims are potentially endogenous with duration of the benefit spell. The benefit spell begins with the retirement age, which itself affects the amount of collected claims: Delayed retirement directly increases the amount of collected claims due to the contributions paid in this time. In order to account for the endogeneity bias, I use an instrument, namely hypothetical benefit claims normalized to a certain age, which I choose to be 60.[13] This can be constructed by subtracting (adding) the claims which (would) have been earned between actual retirement age and the age of 60:

$$\widetilde{x} \quad = \quad x + \frac{x}{\text{years of contr.}}(60 - \text{ret. age}), \tag{3.11}$$

where x is the sum of claims. Although the method used here is fundamentally non–parametric, the approach is similar to 2–stage least squares. As a first step, I estimate the relationship between the instrument and the original explanatory variable by least squares: Suppose that x_i is endogenous with respect to y_i; however, there exists a variable $\widetilde{x}_i$ which does satisfy the restriction of conditional orthogonality and which is associated with the original x_i by

$$x_i = \widetilde{x}_i \theta + u_i. \tag{3.12}$$

A linear relation between the instrument 'hypothetical claims' and the original regressor 'actual benefit claims' can be based on the following argument: If the average benefit claims per year of contribution does not vary with the years of contribution (which is justified by the descriptive statistics in Tables 3.1 and

[13] At the age of 60, the average pensioner in the sample retires; yet, this choice is arbitrary and the validity of the instrument does not rely on the normalization age. See Angrist and Lavy (1999) for another example and a justification of the validity of the *prediction* of a variable as its instrument.

3.2), the association between x and $\widetilde{x}$ is linear by construction. Under linearity of $E(\epsilon_i|x_i, u_i) = u_i'\rho$, the relationship among the residuals is

$$\epsilon_i = u_i'\rho + \nu_i, \tag{3.13}$$

where ϵ_i is the residual of the model in Equation (3.2). The endogeneity–adjusted model can then be written as

$$y_i = f(x_i) + u_i'\rho + \nu_i. \tag{3.14}$$

Finally, in the second step, I estimate the model partially linearly. As u_i is not directly observable, it has to be replaced by an estimate, namely by the residual of Equation (3.12) estimated by least squares. The function f can be identified by the partially linear model I propose above, with u as additional parametric regressor. On this identification strategy, see Yatchew (2003, pp. 87) for a textbook treatment, Speckman (1988) for the introduction of partially linear models with smoothed conditional moments as statistical method, and Blundell and Duncan (1998) for a seminal application, including the problem of endogenous regressors.

Significance of the IV–residual in the final partially linear regressions falsifies exogeneity of benefit claims as regressor. In the least squares framework, I can apply formal tests for the validity and strength of the instruments I use.[14] First, the instruments used in the regressions are strong. A standard F–test applied to the first stage regression of the original regressor on the instrument (in a 2SLS setup, which also applies for the first stage regression in the partially linear framework) reaches 5380.57, and the instrument itself is highly significant. The requirement of orthogonality, hence of no correlation between the instruments and the residual of the 'wrong' regression of y on the original regressor x is only partially fulfilled; the correlation coefficient between the instrument or the instrument squared and the residual is (rounded) .0955 and .0151, respectively. These values are very small, but still significantly different from zero.

A related, but different approach to account for the endogeneity bias is the replacement of total benefit claims as explanatory variable in the first place. Using average benefit claims instead, there is no reason anymore for duration (or indirectly, retirement age) to reversely cause the regressor.

[14] In the partially linear regressions, I need only one instrument, namely hypothetical benefit claims $\widetilde{x}$. In the least squares regression framework, however, I fit a polynomial of degree two in the explanatory variable, so I need (at least) two instruments. I use predicted benefit claims squared as second instrument.

3.4.5 Bootstrapped Confidence Interval

I bootstrap confidence bands around the semi–parametrically estimated function $f(x_i)$. The procedure I apply borrows from Yatchew (2003, p.161). First, I produce over–smoothed and under–smoothed estimates $\underline{f}(x_i)$ and $\overline{f}(x_i)$, using $0.9h$ and $1.1h$. Based on the under–smoothed estimate, I calculate the residuals $\widehat{\epsilon_i} = y_i - \underline{f}(x_i)$. Since I cannot rule out heteroscedastic errors, I transform the error vector and apply a so–called wild bootstrap (see Yatchew 2003, pp. 156 or Li and Racine 2007, pp. 289, 308). The transformed errors are then:

$$\widehat{\epsilon_i}^{\text{wild}} = \begin{cases} \widehat{\epsilon_i}\frac{1-\sqrt{5}}{2} & \text{with} \quad \Pr = \frac{5+\sqrt{5}}{10} \\ \widehat{\epsilon_i}\frac{1+\sqrt{5}}{2} & \text{with} \quad \Pr = \frac{5-\sqrt{5}}{10} \end{cases} \tag{3.15}$$

From all $\widehat{\epsilon_i}^{\text{wild}}$, I draw new errors ϵ_i^B with replacement, and the bootstrap–sample I construct by

$$y_i^B = \overline{f}(x_i) + \epsilon_i^B. \tag{3.16}$$

Based on the bootstrap–sample and the original bandwidth h, I estimate a new semi–parametric function $f_B(x_i)$. I repeat the drawing of ϵ_i^B and the subsequent estimation of $f_B(x_i)$ several times, such that the α–confidence band is finally given by the $\alpha/2$ and the $1 - \alpha/2$ quantile of the empirical distribution of all $f_B(x_i)$.

3.4.6 Approximate Confidence Interval

Additionally, I approximate a pointwise confidence interval around the estimate of $f(x_i)$ using conditional standard errors $\sigma(x_i)$ at each grid point of x. The confidence bounds are given by (see Härdle et al. 2004, pp.119)

$$f_{\text{CB}}(x_i) = \widehat{f}(x_i) \pm z_\alpha \sqrt{\frac{\int_v K^2(v)dv\widehat{\sigma}^2(x_i)}{nh\widehat{p}(x_i)}}, \tag{3.17}$$

where z_α is 2.58, given the number of observations and the desired confidence level of 99%. The estimated conditional (or local) standard deviation is

$$\widehat{\sigma}^2(x_i) = \frac{1}{n}\sum_i w(x_i)\left[y_i - \widehat{f}(x_i)\right]^2, \tag{3.18}$$

the density $\widehat{p}(x_i)$ of x_i is a non–parametric estimate applying the Gaussian kernel, and $\int_v K^2(v)dv$ is 4.37335 in the case of the normal kernel.

3.4.7 Significance Test

Härdle et al. (2004, pp. 124) propose a fully non–parametric significance test for the impact of x on y. The null hypothesis is that x does not have an impact on y, hence that $f(x_i) = \overline{y}$, where $\overline{y}$ denotes the sample average of y. The test statistic is

$$\widetilde{T} = \sqrt{h} \sum_i \left[\widehat{f}(x_i) - \overline{y} \right]^2 - \frac{\widehat{\sigma}^2(x)}{\sqrt{h}} \left(\int_v K^2(v)dv \right)^2 . \tag{3.19}$$

For $n \to \infty$, $\widetilde{T}$ converges in distribution to $N(0, S)$. As compared to the original proposal by Härdle et al. (2004), the notation I use is simpler, because I apply no additional weights, such that $S = 2\sigma_x^4 \left(\int_v K^2(v)dv \right)^2$. The null hypothesis has to be rejected for values of $\widetilde{T}/\sqrt{S}$ being larger than a respective α–quantile of the normal distribution.

3.4.8 Least Squares Regression

In order to quantify the impact of the independent variables I apply a standard least squares approach as well. The main independent variable 'total claims' or 'claims per year' x enter as a polynomial of degree two, the remaining set of covariates (including a constant) form the matrix z, such that the regression equation is

$$y_i \quad = \quad \beta_1 x_i + \beta_2 x_i^2 + z_i \gamma + \epsilon_i. \tag{3.20}$$

In the case of total claims as independent variable, I use two versions of the matrix z, first the one with original regressors, and second, one with fitted regressors $\hat{z}$ from a first stage regression of z on the instrument for benefit claims, $\widetilde{x}$. As I need at least as many instruments as endogenous variables, I compute the squared values of my instrument as well. In general, I compare two specifications, ordinary least squares, and weighted least squares. The latter imposes weights as constructed in Equation (3.1) to account for over–sampling of early deaths.

3.5 Results and Policy

3.5.1 General Remarks and Differential Results by Sex

I apply the following basic estimation strategies and report their results:[15]

(a) Uni–variate locally linear estimation, stratified along certain control variables, with benefit claims—either without (a1) or with instrument (a2)—and average claims as explanatory variable.

(b) Partially linear estimation, with benefit claims (instrumented) and average claims as explanatory variable, and with months in ill–health, unemployment, and years of contribution as additional parametric controls. See Table 3.3 for the coefficients of the parametric regressors.

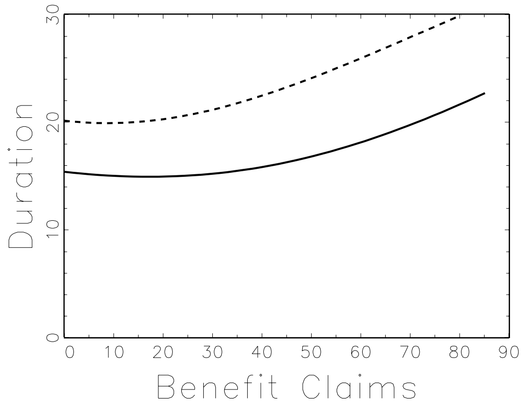
(c) Multi–variate locally linear estimation, with benefit claims (either instrumented or not) and average claims as explanatory variable, and with months in ill–health, unemployment, and years of contribution as additional controls. The results are expectations of the dependent variable 'duration' with respect to benefit claims, contingent on the outcome of additional regressors—which I hold constant at their sample averages, if not indicated otherwise.

(d) Ordinary and weighted least squares with different sets of covariates, see Tables 3.4 and 3.5 for the results.

The specification I estimate on the whole data set is locally linear regression (a) of duration on total benefit claims, see Figure 3.1. Both, men and women, show an upward sloping pattern, hence an indication of redistribution from poor to rich. The pattern is more pronounced for men as for women: The difference between the highest and lowest income groups for men amounts to 7.8 years, while the difference for women is 11.0 years. Regressions on the restricted data set yield similar results (see Figure 3.2), however, on a different level. The difference between the lowest income group and the high income group at $x = 70$ is 6.9 years for men, and 12.1 years for women. Whether total benefit claims are instrumented or not (a1, a2), the general pattern stays the same and provides evidence for major differences in the duration of the pension benefit spells. The results are non–linear and justify the non–parametric approach.

If additional variables enter the regressions as controls, the shape changes; in the partially linear specification (b), differences between income groups almost vanish, whereas the multi–variate locally linear specification (c) even produces a

[15]For the size of the respective sub–samples and optimal bandwidths, refer to Table 3.6.

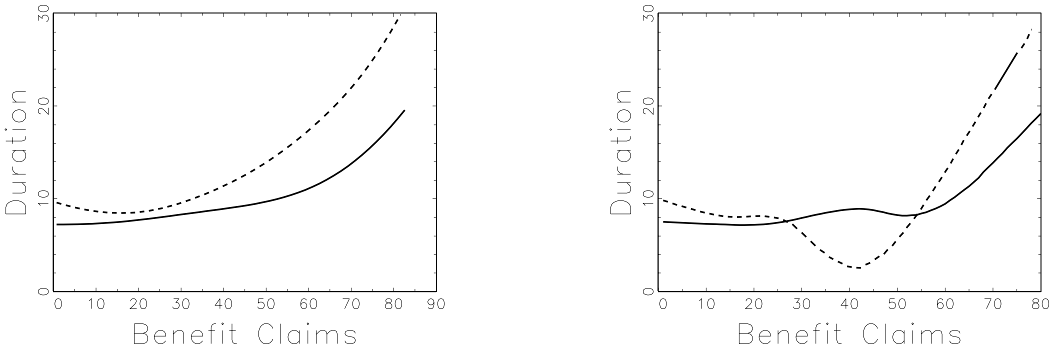


Solid: male, dashed: female. Complete data set without restrictions on years of contribution. Explanatory variable: total benefit claims.

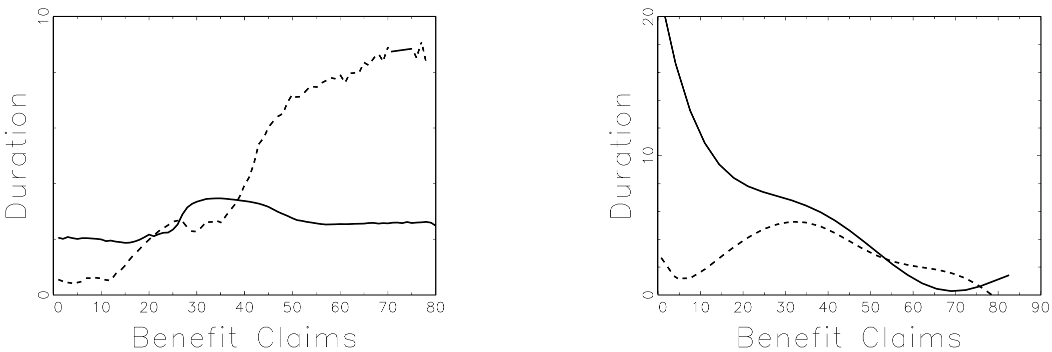
Figure 3.1: Results by Sex, Complete Data

negative relationship between duration and benefit claims for men. The replacement of total benefit claims with average benefit claims in order to avoid the need for an instrument does not substantially alter the results (see Figure 3.3); without additional regressors (a), the relationship is unambiguously upward–sloping, while the partially linear regression (b) reduces the difference between highest and lowest income groups. In the multi–variate specification (c), the shape of the relationship between benefit claims and duration is slightly negative.

As a robustness check, I bootstrap and approximate confidence intervals around the estimate with men only, based on total benefit claims as regressor, instrumented with hypothetical claims. The result corroborates the non–linearly increasing relationship, see Figure 3.4. Following Härdle et al. (2004, p. 124), using confidence bands to infer on significance of the regressor is usually inefficient and too conservative, nevertheless, the confidence band is upward–sloping. Even more, both the bootstrapped and approximate confidence bands are upward–sloping in the sense that the upper confidence bound at x_{min} is smaller than the lower confidence bound at x_{max}. The significance test I introduce in Section 3.4.7 (on the same specification) clearly rejects the null of $f(x_i) = \overline{y}$, such that I infer on a significant impact of benefit claims x on y. The test statistic $T/\sqrt{S}$ is 3.96×10^8, which is larger than any reasonable quantile of the normal distribution. The reason for longer duration for the rich is twofold; wealthier individuals live longer and retire earlier. The first argument is—on average—corroborated in Chapter 4. For the second argument, refer to Berkel and Boersch-Supan (2004). Individuals with higher benefit claims had also better possibilities to accumulate wealth other than social security claims and can therefore afford to retire early and bear the retirement discounts.



Solid: male, dashed: female. Left panel (a1): not instrumented, right panel (a2): instrumented. Explanatory variable: total benefit claims.



Solid: male, dashed: female. Left panel (b): partially linear with added controls, right panel (c): multi–variate. Explanatory variable: total benefit claims.

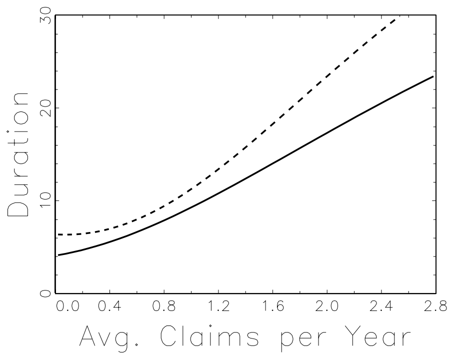
Figure 3.2: Results by Sex I

3.5.2 Differential Results by Health Status

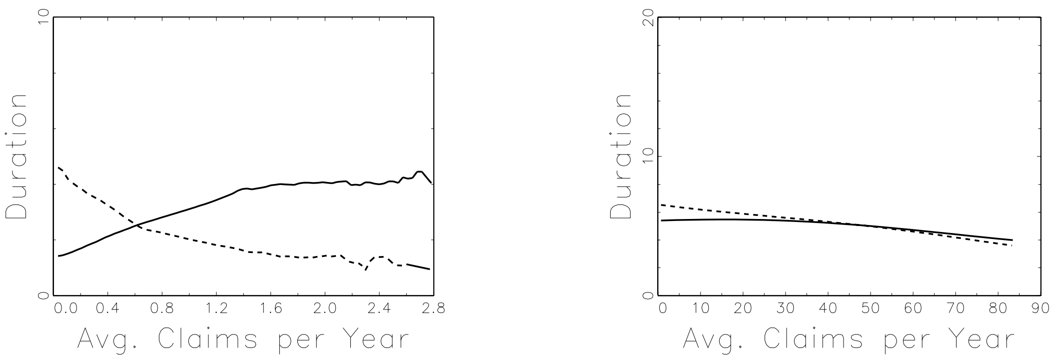
When I analyze the relationship of duration and total benefit claims with special emphasis on different health groups, the results of some specifications are inconclusive (see Figure 3.5). In specification without additional controls (a1, a2) I stratify the population into three groups, with either no months in ill–health at all, with less than six months, or with six months and more.[16] The fundamental relationship between benefit claims and duration is not affected by stratification.

This is corroborated for men by the multi–variate specification (c), where I find little differences along the ill–health dimensions. For women, there are slight differences along the ill–health axis, however, these differences do not affect the general pattern between duration and benefit claims (which is negative).

[16]The distribution of months in ill–health is skewed to the right and the majority of observations had not to suffer from spells of ill–health at all. Although the spells are capped at no less than 48 months, stratification of the sample in the area to the right of six months would yield very small sub–groups.



(a) Solid: male, dashed: female. Explanatory variable: average benefit claims.

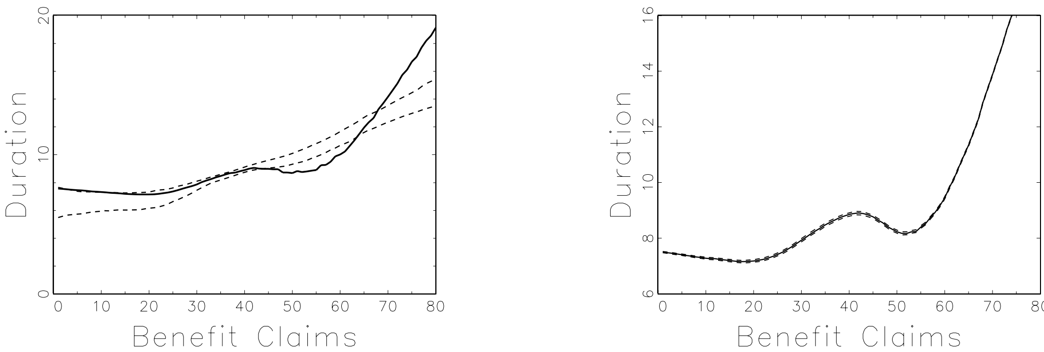


Solid: male, dashed: female. Left panel (b): Partially linear with added controls, right panel (c): multi–variate. Explanatory variable: average benefit claims.

Figure 3.3: Results by Sex II

With average claims as explanatory variable (see Figure 3.6), I identify a subtle, but important impact of health: In the partially linear specification (a), the benefit claims–gradient is steeper, the worse the health status of the individual is, and at the same time, average duration is higher. The latter is confirmed by multi–variate regressions (c), especially for men. Controlling for health does not alter the general shape of the benefit claims—duration relationship, but it raises the conditional expectations of duration to higher levels. For women, even the general shape changes. While women in good health share the common pattern of a negative relationship (once additional controls enter the model), the impact of benefit claims on duration for women in worse health is positive. This finding is of course compatible with the steeper health gradient for worse health I find for men.

The level–effect of health is also confirmed by the partially linear regressions (see Table 3.3), where the sign of the ill–health coefficient is significantly positive. This also holds for almost all specifications of the least squares regressions (see Tables 3.4 and 3.5). I find the coefficient of ill–health only once insignificantly different



Dashed: confidence band. Left panel: Bootstrapped confidence band on the 95% level, based on 100 repetitions and a 5% sub–sample (5,691 observations). Right panel: Approximate confidence band on the 99% level. Men only, explanatory variable: total benefit claims, instrumented (a2).

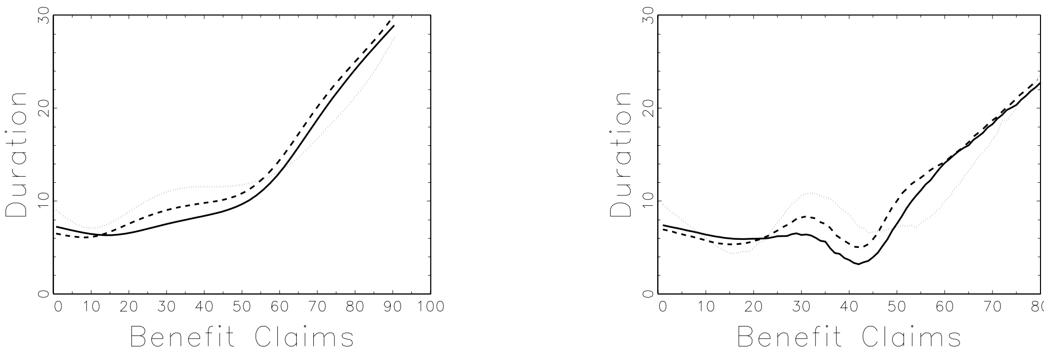
Figure 3.4: Confidence Bands Around $\widehat{f}(x_i)$

from zero—in the full data set, weighted least squares IV–specification. So there is evidence for the fact that individuals in worse health, but whose health status is rather transitory (as opposed to individuals who receive disability pensions) suffer from a stronger impact of socio–economic status on their duration. If you are in ill–health, it really pays to be rich, at least in terms of the benefits the pension system provides. This can e.g. be due to a higher willingness–to–pay for medical treatments, or due to an argument I cannot empirically corroborate: Individuals with higher income are better educated and *financially literate*, such that they can utilize early retirement schemes due to occupational disabilities better than uninformed individuals with poor financial literacy.[17]

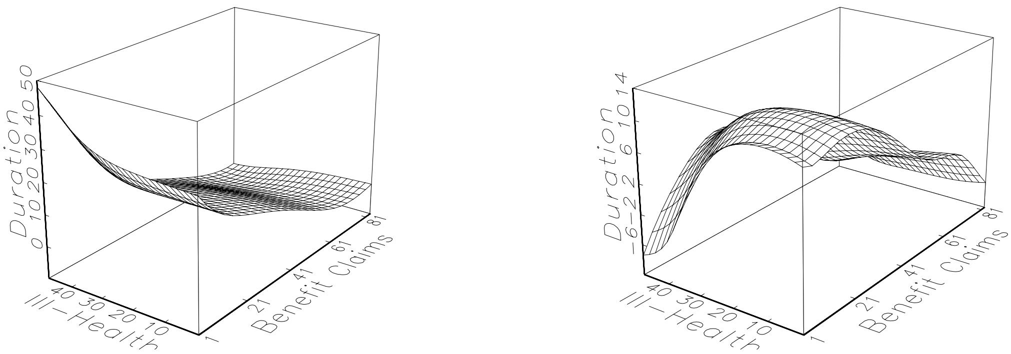
3.5.3 Differential Results by Other Variables

Type of Pension The impact of the type of pension on the analyzed relationship is interesting for two reasons: First, pensioners who died receiving an old–age pension benefit from a longer duration, which is certainly due to their better health status. The alternative—receiving a pension due to a reduction of the earnings capacity or disability pension—is an (almost) ultimate means of helping those who are unable to work, beyond any temporary rehabilitation. In addition, at the legal retirement age, all disability pensions are transformed into old–age pensions, such that there is legal restriction of the age at death for those who receive a disability pension and die during this time. (See Figure 3.7, the results are confirmed by the signs of the parametric identification).

[17]See Lusardi and Mitchell (2007) on the impact of financial literacy on retirement–related behavior. The authors can show that education determines financial literacy, especially the necessary knowledge for retirement planning.



Solid: mts. = 0, dashed: mts. < 6, dotted: mts. ≥ 6. Left panel (a1): not instrumented, right panel: instrumented (a2). Men only, explanatory variable: total benefit claims.

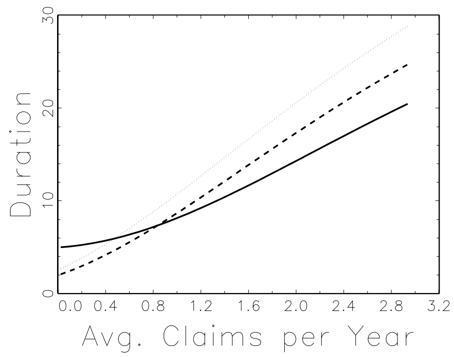


(c) Multi–variate, Left panel: men, right panel: women. Explanatory variable: total benefit claims.

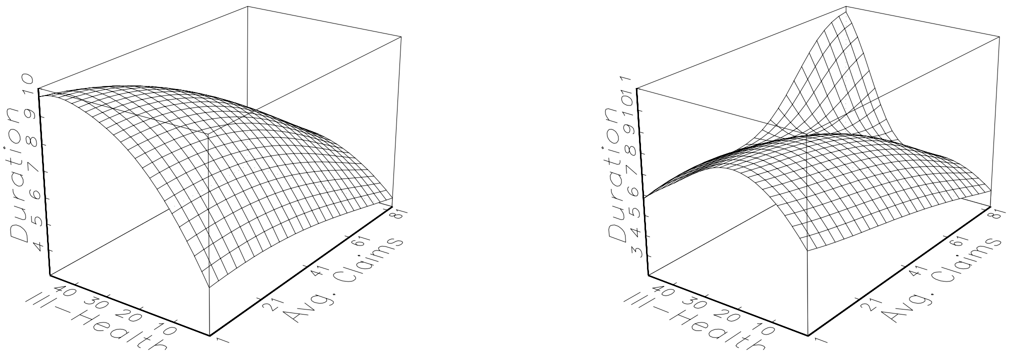
Figure 3.5: Results by Health–Status I

Second, the impact of benefit claims differs by the kind of pension. While I find for those with old–age pensions the already familiar positive impact on duration, there is virtually no influence of benefit claims on people with disability pensions. This corresponds to a recent finding by Kiuila and Mieszkowski (2007), who find that life expectancy is—under some circumstances— not related to income at all, but ultimately determined by independent health factors. A similar argument I find in Adams et al. (2003), who claim that the impact of income on health depend on the type of illness. These arguments carry over to the analysis of the duration of the benefit spell: Conditional on being in very bad health, benefit claims lose their impact on duration, as opposed to my findings with respect to transitory spells of ill–health.

Unemployment I apply a similar stratification to the months spent in unemployment (hence, the individual had not been unemployed at all, less than six months, or six months and more, see Figure 3.8), and the general shape of the benefit claims–duration relationship is not affected, no matter if total benefit claims (instrumented



(a) Solid: mts. = 0, dashed: mts. < 6, dotted: mts. ≥ 6. Men only, explanatory variable: average benefit claims.



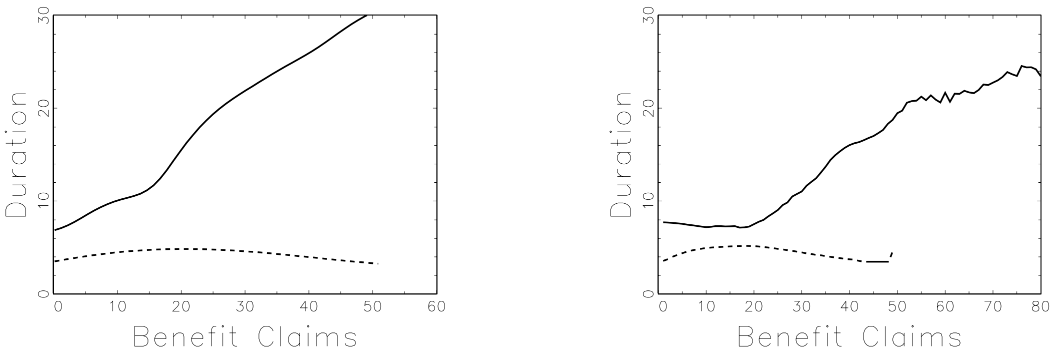
(c) Multi–variate, left panel: men, right panel: women. Explanatory variable: average benefit claims.

Figure 3.6: Results by Health–Status II

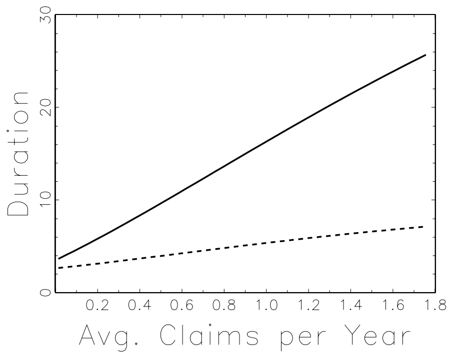
or not) or average benefit claims serve as explanatory variable. Differences in between different unemployment groups are virtually not present.

Still, the coefficients of the unemployment variable (see Tables 3.3, 3.4, and 3.5) are significantly negative in most specifications, hence, pensioners who had to suffer from unemployment spells have shorter pension benefit spells. This result is not universal, as one partially linear specification produces an insignificant coefficient, and least squares regressions yield once an insignificant and once even a significantly positive sign. In all cases, the absolute value is small, especially compared to the ill–health variable, the largest coefficient indicates that each month in unemployment reduces duration of the benefit spell approximately 14 days. Shorter duration of individuals in unemployment is compatible with the negative effect of unemployment on life expectancy I find in Chapter 4.

Residence There is little to no difference in the impact of benefit claims on duration if the sample is divided in West and East Germany. If I use total benefit claims, it is apparent that at the right end of the benefit claims distribution, the gradient



Solid: old–age pension, dashed: disability pension. Left panel: not instrumented, right panel: instrumented (partially linear). Men only, explanatory variable: total benefit claims.

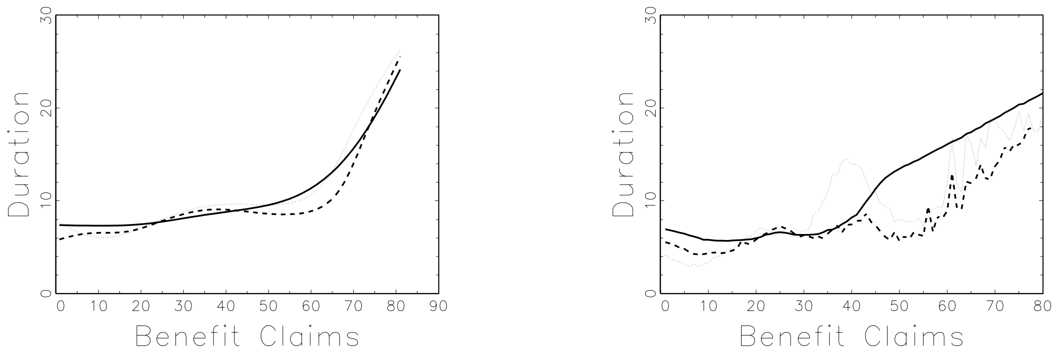


Solid: old–age pension, dashed: disability pension. Men only, explanatory variable: average claims per year.

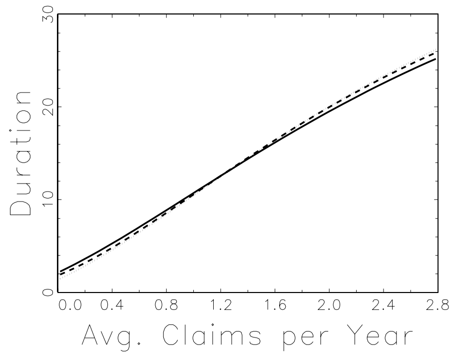
Figure 3.7: Results by Type of Pension

for East Germany is very steep. Notice, however, that there are still more than 6% of all observations in the area with more than 60 points of total benefit claims, such that the steep increase in this area is reliable, and not an artifact of outliers, see Figure 3.9. The sign of the coefficients of the East–dummy (see Tables 3.4 and 3.5) are ambiguous; depending on the exact specification, the sign ranges from significantly negative over insignificant to significantly positive. The most reliably specifications—namely, regressions on the restricted data set, weighted and either IV or average claims as explanatory variable—yield a positive or insignificant sign.

This is certainly due to (1) institutional factors, because large unemployment in East Germany corresponds to the utilization of early retirement schemes wherever possible, and (2) due to the fact that conditional on reaching retirement age, life expectancy in East Germany is not so different from West Germany anymore, and in some instances even higher (see Chapter 4 on life expectancies as a function of benefit claims).



Solid: mts. = 0, dashed: mts. < 6, dotted: mts. ≥ 6. Left panel: not instrumented (a1), right panel: instrumented (a2). Men only, explanatory variable: total benefit claims.



(a) Solid: mts. = 0, dashed: mts. < 6, dotted: mts. ≥ 6. Men only, explanatory variable: average claims per year.

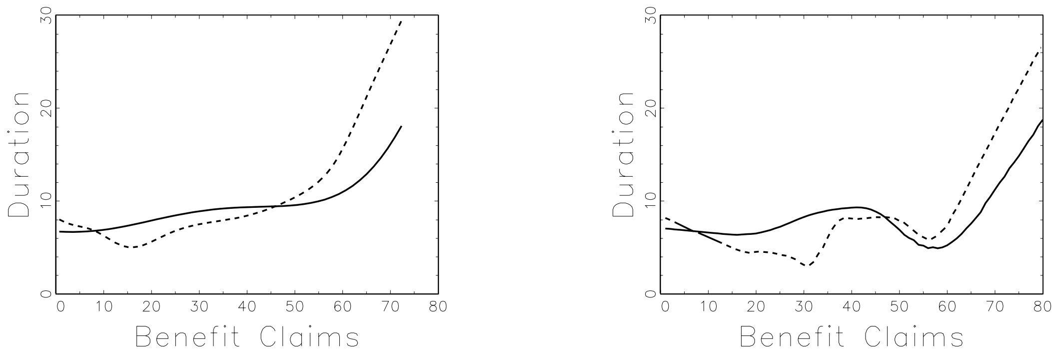
Figure 3.8: Results by Unemployment

Birth Cohorts The weighting scheme I introduce in Equation (3.1) is not necessary, if the whole analysis is stratified along birth cohorts. A specific life expectancy at birth is related to each birth cohort, such that there is no need to account for over or under–sampling of different cohorts.

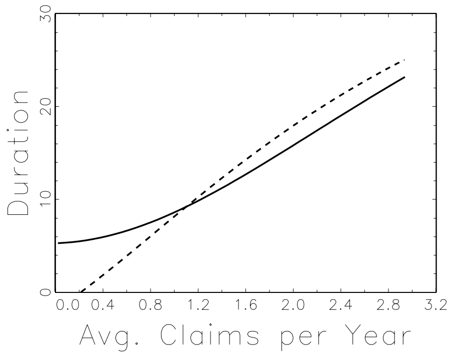
I distinguish cohorts of individuals born between 1920 and 1929 and 1930 to 1939, respectively. The level effect between the both cohorts is a result of the construction and the absence of the weighting scheme *within* each cohort, as the later cohort had to die relatively early in order to be part of the sample, which has a negative effect on duration as well.

In each cohort, I find an increasing relationship between benefit claims and duration, though the impact of benefit claims is less pronounced for the later birth cohort. (See Figure 3.10).

Years of Contribution Finally, the analysis takes years of contribution into account. If years of contributions are considered parametrically (b, d), see Tables 3.3, 3.4, and 3.5, they enter almost always negatively (except of the full, instrumented



Solid: West, dashed: East. Left panel: not instrumented (1), right panel: instrumented (a2). Men only, explanatory variable: total benefit claims.



(a) Solid: West, dashed: East. Men only, explanatory variable: average claims per year.

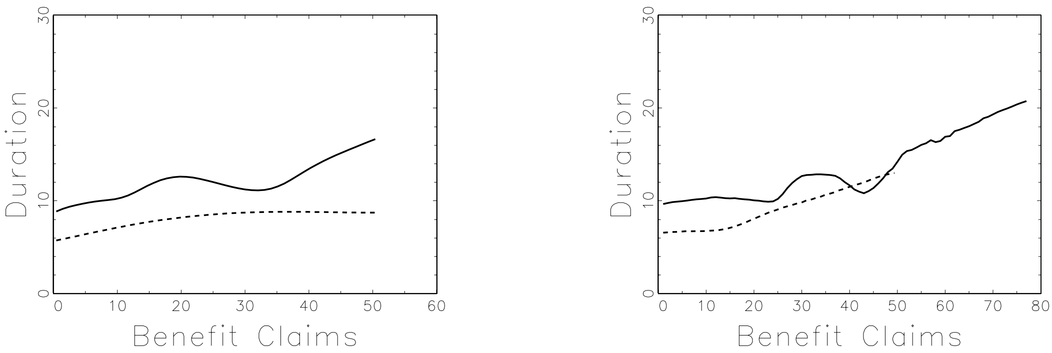
Figure 3.9: Results by Residence

and unweighted specification). This is confirmed by the locally linear specifications (see Figure 3.11), where I find a monotone ordering of the three groups (less than 30 years at the top, between 30 and 40 years in the middle, and more then 40 years at the bottom). The link is easily identified, because if pensioners contributed longer to the pension system, this can at least partially be attributed to delayed retirement, which in turn reduces the duration of the benefit spell.

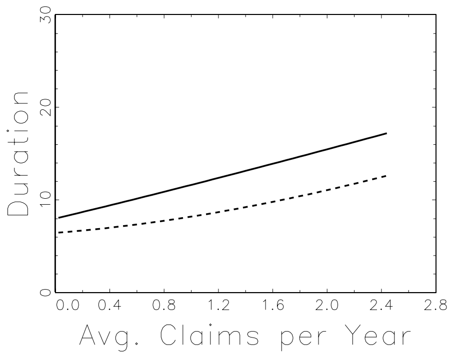
3.5.4 Policy Implications

Although I cannot answer whether the redistribution I find *should* be neutralized within the public pension system, the following entails some means of what *could* be done in order to achieve another distributional outcome.[18] Still, the question

[18]Compare Chapter 5 on objectives of the pension system. If undistorted labor supply is the objective, one could argue that the issue of redistribution of any kind has to be addressed by income taxation; the argument made here would then be in favor of a (ceteris paribus) more progressive income taxation, which, however, then distorts instantaneous labor supply instead of life time labor supply. Additionally, Joaquim Oliveira Martins kindly discussed these findings on the 5th Workshop on Pension and Saving and argued that a higher rate of return from the pension system for



Solid: yr. of birth ∈ (1920, 1930], dashed: yr. of birth ∈ (1930, 1940]. Left panel: not instrumented (a1), right panel: instrumented (a2). Men only, explanatory variable: total benefit claims, data not weighted.



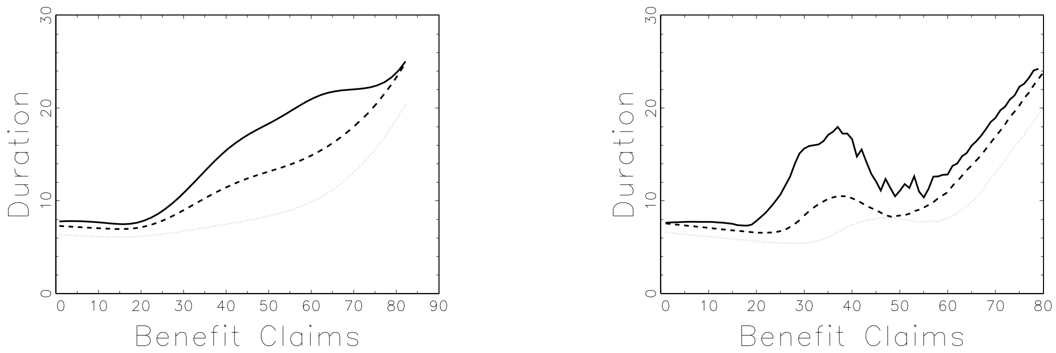
(a) Solid: yr. of birth ∈ (1920, 1930], dashed: yr. of birth ∈ (1930, 1940]. Men only, explanatory variable: average claims per year, data not weighted.

Figure 3.10: Results by Cohort

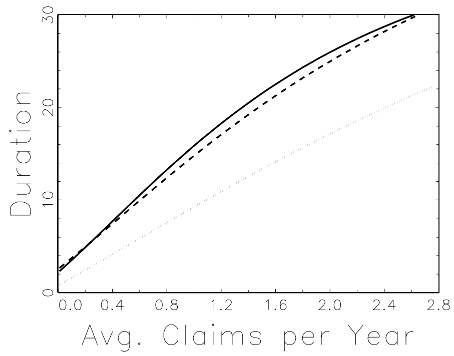
of policy implications of heterogeneous rates of return from the pension system has no easy answer, for the following reason: The shape of the benefit claims–gradient depends on the inclusion of additional control variables. If no controls are added, the relationship is clearly positive. With controls, the positive impact either vanishes or is even transformed into a negative one. From an analytical point of view, this is an indication for the interpretation that higher benefit claims as a measure for income do not *cause* longer duration. However, from a political point of view, this might not be relevant, as long as the pure association of benefit claims with duration prevails. The impact of control variables still serves as guidance for the question, where further policy instruments could approach at.

Benefit claims surely is a variable which is easy to observe and on which redis-

high–income individuals could actually be optimal, once education is taken into account. Education both increases income and life–expectancy (see Grossman 2000 for a theoretical treatment, and Cutler et al. 2006 for an overview of the empirical evidence), such that a higher rate of return due to higher life–expectancy can be understood as a fraction of the education premium, which therefore provides further incentives to educate. See Gorski et al. (2007) and Lau and Poutvaara (2006) for a theoretical treatment of this phenomenon.



Solid: yrs. $<$ 30, dashed: 30 $\leq$ yrs. $<$ 40, dotted: yrs. $\geq$ 40. Left panel: not instrumented (a1), right panel: instrumented (a2). Data non-weighted. Men only, explanatory variable: total benefit claims.



(a) Solid: yrs. $<$ 30, dashed: 30 $\leq$ yrs. $<$ 40, dotted: yrs. $\geq$ 40. Men only, explanatory variable: average claims per year.

Figure 3.11: Results by Years of Contribution

tributive policy can therefore easily condition, in order to imply a certain distributional outcome. Yet, the instrument is not perfect, a phenomenon expressed in the low R^2 of all least squares regressions, even despite the high significance of all benefit claims measures. A major part of the variation in duration has to be explained by variables other than benefit claims.

One instrument within the nexus of benefit claims—which is already applied in reality—are discounts for early retirement. In Germany, each month the retirement age falls short of the age of 65 (for males) reduces the monthly pension benefits about 0.3%. I can use this device and adjust the monthly discounts in a way to approximate a rate of return which is independent of the total benefit claims. An additional degree of freedom—conditioning these discounts on other parameters—would allow to come even closer to the ideal of an income–independent rate of return. Without addressing legal impediments (or even the political process), an obvious dimension of discrimination would be the sex of the beneficiary. Furthermore, discounts may not be constant, but varying with each additional month of early retirement (see Fenge et al. 2006 for a theoretical treatment). The computation of neutral discounts is subject of Chapter 5. Further possibilities of adjusting the benefit formula include a progressive element, meaning that benefit claims increase less than proportionally with income. E.g. in the United States, such an instrument is in place, and it is able to overcome the otherwise regressive effect of longer benefit spells, see Hurd and Shoven (1986).

If the benefit formula is to be left unaltered, incentives and possibilities for early retirement (either into the old–age pension or into any kind of disability pension) for poor or morbid individuals add to an income–neutral rate of return. Such instruments may utilize further results of this analysis, e.g. that people in ill–health slightly profit from the pension system, as their duration is—on average—higher. This can easily be explained by the existence of pensions due to (occupational) disability, which, however, has been restricted in the year 2001 to those born before 1961. Yet, the sub–sample I observe of individuals born afterwards is small and nobody out of this cohort has reached the usual retirement age so far, such that I cannot fully infer the redistributive impact of this legislation. But as eligibility for those born later is subject to stricter conditions of occupational disability (the retirement age for the severely disabled has been increased from 60 to 63),[19] the effect of both policy measures is not fully observable yet, though the conjecture is obvious that people in ill–health will most likely suffer from it in terms of a reduced duration. At the same time, the effect of income on duration and hence redistribution from poor to rich becomes more pronounced for individuals in a poor health status. Analogously, the increased minimum age for a pension after unemployment will most likely reduce the duration and therefore the return from the pension system for the unemployed.

3.6 Conclusion

I address the question whether annuity payments in the German public pension system lead to redistribution from poor to rich. Estimation strategies that account for possible endogeneity of benefit claims yield a clearly positive relationship, and the comparison with standard techniques shows that this impact is also significant, although a major fraction of variation in duration cannot be explained by income alone. I disentangle the influence of income on duration for different sub–groups, especially stratified along the health dimension. In my sample, individuals in ill–health could still benefit from a policy of disability pensions, which has been tightened in the meantime. However, conditional on a certain health status, the positive impact of income on duration is more pronounced, and hence is redistribution.

[19]See e. g. Deutsche Rentenversicherung Bund (2005, pp. 263) for a short chronological overview of changes in the pension system.

Beyond the scope of this paper are other major forces behind differential duration. Differences in retirement age and differences in realized life expectancy (see Chapter 4) both drive duration and have to be analyzed separately, especially with respect to their sensitivity to measures of socio–economic status and health. Furthermore, the revealed relation gives rise to policy interventions to weaken the redistribution. A feasible instrument has to be found, and I propose to adjust the discounts for early retirement, which is subject of Chapter 5: In a joint work with Friedrich Breyer, we calculate discounts which minimize the impact of income on the ratio of benefits to contributions to social security.

3.A Appendix: Tables

Descriptive Statistics, Unweighted

data set:	complete		restricted	
n:	209,751		112,369	
Variable	Mean	St. Dev.	Mean	St. Dev.
duration	8.23	6.70	7.86	5.94
benefit claims	33.47	17.67	44.14	13.23
claims/year	1.10	.73	1.13	.31
contr. years	29.89	13.58	39.03	5.82
ill–health	2.46	5.82	3.04	6.25
unempl.	6.20	17.59	6.3	17.11
old–age pension	.78	—	.77	—
residence west	.73	—	.70	—
residence east	.21	—	.28	—
residence abroad	.06	—	.03	—
female	.32	—	.00	—

Restricted data set include only male observations with at least 25 years of contributions.

Table 3.1: Descriptive Statistics, Unweighted

Descriptive Statistics, Weighted				
data set:	complete		restricted	
n:	209,751		112,369	
Variable	Mean	St. Dev.	Mean	St. Dev.
duration	8.82	6.98	8.33	6.15
benefit claims	33.24	18.20	44.53	13.51
claims/year	1.07	.66	1.13	.31
contr. years	29.73	13.88	39.21	5.74
ill–health	2.47	5.8	3.10	6.33
unempl.	5.90	17.19	6.20	16.93
old–age pension	.83	—	.81	—
residence west	.73	—	.69	—
residence east	.21	—	.28	—
residence abroad	.06	—	.03	—
female	.32	—	.00	—

Restricted data set include only male observations with at least 25 years of contributions.

Table 3.2: Descriptive Statistics, Weighted

Partially Linear Regressions — Dependent Variable: Duration								
specification	male				female			
non–param. regressor	total claims		avg. claims		total claims		avg. claims	
x	β	S.E.	β	S.E.	β	S.E.	β	S.E.
contr. yrs.	-.192***	.005	-.170***	.004	-.232***	.010	-.065***	.008
ill–health	.126***	.003	.135***	.003	-.013*	.009	.141***	.008
unempl.	-.012***	.001	-.016***	.001	-.026***	.002	-.028***	.002
IV–resid.	-.016***	.001	—	—	-.122***	.001	—	—
R^2	.20		.012		.015		.016	
N	112,346		112,346		29,857		29,857	

Non–parametric regressor is either total benefit claims or average benefit claims. *** denotes significance on the .99 level, ** on the .95 level, and * on the .90 level (robust standard errors).

Table 3.3: Coefficients of the Partially Linear Regressions

Least Squares Estimation — Dependent Variable: Duration

data set	full				full, IV			
x	β	S.E.	β	S.E.	β	S.E.	β	S.E.
benefit claims	.315***	.004			-1.45	1.52	-2.14***	.369
benefit claims^2/10	-.029***	.001			.240	.242	.345***	.053
claims/yr.			1.84***	.106				
claims/yr. squ.			-.040***	.008				
contr. yrs.	-.176***	.002	-.049***	.001			.020	.051
ill–health	.149***	.003	.159***	.003			.349***	.031
unempl.	-.017***	.001	-.014***	.001			.072***	.013
residence east	.081**	.035	.292***	.036			-1.00***	.153
female	.601***	.035	-.059	.040			3.45***	.320
old–age pension	6.35***	.027	5.87***	.029			-1.24	1.12
constant	1.63***	.064	2.76***	.135	22.2	16.3	28.58***	3.9
R^2	.20		.18		—		—	
N	209,752		202,181		202,181		202,181	

data set	restricted				restricted, IV			
x	β	S.E.	β	S.E.	β	S.E.	β	S.E.
benefit claims	.190***	.007			-.248***	.011	.097***	.013
benefit claims^2/10	-.010***	.001			.043***	.002	.009***	.002
claims/yr.			1.94***	.288				
claims/yr. squ.			.830***	.129				
contr. yrs.	-.402***	.004	-.277***	.004			-.466***	.004
ill–health	.114***	.003	.118***	.003			.118***	.003
unempl.	-.028***	.001	-.026***	.001			-.024***	.001
residence east	.032	.036	-.003	.036			-.110***	.037
old–age pension	6.44***	.035	6.36***	.034			6.18***	.038
constant	12.18***	.169	10.24***	.187	9.67***	.186	14.86***	.287
R^2	.26		.26		—		.23	
N	112,369		112,369		112,369		112,369	

Restricted data set includes only male observations with at least 25 years of contributions. *** denotes significance on the .99 level, ** on the .95 level, and * on the .90 level (robust standard errors). IV–regressions: benefit claims instrumented with hypothetical claims. Default categories are 'residence west', 'pension due to reduction of earnings capacity', and 'male'.

Table 3.4: Results of the Least Squares Regressions

Weighted Least Squares Estimation — Dependent Variable: Duration

data set	full				full, IV			
x	β	S.E.	β	S.E.	β	S.E.	β	S.E.
benefit claims	.342***	.005			-1.02***	.200	1.31**	.599
benefit claims^2/10	-.030***	.001			.172***	.031	-.138*	.083
claims/yr.			2.60***	.116				
claims/yr. squ.			-.049***	.008				
contr. yrs.	-.197***	.002	-.049***	.001			-.532***	.102
ill–health	.142***	.003	.149***	.003			.051	.054
unempl.	-.018***	.001	-.015***	.001			-.043**	.020
residence east	.307***	.040	.529***	.042			.465	.378
female	.946***	.040	.211***	.045			1.36***	.395
old–age pension	6.43***	.027	5.95***	.028			8.79***	1.69
constant	1.60***	.070	2.02***	.141	17.68***	2.14	-6.767	5.99
R^2	.19		.16		—		—	
N	209,752		202,181		202,181		202,181	

data set	restricted				restricted, IV			
x	β	S.E.	β	S.E.	β	S.E.	β	S.E.
benefit claims	.183***	.008			-.273***	.011	.090***	.014
benefit claims^2/10	-.008***	.001			.047***	.002	.012***	.002
claims/yr.			1.18***	.307				
claims/yr. squ.			1.42***	.136				
contr. yrs.	-.439***	.004	-.293***	.004			-.505***	.005
ill–health	.105***	.003	.109***	.003			.11***	.003
unempl.	-.029***	.001	-.027***	.001			-.025***	.001
residence east	.227***	.040	.167***	.040			-.003	.041
old–age pension	6.54***	.036	6.46***	.035			6.26***	.038
constant	13.55***	.184	11.11***	.201	10.23***	.190	16.30***	.293
R^2	.26		.26		—		.23	
N	112,369		112,369		112,369		112,369	

Restricted data set includes only male observations with at least 25 years of contributions. *** denotes significance on the .99 level, ** on the .95 level, and * on the .90 level (robust standard errors). IV–regressions: benefit claims instrumented with hypothetical claims. Default categories are 'residence west', 'pension due to reduction of earnings capacity', and 'male'.

Table 3.5: Results of the Weighted Least Squares Regressions

Size of Sub–Populations and Optimal Bandwidths

Specification	**Obs.**	h_1	h_2	**Specification**	**Obs.**	h_1	h_2
$m_u(x)$ expected error from	142,203	20.4		by residence			
1st stage, conditional on x				– west	78,519	11.8	477.8
by sex				– east	30,919	5.0	195.3
– male	112,346	15.3	890.1	by unempl.			
– female	29,857	11.4	578.8	– mts. $= 0$	79,091	11.1	410.4
by health				– mts. < 6	12,049	5.6	214.6
– mts. $= 0$	60,036	10.3	398.8	– mts. ≥ 6	21,206	6.9	258.6
– mts. < 6	35,639	8.1	333.9	by contr.			
– mts. ≥ 6	16,671	6.7	277.6	– yrs. < 30	10,705	6.8	185.8
by cohort				– yrs. < 40	39,129	9.9	359.3
– > 1920, ≤ 1930	22,317	5.8	280.2	– yrs. ≥ 40	62,512	12.8	579.9
– > 1930, ≤ 1940	66,120	9.4	529.4	multi–dimensional, male	29,857	2.1	2.4
by pension type				multi–dimensional, female	112,346	2.2	2.3
– old–age	86,615	10.8	432.8	full sample, male	388,754	33.1	
– disability	25,731	12.1	419.7	full sample, female	377,577	42.6	

Bandwidths h_1 with total benefit claims as explanatory variable, h_2 with average claims as explanatory variable. Pilot bandwidths are chosen to be $h = 1.06\sigma_x n^{-1/5}$ (Silverman's Rule–of–Thumb).

Table 3.6: Size of Sub–Populations and Optimal Bandwidths

References of Chapter 3

ADAMS, P., M. D. HURD, D. MCFADDEN, A. MERRILL, AND T. RIBEIRO (2003): "Healthy, Wealthy and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status," *Journal of Econometrics*, 112, 3–56.

ANGRIST, J. AND V. LAVY (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533–575.

ANTOLIN, P. AND S. SCARPETTA (1998): "Microeconometric Analysis of the Retirement Decision: Germany," *OECD Economics Department Working Papers*, No. 204.

BERKEL, B. AND A. BOERSCH-SUPAN (2004): "Pension Reform in Germany: The Impact on Retirement Decisions," *FinanzArchiv*, 60, 393–421.

BLOOM, D. E., D. CANNING, AND M. MOORE (2005): "The Effect of Improvements in Health and Longevity on Optimal Retirement and Saving," *Harvard Initiative of Global Health Working Paper Series*, No. 2.

BLOOM, D. E., D. CANNING, M. MOORE, AND Y. SONG (2006): "The Effect of Subjective Survival Probabilities on Retirement and Wealth in the United States," *NBER Working Paper*, No. 12688.

BLUNDELL, R. AND A. DUNCAN (1998): "Kernel Regression in Empirical Microeconomics," *Journal of Human Resources*, 33, 62–87.

BRATBERG, E., T. H. HOLMAS, AND O. THOGERSEN (2004): "Assessing the Effects of an Early Retirement Program," *Journal of Population Economics*, 17, 387–408.

BREYER, F. AND M. KIFMANN (2002): "Incentives to retire later – a solution to the social security crisis?" *Journal of Pension Economics and Finance*, 1, 111–130.

——— (2004): "The German Retirement Benefit Formula: Drawbacks and Alternatives," *FinanzArchiv*, 60, 63–82.

BUTRICA, B. A., R. W. JOHNSON, K. E. SMITH, AND C. E. STEUERLE (2006): "The Implicit Tax on Work at Older Ages," *National Tax Journal*, LVIV, 211–234.

CLEVELAND, W. S. (1974): "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

CUTLER, D., A. DEATON, AND A. LLERAS-MUNEY (2006): "The Determinants of Mortality," *Journal of Economic Perspectives*, 20, 97–120.

DENTON, F. T., C. H. FEAVER, AND B. G. SPENCER (2005): "Time Series Analysis and Stochastic Forecasting: An Econometric Study of Mortality and Life Expectancy," *Journal of Population Economics*, 18, 203–227.

DEUTSCHE RENTENVERSICHERUNG BUND (2005): "Rentenversicherung in Zeitreihen," *DRV Schriften*, 22.

DIAMOND, P. A. (2003): *Taxation, Incomplete Markets, and Social Security*, Cambridge, MA.: The MIT Press.

DIAMOND, P. A. AND J. A. MIRRLEES (1978): "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics*, 10, 295–336.

FAN, J. (2000): "Prospects of Nonparametric Modeling," *Journal of the American Statistical Association*, 95, 1296–1300.

FAN, J. AND I. GIJBELS (2003): *Local Polynomial Modelling and Its Applications*, Boca Raton: CRC Press.

FAN, Y., D. LI, AND Q. LI (2004): "Nonlinearity in Medical Expenditures: A New Semiparametric Approach," *Applied Economics*, 36, 911–916.

FDZ-RV (2007): "Scientific Use File Demografiedatensatz Rentenwegfall 1993-2005, SUFRTWFjjXVSTDemo," Data set from Deutsche Rentenversicherung Bund.

FELDSTEIN, M. S. (1974): "Social Security, Induced Retirement, and Aggregate Capital Accumulation," *Journal of Political Economy*, 82, 905–926.

FELIPE, A., M. GUILLEN, AND J. P. NIELSEN (2001): "Longevity Studies Based on Kernel Hazard Estimation," *Insurance: Mathematics and Economics*, 28, 191–204.

FENGE, R., S. UEBELMESSER, AND M. WERDING (2006): "On the Optimum Timimg of Implicit Social Secrity Taxes Over the Life Cycle," *FinanzArchiv*, 62, 68–107.

GORSKI, M., T. KRIEGER, AND T. LANGE (2007): "Pensions, Education and Life Expectancy," *Working Paper of the DFG Research Group Heterogeneous Labor*, No. 07/04.

GROSSMAN, M. (2000): "The Human Capital Model," in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, New York and Oxford: Elsevier Science, North-Holland, vol. 1A, 347–408.

HÄRDLE, W., M. MÜLLER, S. SPERLICH, AND A. WERWATZ (2004): *Nonparametric and Semiparametric Models*, Heidelberg: Springer.

HUMAN MORTALITY DATABASE (2005): *Life Expectancy at Birth*, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.humanmortality.de (data downloaded on Jan. 10, 2008).

HURD, M. D. AND J. B. SHOVEN (1986): "The Distributional Impact of Social Security," *NBER Working Paper*, No. 1155.

KIUILA, O. AND P. MIESZKOWSKI (2007): "The Effects of Income, Education and Age on Health," *Health Economics*, 16, 781–798.

LAU, M. I. AND P. POUTVAARA (2006): "Social Security Incentives and Human Capital Investment," *Finnish Economic Papers*, 19, 16–24.

LAUTERBACH, K., M. LÜNGEN, B. STOLLENWERK, A. GERBER, AND G. KLEVER-DEICHERT (2006): "Zum Zusammenhang zwischen Einkommen und Lebenserwartung," *Studien zu Gesundheit, Medizin und Gesellschaft, Köln*, No. 01.

LI, Q. AND J. S. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*, Princeton: Princeton University Press.

LINDAHL, M. (2005): "Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income," *Journal of Human Resources*, 40, 144–168.

LIU, L. AND A. J. RETTENMAIER (2003): "Social Security Outcomes by Racial and Education Groups," *Southern Economic Journal*, 69, 842–864.

LOADER, C. (2004): "Smoothing: Local Regression Techniques," in *Handbook of Computational Statistics*, ed. by J. Gentle, W. Härdle, and Y. Mori, Springer.

LUSARDI, A. AND O. S. MITCHELL (2007): "Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education," *Business Economics*, 42, 35–44.

MEER, J., D. L. MILLER, AND H. S. ROSEN (2003): "Exploring the Health-Wealth Nexus," *Journal of Health Economics*, 22, 713–730.

MEGHIR, C. AND E. WHITEHOUSE (1997): "Labour Market Transitions and Retirement of Men in the UK," *Journal of Econometrics*, 79, 327–354.

MITTELHAMMER, R. C., G. G. JUDGE, AND D. J. MILLER (2000): *Econometric Foundations*, Cambridge: Cambridge University Press.

PAGAN, A. AND A. ULLAH (1999): *Nonparametric Econometrics*, Cambridge: Cambridge University Press.

PORTRAIT, F., M. LINDEBOOM, AND D. DEEG (2001): "Life Expectancies in Specific Health States: Results from a Joint Model of Health Status and Mortality of Older Persons," *Demography*, 38, 525–536.

SCHILS, T. (2005): *Early Retirement Patterns in Europe: A comparative Panel Study*, Amsterdam: Dutch University Press.

SHESHINSKI, E. (1978): "A Model of Social Security and Retirement Decisions," *Journal of Public Economics*, 10, 337–360.

SPECKMAN, P. (1988): "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society*, Series B, 50, 413–436.

STATISTISCHES BUNDESAMT (2007): *Statistisches Jahrbuch 2007 für die Bundesrepublik Deutschland*, Wiesbaden: Statistisches Bundesamt.

VON GAUDECKER, H.-M. AND R. D. SCHOLZ (2007): "Differential Mortality by Lifetime Earnings in Germany," *Demographic Research*, 17, 83–108.

YANG, L. AND R. TSCHERNIG (1999): "Multivariate Bandwidth Selection for Local Linear Regression," *Journal of the Royal Statistical Society. Series B*, 61, 793–815.

YATCHEW, A. (2003): *Semiparametric Regression for the Applied Econometrician*, Cambridge: Cambridge University Press.

Chapter 4

Non–Monotonicity in the Income–Longevity Relationship

4.1 Introduction

There is ample evidence supporting a positive relationship between income and life expectancy. Internationally, this finding is confirmed e.g. by Attanasio and Emmerson (2003) for Great Britain or Deaton and Paxson (2004) for the United States, whereas for Germany, a positive relation is found by Reil-Held (2000) and more recently by von Gaudecker and Scholz (2007). Additionally, there are well established theoretical foundations for such a phenomenon, among the earliest the model of health capital introduced by Grossman (1972) and extensively discussed in Grossman (2000), and as a second strand of theory Ehrlich and Chuma (1990).

Both theory and empirical analysis establish the link between income and life expectancy via health (or health capital). In the Grossman (1972) model of health capital, benefits of investments in health capital accrue directly in terms of utility and indirectly in terms of improved productivity. The age at death is in turn implicitly defined by the stock of health capital—death occurs once health capital falls short of a certain threshold.[1] Under some assumptions (namely, stressing the investment property of health capital), Grossman (1972) derives a positive relation between productivity and health (and therefore life expectancy). Yet, it is not undisputed that the positive relation between income and health is really causal. Empirically, the findings are controversial, as Meer et al. (2003) do not find a causal link from income (or socio–economic status in general) to health, while Lindahl (2005) indeed finds such a causality. Adams et al. (2003) produce mixed evidence and find that income or wealth might be causal for some health incidences, but not for all.

The major impetus for the present paper is the result concerning the relationship between the duration of the pension benefit spell and benefit claims, which if positive would disclose the redistributive nature of annuity–based pension systems. In Chapter 3, I find this relation to be monotonously positive. Here I try to answer whether a positive relation between life expectancy and benefit claims is the driving force behind this phenomenon, as age at death as the end of the benefit spell partially determines duration, even more if retirement age as the starting point of the benefit spell does not vary as much (because of legal impediments). Nevertheless, this work is not only a derivative of Chapter 3, but a study in its own right, as non–monotonicity between income and longevity has not been addressed explicitly yet—though some authors actually find a downward–sloping area of the longevity–income relation. So e.g. von Gaudecker and Scholz (2007) examine the

[1] On the question whether age at death in the Grossman model (as the planning horizon of the individual) is under direct control of the invidual, see especially Ried (1998) and in addition Grossman (2000) for a short discussion.

average remaining life expectancy after the age of 65 for different income groups and find that not the lowest income group, but the third or fourth income group (out of a total of 11) experiences the lowest life expectancy (still, von Gaudecker and Scholz 2007 explain the downward–sloping area as being an artifact of the data.).[2] A similar phenomenon can be found in Clark (2007, p. 98) for English males in the 17th century.

So the predominant focus of this work is the relationship between two variables, namely age at death as dependent variable and collected pension benefit claims being a proxy for life time income as explanatory variable. I observe pensioners in the German public pension system who died between 1994 and 2005. In the following analysis I show that a positive relationship between benefit claims and life expectancy is neither necessarily established for all sub–populations, nor does it have to hold for every period of time. For major sub–groups I find that the non–parametric function that links life expectancy to benefit claims is not monotonous, but U–shaped. This is not based on anomalies of the data or an artifact of the estimation technique, so I provide a fundamental conjecture for this shape: Assuming that labor supply elasticities with respect to wages are higher at the lower end of the income (or ability) distribution, higher total income is rather the result of more work than of higher hourly wages, and the former does not add to better health or a longer life.

The remainder of the essay is organized as follows: In Section 4.2, I introduce the data set, in Section 4.3 I propose the econometric method, and the application of this method and the resulting insights are presented in section 4.4. In Section 4.5, I offer a conjecture about the underlying mechanism driving the results, while Section 4.6 concludes.

[2]The same holds for the estimated probability of reaching the age of 74, conditional on having reached the age of 65, which is lowest for the third lowest income group. In their analysis, von Gaudecker and Scholz (2007) actually use a data set similar to the one analyzed here, namely deaths of pensioners in the German public pension system. Their explanation, however, is not convincing. The authors claim that benefit claims as a proxy for life time income may only work well for certain individuals, namely for those who spent most of their career contributing to the pension system. So for some parts of their analysis, they just include individuals with a certain minimum amount of years of contribution. Still, the downward–sloping area survives; their explanation that "[*we*] find production line workers next to their company's CEO in these [*low income*] groups" is wrong, as both (worker and CEO) contribute to the pension system during their whole career (only self–employed and public servants are excluded), and the latter contributes more and is therefore found in a higher income group. So either the analysis after the restriction utilizes the right proxy for income, then the U–shape is present, or the restriction is wrong, then one should have doubts not only with respect to the downward–sloping area of the relationship, but to the upward–sloping area as well. (In addition, von Gaudecker and Scholz 2007 present some results excluding the lowest income groups, which yields a monotonous order of mortality rates of the remaining income groups.)

4.2 The Data

4.2.1 The Variables and Excluded Observations

The data I use here is the collection of pension discontinuations due to death of the beneficiary beginning in 1994 and ending in 2005 (the SUF Demographie Rentenwegfall 1993–2005, see FDZ-RV 2007). The data set contains a 10% stratified sample (based on the federal states) of all pensions that were discontinued, which adds to a total of 828,000 observations. Yet, the data is not based on individuals, but on pensions, as the individual is not the main subject of interest for the pension system. Sometimes both concepts coincide, but accounting for benefits paid to widows and orphans, an individual may receive more than one pension at a time. All these double payments are excluded, and the benefit claims I use as a proxy for life time income are based on own contributions only. The variables I use are the following:

Age at death: calculated from the date of birth and the date of pension discontinuation due to death.

Benefit claims: total amount of benefit claims in points ("*Persönliche Entgeltpunkte*"), capped at 70. Contributions are a constant fraction of income up to a cap, and are transformed into abstract "points", where one point corresponds to contributions based on the average income in a specific year. Pension benefits are then paid according to the current point value.

Residence: aggregated to West Germany, East Germany, and abroad (Berlin adds to West Germany).

Years of contribution: the number of years the pensioner contributed to the pension system, capped at 45.

Months in ill–health: the time spent in ill–health or rehabilitation, as long as relevant for the calculation of pension benefits, capped at 48.

Months in unemployment: the time spent in unemployment, capped at 120.

Type of pension: either old–age (the standard case) or pension paid due to a reduction of the earnings capacity (for readability, from here on denoted as *disability* pension, although from an administrative point of view, this term is not correct)

Factual anonymity of all pensioners in the data set requires that very sparse areas (especially at the right end of the distributions of the respective variables) are

Descriptive Statistics (restricted: males, at least 25 yrs. of contribution, weighted: corrects selection bias)

Data Set	unrestr., unweighted		unrestr., weighted		restr., unweighted		restr., weighted	
	Mean	**St. Dev.**	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**	**Mean**	**St. Dev.**
age at death	77.58	10.83	79.46	10.04	75.58	10.27	77.31	9.86
benefit claims	30.01	18.85	29.42	19.10	42.84	16.23	42.92	16.59
birth	1922.51	11.20	1920.47	10.41	1924.38	10.66	1922.47	10.25
sex female	49.27%	—	51.52%	—	0%	—	0%	—
residence west	75.44%	—	75.57%	—	76.12%	—	76.30%	—
residence east	20.02%	—	19.96%	—	18.50%	—	18.09%	—
residence foreign	4.54%	—	4.47%	—	5.38%	—	5.60%	—
contr. years	29.89	13.58	29.73	13.88	39.03	5.82	39.21	5.74
months ill	2.46	5.81	2.47	5.83	3.04	6.25	3.10	6.33
months unempl.	6.20	17.59	5.90	17.19	6.34	17.11	6.19	16.93
disability pension	8.77%	—	5.78%	—	11.38%	—	8.13%	—
old–age pension	91.23%	—	94.22%	—	88.62%	—	91.87%	—
	$n = 766,311$		$n = 766,311$		$n = 358,173$		$n = 358,173$	

Table 4.1: Descriptive Statistics

empty, such that the econometrician cannot infer on the actual person behind the observation. Therefore, some variables are capped.

Other variables known to have an impact on life expectancy such as number of children, education (or profession), and marital status[3] cannot be used here, as they are not reliable. To give an example: the number of children is only recorded to be larger than zero if it is relevant for the calculation of pension benefits. Yet, only one parent can utilize parenting time for his or her pension, so children are (at best) only recorded for one parent, and not for both.

Female pensioners are excluded in some of the estimations as well, as those who died between 1994 and 2005 belong to generations where female career patterns were distinct from their male counterparts and therefore not comparable. Additionally, I exclude all individuals who worked less than 25 years in a job with mandatory contributions to the pension system in most parts of the analysis for the following reason: I do not observe the actual life time income, but only the pension benefit claims earned by contributions based on a major, yet specific type of income. In Germany, only employed workers have to contribute to the pension system, whereas public servants and self–employed are either excluded from the system, or their contributions are not mandatory. If a pensioner has worked a major part of his career as a civil servant or in self–employment, the benefit claims are only a severely diluted proxy for total income. Additionally, I exclude observations with missing values in key variables as well; the variables not allowed to have missing values are date of birth, nationality, place of residence, date of retirement, and benefit claims. Together with potential double pensions, this amounts to 61,926 observations I exclude. One major drawback arises once I explicitly include years of contribution, months in ill–health, or months in unemployment in the analysis, as they suffer from a lot of missing values as well, together a total of 556,579. Selection effects cannot be ruled out by the exclusion of so many observations, so the main results are presented for the complete and the restricted data set.

Including all observations, the analysis is based on 766,311 observations. Allowing for male pensioners with less than 25 years of contributions only, this number reduces to 358,173. If all control variables are added, the number reduces further down to 110,472. Descriptive statistics of the data set are presented in Table 4.1, with means and standard deviations for all variables used in the analysis.

[3]On the positive impact of education, see e.g. Deaton and Paxson (2004); on the positive impact of marriage or household composition in general see e.g. Martikainen et al. (2005) and Adams et al. (2003). A further factor is the type of occupation (see Hayward et al. 1989 and Moore and Hayward 1990 for its impact on mortality); unfortunately, the number of missing values (83%) in the occupation variable—which is usually not necessary to calculate the pension, and collection of this variable has seized in 2005—is prohibitively high.

4.2.2 Selection Bias and Weighting Function

The sample potentially suffers from a selection bias. Since I observe a death cohort, life expectancies may be biased downwards for the following reason: In each death cohort, a large variety of birth cohorts are included, and life expectancy is known to be increasing with the year of birth almost linearly, see e.g. Statistisches Bundesamt (2007, p. 54) and Human Mortality Database (2005). However, this increase is only partially captured in the sample, as especially individuals from younger birth cohorts (whose ex ante life expectancy should be higher) only appear in the sample if they died relatively young.

The approach to correcting this selection bias is very basic. The selection is not based on individual decision making—it is a matter of data selection alone. Among the later birth cohorts, deaths at young age are over–represented. Since this relationship is empirically linear, a linear weighting function, which decreases with the year of birth, corrects the potential bias. If b_i denotes the year of birth (normalized to zero for the earliest birth cohort), the function ω that assigns the relative weights takes the following form, with s being the slope parameter:

$$\omega_i(b_i) \quad = \quad 1 - s \cdot b_i \tag{4.1}$$

The parameter of choice is only the slope, while the intercept does not matter, because the weights are normalized afterwards. I select the weighting function that minimizes the difference of the weighted average life expectancy in the data officially reported figures for Germany, and still ensures that all weights are non–negative, which is at $s = 0.0103$.[4]

[4]The life expectancy I observe in the data is conditional on reaching a certain age (the specific retirement age). The remaining life expectancies therefore depend on an assumption on the retirement age; I choose the sample average of 60.01, which includes all transitions into old–age pensions and pensions due to a reduction of the earnings capacity (*disability pension*). Taking the complete, unrestricted sample, the remaining life expectancy for men and women (born on average in 1920) at the age of 60 (hence in 1980) was 18.67 years, adding to a total life expectancy of 78.66, which is slightly lower than the 79.46 years reported in Table 4.1, column 2 (after the application of weights). The remaining life expectancy for males only, born in 1922 (the average year of birth in the restricted data set) at the age of 60 was 16.78 years, adding up to a total life expectancy (conditional on reaching 60) of 76.78, which is only marginally lower than the weighted sample average age at death of 77.31 (see Table 4.1, column 4). For the figures of remaining life expectancies in the year 1980, conditional on reaching the age of 60, see Statistisches Bundesamt (2004). The maximum average age at death obtained with this method (i.e., the one which results from the steepest weighting function) does not match exactly the life expectancy observed in population statistics, which can be explained by the fact that my sample does not include self–employed and public servants, on the difference see e.g. Luy (2006).

4.3 Methodology

4.3.1 General Remarks

From here on, I denote the respective dependent variable *age at death* by the scalar y_i (with i being the individual observation) and the main explanatory variable *benefit claims* (a scalar again) by x_i. In order to infer the nature of the relationship between x_i and y_i, it is convenient to estimate this relation non–parametrically, circumventing an imposed linear or higher polynomial structure. Although the focus of this work is on the relationship of these two variables only, there are more variables that should be included in the analysis as well. However, it is generally not convenient to estimate the influence of all covariates non–parametrically, due to the so–called curse of dimensionality. This means that the required amount of observations increases exponentially with the number of regressors, or vice versa, the approximation error increases more than proportionally if the number of observations is held constant, but the dimension of the regressor matrix is increased, see e.g. Yatchew (2003, p. 17). In addition to benefit claims, the only continuous variables I include in the multi–variate non–parametric estimations are years of contribution, months in ill–health, and months in unemployment. The other controls (type of pension, residence, and the birth cohort) are discrete, and I let them enter only in least squares specifications, as artificial smoothing of discrete data may lead to a bias (see Li and Racine 2007, pp.125); furthermore, I stratify the non–parametric analysis along the discrete variables in order to capture their influence.

4.3.2 Locally Linear Estimation and Bandwidth Choice

Denote the non–parametric estimate of y_i by the function $m(x_i)$, which is the solution to the following problem:

$$m(x_i) = \arg\min_{m,\beta} \sum_{i=1}^{n} \left[y_i - m(x_i) - (x_i - x)\beta_1\right]^2 K\left(\frac{x_i - x}{h}\right) \tag{4.2}$$

The estimator $m(x_i)$ is therefore the constant of a linear fit around each x_i, weighting the neighboring observations around x_i with the kernel function $K(\cdot)$. Another representation of the estimation of the local coefficient vector is (Loader 2004)

$$\widehat{\beta} = (x'Wx)^{-1}x'Wy, \tag{4.3}$$

where $\widehat{\beta} = [m(x_i), \beta_1]$ and W is a diagonal matrix with the respective kernel weights on the main diagonal. Note that kernel weights are distinguished from the weights ω in Equation (4.1). In this case of a linear fit, the asymptotic bias of the estimated function $m(x_i)$ is zero, which is not the case for a local constant estimator (Nadaraya-Watson estimator), see Mittelhammer et al. (2000, pp. 622). As Loader (2004) shows, the asymptotic bias will vanish whenever the degree of the polynomial is odd, and especially the bias at the boundaries of the data set will decrease, compared to the Nadaraya-Watson estimator. This property is especially useful in the setting applied here, as a potentially downward–sloping area of the estimated function $m(x_i)$ at the left boundary of x is analyzed. See e.g. Fan and Gijbels (1992), Fan (1992), Pagan and Ullah (1999, pp.105), and Fan and Gijbels (2003, pp.60) for a discussion on this topic: The bias of $m(x_i)$ from locally linear regression does not depend on the density of x, hence it is not subject to the question whether the local regression is performed at the boundaries or in the interior of x.

Still, I have to choose the weighting kernel $K(\cdot)$. There are several major proposals for a weighting scheme, among them the Gaussian and the Epanechnikov kernel. The latter proves to be the efficient one, see e.g. Pagan and Ullah (1999, p. 28). Using the Gaussian kernel, however, no observation (no matter how far from x_i) ever receives a weight of zero, which eliminates some computational burden[5], and is therefore applied here. In general, a kernel function has only to fulfill non–negativity and symmetry around x_i at the center,[6] and the choice of kernel function is only a minor determinant of the later results. The main difference in the application of kernels is their relative efficiency (as compared to the Epanechnikov kernel), where the Gaussian kernel I apply here reaches .9512. On properties of kernels and their efficiency, see e.g. the discussion in Mittelhammer et al. (2000, pp.602).

After the choice of the kernel, a bandwidth h has to be determined. A bandwidth chosen too high will leave the estimate 'over–smoothed' and potentially ignores specific patterns, whereas an under–smoothed estimate may hide the pattern of interest behind erratic components, leading in the limit (as $h \to 0$) to an exact replication of the unfitted data. This phenomenon is known as the bias–variance–tradeoff (see e.g. Yatchew 1998). The optimal bandwidth can be approximated by a rule–of–thumb, which may be advisable while using large data sets. The method I apply can be understood as a refinement of Silverman's Rule–of–Thumb, which is

[5] A large number of zero weights may yield a computational difficulty, which is due to singular matrices. The matrix $x'Wx$ may be singular for certain outcomes of the kernel weights, and thus not invertible.

[6] In addition, as the kernel determines weights, it has to integrate to one, and with exception of the center x_i, it has to be continuous (this includes, e.g., the triangular kernel with a non–differentiable kink at x_i).

broadly discussed in the literature.[7] Assuming a normal kernel, Silverman (1986) proposes the bandwidth to be

$$h = 1.06\sigma_x n^{-1/5}, \tag{4.4}$$

where σ_x is the standard deviation of the x–variable and n the number of observations. Yet, this formula relies on parametrical distributional assumptions.[8] These assumptions can be replaced by distributional properties of the kernel function and the data itself, i.e. by measures of the variance and skewness, to derive an improved plug–in method.

One specific plug–in method specifying the bandwidth is characterized by Loader (1999) and Loader (2004), who proposes the optimal bandwidth to be

$$h = \left(\frac{\sigma^2 (b-a)^2 \int K(v)^2 dv}{n \left(\int v^2 K(v) dv \right)^2 \int m''(x)^2 dx} \right)^{1/5}, \tag{4.5}$$

where σ^2 is the error variance, $m''(x)$ is the second derivative of the estimated function, and a and b are the lower and upper bounds of x. Using a first stage or pilot estimate, the error variance can be estimated by

$$\widehat{\sigma^2} = \frac{1}{n - 2\nu_1 + \nu_2} \sum_i^n \left[y_i - m(x_i) \right]^2, \tag{4.6}$$

with ν_1 and ν_2 adjusting the degrees of freedom (see Loader 2004 for the computation). The second derivative $m''(x)$ of the estimate is obtained by fitting a local *quadratic* function (as the pilot estimate) to the data first, hence by solving

$$m(x_i) = \arg \min_{m,\beta_1,\beta_2} \sum_{i=1}^n \left[y_i - m(x_i) - (x_i - x)\beta_1 - (x_i - x)^2 \beta_2 \right]^2 K(\cdot). \tag{4.7}$$

An estimate for the second derivative is then given by the vector $2\widehat{\beta_2}$. However, there remains a pilot bandwidth to be chosen, as the respective $\widehat{\beta_2}$ and $m''(x)$ are sensitive to the bandwidth as well. Following Silverman's Rule–of–Thumb, I choose the pilot bandwidth to be $1.06\sigma_x n^{-1/5}$. Of course, the final bandwidth of

[7] See e.g. the textbooks by Fan and Gijbels (2003, pp. 47) or Li and Racine (2007, pp. 14).

[8] See also Pagan and Ullah (1999, p.103), who propose the bandwidth to be of the order of magnitude $n^{-1/5}$.

Equation (4.5) varies with the pilot bandwidth. To give a short overview, a test with the data analyzed here resulted in an under–proportional inverse relation of pilot bandwidth with optimal bandwidth, meaning that the optimal bandwidth does not vary as much as the pilot bandwidth.[9]

Applying this method unconditionally or conditionally on certain outcomes of control variables (stratification), I denote as strategy (I).

4.3.3 Approximate Confidence Interval

I approximate a point–wise confidence interval around the estimate of $m(x_i)$ using conditional standard errors $\sigma(x_i)$ at each grid point of x. The confidence bounds are given by (see Härdle et al. 2004, pp.119)

$$m_{\text{CB}}(x_i) = m(x_i) \pm z_\alpha \sqrt{\frac{||K||_2 \widehat{\sigma}^2(x_i)}{nh\widehat{f}(x_i)}}, \tag{4.8}$$

where z_α is 2.58, given the number of observations and the desired confidence level of 99%. The estimated conditional (or local) standard error is

$$\widehat{\sigma}^2(x_i) = \frac{1}{n} \sum w(x_i) \left[y_i - m(x_i)\right]^2, \tag{4.9}$$

the density $\widehat{f}(x_i)$ of x_i is a non–parametric estimate applying the Gaussian kernel, and $||K||_2$ denotes $\int_u K^2(u)du$, which is 4.37335 in the case of the normal kernel. The kernel weights $w(x_i)$ are the respective elements of the weighting matrix W, see Equation (4.3).

4.3.4 Locally Linear Estimation in Higher Dimensions

In order to include further control variables, x is not bound to be one–dimensional. If a graphical inspection of the complete result is desired, the largest number of independent variables is two—the result is then a surface $m(x_{1i}, x_{2i})$ above the (x_1, x_2)–plane. Yet, higher dimensions are also possible, despite the drawback of the 'curse of dimensionality', see Yatchew (2003, p. 17). But due to the large number of observations in the data set and the small number of regressors I use here, the rate of convergence is not a restriction. If the dimension of x is greater than

[9] In an exemplarily chosen subset of the data, Silverman's Rule–of–Thumb yields $h^{\text{PILOT}} = 2.36$, and the optimal bandwidth of an exemplary estimation is $h^{\text{OPT}} = 4.644$. A smaller pilot bandwidth yields a higher optimal bandwidth and vice versa: $h^{\text{PILOT}} = 4.72 \Leftrightarrow h^{\text{OPT}} = 3.59$ and $h^{\text{PILOT}} = 1.18 \Leftrightarrow h^{\text{OPT}} = 7.92$.

two, either a surface or a vector can be sliced out of the array of results, holding the remaining variables constant at a certain level, such that the result (in the case of a vector m) can be written as $m(x_{1i}|x_{2i} = X_2, x_{3i} = X_3, \ldots)$. In Section 4.4, I will refer to this strategy as strategy (II).

For the case of higher dimensions, there also exist methods for the bandwidth choice. They are discussed e.g. by Yang and Tschernig (1999), who propose rule–of–thumb and plug–in bandwidths. I adopt their notion that the optimal bandwidth in the multivariate case is of the order of magnitude of $n^{-1/(4+d)}$, with d being the number of regressors, which will be four in this analysis. Yet, beyond this optimal order of magnitude, I choose $h = 1.06\overline{\sigma_x}n^{-1/(4+d)}$, with $\overline{\sigma_x}$ being the average sample standard deviation of the regressors, deliberately omitting further refinements due to their computational burden.

4.3.5 Comparative Least Squares Estimation

To compare and to quantify the non–parametric results, I perform least squares regressions in different specifications. First of all, the specifications differ with respect to the method applied, while secondly, different sets of control variables are implemented. In the first dimension, I compare ordinary least squares to weighted least squares. In the second dimension, benefit claims always enter as a polynomial of degree two, augmented with different sets of covariates. In addition, I include the complete data set in some specifications, while in others, only male pensioners with at least 25 years of contributions are considered. Some of the covariates are dummies; in the case of residence, I choose *west* to be the reference category, in the case of pension type, the reference category is *old–age pension*. The model without additional controls for the unweighted data set is

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon, \tag{4.10}$$

while the model in Equation (4.10) estimated by weighted least squares is

$$y_i\sqrt{\omega_i} = \beta_0\sqrt{\omega_i} + \beta_1\sqrt{\omega_i}x_i + \beta_2\sqrt{\omega_i}x_i^2 + \widetilde{\epsilon}, \tag{4.11}$$

with w_i as defined in Equation (4.1). The same applies to the models with additional control variables. In Section 4.4, I will refer to this strategy as strategy (III).

4.4 Implementation and Results

4.4.1 General Results and Least Squares Regressions

Most of the specifications of the uni–variate strategy (I) discover a non–monotonous relationship between age at death and benefit claims (see Figures 4.1 through 4.6). Refer to Table 4.2 for the respective size of the sub–groups and the resulting optimal bandwidths. In some specifications, the multi–variate strategy (II) cannot corroborate this result, while strategy (III) confirms that—in all specifications—age at death is not monotonously increasing in benefit claims, as benefit claims enter negatively, while benefit claims squared enter positively. These signs are all significant. Still, the majority of results discovered by the non–parametric and multi–variate strategy (II) does not reproduce a monotonously positive relation either: The conditional expectations plotted in Figures 4.4 to 4.6 exert a slightly hump–shaped pattern, with decreasing parts to the right of the benefit claims distribution.

For a complete overview of the least squares results, see Table 4.3. It is important to see that the general shape of the benefit claims–age at death relationship does not vary with the inclusion or exclusion of control variables, although they highly improve the predictive quality of the estimated model. The R^2 without additional controls is very low, despite the significance of benefit claims.

4.4.2 Results by Sex and by the Application of Weights and Restrictions

In summary, the existence of a downward–sloping area in the analyzed relation is not subject to the weighting scheme or the imposition of restrictions on the data set to individuals with at least 25 years of contribution (see Figure 4.1). Without weights and restrictions, the downward–sloping area is present for men, but even dominates for women, which is a rationale for the exclusion of women from stratified estimations below. The other specifications yield U–shaped relationships for men and women likewise, yet, on slightly different levels with slightly different minimums. In general, women live longer than men, a fact confirmed by the least squares regressions. If the restriction to at least 25 years of contributions is in place to guarantee that benefit claims serve as a good proxy for income, the minimum life expectancy for men is around benefit claims of 15 points (unweighted) or 12 points (weighted). In this area of benefit claims, the number of observations is admittedly small: in the specification with the smallest downward–sloping area (males, weighted and restricted data set), I find that 2.5% of all observations lie there, while in other specifications up to 11% of the male population and the vast majority of the female population lie on the decreasing part. Yet, the result can never be neglected

Size of Sub–Populations and Optimal Bandwidths (Strategy [I])

Specification	Obs.	Bandwidth	Specification	Obs.	Bandwidth
by sexs, Fig. 4.1			by residence, Fig. 4.3		
– upper left, male	388,754	34.6	– lower, west	78,529	15.2
– upper left, female	377,577	36.1	– lower, east	30,921	8.0
– upper right, male	358,173	36.0	– lower, abroad	2,919	3.6
– upper right, female	340,634	34.7	by health, Fig. 4.4		
– lower left, male	388,754	36.5	– left, mts. $= 0$	60,055	12.0
– lower left, female	377,577	45.4	– left, mts. < 6	32,862	8.2
– lower right, male	358,173	51.7	– left, mts. ≥ 6	19,452	6.8
– lower right, female	340,634	62.0	by unempl., Fig. 4.5		
by cohort, Fig. 4.3			– left, mts. $= 0$	79,112	13.8
– upper left, > 1920, ≤ 1930	128,830	19.1	– left, mts. < 6	26,294	8.4
– upper left, > 1930, ≤ 1940	79,921	13.4	– left, mts. ≥ 6	22,554	8.2
by pension type, Fig. 4.3			by contr. years, Fig. 4.6		
– upper right, old–age	86,635	12.0	– left, yrs. < 30	10,710	7.1
– upper right, disability	25,734	8.5	– left, yrs. < 40	39,142	10.3
			– left, yrs. ≥ 40	62,517	12.5
			multi–variate, Fig. 4.4–4.6	112,369	2.8

Pilot bandwidths are chosen to be $h = 1.06\sigma_x n^{-1/5}$ (Silverman's Rule–of–Thumb)

Table 4.2: Size of Sub–Populations and Optimal Bandwidths

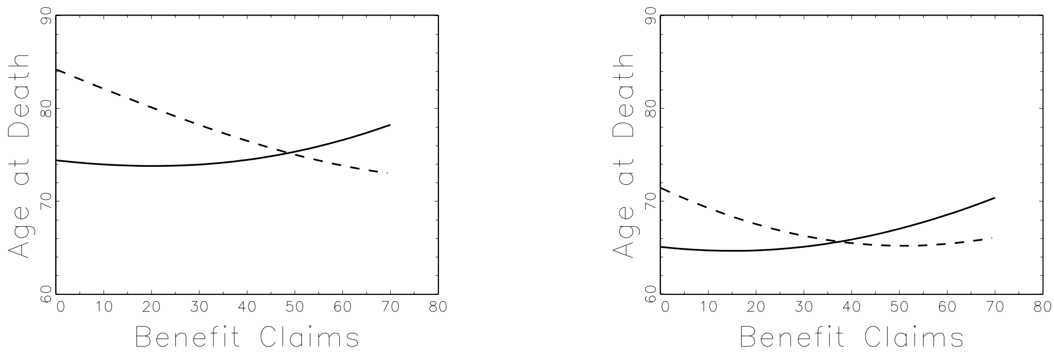
Results of the Least Squares Estimations—Different Specifications

Specification	LS, complete data				LS, restr. data				WLS, complete data				WLS, restr. data			
	β	S. E.	β	S. E.	β	S. E.	β	S. E.	β	S. E.	β	S. E.	β	S. E.	β	S. E.
benefit claims	-.535	.002	-.305	.005	-.501	.004	-.358	.008	-.430	.002	-.210	.005	-.427	.003	-.375	.009
benefit claims squ.	.007	.000	.004	.000	.007	.000	.005	.000	.005	.000	.003	.000	.006	.000	.005	.000
contr. years			.011	.002			.004*	.004			-.068	.002			-.064	.004
months ill			.029	.002			.043	.002			.001*	.002			.023	.002
months unempl.			-.015	.001			-.003	.001			-.020	.001			-.008	.001
residence east			.312	.038			.855	.039			.842	.040			1.24	.043
residence foreign			.340	.072			-.868	.149			.558	.079			-1.04	.17
female			.402	.038							.854	.040				
old–age pension			14.19	.037			10.90	.035			13.49	.034			10.96	.037
constant	85.30	.030	59.65	.074	82.76	.071	63.54	.184	85.80	.027	61.08	.077	83.52	.064	66.61	.203
R^2	.075		.521		.050		.501		.063		.456		.039		.459	
observations	766,331		209,752		356,276		110,472		766,331		209,752		356,276		110,472	

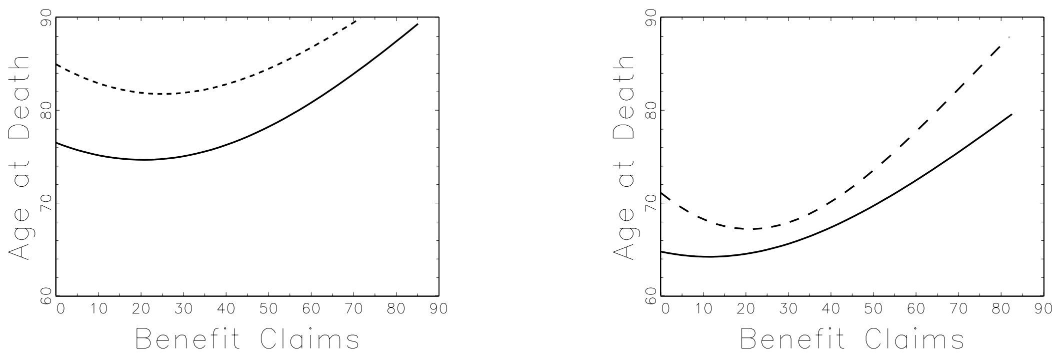
Against the usual conventions, the asterisk (*) denotes that the respective coefficient is *not* significantly different from zero. All other coefficient are significant at least on the .99 level. Standard errors are robust.

Table 4.3: Results of the Least Squares Estimations

for the following reasons: First, this downward–sloping area is corroborated by a variety of results of different estimation strategies and specifications. Second, the estimation technique is explicitly applied because of its unbiasedness at the boundaries.



By sex (strategy [I]); solid: male, dashed: female. Left panel: unrestricted and unweighted. Right panel: restricted and unweighted.



By sex (strategy [I]); solid: male, dashed: female. Left panel: unrestricted and weighted. Right panel: restricted and weighted.

Figure 4.1: Results by Sex

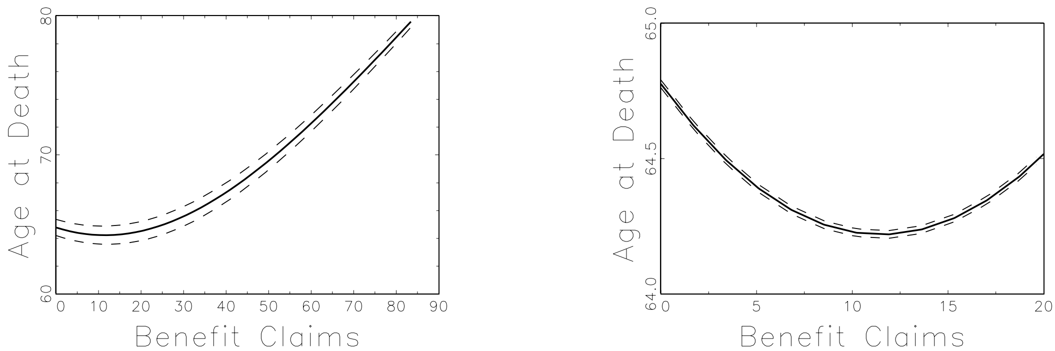
For a robustness check, refer to Figure 4.2. I plot approximate point–wise confidence bands on the 99% level around the original estimate (including males with at least 25 years of contribution). In the left panel, I use naïve confidence bands based on standard errors only,[10] whereas in the right panel, the confidence bands follow Härdle et al. (2004, pp.119), see also Section 4.3.3. The confidence interval in the left panel does not confirm the significance of the downward–sloping leg yet, because the maximum of the lower confidence band (at the very left bound,

[10]The naïve confidence bands are given by

$$m_{\mathrm{CB}}(x_i) = m(x_i) \pm z_\alpha \sqrt{\frac{1}{n} \sum w(x_i) \left[y_i - m(x_i)\right]^2},$$

where $w(x_i)$ denote the kernel weights.

$m[x_i = 1] = 64.2$) is not larger than the minimum of the upper confidence band (with $m[x_i = 12] = 64.9$), such that the function $m(x_i)$ could also increase between $x = 1$ and $x = 12$. The confidence bands in the right panel, however, are so narrow that I present only a zoom on the area of x between zero and 20, where I find the downward–sloping leg (otherwise, the reader could not distinguish the confidence bands and the original estimate). This version of the confidence bands clearly corroborates the U–shaped relationship. Varying the bandwidth as a further robustness check shows that the globally optimal bandwidth works against the case of non–monotonicity. Restricting the analysis to the left part of the x distribution and choosing a new optimal bandwidth for this sub–population considerably shifts the minimum to the right and increases the fraction of observations on the negative slope above 11%, even in the weighted and restricted case.

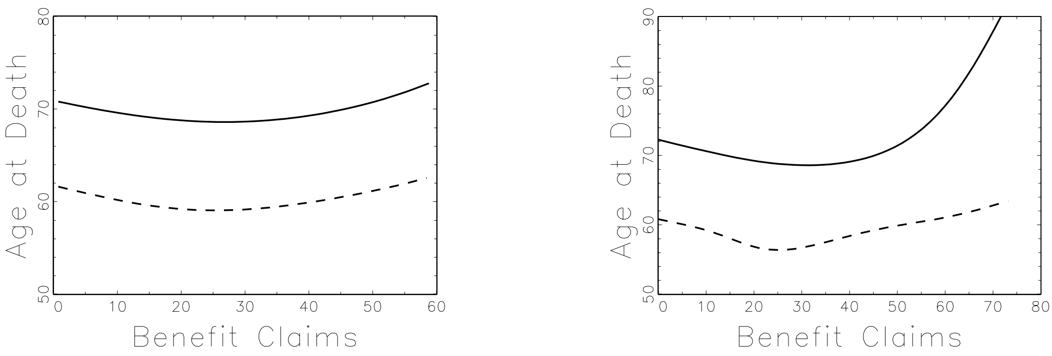


Strategy [I], males only. Solid: original estimate, dashed: approximate confidence bands (on the 99% level). Left panel: with naïve standard errors. Right panel: with standard errors following Härdle et al. (2004), scaled. (Data set: weighted and restricted)

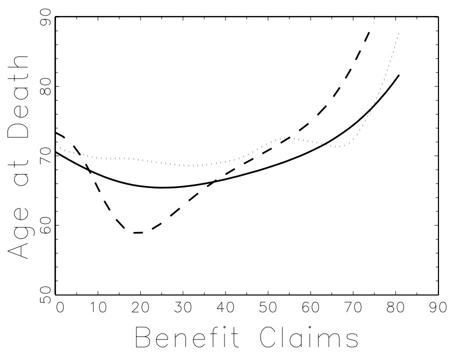
Figure 4.2: Approximate Confidence Bands

4.4.3 Results by Birth Cohorts, by Type of Pension, and by Residence

Another possibility to control for increasing life expectancy over time—instead of weighting the different birth cohorts—is a restriction of the analysis to specific birth cohorts in the first place. In Figure 4.3, I apply this stratification and partition the data in two cohorts born between 1921 and 1930 and 1931–1940, respectively. Considering that deaths occurred between 1994 and 2005, these cohorts are chosen because their realized age at death of the older cohort is by construction relatively close to the age at death which can be expected, conditional on reaching retirement age. The sample is not weighted, but restricted to males with at least 25 years of contribution, and the U–shaped result survives the stratification. The difference in the level of longevity between the two cohorts is an artifact of the construction of the cohorts and is due to the absence of weighting: In the younger cohort, early deaths are relatively over–sampled as compared to the older cohort.



Left panel: By birth cohort (unweighted and restricted; strategy [I]); solid: yr. of birth $\in$ (1920, 1930], dashed: yr. of birth $\in$ (1930, 1940]. Right panel: By type of pension (weighted and restricted; strategy [I]); solid: old–age pension, dashed: disability pension. (Data set: unweighted and restricted, weighted and restricted)



By residence (unweighted and restricted; strategy [I]); solid: west, dashed: east, dotted: abroad. (Data set: unweighted and restricted, weighted and restricted)

Figure 4.3: Results by Cohorts, Type of Pension, and Residence

The results stratified along the type of pension (see also Figure 4.3) does not come as a surprise: men who die receiving an old–age pension live longer than men who die receiving a disability pension. The explanation is two–fold; first, individuals with a disability pension suffer from a shorter life expectancy as a result of their respective illness, an illness so severe to prohibit a later retirement into the regular old–age pension.

Second, an institutional factor shapes this result, because once eligible, disability pensions are transformed into old–age pensions, such that an individual receiving a disability pension first, but reaching the age of 65 will be characterized by the outcome 'old–age pension'.[11] The sign of the respective coefficient in the least

[11]Under certain circumstances, a pensioner receiving a pension due to a reduction of his earnings capacity can already be transformed into an old–age pension at the age of 60 (only after an application of the pensioner).

squares regressions corroborates that old–age pensioners benefit from higher life expectancy. Altogether, the U-shape emerges for both groups.

Although the majority of observations live in West Germany, I provide estimation results for East Germany and abroad as well (Figure 4.3). All groups exert the U-shape, however, to a different degree. The dip is strongest for East Germany, and least pronounced for those living abroad. Further differences between the three residence groups are not conclusive; the non–parametric estimates do not disclose a clear order between all groups, while the signs of the least squares regressions for pensioners living abroad depend on the respective specifications (though the coefficients are always significant).

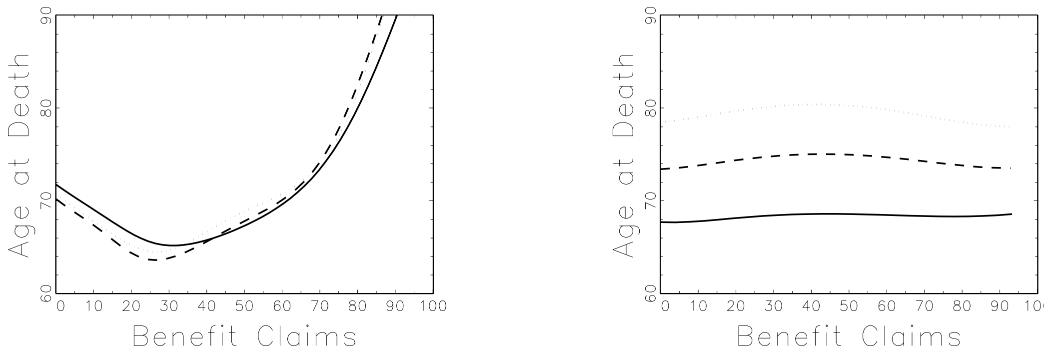
Interestingly, I find in the parametric estimates that people in East Germany live slightly longer, despite the fact that living standards in East Germany are ceteris paribus lower than in West Germany.[12] Yet, this finding is compatible e.g. with von Gaudecker and Scholz (2007), who find that for some income groups, life expectancy is higher in East Germany. Furthermore, official records (see Statistisches Bundesamt 2007, p. 54) show that in 2005, remaining life expectancy conditional on reaching 60 was only slightly higher in the West (20.4 years compared to 19.7 years in the East), and there are birth cohorts whose unconditional life expectancy at birth in the year was even higher in the East.[13] In contrast to population statistics, the population I observe consists of pensioners of the public pension system and is conditioned on 25 years of work, and since different legislation applied for the collection of benefit claims in both parts of the country (until reunification), the observed sub–populations in East and West may differ in their relationship to the particular total populations. Although benefit claims for men are less dispersed in East Germany (15.03 in the West, compared to 11.75 in the East), the income gradient as graphed in Figure 4.3 is stronger: there are less income differences, but if there are differences, their impact on life expectancy is stronger in the East.

4.4.4 Results by Months in Ill–Health

I apply three distinct strategies to identify the impact of health status on the relationship between life–expectancy and benefit claims (see Figure 4.4). First, applying strategy (I), the uni–variate analysis is stratified along the ill–health dimension, hence, I estimate the relationship three times, contingent on the outcome of the stratification variable—namely, months spent in ill–health being equal to zero;

[12] Average income for men in the years 2002 to 2006 in the East reached only 71.0% to 75.2% of the income in West Germany, see Statistisches Bundesamt (2007, p. 523).

[13] E.g. the birth cohort of 1951 and 1952, covering pensioners who also appear in the data I analyze here. The life expectancy at birth in the Eastern part of Germany was 65.1 years compared to 64.6 years in the West.



Left panel: Uni–variate (strategy [I]); solid: mts. $= 0$, dashed: mts. < 6, dotted: mts. ≥ 6. Right panel: Multi–variate (strategy [II]); solid: mts. $= 0$, dashed: mts. < 6, dotted: mts. ≥ 6. (Data set: weighted and restricted)

Figure 4.4: Results by Months in Ill–Health

smaller than six, but strictly positive; and greater or equal than six. The same groups are constructed for strategy (II).[14] Denote benefit claims by x_{1i}, x_{2i} denotes months in ill–health, x_{3i} denotes months in unemployment, and x_{4i} denotes years of contribution, and finally, $\overline{x_{ji}}$ is the respective sample average over all x_{ji}. Hence, the three plots in the right panel of Figure 4.4 represent the conditional moments

$$\begin{aligned} m_1(x_{1i} \quad &| \quad x_{2i} = 0, x_{3i} = \overline{x_{3i}}, x_{4i} = \overline{x_{4i}}) \\ m_2(x_{1i} \quad &| \quad x_{2i} \in (0,6), x_{3i} = \overline{x_{3i}}, x_{4i} = \overline{x_{4i}}) \\ m_3(x_{1i} \quad &| \quad x_{2i} \geq 6, x_{3i} = \overline{x_{3i}}, x_{4i} = \overline{x_{4i}}), \end{aligned} \tag{4.12}$$

The uni–variate procedure (I) produces little to no difference in the resulting relationships, and the downward–sloping area for low benefit claims survives, despite the stratification.

The ordering of conditional moments extracted from the array of results (strategy [II]) for three different groups is unique, with individuals with zero months in ill–health at the bottom, and the group with the longest spell in ill–health at the top. This is corroborated by strategy (III), which yields either insignificant results for the influence of ill–health (WLS estimation on the complete data set) or significantly positive impact (all other specifications), meaning that individuals with rehabilitation spells or months in ill–health actually live longer. This result is explicitly not an artifact of individuals in bad health being more likely to claim disability pensions

[14]The three different groups in specification (II) match as closely as possible the groups of specification (I); a perfect match, however, is not possible, for the following reason: Applying strategy (I), the sub–groups are extracted *before* the estimation, while in strategy (II), the sub–groups are constructed afterwards, based on a grid of 25 points over each x_{ji}. Specifying the sub–groups on the grid after the estimation may slightly shift the cut–off limits.

than people in good health, which is not the case (the average number of months spent in ill–health is 3.2 for individuals with an old–age pension, and only 2.6 for individuals with a disability pension). Potential explanations for this seemingly counter–intuitive result are the following: first, means of rehabilitation are actually effective and extend life–expectancy. Second, disability pensions and times in ill–health are substitutes, implemented for the same general reason (ill–health), but at different levels of the outcome. While months in ill–health are by definition a temporary means to improve the situation of ill individuals, the disability pension is more likely to be ultimate and applied for more severe degrees.[15]

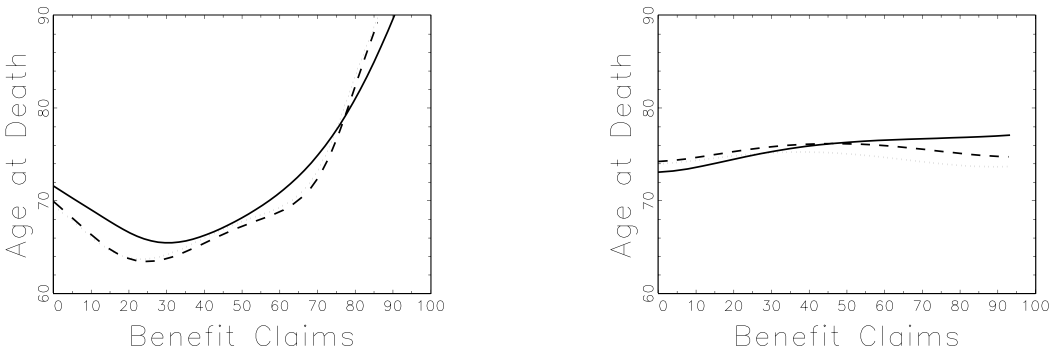
The latter explanation is compatible with results stratified by the type of pension (Figure 4.3 and Section 4.4.3), because individuals receiving a disability pension live shorter as compared to individuals with old–age pensions. The former explanation is also compatible with results produced of Chapter 3, where I find that—on average—bad health increases the duration of the benefit spell (while the benefit claims–gradient is stronger for worse health). However, including health as explanatory variable in strategy (III) does not alter the U–shape, whereas it vanishes in strategy (II).

4.4.5 Results by Months in Unemployment

The procedures (I), (II), and (III) applied to identify the impact of unemployment on the longevity–benefit claims nexus are the same as proposed in detail in Section 4.4.4. The cut–off limits are the same as well, hence months in unemployment being zero, smaller than six, and greater or equal than six for strategy [I], and the conditional moments in strategy [II] are constructed very similar to Equation (4.12), this time holding x_{2i} constant at its sample average and varying x_{3i}. While strategy (I) produces again little to no difference for different unemployment groups, the results following strategy [II] are not as nicely ordered as in the case for ill–health. If there was no time in unemployment at all, the discovered relationship is indeed monotonously increasing (which is not true for the groups with strictly positive unemployment spells).

Strategy (III) discovers a significantly negative impact of unemployment on longevity, which is in line with studies explicitly analyzing unemployment as explanatory factor of mortality or life–expectancy, such as Gerdtham and Johannesson (2003).

[15] Although, from an institutional point of view, the pension paid due to a reduction of the earnings capacity is also intended to be temporary—an intention partially contradicted by the relatively large number of pensioners actually dying while they receive this type of pension, compare Table 4.1.

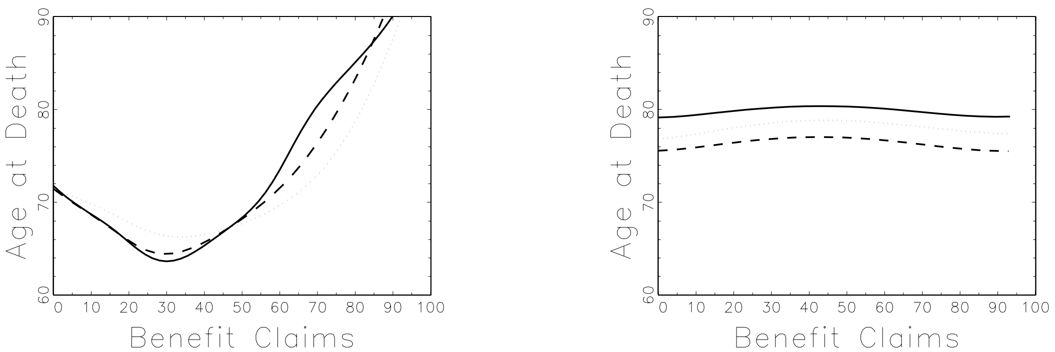


Left panel: Uni–variate (strategy [I]); solid: mts. = 0, dashed: mts. < 6, dotted: mts. ≥ 6. Right panel: Multivariate (strategy [II]); solid: mts. = 0, dashed: mts. < 6, dotted: mts. ≥ 6. (Data set: weighted and restricted)

Figure 4.5: Results by Months in Unemployment

4.4.6 Results by Years of Contribution

Again, strategy (I) is not able to find systematic differences in the analyzed base relationship, stratified by years of contribution to the pension system (the different groups are composed as follows: individuals with strictly less than 30 years of contributions, individuals with between 30 and 40 years, and individuals who contributed at least 40 years.).



Left panel: Uni–variate (strategy [I]); solid: yrs. < 30, dashed: 30 ≤ yrs. < 40, dotted: yrs. ≥ 40. Right panel: Multi–variate (strategy [II]); solid: yrs. < 30, dashed: 30 ≤ yrs. < 40, dotted: yrs. ≥ 40. (Data set: weighted and restricted)

Figure 4.6: Results by Years of Contribution

Strategy (II) fails to deliver a simple answer with respect to the impact of the length of the work life: Individuals in the group with the fewest years of contribution have on average the highest life expectancy, which is consistent with the view that early retirement (hence, a shorter career) serves as a means of investment in longevity (see Wolfe 1985 for a theoretical treatment). Yet, the order of the remaining two groups does not perfectly fit in, because the intermediate group enjoys on average a strictly lower life expectancy than the group with the most years of contribution. This phenomenon together with the general finding of a non–monotonous link between benefit claims and longevity gives rise to the theoretical conjecture proposed in Section 4.5. The three distinct groups for strategy (II) are constructed by holding x_{2i} and x_{3i} constant at their sample averages and setting the cut–off levels for x_{4i} as in strategy (I).

The results from strategy (III) are mixed; unweighted least squares finds no impact (data set restricted to males with at least 25 years of contributions) or a positive impact (complete data), while the WLS regressions discover a negative impact, which corroborates the sequence derived from strategy (II), namely that the group with the fewest years of contribution lives longer than all other groups.

4.5 A Theoretical Conjecture on the Non–Monotonous Longevity–Income–Relationship

So far, theory has provided little or no explanation for the U–shaped link between income (in the present analysis measured by pension benefit claims) and life expectancy. In the following I provide a conjecture which also utilizes the non–monotonous relationship between life expectancy and the length of the work life (see Figure 4.3).[16]

Assume that individual productivity is denoted by a, and that individuals are paid according to their marginal product. Life time utility depends on consumption (or income) and labor, such that utility maximization yields life time labor supply $L(a)$. Life time income from labor is therefore $Y(a) = aL(a)$, and assume that $Y(a)$ is non–decreasing in a (agent monotonicity). Denote age at death by T, and let T be a function of (instantaneous) income a and labor supply $L(a)$, such that $T = T[a, L(a]$. It is straightforward to assume that $T[a, \cdot]$ unambiguously increases in its first argument, hence in productivity a, compare e.g. Grossman (2000). Assume further that $T[\cdot, L(a)]$ decreases in $L(a)$; the results in Figure 4.6 and Table 4.3 support this assumption.[17] Then, the derivative of $T[\cdot, \cdot]$ with respect to a is

$$\frac{dT[\cdot, \cdot]}{da} = \frac{\partial T}{\partial a} + \frac{\partial T}{\partial L} \frac{L(a)}{a} \epsilon_{L,a}. \tag{4.13}$$

The sign of Equation (4.13) is ambiguous since it crucially depends on the mag-

[16] A further rationale for the U–shape phenomenon I provide in Chapter 2 in a multi–task moral–hazard framework.

[17] Furthermore, Ruhm (2000) and Ruhm (2007) justifies this assumption: For aggregate variables he finds that in times of higher unemployment (hence, in times with less work), mortality declines. Yet, Johansson (2004) challenges this result and finds a negative effect of hours worked on mortality.

nitude of the wage elasticity of labor supply $\epsilon_{L,a}$. Because of a time constraint, e.g. due to retirement legislation which does not allow to retire later than a certain threshold, or due to a 'natural' upper limit of $T[\cdot,\cdot]$, the wage elasticity is small (possibly even negative) for large a. In this case, the first term on the right hand side of Equation (4.13) dominates (or both terms have the same sign), and $dT[\cdot,\cdot]/da$ is positive. If $\epsilon_{L,a}$ is positive and relatively large for small a, the sign of $dT[\cdot,\cdot]/da$ is negative, such that over the whole support of a, a U–shaped age at death function $T[\cdot,\cdot]$ emerges. Agent monotonicity, hence $dY(a)/da > 0$, completes the argument, because the sign of $dT[\cdot,\cdot]/dY$ can be directly derived, and the assumed properties allow for a U–shaped relation between life expectancy and life time income.

This theoretical conjecture is supported by differential results of labor supply with respect to sex, as men and women differ in wage elasticities—the wage elasticity of women is known to be greater than the one of men.[18] A higher elasticity of labor supply together with a low realization of a yields a large second term on the right hand side of Equation (4.13), predicting a steeper decrease of life expectancy in this area. Furthermore, a higher a may not offset the large $\epsilon_{L,a}$, such that $T[\cdot,\cdot]$ may be decreasing over the whole support of a and hence over $Y(a)$.

The first prediction (a steeper decrease of life expectancy for women with low a as compared to men) is satisfied for three out of four specifications in Figure 4.1 (all except of the 'unrestricted and weighted' specification in the lower left panel), and even the second prediction (a monotonously *decreasing* relationship) can be found, in the 'unrestricted and unweighted' estimation (upper left panel of Figure 4.1).

4.6 Summary and Outlook

Non–parametric estimation techniques applied to a sample of pensioners receiving a pension from the German public pension system who died between 1994 and 2005 challenge the perceived monotonicity of the income–longevity–relation. Without any correction of the selection bias, which results from under–sampling of later–born individuals with ex ante higher life expectancy, the relationship I observe is clearly non–monotonous, but U–shaped. If a weighting scheme takes this bias into account, the downward–sloping area at the left end of the benefit claims distribution is smaller, but still present. More importantly, this pattern is found for all stratifications applied to the uni–variate estimation and for all least squares specifications. The same non–monotonous link is also found for specific birth cohorts, which are approximations to self-contained cohorts. Applying non–parametric multi–variate methods, the U–shape vanishes, yet, not in favor of a monotone in-

[18]See e.g. Heckman (1995) or van Soest (1995).

creasing function. Confirming the robustness of this result with respect to sub–groups of the population and the significance levels in the least squares regressions, a confidence band around the downward–sloping area is very narrow and uniformly decreasing at the low end of the benefit claims distribution. Yet, benefit claims have quite low explanatory power for life expectancy in the parametric regressions. The fact that the distribution of age at death contingent on benefit claims is still widespread deserves further research, e.g. via the application of quantile regressions. Furthermore, it has to be analyzed whether the extracted pattern can be found in other countries as well.

In a conjecture for an explanation for this pattern, I apply the concept of different elasticities of labor supply with respect to wages. As there is a natural upper limit to labor supply, the wage elasticity has to decline at the upper limit of the productivity distribution, independent of the question whether labor supply is monotonously increasing in wages or not. Beginning at the bottom of the income distribution, a higher life time income may therefore be caused either by higher wages or by higher labor supply, and if the latter reduces life expectancy, the analyzed relationship between life time income and life expectancy may be decreasing at least for low productivity/income individuals. An elaborate theoretical model on this phenomenon—including weekly and life–time labor supply decisions—is an object of my future research, together with a detailed empirical corroboration.

References of Chapter 4

ADAMS, P., M. D. HURD, D. MCFADDEN, A. MERRILL, AND T. RIBEIRO (2003): "Healthy, Wealthy and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status," *Journal of Econometrics*, 112, 3–56.

ATTANASIO, O. P. AND C. EMMERSON (2003): "Mortality, Health Status, and Wealth," *Journal of the European Economic Association*, 1, 821–850.

CLARK, G. (2007): *A Farewell to Alms*, Princeton: Princeton University Press.

DEATON, A. AND C. PAXSON (2004): "Mortality, Education, Income, and Inequality among American Cohorts," *NBER Working Paper*, No. 7140.

EHRLICH, I. AND H. CHUMA (1990): "A Model of the Demand for Longevity and the Value of Life Extension," *Journal of Political Economy*, 98, 761–782.

FAN, J. (1992): "Design Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

FAN, J. AND I. GIJBELS (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics*, 20, 2008–2036.

——— (2003): *Local Polynomial Modelling and Its Applications*, Boca Raton: CRC Press.

FDZ-RV (2007): "Scientific Use File Demografiedatensatz Rentenwegfall 1993-2005, SUFRTWFjjXVSTDemo," Data set from Deutsche Rentenversicherung Bund.

GERDTHAM, U.-G. AND M. JOHANNESSON (2003): "A Note on the Effect of Unemployment on Mortality," *Journal of Health Economics*, 22, 505–518.

GROSSMAN, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80, 223–255.

——— (2000): "The Human Capital Model," in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, New York and Oxford: Elsevier Science, North-Holland, vol. 1A, 347–408.

HAYWARD, M. D., W. R. GRADY, M. A. HARDY, AND D. SUMMERS (1989): "Occupational Influences on Retirement, Disability, and Death," *Demography*, 36, 393–409.

HECKMAN, J. J. (1995): "What Has Been Learned About Labor Supply in the Past Twenty Years?" *American Economic Review*, 83, 116–121.

HÄRDLE, W., M. MÜLLER, S. SPERLICH, AND A. WERWATZ (2004): *Nonparametric and Semiparametric Models*, Heidelberg: Springer.

HUMAN MORTALITY DATABASE (2005): *Life Expectancy at Birth*, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.humanmortality.de (data downloaded on Jan. 10, 2008).

JOHANSSON, E. (2004): "A Note on the Impact of Hours Worked on Mortality in OECD Countries," *European Journal of Health Economics*, 49, 335–340.

LI, Q. AND J. S. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*, Princeton: Princeton University Press.

LINDAHL, M. (2005): "Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income," *Journal of Human Resources*, 40, 144–168.

LOADER, C. (1999): *Local Regression and Likelihood*, New York: Springer.

——— (2004): "Smoothing: Local Regression Techniques," in *Handbook of Computational Statistics*, ed. by J. Gentle, W. Härdle, and Y. Mori, Springer.

LUY, M. (2006): "Differentielle Sterblichkeit: die ungleiche Verteilung der Lebenserwartung in Deutschland," *Rostock Center Discussion Paper*, No. 6.

MARTIKAINEN, P., T. MARTELIN, E. NIHTILA, K. MAJAMAA, AND S. KOSKINEN (2005): "Differences in Mortality by Marital Status in Finland from 1976 to 2000: Analyses of Changes in Marital-Status Distributions, Socio-demographic and Household Composition, and Cause of Death," *Population Studies*, 59, 99–115.

MEER, J., D. L. MILLER, AND H. S. ROSEN (2003): "Exploring the Health-Wealth Nexus," *Journal of Health Economics*, 22, 713–730.

MITTELHAMMER, R. C., G. G. JUDGE, AND D. J. MILLER (2000): *Econometric Foundations*, Cambridge: Cambridge University Press.

MOORE, D. E. AND M. D. HAYWARD (1990): "Occupational Careers and Mortality of Elderly Men," *Demography*, 27, 31–53.

PAGAN, A. AND A. ULLAH (1999): *Nonparametric Econometrics*, Cambridge: Cambridge University Press.

REIL-HELD, A. (2000): "Einkommen und Sterblichkeit in Deutschland: Leben Reiche länger?" *Beiträge zur angewandten Wirtschaftsforschung*, 580-00.

RIED, M. (1998): "Comparative Dynamic Analysis of the Full Grossman Model," *Journal of Health Economics*, 17, 383–426.

RUHM, C. J. (2000): "Are Recessions Good for Your Health?" *Quarterly Journal of Economics*, 115, 617–650.

——— (2007): "A Healthy Economy Can Break Your Heart," *Demography*, 44, 829–848.

SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

STATISTISCHES BUNDESAMT (2004): *Periodensterbetafeln für Deutschland: Allgemeine und abgekürzte Sterbetafeln von 1871/1881 bis 2001/2003*, Wiesbaden: Statistisches Bundesamt.

——— (2007): *Statistisches Jahrbuch 2007 für die Bundesrepublik Deutschland*, Wiesbaden: Statistisches Bundesamt.

VAN SOEST, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach," *Journal of Human Resources*, 30, 63–88.

VON GAUDECKER, H.-M. AND R. D. SCHOLZ (2007): "Differential Mortality by Lifetime Earnings in Germany," *Demographic Research*, 17, 83–108.

WOLFE, J. R. (1985): "A Model of Declining Health and Retirement," *Journal of Political Economy*, 93, 1258–1267.

YANG, L. AND R. TSCHERNIG (1999): "Multivariate Bandwidth Selection for Local Linear Regression," *Journal of the Royal Statistical Society. Series B*, 61, 793–815.

YATCHEW, A. (1998): "Nonparametric Regression Techniques in Economics," *Journal of Economic Literature*, 36, 669–721.

——— (2003): *Semiparametric Regression for the Applied Econometrician*, Cambridge: Cambridge University Press.

Chapter 5

On the Fairness of Early Retirement Provisions

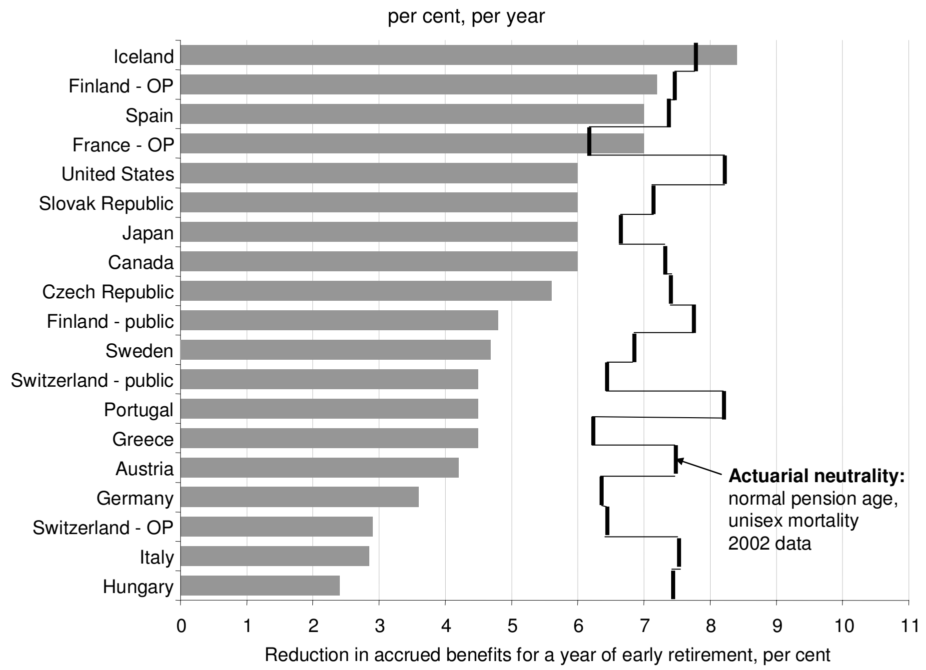
5.1 Introduction

Declining fertility and increasing longevity have rendered public pension systems in many OECD countries unsustainable and have triggered substantial reforms of these systems. One of the officially declared reform objectives is to increase the average retirement age. Crucial parameters for this endeavor are first the legal retirement age and secondly the early retirement provisions inherent in the public pension system. In Germany, e.g., legal retirement age will be gradually increased from 65 to 67 years over the period 2012 to 2029.

When this reform was enacted, there was vigorous criticism by the trade unions who claimed that for physically demanding occupations such as roofers it would be unacceptable to work beyond age 65, and it was emphasized that presently, labor force participation of males aged 60–65 is still pretty low. Of course, in a free society nobody can be forced to work. Therefore, in Germany as in any public pension system workers are allowed to retire up to five years before reaching the legal retirement age, but then their pension level is cut by 3.6 per cent per year of early retirement. Similar regulations exist in other OECD countries with discounts between 4 and 7 per cent per year in the majority of countries (see Figure 5.1). Taking these early retirement provisions into account, it is argued that the increase of the legal retirement age amounts to nothing but a cut in the level of retirement benefits by 7.2 per cent because many workers could not react to the reform by working longer but had to suffer the early retirement discount instead.

Furthermore, it is well–known that life expectancy after reaching age 60 is positively correlated with previous earnings (see, e.g. Reil-Held 2000). Thus, workers at the low end of the earnings distribution are said to be faced with Hobson's choice: either they work until legal retirement age and accept an extremely short (expected) duration of benefits or they retire as early as possible and accept the maximum cut in the benefit level. Hence it seems that by lowering early retirement discounts their plight could be eased and the extent of implicit redistribution from the poor to the rich due to the mentioned correlation could be reduced.

It is the purpose of this paper to examine whether a cut in early retirement discounts is suitable to reduce the extent of 'unfair' income redistribution in the German social security system. To this end we must first develop an appropriate notion of fairness. It turns out that in the relevant literature reviewed in Section 5.2 there are several quite different concepts of fairness and that the 'right' notion of fairness depends upon the objectives pursued in the design of pension systems, which can range from the pure efficiency goal of achieving a 'distortion–free' retirement decision to the very ambitious equity goal implicit in maximizing a social welfare function in the tradition of optimal taxation theory. In Section 5.3, we point



Source: Queisser and Whitehouse (2006)

Figure 5.1: Discounts for Early Retirement in OECD Countries

out the problems attached to both of these 'extreme' positions and propose a more modest concept of fairness called 'distributive neutrality', which is inherent in the constitution of the German pension system and which says that the rate of return on total contributions to the pension system should not depend systematically on the individual's ability. In Section 5.4, we examine the implications of this concept for the possibility to lower the extent of redistribution by changing the size of the early retirement discounts. To this end we explore the relation between lifetime earnings and the benefit–contribution ratio in a large data set provided by the German social security administration. Finally, Section 5.5 concludes.

5.2 Concepts of Fairness

5.2.1 Homogeneous Workers: Concepts of Efficiency

There is widespread agreement that social insurance systems should be so designed as to achieve a given distributive target with the least degree of distortions to individual decisions on education, labor supply, savings, and other behavior. As an example, the German Old Age Insurance system is based on a tight tax–benefit linkage called 'Teilhabe–Äquivalenz' (fairness within cohorts), a feature which is explicitly aimed at minimizing labor–supply disincentives. Such efficiency criteria are particularly important in a world of equals, in which distributional concerns play no role. However, we shall show that in the design of social security systems there is more than one possible efficiency rule.

5.2.1.1 No Distortion of Work Incentives

A straightforward target is the neutrality of the early retirement provision with respect to the labor supply decision of the worker: the pension system should not distort the choice of retirement age (Boersch-Supan 2000 and Boersch-Supan 2004). This implies that net social security wealth, i.e. the present value of all future retirement benefits minus contributions, is not changed when the worker retires one period later (or earlier). This feature of a pension system is also called 'marginal fairness'. The normative appeal of marginal fairness is strongest in a world of equals because in this case equity concerns do not play a role and thus the pure efficiency goal of an undistorted choice prevails as the single objective. Moreover, marginal fairness has unambiguous implications only when the length of remaining life is certain because only in this case can the present value of extra retirement benefits be calculated so that it exactly matches the 'pension costs' of retiring one year later.

In theory, the implications of marginal fairness are straightforward: The costs of retiring one year later are composed of the contributions paid to the pension system and the forgone benefits during the extra work year. If the length of the retirement period were known, the additional benefits could be calculated so that they exactly match this amount in present value. The discount rate to be applied in this calculation should be the 'market interest rate', preferably the rate at which workers can shift their consumption possibilities over time. In the case of a worker who already disposes of savings which he can adjust to the changing stream of pension benefits (and contributions), the interest rate on government bonds seems to be the appropriate one. Matters become more complicated for a worker who has no savings apart from his social security wealth and who does not want to change his consumption pattern when he decides to work another year. This person will want to shift consumption from the retirement period to the present period by borrowing against his pension entitlements, which would require a much higher interest rate such as the one banks charge for overdraft loans.

In the practice of the German pension system, matters are complicated by the fact that retirement benefits accrue in proportion to total earnings during working life. As a consequence, the contributions paid in an extra year of working life already translate into additional benefits, where the 'rate of return' equals the implicit rate of return of the pay–as–you–go system, viz. the growth rate of earnings, which is considerably smaller than the interest rate. To achieve marginal fairness of the total return on the sum of contributions and forgone benefits, therefore, the rate of return on the forgone benefits must be much higher than the interest rate.

5.2.1.2 Minimizing the Burden on Other Generations

Incentive compatibility may be a sensible target in a one–household economy but it becomes questionable as soon as an infinite sequence of overlapping generations is considered. A much more convincing objective for this case is the requirement that the behavior of the retiree does not place a burden on others, in particular on later generations of tax–payers. With this consideration Ohsmann et al. (2004) justify the claim that the discount rate used for making present–value calculations should equal the rate of return of the PAYG system, viz. the growth rate of earnings, g. Their reasoning says that, as any Euro paid in period t as a contribution to a PAYG–financed social security scheme yields $(1+g)$ Euros in additional retirement benefits in period $t+1$—holding everything else constant, the same should be true of an additional Euro paid or forgone due to postponing retirement by one period. Furthermore, they argue that the adjustment rate currently in place in Germany of 3.6 per cent per year comes close to this figure.

To assess the validity of this claim, we must make a distinction between two types of PAYG systems:

a) a pure PAYG system that never holds any fund balances (positive or negative) but adjusts the contribution rate instantaneously to keep total contributions and total payouts of retirement benefits in line at every moment in time,

b) a mixed system in which the pension administration were allowed to borrow and save on the capital market to smooth short–run fluctuations of the contribution rate.

In case b), additional contributions and forgone benefits of a person who postponed retirement by one period could be accumulated by the fund and used to pay out the additional claims accruing to that individual over the course of his retirement period. But then it is again the interest rate on the capital market, r, which is the appropriate rate of return. Clearly, it is questionable if such a system can be called PAYG and the procedure described here requires that 'additional' revenues due to changes in retirement age be distinguished from 'ordinary' revenues. On the other hand, it can be argued that this case is relevant for the German situation in which almost 30 per cent of all pension outlays are financed by subsidies from the federal budget. Provided that fluctuations in net revenues do not lead to changes in the contribution rate but rather adjustments of the state subsidies and indirectly of government debt, the opportunity cost of paying one Euro in period t is in fact paying $(1 + r)$ Euro in period $t + 1$.

In contrast, in a *pure* PAYG system of type a), a shift of the retirement age of a particular individual i from period t to $t + 1$, holding everything else constant, translates into a cut in the contribution rate in t but an increase in this rate in the s periods until the death of this individual. Thus it is impossible to leave all other participants in the system unaffected because it makes all contributors (workers) in period t better off and all workers in the periods up to $t + s$ worse off, so it affects participants differently according to their birth year.

Instead of the impossible target of sheltering everybody else from any consequences of individual i's behavior, a more modest target could be achieved, viz. keeping the contribution rate and the implicit taxes due to participating in the PAYG system from rising in a new steady state when *all* workers staring with a particular cohort increase their retirement age by one year. This question has been analyzed by Breyer and Kifmann (2002), and the answer is that the rate of return must not exceed the growth rate g to keep the long–run contribution rate and implicit tax rate constant. Of course, a number of cohorts in the transition period benefit from lower contribution and implicit tax rates.

5.2.2 Heterogeneous Workers: Concepts of Welfare Maximization

With inequality in initial endowments of productivity, health or life expectancy, efficiency is not the only objective in designing a pension system, and equity considerations come into play. The usual procedure chosen in the optimal taxation literature is to first propose an (Utilitarian) social welfare function and to derive a first–best allocation, and in a second step to make realistic assumptions on the observability of distinguishing characteristics and derive a second–best solution and propose a system of incentives which are suitable to bring about the second–best allocation in the presence of these informational constraints.

5.2.2.1 Heterogeneity in Productivity and Health

Cremer et al. (2004)[1] consider a world in which workers differ in two unobservable characteristics, productivity and health, whereas life expectancy is still the same for everybody. Health status is here distinguished by the rate at which disutility from working increases over the life cycle, with faster growth indicating worse health. In a first–best solution, consumption is the same for all types, but sick people are

[1] For a similar model see Sheshinski (2003).

allowed to retire earlier than healthy ones, and the differences in income are equalized using person–specific lump–sum transfers.

With asymmetric information, when productivity and health are positively correlated but unobservable and period income and retirement age are observable, the desired redistribution from the high–productivity and the healthy to the low–productivity and ill types can be accomplished by positive marginal taxes both on period income and on the length of the working life (ibid., p. 2272). By taxing longer stays in the job (i.e. subsidizing early retirement), the ill type can be induced to retire earlier whereas the healthy type, who would lose more income from retiring early, can be discouraged from mimicking the ill type and thus, by using this additional incentive, the self–selection constraint can be relaxed, which means that the tax rate on period income can be lowered. Interestingly, the same result obtains if individuals differ in either productivity or health but not both.

According to this result, generous early retirement provisions can be interpreted as some kind of disability insurance in a world in which health and thus disability can not be (perfectly) monitored. The result is the more remarkable as it is not based on any differences in life expectancy in the population.

5.2.2.2 Heterogeneity in Life Expectancy

Another potential source of inequality is life expectancy. This is particularly relevant in the context of social security systems because total retirement benefits depend as much on per–period benefits as they do on the length of the retirement period, a fact that is often overlooked in the design of these systems.

This point is taken up by Bommier et al. (2005) who assume that length of life is certain but varies across individuals. The authors consider a benevolent social planner who maximizes a utilitarian welfare function which is concave in individual utilities, which can be justified either with inequality aversion or with risk aversion with respect to length of life. If length of life were public knowledge, (first–best) welfare maximization would entail that the long–lived retire later and consume less per period than the short–lived.

When length of life is private knowledge, a typical optimal taxation situation occurs in which the social planner can only achieve a second–best optimal allocation in which various pairs of consumption and retirement age are offered in such a way that the long–lived do not benefit from mimicking the short–lived. The screening instrument proposed by the authors is a (positive or negative) 'retirement bonus' $B(z)$ which depends upon retirement age z and is added to an individual's gross earnings. The central result of the paper (ibid., p. 14) states that when disutility from work is linear in the length of the working–life, then $B'(z) < 0$, i.e. the retirement bonus is falling in retirement age, which means that there is an implicit tax on working more years. The intuition behind the result is that the desired redistribution from the long–lived to the short–lived can be accomplished by taxing continued activity because the long lived have a stronger demand for retirement consumption and therefore more incentives to work longer.

5.3 Fairness when Income and Life Expectancy are Correlated: The Concept of Distributive Neutrality

The concepts of pure efficiency discussed in Section 5.2.1 are not appropriate in a world of heterogeneous individuals. On the other hand, the welfare criteria used in the approaches described in Section 5.2.2 are based on highly controversial normative foundations. First, individual utilities must be assumed to be measurable on a cardinal scale and interpersonally comparable. Secondly, a specific functional form of the social welfare function must be given. Finally, specific policy implications can only be derived if the functional form of the individual utility functions is given as well. Thus while these approaches are useful in uncovering the relationship between certain widely held value judgments concerning equity and the general design of social security systems, more specific implications on the size of adjustment rates for early retirement can not be expected from these exercises.

Therefore, in the following we shall propose a more modest concept of 'fairness' of social security systems, which is consistent with the usual concept of fairness as distributive neutrality and has the advantage of giving rise to specific propositions on the 'fair' size of early retirement discounts.

The principle of 'Teilhabe–Äquivalenz' underlying the design of the German social security system is based on the general notion of distributive neutrality: within a cohort, the expected retirement benefits shall be proportional to total contributions paid over the working life. The specific way in which this principle is implemented, however, consists in making per period retirement benefits proportional to total contributions, disregarding the length of the benefit spell. This is innocuous as long as there is no systematic variation in life expectancy across social groups. However, it becomes highly questionable when life expectancy is positively correlated with income, education and other indicators of social status (Breyer 1997), and there is ample evidence from many countries that this correlation indeed exists (for Germany, see, e.g., Reil-Held 2000, von Gaudecker and Scholz 2007).

Given these observations, we postulate the following 'fairness' criterion:

Definition: *'Distributive neutrality' is satisfied in a social security system if the ratio between total benefits and total contributions does not vary systematically with average annual earnings.*

This criterion is modest insofar as it does not advocate a specific equity norm, but only reformulates the principle of 'Teilhabe–Äquivalenz' in such a way as to leave room for taking certain well–established empirical relationships between longevity, wages and retirement age into account.

5.4 Distributive Neutrality and Early–Retirement Discounts in the German Pension System

5.4.1 Theory

In this Section we shall examine whether the ideal of distributive neutrality among different ability groups in the population is met in the German public pension system and, if not, whether early–retirement discounts can be an instrument to achieve this goal. Obviously, for this to be possible, a number of restrictive assumptions must be met. In particular, the processes governing retirement and death must be 'laws of nature' linking ability to work and life expectancy to labor productivity, measured by wages. In other words, we assume that the separation from the labor force is not a conscious decision following a rational trade–off between cost and benefits of retirement but is dictated e.g. by insufficient health.[2]

We first calculate the benefit–contribution ratio of an individual i with labor productivity a_i who retires at age E_i and dies at age L_i as follows. Let E^0 be the age at which a worker becomes eligible to early retirement without taking any health related contingencies into account. After this date, potential future contributions and benefits are discounted with the real interest rate ρ. At age E^0, his accumulated life time income is denoted Y_i^0. According to the benefit formula valid in this system, annual benefits B_i are proportional to his total (taxable) lifetime income, Y_i, and are subject to a discount rate x for every year of retiring earlier than at age 65. Therefore, if they are discounted to Age E^0, they are given by

$$B_i \quad = \quad bY_i\left[1 - x(65 - E_i)\right] \int\limits_{E_i}^{L_i} e^{-\rho(t-E^0)}\mathrm{d}t. \qquad (5.1)$$

[2]Wolfe (1983) and Hurd et al. (2002) find that the lower an individual's life expectancy, the earlier he or she will retire.

On the other hand, total contributions C_i are proportional to lifetime income and consist of two parts: those contributions which were paid before age E^0 and which are proportional to total income up to this age, Y_i^0, and the discounted value of future contributions up to the chosen retirement age E_i,

$$C_i \;=\; c\left[Y_i^0 + a_i \int\limits_{E^0}^{E_i} e^{-\rho(t-E^0)}\mathrm{d}t\right], \tag{5.2}$$

where c denotes the contribution rate. We do not discount previous contributions for two reasons. First, this is consistent with German pension law, which treats all contributions equally, no matter when they were paid; and secondly we can not observe the time–path of contributions but only the sum, so we could not implement discounting in our data set.

Hence the ratio of total benefits and total contributionsfor individual i, r_i, is determined by

$$\begin{aligned} r_i &= B_i/C_i \\ &= \frac{bY_i\left[1 - x(65 - E_i)\right] e^{-\rho[E_i - E^0 + L_i]}\left(e^{\rho E_i} - e^{\rho L_i}\right)}{c\left[a_i\left(e^{-\rho(E_i - E^0)} - 1\right) - \rho Y_i^0\right]}. \end{aligned} \tag{5.3}$$

Distributive neutrality is then satisfied if there is no systematic (monotonous) relationship between the benefit–contribution ratio r and ability a, while the system is redistributive in a regressive (progressive) way if r is an increasing (decreasing) function of a. Observe that by Equation (5.3), the relation between r and a depends upon the value of the early–retirement discount rate x. Now, to examine the relationship between r and a empirically, we can choose between two different estimation strategies, which we will call

a) the indirect method and

b) the direct method.

The indirect method is based on postulating the existence of systematic relationships $E(a)$, $L(a)$, $Y(a)$, and $Y^0(a)$, and estimating these four functions using our data to be described below. Inserting the estimated functions into Equation (5.3), we can then synthetically construct the relationship $r(a|x)$. In contrast, the direct method consists in estimating the function $r(a|x)$ directly using (5.3) and then regressing it on a.

Preliminary tests show that the direct method gives slightly more reliable results because the goodness of fit of the regression equations for $E(a)$ and $L(a)$ turns out to be fairly low, despite the significance of a. In the following we thus present only the results generated by the direct method of estimation.

5.4.2 Empirical Estimation

5.4.2.1 Data

The variables used in this analysis are taken from a data set on pension discontinuations from 1994 to 2005, FDZ-RV (2007), published by the Federation of German Pension Insurance Institutes (Deutsche Rentenversicherung Bund). It contains a 10% sample of all discontinued public pensions due to the death of the beneficiary, which amounts to roughly 828,000 observations. However, each observation corresponds to a pension, and not to an individual retiree, who can (subsequently or even simultaneously) benefit from more than one pension. Taking this into account, we are left with a sample of 209,752 pensioners whose benefits are based on own contributions.[3] The most important variables are the sum of pension benefit claims (in points), the length of the work life, the retirement age, and the age at death. From the first two variables we construct the average claims earned per year of work. One point corresponds to contributions based on one year of the average annual income. Other variables which are contained in the data set have to be taken with care—they are only reliable when they have been used for the calculation of benefits, otherwise they are either unreliable or missing. See Table 5.1 for descriptive statistics of the variables used.

5.4.2.2 Weighting Function

Our sample suffers from a selection bias. Since we observe a death cohort, life expectancies are biased downwards. In each death cohort, a large variety of birth cohorts are included, and we know that life expectancy has been increasing with the year of birth.[4] However, this increase is only partially taken into account in the sample, as especially individuals from younger birth cohorts (whose ex ante life expectancy should be higher) only appear in the sample if they died relatively young. Ideally, we would like to observe a birth cohort of which all individuals have already died; obviously, this is only possible for very old birth cohorts (born around 1900) in order to get unbiased estimates. However, as life expectancy has

[3]This number already takes also into account that we excluded observations with missing values in the variables of importance for our analysis.

[4]See Chapter 4 for a detailed treatment on the problem of the present selection bias and its solution.

Descriptive Statistics

	all obs.		restr.		restr., weighted	
	mean	st. dev.	mean	st. dev.	mean	st. dev.
retirement age E	58.54	7.53	58.62	5.44	59.07	5.35
age at death L	66.77	8.98	66.48	7.18	67.40	7.13
total points Y	33.47	17.67	44.14	13.23	44.53	13.51
— till $E_0 = 60$, Y_0	37.10	27.08	45.95	13.86	45.80	14.00
points per year a	1.10	.73	1.13	.31	1.13	.31
sex = female	31.85%	—	0%	—	0%	—

Based on FDZ-RV (2007). With all observations, $n = 209,752$. Restricted to male observations with at least 25 years of contributions, $n = 112,369$.

Table 5.1: Descriptive Statistics

been increasing over time, these very early birth cohorts may not be representative for more recent cohorts and therefore not suitable for drawing policy conclusions.

Our approach to correcting the selection bias is the following. The selection that occurs is not based on individual decision making—it is solely a matter of data selection. Among the later birth cohorts, deaths at young age are over–represented. The relationship is empirically linear (which corresponds to the usually perceived increase of life expectancies)[5], so a linear weighting function, which decreases with the birth year, can correct this bias. However, ex ante we cannot be sure about the slope of weighting function; we only know that the weights have to be linear and non–negative over the whole support. The parameter of choice is therefore only the slope, while the intercept serves as a normalizing constant that limits the range of the potential slopes in order to ensure the non–negativity constraint. If GBJ denotes the year of birth (normalized to zero for the earliest birth cohort), the weighting function w takes the following form, with s being the slope parameter:

$$w(GBJ) \quad = \quad 1 - s \cdot GBJ \tag{5.4}$$

With the intercept set to one, s can vary between zero (hence, a weight of one for all birth cohorts) and .0103, which just ensures that the weight for the latest birth cohort is still positive. The selection criterion for our choice of the slope parameter remains to be determined. We select the weighting function which minimizes the difference between the weighted average life expectancy in our data and the ex-

[5] See e.g.Statistisches Bundesamt (2007, p. 54) and Human Mortality Database (2005), and own calculations.

ogenously known life expectancy. Yet, the maximum average age at death obtained with this method (i.e. the one which results from the steepest weighting function) is still lower than the value of life expectancy observed in population statistics.[6]

5.4.2.3 Data Requirements

For the construction of our variable 'ability', we have to observe the years of contribution. Due to a change in legislation, the sample lacks this variable for every pensioner who retired into the old–age pension before 1992. In 1992, the calculation of pension benefits changed, one reason was the introduction of early retirement discounts, which had not been implemented before. Also before 1992, benefits and the application for a pension were contingent on the years the applicant contributed to the pension system, however, we are unable to retrieve this data.

This phenomenon aggravates the selection bias we introduce above: The earliest possible retirement age into an old–age pension in 1992 was 60 for women (and under certain health–related contingencies) and 63 for men. The latest death cohort we observe is the one of 2005, which delimits by construction realized life expectancy of the pensioners in our sample. This explains the relatively low average life expectancy we observe, although we impose a weighting scheme. Despite this bias, we continue with our analysis. This is justified on the grounds of the preceding Chapters; the bias affects as well duration under the benefit spell (Chapter 3) and life expectancy (Chapter 4) only in their level, the shape of the respective relationships is fairly robust against the selection.

We leave a further possibility of dissolving this bias to future research: A related data set records living pensioners in the years of our death cohort. For each retirement cohort, we could principally estimate the relationship between the total sum of benefit claims and years of contribution. Utilizing this relationship, we can infer from observed benefit claims on the years of contribution which were necessary to collect these claims in order to replace the missing values. Finally, we can either compare the so–constructed sample with the one we originally observe and conclude whether the results are robust, or we can base the analysis on the constructed sample in the first place.

5.4.2.4 Regression Results

In principle, more than one definition of retirement and therefore of the benefit spell can be distinguished. Our variable retirement age E is the age of the first receipt of any pension based on own contributions, which can be the old–age pension,

[6]Notice however that the concept of life expectancy in a given year always refers to age–specific death rates of this year and not to the average age at death of the death cohort of this very year.

but also disability pensions. This notion is in line with our theoretical approach because it takes all paths into retirement including disability pensions into account. Insofar as claiming disability benefits carries some information on the innate ability (including the health capital) of the individual, this is certainly the superior concept compared to the alternative of taking the first receipt of an old–age pension as the age of retirement.

Furthermore, the following procedures were performed with the data. First, observations on women were excluded. Since ability (or the earnings capacity) cannot be observed directly, it has to be ensured that the average benefit claims are a good proxy. In the simplest case, namely when an individual has worked during his whole career and contributed to the public pension system, benefit claims are a linear transformation of income.[7] This even holds if the individual under observations had longer times of education before starting to work or if he or she raised children. The measure is then only slightly diluted, as claims are increased by these activities in order to compensate for the loss of regular contributions. The close relationship between total income and benefit claims, however, is not guaranteed once the individual has been self–employed or has worked as a civil servant for some time in his career. During these times, usually no contributions are paid, as membership in the public pension system is not mandatory (or even possible) anymore. We therefore restrict our sample to male pensioners who worked at least 25 years in a job where contributions are mandatory. This sample contains 112,369 observations. Our results differ compared to the ones using the whole sample, but are robust with respect to the exact choice of the number of years required.

In this data set we do not observe the value of Y^0, which we can construct by

$$Y_i^0 \quad = \quad Y_i - a_i(E_i - E^0). \tag{5.5}$$

The ability variable a lies in the interval $(0, 2.8]$. The upper bound is higher than what could have been achieved by contributions based on work only, in which case we had $a^{\max} = 2.15$. However, we cannot unambiguously distinguish between claims earned because of own work or because of times of education, parenting, and other reasons, which slightly augments the average claims per year of work.

To estimate the relationship between the ration of benefits to contributions and ability for a given value of the discount x, we assume a polynomial of degree five to account for possible non–linearity and non–monotinicity:

[7] Up to a certain income, beyond which contributions (and therefore claims) are capped. The maximum contributions are based (in 2006) on a monthly gross income of EUR 5250 and are adjusted on a yearly basis.

$$r = \rho_0 + \rho_1 a + \rho_2 a^2 + \rho_3 a^3 + \rho_4 a^4 + \rho_5 a^5 + \mu \quad (5.6)$$

See Table 5.2 for the estimation results and Figure 5.2 for the respective plotted functions, with discounts ranging from $x = .00$ to $x = .08$, where the fitted ratio–functions $r(a)$ are ordered from top to bottom with with respect to x, and we find the ratio with the smallest early retirement discount at the top. At current discounts of $x = .036$, the benefit–contribution ratio is an increasing function over the bigger part of a, hence we find a regressive effect of the pension system.

Estimation Results—Dep. Var.: Benefit–Contribution Ratio

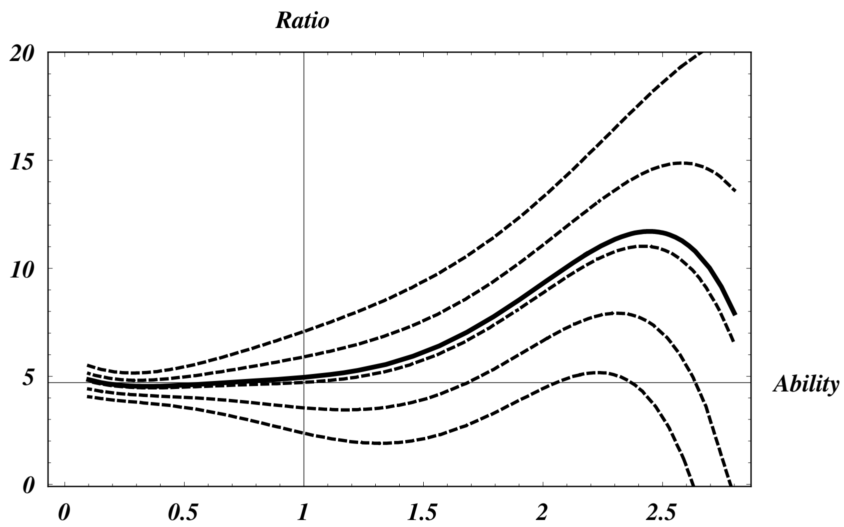
discounts	const.	a	a^2	a^3	a^4	a^5	R^2
$x = .00$	6.04***	-7.23**	18.28**	-15.55*	6.49*	-.96	.003
	(.24)	*(2.96)*	*(7.80)*	*(8.33)*	*(3.95)*	*(.68)*	
$x = .02$	5.62***	-6.33***	16.37***	-16.14***	7.65***	-1.27***	.003
	(.14)	*(1.37)*	*(3.63)*	*(3.90)*	*(1.85)*	*(.32)*	
$x = .036$	5.28***	-5.62***	14.85***	-16.61***	8.57***	-1.51***	.003
	(.15)	*(1.64)*	*(4.30)*	*(4.56)*	*(2.15)*	*(.37)*	
$x = .04$	5.19***	-5.44***	14.46***	-16.73***	8.80***	-1.57***	.002
	(.16)	*(1.94*	*(5.08)*	*(5.39)*	*(2.53)*	*(.16)*	
$x = .06$	4.76***	-4.55	12.56	-17.31*	9.96**	-1.88**	>.000
	(.29)	*(3.79)*	*(9.93)*	*(10.53)*	*(4.95)*	*(.85*	
$x = .0694$	4.56***	-4.13	11.66	-17.59	10.50*	-2.02*	>.000
	(.36)	*(4.73)*	*(12.39)*	*(13.14)*	*(6.19)*	*(1.06)*	
$x = .08$	4.33***	-3.66	10.65	-17.90	11.11	-2.18*	>.000
	(.44)	*(5.80)*	*(15.22)*	*(16.14)*	*(7.60)*	*(1.30)*	

Data set includes only male observations with at least 25 years of own contributions, $n = 112,369$. *** denotes significance on the .99 level, ** on the .95 level, and * on the .90 level. Robust standard errors in parenthesis.

Table 5.2: Direct Estimation Results

5.4.2.5 Achieving Distributional Neutrality

If we want to apply our criterion of distributional neutrality to the different $r(a|x)$ functions, given that x is a constant, the return functions have to be linearized. We then compare the linear return functions $r_{\text{lin}}(a|x)$ with respect to their slope parameter and choose the discounts x that minimize the absolute value of this slope. As the method of linearization we choose least squares, because it inherently takes the distribution of ability a into account. By this method, we fit straight lines to

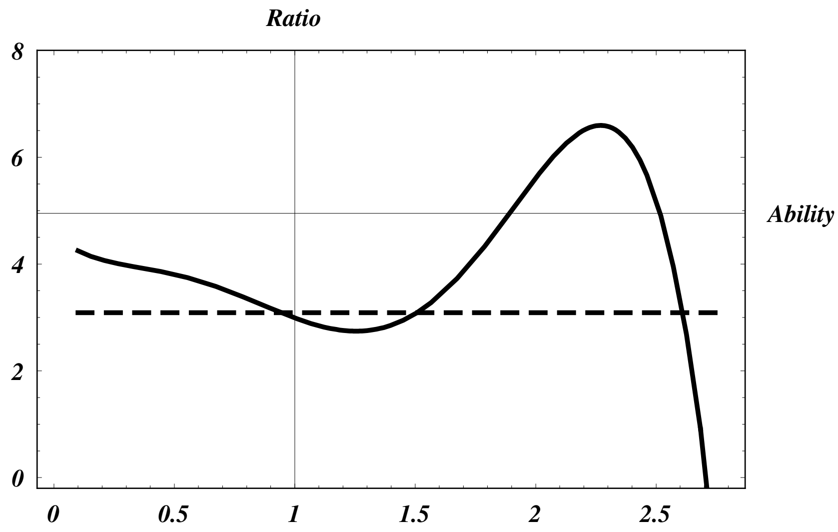


Solid: $x = .036$. Dashed: Ordered from top to bottom with $x = .00$ at the top and $x = .08$ at the bottom, $\Delta = .02$. Axes drawn at $a = 1$ and $r(1|x = .036)$.

Figure 5.2: Direct Estimation of the Benefit–Contribution–Ratio as a Function of Ability

the return functions shown in Figure 5.2 (the direct estimates of r) based on different discount rates. The discount rate that minimizes the slope of r_{lin} turns out to be .0694. See Figure 5.3, where the original return function $r(a|x = .0694)$ is compared to the resulting linearized form. In general, the slope of the linearized forms is decreasing over the choice of discounts. These higher discounts flatten the benefit–contribution ratio, but at the same time, lower the average r as compared to the current discounts of $x = .036$.

Hence, we find that the current discounts of $x = .036$ are too low to achieve distributional neutrality. The main reason for this result is the negative relationship between ability and retirement age. However, this might be a consequence of the data set we use. The individuals under observation died between 1994 and 2005. With an average benefit duration of approximately 8 years, many retired between 1986 and 1997, a period in which the federal government allowed the rather excessive use of early retirement schemes. Additionally, these early retirement schemes were offered mainly by large companies, which are known to pay higher wages for the same level of qualification. So our measure a does not only capture ability, but also differences in firm size, economic sector etc., and along these dimensions possibilities to retire early differed for (otherwise equal) individual workers. We therefore propose to see our results as an exemplary application of a method to achieve distributional neutrality within the public pension system, whereas actual policy advice should be based on more recent data, which allows inferring on the



Benefit–Contribution Ratio and Linearized Form, at $x = .0694$. Dashed: Linearized Benefit–Contribution Ratio.

Figure 5.3: Neutralizing a Linearized Benefit–Contribution–Ratio with Adequate Discounts

behavior of future retirees.

Despite the very low R^2 we already observe in the estimations, we can interpret distributional neutrality also statistically; a regression of r at neutral discounts of 6.94% on a alone (without adding a higher polynomial in a) yields an R^2 of zero, and the impact of a is not only zero with respect to its size, but also not significantly different from zero as well.

5.5 Concluding Remarks

In this paper we discussed several notions of 'fairness' of early retirement provisions in pay–as–you–go financed public pension systems. We advanced the thesis that the 'right' notion of fairness depends upon the objectives pursued in the design of pension systems, which can range from the pure efficiency goal of achieving a 'distortion–free' retirement decision to the very ambitious equity goal implicit in maximizing a social welfare function in the tradition of optimal taxation theory. We pointed out the problems attached to both of these 'extreme' positions and proposed a more modest concept of equity, called 'distributive neutrality', which is based on the notion that the rate of return on total contributions to the pension system should not depend systematically on the individual's ability.

By applying this concept to the German retirement benefit formula and taking empirically estimated relationships between average annual income (as a proxy for ability), life expectancy and retirement age into account, we were able to calculate the relationship between average annual income and the benefit–contribution ratio which is increasing over a wide range of parameter values. Thus distributive neutrality is presently violated but instead there is systematic redistribution in favor of high–ability persons. As this group is not only enjoying higher life expectancy but—at least according to our data—also retires earlier, lowering early–retirement discounts, as e. g. proposed by Sheshinski (2003), would in this case exacerbate this redistribution.

It should be emphasized that our empirical approach is based on the unrealistic assumption that the choice of retirement age is not already affected by the existing early–retirement discounts. If this were indeed the case, as could be expected, we would have to replace the implicitly assumed $E(a)$ function by a relationship of the form $E(a;x)$. The present data set does not allow estimating such a function as the discounts were phased–in gradually and thus a corresponding variable would be perfectly correlated with a time trend. Moreover, different groups of persons were subject to different values of x, but we did not have this information.

References of Chapter 5

BOERSCH-SUPAN, A. (2000): “A Model under Siege: A Case Study of the German Retirement Insurance System,” *Economic Journal*, 110, 24–45.

——— (2004): “Faire Abschläge in der gesetzlichen Rentenversicherung,” *Sozialer Fortschritt*, 53, 258–261.

BOMMIER, A., M.-L. LEROUX, AND J.-M. LOZACHMEUR (2005): “Varying Life Expectancy and Social Security,” *mimeo*.

BREYER, F. (1997): “Sind äquivalente Renten fair?” in *Reform des Sozialstaats I*, ed. by R. Hauser, Berlin, 169–180.

BREYER, F. AND M. KIFMANN (2002): “Incentives to retire later – a solution to the social security crisis?” *Journal of Pension Economics and Finance*, 1, 111–130.

CREMER, H., J.-M. LOZACHMEUR, AND P. PESTIEAU (2004): “Social Security, Retirement Age and Optimal Income Taxation,” *Journal of Public Economics*, 88, 2259–2281.

FDZ-RV (2007): “Scientific Use File Demografiedatensatz Rentenwegfall 1993-2005, SUFRTWFjjXVSTDemo,” Data set from Deutsche Rentenversicherung Bund.

HUMAN MORTALITY DATABASE (2005): *Life Expectancy at Birth*, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.humanmortality.de (data downloaded on Jan. 10, 2008).

HURD, M., J. SMITH, AND J. ZISSIMOPOULOS (2002): “The Effects of Subjective Survival on Retirement and Social Security Claiming,” *NBER Working Paper*, No. 9140.

OHSMANN, S., U. STOLZ, AND R. THIEDE (2004): “Rentenabschläge bei vorzeitigem Rentenbeginn: Was ist versicherungsmathematisch fair?” *Sozialer Fortschritt*, 53, 267–271.

QUEISSER, M. AND E. WHITEHOUSE (2006): "Neutral or Fair? Actuarial Concepts and Pension-System Design," *OECD Social, Employment and Migration Working Papers*, No. 40.

REIL-HELD, A. (2000): "Einkommen und Sterblichkeit in Deutschland: Leben Reiche länger?" *Beiträge zur angewandten Wirtschaftsforschung*, 580-00.

SHESHINSKI, E. (2003): "Optimum Delayed Retirement Credit," *CESifo Working Paper*, No. 889.

STATISTISCHES BUNDESAMT (2007): *Statistisches Jahrbuch 2007 für die Bundesrepublik Deutschland*, Wiesbaden: Statistisches Bundesamt.

VON GAUDECKER, H.-M. AND R. D. SCHOLZ (2007): "Differential Mortality by Lifetime Earnings in Germany," *Demographic Research*, 17, 83–108.

WOLFE, J. R. (1983): "Perceived Longevity and Early Retirement," *Review of Economics and Statistics*, 65, 544–551.

Complete References

ADAMS, P., M. D. HURD, D. MCFADDEN, A. MERRILL, AND T. RIBEIRO (2003): "Healthy, Wealthy and Wise? Tests for Direct Causal Paths Between Health and Socioeconomic Status," *Journal of Econometrics*, 112, 3–56.

ADDA, J., T. CHANDOLA, AND M. MARMOT (2003): "Socio–Economic Status and Health: Causality and Pathways," *Journal of Econometrics*, 112, 57–63.

ANGRIST, J. AND V. LAVY (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 114, 533–575.

ANTOLIN, P. AND S. SCARPETTA (1998): "Microeconometric Analysis of the Retirement Decision: Germany," *OECD Economics Department Working Papers*, No. 204.

ATTANASIO, O. P. AND C. EMMERSON (2003): "Mortality, Health Status, and Wealth," *Journal of the European Economic Association*, 1, 821–850.

BAKER, M. AND D. BENJAMIN (1999): "Early Retirement Provisions and the Labor Force Behavior of Older Men: Evidence from Canada," *Journal of Labor Economics*, 17, 724–756.

BANKS, J., M. MARMOT, Z. OLDFIELD, AND J. P. SMITH (2006): "The SES Health Gradient on Both Sides of the Atlantic," *NBER Working Paper*, No. 12674.

BARNAY, T. (2007): "Redistributive Impact of Differential Mortality in the French Pay-as-You-Go System," *Geneva Papers on Risk and Insurance: Issues and Practice*, 32, 570–582.

BECKMANN, K. (2000): "A Note on the Tax Rate Implicit in Contributions to Pay–As–You–Go Public Pension System," *FinanzArchiv*, 57, 63–76.

BERKEL, B. AND A. BOERSCH-SUPAN (2004): "Pension Reform in Germany: The Impact on Retirement Decisions," *FinanzArchiv*, 60, 393–421.

BLOOM, D. E., D. CANNING, AND M. MOORE (2005): "The Effect of Improvements in Health and Longevity on Optimal Retirement and Saving," *Harvard Initiative of Global Health Working Paper Series*, No. 2.

——— (2007): "A Theory of Retirement," *NBER Working Paper*, No. 13630.

BLOOM, D. E., D. CANNING, M. MOORE, AND Y. SONG (2006): "The Effect of Subjective Survival Probabilities on Retirement and Wealth in the United States," *NBER Working Paper*, No. 12688.

BLUNDELL, R. AND A. DUNCAN (1998): "Kernel Regression in Empirical Microeconomics," *Journal of Human Resources*, 33, 62–87.

BLUNDELL, R., A. DUNCAN, AND K. PENDAKUR (1998): "Semiparametric Estimation and Consumer Demand," *Journal of Applied Econometrics*, 13, 435–461.

BOERSCH-SUPAN, A. (2000): "A Model under Siege: A Case Study of the German Retirement Insurance System," *Economic Journal*, 110, 24–45.

——— (2004): "Faire Abschläge in der gesetzlichen Rentenversicherung," *Sozialer Fortschritt*, 53, 258–261.

BOERSCH-SUPAN, A. AND H. JÜRGES (2006): "Early Retirement, Social Security and Well-Being in Germany," *NBER Working Paper*, No. 12303.

BOLTON, P. AND M. DEWATRIPONT (2005): *Contract Theory*, Cambridge, MA.: The MIT Press.

BOMMIER, A., M.-L. LEROUX, AND J.-M. LOZACHMEUR (2005): "Varying Life Expectancy and Social Security," *mimeo*.

BORCK, R. (2007): "On the Choice of Public Pensions When Income and Life Expectancy Are Correlated," *Journal of Public Economic Theory*, 9, 711–725.

BOVENBERG, A. L. AND P. B. SORENSEN (2006): "Optimal Taxation and Social Insurance in a Lifetime Perspective," *CESIfo Working Paper*, No. 1690.

BRATBERG, E., T. H. HOLMAS, AND O. THOGERSEN (2004): "Assessing the Effects of an Early Retirement Program," *Journal of Population Economics*, 17, 387–408.

BREYER, F. (1997): "Sind äquivalente Renten fair?" in *Reform des Sozialstaats I*, ed. by R. Hauser, Berlin, 169–180.

BREYER, F. AND M. KIFMANN (2002): "Incentives to retire later – a solution to the social security crisis?" *Journal of Pension Economics and Finance*, 1, 111–130.

——— (2004): "The German Retirement Benefit Formula: Drawbacks and Alternatives," *FinanzArchiv*, 60, 63–82.

BREYER, F., P. ZWEIFEL, AND M. KIFMANN (2005): *Gesundheitsökonomik*, Berlin: Springer.

BROWN, J. R. (2003): "Redistribution and Insurance: Mandatory Annuitization with Mortality Heterogeneity," *Journal of Risk and Insurance*, 70, 17–41.

——— (2007): "Rational and Behavioral Persepctives on the Role of Annuities in Retirement Planning," *NBER Working Paper*, No. 13537.

BROWNING, E. K. (1975): "Why the Social Insurance Budget Is Too Large in a Democracy," *Economic Inquiry*, 13, 373–388.

BUTRICA, B. A., R. W. JOHNSON, K. E. SMITH, AND C. E. STEUERLE (2006): "The Implicit Tax on Work at Older Ages," *National Tax Journal*, LVIV, 211–234.

CASE, A., D. LUBOTSKY, AND C. PAXSON (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92, 1308–1334.

——— (2007): "The Income Gradient in Children's Health: A Comment on Currie, Shields and Wheatley Price," *NBER Working Paper*, No. 13495.

CHANG, F.-R. (1996): "Uncertainty and Investment in Health," *Journal of Health Economics*, 15, 369–376.

CLARK, G. (2007): *A Farewell to Alms*, Princeton: Princeton University Press.

CLEVELAND, W. S. (1974): "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

CREEDY, J., R. DISNEY, AND E. WHITEHOUSE (1993): "The Earnings-Related State Pension, Indexation and Lifetime Redistribution in the U.K." *Review of Income and Wealth*, 39, 257–278.

CREMER, H., J.-M. LOZACHMEUR, AND P. PESTIEAU (2004): "Social Security, Retirement Age and Optimal Income Taxation," *Journal of Public Economics*, 88, 2259–2281.

——— (2006a): "Retirement Age and Health Expenditures," *Annales d'Economie et de Statistique*, 167–186.

——— (2006b): "Social Security and Retirement Decision: A Positive and Normative Approach," *CEPR Discussion Paper*, No. 5542.

——— (2007): "Disability Testing and Retirement," *B.E. Journals in Economic Analysis and Policy: Contributions to Economic Analysis and Policy*, 7, 1–32.

CREMER, H. AND P. PESTIEAU (1996): "Redistributive Taxation and Social Insurance," *International Tax and Public Finance*, 3, 281–295.

CURRIE, A., A. SHIELDS, AND W. PRICE (2007): "The ChildHealth/Family Income Gradient: Evidence from England." *Journal of Health Economics*, 26, 213–232.

CURRIE, J. AND M. STABILE (2003): "Socioeconomic Status and Health: Why is the Relationship Stronger for Older Children?" *American Economic Review*, 93, 1813–1823.

CUTLER, D., A. DEATON, AND A. LLERAS-MUNEY (2006): "The Determinants of Mortality," *Journal of Economic Perspectives*, 20, 97–120.

DARDANONI, V. AND A. WAGSTAFF (1987): "Uncertainty, Inequalities in Health and the Demand for Health," *Journal of Health Economics*, 6, 283–290.

DATAR, A., A. GHOSH, AND N. SOOD (2007): "Mortality Risks, Health Endowments, and Parental Investments in Infancy: Evidence from Rural India," *NBER Working Paper*, No. 13649.

DAVIDOFF, T., J. R. BROWN, AND P. A. DIAMOND (2005): "Annuities and Individual Welfare," *American Economic Review*, 95, 1573–1590.

DAVIES, J. B. AND P. KUHN (1992): "Social Security, Longevity, and Moral Hazard," *Journal of Public Economics*, 49, 91–106.

DEATON, A. (2006): "Global Patterns of Income and Health: Facts, Interpretations, and Policies," *NBER Working Paper*, No. 12735.

DEATON, A. AND C. PAXSON (2004): "Mortality, Education, Income, and Inequality among American Cohorts," *NBER Working Paper*, No. 7140.

DENTON, F. T., C. H. FEAVER, AND B. G. SPENCER (2005): "Time Series Analysis and Stochastic Forecasting: An Econometric Study of Mortality and Life Expectancy," *Journal of Population Economics*, 18, 203–227.

DEUTSCHE RENTENVERSICHERUNG BUND (2005): "Rentenversicherung in Zeitreihen," *DRV Schriften*, 22.

DIAMOND, P. A. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95.

——— (2003): *Taxation, Incomplete Markets, and Social Security*, Cambridge, MA.: The MIT Press.

——— (2004): "Social Security," *American Economic Review*, 94, 1–24.

DIAMOND, P. A. AND J. A. MIRRLEES (1971a): "Optimal Taxation and Public Production I: Production Efficiency," *American Economic Review*, 61, 8–27.

——— (1971b): "Optimal Taxation and Public Production II: Tax Rules," *American Economic Review*, 61, 261–278.

——— (1978): "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics*, 10, 295–336.

DINARDO, J. AND J. L. TOBIAS (2001): "Nonparametric Density and Regression Estimation," *Journal of Economic Perspectives*, 15, 11–28.

DORN, D. AND A. SOUSA-POZA (2005): "Early Retirement: Free Choice or Forced Decision?" *CESIfo Working Paper*, No. 1542.

DUGGAN, J. E., R. DILLINGHAM, AND J. S. GREENLEES (2007): "Mortality and Lifetime Income: Evidence from U.S. Social Security Records," *IMF Working Paper*, No. 07/15.

EHRLICH, I. (2000): "Uncertain Lifetime, Life Protection, and the Value of Life Saving," *Journal of Health Economics*, 19, 341–67.

——— (2001): "Erratum to 'Uncertain Lifetime, Life Protection, and the Value of Life Saving'," *Journal of Health Economics*, 20, 459–460.

EHRLICH, I. AND H. CHUMA (1990): "A Model of the Demand for Longevity and the Value of Life Extension," *Journal of Political Economy*, 98, 761–782.

ENGLE, R. F., C. W. J. GRANGER, J. RICE, AND A. WEISS (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310–320.

EROSA, A. AND M. GERVAIS (2002): "Optimal Taxation in Life-Cycle Economies," *Journal of Economic Theory*, 105, 338–369.

ETTNER, S. L. (1996): "New Evidence on the Relationshio Between Income and Health," *Journal of Health Economics*, 15, 67–85.

EUWALS, R., D. VAN VUUREN, AND R. WOLTHOFF (2006): "Early Retirement Behavior in the Netherlands—Evidence from a Policy Reform," *CEPR Discussion Paper*, No. 5596.

FAN, J. (1992): "Design Adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 87, 998–1004.

——— (2000): "Prospects of Nonparametric Modeling," *Journal of the American Statistical Association*, 95, 1296–1300.

FAN, J. AND I. GIJBELS (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics*, 20, 2008–2036.

——— (2003): *Local Polynomial Modelling and Its Applications*, Boca Raton: CRC Press.

FAN, Y., D. LI, AND Q. LI (2004): "Nonlinearity in Medical Expenditures: A New Semiparametric Approach," *Applied Economics*, 36, 911–916.

FDZ-RV (2007): "Scientific Use File Demografiedatensatz Rentenwegfall 1993-2005, SUFRTWFjjXVSTDemo," Data set from Deutsche Rentenversicherung Bund.

FEHR, H. AND C. HABERMANN (2007): "Welfare Effects of Life Annuities: Some Clarifications," *Economics Letters (forthcoming)*.

FELDSTEIN, M. S. (1974): "Social Security, Induced Retirement, and Aggregate Capital Accumulation," *Journal of Political Economy*, 82, 905–926.

FELIPE, A., M. GUILLEN, AND J. P. NIELSEN (2001): "Longevity Studies Based on Kernel Hazard Estimation," *Insurance: Mathematics and Economics*, 28, 191–204.

FENGE, R., S. UEBELMESSER, AND M. WERDING (2006): "On the Optimum Timimg of Implicit Social Secrity Taxes Over the Life Cycle," *FinanzArchiv*, 62, 68–107.

FERREIRA, P. C. AND S. D. A. PESSOA (2007): "The Effects of Longevity and Distortions on Education and Retirement," *Review of Economic Dynamics*, 10, 472–493.

FLEISCHHAUER, K.-J. (2007): "The Effects of Pension Reform on Retirement and Human Capital Formation," *University of St. Gallen Discussion Paper*, 2007–03.

FLORENS, J.-P. (2003): "Some Technical Issues in Defining Causality," *Journal of Econometrics*, 112, 127–128.

FRIJTERS, P., J. P. HAISKEN-DENEW, AND M. A. SHIELDS (2005): "The Causal Effect of Income on Health: Evidence from German Reunification," *Journal of Health Economics*, 24, 997–1017.

GERDTHAM, U.-G. AND M. JOHANNESSON (2003): "A Note on the Effect of Unemployment on Mortality," *Journal of Health Economics*, 22, 505–518.

——— (2004): "Absolute Income, Relative Income, Income Inequality, and Mortality," *Journal of Human Resources*, 39, 228–247.

GEWEKE, J. (2003): "Econometric Issues in Using the AHEAD Panel," *Journal of Econometrics*, 112, 115–120.

GORSKI, M., T. KRIEGER, AND T. LANGE (2007): "Pensions, Education and Life Expectancy," *Working Paper of the DFG Research Group Heterogeneous Labor*, No. 07/04.

GRANGER, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–438.

GROSSMAN, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80, 223–255.

——— (2000): "The Human Capital Model," in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, New York and Oxford: Elsevier Science, North-Holland, vol. 1A, 347–408.

GROSSMAN, S. J. AND O. D. HART (1983): "An Analysis of the Principal-Agent Problem," *Econometrica*, 51, 7–45.

GRUBER, J. AND D. WISE (1998): "Social Security and Retirement: An International Comparison," *American Economic Review*, 88, 158–163.

GUSTMAN, A. L. AND T. L. STEINMEIER (2001): "Retirement Outcomes in the Health and Retirement Study," *Social Security Bulletin*, 64, 3–17.

——— (2005): "The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth," *Journal of Public Economics*, 89, 441–463.

HANEL, B. AND R. RIPHAHN (2006): "Financial Incentives and the Timing of Retirement: Evidence from Switzerland," *IZA Discussion Paper*, No. 2492.

HAUSMAN, J. A. (2003): "Triangular Structural Model Specification and Estimation with Application to Causality," *Journal of Econometrics*, 112, 107–113.

HAYWARD, M. D., W. R. GRADY, M. A. HARDY, AND D. SUMMERS (1989): "Occupational Influences on Retirement, Disability, and Death," *Demography*, 36, 393–409.

HECKMAN, J. (2003): "Conditioning, Causality and Policy Analysis," *Journal of Econometrics*, 112, 73–78.

HECKMAN, J. J. (1995): "What Has Been Learned About Labor Supply in the Past Twenty Years?" *American Economic Review*, 83, 116–121.

HENRIET, D. AND J.-C. ROCHET (2006): "Is Public Health Insurance an Appropriate Instrument for Redistribution?" *Annales d'Economie et de Statistique*, 61–88.

HOLMSTROM, B. AND P. MILGROM (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, 55, 303–328.

——— (1991): "Multitask Principle-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design," *Journal of Law, Economics and Organization*, 7, 24–51.

——— (1994): "The Firm as an Incentive System," *American Economic Review*, 84, 972–991.

HÄRDLE, W., M. MÜLLER, S. SPERLICH, AND A. WERWATZ (2004): *Nonparametric and Semiparametric Models*, Heidelberg: Springer.

HUMAN MORTALITY DATABASE (2005): *Life Expectancy at Birth*, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.humanmortality.de (data downloaded on Jan. 10, 2008).

HURD, M., J. SMITH, AND J. ZISSIMOPOULOS (2002): "The Effects of Subjective Survival on Retirement and Social Security Claiming," *NBER Working Paper*, No. 9140.

HURD, M. D. AND J. B. SHOVEN (1986): "The Distributional Impact of Social Security," *NBER Working Paper*, No. 1155.

JEWITT, I. (1988): "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, 56, 1177–1190.

JOHANSSON, E. (2004): "A Note on the Impact of Hours Worked on Mortality in OECD Countries," *European Journal of Health Economics*, 49, 335–340.

JONES, A. M. A. (2006): "Health and Wealth: Empirical Findings and Political Consequences," *Perspektiven der Wirtschaftspolitik*, 7, 93–112.

KALEMLI-OZCAN, S. AND D. N. WEIL (2002): "Mortality Change, the Uncertainty Effect, and Retirement," *NBER Working Paper*, No. 8742.

KIFMANN, M. (2008): "Age-Dependent Taxation and the Optimal Retirement benefit Formula," *German Economic Review (forthcoming)*.

KIUILA, O. AND P. MIESZKOWSKI (2007): "The Effects of Income, Education and Age on Health," *Health Economics*, 16, 781–798.

KOSTORIS PADOA SCHIOPPA, F. (1990): "Undesirable Redistributions in the Retirement-Public Pension Schemes: The Italian Case Study," *CEPR Discussion Paper*, No. 463.

LAU, M. I. AND P. POUTVAARA (2006): "Social Security Incentives and Human Capital Investment," *Finnish Economic Papers*, 19, 16–24.

LAUTERBACH, K., M. LÜNGEN, B. STOLLENWERK, A. GERBER, AND G. KLEVER-DEICHERT (2006): "Zum Zusammenhang zwischen Einkommen und Lebenserwartung," *Studien zu Gesundheit, Medizin und Gesellschaft, Köln*, No. 01.

LI, Q. AND J. S. RACINE (2007): *Nonparametric Econometrics: Theory and Practice*, Princeton: Princeton University Press.

LINDAHL, M. (2005): "Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as an Exogenous Source of Variation in Income," *Journal of Human Resources*, 40, 144–168.

LINDBECK, A. AND M. PERSSON (2003): "The Gains from Pension Reform," *Journal of Economic Literature*, 41, 74–112.

LINDEBOOM, M. (2006): "Health and Work of Older Workers," in *The Elgar Companion to Health Economics*, ed. by A. M. Jones, Cheltenham, U.K. and Northampton, Mass., 26–35.

LIU, L. AND A. J. RETTENMAIER (2003): "Social Security Outcomes by Racial and Education Groups," *Southern Economic Journal*, 69, 842–864.

LOADER, C. (1999): *Local Regression and Likelihood*, New York: Springer.

——— (2004): "Smoothing: Local Regression Techniques," in *Handbook of Computational Statistics*, ed. by J. Gentle, W. Härdle, and Y. Mori, Springer.

LUSARDI, A. AND O. S. MITCHELL (2007): "Financial Literacy and Retirement Preparedness: Evidence and Implications for Financial Education," *Business Economics*, 42, 35–44.

LUY, M. (2006): "Differentielle Sterblichkeit: die ungleiche Verteilung der Lebenserwartung in Deutschland," *Rostock Center Discussion Paper*, No. 6.

MARTIKAINEN, P., T. MARTELIN, E. NIHTILA, K. MAJAMAA, AND S. KOSKINEN (2005): "Differences in Mortality by Marital Status in Finland from 1976 to 2000:

Analyses of Changes in Marital-Status Distributions, Socio-demographic and Household Composition, and Cause of Death," *Population Studies*, 59, 99–115.

MEALLI, F. AND D. B. RUBIN (2003): "Assumptions Allowing the Estimation of Direct Causal Effects," *Journal of Econometrics*, 112, 79–87.

MEER, J., D. L. MILLER, AND H. S. ROSEN (2003): "Exploring the Health-Wealth Nexus," *Journal of Health Economics*, 22, 713–730.

MEGHIR, C. AND E. WHITEHOUSE (1997): "Labour Market Transitions and Retirement of Men in the UK," *Journal of Econometrics*, 79, 327–354.

MENCHIK, P. L. (1993): "Economic Status as a Determinant of Mortality Among Black and White Older Men: Does Poverty Kill?" *Population Studies*, 47, 427–436.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, 38, 175–208.

MITTELHAMMER, R. C., G. G. JUDGE, AND D. J. MILLER (2000): *Econometric Foundations*, Cambridge: Cambridge University Press.

MOORE, D. E. AND M. D. HAYWARD (1990): "Occupational Careers and Mortality of Elderly Men," *Demography*, 27, 31–53.

MULLIGAN, C. B. AND X. SALA–I–MARTIN (2004): "Internationally Common Features of Public Old-Age Pensions, and their Implications for Models of the Public Sector," *Advances in Economic Analyses & Policy*, 4.

MUURINEN, J.-M. (1982): "Demand for Health," *Journal of Health Economics*, 1, 5–28.

MYLES, G. D. (2002): *Public Economics*, Cambridge: Cambridge University Press.

NELISSEN, J. H. M. (1995): "Lifetime Income Redistribution by the Old-Age State Pension in the Netherlands," *Journal of Public Economics*, 58, 429–451.

——— (1998): "Annual versus Lifetime Income Redistribution by Social Security," *Journal of Public Economics*, 68, 223–249.

OHSMANN, S., U. STOLZ, AND R. THIEDE (2004): "Rentenabschläge bei vorzeitigem Rentenbeginn: Was ist versicherungsmathematisch fair?" *Sozialer Fortschritt*, 53, 267–271.

PAGAN, A. AND A. ULLAH (1999): *Nonparametric Econometrics*, Cambridge: Cambridge University Press.

PHILIPSON, T. J. AND G. S. BECKER (1998): "Old-Age Longevity and Mortality-Contingent Claims," *Journal of Political Economy*, 106, 551–573.

PORTRAIT, F., M. LINDEBOOM, AND D. DEEG (2001): "Life Expectancies in Specific Health States: Results from a Joint Model of Health Status and Mortality of Older Persons," *Demography*, 38, 525–536.

POTERBA, J. M. (2003): "Some Observations on Health Status and Economic Status," *Journal of Econometrics*, 112, 65–67.

QUEISSER, M. AND E. WHITEHOUSE (2006): "Neutral or Fair? Actuarial Concepts and Pension-System Design," *OECD Social, Employment and Migration Working Papers*, No. 40.

REIL-HELD, A. (2000): "Einkommen und Sterblichkeit in Deutschland: Leben Reiche länger?" *Beiträge zur angewandten Wirtschaftsforschung*, 580-00.

RIBEIRO, T. (2002): "Discrete Dynamic Models: A HMM Approach to Estimation and Forecasting Using Panel Survey Sata," *UC Berkeley Working Paper*.

RIED, M. (1998): "Comparative Dynamic Analysis of the Full Grossman Model," *Journal of Health Economics*, 17, 383–426.

ROBINS, J. M. (2003): "General Methodological Considerations," *Journal of Econometrics*, 112, 89–106.

ROBINSON, P. M. (1988): "Semiparametric Econometrics: A Survey," *Journal of Applied Econometrics*, 3, 35–51.

ROGERSON, W. P. (1985): "The First-Order Approach to Prinicpial-Agent Problems," *Econometrica*, 53, 1357–1368.

RUHM, C. J. (2000): "Are Recessions Good for Your Health?" *Quarterly Journal of Economics*, 115, 617–650.

——— (2007): "A Healthy Economy Can Break Your Heart," *Demography*, 44, 829–848.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229.

SALANIÉ, B. (1997): *The Economics of Contracts—A Primer*, Cambridge: MIT Press.

SALM, M. (2007): "The Effect of Penions on Longevity: Evidence from Union Army Veterans," *IZA Discussion Paper*, No. 2668.

SCHILS, T. (2005): *Early Retirement Patterns in Europe: A comparative Panel Study*, Amsterdam: Dutch University Press.

SELDEN, T. (1993): "Uncertainty and Health Care Spending by the Poor: The Health Capital Model Revisited," *Journal of Health Economics*, 12, 109–115.

SHESHINSKI, E. (1978): "A Model of Social Security and Retirement Decisions," *Journal of Public Economics*, 10, 337–360.

——— (2003): "Optimum Delayed Retirement Credit," *CESifo Working Paper*, No. 889.

SHESHINSKI, E. AND Y. WEISS (1981): "Uncertainty and Optimal Social Security Systems," *Quarterly Journal of Economics*, 96, 189–206.

SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

SIMONOVITS, A. (2004): "Designing Benefit Rules for Flexible Retirement with or without Redistribution," *CESifo Working Paper*, No. 1370.

——— (2006): "Optimal Design of Pension Rule with Flexible Retirement: The Two-Type Case," *Journal of Economics (Zeitschrift fur Nationalokonomie)*, 89, 197–222.

SIMONOVITS, A. AND P. ESÖ (2003): "Designing Optimal Benefit rules for Flexible Retirement," *Journal of Pension Economics and Finance*, 2, 273–293.

SMITH, J. P. (1998): "Socioeconomic Status and Health," *American Economic Review*, 88, 192–196.

SOARES, R. R. (2007): "On the Determinants of Mortality Reductions in the Developing World," *Population and Development Review*, 33, 247–287.

SPECKMAN, P. (1988): "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society*, Series B, 50, 413–436.

STAHLBERG, A.-C. (1989): "Redistribution Effects of Social Policy in a Lifetime Analytical Framework," in *The Political Economy of Social Security*, ed. by B. A. Gustafsson and N. A. Klevmarken, Contributions to Economic Analysis, no. 179, 51–65.

STATISTISCHES BUNDESAMT (2004): *Periodensterbetafeln für Deutschland: Allgemeine und abgekürzte Sterbetafeln von 1871/1881 bis 2001/2003*, Wiesbaden: Statistisches Bundesamt.

——— (2007): *Statistisches Jahrbuch 2007 für die Bundesrepublik Deutschland*, Wiesbaden: Statistisches Bundesamt.

TURRA, C. M. AND O. S. MITCHELL (2004): "The Impact of Health Status and Out–of–Pocket Medical Expenditures on Annuity Valuation," *PARC Working Paper Series*, No. 04–02.

VAN SOEST, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach," *Journal of Human Resources*, 30, 63–88.

VON GAUDECKER, H.-M. AND R. D. SCHOLZ (2007): "Differential Mortality by Lifetime Earnings in Germany," *Demographic Research*, 17, 83–108.

WALDRON, H. (2001): "Links between Early Retirement and Mortality," *ORES Working Paper (U.S. Social Security Administration)*, No. 93.

WILKINSON, R. (1996): *Unhealthy Societies: The Afflictions of Inequality*, London: Routledge.

WOLFE, J. R. (1983): "Perceived Longevity and Early Retirement," *Review of Economics and Statistics*, 65, 544–551.

——— (1985): "A Model of Declining Health and Retirement," *Journal of Political Economy*, 93, 1258–1267.

WORLD BANK (1994): *Averting the Old Age Crisis—Policies to Protect the Old and Promote Growth*, New York: Oxford University Press.

WREDE, M. (1999): "Pareto Efficient Pay–As–You–Go Pension Systems With Multiperiod Lives," *Jahrbücher für Nationalökonomie und Statistik*, 219, 494–503.

YAARI, M. E. (1965): "Uncertain Lifetime, Life Insurance, and the Theory of the Consumer," *Review of Economic Studies*, 32, 137–150.

YANG, L. AND R. TSCHERNIG (1999): "Multivariate Bandwidth Selection for Local Linear Regression," *Journal of the Royal Statistical Society. Series B*, 61, 793–815.

YATCHEW, A. (1998): "Nonparametric Regression Techniques in Economics," *Journal of Economic Literature*, 36, 669–721.

——— (2003): *Semiparametric Regression for the Applied Econometrician*, Cambridge: Cambridge University Press.

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema

Heterogeneous Ability, Life Expectancy, and Social Security: Four Essays

ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, insbesondere Promotionsberater, waren an der inhaltlich materiellen Erstellung dieser Arbeit nicht beteiligt.[8] Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Konstanz, den 29. Februar 2008

(Stefan Hupfeld)

[8]Siehe hierzu die Abgrenzung auf der folgenden Seite.

Abgrenzung

Ich versichere hiermit, dass ich Kapitel 1 bis 4 der vorliegenden Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Kapitel 5 (*On the Fairness of Early Retirement Provisions*) entstammt einer gemeinsamen Arbeit mit Herrn Prof. Dr. Friedrich Breyer (Universität Konstanz). Die individuelle Leistung im Rahmen dieser Arbeit gliedert sich wie folgt:

i. Die Einleitung, die Gegenüberstellung verschiedener Fairness–Konzepte und die Definition von *Distributive Neutrality* (Abschnitte 5.1 bis 5.3), sowie die Schlussbetrachtung (Abschnitt 5.5) stammen zum überwiegenden Teil von Prof. Dr. Friedrich Breyer.

ii. Der Abschnitt 5.4 *Distributive Neutrality and Early–Retirement Discounts in the German Pension System* stammt zum überwiegenden Teil von Stefan Hupfeld. Darunter fallen zum Teil die theoretischen Berechnungen, sowie zum überwiegenden Teil die ökonometrischen Abschnitte und die formale und grafische Präsentation und Interpretation der Ergebnisse.

Konstanz, den 29. Februar 2008 ______________________________

(Stefan Hupfeld)